Uniwersytet Wrocławski
Wydział Matematyki i Informatyki

Rozprawa Doktorska

mgr Piotr Markowski

Zastosowania dualności łańcuchów Markowa w modelach ruiny gracza i symulacji doskonałej

Promotor
dr hab. Paweł Lorek

Wrocław 2022

University of Wrocław

Departments of Mathematics and Computer Science

PhD Thesis

mgr Piotr Markowski

Applications of Markov chains dualities in gambler's ruin problem and perfect sampling

Supervisor
dr hab. Paweł Lorek

Wrocław 2022

Streszczenie

Poniższa rozprawa skoncentrowana jest wokół problemów związanych z łańcuchami Markowa na skończonych przestrzeniach stanów, które możemy rozwiązać wykorzystując pewne dualności między łańcuchami. W ogólności mówimy, że łańcuch $\mathbf{X}^*$ z macierzą przejścia $\mathbf{P}_{X^*}$ jest łańcuchem dualnym do $\mathbf{X}$ z macierzą przejścia $\mathbf{P}_X$ jeśli zachodzi (pomijamy warunki na rozkłady początkowe)

$$\Lambda \mathbf{P}_X = \mathbf{P}_{X^*}\Lambda,$$

gdzie Λ jest tzw. *linkiem*. Różne linki wyznaczają różne dualności. Dla konkretnych rodzajów dualności potrzebne są dodatkowe założenia na macierze przejść. Założenia te są mocno związane z porządkiem, najczęściej częściowym, przestrzeni stanów.

W pierwszym rozdziale rozprawy rozważamy monotoniczność Möbiusa, jej związek z innymi monotonicznościami (w tym m.in. ze stochastyczną monotonicznością) i z *symulacją doskonałą* (ang. *perfect simulation*) – metodą, która zwraca nieobciążoną próbkę z rozkładu stacjonarnego zadanego ergodycznego łańcucha Markowa. W tym rozdziale prezentujemy nowy algorytm do wspomnianej symulacji doskonałej bazujący na tzw. *mocnej dualności stacjonarnej* (ang. *strong stationary duality*), jego użycie wymaga wspomnianej monotoniczności Möbiusa.

Kolejny rozdział poświęcony jest wielowymiarowym modelom ruiny gracza. Okazuje się, że iloczyny Kroneckera dobrze *współpracują* z macierzowymi wzorami dla dualności, co pozwala uogólnić klasyczne wyniki. W szczególności, pokazujemy dużą rodzinę wielowymiarowych łańcuchów (które odpowiadają jakimś wielowymiarowym wersjom modelu ruiny gracza), które mają takie samo prawdopodobieństwo wygrania (na co podajemy wzór) i/lub taki sam rozkład czasu do wygranej/przegranej (podajemy jego strukturę).

Ostatni rozdział związany jest z rozkładem czasu trwania gry w modelu ruiny gracza (jednowymiarowym) pod warunkiem zwycięstwa bądź porażki – rozważamy model z dowolnymi prawdopodobieństwami wygrania/przegrania w jednym kroku. Pokazujemy m. in. interesujące symetrie wyników dla modelu oryginalnego i modelu z odwróconymi prawdopodobieństwami wygrania/przegrania. Stosując uzyskane wyniki (oraz mocną dualność stacjonarną), polepszamy wyniki dotyczące prędkości zbieżności do stacjonarności dla symetrycznego błądzenia po okręgu.

Abstract

The dissertation focuses on problems related to Markov chains on finite state spaces, which can be solved using certain dualities between the chains. In general, we say that the chain $\mathbf{X}^*$ with the transition matrix $\mathbf{P}_{X^*}$ is dual to the chain $\mathbf{X}$ with the transition matrix $\mathbf{P}_X$ if (we omit the conditions for initial distributions)

$$\Lambda \mathbf{P}_X = \mathbf{P}_{X^*} \Lambda,$$

where Λ matrix is a so-called *link*. Different links define different dualities. For specific types of duality, additional assumptions on transition matrices are needed. These assumptions are closely related to the ordering, most often partial, of the state space.

In the first chapter of the dissertation, we consider the Möbius monotonicity, its relations to other monotonicities (including, among others, stochastic monotonicity) and with *perfect simulation* – a method that returns an unbiased sample from the stationary distribution of a given ergodic Markov chain. In this chapter, we present a new algorithm for this perfect simulation based on the so-called *strong stationary duality*, its use requires the aforementioned Möbius monotonicity.

The next chapter is devoted to multidimensional gambler's ruin models. It turns out that the Kronecker products *work* well with matrix formulas for duality, which allows to generalize the classical results. In particular, we show a large family of multidimensional chains (which correspond to some multidimensional versions of the gambler's ruin model) that have the same probability of winning (we provide the exact formula) and/or have the same distribution of time till win/lose (we provide its structure).

The last chapter is related to the distribution of the game duration in the gambler's ruin model (one-dimensional) conditioned on the event of winning or losing – we consider model with arbitrary winning/losing probabilities in one step. We show, among others, interesting symmetries of the results for the original model and the model with swapped winning/losing probabilities. Using the obtained results (and strong stationary duality), we improve the results on the speed of convergence to stationarity for the symmetric random walk on a circle.

Preface

The dissertation consists of three articles, one article per chapter. The articles included here differ from the original ones, mainly in pagination, typographical details and numbering of equations, theorems, lemmas etc. The articles and corresponding chapters are following:

- **Chapter 1**. Lorek, P., Markowski, P., *Monotonicity requirements for efficient exact sampling with Markov chains.* Markov Processes And Related Fields, **23**(3), 485–514, 2017.
- **Chapter 2**. Lorek, P., Markowski, P., *Absorption time and absorption probabilities for a family of multidimensional gambler models.* ALEA-Latin American Journal of Probability and Mathematical Statistics, **19**, 125–150, 2022.
- **Chapter 3**. Lorek, P., Markowski, P., *Conditional gambler's ruin problem with arbitrary winning and losing probabilities with applications.* Submitted.

The thesis deals mainly with the theory of Markov chains. The first chapter is an in-depth study of several notions of monotonicities in Markov chains. The theoretical findings on relations between the monotonicities allowed to construct a new perfect simulation algorithm. The understanding of monotonicities and dualities was later extended to multidimensional chains – in the second chapter we studied generalizations of a gambler's ruin model. Among others, we consider games with several players involved, at one step we may win/lose with several players, the probabilities can depend e.g., on the current fortune. We provide the winning probabilities and the structure of a game duration of such models. In the last chapter we provide new results on birth and death chains – we provide the formulas (in terms of the parameters of the system) for a conditional (conditioned on winning or losing) game duration. The approach here is matrix-analytics. As a consequence we show interesting symmetries in such models (which were earlier known only for constant birth and death rates). As an application, using dualities, we show the exact rate of convergence (measured in the separation distance) – thus we improve known results – for the symmetric random walk on circle.

In the end I would like to appreciate my supervisor Dr. Paweł Lorek for all help with putting thoughts into words, and being so patient. I am also thankful to all the professors and teachers at the University and from the earlier levels of education for all the wisdom that they tried to pass over to me. Last but not least I want to thank my dear wife and family who were very supportive and thoughtful through the process.

Piotr Markowski,
Wrocław, 2022.

Contents

Introduction

In this thesis we deal with some problems related to Markov chains on a finite state space. We start with describing the *state of the art* of the problems under study, later in Sections 1, 2 and 3 we shortly describe the problems solved in the thesis, which are later on presented in details in corresponding Chapters 1, 2 and 3.

State of the art of the problems under study

We need to introduce some notation. The aforementioned finite state space is denoted by $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, whereas by $\mathbf{X} = \{X_k\}_{k=1}^{\infty} \sim (\nu, \mathbf{P}_X)$ we denote a discrete-time Markov chain with a transition matrix $\mathbf{P}_X$ and an initial distribution ν. By $\nu\mathbf{P}_X^k$ we denote the distribution of the chain at step k, *i.e.,* the distribution of X_k. We assume that $\mathbb{E}$ is partially ordered by $\preceq$. Moreover, throughout the thesis we assume that $\mathbf{e}_1$ is the minimal, and $\mathbf{e}_M$ is the maximal state, *i.e.,* $\forall \mathbf{e} \in \mathbb{E}$ we have $\mathbf{e}_1 \preceq \mathbf{e} \preceq \mathbf{e}_M$. For $A \subseteq \mathbb{E}$ we define $\mathbf{P}_X(\mathbf{e}, A) := \sum_{\mathbf{e}' \in A} \mathbf{P}_X(\mathbf{e}, \mathbf{e}')$, $\{\mathbf{e}\}^{\uparrow} := \{\mathbf{e}' : \mathbf{e} \preceq \mathbf{e}'\}$, $\{\mathbf{e}\}^{\downarrow} := \{\mathbf{e}' : \mathbf{e}' \preceq \mathbf{e}\}$ and $\delta(\mathbf{e}, \mathbf{e}') = \mathbf{1}(\mathbf{e} = \mathbf{e}')$. Whenever the ordering of the state space is linear, we denote the elements of $\mathbb{E}$ as consecutive numbers starting from 0 or 1, *i.e.,* $\mathbb{E} = \{1, \ldots, M\}$, and the ordering by $\leq$.

The crucial components of all chapters constituting the thesis is the **Siegmund duality** and the **strong stationary duality** (SSD). We will start with the Siegmund duality since the latter duality can be define in terms of the first one.

Siegmund duality and winning probabilities in gambler models. Assume $\mathbf{X}$ is an ergodic chain with the stationary distribution π. A *Siegmund dual* $\mathbf{Z}$ (of $\mathbf{X}$) with the transition matrix $\mathbf{P}_Z$ is a chain fulfilling the following:

$$\forall(\mathbf{e}_1, \mathbf{e}_2 \in \mathbb{E}) \qquad \mathbf{P}_X(\mathbf{e}_1, \{\mathbf{e}_2\}^{\downarrow}) = \mathbf{P}_Z(\mathbf{e}_2, \{\mathbf{e}_1\}^{\uparrow}). \tag{1}$$

In [LS16] it is shown that the matrix $\mathbf{P}_Z$ is non-negative iff $\mathbf{X}$ is Möbius monotone (more details below). However, it will be a substochastic matrix – then we add one extra absorbing state, say $\{-\infty\}$ and then we make – still calling it $\mathbf{P}_Z$ – a stochastic matrix on $\mathbb{E}' = \mathbb{E} \cup \{-\infty\}$. It is relatively easy to show that $\mathbf{Z}$ must have two absorbing states, the introduced $\{-\infty\}$ and the maximal one $\mathbf{e}_M$.

