universität wien

MASTERARBEIT

Titel der Masterarbeit

The Asymptotic Behaviour of Denumerable Markov Chains

Verfasser

Manfred Buchacher, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, im März 2015

Studienkennzahl lt. Studienblatt:	A 066 821
Studienrichtung lt. Studienblatt:	Masterstudium Mathematik
Betreuer:	Univ. Prof. Dr. Jiří Černý

THE ASYMPTOTIC BEHAVIOUR OF DENUMERABLE MARKOV CHAINS

CONTENTS

INTRODUCTION

The aim of this text is to survey some concepts for the description of the asymptotic behaviour of denumerable Markov chains, and show how they are related to each other.

Basically there are two classes of Markov chains: Markov chains which are recurrent and Markov chains which are not, the latter ones being called transient. The class of recurrent Markov chains itself consists of two subclasses: those which are positive recurrent, and those which are null-recurrent. For irreducible, aperiodic and positive recurrent Markov chains there is a convergence result, which states the existence of a unique invariant probability measure the distributions of the Markov chain's random variables converge to with respect to the total variation norm. For Markov chains which are not positive recurrent an analogous result does not exist, since the corresponding distributions converge pointwise to zero. See [1] for details.

In this text we present three different approaches in describing a Markov chain's asymptotic behaviour, mainly interesting in the context of irreducible and transient Markov chains:

One approach in doing so is the investigation of its tail- and invariant σ-algebra: both of them describe the possible long-term behaviour of the Markov chain. For recurrent Markov chains the invariant σ-algebra turns out to be trivial; so this approach, at least in what concerns the invariant σ-algebra, is interesting in the setting of transient Markov chains only. See sections 5, 6 and 2.

In a more analytic approach the transition kernel of a Markov chain is considered as an operator which acts on functions of the state space. Functions which are invariant under this action are called harmonic, and the space of bounded harmonic functions turns out to contain much information about the Markov chain's asymptotic behaviour. Indeed, there is a canonical correspondence between bounded harmonic functions and equivalence classes of bounded functions which are measurable with respect to the invariant σ-algebra. Again, in case the Markov chain is recurrent, this approach is not very fruitful either, since in this case every bounded harmonic function is constant. See section 5 and section 2.

The last, more geometric approach we present in this text is given in terms of boundary theory. Here, the idea is to construct a compactification of the state space, the points which are added to the state space being interpreted as those points the Markov chain possibly can converge to. These points are referred to as the boundary of the chain. For an irreducible and transient Markov chain we will establish the existence of a random variable which takes values in the boundary only, and show almost sure convergence of the Markov chain to this random variable in the topology of the compactification. In case the Markov chain is recurrent, the construction, which is presented here, turns out to be not very helpful for the understanding of its

asymptotic behaviour, since in this case its boundary consists of a single point only. In [2] another notion of compactification is introduced, adjusted to the setting of recurrent Markov chains. Again, there is a very close relationship between boundary theory and potential theory: every non-negative super-harmonic function can be represented as the integral of a function - the Martin kernel of the chain - with respect to some Borel-measure on the compactification of the state space. We will show under which condition the corresponding Borel-measure turns out to be unique and deduce a characterization of the triviality of the boundary in terms of the non-existence of non-constant bounded harmonic functions. See section 7.

Whatever approach one follows in investigating the asymptotic behaviour of a Markov chain, the first question always is the question of how to decide its triviality or non-triviality. In general a Markov chain's invariant σ-algebra is trivial if and only if the corresponding space of bounded harmonic functions is trivial, which is the case if and only if its boundary is trivial, and proving every bounded harmonic function to be constant can be a quite simple task.

In the setting of random walks on discrete groups driven by a symmetric probability measure, we will show that the corresponding invariant σ-algebra and tail-σ-algebra are equal. See section 6.

Furthermore, we introduce the notions of drift and asymptotic entropy. While the drift of a random walk on a group can be considered as a measure for how much it diverges from its starting point on the long run, the asymptotic entropy of a random walk on a group driven by a probability measure quantifies how much it is spread out on the long run. In the setting of random walks on groups driven by a symmetric and finitely supported probability measure whose support generates an infinite group, we give a characterization of the triviality of its asymptotic behaviour in terms of these two notions: its asymptotic behaviour (invariant σ-algebra, tail-σ-algebra, space of bounded harmonic functions, boundary) is trivial if and only if its drift is equal to 0, which is the case if and only if its asymptotic entropy is equal to 0. See sections $2, 3, 4$ and 6.

1. MARKOV CHAINS

In this section we explain the meaning of the term Markov chain, introduce the notion of a random walk on a group, illustrate the usefulness of this concept using the example of shuffling cards, show what questions come up naturally, and how they can be answered. We follow the presentation in Laurent Saloff-Coste's survey article about random walks on finite groups, see [4] for more details.

A Markov or transition kernel on a countable set X is a function $P : X \times X \to [0,1]$ such that $\sum_y P(x,y) = 1$ for every $x \in X$. Given a Markov kernel P and a probability measure ν on X, we can construct a discrete-time stochastic process $X_0, X_1, X_2, \ldots$ taking values in X, whose law $\mathbb{P}_\nu$ on $X^{\mathbb{N}}$ is characterized by

$$\mathbb{P}_\nu(X_i = x_i,\, 0 \le i \le n) = \nu(x_0)P(x_0,x_1)\cdots P(x_{n-1},x_n).$$

For a detailed description of the construction see [5]. In this context X is referred to as the state space of the stochastic process. In case the initial probability measure ν is concentrated on $x \in X$, that is $\nu = \delta_x$, we write $\mathbb{P}_x$ instead of $\mathbb{P}_{\delta_x}$, for notational convenience. This stochastic process obviously is a Markov chain, meaning that

$$\mathbb{P}_\nu(X_{n+1} = x_{n+1} \,|\, X_i = x_i,\, 0 \le i \le n) = \mathbb{P}_\nu(X_{n+1} = x_{n+1} \,|\, X_n = x_n),$$

for every $n \in \mathbb{N}$ and all $x_0, \ldots, x_{n+1} \in X$, whenever the event $\{X_i = x_i,\, 0 \le i \le n\}$ has positive probability. It's a time-homogeneous Markov chain, meaning that, additionally to the above,

$$\mathbb{P}_\nu(X_{n+1} = y \,|\, X_n = x) = \mathbb{P}_\nu(X_{m+1} = y \,|\, X_m = x)$$

holds for all $x, y \in X$ and $n, m \in \mathbb{N}$, whenever the events $\{X_n = x\}$ and $\{X_m = x\}$ have positive probability.

It might be helpful having in mind the following picture when working with discrete-time, time-homogeneous Markov chains on countable state spaces: with a given Markov kernel $P : X \times X \to [0,1]$ one can associate a directed graph, the vertices of this graph being the elements of X, and an edge leading from a vertex x to a vertex y if and only if $P(x,y) > 0$. Assuming the initial distribution to be concentrated at some vertex x the corresponding Markov chain describes how a particle located at x at time 0 moves in time, jumping from one vertex to a neighbouring one, the jump only depending on its current position, being independent of its earlier positions.

It is natural to ask about the law of the random variable X_n for a given n, and how it is related to the starting distribution ν and the Markov kernel P.

It is not difficult to see that

$$\mathbb{P}_\nu(X_n = x) = \sum_y \nu(y)P^n(y,x),$$

where the iterated kernel P^n is defined recursively by $P^0(x,y) = \delta_x(y)$ and $P^n(x,y) = \sum_z P^{n-1}(x,z)P(z,y)$. Obviously, considering ν as a row vector and P as a matrix indexed by the elements of X, P^n is just the n-th power of P and the law of X_n is νP^n, the initial distribution ν multiplied by P^n from the right. Considering P as an operator on a suitable space also leads to the question about the connections between the properties of this operator and the properties of the corresponding Markov chain, but more about that later.

Random walks on groups. Above we have seen how to equip the state space of a Markov chain with some additional structure - the structure of a directed graph - and viewed the Markov chain somehow adapted to that. One can go the other way around too: one starts with a set equipped with some geometric or algebraic structure and defines a Markov chain on that set, whose transition probabilities are adapted to that structure. In this case the Markov chain is called a random walk. In the following we are interested in random walks on groups which are driven by some probability measure. Here's the definition:

Consider the following specialization of the above construction: given a discrete group G and a probability measure μ on G, we replace X by G and define the transition kernel P by

$$P(x,y) = \mu(x^{-1}y).$$

For a fixed initial distribution ν the resulting Markov chain $X_0, X_1, X_2, \ldots$ is called the random walk on G driven by μ, moving from position X_n to position X_{n+1} by choosing an element ξ of G according to μ, independently from the elements chosen before, and multiplying X_n by ξ from the right, that is

$$X_{n+1} = X_n \cdot \xi.$$

Again it's not difficult to see that in this case the iterated kernel P^n is given by $P^n(x,y) = \mu^{(n)}(x^{-1}y)$, while the law of X_n is given by $\nu * \mu^{(n)}$. By $\mu^{(n)}$ we denote the n-fold convolution product of μ, the convolution product being defined by

$$f * g(x) = \sum_{z \in G} f(z)g(z^{-1}x)$$

for arbitrary functions $f, g : G \to \mathbb{R}$.

A very popular application of the concept of a random walk on a group are card shuffling models. Now, how can this be used to model card shuffling? Both an arrangement of a deck of cards and its shuffling can be

considered as a permutation, the difference between those two notions being that an arrangement relates face values to positions, while a shuffling is a permutation of positions. Assuming we have a deck of 52 cards, which are marked from 1 to 52, an arrangement of the deck can be described by a 52-tuple giving the face values of the cards in order from top to bottom. Thus we can identify the arrangement $(\sigma_1, ..., \sigma_{52})$ of the deck with the permutation $\sigma : i \mapsto \sigma(i) = \sigma_i$ in S_{52}. Conversely, given a permutation $\sigma \in S_{52}$ the corresponding arrangement of the deck has card i in position $\sigma^{-1}(i)$, whereas $\sigma(i)$ gives the value of the card in position i. From the perspective of shuffling cards, we want permutations to act on positions, not on face values. In the present notation this corresponds to multiplication on the right in S_{52}.

Card shuffling proceeds by repeating several times a fixed procedure where some randomness occurs. This randomness is modelled by a probability measure μ. According to this measure an element θ of S_{52} is chosen, which transfers a given arrangement σ of the deck into the arrangement $\sigma \circ \theta$. So shuffling a deck of 52 cards can be described by a random walk on S_{52} driven by some measure μ.

Having been introduced to the general framework of shuffling cards, there are some questions one might pose immediately. For instance: Shuffling a deck of cards what does it mean for it to get well mixed? Does it ever get well mixed, under which assumptions does it? And how fast does it mix, if it does?

By $X_0, X_1, X_2, ...$ we denote the random walk on S_{52} driven by some probability measure μ with initial distribution concentrated at the identity. Let us consider the distribution of X_n, the arrangement of the deck after it has been shuffled n times. As already mentioned above it is given by $\mu^{(n)}$, the n-th convolution power of the probability measure μ. It is natural to look at

$$\| \mu^{(n)} - u \|_{TV} = \frac{1}{2} \sum_{\sigma \in S_{52}} |\mu^{(n)}(\sigma) - u(\sigma)|$$

as a measure of how well the deck is mixed, where u denotes uniform measure on S_{52} and $\| \cdot \|_{TV}$ denotes total variation norm. Having this at hand we can answer the first question: a deck of cards gets well mixed if and only if $\| \mu^{(n)} - u \|_{TV}$ tends to zero as n tends to infinity.

To partially answer the other two questions we come back to the more general setting of denumerable Markov chains. There are several quite important notions concerning Markov kernels we haven't spoken about yet: a Markov kernel P is said to be irreducible, if for any two states x, y there exists an integer n, which might depend on x and y, such that $P^n(x,y) > 0$. A state x is called aperiodic if $gcd\{n \in \mathbb{N} \mid P^n(x,x) > 0\} = 1$ or equivalently, if $P^n(x,x) > 0$ for all sufficiently large n, while P is called aperiodic if this

is true for every state $x \in X$, that is every state is aperiodic. Clearly, if P is irreducible and there is a state which is aperiodic, then also P is aperiodic.

A measure $\mathfrak{m}$ on X is invariant or stationary with respect to the Markov kernel P, if $\mathfrak{m}P = \mathfrak{m}$, and the Markov chain is called reversible if there is a positive measure $\mathfrak{m}$, such that

$$\mathfrak{m}(x)P(x,y) = \mathfrak{m}(y)P(y,x)$$

for all $x, y \in X$. Obviously, any reversible measure is an invariant measure too. The following theorem, whose proof can be found in [5], relates some of these notions:

1.1. **Theorem.** Let P be an irreducible and aperiodic Markov kernel on a finite set X. Then P admits a unique invariant measure $\mathfrak{m}$ on X. Furthermore there are $k \in \mathbf{N}$ and $\tau < 1$ such that

$$\sum_y |P^n(x,y) - \mathfrak{m}(y)| \leq 2\tau^n$$

for every $x \in X$ and $n \geq k$.

In case of a random walk on a finite group driven by a probability measure, uniform distribution is stationary. The above theorem gives a sufficient condition for the probability measures given by the iterated Markov kernels to converge to uniform distribution with respect to total variation norm and an upper bound for how fast this is done. Those readers who are interested in the investigation of specific models of shuffling cards may be referred to [4] and the references which can be found therein.

2. DRIFT

In this section we explain the notion of drift of a random walk on a group, investigate biased and unbiased random walk on $\mathbb{Z}$ in this respect, explain why the latter one's behaviour is typical in terms of recurrence of the random walk and commutativity of the underlying group, show how the notion of spectral radius relates to the notion of drift, and prove that simple random walk on the homogeneous tree $\mathbb{T}_n$, n being even, moves with non-trivial drift.

Given a random walk $X_0, X_1, X_2, ...$ on a group G driven by a probability measure μ, we define its drift $\sigma(\mu)$ as

$$\sigma(\mu) = \lim_{n\to\infty} \frac{\mathbb{E}_e[d(X_0, X_n)]}{n},$$

where the expectation is taken with respect to the probability measure $\mathbb{P}_e$, $e \in G$, on $G^{\mathbb{N}}$ and $d(X_0, X_n)$ denotes the distance between the position of the random walk at time 0 and its position at time n in the graph metric of the corresponding directed graph. Convergence of this series can be seen as follows. Although the sequence $\{\mathbb{E}[d(X_0, X_n)]\}$ is not subadditive, in which case we could have applied Fekete's lemma - see the appendix for that -, it is similar to being subadditive. Writing $d_n(X_0, X_1, ...)$ for the distance between the position of the random walk at time 0 and its position at time n, and introducing the shift operator $S : G^{\mathbb{N}} \to G^{\mathbb{N}}$ defined by $(x_0, x_1, x_2, ...) \mapsto (x_1, x_2, x_3, ...)$, we have the following inequality:

$$d_{n+m}(X_0, X_1, ...) \leq d_n(X_0, X_1, ...) + d_m(S^n(X_0, X_1, ...)).$$

This might just look like an elaborate formulation of the triangle inequality, but it shows that we can apply Kingman's subadditive ergodic theorem - see the appendix for its formulation and proof -, which implies almost sure convergence of $\{d_n(X_0, X_1, ...)/n\}$. Since the random variable $d_n(X_0, X_1, ...)/n$, that is $d(X_0, X_n)/n$, is almost surely bounded by 1, the theorem of dominated convergence implies that the notion of drift is well-defined.

2.1. **Example.** Consider biased random walk on $\mathbb{Z}$, the random walk on $\mathbb{Z}$ generated by the probability measure μ with $\mu(1) = p$ and $\mu(-1) = 1 - p$, where $p \neq 1/2$. Denoting by $\xi_1, \xi_2, ...$ a sequence of independent and μ-distributed random variables, the random walk's position at time n is given by $X_n = \sum_{k=1}^{n} \xi_k$. Since

$$\mathbf{E}[|X_n|] \geq |\mathbf{E}[X_n]| = n|2p - 1|,$$

we have $\sigma(\mu) \geq |2p - 1| > 0$, so biased random walk on $\mathbb{Z}$ moves with non-trivial drift.

Having investigated biased random walk on $\mathbb{Z}$, it is natural to ask about the unbiased one. Just as above intuition made us expect drift to be non-trivial, we expect drift to be trivial in this case, in the case of unbiased

random walk. Indeed, it is equal to 0 demonstrating that our notion of drift captures intuition quite good.

2.2. **Example.** Unbiased or simple random walk on $\mathbb{Z}$ is generated by the probability measure μ with $\mu(1) = 1/2 = \mu(-1)$. Again we write $X_n = \sum_{k=1}^{n} \xi_k$ for the position of the random walk at time n, where $\xi_1, \xi_2, \ldots$ are independent and μ-distributed random variables. Since the increments of the random walk are integrable and pairwise independent we can apply the strong law of large numbers to deduce that

$$\lim_{n\to\infty} \frac{|X_n|}{n} = 0 \quad \text{almost surely,}$$

hence, by the theorem of dominated convergence, simple random walk on $\mathbb{Z}$ moves without drift.

Recurrence. Unbiased random walk on $\mathbb{Z}$ is a typical example for a random walk which moves without drift, it is typical in at least two ways: Firstly, it is an example of a recurrent Markov chain; a Markov chain is called recurrent if every state is visited infinitely often almost surely. To see that unbiased random walk is recurrent we follow the presentation in [7]:

Let $x, y, z \in \mathbb{Z}$ and $0 \leq y \leq x \leq z$, without loss of generality. Let $p(y,x,z)$ denote the probability the random walk starting from x hits y before z, that is

$$p(y,x,z) = \mathbb{P}_x(\min\{j \mid X_j = y\} < \min\{j \mid X_j = z\}).$$

Conditioning on the first step leads to the following recursion

$$p(y,x,z) = \frac{1}{2}p(y,x-1,z) + \frac{1}{2}p(y,x+1,z), \quad y < x < z,$$

whose unique solution with respect to the boundary condition $p(y,y,z) = 1$ and $p(y,z,z) = 0$ is given by $p(y,x,z) = (z-x)/(z-y)$. Hence, for any $\varepsilon > 0$ there is a positive integer z_0, depending on ε, such that $p(y,x,z) = 1-(x-y)/(z-y) \geq 1-\varepsilon$ if $z \geq z_0$. This then implies that $\mathbb{P}_x(X_n = y, \text{ for some } n > 0) = 1$, which in turn implies that $\mathbb{P}_x(X_n = y, \text{ for infinitely many } n) = 1$. This can be seen in the following way: $\{X_n = y \text{ for at least } m \text{ time instants } n > 0\}$ is decreasing in m and approaches $\{X_n = y \text{ for infinitely many } n\}$ as m tends to infinity, so

$$\begin{aligned}\lim_{m\to\infty} &\mathbb{P}_x(X_n = y \text{ for at least } m \text{ time instants } n > 0) \\ &= \mathbb{P}_x(X_n = y \text{ for infinitely many } n).\end{aligned}$$

Furthermore it's not difficult to see that

$$\begin{aligned}&\mathbb{P}_x(X_n = y \text{ for at least } m+1 \text{ time instants } n > 0) \\ &= \mathbb{P}_x(X_n = y, \text{ for some } n > 0)\, \mathbb{P}_y(X_n = y \text{ for at least } m \text{ time instants } n > 0),\end{aligned}$$

via conditioning on the first positive hitting time of y and using the (strong) Markov property of the chain. In case $x = y$ this equation is a recursion in m, its solution being given by

$$\begin{aligned}&\mathbb{P}_x(X_n = x \text{ for at least } m \text{ time instants } n > 0)\\ &\quad = \mathbb{P}_x(X_n = x, \text{ for some } n > 0)^m.\end{aligned}$$

Letting m tend to infinity we see that simple random walk on $\mathbb{Z}$ is recurrent.

Intuitively we expect recurrent random walks to move without drift. In case the random walk is driven by a symmetric and finitely supported probability measure, whose support generates an infinite group, this is not difficult to see having at hand its characterization via bounded harmonic functions. We first explain the meaning of harmonicity in this context in the next definition.

2.3. **Definition.** Given a Markov chain (X,P), a function $f : X \to \mathbb{R}$ is called harmonic with respect to the Markov kernel P, if

$$Pf(x) := \sum_{y\in X} P(x,y)f(y) = f(x)$$

holds for every $x \in X$.

2.4. **Theorem.** A recurrent Markov chain $X_0, X_1, \ldots$ has no non-constant bounded harmonic functions.

Proof. If f is a bounded harmonic function, it's not difficult to see that $M_n = f(X_n)$ defines a bounded martingale. Indeed

$$\begin{aligned}&\mathbb{E}[M_{n+1} \mid M_n, \ldots, M_0]\\ &\quad = \sum_{x_i \in f^{-1}(M_i)} \mathbb{E}[f(X_{n+1}) \mid X_i = x_i \ \forall i]\, \mathbb{P}(X_i = x_i \ \forall i \mid M_n, \ldots, M_0)\\ &\quad = \sum_{x_i \in f^{-1}(M_i)} \mathbb{E}[f(X_{n+1}) \mid X_n = x_n]\, \mathbb{P}(X_i = x_i \ \forall i \mid M_n, \ldots, M_0)\\ &\quad = \sum_{x \in f^{-1}(M_n)} f(x_n)\, \mathbb{P}(X_i = x_i \ \forall i \mid M_n, \ldots, M_0) = M_n,\end{aligned}$$

where the second equation holds because of the Markov property and the third one because of f being a harmonic function. The martingale convergence theorem implies almost sure convergence of $\{f(X_n)\}$. Assuming f to be non-constant there are states x and y such that $f(x) \neq f(y)$ and which both are visited infinitely often almost surely by recurrence, which is a contradiction to almost sure convergence. □

The above theorem can not be used to deduce triviality of drift for simple random walk on $\mathbb{Z}^d$ for arbitrary d, since it is recurrent in dimensions $d = 1, 2$ only. But interestingly recurrence of simple random walk on $\mathbb{Z}$ can be used to show that every bounded harmonic function on $\mathbb{Z}^d$ is constant. The proof relies on a technique known as "coupling" and the following lemma.

2.5. **Lemma.** Given a Markov chain, if for all $x, y \in X$ there is a coupling (X_n, Y_n) of random walks starting from (x, y) such that $\mathbb{P}[X_n \neq Y_n]$ tends to 0 as n tends to infinity, then every bounded harmonic function is constant.