Relation (1) implies that we actually have $\mathbf{P}_X^n(\mathbf{e}_i, \{\mathbf{e}_j\}^{\downarrow}) = \mathbf{P}_Z^n(\mathbf{e}_j, \{\mathbf{e}_i\}^{\uparrow})$ for any $n \geq 1$ and any $1 \leq i, j \leq M$. Then, calculating $\lim_{n\to\infty}$ we have

$$\pi(\{\mathbf{e}_j\}^{\downarrow}) = P(\tau_{\mathbf{e}_M} < \tau_{-\infty} | Z_0 = \mathbf{e}_j) =: \rho(\mathbf{e}_j), \tag{2}$$

where $\tau_{\mathbf{e}} := \inf\{n : Z_n = \mathbf{e}\}$ (first hitting time). This is the fundamental relationship between the stationary distribution of an egodic chain and the absorption (winning) probabilities of its Siegmund dual. We can also start with a chain $\mathbf{Z}$ having already two absorbing states, then we

i) remove the row and column corresponding to the losing state $\{-\infty\}$ from $\mathbf{P}_Z$;

ii) solve (1) for $\mathbf{P}_X$;

iii) compute the stationary distribution of $\mathbf{X} \sim \mathbf{P}_X$;

iv) use (2) to compute the winning probabilities.

Thus, the main application of Siegmund duality is to provide the formula for winning probabilites in gambler's ruin-like problems (chains with two absorbing states).

Strong stationary duality and rate of convergence to stationarity for ergodic Markov chains. Consider an ergodic Markov chain $\mathbf{X} \sim (\nu, \mathbf{P}_X)$ on $\mathbb{E}$ with a stationary distribution π. By the rate of convergence of the chain to its stationary distribution we understand a knowledge on number of steps guaranteeing that the distance between the distribution at this step and π is small. Often total variation distance is used

$$d_{TV}(\nu \mathbf{P}_X^k, \pi) := \sup_{A \subset \mathbb{E}} |\nu \mathbf{P}_X^k(A) - \pi(A)|.$$

By "mixing time" we usually understand

$$\tau(\varepsilon) = \min\{k : d_{TV}(\nu \mathbf{P}_X^k, \pi) \leq \varepsilon\}.$$

We say that $\mathbf{X}^* \sim (\nu^*, \mathbf{P}_{X^*})$ on $\mathbb{E}$ is the **strong stationary dual** (SSD) chain if:

a) $\mathbf{P}_{X^*}$ has one absorbing state, say $\mathbf{e}_M$, and

b) there exists a stochastic matrix Λ such that:

$$\Lambda(\mathbf{e}_M, \cdot) = \pi(\cdot), \quad \nu = \nu^* \Lambda, \quad \Lambda \mathbf{P}_X = \mathbf{P}_{X^*} \Lambda. \tag{3}$$

Then in [DF90b] it is shown that the absorption time T^* for $\mathbf{X}^*$ gives the following bound:

$$sep(\nu \mathbf{P}_X^k, \pi) \leq P(T^* > k), \text{ where } sep(\nu \mathbf{P}_X^k, \pi) := \max_{\mathbf{e} \in \mathbb{E}} (1 - \nu \mathbf{P}_X^k(\mathbf{e}) / \pi(\mathbf{e}))$$

(*sep* is called the *separation distance*). On the other hand, it is known that $d_{TV}(\nu \mathbf{P}_X^k, \pi) \leq sep(\nu \mathbf{P}_X^k, \pi)$. In other words, once we have an SSD, the problem of studying the rate of convergence translates to studying the absorption time (usually easier to handle). In [LS12b] it is shown that for a given ergodic $\mathbf{X}$ there exists its SSD $\mathbf{X}^*$ iff time reversed chain $\overleftarrow{\mathbf{X}}$ is Möbius monotnone (w.r.t. the partial ordering $\preceq$). The time reversed chain $\overleftarrow{\mathbf{X}}$ is the one with transition matrix $\overleftarrow{\mathbf{P}}_X(\mathbf{e}_i, \mathbf{e}_j) = \frac{\pi(\mathbf{e}_j)}{\pi(\mathbf{e}_i)} \mathbf{P}(\mathbf{e}_j, \mathbf{e}_i)$ (many examples are reversible chains, then $\mathbf{P}_X = \overleftarrow{\mathbf{P}}_X$). The transitions of the SSD are following

$$\mathbf{P}_{X^*}(\mathbf{e}_i, \mathbf{e}_j) = \frac{H(\mathbf{e}_i)}{H(\mathbf{e}_j)} \sum_{\mathbf{e}:\mathbf{e} \succeq \mathbf{e}_j} \mu(\mathbf{e}_j, \mathbf{e}) \overleftarrow{\mathbf{P}}_X(\mathbf{e}, \{\mathbf{e}_i\}^{\downarrow}), \tag{4}$$

where $H(\mathbf{e}) = \sum_{\mathbf{e}':\mathbf{e}' \preceq \mathbf{e}} \pi(\mathbf{e}')$ and μ is the Möbius function of the ordering $\preceq$. The nonnegativity of $\mathbf{P}_{X^*}$ corresponds (iff condition) to Möbius monotonicity of $\overleftarrow{\mathbf{X}}$ (see below). Later, in [LS16] it was shown that this construction yields optimal SSD: the absorption time T^* is a FSST for $\mathbf{X}$, *i.e.,* we have

$$sep(\nu \mathbf{P}_X^k, \pi) = P(T^* > k). \tag{5}$$

Möbius monotonicity. The existence of a Siegmund dual chain and strong stationary dual chain for *linearly ordered* state spaces was known for quite a long time. Namely, the usual stochastic monotonicity was required. However, it was not known how to extend the results to partial orderings. As already mentioned, it was done in [LS12b], [Lor18]. It turns out that the *Möbius monotonicity* (which is equivalent to stochastic monotonicity for linearly ordered state space, but it is quite different notion of monotonicity for partial non-linear ordering) is required (of the chain – in case of Siegmund dual, or of the time-reversed chain – in case of an SSD), we will thus provide some details.

Recall, we consider partial ordering $\preceq$ on $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$. Let $\mathbf{C}(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{1}(\mathbf{e}_i \preceq \mathbf{e}_j)$. We can always rearrange the states so that $\mathbf{e}_i \preceq \mathbf{e}_j$ implies $i \leq j$. Then the matrix $\mathbf{C}$ is 0–1 valued, upper triangular, and thus invertible. The inverse $\mathbf{C}^{-1}$ is often denoted by μ and is called *the Möbius function.* Let $f, \bar{F} : \mathbb{E} \to \mathbb{R}$. The famous *Möbius inversion formula* states:

$$\text{Let } \bar{F}(\mathbf{e}) = \sum_{\mathbf{e}' \succeq \mathbf{e}} f(\mathbf{e}'), \quad \text{then } f(\mathbf{e}) = \sum_{\mathbf{e}' \succeq \mathbf{e}} \mu(\mathbf{e}, \mathbf{e}') \bar{F}(\mathbf{e}').$$

We say that the function $\bar{F} : \mathbb{E} \to \mathbb{R}$ is Möbius monotone if $\sum_{\mathbf{e}' \succeq \mathbf{e}} \mu(\mathbf{e}, \mathbf{e}') \bar{F}(\mathbf{e}') \geq 0$ for all $\mathbf{e} \in \mathbb{E}$. Let us define $\bar{F}_{\mathbf{e}_2}(\mathbf{e}') = \mathbf{P}_X(\mathbf{e}', \{\mathbf{e}_2\}^{\downarrow})$, for any states $\mathbf{e}_2, \mathbf{e}' \in \mathbb{E}$. We say that the chain $\mathbf{X}$ is **Möbius monotone if** $\bar{F}_{\mathbf{e}}$ are Möbius monotone for all $\mathbf{e} \in \mathbb{E}$, *i.e.,*

$$\forall (\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}) \quad \sum_{\mathbf{e} : \mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) \mathbf{P}_X(\mathbf{e}, \{\mathbf{e}_j\}^{\downarrow}) \geq 0. \tag{6}$$

1 Möbius monotonicity, its relations to other monotonicities and perfect simulation

Markov chain Monte Carlo (MCMC) methods are a class of tools for approximate sampling from a prescribed distribution π. The main application is for distributions on large state spaces which are hard/impossible to sample with other methods. Rougly speaking, we construct a chain with π being its stationary distribution. Then, running the chain *long enough* we approximately sample from π. **Perfect simulation** is the art of converting a Markov chain into an algorithm which returns an *unbiased* sample from its stationary distribution. There are two popular (the first one being the most popular) perfect sampling algorithms:

- **Coupling from the past** (CFTP). It was introduced in a ground breaking paper [PW96]. The ingenious idea of the algorithm is to realize the chain as a stochastic flow and evolve it *from the past* (rather than into the future). In order to be able to use it *efficiently*, the chain must be **realizable monotone**.

- **Fill's rejection algorithm**. It was introduced in [Fil98]. A general version of this algorithm can be applied assuming **stochastic monotonicity** of the time reversed chain. Although in theory it has wider applications (realizable monotonicity implies stochastic monotonicity), in practice it is harder to apply (compared to CFTP). Some more details and extensions are provided also in [FMMR00].

In Chapter 1 we also present a new perfect sampling algorithm:

- **Strong stationary dual**-based perfect sampling algorithm. It turns out that once we compute SSD for a given ergodic chain, then we may be able to perform perfect sampling. Thus, in principle, we are able to construct an SSD if the time reversed chain (of the ergodic chain) is **Möbius monotone**.

In Chapter 1 we study in-depth relations between several notions of monotonicities in Markov chains: usual stochastic monotonicity; weak monotonicity (two versions); Möbius monotonicity (also two versions) and realizable monotonicity. We mentioned that stochastic monotonicity and Möbius monotonicity are equivalent for linear orderings (actually they are equivalent, also with realizable monotonicity, for "tree-like" orderings) and in general they are different for partial orderings. Surprisingly, we show that we can have e.g.,

- a chain which is stochastically, but not Möbius monotone,
- a chain which is Möbius, but not stochastically monotone (thus also not realizable monotone).

Note that the latter example means that for some chains only *Strong stationary dual*-based perfect algorithm out of three mentioned ones may be applied.