Proof. Assume f is a bounded harmonic function, bounded by some constant B. Since $f(X_n)$ and $f(Y_n)$ are martingales, we have $\mathbb{E}[f(X_n)] = f(x)$ and $\mathbb{E}[f(Y_n)] = f(y)$. So, for every n we have

$$\begin{aligned} |f(x) - f(y)| &= |\mathbb{E}[f(X_n)] - \mathbb{E}[f(Y_n)]| \\ &\leq \mathbb{E}[|f(X_n) - f(Y_n)] \\ &\leq 2B\,\mathbb{P}(X_n \neq Y_n). \end{aligned}$$

So $f(x) = f(y)$ and therefore f is constant. □

2.6. **Theorem.** Every bounded harmonic function on $\mathbb{Z}^d$ is constant, where harmonicity refers to the Markov operator associated to simple random walk on $\mathbb{Z}^d$.

Proof. We quickly sketch the proof. Let P denote the Markov kernel associated to simple random walk on $\mathbb{Z}^d$. A function f is harmonic with respect to P if and only if it is harmonic with respect to $\frac{I+P}{2}$, where I denotes the identity operator. The Markov chain which is associated to $\frac{I+P}{2}$ is referred to as lazy simple random walk on $\mathbb{Z}^d$. To apply the last lemma consider the following coupling: X_n and Y_n always move in the same coordinate, the coordinate being chosen at random according to equidistribution on the coordinates. If their distance in that coordinate is 0, it remains 0. If it is not, then when the first one moves, the second one does not, or vice versa. Each of the chains represents a lazy random walk, when considered independently. Having a look at a particular coordinate we see that, if the distance of X_n and Y_n in that coordinate is different from zero it is described by simple random walk on $\mathbb{Z}$, so with probability 1 this distance tends to 0 and remains 0 once it is attained. So with probability 1 there's an N such that $X_n = Y_n$ for all $n \geq N$. □

Destilling the essence of the above proof we arive at the following statement.

2.7. **Theorem.** Assume $X_0, X_1, X_2, \ldots$ is a Markov chain on a discrete metric space (X, d), the metric d taking values in $\mathbb{Z}_+$ only, and $d(X_n, X_{n+1})$ being equal to 1 almost surely for every n. Furthermore, 0 is a recurrent state for the stochastic process $\{d(X_n, X_{n+1})\}$ on $\mathbb{Z}_+$. Then every bounded harmonic function on X^k - the k-fold direct product - is constant, where harmonicity refers to the Markov operator associated to the Markov chain on X^k defined by flipping a k-sided coin to determine which coordinate to move in and then moving according to the Markov chain on X.

Algebraic imply stochastic properties. As promised above, here is the second reason why simple random walk on $\mathbb{Z}$ was typical in our context. We present a theorem by Choquet and Deny, following the proof of Albert Raugi - see [8] - adapted by Gabor Pete - see [3] - to our needs. It gives the definite answer concerning drift for random walks on abelian groups generated by symmetric and finitely supported probability measures, and is an example of a theorem which relates stochastic properties of a random walk and algebraic properties of the underlying group.

2.8. **Theorem.** Let μ be a probability measure on a discrete abelian group G whose support generates G. Then every bounded function which is harmonic with respect to the Markov chain driven by μ is constant.

Proof. Let $h : G \to \mathbb{R}$ be a bounded harmonic function and define $u_n(x) = \mathbb{E}_x[(h(X_n) - h(X_{n-1}))^2]$ for $n \geq 1$. If we denote the increments of the random walk generated by μ by $\xi_1, \xi_2, ...$ such that its position X_n at time n is given by $X_n = X_0 + \xi_1 + ... + \xi_n$, we can write

$$u_n(x) = \sum_{t_1,...,t_n \in G} (h(x+t_1+...+t_n) - h(x+t_1+...+t_{n-1}))^2 \mu(t_1) \cdot ... \cdot \mu(t_n).$$

For $n \geq 2$ we have

$$\begin{aligned} u_n(x) &= \mathbb{E}[\mathbb{E}_x[(h(X_n) - h(X_{n-1}))^2 | \xi_2, ...\xi_n]] \\ &\geq \mathbb{E}[(\mathbb{E}_x[h(X_n) - h(X_{n-1}) | \xi_2, ..., \xi_n])^2] \\ &= \mathbb{E}[(h(x+\xi_2+...+\xi_n) - h(x+\xi_2+...+\xi_{n-1}))^2] \\ &= \mathbb{E}_x[(h(X_{n-1}) - h(X_{n-2})))^2] = u_{n-1}(x), \end{aligned}$$

where $\mathbb{E}$ denotes the expectation with respect to the law of $(\xi_2, ..., \xi_n)$. The inequality in the second line follows from Jensen's inequality for conditional expectations and the equality in the third line follows from harmonicity of h and commutativity of G.

Since $u_n(x) = \mathbb{E}_x[h(X_n)^2] - \mathbb{E}_x[h(X_{n-1})^2]$ by the orthogonality of martingale increments, we have $\sum_{n=1}^{N} u_n(x) = \mathbb{E}_x[h(X_N)^2] - h(x)^2$. By the above inequality this is a sum of non-decreasing non-negative terms that remains bounded as N tends to infinity, hence all terms must be identically 0. Having a look at the specific form of u_n for $n = 1$ this implies that $h(x) = h(x+y)$ for all x in G and all y which lie in the support of μ. Since its support was supposed to generate all of G the function h has to be constant. □

Spectral radius and drift. The following subsection relates the notions of spectral radius and drift, see [6] for a more detailed presentation.

Let (X, P) be an irreducible Markov chain. For $x, y \in X$ and $z \in \mathbb{C}$ we define Green's function $G(x, y|z)$ by

$$G(x, y|z) = \sum_{n=0}^{\infty} p^{(n)}(x, y) z^n,$$

$p^{(n)}(x,y)$ denoting the probability $\mathbb{P}_x(X_n = y)$. Note that for $z = 1$ the value $G(x,y|1)$ is the expected number of visits of y, given the Markov chain starts at x. The spectral radius $\rho(P)$ of the chain is then defined as the reciprocal of the radius of convergence of this power series:

$$\rho(P) = \limsup_{n\to\infty} p^{(n)}(x,y)^{1/n}.$$

By irreducibility this is well defined, i.e. independent of $x,y \in X$. Indeed, let $x_1,x_2,y_1,y_2 \in X$ and $k,l \in \mathbb{N}$ such that $p^{(k)}(x_1,x_2) > 0$ and $p^{(l)}(y_2,y_1) > 0$. Then

$$p^{(k+n+l)}(x_1,y_1) \geq p^{(k)}(x_1,x_2)p^{(n)}(x_2,y_2)p^{(l)}(y_2,y_1),$$

and hence, for $z > 0$,

$$G(x_1,y_1|z) \geq p^{(k)}(x_1,x_2)p^{(l)}(y_2,y_1)z^{k+l}G(x_2,y_2|z).$$

So we see, if $G(x_1,y_1|z)$ converges, then so does $G(x_2,y_2|z)$. But since x_1,x_2,y_1,y_2 were chosen arbitrarily, this is true the other way around too. So their radii of convergence have to be equal.

The period $d = d(P)$ of an irreducible Markov chain (X,P) is defined as $d = gcd\{n \in \mathbb{N} \mid p^{(n)}(x,x) > 0\}$, where x is an arbitrary element of X. Again, by irreducibility this definition is independent of x. Obviously, for $p^{(n)}(x,x) > 0$, n necessarily has to be of the form $n = md$ for some $m \in \mathbb{N}$. In general n being divisible by d is not sufficient for $p^{(n)}(x,x) > 0$, but there is a constant m_x such that $p^{(md)}(x,x) > 0$ for all $m > m_x$. Furthermore, with respect to P^d the state space decomposes into d irreducible, aperiodic classes $C_0,\ldots,C_{d-1}$ which are visited in cyclical order by the originial Markov chain. So in many cases it's no restriction to assume the Markov chain to be aperiodic. For details and proofs see, for example, [5].

The first goal of this subsection is the following theorem, which in particular shows how the return probabilities $p^{(n)}(x,x)$ decrease as n increases in terms of the spectral radius.

2.9. **Theorem.** Let (X,P) be an irreducible Markov chain with period d and $x \in X$. Then

$$\lim_{n\to\infty} p^{(nd)}(x,x)^{1/(nd)} = \rho(P) \quad \text{and} \quad p^{(n)}(x,x) \leq \rho(P)^n \quad \text{for all } n \in \mathbb{N}.$$

We first proof the following Proposition.

2.10. **Proposition.** For a sequence (a_n) of non-negative real numbers satisfying $a_n > 0$ for all $n \geq n_0$ for some n_0 and $a_m a_n \leq a_{m+n}$ for all $m,n \in \mathbb{N}$, there is some $\rho > 0$ such that

$$\lim_{n\to\infty} a_n^{1/n} = \rho \quad \text{and} \quad a_n \leq \rho^n \quad \text{for all } n \in \mathbb{N}.$$

Proof. It is sufficient to remark that $-\log a_n$ defines a subadditive sequence for which Fekete's lemma implies the existence of the limit of $-\log(a_n)/n$ and identifies it as the infimum of $-\log(a_n)/n$ over all n. □

We now apply this to prove the above theorem.

Proof. Following the above remarks concerning the period of a Markov chain we can apply the last proposition to the sequence $(a_n) = (p^{(nd)}(x,x)^{1/d})$ to conclude that

$$\lim_{n\to\infty} p^{(nd)}(x,x)^{1/(nd)} = \rho(P) \quad \text{and} \quad p^{(nd)}(x,x) \leq \rho(P)^{nd}$$

Since $p^{(n)}(x,x) = 0$ if n is not divisible by d, we have $p^{(n)}(x,x) \leq \rho(P)^n$ for every $n \in \mathbb{N}$. □

Next we extend the inequality $p^{(n)}(x,x) \leq \rho(P)^n$ under the more restrictive assumption of uniform irreducibility and show how the spectral radius $\rho(P)$ relates to the drift of the Markov chain.

2.11. **Definition.** An irreducible Markov chain (X,P) is called uniformly irreducible if there are $\varepsilon > 0$ and $K < \infty$ such that $p(x,y) > 0$ implies $p^{(k)}(x,y) \geq \varepsilon$ for some $k \leq K$.

2.12. **Lemma.** Let (X,P) be a uniformly irreducible Markov chain. Then there is a constant $A > 0$ sucht that $p^{(n)}(x,y) \leq A^{d(x,y)}\rho(P)^n$ for every $n \in \mathbb{N}$.

Proof. Let $x,y \in X$ and $r = d(x,y)$, and let ε be as in the definition of uniform irreducibility. Applying the assumption of uniform irreducibility to the successive neighbours of a path of lenght r from x to y, we know that there are $k_1,\ldots,k_r \geq 1$ such that $p^{(k_1+\ldots+k_r)}(y,x) \geq \varepsilon^r$. Defining $k = k_1 + \ldots + k_r$, and using the last theorem and the fact that $\rho(P) \leq 1$ we have

$$p^{(n)}(x,y)\varepsilon^r \leq p^{(n)}(x,y)p^{(k)}(y,x) \leq p^{(n+k)}(x,x) \leq \rho(P)^{n+k} \leq \rho(P)^{n+r},$$

which shows that the statement is true for $A = \rho(P)/\varepsilon$. □

2.13. **Theorem.** Assuming (X,P) to be a uniformly irreducible Markov chain, whose spectral radius $\rho(P)$ is smaller than 1, there is a lower bound $b > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} d(X_0,X_n) \geq b \quad \mathbb{P}_x\text{-almost-surely for every } x \in X.$$

Proof. By uniform irreducibility of the Markov chain, the associated graph not only is locally finite, but even of bounded geometry. Let M be an upper bound for the vertex degrees, and let A be as in the above lemma. Without loss of generality we may assume $A > 1$; choose $b > 0$ such that

$(MA)^b\rho(P) < 1$. We have the following estimate:

$$\mathbb{P}_x(\inf_{k\geq n}\frac{1}{k}d(X_0,X_k) < b) \leq \sum_{k=n}^{\infty}\sum_{y:d(y,x)<bk} p^{(k)}(x,y)$$
$$\leq \sum_{k=n}^{\infty}\sum_{y:d(y,x)\leq bk} A^{d(x,y)}\rho(P)^k \leq c\left((MA)^b\rho(P)\right)^n,$$

for some $c > 0$, implying that $\mathbb{P}_x(\inf_{k\geq n}\frac{1}{k}d(X_0,X_k) < b)$ tends to 0 as n tends to infinity. We finally observe that

$$\mathbb{P}_x(\liminf_{n\to\infty}\frac{1}{k}d(X_0,X_k)\geq b) \geq \mathbb{P}_x(\inf_{k\geq n}\frac{1}{k}d(X_0,X_k)\geq b)$$
$$= 1-\mathbb{P}_x(\inf_{k\geq n}\frac{1}{k}d(X_0,X_k) < b)$$

holds for every n, so $\mathbb{P}_x(\liminf_{n\to\infty}\frac{1}{k}d(X_0,X_k)\geq b)$ has to be equal to 1, and the proof has come to an end. □

Recall the case of random walks on groups: we already know almost sure convergence of $d(X_0,X_n)/n$ in this situation. So what we have found above is a sufficient condition for non-triviality of drift.

In the following we denote the free group generated by n elements by $\mathbb{F}_n = \langle e_1,...,e_n\rangle$, and the probability measure on it given by equidistribution on its generators and its inverses by μ. The next goal of this text is to investigate the random walk driven by μ with respect to triviality or non-triviality of its drift. Since this random walk is nothing else but simple random walk on the homogeneous tree $\mathbb{T}_{2n}$, we can consider simple random walk on $\mathbb{T}_n$ for arbitrary n without more ado. Recall the definition of the homogeneous tree $\mathbb{T}_n$ as the tree whose vertices have degree n. Of course we want to apply the theorem we have proved last. Obviously simple random walk on $\mathbb{T}_n$ is uniformly irreducible, so in what is left we have to show that its spectral radius is smaller than 1. Before we compute the corresponding Green function we collect some useful things in the next two lemmata whose proofs are left to the reader, or can be found in [6].

2.14. **Lemma.** Given a Markov chain (X,P) we define stopping times

$$\mathbf{s}^y = \min\{n\geq 0\,|\,X_n = y\} \quad \text{and} \quad \mathbf{t}^y = \min\{n > 0\,|\,X_n = y\} \quad \text{for } y\in X,$$

and functions - the generating functions of the corresponding hitting probabilities -

$$F(x,y|z) = \sum_{n=0}^{\infty}\mathbb{P}_x[\mathbf{s}^y = n]z^n \quad \text{and} \quad U(x,y|z) = \sum_{n=0}^{\infty}\mathbb{P}_x[\mathbf{t}^y = n]z^n \quad \text{for } x,y\in X.$$

They satisfy the following identities:

(1) $G(x,x|z) = \frac{1}{1-U(x,x|z)}$,

(2) $G(x,y|z) = F(x,y|z)G(y,y|z)$,

(3) $U(x,x|z) = \sum_y p(x,y)zF(y,x|z)$ and,
(4) if $x \neq y$, $F(x,y|z) = \sum_w p(x,w)zF(w,y|z)$.

2.15. **Lemma.** Given a nearest neighbour random walk on a tree, for x and y being two vertices and w a vertex on the unique path of length $d(x,y)$ from x to y, the following holds:

$$F(x,y|z) = F(x,w|z)F(w,y|z).$$

2.16. **Proposition.** For simple random walk on $\mathbb{T}_n$ the Green function is given by

$$G(x,y|z) = \frac{2(n-1)}{n-2+\sqrt{n^2-4(n-1)z^2}} \left(\frac{n-\sqrt{n^2-4(n-1)}}{2(n-1)z} \right)^{d(x,y)}.$$

Proof. For fixed z the value of $F(x,y|z)$ does only depend on the distance $d(x,y)$ between x and y, by symmetry of the tree. Write $F(z) = F(x,y|z)$ for two neighouring vertices x and y. Then, by the last lemma,

$$F(x,y|z) = F(z)^{d(x,y)}$$

for arbitrary x and y. Furthermore, by Lemma 2.13(4), we have the following second order equation for $F(z)$

$$F(z) = F(x,y|z) = \sum_{w\sim x} \frac{1}{n} zF(z)^{d(y,w)} = \frac{1}{n}z + \frac{n-1}{n} zF(z)^2,$$

whose - in this context correct - solution is given by

$$F(z) = \frac{1}{2(n-1)z} \left(n - \sqrt{n^2-4(n-1)z^2} \right),$$

since F is continuous in z and $F(0) = 0$.

Using Lemma 2.13(3), (1) and (2) one can first compute $U(x,x|z)$ and $G(x,x|z)$, and subsequently $G(x,y|z)$. □

2.17. **Corollary.** The spectral radius $\rho(P)$ of simple random walk on $\mathbb{T}_n$ is given by $\rho(P) = \frac{2\sqrt{n-1}}{n}$.

Proof. Green's function $G(x,y|z)$ is a power series with non-negative real coefficients; following Pringsheim's theorem it is sufficient to compute its smallest positive singularity to determine its radius of convergence; having a look at the above representation of $G(x,y|z)$, this is that $z > 0$ for which $\sqrt{n^2-4(n-1)z^2}$ is zero, that is $z = \frac{n}{2\sqrt{n-1}}$. □

2.18. **Corollary.** The random walk on $\mathbb{F}_n$ we have defined above moves with non-trivial drift for $n \geq 2$.

Concerning the case $\mathbb{F}_n$ for $n = 1$, we only mention that we have already considered this one, since $\mathbb{F}_1$ is isomorphic to $\mathbb{Z}$.

3. ENTROPY

We explain the notion of the entropy of a random variable / measure, give an upper bound in terms of its support, introduce the notion of asymptotic entropy for a random walk on a group driven by a probability measure, and compute the asymptotic entropy of biased and unbiased random walk on $\mathbb{Z}$.

We now introduce one of the central notions of ergodic and information theory: for a random variable X distributed according to some probability measure μ we define its entropy by

$$H(X) = -\sum_x \mu(x) \log \mu(x),$$

which we also denote by $H(\mu)$. Note that in the above definition we use the convention that $0 \log 0 = 0$. In the context of information theory logarithm is usually taken to base 2, in which case entropy is expressed in bits and can be interpreted as the average number of bits required to describe the random variable, see for example [9] or [10]. In the context of a random walk $X_0, X_1, X_2, \ldots$ the entropy $H(X_n)$ can therefore be regarded as a measure of how much the walk is spread out.

A useful upper bound for the entropy of a probability measure can be derived by using Jensen's inequality: let μ be an arbitrary but finitely supported probability measure and u the probability measure which is uniform on the support of μ, that is $u(x)$ equals $1/|\operatorname{supp} \mu|$ whenever x lies in the support of μ and 0 if this is not the case. Then

$$\sum_x \mu(x) \log \frac{u(x)}{\mu(x)} \leq \log \sum_x u(x) = \log 1 = 0.$$

Since the left side of this inequality is $H(\mu) - \log |\operatorname{supp} \mu|$ we have just found $\log |\operatorname{supp} \mu|$ to be an upper bound for $H(\mu)$.

Given a random walk $X_0, X_1, \ldots$ on a group, driven by some probability measure μ, we will be interested in its asymptotic entropy $h(\mu)$ defined by

$$h(\mu) = \lim_{n \to \infty} H(\mu^{(n)})/n = \lim_{n \to \infty} H(X_n)/n,$$

The limit exists since the sequence $H(\mu^{(n)})$ is subadditive. To see this assume we have two probability measures μ and ν. Then the entropy $H(\mu * \nu)$ of their convolution can be bounded from above by the entropy $H(\mu \times \nu)$ of the corresponding product measure, which is obviously equal to $H(\mu) + H(\nu)$. The upper bound holds because changing a probability measure by concentrating mass from different points on a single point doesn't increase its entropy. So indeed $H(\mu^{(n+m)}) \leq H(\mu^{(n)}) + H(\mu^{(m)})$ for all n and m and we can apply Fekete's lemma to conclude that asymptotic entropy is well defined.

3.1. **Example.** Consider biased or unbiased random walk on $\mathbb{Z}$ and let μ be the probability measure associated with it. The asymptotic entropy of μ is zero. Indeed, we have

$$H(\mu^{(n)}) \leq \log |\operatorname{supp} \mu^{(n)}| \leq \log(2n+1),$$

where for the first inequality we just use the upper bound for the entropy from above, while the second one should be clear since $\mu^{(n)}$ is the distribution of X_n, the position of the random walk at time n. So dividing by n and letting n tend to infinity we see that $h(\mu)$ has to be zero.

4. Drift and Entropy

The aim of this section is to show that positivity of asymptotic entropy $h(\mu)$ and drift $\sigma(\mu)$ are equivalent under certain assumptions. Here is the precise statement:

4.1. **Theorem.** Assume we have given a group and a symmetric and finitely supported probability measure μ, whose support generates an infinite subgroup. Then the corresponding random walk has non-trivial drift $\sigma(\mu)$ if and only if its asymptotic entropy $h(\mu)$ is positive.

In case μ is not symmetric this equivalence does not hold: biased random walk on $\mathbb{Z}$ represents a counter-example.

Two inequalities. The proof of this equivalence relies on the following two inequalities:

$$\sigma^2/2 \leq h \leq \sigma \cdot v,$$

where $v = v(\mu)$ denotes the logarithmic growth of the underlying group with respect to the support of μ, that is v is defined by

$$v = \lim_{n\to\infty} \log(|B_n|)/n$$

where $B_n = B_n(\mu)$ is the set of those elements which can be written as a product of at most n elements of the support of μ, or differently expressed B_n is the ball of radius n centered at the identity element in the graph associated with the Markov chain. Again, the limit exists because of the subadditivity of the sequence $\log|B_n|$. The first inequality is based on a result of Varopoulos, generalized and improved by Carne and again improved a little bit by Lyons and Peres, see [11], [12] and [13].