My contribution

- My main contribution is Theorem 5.1 (page 19), i.e., the relation between several notions of monotonicities of Markov chains defined w.r.t. some fixed partial ordering of the state space: realizable monotonicity $\mathcal{R}$; stochastic monotonicity $\mathcal{S}$; two versions of weak monotonicities ($\mathcal{W}^{\uparrow}$ and $\mathcal{W}^{\downarrow}$) and two version of Möbius monotonicities $\mathcal{M}^{\uparrow}$ and $\mathcal{W}^{\downarrow}$). To obtain this result I had to combine and transform definitions appropriately, mainly to be able to work in terms of Möbius function. The real challenge I was trying to get a full picture of relations between those monotonicities. The relations between the aforementioned monotonicities are presented in the Figure 1 (page 21). I showed almost full characterization. I provided examples (in Appendix 6 in 26) for each case except one (which is – until today – an open problem): I was unable to show if there is a chain which is in $\mathcal{M}^{\uparrow}$, $\mathcal{M}^{\downarrow}$ and $\mathcal{S}^{\uparrow}$ but not in $\mathcal{M}^{\uparrow}$. I came up with those examples on the train from Rawicz to Wrocław. I am grateful to my coauthor for making me write the proper checker in the `Julia language` (some of those examples were wrong in the prepublished version).
- Also Theorem 5.2 (page 22) is my contribution. It is a special case of Theorem 5.1 for a specific (tree-like) partial ordering. It turns out that in such a case the monotonicites actually reduce to only three: $\mathcal{S}, \mathcal{R}$ and $\mathcal{M}^{\downarrow}$ (each of any other is equivalent to one of these).
- The whole idea for studying the relation between monotonicities and relating them to perfect simulation algorithms came from my coauthor. Other then that, we have discussed and developed all of the other parts together, it is hard to single out any of the results as my own. For example for our more general version of Fill's rejection algorithm – my coauthor have adapted the algorithm and checked that it is still mathematically correct.

2 Multidimensional gambler models – winning probabilities and absorption time

In [Lor17] the following multidimensional gambler's ruin model was studied: we play with $d \geq 1$ other players. Assume our current fortune with each of the players is $(i_1, \ldots, i_d)$ (*i.e.,* i_j is the amount of money on j-th "table"), where $i_j \leq N_j$ (fixed integers). At each step we may: win one dollar with player j with probability $p_j(i_j)$, lose one dollar with probability $q_j(i_j)$, we assume that $p_j(i_j) > 0, q_j(i_j) > 0$ and $\sum_{k=1}^{d}(p_k(i_k) + q_k(i_k)) \leq 1$. We also assume that we win the whole game if we win with all the players (*i.e.,* the state $(N_1, \ldots, N_d)$ is the *winning* state) and we lose the whole game if we lose with at least one player (additionally introduced *losing* state $-\infty$). Denote the chain by $\mathbf{Z}$. The following winning probability (provided we start at $(i_1, \ldots, i_j)$) was derived in [Lor17]:

$$\rho((i_1, \ldots, i_d)) = \frac{\prod_{j=1}^{d}\left(\sum_{n_j=1}^{i_j}\prod_{r=1}^{n_j-1}\left(\frac{q_j(r)}{p_j(r)}\right)\right)}{\prod_{j=1}^{d}\left(\sum_{n_j=1}^{N_j}\prod_{r=1}^{n_j-1}\left(\frac{q_j(r)}{p_j(r)}\right)\right)}.$$

The Möbius monotonicity w.r.t. coordinate-wise ordering was used. An ergodic chain $\mathbf{X}$ was found such that it was Möbius monotone w.r.t. coordinate-wise ordering and such that $\mathbf{Z}$ is its Siegmund dual. Then the result was derived exploiting (1).

In Chapter 2 we construct multidimensional gambler models in such a way that its winning probability is a product of underlying one-dimensional chains. As a special case (when these underlying chains are birth and death chains) we show the whole family of multidimensional gambler models such that their winning probability is still given by formula (1.3) – for example we may have the following scenarios:

- We may play with $r \leq d$ players at one step.
- If our current total fortune (with all the players) is less than 100$ we play with say at most 2 players at one step. But once we have $\geq$ 100$ we may play with all the players at one step.

Moreover, in this Chapter 2 we also study the (structure of) absorption time of such multidimensional models. We show that, in a sense, the structure of absorption time of some chains (under suitable assumptions) is similar to the absorption time of birth and death chains. Roughly speaking, we show that it may be expressed in terms of the absorption time of another *pure-birth* chain. By pure-birth multidimensional chain we mean that at any step no coordinate may decrease its value. To achieve the results we heavily exploit the notion of Kronecker sums and products. For winning probability results we use Siegmund duality, whereas for results related to absorption time we use some intertwining between chains.

My contribution

- Theorem 2.1 (page 40) is my main contribution. In the article of my coauthor [Lor17] (which was a starting point for generalised gambler models) the winning probability for some specific multidimensional game was provided. The aforementioned Theorem 2.1 is a (significant) extension: it provides a way to construct a variety of multidimensional gambler models out of one-dimensional games in such a way, that if we know winning probabilities of these one dimensional games, we know the winning probability of the resulting multidimensional game. The proof of the theorem is based on careful operations on matrices, eigenvectors and Kronecker sums and products. For clarity, it was split into Lemma 4.1 (page 48) and and Corollary 4.2 (page 49), whose proofs are mine.

- Theorem 2.3 (page 42) deals with the structure of absorption time T for multidimensional models. My coauthor spotted that for some subclass of models resulting from Theorem 2.1 the probability generating function pgf of T can be expressed as a mixture of absorption times of some *simpler* (in a sense) models. I have proposed the formulation of assumptions and proven those with Kronecker product operations (Section 4.3 page 50), while my coauthor suggested the other way – using intertwining between absorbing chain and ergodic chain (Section 5 page 51) .

- The most interesting point in Theorems 2.1 and 2.3 are the assumptions that are very general (it takes over half a page to state them for each of the theorems). Basically, in Theorem 2.1 we propose whole family of processes for which the winning probability have product form (2.3) page 40:

$$\rho(i_1, \ldots, i_d) = \prod_{j=1}^{d} \rho_j(i_j).$$

  In Theorem 2.3 we show the following: we start with d one dimensional birth and death chains (gambler models) and construct a family of d dimensional multivariate gambler models (using weighted sums of Kronecker products). Denote its absorption time in $\mathbf{N} = (N_1, \ldots, N_d)$, provided it started with distribution ν^*, by $T^*_{\nu^*,N}$. Then, for each of such models, we show another d dimensional model which has only upward transitions. Denote its absorption time in $\mathbf{N}$, provided it started in $\hat{\mathbf{e}} = (\hat{e}_1, \ldots, \hat{e}_d)$, by $\hat{T}_{\hat{\mathbf{e}}}$. Then both of those chains are complicated but the formula connecting they're pgf's is quite simple:

$$\mathsf{pgf}_{T^*_{\nu^*,\mathbf{N}}}(s) = \sum_{\hat{\mathbf{e}} \in \mathbb{E}} \hat{\nu}(\hat{\mathbf{e}}) \mathsf{pgf}_{\hat{T}_{\hat{\mathbf{e}},\mathbf{N}}}(s) \left(\prod_{j=1}^{d} \rho_j(1) \right).$$

  To show how we can work with such complicated assumptions we had to show some examples. First two of them (provided in subsections 6.1 and 6.2) were proposed by my coauthor and calculated by me, while I have came up (and calculated) myself with the other two (provided in subsections 6.3 and 6.4). I am satisfied with those, because it is hard to get results for multidimensional Markov chains with useful non-mathematical description.

- Discussion in the end of Section 2 on consequences of Theorem 2.3, as well as discussion on dualities in Section 3 are results of our joint work, but they were properly written down mostly by my coauthor.

- I would like to acknowledge Bartłomiej Błaszczyszyn for suggesting Kronecker products as a tool which turned out to work nicely with multidimensional Markov chains.

3 Conditional gambler's ruin problem

Let us go back to one dimensional gambler's ruin problem, *i.e.,* the birth and death chain on $\{0, 1, \ldots, N\}$ with the probability of winning one dollar in one step $p(j)$ and losing it with probability $q(j), j = 1, \ldots, N-1$. As usual, we identify state 0 with *losing* and state N with *winning*. Denote the chain by $\mathbf{Z}$. This time we are interested in *absorption time* T_i and *conditional absorption time* W_i (conditioned on *winning*), both in case when we start at i. To be more specific, for a gambler's model Z we define

$$\begin{aligned} T_i &= \inf\{n \geq 0 : Z_n = 0 \text{ or } Z_n = N,\ Z_0 = i\}, \\ W_i &= T_i \text{ conditioned on } Z_{T_i} = N. \end{aligned}$$

In Chapter 3 our main results are the formulae for ET_i and EW_i, both stated only in terms of the parameters of the model, *i.e.,* in terms of $p(j)$ and $q(j)$. As far as we are aware, these formulas were not known previously.

Let $W_i^{\mathbf{p}\leftrightarrow\mathbf{q}}$ be the conditional absorption time for a chain with swapped rates, *i.e.,* $p'(j) = q(j)$ and $q'(j) = p(j)$. An interesting symmetry was earlier noted for the classical gambler's ruin problem, *i.e.,* with constant rates $p(j) = p, q(j) = q$. To be more exact first in [Ste75] it was shown that $EW_{N/2} = EW_{N/2}^{\mathbf{p}\leftrightarrow\mathbf{q}}$ for even N and in a model with no ties, *i.e.,* $p + q = 1$. Later in [BW77] it was shown that actually in such a case a symmetry in a distribution holds, *i.e.,* $W_i \stackrel{d}{=} W_i^{\mathbf{p}\leftrightarrow\mathbf{q}}$ (and any i). Authors said that it was a *surprising and paradoxical result*.

We extend the result: the formula for EW_i we provide in Chapter 3 implies that $EW_i = EW_i^{\mathbf{p}\leftrightarrow\mathbf{q}}$ as long as $r = \frac{q(j)}{p(j)}$ is constant (and we conjecture that actually in this case the the distribution of conditional absorption time is symmetric, i.e., $W_i \stackrel{d}{=} W_i^{\mathbf{p}\leftrightarrow\mathbf{q}}$).

Moreover, in Chapter 3, as a *side effect* we provide detailed analysis of the rate of convergence to stationarity of a random walk $\mathbf{X}$ on a circle (*i.e.,*, on integers $0, 1, \ldots, d$, we cyclically move to the left or to the right with probability p). Roughly speaking, we introduce some non-trival ordering on the circle and compute its SSD $\mathbf{X}^*$ using formula (4). It turns out that $\mathbf{X}^*$ is a birth and death chain with non-constant birth and death rates, for which we are able to compute expectation of its absorption time. Using the fact that such a construction yields an optimal SSD, we have (5). As a consequence, for example for $p = 1/3$ and d being some power of 2, we have

$$ET = \frac{1}{8}d^2 + \frac{1}{4}.$$

Thus we improve upon a result of Fill and Diaconis [DF90b], where they show a construction of not optimal T_0 such that $ET_0 = \frac{1}{8}d^2 + 1$. It is worth noting that ET and ET_0 differ by $\frac{3}{4}$ independently of d.