4.2. **Theorem.** For a reversible Markov chain with state space X and Markov kernel P and reversible measure π, we have

$$p^{(n)}(x,y) \leq 2\left(\frac{\pi(y)}{\pi(x)}\right)^{1/2} |P|^n \, e^{-d(x,y)^2/2n} \quad \text{for all } x,y \in X,$$

where $|P|$ denotes the operator norm of P on the Hilbert space $L^2(\pi)$

$$L^2(\pi) = \{f : X \to \mathbb{C} \mid \sum_x |f(x)|^2 \pi(x) < \infty\}, \quad (f,g)_\pi = \sum_x f(x)\overline{g(x)}\pi(x).$$

The second inequality - $h \leq \sigma \cdot v$ - is due to Vershik and we quickly prove it here following the presentation in his article [14]:

Proof. We already know that Kingman's subadditive ergodic theorem implies almost sure convergence of $d(X_0,X_n)/n$. We can also say something about its limit: since the increments of the random walk are chosen independently and the limit is invariant under a change of a finite number of increments we can use Kolmogorov's zero-one law to deduce that the limit has to be constant. Obviously, by the definition of the drift of a random walk this constant has to be equal to σ (using the dominated convergence

theorem). So, for any $\varepsilon > 0$ and for almost every sequence $x_0, x_1, \ldots$ there's an N such that for all $n \geq N$

$$|d(x_0, x_n)/n - \sigma| \leq \varepsilon.$$

In the following we write $\mu^{(n)}$ as a convex combination of a probability measure μ_1, the normalized restriction of $\mu^{(n)}$ to the set V of elements whose length lies in the interval $[(1-\varepsilon)\sigma n, (1+\varepsilon)\sigma n]$, and a probability measure μ_2, its normalized restriction to the complement of V. Using the fact that almost sure convergence implies convergence in probability it's not difficult to see that we have the following upper bound for the entropy of $\mu^{(n)}$,

$$H(\mu^{(n)}) \leq H(\mu_1) - \log(1-\varepsilon) + \varepsilon \log|B_n| + C.$$

While the constant C is independent of n, $H(\mu_1)$ is not, so let's have a closer look at this term. Using Jensen's inequality and the definition of V and logarithmic volume we deduce $H(\mu_1) \leq \log|V| \leq \log(|B_{(1+\varepsilon)\sigma n}|) = (1+\varepsilon)\sigma n(v + o(1))$, which finally allows us to conclude that

$$h \leq (1+\varepsilon)\sigma \cdot v + \varepsilon \cdot v,$$

and since ε was chosen arbitrarily we have proven $h \leq \sigma \cdot v$. □

Before proving Carne's upper bound for the transition probabilities of certain Markov chains we show how it implies the lower bound for the asymptotic entropy of a symmetric and finitely supported probability measure on a group. Obviously $\pi \equiv 1$ is a reversible measure in this setting, so by the above theorem we have $\mu^{(n)}(X_n) \leq 2|P|^n e^{-d(e,X_n)^2/2n}$, and hence

$$\begin{aligned} H(\mu^{(n)})/n &\geq \mathbb{E}[-\log(2|P|^n e^{-d(e,X_n)^2/2n})]/n \\ &= -\log(2)/n - \log|P| + \mathbb{E}[d(e,X_n)^2]/2n^2 \\ &\geq -\log(2)/n + \mathbb{E}[d(e,X_n)^2/2n^2]. \end{aligned}$$

The last equality is true because the norm of a Markov kernel is lower than or equal to 1. Letting n tend to infinity we arrive at

$$\sigma^2/2 \leq h.$$

There are at least two known different approaches in proving Carne's upper bound. The first one is of spectral analytic kind, while the second one is probabilistic. We sketch both proofs. For the first one we follow the presentation in [13]. We begin with a sketch of what is going to happen on the next few pages.

The transition probability $p^{(n)}(x,y)$ of a Markov chain can be represented as $(f_x, P^n f_y)_\pi$, for appropriate unit vectors f_x and f_y. Basic properties of Chebyshev polynomials Q_k for $k \in \mathbb{Z}$ allow to write the iterated transition kernel P^n as

$$P^n = |P|^n \sum_{k \in \mathbb{Z}} q_n(k) Q_k(P/|P|),$$

where $q_n(k)$ is the probability of simple random walk on $\mathbb{Z}$ to be at k at time n when started at 0. So replacing P^n in the above inner product by this expression and applying the Cauchy-Schwarz-inequality we end up in a situation in which we would like to have an upper bound for the norms of the operators $Q_k(P/|P|)$ and certain sums of the $q_n(k)$'s. An upper bound for the first one can be achieved by using spectral theoretic methods for self-adjoint operators on Hilbert spaces. An upper bound for the second one by using an inequality discovered independently by Hoeffding and Azuma in the 1960's, see [15] and [16].

We start with presenting the theorem of Hoeffding and Azuma.

4.3. **Theorem.** Let $X_1,...,X_n$ be bounded random variables such that the expectations of products of the form $X_{i_1}\cdots X_{i_k}$ for $1 \leq i_1 < ... < i_k \leq n$ and $1 \leq k \leq n$ are zero. Then for any $L > 0$,

$$\mathbb{P}[\sum_{i=1}^{n} X_i \geq L] \leq \exp(-\frac{L^2}{2\sum_{i=1}^{n} ||X_i||_\infty^2}).$$

Proof. By convexity of $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto e^{ax}$, a being real, for any $x \in [-1,1]$ we have

$$e^{ax} = f(x) \leq \frac{1-x}{2} f(-1) + \frac{x+1}{2} f(1) = \cosh a + x \sinh a.$$

So defining $x_i = X_i/||X_i||_\infty$ and $a_i = t||X_i||_\infty$, where t is specified later, this implies

$$\exp(t\sum_{i=1}^{n} X_i) \leq \prod_{i=1}^{n} (\cosh(t||X_i||_\infty) + \frac{X_i}{||X_i||_\infty} \sinh(t||X_i||_\infty)).$$

Taking expectations and using our assumptions, we have

$$\mathbb{E}[\exp(t\sum_{i=1}^{n} X_i)] \leq \prod_{i=1}^{n} \cosh(t||X_i||_\infty),$$

and combining this with

$$\cosh x = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{x^{2k}}{2^k k!} = e^{x^2/2}$$

leads to

$$\mathbb{E}[\exp(t\sum_{i=1}^{n} X_i)] \leq \exp(\frac{t^2}{2} \sum_{i=1}^{n} ||X_i||_\infty^2).$$

Markov's inequality allows to conclude that for any $t > 0$

$$\mathbb{P}[\sum_{i=1}^{n} X_i \geq L] = \mathbb{P}[\exp(t\sum_{i=1}^{n} X_i) \geq e^{Lt}] \leq e^{-Lt} \exp(\frac{t^2}{2} \sum_{i=1}^{n} ||X_i||_\infty^2).$$

Choosing $t = L/(\sum_{i=1}^{n} ||X_i||_\infty^2)$ finally proves the statement. □

4.4. **Example.** For simple random walk $X_0, X_1, \ldots$ on $\mathbb{Z}$ we have already noticed that the strong law of large numbers implies almost sure convergence, and hence in particular convergence in probability of X_n/n to 0. The theorem of Hoeffding and Azuma gives a rate for how fast this is done. Adapting the variables to our setting gives

$$\mathbb{P}[|X_n|/n \geq L] \leq 2\exp(-\frac{nL^2}{2}).$$

For the proof of Carne's upper bound we will need the estimate in the following form

$$\mathbb{P}[|X_n| \geq d] \leq 2\exp(-\frac{d^2}{2n}).$$

The next theorem is stated only, for its proof the reader is referred to Emmanuel Kowalski's lectures notes on spectal theory in Hilbert spaces, see [17].

4.5. **Theorem.** For any bounded and selfadjoint operator T on a Hilbert space there is a projection-valued measure Π which allows to represent T as

$$T = \int_{-||T||}^{||T||} s\, d\Pi(s).$$

Furthermore, for any polynomial P we have

$$P(T) = \int_{-||T||}^{||T||} P(s)\, d\Pi(s).$$

The reason why we are interested in this theorem is, because it allows us to deduce an upper bound for the operator norm of $P(T)$

$$||P(T)|| \leq \max_{|s|\leq||T||} |P(s)|.$$

We roughly sketch the argument how this can be deduced from the above theorem: first $||P(T)|| = r(P(T))$, where the range $r(P(T))$ of the operator $P(T)$ is defined by $\max\{|\lambda| \,|\, \lambda \in \sigma(P(T))\}$. The spectrum $\sigma(P(T))$ of $P(T)$ is defined by $\{\lambda \in \mathbb{C} \,|\, P(T) - \lambda \cdot Id \notin \mathbf{G}(H)\}$ and contained in the closure of the numerical range $N(P(T))$ given by $\{(P(T)v, v) \,|\, ||v|| = 1\}$. For the modulus of $(P(T)v, v)$ we have the following bound:

$$\begin{aligned} |(P(T)v,v)| &\leq |\int_{-||T||}^{||T||} P(s)\, d\mu_v(s)| \leq \int_{-||T||}^{||T||} |P(s)|\, d\mu_v(s) \\ &\leq \max_{|s|\leq||T||} |P(s)|\, \mu_v([-||T||, ||T||]) \leq \max_{|s|\leq||T||} |P(s)|, \end{aligned}$$

where the measure μ_v is defined by $\mu_v(A) = (\Pi_A v, v)$ for $A \in \mathbf{B}(\mathbb{R})$. So the upper bound for the norm of the operator does indeed hold.

To proof Carne's upper bound we have to introduce Chebyshev polynomials: for every non-negative integer k there's a unique polynomial Q_k of degree k such that

$$Q_k(\cos\theta) = \cos k\theta \quad \text{for } \theta \in \mathbb{R}.$$

In particular $|Q_k(s)| \leq 1$ for $-1 \leq s \leq 1$. We start with proving their existence. First of all any complex number z can be written as $z = (w+w^{-1})/2$ for some complex number w. This is not so difficult to see. Writing down the equation in a nice way one realizes that this statement reduces to the existence of real roots for real polynomials of degree 3. Induction on k proves the existence of polynomials Q_k such that

$$Q_k(z) = Q_k((w+w^{-1})/2) = (w^k + w^{-k})/2.$$

For $k = 0$ the statement obviously is true after what we have proved so far. So let k be greater than 0 and assume the statement to be true for $l < k$. The binomial theorem shows that $(w^k + w^{-k})/2 - (2z)^k/2$ is of the form $-\sum_{l=1}^{k-1} \binom{k}{l} w^{-k+2l}/2$, so the induction hypothesis can be applied to show that there is a polynomial Q_k of degree k with the desired property. Choosing w as $e^{i\theta}$ shows that the polynomials Q_k are actually those we are looking for, namely Chebyshev polynomials. They are unique since any two polynomials which agree on infinitely many points have to be equal. Since any $s \in [-1,1]$ is of the form $\cos\theta$ for some $\theta \in \mathbb{R}$, the last statement follows immediately.

4.6. **Lemma.** Denoting the Chebyshev polynomial of degree k by Q_k, for the Markov operator P associated with a reversible random walk, we have

$$P^n = |P|^n \sum_{k\in\mathbb{Z}} q_n(k) Q_k(P/|P|),$$

where $|.| = |.|_\pi$ denotes the operator norm induced by a reversible measure for the random walk and $q_n(k)$ the probability of simple random walk on $\mathbb{Z}$ to be at k at time n when started at 0. Moreover for any k the operator norm of $Q_k(P/|P|)$ is bounded by 1.

Proof. Using again that any complex number z can be written as $(w + w^{-1})/2$ for some complex number w, using the binomial theorem and the characterizing identity for Chebyshev polynomials we can deduce that

$$\begin{aligned} z^n = (w+w^{-1})^n/2^n &= \sum_{k\in\mathbb{Z}} q_n(k) w^k \\ &= \sum_{k\in\mathbb{Z}} q_n(k)(w^k + w^{-k})/2 = \sum_{k\in\mathbb{Z}} q_n(k) Q_k(z). \end{aligned}$$

Replacing z by the Markov operator $P/|P|$ we end up at the desired representation for P^n. □

We finally are able to proof Carne's upper bound for the transition probabilities of a reversible Markov chain:

Proof. Let d be the distance between two states x and y in the graph metric of the graph associated with P and denote by f_x and f_y the functions $\mathbf{1}_{\{x\}}/\sqrt{\pi(x)}$ and $\mathbf{1}_{\{y\}}/\sqrt{\pi(y)}$, respectively.

Now, $p^{(n)}(x,y)$ can be written as

$$p^{(n)}(x,y) = \left(\frac{\pi(y)}{\pi(x)}\right)^{1/2} (f_x, P^n f_y)_\pi.$$

Using the representation for P^n from above, the fact that

$$(f_x, Q_k(P/|P|) f_y)_\pi = 0 \quad \text{for } |k| < d,$$

which holds since Q_k has degree k and $p^{(i)}(x,y) = 0$ for $i < d$, and the bound

$$|(f_x, Q_k(P/|P|) f_y)_\pi| \leq |Q_k(P/|P|)|||f_x||_\pi ||f_y||_\pi \leq 1,$$

for which we use Theorem 4.5, we arrive at

$$p^{(n)}(x,y) \leq \left(\frac{\pi(y)}{\pi(x)}\right)^{1/2} |P|^n \sum_{|k|\geq d} q_n(k).$$

Applying the bound for $\sum_{|k|\geq d} q_n(k)$ we derived from the Azuma-Hoeffding inequality we end up with the desired upper bound for the transition probabilities $p^{(n)}(x,y)$. □

A probabilistic approach to Carne's bound. The presentation of the probabilistic prove follows Remi Peyre's article, see [18]:

Let $X_0, X_1, X_2, \ldots$ be an irreducible and reversible Markov chain with state space X, transition kernel P and reversible measure μ. To proof that

$$p^{(n)}(x,y) \leq 2\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P|^n \exp\left(-\frac{d(x,y)^2}{2n}\right),$$

for $n \geq 1$ and $x, y \in X$, we first show that the upper bound

$$p^{(n)}(x,y) \leq \sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \exp\left(-\frac{(d(x,y)-1)^2}{2(n-1)}\right)$$

holds for all $n \geq 2$ and $x \neq y \in X$.

By reversibility of the Markov chain, we have

$$p^{(n)}(y,x) = \frac{\mu(x)}{\mu(y)} p^{(n)}(x,y),$$

so to prove the above upper bound it's sufficient to prove

$$p^{(n)}(x,y)p^{(n)}(y,x) \leq \exp\left(\frac{(d(x,y)-1)^2}{n-1}\right).$$

The next lemma is essential for achieving this.

4.7. **Lemma.** Let X be a centered real-valued random variable such that for every $\lambda \in \mathbb{R}$

$$\mathbb{E}[e^{\lambda X}] \leq e^{k\lambda^2/2}$$

for some $k > 0$.
If A is an event such that

$$\mathbb{E}[X|A] \geq C$$

for some constant $C \geq 0$, then

$$\mathbb{P}(A) \leq \exp\left(-\frac{C^2}{2k}\right).$$

Proof. Our assumptions and Jensen's inequality imply the following sequence of equalities and inequalities:

$$C \leq \mathbb{E}[X|A] = \mathbb{E}[\frac{1}{\lambda}\ln\left(e^{\lambda X}\right)|A] \leq \frac{1}{\lambda}\ln\mathbb{E}[e^{\lambda X}|A]$$
$$= \frac{1}{\lambda}\ln\frac{\mathbb{E}[\mathbf{1}_A e^{\lambda X}]}{\mathbb{P}[A]} \leq \frac{1}{\lambda}\ln\frac{e^{k\lambda^2/2}}{\mathbb{P}[A]},$$

and hence

$$\mathbb{P}[A] \leq e^{-C\lambda + k\lambda^2/2}.$$

For $\lambda = C/k$ we get the desired result. □

Obviously we would like to apply this lemma to prove the upper bound for the transition probabilities. It's clear what the event A is and how the constants C and k should look like in this context. From this we are lead to the following construction: having fixed two states $x, y \in X$, we introduce a function $\xi : X \to \mathbb{R}$ by $\xi(z) = d(x,z)$, and a function $m : X \to \mathbb{R}$ by $m(z) = \mathbb{E}_z[\xi(X_1)] - \xi(z)$, the expectation of the variation of the distance after one step of the Markov chain, starting at z. We then define a martingale $(M_n)_{n\geq 1}$ by

$$M_n = \xi(X_n) - \xi(X_1) - \sum_{k=1}^{n-1} m(X_k) = \sum_{k=1}^{n-1}(\xi(X_{k+1}) - \xi(X_k) - m(X_k)).$$

Conditioned on the event $\{X_n = y\}$ we have $\xi(X_n) - \xi(X_1) \geq d(x,y) - 1$, and therefore

$$\mathbb{E}_x[M_n|X_n = y] \geq d(x,y) - 1 - \mathbb{E}_x[\sum_{k=1}^{n-1} m(X_k)|X_n = y],$$

and similarly for the chain starting at y,

$$\mathbb{E}_y[M_n|X_n = x] \leq -(d(x,y) - 1) - \mathbb{E}_y[\sum_{k=1}^{n-1} m(X_k)|X_n = x].$$

Reversibility of the Markov chain ensures that

$$\mathbb{E}_x[m(X_k)|X_n = y] = \mathbb{E}_y[m(X_{n-k})|X_n = x],$$

for every $k \in \{1,...,n-1\}$. So let us consider two independent Markov chains X^x and X^y - one starting at x, the other at y - and denote their joint law by $\mathbb{P}_{x\otimes y}$. Denoting the corresponding martingales by M_n^x and M_n^y, we have

$$\mathbb{E}_{x\otimes y}[M_n^x - M_n^y | X_n^x = y, X_n^y = x] \geq 2(d(x,y)-1).$$

To be able to apply the above measure concentration lemma, we notice that $M_n^x - M_n^y$ can be written as $\sum_{k=1}^{2(n-1)} Y_k$, where $(Y_k)_{k\geq 0}$ satisfies the assumptions of the following lemma:

4.8. **Lemma.** Let $(\mathscr{F}_n)_{n\geq 1}$ be a filtration and $(X_n)_{n\geq 1}$ an adapted real-valued process such that $\mathbb{E}[X_{n+1}|\mathscr{F}_n] = 0$. If the law of $(X_{n+1}|\mathscr{F}_n)$ is supported by an interval of lenght 2 almost surely, then

$$\mathbb{E}[\exp\left(\lambda \sum_{k=1}^{n} X_k\right)] \leq \exp\left(n\frac{\lambda^2}{2}\right),$$

for $n \geq 0$ and all $\lambda \in \mathbb{R}$.

Proof. To proof this lemma we use an inequality of Hoeffding:

4.9. **Lemma.** For a centered real-valued random variable X, whose law is supported by an interval of length 2, the following inequality holds for every $\lambda \in \mathbb{R}$

$$\mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2/2}.$$

Taking this for granted for a moment, we proceed by induction. The statement obviously is true for $n = 0$, so let n be greater than 0 and suppose it's true for $n-1$. Then

$$\begin{aligned}
\mathbb{E}[\exp\left(\lambda \sum_{k=1}^{n} X_k\right)] &= \mathbb{E}[\exp\left(\lambda \sum_{k=1}^{n-1} X_k\right) \mathbb{E}[e^{\lambda X_n}|\mathscr{F}_{n-1}]] \\
&\leq \mathbb{E}[\exp\left(\lambda \sum_{k=1}^{n-1} X_k\right)] \cdot ||\mathbb{E}[e^{\lambda X_n}|\mathscr{F}_{n-1}]||_\infty \\
&\leq e^{(n-1)\lambda^2/2} \cdot e^{\lambda^2/2} = e^{n\lambda^2/2}
\end{aligned}$$

□

Applying this lemma on $M_n^x - M_n^y$ gives

$$\mathbb{E}_{x\otimes y}[e^{\lambda(M_n^x - M_n^y)}] \leq e^{(n-1)\lambda^2},$$

for every $\lambda \in \mathbb{R}$. And using the measure concentration lemma we finally arrive at

$$\mathbb{P}_{x\otimes y}[X_n^x = y,\, X_n^y = x] \leq \exp(-\frac{(d(x,y)-1)^2}{n-1}).$$

As we have pointed out above, this implies

$$p^{(n)}(x,y) \leq \left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \exp\left(-\frac{(d(x,y)-1)^2}{2(n-1)}\right),$$

for $n \geq 2$ and $x \neq y \in X$, which again easily implies that

$$p^{(n)}(x,y) \leq \sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} \exp\left(-\frac{d(x,y)^2}{2n}\right),$$

for $n \geq 1$ and all $x, y \in X$.

At last we prove Hoeffding's lemma in the following, a little bit more general form:

4.10. **Lemma.** Let X be a centered real random variable, taking values in an interval $[a,b]$ almost surely. Then

$$\mathbb{E}[e^{sX}] \leq \exp\left(\frac{1}{8}s^2(b-a)^2\right).$$

Proof. Using that $x \mapsto e^{sx}$ is a convex function we have for all $x \in [a,b]$

$$e^{sx} \leq \frac{b-x}{b-a}e^{sa} + \frac{x-a}{b-a}e^{sb}$$

and hence, for the expectation

$$\begin{aligned}\mathbb{E}[e^{sX}] &\leq \frac{b-\mathbb{E}[X]}{b-a}e^{sa} + \frac{\mathbb{E}[X]-a}{b-a}e^{sb} \\ &= \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb}.\end{aligned}$$

The term on the right side can be written as $(1 + \frac{a}{b-a} - \frac{a}{b-a}e^{s(b-a)})e^{sa}$ or as $(1-\theta+\theta e^{s(b-a)})e^{-s\theta(b-a)}$, for $\theta = -a/(b-a)$. Defining $\phi : \mathbb{R} \to \mathbb{R}$ by $\phi(u) = -\theta u + \log(1-\theta+\theta e^u)$ we obviously have

$$\mathbb{E}[e^{sX}] \leq e^{\phi(s(b-a))}.$$

To finish the proof of this lemma it's sufficient to notice that by Taylor's theorem

$$\phi(u) \leq \frac{1}{8}u^2,$$

since $\phi(0) = \phi'(0) = 0$ and $\phi''(u) \leq 1/4$ for every $u \in \mathbb{R}$.

□

In what follows we deduce Carne's bound from what we have proven so far. Starting point is the following observation: $|P| < 1$ expresses the possibility of the Markov chain to "flee to infinity". The idea is to manipulate the chain such that this property is removed, find out how this manipulation changes the probabilities of the events we are interested in and apply the upper bound we have proven so far.

By $\mathscr{R}_n$ we denote the event of visiting x at or after a fixed time n, and by τ_x the hitting time of x of the Markov chain, that is

$$\tau_x = \inf\{k \geq 0 \,|\, X_k = 0\}.$$

For $z \in X$ we finally denote

$$R(z) = \mathbb{E}_z[\mathbf{1}_{\tau_x<\infty}|P|^{-\tau_x}].$$

For the moment we make the following assumptions, which will be removed again later on:

(1) X is finite,
(2) there's a state $\dagger \in X$ such that $p(\dagger,\cdot) = \delta_\dagger$ - $X \setminus \{\dagger\}$ will be denoted by $\tilde{X}$,
(3) the Markov chain is aperiodic on $\tilde{X}$.