My contribution

- Theorem 2.1 (page 62) with equation (2.1) are my contribution. In Chapter 3 one-dimensional gambler ruin problem is considered. The main result of the theorem – equation (2.1) – is the expectation of absorption time (*aka* game duration). Although the idea for the proof is straightforward – apply "*first step analysis*", the computations are far from obvious. The formula for expectation was known e.g., in terms of the eigenvalues of the transition matrix, whereas in Chapter 3 a formula in terms of the parameters of the system is provided:

$$ET_{j:i:k} = \frac{\sum_{n=j+1}^{k-1}[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}]}{\sum_{n=j}^{k-1} d_n} \sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1} \left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s} \right]$$

  where $d_s = \prod_{i=j+1}^{s} \frac{q(i)}{p(i)} = \prod_{i=j+1}^{s} r(i)$ (with convention $d_j = 1$). It is a case $j \leq i \leq k$ where j is the *losing* state and k is the *winning* state.

- Theorem 2.2 (page 62) was also prooven by me – it is a consequence of Theorem 2.1 for only one absorbing state.

- Theorem 2.3 (page 63), which is the main result of Chapter 3, was done be me. It is about expected conditional game duration (conditioned on winning, or symmetrically on losing). The proof in principle also uses first step analysis, but is much more involved. E.g., it uses some matrix-analytics approach for indices inferred from the aforementioned first step analysis (see Lemma 5.1). The formula itself requires introducing lengthy notations, it itself is

$$EW_{0:1:i} \quad = \quad \sum_{n=1}^{i-1} \frac{\rho_{0:n:i}}{p(n)} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \xi_s^{n+1,i-1}.$$

  Here $1 \leq i$, this is the formula for game duration conditioned on winning (which is reaching state i, state 0 is the losing state). Because of the complexity of the formula, it was important to provide all-telling examples. The ideas for those (in Sections 2.1, 2.2, 2.3 and 2.4) were not mine, but most of the calculations were mine.

- Similarly with applications for *random walk on a polygon* and *symmetric random walk on a circle*, the ideas came from the coauthor, but calculations (namely: proofs of Theorem 3.1 on page 72 and Lemma 4.1 on page 75) were mine. To be precise, coauthor came up with the whole concept of computing an optimal strong stationary dual chain for the problem, which turned out to be absorbing birth and death chain. Then using Theorem 2.2 (page 62) we computed the expectation of the so-called *fastest strong stationary time* for this random walk, improving this way the results of Diaconis and Fill [DF90b].

Chapter 1. Monotonicity requirements for efficient exact sampling with Markov chains

1 Introduction

Monte Carlo Markov Chain (MCMC) methods are a class of tools for approximate sampling from a given distribution (usually intractable by other methods). The method is based on constructing an ergodic Markov chain that has the desired distribution as its stationary distribution. Then the algorithm outputs a state after simulating the chain for some number of steps. The ergodicity implies that the more steps are performed, the closer it is to the stationary distribution. However, to say something about the error, one needs to have some theoretical bounds on the rate of convergence, e.g., the mixing time. In many practical problems this is an obstacle that is hard to overcome.

Exact (or perfect) simulation refers to the art of converting a Markov chain (usually obtained from MCMC methods) into an algorithm which returns an unbiased sample from its stationary distribution. In this chapter we briefly present three such algorithms. Our main focus is on the monotonicity requirements for *efficient* application of these algorithms. Each of the algorithms requires then a different notion of monotonicity. The monotonicities will be defined with respect to a partial ordering (in the applications, a state space usually has some natural underlying partial ordering). The idea of all the algorithms is based on a coupling.

Coupling from the past (CFTP) is probably the most famous exact sampling algorithm, introduced in a ground breaking paper [PW96]. The ingenious idea of the algorithm is to realize the chain as a stochastic flow and evolve it *from the past* (rather than into the future). Doing so requires considering coupled realizations of chains started at all possible starting points. This is infeasible in most cases. However, if the chain is so-called **realizable monotone**, we need only to simulate two chains, thus making the algorithm very effective. Although many variations of the algorithm have been invented, often with slightly different requirement for monotonicity, see, e.g., [HN98, Hub03, Ken98, KM00], we focus on this (widely used) monotonicity.

The second exact sampling algorithm is *Strong Stationary Dual*-based. The notion was introduced in [DF90b] and exploited mainly for studying the rate of convergence. However, having such a dual chain (including a so-called *link*), we can couple two chains in such a way that when the dual chain hits a specific state, then the original chain has the stationary distribution. The point is that there is no general way to come up with such a dual chain. In [DF90b], the authors give a recipe only for chains whose time reversals are stochastically monotone. This duality for total ordering has been exploited in many contexts, see, e.g., [DSC06], [Fil09a, Fil09b, FL14] (most of them deal only with birth and death chains). However, many interesting distributions are given on a state space possessing a natural non-linear ordering. The existence (and recipe) of such a dual for a partial ordering was given in [LS12a]. We briefly present an exact sampling algorithm based on the coupling from [DF90b] and the dual given in [LS12a]. This dual exists (and thus we can use the algorithm) if and only if the time reversal is **Möbius monotone**. For examples of Möbius monotone chains, see [LS16], for connections with Siegmund duality on partially ordered state spaces, see [Lor18] and [Lor17].

The last algorithm we will present is the so-called **Fill's rejection algorithm** [Fil98]. The author presents the algorithm already assuming some monotonicity. We present here a more general version of the algorithm, together with a short proof of correctness. Similarly to the previous algorithms, it is hard to apply it to a general chain. However, the algorithm can be applied assuming **stochastic monotonicity** (w.r.t. the partial ordering) of the time reversed chain. For more details and/or extensions of the algorithm, consult, e.g., [CLR01, Dim01, FMMR00].

The three exact sampling algorithms presented, require, as already mentioned, three different monotonicities for efficient application: realizable monotonicity, Möbius monotonicity, and stochastic monotonicity. The relation between the first and the last one is already known: for a general partial ordering, realizable monotonicity implies stochastic monotonicity, whereas they are equivalent for total or tree-orderings, see [FM01], [Mac01]. However, Möbius monotonicity has not been studied as extensively as the other two. Theorem 5.1 and Fig. 1 show all the relationships between the orderings (including also weak monotonicities and distinguishing between Möbius-$^{\uparrow}$ and Möbius-$^{\downarrow}$ monotonicities). In particular, one interesting (from both points of view: theoretical and practical) implication is that the chain (and/or its time reversal) does not have to be stochastically monotone (and thus realizable monotone) but can be Möbius monotone. Thus we can still use one of the exact sampling algorithms. We also present several examples of orderings and chains showing all possible cases of being/not being monotone in a specific sense. However we are forced to leave one **open problem** (see open problem 1): we cannot prove or disprove that there exists a chain and ordering such that the chain is Möbius-$^{\uparrow}$, Möbius-$^{\downarrow}$, stochastically monotone, but is not realizable monotone.

The organization of this chapter is as follows: In Section 2 we present the aforementioned exact sampling algorithms in their full generality. In Section 3 we formally introduce the monotonicities. Section 4 contains applications of the monotonicities the three algorithms mentioned above (i.e., to their efficient applications). In Section 5 we present the relationship between the monotonicities, whereas the examples are postponed to the Appendix.

2 Three general methods for exact sampling

We consider an ergodic Markov chain $\mathbf{X} = \{X_k\}_{k\geq 0}$ with the initial distribution ν, finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, transition matrix $\mathbf{P} = [\mathbf{P}(\mathbf{e}, \mathbf{e}')]_{\mathbf{e},\mathbf{e}'\in\mathbb{E}}$, and stationary distribution π. The distribution of the chain at step k, started with the initial distribution ν, is denoted by $\nu\mathbf{P}^k(\cdot)$. For a measure f on $\mathbb{E}$ we write

$$f(A) = \sum_{\mathbf{e}\in A} f(\mathbf{e}) \text{ for } A \subseteq \mathbb{E}.$$

It is said that $\overleftarrow{\mathbf{X}}$ is the time reversed chain of the chain $\mathbf{X}$ if $\overleftarrow{\mathbf{X}}$ is defined on the same space as $\mathbf{X}$ and has t.m.

$$\overleftarrow{\mathbf{P}}(\mathbf{e}, \mathbf{e}') = \frac{\pi(\mathbf{e}')}{\pi(\mathbf{e})}\mathbf{P}(\mathbf{e}', \mathbf{e}).$$

One can simulate the Markov chain using an update rule:

Definition 2.1. A function $\phi : \mathbb{E} \times [0,1] \to \mathbb{E}$ is an **update rule** for the chain $\mathbf{X}$ with the transition matrix $\mathbf{P}$ if:

1. for fixed $\mathbf{e} \in \mathbb{E}$ the function $\phi(\mathbf{e}, u)$ is piecewise constant, and
2. for all $\mathbf{e}, \mathbf{e}' \in \mathbb{E}$ we have $\int_0^1 \mathbf{1}(\phi(\mathbf{e}, u) = \mathbf{e}')du = \mathbf{P}(\mathbf{e}, \mathbf{e}')$.

Note that for the uniformly distributed random variable $U \sim Unif[0,1]$ we have $\mathbf{P}(\mathbf{e}, \mathbf{e}') = Pr(\phi(\mathbf{e}, U) = \mathbf{e}')$. Having such an update rule, one can recursively simulate the chain:

$$X_0 \sim \nu, \quad X_{k+1} = \phi(X_k, U_{k+1}),$$

where $U_1, U_2, \ldots$ is an iid sequence of random variables uniformly distributed on $[0,1]$.

The usual Monte Carlo Markov Chain (MCMC) provides methods for approximate sampling from the desired distribution π, usually intractable by other methods (such as, e.g., inverting the distribution function). Roughly speaking, the methods include constructing an ergodic chain with π being its stationary distribution. Thus, simulating $X_0, X_1, \ldots$ *long enough*, the distribution of X_k will be close to the stationary distribution (since ergodicity implies that $\lim_{k\to\infty} \nu\mathbf{P}^k(\cdot) = \pi(\cdot)$). Note that: i) usually it will never be exactly the stationary distribution; ii) to know how close it is to the stationarity the distribution of X_k, one needs to know the rate of convergence.

We briefly recall three methods for **exact sampling** (often called *perfect sampling*), i.e., obtaining an unbiased sample from π. All of them rely on the concept of *coupling*. A coupling of a pair of Markov chains with a common transition matrix $\mathbf{P}$ is a bivariate process $\{(X_k, Y_k)\}_{k\geq 0}$ such that marginally $\{X_k\}_{k\geq 0}$ and $\{Y_k\}_{k\geq 0}$ are Markov chains with the transition matrix $\mathbf{P}$ (in particular, the processes may be, and usually are, dependent, and have different initial distributions).

2.1 Method 1: Coupling from the past

One of the most known algorithms for exact sampling is called coupling from the past (CFTP) (or the Propp–Wilson algorithm, cf. [PW96]). Given an increasing sequence $N_1, N_2, \ldots$ of positive integers (usually $N_r = 2^{r-1}$), the algorithm is as follows:

Algorithm 1 Coupling from the past (CFTP).