4.11. **Lemma.** For an irreducible and reversible Markov chain satisfying the above assumptions, it is true that

(1) there are constants $0 < c_1 \leq c_2 < \infty$, such that

$$c_1|P|^n \leq \mathbb{P}_x[\mathscr{R}_n] \leq c_2|P|^n,$$

for every $n \geq 0$.
(2) $\mathbb{P}_x[\mathscr{R}_{n+1}]/\mathbb{P}_x[\mathscr{R}_n] \to |P|$ as n tends to infinity,
(3) $R(z) < \infty$ for all $z \in \tilde{X}$.

The proof of this lemma is rather technical; we do not present it here and refer to [18]. Based on this lemma we prove the following proposition:

4.12. **Proposition.** If we condition on the event $\mathscr{R}_m$ and let m tend to infinity, the law of $(X_k)_{0\leq k\leq 2n}$ converges - with respect to the total variation norm on $X^{\{0,\ldots,2n\}}$ - to the corresponding law of a Markov chain on $\tilde{X}$, whose transtition probabilities are given by

$$p'(z,v) = \frac{p(z,v)R(v)}{\sum_{w\sim z} p(z,w)R(w)},$$

for $z, v \in \tilde{X}$.

Proof. Using the Markov property of the chain we easily see that

$$\mathbb{P}_x[X_{k+1} = v \,|\, X_k = z, \mathscr{R}_m] = \frac{p(z,v)\mathbb{P}_v[\mathscr{R}_{m-k-1}]}{\sum_{w\sim z} p(z,w)\mathbb{P}_w[\mathscr{R}_{m-k-1}]},$$

Hence, it is sufficient to show that

$$\frac{\mathbb{P}_z[\mathscr{R}_m]}{\mathbb{P}_x[\mathscr{R}_m]} \longrightarrow R(z),$$

as m tends to infinity. For achieving this we write

$$\mathbb{P}_z[\mathscr{R}_m] = \sum_{k=0}^{m} \mathbb{P}_z[\tau_x = k]\mathbb{P}_x[\mathscr{R}_{m-k}] + \sum_{k\geq m+1} \mathbb{P}_z[\tau_x = k],$$

hence

$$\frac{\mathbb{P}_z[\mathscr{R}_m]}{\mathbb{P}_x[\mathscr{R}_m]} = \sum_{k=0}^{m} \mathbb{P}_z[\tau_x = k]\frac{\mathbb{P}_x[\mathscr{R}_{m-k}]}{\mathbb{P}_x[\mathscr{R}_m]} + \sum_{k>m} \frac{\mathbb{P}_z[\tau_x = k]}{\mathbb{P}_x[\mathscr{R}_m]}.$$

The third item of the last lemma implies that, given $\varepsilon > 0$, an m_0 can be chosen such that $\sum_{k>m_0} \mathbb{P}_z[\tau_x = k] \le \varepsilon$. This and the first item of the lemma then imply

$$\sum_{k=m_0+1}^{m} \mathbb{P}_z[\tau_x = k] \frac{\mathbb{P}_x[\mathscr{R}_{m-k}]}{\mathbb{P}_x[\mathscr{R}_m]} + \sum_{k>m} \frac{\mathbb{P}_z[\tau_x = k]}{\mathbb{P}_x[\mathscr{R}_m]} \le \frac{c_2 \vee 1}{c_1} \sum_{k>m_0} \mathbb{P}_z[\tau_x = k] |P|^{-m} \le \frac{c_2 \vee 1}{c_1}$$

For the remaining terms the second item of the lemma implies

$$\sum_{k=0}^{m_0} \mathbb{P}_z[\tau_x = k] \frac{\mathbb{P}_x[\mathscr{R}_{m-k}]}{\mathbb{P}_x[\mathscr{R}_m]} \longrightarrow \sum_{k=0}^{m_0} \mathbb{P}_z[\tau_x = k] |P|^{-k},$$

as m tends to infinity. Altogether this gives

$$\limsup_{m\to\infty} |R(z) - \frac{\mathbb{P}_v[\mathscr{R}_m]}{\mathbb{P}_x[\mathscr{R}_m]}| \le \left(1 + \frac{c_2 \vee 1}{c_1}\right) \varepsilon,$$

finally proving the proposition since ε can be chosen arbitrarily small. □

The modified Markov chain we have constructed above obviously is irreducible. The next proposition shows that it also is reversible, and identifies a reversible measure:

4.13. **Proposition.** The above Markov chain is reversible and a reversible measure is given by

$$\mu'(z) = \begin{cases} R(z)^2 \mu(z) & \text{if} \quad z \neq x, \\ R(x) R^+(x) \mu(x) & \text{if} \quad z = x, \end{cases}$$

where μ is a reversible measure for the original Markov chain and $R^+(x)$ is defined by

$$R^+(x) = \mathbb{E}_x[\mathbf{1}_{\tau_x^+ < \infty} |P|^{-\tau_x^+}],$$

τ_x^+ being the first positive return time of the chain,

$$\tau_x^+ = \inf\{k \ge 1 \mid X_k = x\}.$$

Denote the law of the new Markov chain, starting at x, by $\mathbb{P}'_x$, and let $\mathbb{P}_x[A]$ denote $p^n(x,y)p^n(y,x) = \mathbb{P}_x[X_n = y, X_{2n} = x]$. For $m \ge 2n$ we have

$$\frac{\mathbb{P}_x[A]}{\mathbb{P}_x[A|\mathscr{R}_m]} = \frac{\mathbb{P}_x[\mathscr{R}_m]}{\mathbb{P}_x[\mathscr{R}_m|A]} = \frac{\mathbb{P}_x[\mathscr{R}_m]}{\mathbb{P}_x[\mathscr{R}_{m-2n}]},$$

and hence

$$\frac{\mathbb{P}_x[A]}{\mathbb{P}'_x[A]} = |P|^{2n}.$$

Since $\mathbb{P}'_x$ is the law of an irreducible and reversible Markov chain, we know that $\mathbb{P}'_x[A] \le e \cdot \exp(-d(x,y)^2/n)$, therefore $\mathbb{P}_x[A] \le e|P|^{2n} \exp(-d(x,y)^2/n)$ and

$$p^{(n)}(x,y) \le \sqrt{e} \left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P|^n \exp(-d(x,y)^2/2n).$$

For the rest of the section we show that we can remove the additional assumptions we have made above, and begin with doing that for the aperiodicity of the Markov chain via a continuity argument. We recall that X is a finite set on which the transition matrix $P = (p(x,y))_{x,y\in X}$ defines a Markov chain, for which there is a cemetery point $\dagger \in X$, and which is irreducible and reversible on $\tilde{X} = X \setminus \{\dagger\}$.

For $\varepsilon \in [0,1)$, let P_ε be the transition kernel defined by

$$p_\varepsilon(x,y) = \begin{cases} p(x,x) + \varepsilon(1-p(x,x)) & \text{if} \quad y = x, \\ (1-\varepsilon)p(x,y) & \text{if} \quad y \neq x. \end{cases}$$

The in such a manner defined Markov chain is irreducible and reversible and aperiodic on the complement of the cemetery point. Furthermore, the reversible measure is independent of ε as is the graph metric of the corresponding graph of the Markov chain. So applying the above estimate we arrive at

$$p_\varepsilon^{(n)}(x,y) \leq \sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P_\varepsilon|^n \exp(-d(x,y)^2/2n).$$

It is sufficient to remark that $p_\varepsilon^{(n)}(x,y)$ is continuous in ε as is the operator norm $|P_\varepsilon|$, being the spectral radius of a matrix whose entries depend continuously on ε.

Having shown we can drop the aperiodicity condition, we next show we can drop the assumption of the state space being finite. The idea is to replace a general Markov chain by a sequence of finite ones which approximate the original one. Consider a general irreducible and reversible Markov chain with state space X. Let $\nu : X \to (0,\infty)$ be a weight such that for any $\varepsilon > 0$ the cardinality of $\{z \in X \mid \nu(z) > \varepsilon\}$ is a finite one. Fix $x, y \in X$ and let $\varepsilon > 0$ such that $\varepsilon < \nu(x), \nu(y)$. We identify all points in X whose weight is $\leq \varepsilon$. By $\dagger$ we denote the point we have received in this way. This point is the cemetery point of the following finite Markov chain:

4.14. **Definition.** Let X_ε be the set defined by

$$X_\varepsilon = \{z \in X \mid \nu(z) > \varepsilon\} \cup \{\dagger\},$$

and define a Markov chain via the following specification of its transition probabilities

$$p_\varepsilon(z,v) = \begin{cases} p(z,v) & \text{if} \quad z,v \in \tilde{X}_\varepsilon = X_\varepsilon \setminus \{\dagger\}, \\ 0 & \text{if} \quad z = \dagger,\ v \in \tilde{X}_\varepsilon, \\ 1 & \text{if} \quad z = v = \dagger, \\ \sum_{\nu(w)<\varepsilon} p(z,w) & \text{if} \quad z \in \tilde{X}_\varepsilon,\ v = \dagger. \end{cases}$$

If we restrict our Markov chain to the irreducible class which contains x, all necessary assumptions are fullfilled to conclude

$$p_\varepsilon^{(n)}(x,y) \leq \sqrt{e}\left(\frac{\mu(y)}{\mu(x)}\right)^{1/2} |P_\varepsilon|^n \exp\left(-\frac{d_\varepsilon(x,y)^2}{2n}\right).$$

In what is left to finish the proof we have to show that $d_\varepsilon(x,y) \to d(x,y)$, $p_\varepsilon^{(n)}(x,y) \to p^{(n)}(x,y)$ and $|P_\varepsilon| \to |P|$ as ε tends to 0; instead of the last it is sufficient to show $|P_\varepsilon| \leq |P|$. The first one of these three statements is obvious by the very construction of the transition kernel. For the third one we recall the following representation of the norm of a Markov operator,

$$|P_\varepsilon| = \limsup_{n\to\infty} p_\varepsilon^{(n)}(x,x)^{1/n} = \sup_{n\geq 1} p_\varepsilon^{(n)}(x,x)^{1/n},$$

see [13]. Since $p_\varepsilon^{(n)}(z,v) \leq p^{(n)}(z,v)$, and in particular $p_\varepsilon^{(n)}(x,x) \leq p^{(n)}(x,x)$, this implies $|P_\varepsilon| \leq |P|$. To see that the second statement holds, let $\mathbb{P}_x$ and $\mathbb{P}_x^\varepsilon$ denote the law of $(X_k)_{0\leq k\leq n}$ on X and X_ε. Making no formal difference between the elements of X and the corresponding elements in X_ε we can consider $\mathbb{P}_x$ and $\mathbb{P}_x^\varepsilon$ as two probability measures on the same probability space; so it does make sense to consider their total variation distance and deduce the following estimate

$$\begin{aligned}\|\mathbb{P}_x^\varepsilon \restriction X_\varepsilon^{\{0,\dots,n\}} - \mathbb{P}_x \restriction X^{\{0,\dots,n\}}\|_{TV} &\leq \mathbb{P}_x[\exists k \in \{0,\dots,n\} \text{ s.t. } X_k \notin \tilde{X}_\varepsilon] \\ &\leq \sum_{k=0}^{n} \mathbb{P}_x[\nu(X_k) < \varepsilon].\end{aligned}$$

The upper sum converges to 0 as ε tends to 0, by the theorem of dominated convergence. In particular $p_\varepsilon^{(n)}(x,y)$ converges to $p^{(n)}(x,y)$ as ε approaches 0, which finally shows that we can drop the finiteness condition concerning the state space of the Markov chain.

5. THE INVARIANT σ-ALGEBRA

We introduce the tail- and invariant σ-algebra, explain how they are related to each other, show that there is a very close connection between bounded invariant functions and bounded harmonic functions, and give a criterium for the triviality of the invariant σ-algebra in terms of the non-existence of non-constant bounded harmonic functions.

Let $X_0, X_1, \ldots$ be a sequence of random variables. Consider $\sigma(X_n, X_{n+1}, \ldots)$, the smallest σ-algebra with respect to the random variables X_k, $k \geq n$ are measurable. This σ-algebra describes the future after and including n. The tail-σ-algebra $\mathscr{T}$ of the sequence is defined by

$$\mathscr{T} = \bigcap_{n \geq 0} \sigma(X_n, X_{n+1}, \ldots),$$

and describes the asymptotic behaviour of the sequence.

We recall the definition of the shift operator $S : X^{\mathbb{N}} \to X^{\mathbb{N}}$ which maps a sequence $x_0, x_1, \ldots$ of elements of X to the sequence $x_1, x_2, \ldots$. The set of events which are invariant with respect to S form a σ-algebra, which is called the invariant σ-algebra and which we denote by $\mathscr{I}$. In symbols

$$\mathscr{I} = \{A \mid A \subseteq X^{\mathbb{N}} \text{ is measurable and } S^{-1}(A) = A\}$$

The tail-σ-algebra does describe the asymptotic behaviour of the sequence $X_0, X_1, \ldots$ as does the invariant σ-algebra. For an invariant event A we have $A = S^{-n}(A) = X^n \times A \in \sigma(X_n, X_{n+1}, \ldots)$ for every n, so the invariant σ-algebra is contained in the tail-σ-algebra. Although the other inclusion is not necessarily true, in many situations one can neglect this difference. This will be the content of the next section.

A function $f : X^{\mathbb{N}} \to \mathbb{R}$ is called a tail function, respectively an invariant function, if f is measurable with respect to $\mathscr{T}$, respectively $\mathscr{I}$. We call two functions f and g equivalent if they agree almost surely with respect to the probability measures $\mathbb{P}_x$ for all $x \in X$.

Harmonicity and Invariance. The next theorem gives a description of bounded invariant functions in terms of bounded harmonic functions in the setting of Markov chains.

5.1. **Theorem.** There is a bijective correspondence between equivalence classes of bounded invariant functions on $X^{\mathbb{N}}$ and bounded harmonic functions on X.

Proof. Consider the following assignment which maps an equivalence class of bounded invariant functions represented by $f : X^{\mathbb{N}} \to \mathbb{R}$ to the bounded harmonic function $h : X \to \mathbb{R}$ defined by

$$h(x) = \mathbb{E}_x[f(X_0, X_1, \ldots)].$$

This assignment is well defined, since the expectation $\mathbb{E}_x$ neglects sets of $\mathbb{P}_x$-measure zero, and h is bounded since f is. Invariance of f implies harmonicity of h: for simplicity, but without loss of generality, assume f to be the indicator function of an invariant set A, $f = \mathbf{1}_A$. By invariance of A, we then have

$$h(x) = \mathbb{P}_x[A] = \sum_{y \in X} P(x,y)\, \mathbb{P}_y[A] = Ph(x).$$

An inverse is given by the map which assigns to a bounded harmonic function $h : X \to \mathbb{R}$ the equivalence class of bounded invariant functions represented by $f : X^{\mathbb{N}} \to \mathbb{R}$ defined by

$$f(x_0, x_1, ...) = \limsup_{n \to \infty} h(x_n).$$

We first show that, given a bounded harmonic function h, we have

$$h(x) = \mathbb{E}_x[\limsup_{n \to \infty} h(X_n)].$$

Since $(h(X_n))$ is a bounded martingale, we have

$$\begin{aligned} \mathbb{E}_x[\limsup_{n \to \infty} h(X_n)] &= \mathbb{E}_x[\lim_{n \to \infty} h(X_n)] = \lim_{n \to \infty} \mathbb{E}_x[h(X_n)] \\ &= \mathbb{E}_x[h(X_0)] = h(x), \end{aligned}$$

where the martingale convergence theorem is used in the first equation and the theorem of dominated convergence in the second one.

Conversely, given a bounded invariant function $f : X^{\mathbb{N}} \to \mathbb{R}$ we have to show that

$$f(x_0, x_1, ...) = \limsup_{n \to \infty} \mathbb{E}_{x_n}[f(X_0, X_1, ...)]$$

holds $\mathbb{P}_x$-almost surely for every $x \in X$.

The σ-algebras $\mathscr{F}_n$ defined by $\mathscr{F}_n = \sigma(X_0, ..., X_n)$ form a filtration such that $\mathscr{F}_n \nearrow \mathscr{F}$, for $\mathscr{F} = \sigma(X_0, X_1, ...)$, meaning that $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$ for every n and $\mathscr{F} = \sigma(\bigcup \mathscr{F}_n)$. Given the (bounded) integrable random variable f, we define random variables X_n by $X_n = \mathbb{E}_x[f|\mathscr{F}_n]$. Obviously, these random variables form a martingale by the tower law for conditional expectations; since f was chosen to be integrable the convergence theorem for martingales implies the existence of a random variable X the martingale converges to almost surely. This limit can be expressed in terms of conditional expectation, that is $X = \mathbb{E}_x[f|\mathscr{F}]$: we just observe that X has the properties which characterize the conditional expectation of f given $\mathscr{F}$. Firstly, X is measurable with respect to $\mathscr{F}$, and secondly, $\int_A X\, d\mathbb{P}_x = \int_A f\, d\mathbb{P}_x$ for every $A \in \mathscr{F}$. The first statement is obviously true. The second one is true, since the set of all $A \in \mathscr{F}$ for which the above equation holds is a Dynkin-system which contains $\bigcup_n \mathscr{F}_n$; hence it also contains the Dynkin-system generated by $\bigcup_n \mathscr{F}_n$, which is equal to $\mathscr{F}$, the σ-algebra generated by $\bigcup_n \mathscr{F}_n$, since $\bigcup_n \mathscr{F}_n$ is stable under intersections. To finish the proof we just observe that $X = f$, since f is measurable with respect to $\mathscr{F}$, and $\mathbb{E}_x[f|\mathscr{F}_n] = \mathbb{E}_x[f \circ S^n|\mathscr{F}_n] = \mathbb{E}_{X_n}[f]$, since f is invariant. □

5.2. **Remark.** The last statement in the above proof - given a filtration $(\mathscr{F}_n)$ such that $\mathscr{F}_n \nearrow \mathscr{F}$ and an integrable random variable X, the conditional expectations $\mathbb{E}[X|\mathscr{F}_n]$ converge to $\mathbb{E}[X|\mathscr{F}]$ almost surely - is known as Levy's 0-1-law. This comes from the following specialization: taking X to be the characteristic function $\mathbf{1}_A$ of an event $A \in \mathscr{F}$ implies almost sure convergence of $\mathbb{E}[X|\mathscr{F}_n]$ to 0 and 1. In case A is independent of $\mathscr{F}_n$ for every n, the conditional expectation of $\mathbf{1}_A$ given $\mathscr{F}_n$ is almost surely equal to $\mathbb{P}(A)$, hence the probability of the event A has to be 0 or 1. Hence, Levy's 0-1-law implies Kolmogorov's 0-1-law.

5.3. **Corollary.** Given an irreducible Markov chain, its invariant σ-algebra $\mathscr{I}$ is trivial with respect to the probability measures $\mathbb{P}_x$ for all $x \in X$ if and only if all bounded harmonic functions are constant.

Proof. Let's assume that $\mathscr{I}$ is trivial, and let $h : X \to \mathbb{R}$ be a bounded harmonic function. We already know that h can be written as $h(x) = \mathbb{E}_x[f]$ for a bounded and invariant function $f : X^{\mathbb{N}} \to \mathbb{R}$. Triviality of $\mathscr{I}$ forces f to be almost surely constant, that is $f \equiv C$ $\mathbb{P}_x$-almost surely, where $C = C(x) = \inf\{c \mid f \geq c \ \mathbb{P}_x\text{-almost surely}\}$ is a constant which apriori might depend on x. Irreducibility of the Markov chain and triviality of the invariant σ-algebra ensure the constant to be independent of x: let x and y be elements of X, and A an invariant event of $\mathbb{P}_x$-measure 0 (if it were not, take the complement of A). By irreducibility there is a sequence $x, x_1, ..., x_{n-1}, y$ of states such that $P(x_i, x_{i+1}) > 0$ for $i = 0, ..., n-1$, hence

$$\begin{aligned}\mathbb{P}_x[A] = \mathbb{P}_x[S^{-n}(A)] &= \mathbb{P}_x[X^n \times A] \\ &\geq P(x, x_1) \cdot ... \cdot P(x_{n-1}, y)\mathbb{P}_y[A]\end{aligned}$$

and A has to be of $\mathbb{P}_y[A]$-measure 0 too.

For the other direction, assume every bounded harmonic function to be constant and let $A \in \mathscr{I}$ be an invariant event. Since $\mathbf{1}_A$ is bounded and invariant it can be written as $\mathbf{1}_A(x_0, ...) = \limsup_{n\to\infty} h(x_n)$ for a bounded harmonic function $h : X \to \mathbb{R}$. By assumption $h \equiv C$, for a constant C, so C has to be either 0 or 1. Which allows to deduce

$$\mathbb{P}_x[A] = \mathbb{E}_x[\mathbf{1}_A] = C \in \{0, 1\}.$$

□

6. ASYMPTOTIC ENTROPY AND THE INVARIANT σ-ALGEBRA

We give a criterium for the triviality of the invariant σ-algebra of a random walk on a group generated by a symmetric probability measure in terms of asymptotic entropy. The proof relies very much on the equality of the tail- and invariant σ-algebra in this setting. Establishing this equality will be the most difficult and lengthiest part of the proof.

We first give the following definition.

6.1. **Definition.** Given two random variables X and Y, we define $H(X|Y)$, the conditional entropy of X given Y, by

$$H(X|Y) = -\sum_{x,y} \mathbb{P}[X=x, Y=y] \log \mathbb{P}[X=x \mid Y=y] = H(X,Y) - H(Y).$$

6.2. **Theorem.** Given a symmetric measure μ on a group, with finite entropy $H(\mu)$, its asymptotic entropy $h(\mu)$ is zero if and only if the invariant σ-algebra of the corresponding random walk is trivial.

Proof. We first show that the tail σ-algebra $\mathscr{T}$ is trivial if and only if the aymptotic entropy $h(\mu)$ is zero. This is done via interpreting vanishing asymptotic entropy as asymptotic independence of $(X_1, ...X_k)$ of X_n as n tends to infinity.

Applying Jensen's inequality to

$$\sum_{x,y} \mathbb{P}[X=x, Y=y] \log \frac{\mathbb{P}[X=x]\mathbb{P}[Y=y]}{\mathbb{P}[X=x, Y=y]},$$

shows that this expression, which is equal to $H(X|Y) - H(X)$, is non-positive and equal to 0 if and only if X and Y are independent. So $H(X|Y) = H(X)$ if and only if X and Y are independent. It's not difficult to see that

$$H(X_1, ..., X_k|X_n) = k \cdot H(X_1) + H(X_{n-k}) - H(X_n),$$

the Markov property of the random walk implies

$$H(X_1, ..., X_k|X_n, X_{n+1}) = H(X_1, ..., X_k|X_n)$$

and Jensen's inequality - similarly as above - that

$$H(X_1, ..., X_k|X_n, X_{n+1}) \leq H(X_1, ..., X_k|X_{n+1}).$$

Combining these three (in-)equalities for the case of $k = 1$ shows that

$$H(X_n) - H(X_{n-1}) \geq H(X_{n+1}) - H(X_n).$$

We use the monotonicity of the increments to show that $H(X_n) - H(X_{n-1})$ converges to $h(\mu)$ as n tends to infinity. For notational convenience we abbreviate $H(X_n)$ by h_n. By monotonicity, the increments $h_n - h_{n-1}$ converge to some h which might be equal to $-\infty$. Let's assume $h \neq h(\mu)$, and

$|h-h(\mu)| \geq \varepsilon > 0$. So w.l.o.g. there's an N such that for all $n \geq N$ we have $h_n - h_{n-1} - h(\mu) \geq \varepsilon/2$. Writing h_n/n as

$$\frac{1}{n}\sum_{k=2}^{N}(h_k - h_{k-1} - h(\mu)) + \frac{1}{n}\sum_{k=N+1}^{n}(h_k - h_{k-1} - h(\mu)) + \frac{n-1}{n}h(\mu) + \frac{h_1}{n},$$

we now easily see that we end up at a contradiction to $\lim h_n/n = h(\mu)$. So the sequence of differences $H(X_n) - H(X_{n-1})$ does converge to $h(\mu)$.

Taking the limit in $H(X_1,...,X_k|X_n) = k \cdot H(X_1) + H(X_{n-k}) - H(X_n)$ gives

$$\begin{aligned}\lim_{n\to\infty} H(X_1,...,X_k|X_n) &= k \cdot H(X_1) - k \cdot h(\mu) \\ &= H(X_1,...,X_k) - k \cdot h(\mu),\end{aligned}$$

where we used that $H(X_1,...,X_k) = k \cdot H(X_1)$. This gives the following expression for the asymptotic entropy

$$h(\mu) = (H(X_1,...,X_k) - \lim_{n\to\infty} H(X_1,...,X_k|X_n))/k,$$

showing that $h(\mu) = 0$ if and only if $\lim_n H(X_1,...,X_k|X_n) = H(X_1,...,X_k)$, which means, as we have seen above using Jensen's inequality, that $(X_1,...,X_k)$ is asymptotically independent of X_n as n tends to infinity.

Let's assume asymptotic entropy to be zero, that is asymptotic independence of $(X_1,...,X_k)$ of (X_n) as n tends to infinity. In general, conditioned on X_n, a tail event $A \in \mathscr{T}$ is independent of $(X_1,...,X_k)$. Under the assumption of asymptotical independence of $(X_1,...,X_k)$ of X_n the event A is independent of $(X_1,...,X_n)$. Applying Kolmogorov's 0-1-law yields triviality of the tail-σ-algebra $\mathscr{T}$.

To see that triviality of the tail-σ-algebra implies asymptotic entropy to be zero we use the following lemma we have taken from Steven Orey's lecture notes on limit theorems for Markov chain transition probabilities, see [19].

6.3. **Lemma.** Let $X_0, X_1, ...$ be a stochastic process, $(\Omega, \mathscr{F}, \mathbb{P})$ the underlying probability space, and define σ-algebras $\mathscr{F}_n = \sigma(X_n, X_{n+1}, ...)$ and $\mathscr{T} = \bigcap_{n\geq 0} \sigma(X_n, X_{n+1,...})$. Then

$$\lim_{n\to\infty} \sup_{A\in\mathscr{F}_n} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| = 0$$

for every $B \in \mathscr{F}$ if and only if $\mathscr{T}$ is trivial.

Proof. In case the above equation does hold, the tail-σ-algebra $\mathscr{T}$ obviously is trivial: let $B \in \mathscr{T}$ and $A = B$. Then $\mathbb{P}(B) = \mathbb{P}(B)^2$, so $\mathbb{P}(B) \in \{0,1\}$.

Now assume the tail-σ-algebra to be trivial. Given $B \in \mathscr{F}$ and $A \in \mathscr{F}_n$, we have

$$\begin{aligned}|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| &= |\int_A \mathbb{P}(B \mid \mathscr{F}_n) - \mathbb{P}(B) d\mathbb{P}| \\ &\leq \int_\Omega |\mathbb{P}(B \mid \mathscr{F}_n) - \mathbb{P}(B)| d\mathbb{P}.\end{aligned}$$

The backward martingale convergence theorem - see the appendix for that - implies convergence of $\mathbb{P}(B \mid \mathscr{F}_n)$ to $\mathbb{P}(B \mid \mathscr{T})$, which is almost surely equal to $\mathbb{P}(B)$ since $\mathscr{T}$ is trivial. Taking the supremum over all events A in $\mathscr{F}_n$ on the left hand side of the above equation - the estimate being independent of A - and letting n tend to infinity, using the theorem of dominated convergence on the right hand side, finally finishes the proof. □

Having at hand this lemma, we see: if the tail σ-algebra $\mathscr{T}$ is trivial, then the random variables $X_1, ..., X_k$ are asympotically independent of X_n as n tends to infinity, hence the asymptotic entropy $h(\mu)$ is zero.

To finish the proof we have to show that the elements of the tail-σ-algebra $\mathscr{T}$ and the invariant σ-algebra $\mathscr{I}$ differ by sets of measure 0 or 1 only. This is the purpose of the rest of this section. □

First we establish a criterium for the equality of the tail and invariant σ-algebra in our setting. The following corresponding theorem is true in a more general setting, see Yves Derriennic's article [20] for its formulation and proof.

6.4. **Theorem.** Let G be a discrete group and P a Markov kernel defined by a symmetric probability measure on G. Define a function α on G by

$$\alpha(x) = \lim_{n\to\infty} ||\delta_x P^n - \delta_x P^{n+1}||.$$

The supremum of α over all $x \in G$ can be expressed by

$$\sup_{x\in G} \alpha(x) = \sup_{\mu\in\mathscr{M}_1} \lim_{n\to\infty} ||\mu P^n - \mu P^{n+1}||_{TV},$$

and can attain two values only, either 0 or 2. It is equal to 0 if and only if the tail- and the invariant σ-algebra are equal with respect to $\mathbb{P}_\mu$ for every probability measure μ on G. Note that the limits of the above sequences do exist since they are decreasing in n.

We next three lemmata will allow us to prove the above theorem.

6.5. **Lemma.** For a given irreducible Markov chain (X, P), the following statements are equivalent:

(1) For every probability measure μ on X, tail and invariant σ-algebra are equal with respect to $\mathbb{P}_\mu$.
(2) For every probability measure μ on X and every bounded tail function f, we almost surely have $f = f \circ S$ with respect to $\mathbb{P}_\mu$.
(3) For every bounded tail function f and every $x \in X$ we have $\mathbb{P}_x[f \neq f \circ S] = 0$.
(4) For every bounded tail function f there's a bounded invariant function f' such that $f = f'$ holds $\mathbb{P}_\mu$-almost surely for every probability measure μ on X.
(5) Every uniformly bounded sequence $\{g_n\}$ of functions on X for which $Pg_{n+1} = g_n$ for every n, is constant in n.

Proof. Obvously (3) and (2) are equivalent and (2) implies (1).

We show that (1) implies (2). Let μ be an arbitrary probability measure on X and f a tail function. Without loss of generality we can assume f to be the characteristic function of a tail event F. By (1) there is an invariant event F' such that F and F' agree up to sets of $\mathbb{P}_\mu$-measure 0, that is $\mathbb{P}_\mu(F \triangle F') = 0$. Since

$$\begin{aligned}\mathbb{P}_\mu(S^{-1}(F) \triangle S^{-1}(F')) = \mathbb{P}_\mu(X \times (F \triangle F')) &= \sum_{x\in X} \mu(x)\mathbb{P}_x(X \times (F \triangle F')) \\ &= \sum_{x,y\in X} \mu(x)P(x,y)\mathbb{P}_y(F \triangle F') \\ &= \sum_{y\in X} \mu P(y)\mathbb{P}_y(F \triangle F') = \mathbb{P}_{\mu P}(F \triangle F')\end{aligned}$$

we see that $\mathbb{P}_\mu(S^{-1}(F) \triangle S^{-1}(F')) = 0$ in case μP is absolutely continuous with respect to μ. In case it is not, we replace μ by $\mu' = \sum_{n\geq 0} \frac{1}{2^{n+1}} \mu P^n$ and observe that $\mu' P$ is absolutely continuous with respect to μ' and $\mathbb{P}_\mu$ with respect to $\mathbb{P}_{\mu'}$. In any case there is an invariant event F' such that

$$\mathbf{1}_F \circ S = \mathbf{1}_{S^{-1}(F)} = \mathbf{1}_{S^{-1}(F')} = \mathbf{1}_{F'} = \mathbf{1}_F$$

holds almost surely with respect to $\mathbb{P}_\mu$.

To see that (4) implies (2) one argues similarly as above. Details are left to the reader.

To see that (5) implies (4) we use the existence of an invertible correspondence between equivalence classes of bounded tail functions f and sequences $\{g_n\}$ of functions on X having the property that $Pg_{n+1} = g_n$ and $\sup_n \sup_x g_n(x) < \infty$. Recall that two functions are called equivalent if they agree almost everywhere with respect to $\mathbb{P}_x$ for every $x \in X$. This correspondence is given by the following assignment, which maps an equivalence class of bounded tail functions represented by f to the sequence $\{g_n\}$ whose elements are defined by

$$g_n(x) = \int_{X^{\mathbf{N}}} f \circ S^{-n}\, d\mathbb{P}_x.$$

The function $f \circ S^{-n}$ maps a sequence $x_0, x_1, \ldots$ in X to $f(y_0, \ldots, y_{n-1}, x_0, \ldots)$ where $y_0, \ldots y_{n-1}$ are arbitrary elements of X; this is well defined since f is a tail function. The inverse of this assignment maps a sequence $\{g_n\}$ of the above kind to

$$\liminf_{n\to\infty} g_n(X_n).$$

We first show that given a bounded tail function f, the construction of g_n does only depend on its equivalence class: let f' be another bounded tail function such that $A = \{f \neq f'\} \in \mathscr{T}$ has $\mathbb{P}_x$-measure zero for every $x \in X$ and consider $\{f \circ S^{-1} \neq f' \circ S^{-1}\} = S(A) \in \mathscr{T}$. Since

$$0 = \mathbb{P}_x(A) = \sum_{y\in X} P(x,y)\mathbb{P}_y(S(A))$$

we see: if $P(x,y) > 0$, then $\mathbb{P}_y(S(A)) = 0$. By irreducibility, for every $y \in X$ there is some $x \in X$ for which $P(x,y) > 0$. Hence, $\mathbb{P}_y(S(A)) = 0$ for every $y \in X$.

Let us have a closer look at the sequence $\{g_n\}$ constructed from a bounded tail function f. Of course, g_n is bounded by the the essential supremum of f uniformly in n. To see that $Pg_{n+1} = g_n$ assume, for simplicity, that f is the characteristic function of a tail event A. Then

$$\begin{aligned} Pg_{n+1}(x) &= \sum_{y\in X} P(x,y) \int_{X^{\mathbb{N}}} \mathbf{1}_A \circ S^{-n-1}\, d\mathbb{P}_y = \sum_{y\in X} P(x,y)\mathbb{P}_y(S^{n+1}(A)) \\ &= \mathbb{P}_x(X \times S^{n+1}(A)) = \mathbb{P}_x(S^n(A)) = \int_{X^{\mathbb{N}}} \mathbf{1}_A \circ S^n\, d\mathbb{P}_x \\ &= g_n(x). \end{aligned}$$

The sequence $(g_n(X_n))$ is a bounded martingale, since

$$\begin{aligned} \mathbb{E}_x[g_{n+1}(X_n+1)|X_n,...,X_0] &= \sum_{y\in X} g_{n+1}(y)\mathbb{P}_x(X_{n+1} = y|X_n,...,X_0) \\ &= \sum_{y\in X} P(X_n,y)g_{n+1}(y) = g_n(X_n), \end{aligned}$$

hence $\mathbb{P}_x$-almost surely convergent by the martingale convergence theorem. Obviously, the limit is a bounded tail function.

We now show that these two assignments are mutual inverse. Given a bounded tail function f, we notice that

$$g_n(X_n) = \mathbb{E}_{X_n}[f \circ S^{-n}] = \mathbb{E}_x[f|X_n,...,X_0],$$

and

$$\lim_{n\to\infty} \mathbb{E}_x[f|X_n,...,X_0] = \mathbb{E}_x[f|X_0,X_1,...] = f,$$

see the proof of Theorem 5.2 for more details.

For a sequence $\{g_n\}$ of the above kind, let f be the corresponding bounded tail function, and $\{\tilde{g}_n\}$ the sequence constructed from f. Then

$$\tilde{g}_n(x) = \int_{X^{\mathbb{N}}} f \circ S^{-n}\, d\mathbb{P}_x = \lim_{k\to\infty} \int_{X^{\mathbb{N}}} g_k(X_{k-n})\, d\mathbb{P}_x = g_n(x),$$

where the second equality follows from the theorem of dominated convergence and the third one from $(g_k(X_{k-n}))_{k\geq n}$ being a martingale.

Under the assumption of (5), the sequence $\{g_n\}$, which is constructed from a given bounded tail function f, is constant in n, that is, there is a function g on X such that $g_n = g$ for every n. Hence,

$$f \circ S = \lim_{n\to\infty} g(X_n) \circ S = \lim_{n\to\infty} g(X_{n+1}) = f$$

holds almost surely with respect to $\mathbb{P}_x$ for every $x \in X$.

To see that (2) implies (5), take a function g for which there is a sequence of functions of the above kind. Then there is a bounded tail function f such that $g(x) = \int_{X^{\mathbb{N}}} f\, d\mathbb{P}_x$ for every $x \in X$. By (2) the function f is almost surely invariant; invariance easily implies harmonicity of g. □

6.6. **Lemma.** Let B be a Banach space, B^* its dual, S^* the unit ball in B^* and T a contraction on B. Then

$$\lim_{n\to\infty} ||T^n x|| = \sup\{|\langle x, x^*\rangle| \,|\, x^* \in \bigcap_{n\geq 0} T^{*n} S^*\}$$

Proof. Since T is a contraction, the sequence $||T^n x||$ is decreasing and hence converging. Using basic functional analysis we find

$$\begin{aligned} ||T^n x|| &= \sup_{x^*\in S^*} |\langle T^n x, x^*\rangle| = \sup_{x^*\in T^{*n}S^*} |\langle x, T^{*n} x^*\rangle| \\ &\geq \sup\{|\langle x, x^*\rangle| \,|\, x^* \in \bigcap_{n\geq 0} T^{*n} S^*\}, \end{aligned}$$

and hence

$$\lim_{n\to\infty} ||T^n x|| \geq \sup\{|\langle x, x^*\rangle| \,|\, x^* \in \bigcap_{n\geq 0} T^{*n} S^*\}$$

To see that the other inequality holds, let ε be greater 0. For every n there is an $x_n^* \in S^*$ such that $|\langle x, T^{*n} x_n^*\rangle| = |\langle T^n x, x_n^*\rangle| \geq ||T^n x|| - \varepsilon$. Compactness of S^* with respect to the weak topology, continuitiy of T^* with respect to this topology and hence compactness of $T^{*n}S^*$ and the fact that $T^{*(n+1)}S^*$ is contained in $T^{*n}S^*$ imply that the sequence $\{T^{*n} x_n^*\}$ has an accumulation point y^*, which is contained in $\bigcap_n T^{*n} S^*$. So we find a subsequence such that

$$|\langle x, y^*\rangle| = \lim_{k\to\infty} |\langle x, T^{*n_k} x_{n_k}^*\rangle| \geq \lim_{n\to\infty} ||T^n x|| - \varepsilon,$$

which shows that the proposed equation holds. □

6.7. **Lemma.** Given a random walk on a discrete group G driven by a symmetric probability measure, the following series of equalities holds

$$\lim_{n\to\infty} ||\mu P^n|| = ||\mathbb{P}_\mu||_{\mathscr{T}} = \sup_{D^1} \int_G g \, d\mu,$$

writing μ for $\delta_x - \delta_x P$, $||.||$ for the total variation norm on G, $||.||_{\mathscr{T}}$ for the total variation norm on $\mathscr{T}$ and D^1 for the set of functions g on G for which there is a sequence (g_n) of functions on G such that $g_0 = g$ and $P g_{n+1} = g_n$ and $\sup_x g_n(x) \leq 1$.

Proof. By assumption P is reversible and every positive constant defines a positive reversible measure. Let m be the reversible measure defined by $m(x) = 1$ for every $x \in G$. By symmetry of P we have $||\mu P^n|| = ||P^n \mu||_{L^1}$ for every n, where $L^1 = L^1(m)$ is given by $L^1 = \{g : G \to \mathbb{R} \,|\, ||g||_{L^1} := \sum_{x\in G} |g(x)| < \infty\}$. Hence, we can apply the last theorem to conclude

$$\lim_{n\to\infty} ||\mu P^n|| = \sup\{\int_G g \, d\mu \,|\, g \in \bigcap_n P^n S^*\},$$

where S^* is the unit ball in $L^\infty = \{g : G \to \mathbb{R} \,|\, ||g||_{L^\infty} := \sup_x |g(x)| < \infty\}$, the dual space of L^1.

One of the stated equations follows from the fact that $D^1 = \bigcap_n P^n S^*$: obviously $D^1 \subseteq \bigcap_n P^n S^*$; the other inclusion follows from surjectivity of $P : \bigcap_n P^n S^* \to \bigcap_n P^n S^*$, see [21].

Equality between $||P_\mu||_{\mathscr{T}}$ and $\sup_{D^1} \int_{G^{\mathbb{N}}} g \, d\mu$ follows from the canonical correspondence between elements of D^1 and tail functions which are bounded by 1: let g be an element of D^1 and f the corresponding tail function (which is bounded by 1), then

$$\int_G g \, d\mu = \sum_{x \in G} \mathbb{E}_{\mathbb{P}_x}[f] \, \mu(x) = \mathbb{E}_{\mathbb{P}_\mu}[f].$$

By the theorem of Hahn and Jordan for the decomposition of signed measures, we can write $\Omega = G^{\mathbb{N}}$ as the disjoint union of two sets Ω^- and Ω^+, both measurable with respect to $\mathscr{T}$, such that $\mathbb{P}_\mu(A) \leq 0$ for $A \subseteq \Omega^-$ and $\mathbb{P}_\mu(B) \geq 0$ for $\mathbb{B} \subseteq \Omega^+$. This decomposition is unique up to sets of $\mathbb{P}_\mu$-measure zero. Hence, we have shown that

$$\sup_{D^1} \int_G g \, d\mu = ||P_\mu||_{\mathscr{T}},$$

and finished the proof. □

We are finally able to prove Theorem 6.4.

Proof. In case $\mathscr{T}$ and $\mathscr{I}$ are equal with respect to $\mathbb{P}_\mu$, μ being a probability measure on G, the probability measures $\mathbb{P}_\mu$ and $\mathbb{P}_{\mu P}$ agree on $\mathscr{T}$. To see this, let A be a tail event. We then have

$$\begin{aligned}\mathbb{P}_\mu(A) &= \mathbb{E}_{\mathbb{P}_\mu}[\mathbf{1}_A] = \mathbb{E}_{\mathbb{P}_\mu}[\mathbf{1}_A \circ S] = \mathbb{E}_{\mathbb{P}_\mu}[\mathbf{1}_{G \times A}] = \mathbb{P}_\mu[G \times A] \\ &= \sum_{x \in G} \mu(x) \mathbb{P}_x[G \times A] = \sum_{x,y \in G} \mu(x) P(x,y) \mathbb{P}_y(A) = \sum_{y \in G} \mu P(y) \mathbb{P}_y(A) \\ &= P_{\mu P}(A),\end{aligned}$$

where the second equation holds because of Lemma 6.6. So, if this is true for every probability measure μ on G, in particular for $\mu = \delta_x$ for $x \in G$, the last theorem shows that

$$\alpha(x) = \lim_{n \to \infty} ||(\delta_x - \delta_x P) P^n|| = ||\mathbb{P}_{\delta_x} - \mathbb{P}_{\delta_x P}||_{\mathscr{T}} = 0.$$

Conversely, assume there is a probability measure μ on G such that $\mathscr{T}$ and $\mathscr{I}$ do not agree with respect to $\mathbb{P}_\mu$ - again, without loss of generality we assume μP to be absolutely continuous with respect to μ. By Lemma 6.6 there is a tail event F' such that $\mathbb{P}_\mu(F' \triangle S^{-1}(F')) > 0$. Let F be the tail event which is defined by $F = F' - S^{-1}(F')$ if $\mathbb{P}_\mu(F' - S^{-1}(F')) > 0$ and by $F = F' - S(F')$ if this is not the case. In any case $\mathbb{P}_\mu(F) > 0$ and $F \cap S^{-1}(F) = \emptyset$. The first statement is true, since

$$\begin{aligned}\mathbb{P}_\mu(S^{-1}(F') - F') &= \mathbb{P}_\mu(X \times F' - F') = \mathbb{P}_\mu(X \times F' - X \times S(F')) \\ &= \mathbb{P}_{\mu P}(F' - S(F')).\end{aligned}$$

The second statement can be proved by elementary set theoretic manipulations, and is left to the reader.

Define a tail function f by $f = \mathbf{1}_F - \mathbf{1}_{S^{-1}(F)}$ and consider the corresponding sequene (g_n) of functions on G defined by $g_n(x) = \mathbb{E}_x[f \circ S^{-n}]$ for $x \in G$. We already know that $\mathbb{P}_\mu$-almost surely we have

$$\lim_{n\to\infty} g_n(X_n) = f$$

and

$$\lim_{n\to\infty} g_{n+1}(X_n) = f \circ S^{-1}.$$

Since $|f \circ S^{-1} - f| = 2$ on F (which follows from $F \cap S^{-1}(F) = \emptyset$ and F being a tail event) and F does not have $\mathbb{P}_\mu$-measure zero, for every $\varepsilon > 0$ there is an $x_0 \in G$ and an l such that $|g_{l+1}(x_0) - g_l(x_0)| > 2 - \varepsilon$. Since the sequence (g_n) satisfies $||g_n|| \leq 1$ and $Pg_{n+1} = g_n$ we see - using again the last theorem - that

$$\begin{aligned}\alpha(x_0) &= \lim_{k\to\infty} ||(\delta_{x_0} - \delta_{x_0}P)P^k|| = \sup_{D^1} \int_G g\, d(\delta_{x_0} - \delta_{x_0}P) \\ &\geq |g_{l+1}(x_0) - g_l(x_0)| > 2 - \varepsilon,\end{aligned}$$

hence $\sup_x \alpha(x)$ has to be equal to 2. □

The next theorem allows to deduce triviality of the invariant σ-algebra in the setting of Theorem 6.1.