Require: State space $\mathbb{E}$, ergodic chain $\mathbf{X}$ with update rule ϕ.
1: Set $n = 1$.
2: For each $\mathbf{e} \in \mathbb{E}$ simulate the Markov chain starting at time $-N_n$ in state $\mathbf{e}$ and run it till time 0 using the same update rule ϕ and iid random variables $U_{-N_n+1}, U_{-N_n+2}, \ldots, U_{-1}, U_0$ uniformly distributed on $[0,1]$ (the same for each chain).
3: If all chains in the previous step end up in the same state $\mathbf{e}_0$ at time 0, then output $\mathbf{e}_0$ and **stop**.
4: Set $n = n + 1$ and go to Step 2 (keep the previously used $\{U_i\}_{0\leq i\leq -N_n+1}$ for new n).

The very rough idea of the CFTP algorithm is the following: assume at some time "in the past," say at $-N_n$, for each $\mathbf{e} \in \mathbb{E}$ we started a chain. Later on, using the same update rule and the same uniform random variables driving the chains, all the chains have coalesced before time 0. If we had started the chains earlier, even at "minus infinity," but from $-N_n$ on using the same uniform random variables, we would still end up in the same state at time 0. And from "minus infinity" till 0 it surely has already "reached" stationarity, thus the output of the algorithm is a random variable with the distribution being the stationary distribution of the chain. For more technical details (e.g., that it always terminates), see [PW96].

2.2 Method 2: Fill's rejection algorithm

Let $p(\cdot), q(\cdot)$ be two probability distributions on $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$. Assume that for some $c \in \mathbb{R}$ we have $p(\mathbf{e}) \leq cq(\mathbf{e})$ for all $\mathbf{e} \in \mathbb{E}$ (then $c \geq 1$, but $c > 1$ if the distributions are different). The classical acceptance-rejection algorithm lets us simulate from $p(\cdot)$ if we are able to simulate from $q(\cdot)$:

Algorithm 2 Acceptence-rejection algorithm.

Require: Distributions $p(\cdot), q(\cdot)$ on $\mathbb{E}$, constant $c : p(\mathbf{e}) \leq cq(\mathbf{e})$.

1: Generate $Y \sim q(\cdot)$.
2: Flip a coin with Head probability $\frac{p(Y)}{cq(Y)}$. If Head, then return $X := Y$.
3: Else go to Step 1.

Based on this algorithm, Fill [Fil98] came up with a tricky idea for simulating from a stationary distribution π of an ergodic Markov chain. We will present here its slightly generalized version (without specific assumptions on the transition matrix). We include also a short proof of its correctness (although it is similar to that of Fill). Fix an integer $k \geq 1$ (time instance) and a state $\mathbf{e}_1 \in \mathbb{E}$. In the above settings, we want to simulate from $p(\cdot) = \pi(\cdot)$, given that we are able to simulate from $q(\cdot) = P(X_k = \cdot | X_0 = \mathbf{e}_1) = \mathbf{P}^k(\mathbf{e}_1, \cdot)$ (which can be done straightforwardly). Let $\overleftarrow{\phi}$ be an update function of the time reversed chain $\overleftarrow{\mathbf{X}}$ with t.m. $\overleftarrow{\mathbf{P}}(\mathbf{e}, \mathbf{e}') = \frac{\pi(\mathbf{e}')}{\pi(\mathbf{e})}\mathbf{P}(\mathbf{e}', \mathbf{e})$. Similarly as in the CFTP algorithm, we can use a sequence $N_r = 2^{r-1}$.

Algorithm 3 Fill's rejection algorithm.

Require: State space for all the following chains $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, ergodic chain $\mathbf{X}$ with a transition matrix $\mathbf{P}$.

1: Set $n = 1$.
2: Simulate the Markov chain $\mathbf{X}$ starting at time 0 in state $\mathbf{e}_1$ and run it till time $k = N_n$ using the transition matrix $\mathbf{P}$. Denote $X_k = \mathbf{e}_z$.
3: Treat $(X_k = \mathbf{e}_z, X_{k-1}, \ldots, X_0 = \mathbf{e}_1)$ as a path of the time reversed chain $(\overleftarrow{X}_0, \ldots, \overleftarrow{X}_k)$. For $s = 0, \ldots, n$ do: Assume $\overleftarrow{X}_s = \mathbf{e}$ and $\overleftarrow{X}_{s+1} = \mathbf{e}'$. Then generate $U_s \sim \mathtt{Unif}\{u : \overleftarrow{\phi}(\mathbf{e}, u) = \mathbf{e}'\}$.
4: Start M chains $\overleftarrow{\mathbf{Y}}^j, j = 1, \ldots, M$ so that $\overleftarrow{Y}^j_0 = j$ and *couple* them simulating $\overleftarrow{Y}^j_{s+1} = \phi(\overleftarrow{Y}^j_s, U_{s+1})$ (using the common update function ϕ and the randomness obtained in the previous step).
5: If all chains $\overleftarrow{\mathbf{Y}}^j, j = 1, \ldots, M$ have coupled before time k (and thus at time k they are all in state $\mathbf{e}_1$), output $\mathbf{e}_z$ and **stop**.
6: Erase all information, set $n = n + 1$ and go to Step 2.

In Step 3 we simulate random variables $U_1, \ldots, U_k$ in such a way that if we started the time reversed chain at $\mathbf{e}_z$ then we would obtain exactly the trajectory $(X_k = \mathbf{e}_z, X_{k-1}, \ldots, X_0 = \mathbf{e}_1)$. And this is what will happen for sure with one of $\overleftarrow{\mathbf{Y}}^j$ (the one starting in $\mathbf{e}_z$). That is why if all $\overleftarrow{\mathbf{Y}}^j, j = 1, \ldots, M$ coalesce, then we must have $\overleftarrow{Y}^j_k = \mathbf{e}_1$ for all $j = 1, \ldots, M$. Let $C_k(\mathbf{e})$ denote the event that all the chains $\overleftarrow{\mathbf{Y}}^j$ have coalesced before time k and that at this time they are all in $\mathbf{e}$. Of course we have for any $\mathbf{e}$ that $\overleftarrow{\mathbf{P}}^k(\mathbf{e},\mathbf{e}_z) \leq Pr[C_k(\mathbf{e}_z)]$. Now we are ready to choose the constant c from Alg. 2.

$$\frac{\pi(\mathbf{e})}{\mathbf{P}^k(\mathbf{e}_1,\mathbf{e})} = \frac{\pi(\mathbf{e}_1)}{\overleftarrow{\mathbf{P}}^k(\mathbf{e},\mathbf{e}_1)} \leq \frac{\pi(\mathbf{e}_1)}{Pr[C_k(\mathbf{e}_1)]} =: c.$$

Thus we simulate from distribution $\mathbf{P}^k(\mathbf{e}_1,\cdot)$, say $\mathbf{e}_z$ was obtained. We should accept $\mathbf{e}_z$ with probability

$$\frac{\pi(\mathbf{e}_z)}{c\mathbf{P}^k(\mathbf{e}_1,\mathbf{e}_z)} = \frac{Pr[C_k(\mathbf{e}_1)]}{\pi(\mathbf{e}_1)}\frac{\pi(\mathbf{e}_z)}{\mathbf{P}^k(\mathbf{e}_1,\mathbf{e}_z)} = \frac{Pr[C_k(\mathbf{e}_1)]}{\overleftarrow{\mathbf{P}}^k(\mathbf{e}_z,\mathbf{e}_1)}.$$

The whole point is that this is exactly the acceptance probability in Alg. 3:

$$\begin{aligned} &Pr(\overleftarrow{Y}^1_k = \ldots = \overleftarrow{Y}^M_k = \mathbf{e}_1 \mid \overleftarrow{X}_0 = \mathbf{e}_z, \overleftarrow{X}_k = \mathbf{e}_1, \overleftarrow{Y}^j_0 = \mathbf{e}_j, j=1,\ldots,M) \\ &= \frac{Pr(\overleftarrow{Y}^1_k = \ldots = \overleftarrow{Y}^M_k = \mathbf{e}_1 \mid \overleftarrow{X}_0 = \mathbf{e}_z, \overleftarrow{Y}^j_0 = \mathbf{e}_j, j=1,\ldots,M)}{Pr(\overleftarrow{X}_k = \mathbf{e}_1 \mid \overleftarrow{X}_0 = \mathbf{e}_z, \overleftarrow{Y}^j_0 = \mathbf{e}_j, j=1,\ldots,M)}. \end{aligned}$$

In this setting, X_0 and $\mathbf{Y}^j$ are independent, thus

$$= \frac{Pr(\overleftarrow{Y}^1_k = \ldots = \overleftarrow{Y}^M_k = \mathbf{e}_1 \mid \overleftarrow{Y}^j_0 = \mathbf{e}_j, j=1,\ldots,M)}{Pr(\overleftarrow{X}_t = \mathbf{e}_1 \mid \overleftarrow{X}_0 = \mathbf{e}_z)} = \frac{Pr[C_k(\mathbf{e}_z)]}{\overleftarrow{\mathbf{P}}^k(\mathbf{e}_z,\mathbf{e}_1)}.$$

2.3 Method 3: Strong stationary duality

A random variable T is a **Strong Stationary Time** (SST) for $\mathbf{X}$ if it is a stopping time independent from X_T such that X_T has the distribution π. It was introduced in [AD86] mainly for studying the rate of convergence of the chain, but it is also applicable for exact sampling, simply by simulating the chain until time T we obtain unbiased sample from π. Although there are many examples where such SST was find (probably the best example is Top-To-Random card shuffling), the problem is that the examples were usually found “ad hoc”, in general it is not easy to come up with SST.

Diaconis and Fill [DF90b] came up with a systematic way of finding an SST, which we will describe here directly with application to exact sampling.