6.8. **Theorem.** Let G be a discrete group, and P the Markov kernel on G defined by a symmetric probability measure whose support generates all of G. Furthermore assume P to be aperiodic. In this case the tail and invariant σ-algebra of the corresponding random walk on G agree with respect to $\mathbb{P}_\mu$ for any starting distribution μ.

Proof. By the last theorem it is sufficient to show that

$$\sup_{x\in G} \alpha(x) < 2,$$

where α is defined as above.

We first show that aperiodicity of P implies that for all $x, y \in G$ there is an integer $n_0 = n_0(x,y)$ such that $p^{(n)}(x,y) > 0$ for all $n > n_0$. Fix $x, y \in G$. Define a set M by $M = \{n \mid p^{(n)}(y,y) > 0\}$. This set is closed under addition, since $p^{(n+m)}(y,y) \geq p^{(n)}(y,y)p^{(m)}(y,y)$, and its greatest common divisor is 1, by aperiodicity. So, M contains all integers exceeding some n_1. By irreducibility there is an r such that $p^{(r)}(x,y) > 0$; hence for $n > n_0 := n_1 + r$ we have $p^{(n)}(x,y) \geq p^{(r)}(x,y)p^{(n-r)}(y,y) > 0$.

Choosing $x = y$, we in particular find a positive interger k such that

$$p^{(k)}(x,x) \cdot p^{(k+1)}(x,x) > 0.$$

Here, k is independent of x by the construction of the Markov chain. Define $\varepsilon := \min\{p^{(k)}(x,x), p^{(k+1)}(x,x)\} > 0$ for a fixed k of the above kind, and notice that $p^{(n)}(x,y) \geq p^{(k)}(x,x) \cdot p^{(n-k)}(x,y) \geq \varepsilon \cdot p^{(n-k)}(x,y)$ and $p^{(n+1)}(x,y)$

$\geq \varepsilon \cdot p^{(n-k)}(x,y)$. For the total variation norm of $\delta_x P^{n+1} - \delta_x P^n$ we now easily get

$$||\delta_x P^{n+1} - \delta_x P^n|| = \sum_{y \in G} |p^{(n+1)}(x,y) - p^{(n)}(x,y)| \leq 2 - \varepsilon.$$

Since this upper bound holds for all $n \geq k$ and every $x \in G$, $2 - \varepsilon$ also is an upper bound for $\sup_x \alpha(x)$. □

7. MARTIN BOUNDARY

In this section, we introduce another, more geometric description of the asymptotic behaviour of an irreducible and transient Markov chain. Given a Markov chain (X,P) of this kind, we construct a compactification $\widehat{X}$ of X and a random variable X_∞ on $\widehat{X}\setminus X$ the Markov chain converges to, in the topology of the compactification, almost surely with respect to $\mathbb{P}_x$ for every $x\in X$. The proof of this statement will be the content of the first part of this section. In the second part we show how this relates to potential thory. We show that every non-negative super-harmonic function can be represented as the integral of a function with respect to some Borel-measure on the compactification of the state space. We discuss under which condition the corresponding Borel-measure turns out to be unique and deduce a characterization of the triviality of the boundary in terms of the non-existence of non-constant bounded harmonic functions.

7.1. **Definition.** A compactification of a set X is a compact Hausdorff space $\widehat{X}$ which contains X as a dense and discrete subset.

7.2. **Theorem.** Given a countable family $\mathfrak{F}$ of bounded functions on an infinite but countable set X, there is a unique - unique up to homeomorphism - compactification $\widehat{X}=\widehat{X}_{\mathfrak{F}}$ of X such that every function in $\mathfrak{F}$ can be extended to a continuous function on $\widehat{X}$ and for all $\xi,\eta\in\widehat{X}\setminus X$, $\xi\neq\eta$ there is an $f\in\mathfrak{F}$ such that $f(\xi)\neq f(\eta)$ for its extension.

Proof. We only sketch the proof. Define $\mathfrak{F}^*=\mathfrak{F}\cup\{\mathbf{1}_x\,|\,x\in X\}$, where $\mathbf{1}_x$ denotes the characteristic function of $x\in X$. By assumption, for every $f\in\mathfrak{F}^*$ there is a constant C_f such that $|f|\leq C_f$. Consider $\Pi_{\mathfrak{F}}=\prod_{f\in\mathfrak{F}^*}[-C_f,C_f]=\{\phi:\mathfrak{F}^*\to\mathbb{R}\,|\,|\phi(f)|\leq C_f\}$ provided with the topology of pointwise convergence. This is a compact Hausdorff space; X can be considered as a subset. Consider the map $\iota:X\to\Pi_{\mathfrak{F}^*}$, which maps x to ϕ_x, defined by $\phi_x(f)=f(x)$ for $f\in\mathfrak{F}^*$; ι is injective and its image $\iota(X)$ is discrete in $\Pi_{\mathfrak{F}}$. Defining $\widehat{X}$ as the closure of $\iota(X)$ in $\Pi_{\mathfrak{F}}$ yields a compactification of X which contains X as a dense and discrete subset.

Every function $f\in\mathfrak{F}$ can be extended to a continuous function $f:\widehat{X}\to\mathbb{R}$ by setting $f(\xi)=\xi(f)$ for $\xi\in\widehat{X}\setminus X$. For $\xi,\eta\in\widehat{X}\setminus X$, $\xi\neq\eta$, there is an $f\in\mathfrak{F}^*$ such that $f(\xi)=\xi(f)\neq\eta(f)=f(\eta)$; since $\xi,\eta\in\widehat{X}\setminus X$ we have $\xi(\mathbf{1}_x)=0=\eta(\mathbf{1}_x)$, so f has to be an element of $\mathfrak{F}$, which proves the first part of the theorem. For the proof of the uniqueness of this construction, see [5]. □

7.3. **Remark.** The above constructed compactification is metrizable, see the general criteria for metrizability; a concrete metric can be constructed in the following way: for every $f\in\mathfrak{F}$ choose a weight $w_f>0$ such that $\sum_{f\in\mathfrak{F}^*}w_fC_f<\infty$. For $x,y\in X$ define their distance $d(x,y)=\sum_{f\in\mathfrak{F}^*}w_f|f(x)-$

$f(y)|$. The completion of (X,d) is homeomorphic with the above constructed compactification, see [5] for more details.

Given an irreducible and transient Markov chain (X,P), the Martin compactification of its state space is the following specialization of the above construction: for fixed $o \in X$, $\mathfrak{F}$ is chosen to be the set of functions

$$K(x,.) = \frac{F(x,.)}{F(o,.)},$$

where x ranges over all elements of X. Recall that $F(x,y) = \mathbb{P}_x[s^y < \infty]$, where $\mathbf{s}^y = \min\{n \geq 0 \,|\, X_n = y\}$. Since $F(o,y) \geq F(o,x)F(x,y)$, we see that $K(x,.)$ is bounded by $1/F(o,x)$. The corresponding compactification $\widehat{X} := \widehat{X}(P) := \widehat{X}_{\mathfrak{F}}$ is called the Martin compactification of (X,P) and $\mathfrak{M} := \widehat{X} \setminus X$ is called the Martin boundary of (X,P).

7.4. **Remark.** For recurrent Markov chains this construction is not of great interest. In this case $K(x,.) \equiv 1$ for every $x \in X$, hence the boundary consists of a single point only and does not contain much information. Recall that recurrence of a state x was characterized by the equation $U(x,x) = \mathbb{P}_x[\exists\, n : X_n = x] = 1$. We show that recurrence and irreducibility of the chain imply that $U(y,x) = \mathbb{P}_y[\exists\, n : X_n = x] = 1$ for any $y \in X$. This is done via induction on the length n of a minimal path connecting x and y. If $n = 0$, then $x = y$ and $U(x,y) = 1$ by assumption; let $n > 0$, and $w \in X$ the state which lies on the path directly before x. According to the induction hypothesis $U(w,x) = 1$. Then

$$1 = U(w,x) = p(w,x) + \sum_{v \neq x} p(w,v)U(v,x)$$

and hence

$$0 = \sum_{v \neq x} p(w,v)(1 - U(v,x)) \geq p(w,y)(1 - U(y,x)) \geq 0.$$

Since $p(w,y) > 0$, this means that $U(y,x) = 1$. Obviously, $F(x,y) = 1$ for $x = y$, and $F(x,y) = U(x,y)$ for $x \neq y$, hence $K(x,.)$ is identically 1 for every $x \in X$.

The next aim is to prove the following theorem.

7.5. **Theorem.** Given an irreducible and transient Markov chain $X_0, X_1, \ldots$ there is a random variable X_∞, taking values in the Martin boundary only, such that in the topology of the Martin compactification

$$\lim_{n \to \infty} X_n = X_\infty$$

almost surely with respect to $\mathbb{P}_x$ for every $x \in X$.

For technical reasons we are going to consider the more general case of sub-stochastic Markov chains. A map $P : X \times X \to [0,1]$ is called sub-stochastic, if $\sum_y P(x,y) \leq 1$ for every $x \in X$; the pair (X,P) is called a sub-stochastic Markov chain. Introducing the so-called tomb state † and extending P to $X \cup \{\dagger\}$ by $P(x,\dagger) = 1 - \sum_{y \in X} P(x,y)$, $P(\dagger,x) = 0$ and $P(\dagger,\dagger) = 1$ for every $x \in X$, P becomes stochastic on the extended state space.

Define a random time $\varepsilon = \mathbf{t}^\dagger - 1$, the last time before the tomb state is visited, a subset $\Omega_\varepsilon = \Omega_\infty \cup \Omega_\varepsilon$ of $X^{\mathbb{N}}$, where Ω_∞ and Ω_ε are defined by

$$\Omega_\infty = \{(x_n) \,:\, \exists\, x_\infty \in \mathfrak{M} \text{ s.t. } x_n \to x_\infty\}$$

and

$$\Omega_\dagger = \{(x_n) : \exists k \geq 1 \text{ s.t. } x_n \in X \Leftrightarrow n \leq k\},$$

and a function $X_\varepsilon : \Omega_\varepsilon \to \widehat{X}(P)$ by

$$X_\varepsilon(\omega) = \begin{cases} X_{\varepsilon(\omega)}(\omega) & \text{if} \quad \omega \in \Omega_\dagger \\ \lim_{n\to\infty} x_n & \text{if} \quad \omega = (x_n) \in \Omega_\infty. \end{cases}$$

In this setting the last theorem reads as follows.

7.6. **Theorem.** Given an irreducible, transient and sub-stochastic Markov chain (X,P), the above defined function $X_\varepsilon : \Omega_\varepsilon \to \widehat{X}(P)$ is a random variable and

$$\lim_{n\to\varepsilon} X_n = X_\varepsilon$$

in the topology of $\widehat{X}(P)$, almost surely with respect to $\mathbb{P}_x$ for every $x \in X$.

The proof uses the so-called up-crossing lemma, which we recall here. See [5] for a proof.

7.7. **Lemma.** Given a non-negative super-martingale (M_n), for any interval $[a,b] \subseteq [0,\infty)$ the expectation of the number of down-crossings with respect to $[a,b]$ is bounded by the reciprocal of the length of the interval times the expectation of M_0, that is

$$\mathbb{E}[D_\downarrow((M_n)|[a,b])] \leq \frac{1}{b-a}\mathbb{E}[M_0].$$

This lemma will allow to control the number of up-crossings of the random sequence $(K(x,X_n))$ for any $x \in X$; since $(K(x,X_n))$ itself is bounded this will imply convergence of $(K(x,X_n))$ for any $x \in X$ and hence convergence of (X_n) with respect to the topology of the Martin compactification.

We recall the definition of the Green function $G(x,y) = \sum_{n\geq 0} \mathbb{P}_x[X_n = y]$ and note that for any fixed $x \in X$ the function $G(x,.)$ defines a measure on X. Using these measures we can construct an excessive measure: let μ be some probability measure on X and define

$$\nu(y) := \mu G(y) := \sum_x \mu(x)G(x,y) \leq \sum_x \mu(x)G(y,y) = G(y,y) < \infty.$$

The first inequalitiy follows from $G(x,y) = F(x,y)G(y,y)$ and $F(x,y) \leq 1$ and the second one from the Markov chain being transient. Excessivity of ν means that $\nu P \leq \nu$ holds, which indeed is true since $\nu P = \mu GP = \mu \sum_{n\geq 1} P^n = \mu(G-I) = \mu G - \mu$. We write $\nu = f_\mu(y)G(o,y)$, where $f_\mu(y) = \sum_{x\in X} \mu(x)K(x,y)$ can be interpreted as the density of ν with respect to $G(o,.)$. Do we have to extend f_μ to $X \cup \{\dagger\}$? Furthermore, for a finite subset V of X, which contains the reference point o, we define the random time $\varepsilon_V = \sup\{n \mid X_n \in V\}$, which is $\mathbb{P}_o$-almost surely finite by transience of (X,P).

For $n \geq 0$ and $\omega \in \Omega_V = \{\omega \in X^{\mathbb{N}} \mid \varepsilon_V(\omega) < \infty\}$ we define

$$X_{\varepsilon_V - n}(\omega) = \begin{cases} X_{\varepsilon_V(\omega)-n}(\omega), & \text{if} \quad n \leq \varepsilon(\omega) \\ \dagger, & \text{if} \quad n > \varepsilon(\omega). \end{cases}$$

Almost surely we have

$$\begin{aligned} &D^{\uparrow}(f_\mu(X_0), f_\mu(X_1), ..., f_\mu(X_{\varepsilon_V}) \mid [a,b]) \\ &\quad = \lim_{N\to\infty} D_{\downarrow}(f_\mu(X_{\varepsilon_V}), f_\mu(X_{\varepsilon_V - 1}), ..., f_\mu(X_{\varepsilon_V - N}) \mid [a,b]) \end{aligned}$$

Before we are able to apply the up-crossing lemma we prove the following two statements.

7.8. **Proposition.** Firstly, the sequence $f_\mu(X_{\varepsilon_V}), f_\mu(X_{\varepsilon_V - 1}), ..., f_\mu(X_{\varepsilon_V - N})$ is a super-martingale with respect to $X_{\varepsilon_V}, X_{\varepsilon_V - 1}, ..., X_{\varepsilon_V - N}$, and secondly, $\mathbb{E}_o[f_\mu(X_{\varepsilon_V})] \leq 1$.

Proof. Since

$$\begin{aligned} \mathbb{P}_x[X_{\varepsilon_V} = y] &= \sum_{n\geq 0} \mathbb{P}_x[\varepsilon_V = n, X_n = y] = \sum_{n\geq 0} p^{(n)}(x,y)\mathbb{P}_y[\varepsilon_V = 0] \\ &= G(x,y)\mathbb{P}_y[\varepsilon_V = 0], \end{aligned}$$

we have

$$\begin{aligned} \mathbb{E}_o[f_\mu(X_{\varepsilon_V})] &= \sum_{y\in V} f_\mu(y)\mathbb{P}_o[X_{\varepsilon_V} = 0] = \sum_{y\in V} f_\mu(y)G(o,y)\mathbb{P}_y[\varepsilon_V = 0] \\ &= \sum_{y\in V} \mu G(y)\mathbb{P}_y[\varepsilon_V = 0] = \sum_{x\in X} \mu(x) \sum_{y\in V} G(x,y)\mathbb{P}_y[\varepsilon_V = 0] \\ &= \sum_{x\in X} \mu(x) \sum_{y\in V} \mathbb{P}_x[X_{\varepsilon_V} = y] \leq 1, \end{aligned}$$

which proves the first statement of the proposition. To prove the second one we choose $x_0 \in V$ and $x_1, ..., x_n \in X$ and notice that

$$\begin{aligned} &\mathbb{P}_o[X_{\varepsilon_V} = x_o, X_{\varepsilon_V - 1} = x_1, ..., X_{\varepsilon_V - 1} = x_n] \\ &\quad = G(o,x_n)p(x_n, x_{n-1}) \cdot ... \cdot p(x_1, x_0)\mathbb{P}_{x_0}[\varepsilon_V = 0], \end{aligned}$$

and hence

$$\begin{aligned}
&\sum_{x\in X} f_\mu(x)\mathbb{P}_o[X_{\varepsilon_V}=x_0,\ldots,X_{\varepsilon_V-n+1}=x_n-1,X_{\varepsilon_V-n=x}]\\
&\quad=\sum_{x\in X} f_\mu(x)G(o,x)p(x,x_{n-1})\cdot\ldots\cdot(x_1,x_0)\mathbb{P}_{x_0}[\varepsilon_V=0]\\
&\quad=\left(\sum_{x\in X}\mu G(x)p(x,x_{n-1})\right)p(x_{n-1},x_{n-2})\cdot\ldots\cdot p(x_1,x_0)\mathbb{P}_{x_0}(\varepsilon_V=0)\\
&\quad\leq(\mu G(x_{n-1}))p(x_{n-1},x_{n_2})\cdot\ldots\cdot p(x_1)\mathbb{P}_{x_0}(\varepsilon_V=0)\\
&\quad=f_\mu(x_{n-1})\mathbb{P}_o(X_{\varepsilon_V}=x_0,\ldots,X\varepsilon_V-n+1=x_{n-1}),
\end{aligned}$$

where we used excessivity of $\nu=\mu G$ in the last inequality. □

As a corollary we get an upper bound for the expected number of upcrossings of the random sequence $(f_\mu(X_n))$ with respect to any intervall $[a,b]$, that is

$$\mathbb{E}_o[D^\uparrow((f_\mu(X_n))_{n\leq\varepsilon}\mid[a,b])]\leq\frac{1}{b-a}.$$

Proof. Let V be a finite subset of X which contains the reference point o; using monotone convergence and the last proposition we are able to deduce that

$$\mathbb{E}_o[D^\uparrow(f_\mu(X_0),f_\mu(X_1),\ldots,f_\mu(X_{\varepsilon_V})\mid[a,b])]\leq\frac{1}{b-a}\mathbb{E}_o[f_\mu(X_{\varepsilon_V})]\leq\frac{1}{b-a}$$

Taking a sequence (V_k) of finite subsets of X such that $V_k\subseteq V_{k+1}$ and $\bigcup V_k=X$ and $o\in V_k$, we have $\lim_{k\to\infty}\varepsilon_{V_k}=\varepsilon$ and

$$\lim_{k\to\infty}D^\uparrow((f_\mu(X_n))_{n\leq\varepsilon_{V_k}}\mid[a,b])]=D^\uparrow((f_\mu(X_n))_{n\leq\varepsilon}\mid[a,b])].$$

To finish the proof of the corollary we take expectations on both sides of this equation and use the theorem of monotone convergence. □

Proof. We now prove the theorem about convergence to the boundary, that is Theorem 7.6; doing so we start with proving the measurability of $\Omega_\varepsilon=\Omega_\infty\cup\Omega_\dagger$. Obviously $\Omega_\dagger$ is a measurable set since it can be written as a countable union of cylinder sets. To see measurability of Ω_∞ we write $\Omega_\infty=\bigcap_x\Omega_x$, where $\Omega_x=\{(x_n)\mid\lim_{n\to\infty}K(x,x_n)\text{ exists in }\mathbb{R}\}$ and $\Omega_x=\bigcap_{a,b\in\mathbb{Q},a<b}A_x([a,b])$, where $A_x([a,b])=\{(x_n)\mid D_\downarrow(K(x,x_n)|[a,b])<\infty\}$. Since $\Omega\setminus A_x([a,b])$ can be written as $\{(x_n)\mid D_\downarrow((K(x,x_n))|[a,b])=\infty\}=\bigcap_k\bigcup_{l,m\geq k}(\{(x_n)\mid K(x,x_l)\geq b\}\cap\{(x_n)\mid K(x,x_m)\leq a\})$ and for fixed l and m the sets $\{(x_n)\mid K(x,x_l)\geq b\}$ and $\{(x_n)\mid K(x,x_l)\leq a\}$ are countable unions of cylinder sets this shows measurability of Ω_∞.

We next show that $\mathbb{P}_x(\Omega_\varepsilon)=1$ for every $x\in X$, which can be done quite quickly: consider the last corollary in case $\mu=\delta_x$, that is, the case in which $f_\mu(y)=K(x,y)$, and conclude that $\lim_{n\to\varepsilon}K(x,X_n)$ exists $\mathbb{P}_o$-almost surely for every x, which means that (X_n) converges $\mathbb{P}_o$-almost surely in the topology of $\widehat{X}(P)$ or $\mathbb{P}_o(\Omega_\varepsilon)=1$. By irreducibility of the Markov chain, for

every $x \in X$ there is a finite sequence $0, y_1, ..., y_{k-1}, x$ from o to x such that $p(o,y_1), .., p(y_{k-1}, x) > 0$; using the identity

$$p(o,y_1) \cdot ... \cdot p(y_{k-1},x)\, \mathbb{P}_x(\Omega \setminus \Omega_\varepsilon) = \mathbb{P}_o(C(o,y_1,...,y_{k-1},x) \cap (\Omega \setminus \Omega_\varepsilon)),$$

we see that almost sure convergence holds with respect to all of the probability measures $\mathbb{P}_x$, $x \in X$, since it does so with respect to $\mathbb{P}_o$.

Finally, we show that $X_\varepsilon : \Omega_\varepsilon \to \widehat{X}(P)$ indeed is a random variable, i.e. measurable with respect to the Borel-σ-algebra on $\widehat{X}(P)$. Obviously, this is true for the restriction of X_ε to $\Omega_\dagger$. To see the measurability of the restriction of X_ε to Ω_∞ with respect to the Borel-σ-algebra of the Martin boundary $\mathfrak{M}$, it is sufficient to show measurability of the sets $\{X_\varepsilon \in B_{x,\xi,\varepsilon}\}$, where $B_{x,\xi,\varepsilon} = \{\eta \in \mathfrak{M} \,|\, |K(x,\eta) - K(x,\xi)| < \varepsilon\}$ for fixed $x \in X, \xi \in \mathfrak{M}, \varepsilon > 0$, since the collection of sets of this form represents a basis for the topology of $\widehat{X}(P)$. This is done analogously as in the first part of the proof. □

We next prove the so-called Poisson-Martin integral representation theorem which states that every non-negative superharmonic function $h : X \to \mathbb{R}$ can be written as an integral of the form

$$h = \int_{\widehat{X}} K(.,\xi) d\nu^h(\xi),$$

where ν^h is a Borel-measure on $\widehat{X}$. In case h is harmonic, the support of ν^h is contained in $\mathfrak{M}$, see the corresponding theorem below for more details. We will give conditions under which this representation is unique and use it to deduce a criterium for triviality of the the asymptotic behaviour in terms of bounded harmonic functions.

The central role is played by the limit distributions ν_x, $x \in X$ of the random variable X_ε:

$$\nu_X(B) = \mathbb{P}_x(X_\varepsilon \in B),$$

B being a Borel-subset of $\widehat{X}$. The next proved property - mutual absolute continuity - is going to be crucial in the following.

7.9. **Theorem.** The measure ν_x is absolutely continuous with respect to ν_o and a realization of its Radon-Nikodym derivative is given by

$$\frac{d\nu_x}{d\nu_o} = K(x,.)$$

Proof. Let V be a finite subset of X, and $0, x \in V$ and $\varepsilon_V = \sup\{n \geq 0 \,|\, X_n \in V\}$. By the Markov property we have

$$\mathbb{P}_x(X_{\varepsilon_V} = y) = G(x,y)\mathbb{P}_y(\varepsilon_V = 0)$$

and

$$\mathbb{P}_o(X_{\varepsilon_V} = y) = G(x,y)\mathbb{P}_y(\varepsilon_V = 0).$$

So we can write

$$\begin{aligned}\mathbb{P}_x(X_{\varepsilon_V} = y) &= \frac{G(x,y)}{G(o,y)}\mathbb{P}_o(X_{\varepsilon_V} = y)\\ &= K(x,y)\mathbb{P}_o(X_{\varepsilon_V} = y),\end{aligned}$$

where the last equation holds because of the identity $G(x,y) = F(x,y)G(y,y)$, see Lemma 2.14. For a continuous function $f : X \to \mathbb{R}$, we hence have

$$\begin{aligned}\mathbb{E}_x[f(X_{\varepsilon_V})] &= \sum_{y\in V} f(y)\mathbb{P}_x(X_{\varepsilon_V} = y)\\ &= \sum_{y\in V} f(y)K(x,y)\mathbb{P}_o(X_{\varepsilon_V} = y)\\ &= \mathbb{E}_o[f(X_{\varepsilon_V})K(x,X_{\varepsilon_V})].\end{aligned}$$

Taking an increasing sequence (V_k) of finite subsets which approaches X, we have $\lim_k X_{\varepsilon_{V_k}} = X_\varepsilon$ almost surely with respect to $\mathbb{P}_x$ and $\mathbb{P}_o$; since f and $K(x,.)$ are continuous on the compact $\widehat{X}$ and hence bounded, the theorem of dominated convergence implies

$$\mathbb{E}_x[f(X_\varepsilon)] = \mathbb{E}_o[f(X_\varepsilon)K(x,X_\varepsilon)].$$

In particular, for the characteristic function $\mathbf{1}_B$ of a Borel-subset B of $\widehat{X}$ (which easily can be approximated by continuous functions, since $\widehat{X}$ is metrizable), we have

$$\nu_x(B) = \int_B K(x,.)\, d\nu_o.$$

□

7.10. **Proposition.** Given a continuous function $f : \widehat{X} \to \mathbb{R}$, we have

$$\begin{aligned}\mathbb{E}_x[f(X_\varepsilon)] &= \sum_{y\in X} f(y)\nu_x(y) + \lim_{n\to\infty} P^n f(x)\\ &= \sum_{y\in X} f(y)G(x,y)p(y,\dagger) + \lim_{n\to\infty} P^n f(x).\end{aligned}$$

Proof. As the statement of the proposition suggests, in taking the expectation of $f(X_\varepsilon)$ one considers the components $f(X_\varepsilon)\mathbf{1}_{\Omega_\infty}$ and $f(X_\varepsilon)\mathbf{1}_{\Omega_\dagger}$ seperately (which is legitimate since $\mathbb{P}_x(\Omega_\varepsilon) = 1$). See [5] for more details. □

Before we state and prove the representation theorem for non-negative superharmonic functions we recall the notion of the support of a non-negative Borel-measure ν as the set of all elements ξ for which $\nu(V) > 0$ for every open neighbourhood V of ξ.

7.11. **Theorem.** Given an irreducible and transient and sub-stochastic Markov chain (X,P), for every non-negative superharmonic function $h : X \to \mathbb{R}$ there is a Borel-measure ν^h on the Martin compactification $\widehat{X}$ such that

$$h = \int_{\widehat{X}} K(.,\xi)\, d\nu^h(\xi).$$

In case h is harmonic, the support of ν is contained in $\widehat{X} \setminus X$.

Proof. The only case of interest is when h is not identically 0. In this case h is a strictly positive function: if this were not true, there would be an $x \in X$ such that $h(x) = 0$, but by irreducibility of P for every $y \in X$ we find a positive integer n such that $p^{(n)}(x,y) > 0$ and by super-harmonicity of h

$$h(x) \geq \sum_y p^{(n)}(x,y)h(y),$$

which shows that $h(y) = 0$ too, and hence $h \equiv 0$, contradictory to the assumption.

Consider the following modified (sub-) Markov chain, whose transtition kernel P^h is given by

$$P^h(x,y) = \frac{P(x,y)h(y)}{h(x)}.$$

The corresponding Martin kernel is given by $K^h(x,y) = K(x,y)h(o)/h(x)$, hence the Martin compactification of X is not affected by the modification of the transition probabilities, i.e. $\widehat{X}(P) = \widehat{X}(P^h)$. Writing $\tilde{\nu}_x$ and $\tilde{\nu}_o$ for the distribution of X_ε with respect to P^h and using the last theorem about mutual absolute continuity we have

$$1 = \tilde{\nu}_x(\widehat{X}) = \int_{\widehat{X}} K^h(x,.)\, d\tilde{\nu}_o.$$

Multiplying by $h(x)$ gives

$$h(x) = \int_{\widehat{X}} K(x,.)\, d\nu^h,$$

ν^h being defined by $\nu^h = h(o)\tilde{\nu}_o$.

To verify the second statement, let h be harmonic. Suppose there is some y in X which does lie in the support of ν^h, which means that $\nu^h(y) > 0$, since X is discrete in $\widehat{X}$. Writing $\nu^h = a\delta_y + \nu'$, where a is chosen such that y does not lie in the support of ν', the representation of h from above reads as

$$h(x) = aK(x,y) + \int_{\widehat{X}} K(x,.)\, d\nu'.$$

This gives the desired contradiction, since both summands are super-harmonic and the first one is strictly super-harmonic at y. □

We face the problem of finding a condition under which the above representation of a non-negative superharmonic fuction becomes unique. We need the following lemma.

7.12. **Lemma.** For two strictly positive superharmonic functions h_1, h_2 and positive real numbers a_1 and a_2, consider the linear combination $h = a_1h_1 + a_2h_2$. Recalling the definition of the measures $\nu^h, \nu^{h_1}, \nu^{h_2}$, the following relation holds:

$$\nu^h = a_1\nu^{h_1} + a_2\nu^{h_2}.$$

Proof. The proof just consists of demonstrating that the two measures agree on cylinder sets. Again, this is left to the reader. □

7.13. **Definition and Remark.** By $\mathfrak{M}_{min}$ we denote the so-called minimal Martin boundary, which consists of all $\xi \in \mathfrak{M}$ for which $K(.,\xi)$ is a minimal harmonic function; a non-negative harmonic function h is called minimal if $h(o) = 1$ and every other non-negative harmonic function h_1 with $h \geq h_1$ is a multiple of h. The minimal Martin boundary turns out to be measurable and the natural condition for a measure of the above representation to become unique is to map $\mathfrak{M} \setminus \mathfrak{M}_{min}$ to 0. Here is the explanation in which sense the term 'natural' is to be understood: the set S^+ of non-negative superharmonic functions forms a cone for which a basis B is given by $B = \{h \in S^+ \mid h(o) = 1\}$; so to understand the cone S^+, it is sufficient to understand the base B. This base is a convex set and its extremal points dB - those points which can not be written as a convex combination of others - is given by the union of $\{K(.,y) \mid y \in X\}$ and the set of all minimal harmonic functions, see Theorem 7.6 in [5]. Since every minimal harmonic function h is of the form $K(.,\xi)$ for some $\xi \in \mathfrak{M}$ - as will be proved in a few moments - this explains the above representation theorem. Uniqueness of this representation under the condition of mapping $\mathfrak{M} \setminus \mathfrak{M}_{min}$ to 0 comes from the fact of B being a simplex.

7.14. **Theorem.** For every minimal harmonic function h there is an element ξ of $\mathfrak{M}$ such that

$$h(x) = K(x,\xi) \quad \text{for all } x \in X.$$

Proof. We show that the (unique) measure ν on $\widehat{X}$ which gives rise to the integral representation of h is given by $\nu = \delta_\xi$ for some $\xi \in \mathfrak{M}$. Being a minimal harmonic function, and hence an extremal point of the base, implies that

$$\int_B h(x)\, d\nu(\eta) = \int_B K(x,\eta)\, d\nu(\eta)$$

holds for every Borel-measurable subset $B \subseteq \widehat{X}$; in particular for each $x \in X$ we have $h(x) = K(x,\eta)$ for ν-almost every η. Let A be the set of all η in $\mathfrak{M}$ (h is harmonic, therefore $\mathrm{supp}(\nu) \subseteq \mathfrak{M}$) for which this is true for all $x \in X$. This set has measure 1, since X is countable, and does consist of a single point only, since $K(.,\xi) \neq K(.,\eta)$ for $\xi, \eta \in \mathfrak{M}$, $\xi \neq \eta$ by construction of the Martin compactification, hence $\nu = \delta_\xi$ for $\xi \in A$. □

7.15. **Corollary.** Among the elements of $\mathfrak{M}$ the elements of the minimal Martin boundary $\mathfrak{M}_{min}$ are characterized by the property $\nu^{K(.,\xi)} = \delta_\xi$.

Proof. Assume ξ is an element of $\mathfrak{M}$ for which $\nu^{K(.,\xi)} = \delta_\xi$, but for which $K(.,\xi)$ is not a minimal harmonic function. We first show that $K(.,\xi)$ is an extremal element of the base of the cone S^+. If this were not the case, there would be two super-harmonic functions h_1, h_2 with $h_1(o) = 1 = h_2(o)$ and constants $a_1, a_2 \in (0,1)$ such that $h = a_1 h_1 + a_2 h_2$. Lemma 7.13 shows that $\delta_\xi = \nu^h = a_1 \nu^{h_1} + a_2 \nu^{h_2}$, hence $\nu^{h_1} = \delta_\xi = \nu^{h_2}$ (as is easily seen since ν^{h_1} and ν^{h_2} are probability measures) and hence $h_1 = K(.,\xi) = h_2$, which means that $K(.,\xi)$ indeed is an extremal element.

If $K(.,\xi)$ is not minimal, it is equal to $K(.,y)$ for some $y \in X$, see Remark 7.14. But this cannot be the case, since $\nu^{K(.,y)} = \delta_y$ is different from $\nu^{K(.,\xi)} = \delta_\xi$. To see this recall the definition of ν^h; for $x \in X$, $\nu^h(x) = G(o,x)(h(x) - Ph(x))$, which implies $\nu^h = \delta_y$ for $h = K(.,y)$. □

The above characterization allows to proof the next lemma.

7.16. **Lemma.** The minimal Martin boundary $\mathfrak{M}_{min}$ is Borel-measurable.

Proof. Measurability follows from the representation

$$\mathfrak{M}_{min} = \{\xi \in \mathfrak{M} \mid \lim_{m\to\infty} \lim_{n\to\infty} \sum_{y\in X} p^{(n)}(o,y) K(y,\xi) e^{-m\cdot d(y,\xi)} = 1\},$$

where d is a metric on $\widehat{X}$ which generates the topology of the compactification. Let $\xi \in \mathfrak{M}$ and $m \in \mathbb{Z}$ and define $h = K(.,\xi)$ and $f_m = e^{-m\cdot d(.,\xi)}$; by Proposition 7.11 we have

$$\int_{\widehat{X}} f_m \, d\nu^h = \sum_{x\in X} f_m(x) \nu^h(x) + \lim_{n\to\infty} \sum_{y\in X} p^{(n)}(o,y) h(y) f_m(y)$$

Letting m tend to infinity, the left hand side of the equation tends to $\nu^h(\xi)$ and the first summand of the right hand side tends to 0 by the theorem of dominated convergence, since f_m converges to $\mathbf{1}_{\{\xi\}}$ as m tends to infinity. So,

$$\nu^{K(.,\xi)}(\xi) = \lim_{m\to\infty} \lim_{n\to\infty} \sum_{y\in X} p^{(n)} K(y,\xi) e^{-m\cdot d(y,\xi)}.$$

The representation now follows from the last corollary. □

7.17. **Theorem.** For every non-negative super-harmonic function h, there is a measure ν on $\widehat{X}$ such that

$$h(x) = \int_{\widehat{X}} K(x,.) \, d\nu \quad \text{for every } x \in X.$$

Under the condition of

$$\nu(\mathfrak{M} \setminus \mathfrak{M}_{min}) = 0,$$

ν is unique and given by ν^h, the measure ν^h being defined as above.

Proof. We first show that ν^h maps $\mathfrak{M} \setminus \mathfrak{M}_{min}$ to 0, and start doing so by establishing the following equation

$$f(\xi) = \int_{\mathfrak{M}} f(\eta)\, d\nu^{K(.,\xi)}(\eta)$$

for ν^h-almost every ξ in $\mathfrak{M}$ and every continuous function f on $\widehat{X}$. This turns out to be very useful: consider the countable family of continuous functions $(f_{m,k})$ on $\widehat{X}$ which is defined by $f_{m,k}(x) = e^{-m \cdot d(x,\xi_k)}$ for $x \in \widehat{X}$; here m is an integer, d denotes a metric which generates the topology of $\widehat{X}$ and $\{\xi_k\}$ is a countable and dense subset of $\mathfrak{M}$. Defining $B_{m,k}$ by the set of all elements $\xi \in \mathfrak{M}$ for which the above equation does not hold for $f = f_{m,k}$, we know that $\nu^h(B_{m,k}) = 0$ and hence also $\nu(B) = 0$ for $B = \bigcup_{m,k} B_{m,k}$. The plan is to show that $\mathfrak{M} \setminus \mathfrak{M}_{min} \subseteq B$, which of course implies $\nu^h(\mathfrak{M} \setminus \mathfrak{M}_{min}) = 0$. So let ξ be an element of B, in which case we have

$$e^{-m \cdot d(\xi,\xi_k)} = \int_{\mathfrak{M}} e^{-m \cdot d(\eta,\xi_k)}\, d\nu^{K(.,\xi)}(\eta),$$

for all m and k. Taking a subsequence of $\{\xi_k\}$ which approaches ξ and letting k tend to infinity gives

$$1 = \int_{\mathfrak{M}} e^{-m \cdot d(\eta,\xi)}\, d\nu^{K(.,\xi)}(\eta),$$

and letting m tend to infinity shows that $d\nu^{K(.,\xi)}$ has to be δ_ξ. Since the elements of $\mathfrak{M}_{min}$ are characterized by the property $\nu^{K(.,\xi)} = \delta_\xi$, this means that ξ is an element of the minimal Martin boundary. To prove that

$$f(\xi) = \int_{\mathfrak{M}} f(\eta)\, d\nu^{K(.,\xi)}(\eta)$$

holds for ν^h-almost every ξ in $\mathfrak{M}$ and every continuous function f on $\widehat{X}$ it's sufficient to show that

$$\int_{\mathfrak{M}} f(\xi) g(\xi)\, d\nu^h(\xi) = \int_{\mathfrak{M}} \left(\int_{\mathfrak{M}} f(\eta)\, d\nu^{K(.,\xi)}(\eta) \right) d\nu^h(\xi)$$

holds for every continuous function g on $\widehat{X}$. This is done via examination of $\mathbb{E}_o^{K(.,\xi)}[f(X_n) g(x_{n+m}) \mathbf{1}_{\{\varepsilon \geq n+m\}}]$ as m and n tend to infinity, see [5] for more details.

Let's show uniqueness. Let h be a super-harmonic function, without loss of generality mapping o to 1 and let ν and ν^h be measures having the stated properties. The proof of Corollary 7.16 shows that for every $\xi \in \mathfrak{M}_{min}$ we have $\nu^{K(.,\xi)} = \delta_\xi$, as we have $\nu^{K(.,y)} = \delta_y$ for every $y \in X$; so, for $\eta \in$

$X \cup \mathfrak{M}_{min}$ a continuous function $f : \widehat{X} \to \mathbb{R}$ and by Proposition 7.11

$$\begin{aligned} f(\eta) &= \int_{\widehat{X}} f(\xi)\, d\nu^{K(.,\eta)} \\ &= \sum_{y\in X} f(y)G(o,y)(K(y,\eta) - \sum_{w\in X} p(y,w)K(w,\eta)) \\ &\quad + \lim_{n\to\infty} \sum_{x\in X} p^{(n)}(o,x)f(x)K(x,\eta). \end{aligned}$$

Integrating over $\widehat{X}$ with repect to ν yields, since $\nu(\mathfrak{M} \setminus \mathfrak{M}_{min}) = 0$, integral and limits can be interchanged and $\int_{\widehat{X}} K(x,\eta)\, d\nu(\eta) = h(x)$ for all $x \in X$,

$$\int_{\widehat{X}} f(\eta)\, d\nu(\eta) = \int_{\widehat{X}} f(\eta)\, d\nu^h(\eta).$$

Since this is true for every continuous function $f : \widehat{X} \to \mathbb{R}$, the two measures ν and ν^h have to agree. □

7.18. **Theorem.** Every bounded harmonic function h on X is of the form

$$h(x) = \int_{\mathfrak{M}} K(x,\xi)\varphi(\xi)\, d\nu_o = \int_{\mathfrak{M}} \varphi(\xi)\, d\nu_x,$$

for a unique (up to ν_o null-sets, of course) bounded measurable function φ on $\mathfrak{M}$; the integral is called the Poisson-integral of φ.

Proof. We only consider the case in which P is stochastic. By assumption there's a constant $c \geq 0$ such that $-c \leq h \leq c$; we write

$$2c \cdot \mathbf{1}_X = h_1 + h_2,$$

where $h_1 = c \cdot \mathbf{1}_X - h$ and $h_2 = h + c \cdot \mathbf{1}_x$ are non-negative harmonic functions, and hence

$$\nu^{h_1} + \nu^{h_2} = 2c \cdot \nu^{\mathbf{1}_X} = 2c \cdot \mathbf{1}_{\mathfrak{M}} \nu_o.$$

In particular we see that ν^{h_1} and ν^{h_2} are absolute continuous with respect to ν_o; let φ_1 and φ_2 be the corresponding Radon-Nikodym derivatives, then $\varphi_1 + \varphi_2$ is the Radon-Nikodym derivative of $\nu^{h_1} + \nu^{h_2} = 2c \cdot \mathbf{1}_{\mathfrak{M}}$ and hence

$$\varphi_1 + \varphi_2 = 2c \cdot \mathbf{1}_{\mathfrak{M}} \qquad \nu_o\text{-almost surely}.$$

So, φ_1 and φ_2 have to be ν_o-almost surely bounded. This finishes the proof, since it allows to write h as the Poisson-integral of the (ν_o-almost surely) bounded and measurable function $\varphi = \varphi_2(\xi) - c \cdot \mathbf{1}_{\mathfrak{M}}(\xi)$, that is

$$h = h_2 - c \cdot \mathbf{1}_X = \int_{\mathfrak{M}} K(.,\xi)(\varphi_2(\xi) - c \cdot \mathbf{1}_{\mathfrak{M}}(\xi))\, d\nu_o.$$

Uniquness follows from the uniqueness statement of the last theorem. □

Finally we are able to give a characterization of triviality of the Martin boundary in potential-theoretic terms. Recall that $\mathfrak{M}$ is called trivial if $\nu_o(B) \in \{0,1\}$ for every Borel set $B \subseteq \mathfrak{M}$.

7.19. **Corollary.** The Martin boundary $\mathfrak{M}$ of an irreducible and transient Markov chain is trivial if and only if every bounded harmonic function is constant.

Proof. In case $(\mathfrak{M}, \nu_o)$ is trivial every measurable function on $\mathfrak{M}$ has to be ν_o-almost surely constant. So, given a bounded harmonic function h and a bounded measurable function φ, ν_o-almost surely equal to a constant c, whose Poisson-integral is h, we get

$$h(x) = \int_{\mathfrak{M}} K(x,\xi)\varphi(\xi)\, d\nu_o = c \cdot \int_{\mathfrak{M}} K(x,\xi)\, d\nu_o = c \cdot \int_{\mathfrak{M}} d\nu_x = c$$

for all $x \in X$.

Now assume that every bounded harmonic function on X is constant. Let $\varphi = \mathbf{1}_A$ for a Borel set $A \subseteq \mathfrak{M}$ and consider it's Poisson-integral $h(x) = \int_{\mathfrak{M}} \varphi\, d\nu_x$, which by assumption is equal to a constant $c \in [0,1]$. The proof of Theorem 7.19 and the stated uniqueness therein show that $\varphi = c \cdot \mathbf{1}_{\mathfrak{M}}$ holds almost surely with respect to ν_o. Since φ takes values in $\{0,1\}$ only and $\mathbf{1}_X$ is ν_o-almost surely constant 1 on $\widehat{X}$, the constant c has to be an element of $\{0,1\}$, and hence

$$\nu_o(A) = \int_{\widehat{X}} \mathbf{1}_A\, d\nu_o = \int_{\widehat{X}} c \cdot \mathbf{1}_{\mathfrak{M}} = c \cdot \nu_o(\mathfrak{M}) = c \in \{0,1\}.$$

□

8. APPENDIX

Kingman's subadditive ergodic theorem. We showed that the notion of drift for random walks was well defined by using Kingman's subadditive ergodic theorem. In the following we present a proof of this theorem due to J. Michael Steele, see [22]. His proof relies on Birkhoff's ergodic theorem, so we first proof Birkhoff's ergodic theorem, following the presentation in Achim Klenke's book [10].