Let $\mathbb{E}^* = \{\mathbf{e}^*_1,\ldots,\mathbf{e}^*_N\}$ be the state space of an absorbing Markov chain $\mathbf{X}^*$ with initial distribution ν^* and transition matrix $\mathbf{P}^*$, whose unique absorbing state is denoted by $\mathbf{e}^*_N$. An $N\times M$ matrix Λ is said to be a *link* if it is a stochastic matrix such that $\Lambda(\mathbf{e}^*_N,\mathbf{e}) = \pi(\mathbf{e})$ for all $\mathbf{e}\in\mathbb{E}$. We say that $\mathbf{X}^*$ is a **strong stationary dual** (SSD) of $\mathbf{X}$ with link Λ if

$$\nu = \nu^*\Lambda \quad \text{and} \quad \Lambda\mathbf{P} = \mathbf{P}^*\Lambda. \tag{2.1}$$

In this chapter we assume that the SSD has the same state space, i.e., $\mathbb{E}^* = \mathbb{E}$. For the general case, see [DF90b]. The sample path of the chain $\mathbf{X}^*$ can be constructed from a sample path of $\mathbf{X}$ as follows. Start with $X_0 = \mathbf{e}^0$ and (using additional randomness) set

$$X^*_0 = \mathbf{e}^{*0} \quad \text{with probability} \quad \frac{\nu^*(\mathbf{e}^{*0})\Lambda(\mathbf{e}^{*0},\mathbf{e}^0)}{\nu(\mathbf{e}^0)}.$$

Then we proceed as follows. Assume $X_0 = \mathbf{e}^0,\ldots,X_{k-1} = \mathbf{e}^{(k-1)}$ and $X^*_0 = \mathbf{e}^{*0},\ldots,X^*_{k-1} = \mathbf{e}^{*(k-1)}$. If $X_k = \mathbf{e}^k$ have been chosen, then set

$$X^*_k = \mathbf{e}^{*k} \quad \text{with probability} \quad \frac{\mathbf{P}^*(\mathbf{e}^{*(k-1)},\mathbf{e}^{*(k)})\Lambda(\mathbf{e}^{*k},\mathbf{e}^k)}{\triangle(\mathbf{e}^{*(k-1)},\mathbf{e}^k)},$$

where $\triangle = \mathbf{P}^*\Lambda$. This construction yields a bivariate chain (X_k^*, X_k) such that $Pr(X_k = \cdot | X_0^* = \mathbf{e}^{*0}, \ldots, X_k^* = \mathbf{e}^{*k}) = \Lambda(\mathbf{e}^{*k}, \cdot)$ (consult [DF90a], [DF90b]). This implies that T, the first time the chain $\mathbf{X}^*$ hits the state $\mathbf{e}_N^*$ (the absorbing one) and the value of X_T are independent. Moreover, the distribution of X_T is $\Lambda(\mathbf{e}_N^*, \cdot) = \pi(\mathbf{e})$.

In summary, we are able to couple two chains in such a way that when one hits a specific state ($\mathbf{e}_N^*$) then the other has a stationary distribution. This way we can obtain an unbiased sample from π, i.e., we can perform an exact sampling. Note that having the SST T lets one also study the rate of convergence (the main application of this duality in [DF90b]): the time to absorption T for $\mathbf{X}^*$ is an SST for $\mathbf{X}$. In many examples, SST have been found ad hoc. The above duality approach provided the first systematic way of finding them. Below we present the above mentioned description of SSD-based exact sampling in algorithmic form.

Algorithm 4 Exact sampling based on SSD.

Require: Ergodic chain $\mathbf{X}$ and absorbing chain $\mathbf{X}^*$ on the same state space $\mathbb{E}$, link Λ.

1: Start with $X_0 = \mathbf{e}^0$ and set $X_0^* = \mathbf{e}^{*0}$ with probability $\frac{\nu^*(\mathbf{e}^{*0})\Lambda(\mathbf{e}^{*0},\mathbf{e}^0)}{\nu(\mathbf{e}^0)}$.
2: If $X_0^* = \mathbf{e}_N^*$ then output X_0 and **stop**.
3: Set $n = 1$.
4: Having $X_{n-1} = \mathbf{e}^{n-1}$ set $X_n = \mathbf{e}^n$ with probability $P(e^{n-1}, e^n)$.
5: Having $X_{n-1} = \mathbf{e}^{n-1}, X_n = \mathbf{e}^n, X_{n-1}^* = \mathbf{e}^{*n-1}$ set $X_n^* = \mathbf{e}^{*n}$ with probability $\frac{\mathbf{P}^*(\mathbf{e}^{*(n-1)},\mathbf{e}^{*(n)})\Lambda(\mathbf{e}^{*n},\mathbf{e}^n)}{\triangle(\mathbf{e}^{*(n-1)},\mathbf{e}^n)}$, where $\triangle = \mathbf{P}^*\Lambda$.
6: If $X_n^* = \mathbf{e}_N^*$ then output X_n and **stop**.
7: Set $n = n + 1$ and go to Step 4 (keep previously simulated $X_n = \mathbf{e}^n, X_n^* = \mathbf{e}^{*n}$ for new n).

3 Monotonicites in Markov chains

In the previous section we briefly described some methods for exact sampling. Note however that CFTP and Fill's rejection algorithm, as they stand, are very inefficient (the number of chains one has to simulate is equal to the cardinality of $\mathbb{E}$) and no concrete way for finding the SSD was given (how to choose/find Λ and $\mathbf{P}^*$). This is where monotonicities come into play. Each of the methods can be **efficiently** applied if the chain is monotone in some way.

So far we did not need any structure on $\mathbb{E}$. However in many examples there is a natural ordering of the state space, e.g., a total ordering, a coordinatewise ordering, etc. From now on we assume that $\mathbb{E}$ is equipped with a partial ordering $\preceq$, making $(\mathbb{E}, \preceq)$ a poset. We also assume that $\mathbf{e}_1$ is the minimum and $\mathbf{e}_M$ is the maximum. We will use the following notation. We say that $U \in \mathbb{E}$ is an **upset** if $(\mathbf{e}_1 \preceq \mathbf{e}_2, \mathbf{e}_1 \in U) \Rightarrow \mathbf{e}_2 \in U$. Similarly, we say that $D \in \mathbb{E}$ is a **downset** if $(\mathbf{e}_1 \preceq \mathbf{e}_2, \mathbf{e}_2 \in D) \Rightarrow \mathbf{e}_1 \in D$. For given $\mathbf{e} \in \mathbb{E}$ we define $\{\mathbf{e}\}^\uparrow := \{\mathbf{e}' : \mathbf{e} \preceq \mathbf{e}'\}$ and $\{\mathbf{e}\}^\downarrow := \{\mathbf{e}' : \mathbf{e}' \preceq \mathbf{e}\}$. Note that each $\{\mathbf{e}\}^\uparrow$ ($\{\mathbf{e}\}^\downarrow$) is an upset (downset), but, in general, not vice versa.

All the monotonicities we are about to define are defined for chains on a common state space $\mathbb{E}$ with respect to a fixed partial ordering $\preceq$. By $\mathbf{X} \in \mathcal{P}$ we mean that $\mathbf{X}$ has monotonicity property $\mathcal{P}$ keeping in mind that it is defined w.r.t. the fixed partial ordering $\preceq$.

Usual and weak stochastic monotonicity.

Definition 3.1. A Markov chain $\mathbf{X}$ with transition matrix $\mathbf{P}$ is **stochastically monotone** (we write $\mathbf{X} \in \mathcal{S}$) if and only if for all upsets U and all $\mathbf{e} \preceq \mathbf{e}' \in \mathbb{E}$ we have $\mathbf{P}(\mathbf{e}, U) \leq \mathbf{P}(\mathbf{e}', U)$.

Remark 3.1. Since the complement of any upset is a downset, the condition for stochastic monotonicity can be equivalently given by: for all downsets D and all $\mathbf{e} \preceq \mathbf{e}' \in \mathbb{E}$ we have $\mathbf{P}(\mathbf{e}, D) \geq \mathbf{P}(\mathbf{e}', D)$.

Stochastic monotonicity can be equivalently defined in the following way. For two random variables Y_1, Y_2 (with distribution functions ν_1, ν_2) on $\mathbb{E}$, we say that $Y_1 \preceq_{st} Y_2$ (or $\nu_1 \preceq_{st} \nu_2$) $\iff E[f(Y_1)] \leq E[f(Y_2)]$ for all nondecreasing (w.r.t. $\preceq$) functions $f : \mathbb{E} \to \mathbb{R}$. Then the Markov chain $\mathbf{X}$ with the transition matrix $\mathbf{P}$ is stochastically monotone if and only if $\nu_1 \preceq_{st} \nu_2$ implies $\nu_1\mathbf{P} \preceq_{st} \nu_2\mathbf{P}$.

Recall that K is an *upward kernel* if it is a Markov kernel such that $K(\mathbf{e}_i, \cdot)$ is supported on $\{\mathbf{e}_j \in \mathbb{E} : \mathbf{e}_i \preceq \mathbf{e}_j\}$. The following lemma goes back to Strassen [Str65] (and is part of Theorem 1 in [KKO77]).

Lemma 3.1. *A Markov chain $\mathbf{X}$ with transition matrix $\mathbf{P}$ is stochastically monotone if and only if for all $\mathbf{e} \preceq \mathbf{e}'$ there exists an upward kernel $K_{\mathbf{e},\mathbf{e}'}$ such that* $\mathbf{P}(\mathbf{e}', \mathbf{e}_j) = \sum_{\mathbf{e}_i : \mathbf{e}_j \preceq \mathbf{e}_i} \mathbf{P}(\mathbf{e}, \mathbf{e}_i) K_{\mathbf{e},\mathbf{e}'}(\mathbf{e}_i, \mathbf{e}_j)$.

Replacing any upset (downset) in Definition 3.1 with a specific one we obtain the notion of weak monotonicity.

Definition 3.2. A Markov chain $\mathbf{X}$ with transition matrix $\mathbf{P}$ is **weakly-$^\uparrow$ monotone** (we write $\mathbf{X} \in \mathcal{W}^\uparrow$) if and only if for all $\mathbf{e} \preceq \mathbf{e}', \mathbf{e}_j \in \mathbb{E}$ we have $\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\uparrow) \leq \mathbf{P}(\mathbf{e}', \{\mathbf{e}_j\}^\uparrow)$.

The chain is **weakly-$^\downarrow$ monotone** (we write $\mathbf{X} \in \mathcal{W}^\downarrow$) if and only if for all $\mathbf{e} \preceq \mathbf{e}', \mathbf{e}_j \in \mathbb{E}$ we have $\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\downarrow) \geq \mathbf{P}(\mathbf{e}', \{\mathbf{e}_j\}^\downarrow)$.

We define $\mathcal{W} := \mathcal{W}^\uparrow \cap \mathcal{W}^\downarrow$.

Realizable monotonicity. This notion of monotonicity is defined in terms of the update rule of the chain given in Definition 2.1.

Definition 3.3. A Markov chain $\mathbf{X}$ with transition matrix $\mathbf{P}$ is **realizable monotone** if there exists a monotone update rule (preserving the ordering), i.e.,

$$\forall (u \in [0,1]) \ \forall (\mathbf{e} \preceq \mathbf{e}') \quad \phi(\mathbf{e}, u) \preceq \phi(\mathbf{e}', u).$$

This definition implies that for any states $\mathbf{e} \preceq \mathbf{e}'$ and upset U we have

$$\phi(\mathbf{e}, u) \in U \Rightarrow \phi(\mathbf{e}', u) \in U. \tag{3.1}$$

Finding a monotone update rule is often a challenging task, and proving that none exist can be even harder.