A measure-preserving dynamical system consists of a probability space and a measurable map on the underlying space which leaves the probability measure invariant, in symbols $(\Omega, \mathbf{A}, \mathbb{P}, T)$, where $(\Omega, \mathbf{A}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ a measurable map such that $\mathbb{P} \circ T^{-1} = \mathbb{P}$. Given a measurable function $f : \Omega \to \mathbb{R}$ we define random variables X_k by $X_k = f \circ T^k$ for $k \geq 0$. Birkhoff's ergodic theorem explains how the arithmetic mean of the first n of those behaves as n becomes very large. For notational convenience we abbreviate the partial sums $\sum_{k=0}^{n-1} X_k$ by S_n for $n \geq 1$.

8.1. **Lemma.** Let X_0 be an integrable random variable and define M_n by $\max\{0, S_1, ..., S_n\}$. Then, for every $n \geq 0$

$$\mathbb{E}[X_0 \, \mathbf{1}_{M_n > 0}] \geq 0.$$

Proof. We have $X_0 \geq \max\{S_1, ..., S_n\} - M_n \circ T$. Indeed for $1 \leq k \leq n$ we obviously have $X_0 + M_n \circ T \geq X_0 + S_k \circ T = S_{k+1}$, that is $X_0 \geq S_{k+1} - M_n \circ T$. And since $S_1 = X_0$ and $M_n \circ T \geq 0$ this is also true for $k = 0$, which proves the inequality. Using this and $\{M_n > 0\}^c \subseteq \{M_n = 0\} \cap \{M_n \circ T \geq 0\} \subseteq \{M_n - M_n \circ T \leq 0\}$ and T being measure-preserving, we can deduce

$$\begin{aligned}
\mathbb{E}[X_0 \, \mathbf{1}_{M_n > 0}] &\geq \mathbb{E}(\max(S_1, ..., S_n) - M_n \circ T) \, \mathbf{1}_{M_n > 0}] \\
&= \mathbb{E}(M_n - M_n \circ T) \, \mathbf{1}_{M_n > 0}] \\
&\geq \mathbb{E}[M_n - M_n \circ T] = \mathbb{E}[M_n] - \mathbb{E}[M_n] = 0.
\end{aligned}$$

□

8.2. **Theorem.** Given a measure-preserving dynamical system, the measure-preserving transformation being denoted by T, and an integrable random variable f, we have

$$\frac{1}{n} \sum_{k=0}^{n-1} X_k \to \mathbb{E}[X_0 | \mathbf{I}],$$

almost surely with respect to $\mathbb{P}$ as n tends to infinity (using the notation from above, where $X_k = f \circ T^k$). Here $\mathbf{I}$ denotes the σ-algebra of events which are invariant with respect to T, that is $\mathbf{I} = \{A \in \mathbf{A} \mid T^{-1}(A) = A\}$. In case T is ergodic, meaning that $\mathbb{P}[A] \in \{0, 1\}$ for every invariant event A, the above specializes to almost sure convergence to $\mathbb{E}[X_0]$.

Proof. Without loss of generality we can assume $\mathbb{E}[X_0 | \mathbf{I}]$ to be zero. If it were not we can replace X_n by $X_n - \mathbb{E}[X_0 | \mathbf{I}]$. In the following we want to

show that $\limsup_{n\to\infty} S_n/n$, and similarly $\liminf_{n\to\infty} S_n/n$, is almost surely zero. So let us define the random variable Z by $\limsup_{n\to\infty} S_n/n$ and the event F by $\{Z > \varepsilon\}$ for some $\varepsilon > 0$. As soon as we have shown $\mathbb{P}[F] = 0$ we are finished using the usual continuity argument. Obviously $Z \circ T = Z$, so Z is measurable with respect to $\mathbf{I}$ (which allows us to use the above lemma). Now define random variables $X_n^\varepsilon = (X_n - \varepsilon)\mathbf{1}_F$, $S_n^\varepsilon = \sum_{k=0}^{n-1} X_k^\varepsilon$, $M_n^\varepsilon = \max\{0, S_1^\varepsilon, ..., S_n^\varepsilon\}$ and events $F_n = \{M_n^\varepsilon > 0\}$. The events F_n are increasing and their union is

$$\bigcup_{n\geq 1} F_n = \{\sup_{k\geq 0} S_k^\varepsilon/k\} = \{\sup_{k\geq 0} S_k/k > \varepsilon\} \cap F = F.$$

The theorem of dominated convergence implies that $\mathbb{E}[X_0^\varepsilon \mathbf{1}_{F_n}]$ converges to $\mathbb{E}[X_0^\varepsilon]$ and applying the last lemma to X_0^ε shows $\mathbb{E}[X_0^\varepsilon] \geq 0$. Together this implies that

$$0 \leq \mathbb{E}[X_0^\varepsilon] = \mathbb{E}[(X_0 - \varepsilon)\mathbf{1}_F] = \mathbb{E}[\mathbb{E}[X_0 \mid \mathbf{I}]\, \mathbf{1}_F] - \varepsilon\mathbb{P}[F] = -\varepsilon\mathbb{P}[F],$$

so $\mathbb{P}[F]$ indeed has to be equal to 0. □

8.3. **Theorem.** Let us assume we have a probability space $(\Omega, \mathbf{F}, \mathbb{P})$, a measure preserving transformation $T : \Omega \to \Omega$ and a sequence $\{g_n\}$ of integrable random variables satisfying $g_{n+m}(x) \leq g_n(x) + g_m(T^n x)$ for almost every x and for all n and m. Then with probability one we have $\lim_{n\to\infty} g_n(x)/n = g(x) \geq -\infty$.

Before presenting the proof we'd only like to mention that Fekete's lemma - which states that for any sequence $\{\alpha_n\}$ with $\alpha_{n+m} \leq \alpha_n + \alpha_m$ the limit $\lim_{n\to\infty} \alpha_n/n$ exists - can easily be deduced from Kingman's subadditive ergodic theorem.

Proof. We start with showing that it's no restriction to assume the random variables to be smaller or equal to 0. This can be seen in the following way: we define a sequence of random variables by $g'_n(x) = g_n(x) - \sum_{i=0}^{n-1} g_1(T^i x)$, which are easily seen to be smaller or equal to 0 and to have the above subadditive property. Now Birkhoff's ergodic theorem implies that g'_n/n converges almost surely if and only if g_n/n does. So without loss of generality we can assume $g_n \leq 0$ for every n.

We next show that $g(x) = \limsup_{n\to\infty} g_n(x)/n$ is an invariant function with respect to T, that is $g(Tx) = g(x)$ for almost every x. Using subadditivity we deduce $g_{n+1}(x)/n \leq g_1(x)/n + g_n(Tx)$, and taking the limes inferior gives the first inequality $g(x) \leq g(Tx)$. So $\{g > \alpha\} \subseteq \{g \circ T > \alpha\} = T^{-1}\{g > \alpha\}$, and since T is measure preserving those two sets differ by a set of measure zero only which finally implies almost sure equality of g and $g \circ T$. So it's no restriction to assume $g(x) = g(T^k x)$ for almost every x and every non-negative integer k. Invariance of g with respect to T is actually equivalent to g being measureable with respect to the invariant σ-algebra

$\mathbf{A}$, which is defined by $\mathbf{A} = \{A \in \mathbf{F} \mid T^{-1}A = A\}$.

For given $\varepsilon > 0$, $1 < N < \infty$ and $0 < M < \infty$, we define a function $G_M(x) = \max\{-M, g(x)\}$ and a set $B(N,M) = \{x \mid g_l(x) > l(G_M(x) + \varepsilon)$ for all $1 \leq l \leq N\}$. For any $x \in \Omega$ and $n \geq N$ we decompose the integer set $[1,n)$ into a union of three classes of intervals in the following way:

If k is the least integer in $[1,n)$ which is not in an interval already constructed, consider $T^k x$. By definition, if $T^k x$ is not in $B(N,M)$, there's an $l \leq N$ such that $g_l(T^k x) \leq l(G_M(T^k x) + \varepsilon) = l(G_M(x) + \varepsilon)$. If $k + l \leq n$ we take $[k, k+l)$ as an element of our decomposition, if $k + l > n$ we take the interval $[k, k+1)$, consisting of a single integer only, and in case $x \notin B(N,M)$ the interval $[k, k+1)$ is taken again.

For any $x \in \Omega$ we arrive at a decomposition of $[1,n)$ into a set of, say u, intervals $[\tau_i, \tau_i + l_i)$ where $g_{l_i}(T^{\tau_i} x) \leq l_i(G_M(x) + \varepsilon)$ with $1 \leq l_i \leq N$, and two sets of singletons: the first one consisting of v singletons $[\sigma_i, \sigma_i + 1)$ for which $T^{\sigma_i}(x)$ lies in $B(N,M)$ and the second one of w singletons contained in $(n-N, n)$.

Using our assumptions of $\{g_n\}$ being subadditive, the decomposition of $[1,n)$, g being invariant and $g_m \leq 0$ we can derive the following bounds:

$$\begin{aligned} g_n(x) &\leq \sum_{i=1}^{u} g_{l_i}(T^{\tau_i} x) + \sum_{i=1}^{v} g_1(T^{\sigma_i} x) + \sum_{i=1}^{w} g_1(T^{n-i}) \\ &\leq (G_M(x) + \varepsilon) \sum_{i=1}^{u} l_i \leq G_M(x) \sum_{i=1}^{u} l_i + n\varepsilon \end{aligned}$$

By construction of the intervals we have

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

and applying Birkhoff's ergodic theorem we find

$$\begin{aligned} \liminf_{n\to\infty} n^{-1} \sum_{i=1}^{n} l_i &\geq \lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} 1_{B(N,M)^c}(T^k x) \\ &= 1 - \mathbb{E}[1_{B(N,M)} \mid \mathbf{A}] \quad \text{a. s.} \end{aligned}$$

By the above estimate for g_n we can conclude that for any M

$$\limsup_{n\to\infty} g_n(x)/n \leq G_M(x)(1 - \mathbb{E}[1_{B(N,M)} \mid \mathbf{A}]) + \varepsilon.$$

The definition of $B(N,M)$ ensures $1_{B(N,M)}$ to converge to 0 as N tends to infinity for any fixed M. Thus, if let N tend to infinity in the above inequality we get

$$\limsup_{n\to\infty} g_n(x)/n \leq G_M(x) + \varepsilon,$$

and since this holds for all $M > 0$ and $\varepsilon > 0$, we know

$$\limsup_{n\to\infty} g_n(x)/n \leq \liminf_{n\to\infty} g_n(x)/x,$$

for almost every x, which proves the theorem. □

Martingales. In the following we present results concerning the theory of martingales; for more details we refer to [23].

A sequence $X_1, X_2, \ldots$ of integrable random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a martingale with respect to a filtration $(\mathscr{F}_n)_{n\geq 1}$, if X_n is measurable with respect to $\mathscr{F}_n$ and $\mathbf{E}[X_{n+1}|\mathscr{F}_n] = X_n$ holds almost surely for every $n \geq 1$. The sequence is called a submartingale if the above equal sign is replaced by a greater-than-or-equal sign.

8.4. **Example.** Given an integrable random variable X on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a filtration $(\mathscr{F}_n)$, then the sequence of conditional expectations $\mathbb{E}[X|\mathscr{F}_n]$ is a martiangale with respect to $(\mathscr{F}_n)$.

The martingale convergence theorem states the following.

8.5. **Theorem.** Given a submartingale $X_1, X_2, \ldots$ such that $\sup_n \mathbb{E}[|X_n|]$ is finite, X_n almost surely converges to a random variable X. Furthermore X is integrable and $\mathbb{E}[|X|] \leq \sup_n \mathbb{E}[|X_n|]$.

8.6. **Theorem.** If $\mathscr{F}_n \nearrow \mathscr{F}_\infty$, meaning that $(\mathscr{F}_n)$ is a non-decreasing sequence of σ-algebras, whose union generates the σ-field $\mathscr{F}_\infty$, and X is an integrable random variable, then

$$\mathbb{E}[X|\mathscr{F}_n] \to \mathbb{E}[X|\mathscr{F}_\infty]$$

almost surely.

8.7. **Definition.** A left-infinite sequence $\ldots, X_{-2}, X_{-1}$ of integrable random variables is a reversed or backward martingale with respect to σ-algebras $\ldots \subseteq \mathscr{F}_{-2} \subseteq \mathscr{F}_{-1}$, if X_{-n} is measurable relative to $\mathscr{F}_{-n}$ for every $n \leq -1$ and $\mathbb{E}[X_{-n+1} \mid \mathscr{F}_{-n}] = X_{-n}$ for every $n < -1$.

An analogous convergence result does exist for reversed martingales and does read as follows.

8.8. **Theorem.** Given a reversed martingale $\ldots, X_{-2}, X_{-1}$, the random variables X_{-n} almost surely converge to a random variable X as n tends to infinity. The limit X is integrable and its expectation $\mathbb{E}[X]$ is given by $\mathbb{E}[X_{-n}]$ for all n.

8.9. **Theorem.** If $\mathscr{F}_n \searrow \mathscr{F}_\infty$, meaning that $\mathscr{F}_n$ is a non-increasing sequence of σ-algebras, whose intersection is the σ-algebra $\mathscr{F}_\infty$, and X is an integrable random variable, then the sequence of conditional expectations $\mathbb{E}[X \mid \mathscr{F}_n]$ defines a backward martingale with respect to $\ldots, \mathscr{F}_2, \mathscr{F}_1$ and

$$\mathbb{E}[X|\mathscr{F}_n] \to \mathbb{E}[X|\mathscr{F}_\infty]$$

almost surely as n tends to infinity.

REFERENCES

[1] Levin, David A., Peres, Yuval, Wilmer, Elizabeth L. Markov Chains and Mixing Times. American Mathematical Society. Providence. Rhode Island. 2009.

[2] Kemeny, J. G., Snell, J. L. Boundary theory for recurrent Markov chains. Trans. Amer. Math. Soc. 106. 1963.

[3] Pete, Gabor. Probability and Geometry on Groups. Lecture notes for a graduate course, available on his homepage `http://www.math.bme.hu/~gabor`

[4] Saloff-Coste, Laurent. Random Walks on Finite Groups. Encyclopaedia of Mathematical Sciences Volume 110. Probability Theory. Probability on Discrete Structures. Springer Verlag 2004

[5] Woess, Wolfgang. Denumerable Markov Chains - Generating Functions, Boundary Theory, Random Walks on Trees. EMS Textbooks in Mathematics. European Mathematical Society Publishing House 2009.

[6] Woess, Wolfgang. Random Walks on Infinite Graphs and Groups. Cambridge Tracts in Mathematics 138. Cambridge University Press 2000.

[7] Revesz, Pal. Random Walk in Random and Non-random Environments. Second Edition. World Scientific Publishing 2005

[8] Raugi, Albert. A general Choquet-Deny theorem for nilpotent groups. Annales de l'I.H.P., section B, 2004. Volume: 40, Issue: 6, page 677-683

[9] T. M. Cover and J. A. Thomas. Elements of Information Theory. John Wiley and Sons Inc. 1991.

[10] Klenke, Achim. Wahrscheinlichkeitstheorie. 2., korrigierte Auflage. Springer Verlag 2008.

[11] N. Th. Varopoulos. Long range estimates for Markov chains. Bul. Sci. Math. (2), 109:225-252, 1985.

[12] T. K. Carne. A transmutation formula for Markov chains. Bul. Sci. Math. (2), 109: 399-405, 1985.

[13] R. Lyons with Y. Peres. Probability on trees and networks. Cambridge University Press. In preparation. Current version available at `http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html`

[14] A. M. Vershik. Dynamic theory of growth in groups: Entropy, boundaries, examples. Russ. Math. Surveys 55 (2000), 667–733.

[15] Hoeffding, W. (1963) Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc., 58, 13–30.

[16] Azuma, K. (1967) Weighted sums of certain dependent random variables. Tohoku Math. J. (2), 19, 357–367.

[17] E. Kowalski. Spectral Theory in Hilbert Spaces. Lecture notes for a course at ETH Zurich. Current version available at `http://www.math.ethz.ch/~kowalski`

[18] R. Peyre. A probabilistic approach to Carne's bound. Potential Analysis 29 (2008), 17– 36. http://www.normalesup.org/ rpeyre/pro/research/carne.pdf

[19] S. Orey. Lectures on Limit Theorems for Markov Chain Transition Probabilities. Van Nostrand Reinhold Company. 1971

[20] Yves Derriennic. Lois "zero ou deux" pour les processus de Markov. Applications aux marches aleatoires. Annales de l'I.H.P., section B, tome 12, $n^\circ 2$ (1976), p. 111-129.

[21] M. Lin. Mixing for Markov operators. Z. Wahrschein., t. 19, 1971, p. 231-242.

[22] J. Michael Steele. Kingman's subadditive ergodic theorem. Annales de l'I.H.P., Vol. 25, $n^\circ 1$, 1989, p. 93-98.

[23] Patrick Billingsley. Probability and Measure. Second Edition. John Wiley and Sons Inc. 1986.

ZUSAMMENFASSUNG

Dieser Text stellt einige Konzepte vor, die das asymptotische Verhalten abzählbarer Markov Ketten erlauben zu beschreiben, und zeigt in welchem Verhältnis diese zueinander stehen.

Grundsätzlich werden zwei Klassen von Markov Ketten unterschieden: Markov Ketten, die rekurrent sind und Markov Ketten, die transient sind. Die Klasse der rekurrenten Markov Ketten besteht selbst wieder aus zwei Klassen: jene der positiv-rekurrenten und jene der null-rekurrenten. Für irreduzible, aperiodische und positiv-rekurrente Markov Ketten gibt es ein Theorem, das die Existenz eines eindeutig bestimmten invarianten Wahrscheinlichkeitsmaßes sicherstellt gegen das die Verteilungen der Zufallsvariablen der Markov Kette bezüglich der Norm der Totalen Variation konvergiert. Für Markov Ketten, die nicht positiv-rekurrent sind existiert ein solches Resultat nicht; stattdessen konvergieren die obigen Verteilungen punktweise gegen 0. Siehe dazu [1].

Wir präsentieren hier drei - andere - Möglichkeiten das asymptotische Verhalten zu beschreiben, hauptsächlich interessant im Kontext irreduzibler und transienter Markov Ketten:

Eine Möglichkeit besteht in der Einführung von σ-Algebren, die das mögliche Langzeitverhalten beschreiben: der asymptotischen und der invarianten σ-Algebra.

In einem analytischen Zugang wird der Übergangskern einer Markov Kette als Operator aufgefasst, der auf Funktionen des Zustandsraumes wirkt. Funktionen, die unter dieser Wirkung invariant sind, heißen harmonisch und der Raum der harmonischen Funktionen enthält sehr viel Information über das asymptotische Verhalten der Markov Kette. So gibt es eine kanonische Korrespondenz zwischen beschränkten harmonischen Funktionen und Äquivalenzklassen beschränkter Funktionen, die messbar bzgl. der invarianten σ-Algebra sind.

Die dritte Möglichkeit, die in diesem Text vorgestellt wird, ist geometrischer Natur: die zugrundeliegende Idee ist eine Kompaktifierung des Zustandsraumes zu konstruieren; die Punkte die dabei dem Zustandsraum hinzugefügt werden, werden dabei als Punkte interpretiert gegen die die Markov Kette konvergieren kann. Die Menge dieser Punkte wird als der Rand der Kette bezeichnet. Für eine irreduzible und transiente Markov Kette wird die Existenz einer Zufallsvariable gezeigt, die nur Werte im Rand annimmt und gegen die die Markov Kette fast sicher, in der Topologie der Kompaktifizierung konvergiert. Es gibt einen engen Zusammenhang zur Potentialtheorie: jede nicht-negative super-harmonische Funktion kann dargestellt werden als Integral einer Funktion - des Martin Kerns - bezüglich eines Borel-Maßes auf der Kompaktifizierung. Dieser Zusammenhang erlaubt die Charakterisierung der Trivialität des Randes durch die Nicht-Existenz nicht-konstanter beschränkter harmonischer Funktionen.

Welcher Zugang auch immer bevorzugt wird, die erste Frage die sich stellt, ist: wie lässt Trivialität bzw. Nicht-Trivialität des asymptotischen Verhaltens entscheiden. Die invariante σ-Algebra ist trivial genau dann wenn jede beschränkte harmonische Funktion konstant ist, was genau dann der Fall ist wenn der Rand trivial ist; und zu zeigen, dass jede beschränkte harmonische Funktion trivial ist, kann in manchen Situation recht einfach sein.

Wir zeigen, dass im Kontext von Irrfahrten auf diskreten Gruppen, die von einem symmetrischen Wahrscheinlichkeitmaß erzeugt werden, die invariante σ-Algebra und die asymptotische σ-Algebra übereinstimmen.

Weiters erklären wir die Bedeutung der Begriffe Drift und asymptotische Entropy. Informell gesprochen: der Drift einer Irrfahrt auf einer Gruppe ist ein Maß dafür inwiefern sich die Markov Kette auf lange Sicht vom Ausgangspunkt entfernt, während die asymptotische Entropy einer von einem Wahrscheinlichkeitsmaß erzeugten Irrfahrt quantifiziert inwiefern sie über die Elemente der Gruppe verstreut ist. Für Irrfahrten auf Gruppen, die von einem symmetrischen und endlich getragenen Wahrscheinlichkeitsmaß erzeugt werden, deren Träger eine unendliche Gruppe erzeugt, kann Trivialität des asymptotischen Verhaltens mit Hilfe dieser beiden Begriffe ausgedrückt werden: das asymptotische Verhalten - i.e. die invariante σ-Algebra, die asymptotische σ-Algebra, der Raum der beschränkten harmonischen Funktionen, der Rand - ist trivial genau dann, wenn der Drift gleich 0 ist, was genau dann der Fall ist, wenn die asymptotische Entropy gleich 0 ist.

CURRICULUM VITAE

Personal Information
Name: Manfred Buchacher
Date of Birth: 12th of May, 1989
Place of Birth: St. Pölten, Austria

Education
10/2012 - 04/2015 Master of Science in Mathematics, Universität Wien.
03/2009 - 09/2012 Bachelor of Science in Mathematics, Universität Wien.
09/1999 - 06/2007 Stiftsgymnasium der Benediktiner in Melk.
09/1995 - 06/1999 Volksschule Raxendorf.

Scholarships
08/2014 - 03/2015 Scholarship for Master Students, FWF.

Teaching Activities
10/2013 - 02/2014 Teaching Assistant, Fakultät für Math., Uni Wien.