Möbius monotonicity.
We can identify the ordering $\preceq$ with the matrix $\mathbf{C}(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{1}(\mathbf{e}_i \preceq \mathbf{e}_j)$. We can always rearrange the states in such a way that $\mathbf{C}$ is upper triangular (keeping in mind that the enumerations of the states in $\mathbf{C}$ and $\mathbf{P}$ must preserve the same order), thus invertible. The inverse of $\mathbf{C}$ is usually denoted by $\mu \equiv \mathbf{C}^{-1}$ and called the **Möbius function**.

Definition 3.4. The function $\mathbf{f} : \mathbb{E} \to \mathbb{R}^M$ is **Möbius-$^\downarrow$** (**Möbius-$^\uparrow$**) monotone if $\mathbf{f}(\mathbf{C}^T)^{-1} \geq 0$ ($\mathbf{f}\mathbf{C}^{-1} \geq 0$), i.e., each entry is nonnegative.

Definition 3.5. A Markov chain $\mathbf{X}$ with transition matrix $\mathbf{P}$ is **Möbius-$^\downarrow$ monotone** (we write $\mathbf{X} \in \mathcal{M}^\downarrow$) if

$$\mathbf{C}^{-1}\mathbf{P}\mathbf{C} \geq 0 \quad \text{(each entry nonnegative).}$$

Equivalently, in terms of the transition probabilities,

$$\forall(\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}) \qquad \sum_{\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e})\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\downarrow) \geq 0.$$

The chain is **Möbius-$^\uparrow$ monotone** (we write $\mathbf{X} \in \mathcal{M}^\uparrow$) if

$$(\mathbf{C}^T)^{-1}\mathbf{P}\mathbf{C}^T \geq 0 \quad \text{(each entry nonnegative).}$$

In terms of the transition probabilities this is

$$\forall(\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}) \qquad \sum_{\mathbf{e} \preceq \mathbf{e}_i} \mu(\mathbf{e}, \mathbf{e}_i)\mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^\uparrow) \geq 0.$$

We define $\mathcal{M} := \mathcal{M}^\uparrow \cap \mathcal{M}^\downarrow$.

In the applications, checking Möbius monotonicity is usually not harder than checking stochastic monotonicity. First, note that the inverse of $\mathbf{C}$ (i.e., the Möbius function of the ordering) is known for many natural partial orderings (however, its derivation is often not trivial). To mention a few:

- **E1** For $\mathbb{E} = \{1, \ldots, M\}$ and a linear ordering $\preceq := \leq$ the Möbius function is given by $\mu(i, i) = 1$, $\mu(i, i+1) = -1$ and $\mu(i, j) = 0$ for $j \notin \{i, i+1\}$.
- **E2** For $\mathbb{E} = \{0, 1\}^d$ with the coordinate-wise partial ordering $\mathbf{e} \preceq \mathbf{e}'$, if $\mathbf{e}(i) \leq \mathbf{e}'(i), i = 1, \ldots, d$, the Möbius function is given by $\mu(\mathbf{e}, \mathbf{e}') = (-1)^{|\mathbf{e}| - |\mathbf{e}'|}$ if $\mathbf{e} \preceq \mathbf{e}'$ and 0 otherwise (where $|\mathbf{e}| = \sum_{i=1}^d \mathbf{e}(i)$).
- **E3** For $\mathbb{E} = \{0, 1, \ldots, N\}^d$ with coordinate-wise partial ordering $\mathbf{e} \preceq \mathbf{e}'$, if $\mathbf{e}(i) \leq \mathbf{e}'(i), i = 1, \ldots, d$, the Möbius function is given by $\mu(\mathbf{e}, \mathbf{e}') = (-1)^{|\mathbf{e}'| - |\mathbf{e}|}$ if $\mathbf{e}'(k) = \mathbf{e}(k)$ or $\mathbf{e}'(k) = \mathbf{e}(k) + 1$ for each $k = 1, \ldots, d$.
- **E4** For a finite set I let $P(I)$ be the set of all partitions of I. Let $\alpha, \beta \in P(I)$. The typically considered partial order is the following: $\alpha \preceq \beta$ if $\forall(A \in \alpha)\exists(B \in \beta)(A \subseteq B)$. As derived in [Com70], the Möbius function is given by $\mu(\alpha, \beta) = \mathbf{1}_{\alpha \preceq \beta}(-1)^{|\alpha| + |\beta|} \prod_{B \in \beta}(l_B^\alpha - 1)!$, where l_B^α is the number of atoms from α in $B \in \beta$.

Checking the Möbius monotonicity of a chain having a Möbius function turns out to be feasible in many cases. For the total ordering (**E1**), exemplary calculations are given in [Lor18]. The computations checking Möbius-$^\downarrow$ monotonicity for some nonsymmetric random walk on the cube (**E3**) are given in [LS12a]. For the chain corresponding to a nonstandard queue network, the computations are given in [Lor18]. The partial ordering on partitions (**E4**) was considered in the context of duality in [HM16].

4 Applications of monotonicites

4.1 Realizable monotonicity and an efficient coupling from the past algorithm

The CFTP algorithm given in Alg. 1 is very inefficient. The number of chains we have to run is equal to the size of the state space. In most cases where CFTP is to be applied, the size of the state space is huge (e.g., exponential in some parameter). The main idea of the algorithm was to run the chains "from the past" and check if all of them have coupled before time 0. Note that if we have a monotone update rule and say $X_{-m} = \mathbf{e} \preceq X'_{-m} = \mathbf{e}'$, then

$$X_{-m+1} = \phi(X_{-m}, U_{-m+1}) \preceq \phi(X'_{-m}, U_{-m+1}) = X'_{-m+1}.$$

Thus it is enough to start only two chains: $X_0^1 = \mathbf{e}_1$ and $X_0^2 = \mathbf{e}_M$. Summarizing, if the chain is *realizable monotone* and has the minimum and the maximum, then we have an efficient CFTP algorithm:

Algorithm 5 Efficient coupling from the past.

Require: State space $\mathbb{E}$, ergodic chain $\mathbf{X}$, monotone update rule ϕ

1. Set $n = 1$
2. Start two chains at time $-N_n$, one at the minimum $\mathbf{e}_1$, the other at the maximum $\mathbf{e}_M$. Run the chains till time 0 using the same update rule ϕ and iid random variables $U_{-N_n+1}, U_{-N_n+2}, \ldots, U_{-1}, U_0$ uniformly distributed on $[0, 1]$ (the same for each chain).
3. If both chains in previous step end up in the same state $\mathbf{e}_0$ at time 0, then output $\mathbf{e}_0$ and **stop**.
4. Set $n = n + 1$ and go to Step 2 (keep previously used $\{U_i\}_{0 \leq i \leq -N_n+1}$ for new n).

4.2 Stochastic monotonicity and an efficient Fill's rejection algorithm

Similarly to the general CFTP algorithm given in Alg. 1, Fill's rejection Alg. 3 is very inefficient. This is due to the fact that we have to start (and simulate) as many chains as there are elements of the state space. It turns out that the algorithm can be made efficient by assuming stochastic monotonicity of the time reversed chain $\overleftarrow{X}$. (This condition is weaker, as forthcoming sections will show, than being realizable monotone).

Assume for the moment that $\overleftarrow{X}$ is *realizable monotone* (this will soon be relaxed to stochastic monotonicity). Assume that $\mathbf{e}_i \preceq \mathbf{e}_j$ and that $\overleftarrow{Y}^i_s = \mathbf{e}_i \preceq \overleftarrow{Y}^j_s = \mathbf{e}_j$ for some $s \leq k$. Realizable monotonicity implies that $\overleftarrow{Y}^i_{s+1} = \phi(\overleftarrow{Y}^i_s, U_s) \preceq \phi(\overleftarrow{Y}^j_s, U_s) = \overleftarrow{Y}^j_{s+1}$. This means that then in Step 5 of the algorithm, checking the coalescence of all M chains is equivalent to checking only that the chain $\overleftarrow{\mathbf{Y}}^M$ (the one started in $\mathbf{e}_M$) has already reached the minimum $\mathbf{e}_1$ at time k. In other words, it is enough to simulate just one chain $\overleftarrow{\mathbf{Y}}^M$.

Now we relax the realizable monotonicity requirement, assuming only that $\overleftarrow{X}$ is stochastically monotone. Similarly, we want to have an efficient version of the algorithm simulating only one chain $\overleftarrow{\mathbf{Y}}^M$ (denoted simply by $\mathbf{Y}$). This time we do not generate U_s as in Step 3 of Alg. 3. We make use of Lemma 3.1 instead. Assume that at some step s we have $\overleftarrow{X}_s = \mathbf{e}_1$, $\overleftarrow{X}_{s+1} = \mathbf{e}_2$ and $\overleftarrow{Y}_s = \mathbf{e}_i$. Then, since $\mathbf{e}_1 \preceq \mathbf{e}_2$, we may choose a state $\mathbf{e}_j$ for $\overleftarrow{Y}_{s+1}$ with probability $K_{\mathbf{e}_1,\mathbf{e}_i}(\mathbf{e}_2, \mathbf{e}_j)$. This constructions ensures that $\overleftarrow{X}_s \preceq \overleftarrow{Y}_s, s = 0, \ldots, k$. Thus, similarly, only the condition $\overleftarrow{Y}_k = \mathbf{e}_1$ must be checked. In summary, we have:

Algorithm 6 Efficient Fill's rejection algorithm.

Require: State space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, ergodic chain $\mathbf{X}$ whose time reversal $\overleftarrow{X}$ is stochastically monotone and set of kernels $K_{\mathbf{e},\mathbf{e}'}$ for all $\mathbf{e} \preceq \mathbf{e}'$.

1: Set $n = 1$
2: Simulate the Markov chain $\mathbf{X}$ starting at time 0 in state $\mathbf{e}_1$ and run it till time $k = N_n$ using the transition matrix $\mathbf{P}$. Denote $X_t = \mathbf{e}_z$
3: Simulate the chain $\overleftarrow{\mathbf{Y}}$ starting at the maximum, i.e., $\overleftarrow{Y}_0 = \mathbf{e}_M$ in the following way: Assume at time s we have $\overleftarrow{X}_s = \mathbf{e}_1, \overleftarrow{Y}_s = \mathbf{e}_i$ and $\overleftarrow{X}_{s+1} = \mathbf{e}_2$. Set $\overleftarrow{Y}_{s+1} = \mathbf{e}_j$ with probability $K_{\mathbf{e}_1,\mathbf{e}_i}(\mathbf{e}_2, \mathbf{e}_j)$.
4: If $\overleftarrow{Y}_k = \mathbf{e}_1$, then output $\mathbf{e}_z$ and **stop**
5: Erase all information, set $n = n + 1$ and go to Step 2.

4.3 Möbius monotonicity and strong stationary duality

In Section 2.3 we presented an exact sampling algorithm based on strong stationary duality. Note however that no recipe was given on how to find such a dual. The duality was introduced in [DF90b], where the recipe was given only in case the time reversed chain $\overleftarrow{\mathbf{X}}$ was stochastically monotone w.r.t. a total ordering. In [LS12a] an extension to partial orderings was given. Surprisingly, it turned out that not the usual stochastic monotonicity, but rather Möbius monotonicity, was required. We recall here the main theorem from [LS12a].

Theorem 4.1 (Lorek and Szekli [LS12a]). *Let $\mathbf{X}$ be an ergodic Markov chain on a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ which is partially ordered by $\preceq$ and has the maximum $\mathbf{e}_M$. For a stationary distribution π and an initial distribution ν we assume that*

(i) $g(\mathbf{e}) = \frac{\nu(\mathbf{e})}{\pi(\mathbf{e})}$ is Möbius-$^\downarrow$ monotone,

(ii) $\overleftarrow{\mathbf{X}}$ is Möbius-$^\downarrow$ monotone.

Then there exists a strong stationary dual chain $\mathbf{X}^*$ on $\mathbb{E}^* = \mathbb{E}$ with link a truncated stationary distribution $\Lambda(\mathbf{e}_j, \mathbf{e}_i) = \mathbf{1}(\mathbf{e}_i \preceq \mathbf{e}_j)\frac{\pi(\mathbf{e}_i)}{H(\mathbf{e}_j)}$, where $H(\mathbf{e}_j) = \sum_{\mathbf{e}:\mathbf{e}\preceq\mathbf{e}_j} \pi(\mathbf{e})$. The initial distribution and the transitions of $\mathbf{X}^*$ are given, respectively, by

$$\nu^*(\mathbf{e}_i) = H(\mathbf{e}_i) \sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) g(\mathbf{e}), \tag{4.1}$$

$$\mathbf{P}^*(\mathbf{e}_i, \mathbf{e}_j) = \frac{H(\mathbf{e}_j)}{H(\mathbf{e}_i)} \sum_{\mathbf{e}:\mathbf{e}\succeq\mathbf{e}_j} \mu(\mathbf{e}_j, \mathbf{e}) \overleftarrow{\mathbf{P}}(\mathbf{e}, \{\mathbf{e}_i\}^\downarrow). \tag{4.2}$$

(The Möbius monotonicity of the function $g(\mathbf{e})$ means that the resulting $\nu^*(\mathbf{e})$ is nonnegative).

Remark 4.1. Note that the existence of the minimum is not required in Theorem 4.1. However, if it exists and if the chain $\mathbf{X}$ starts at the minimum (i.e., $Pr(X_0 = \mathbf{e}_1) = 1$), then so does the dual chain (i.e., $Pr(X_0^* = \mathbf{e}_1)$). Similarly one can construct an SSD chain when there is a minimum $\mathbf{e}_1$ and the time reversed chain $\overleftarrow{\mathbf{X}}$ is Möbius-$^\uparrow$ monotone, see Corollary 3.1 in [LS12a].

More examples of SSDs constructed on partially ordered state spaces can be found in [LS16].

5 Relations between monotonicities in Markov chains

Fix a state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ and partial ordering $\preceq$. Recall that by $\mathbf{X} \in \mathcal{P}$ we mean that the chain has the monotonicity property $\mathcal{P}$, which is defined with respect to this given state space and ordering. For example, the implication "if $\mathbf{X} \in \mathcal{P}_1$ then $\mathbf{X} \in \mathcal{P}_2$" means that if $\mathbf{X}$ is $\mathcal{P}_1$-monotone then it is $\mathcal{P}_2$-monotone with respect to *the same* state space and ordering. For a general ordering $\preceq$, we present the relations between the different concepts of monotonicity in Theorem 5.1.

Theorem 5.1. *For a discrete time Markov chain $\mathbf{X}$ on a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ which is partially ordered by $\preceq$, we have the following implications:*

1. *If $\mathbf{X} \in \mathcal{R}$ then $\mathbf{X} \in \mathcal{S}$*
2. *If $\mathbf{X} \in \mathcal{S}$ then $\mathbf{X} \in \mathcal{W}^{\uparrow}$*
3. *If $\mathbf{X} \in \mathcal{S}$ then $\mathbf{X} \in \mathcal{W}^{\downarrow}$*
4. *If $\mathbf{X} \in \mathcal{M}^{\uparrow}$ then $\mathbf{X} \in \mathcal{W}^{\uparrow}$*
5. *If $\mathbf{X} \in \mathcal{M}^{\downarrow}$ then $\mathbf{X} \in \mathcal{W}^{\downarrow}$*

We derive and recall some useful properties of the Möbius function of a partial ordering. First we will show that $\sum_{\mathbf{e}_i \in \mathbb{E}} \mu(\mathbf{e}_i, \mathbf{e}) = 0$ for any poset with the minimum state (we denote it by $\mathbf{e}_1$) and that $\sum_{\mathbf{e}_i \in \mathbb{E}} \mu(\mathbf{e}, \mathbf{e}_i) = 0$ for any poset with the maximum state (denoted by $\mathbf{e}_M$).

It is known (see [Rot64]) that the matrix $\mathbf{C}^{-1} = \mu$ can be calculated recursively:

$$\mu(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} 1 & \text{if} \quad \mathbf{e}_i = \mathbf{e}_j \\ -\sum_{\mathbf{e}_i \prec \mathbf{e}_k \preceq \mathbf{e}_j} \mu(\mathbf{e}_k, \mathbf{e}_j) & \text{if} \quad \mathbf{e}_i \prec \mathbf{e}_j, \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

or by inverting the matrix $\mathbf{C}$ using Gauss–Jordan elimination by columns instead of rows:

$$= \begin{cases} 1 & \text{if} \quad \mathbf{e}_i = \mathbf{e}_j \\ -\sum_{\mathbf{e}_i \preceq \mathbf{e}_k \prec \mathbf{e}_j} \mu(\mathbf{e}_i, \mathbf{e}_k) & \text{if} \quad \mathbf{e}_i \prec \mathbf{e}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Therefore, using (5.1): for any poset with the minimum state and any state $\mathbf{e}$ which is not the minimum, we have

$$\sum_{\mathbf{e}_i \in \mathbb{E}} \mu(\mathbf{e}_i, \mathbf{e}) = \sum_{\mathbf{e}_i \in \mathbb{E}: \mathbf{e}_i \not\preceq \mathbf{e}} \mu(\mathbf{e}_i, \mathbf{e}) + \sum_{\mathbf{e}_1 \prec \mathbf{e}_i \preceq \mathbf{e}} \mu(\mathbf{e}_i, \mathbf{e}) + \mu(\mathbf{e}_1, \mathbf{e})$$
$$= 0 - \mu(\mathbf{e}_1, \mathbf{e}) + \mu(\mathbf{e}_1, \mathbf{e}) = 0.$$

Similarly, using (5.2): for any poset with the maximum state and for any state $\mathbf{e}$ which is not the maximum, we have

$$\sum_{\mathbf{e}_i \in \mathbb{E}} \mu(\mathbf{e}, \mathbf{e}_i) = \sum_{\mathbf{e}_i \in \mathbb{E}: \mathbf{e} \not\preceq \mathbf{e}_i} \mu(\mathbf{e}, \mathbf{e}_i) + \sum_{\mathbf{e} \preceq \mathbf{e}_i \prec \mathbf{e}_M} \mu(\mathbf{e}, \mathbf{e}_i) + \mu(\mathbf{e}, \mathbf{e}_M)$$
$$= 0 - \mu(\mathbf{e}, \mathbf{e}_M) + \mu(\mathbf{e}, \mathbf{e}_M) = 0.$$

We will write $\mathbf{e}^+$ for an arbitrary successor of $\mathbf{e}$. For a poset $(\mathbb{E}, \preceq)$ with the maximum $\mathbf{e}_M$ we can consider the subspaces $\{\mathbf{e}\}^\uparrow, \{\mathbf{e}^+\}^\uparrow$ with the minimum states $\mathbf{e}$ and $\mathbf{e}^+$ respectively. From the above consideration, we have

$$\sum_{\mathbf{e}_i:\mathbf{e}\preceq\mathbf{e}_i,\mathbf{e}^+\not\preceq\mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}') = \sum_{\mathbf{e}_i:\mathbf{e}\preceq\mathbf{e}_i\preceq\mathbf{e}_M} \mu(\mathbf{e}_i, \mathbf{e}') - \sum_{\mathbf{e}_i:\mathbf{e}^+\preceq\mathbf{e}_i\preceq\mathbf{e}_M} \mu(\mathbf{e}_i, \mathbf{e}') = \tag{5.3}$$

$$\begin{cases} 1-0 & \text{if } \mathbf{e}' = \mathbf{e}, \\ 0-1 & \text{if } \mathbf{e}' = \mathbf{e}^+ \\ 0-0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } \mathbf{e}' = \mathbf{e}, \\ -1 & \text{if } \mathbf{e}' = \mathbf{e}^+, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

Similarly, for a poset $(\mathbb{E}, \preceq)$ with the minimum state $\mathbf{e}_1$ we can consider the subspaces $\{\mathbf{e}\}^\downarrow, \{\mathbf{e}^+\}^\downarrow$ with the maximum states $\mathbf{e}$ and $\mathbf{e}^+$ respectively. We have:

$$\sum_{\mathbf{e}_i:\mathbf{e}_i\preceq\mathbf{e}^+,\mathbf{e}_i\not\preceq\mathbf{e}} \mu(\mathbf{e}', \mathbf{e}_i) = \sum_{\mathbf{e}_i:\mathbf{e}_1\preceq\mathbf{e}_i\preceq\mathbf{e}^+} \mu(\mathbf{e}', \mathbf{e}_i) - \sum_{\mathbf{e}_i:\mathbf{e}_1\preceq\mathbf{e}_i\preceq\mathbf{e}} \mu(\mathbf{e}', \mathbf{e}_i) = \tag{5.5}$$

$$\begin{cases} 0-1 & \text{if } \mathbf{e}' = \mathbf{e}, \\ 1-0 & \text{if } \mathbf{e}' = \mathbf{e}^+, \\ 0-0 & \text{otherwise} \end{cases} = \begin{cases} -1 & \text{if } \mathbf{e}' = \mathbf{e}, \\ 1 & \text{if } \mathbf{e}' = \mathbf{e}^+, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

Proof of Theorem 5.1. Implications 1–3 are known (see, e.g., [FM01]), we include short proofs in our notation for completeness.

1. We want to show that for any states $\mathbf{e}_i \preceq \mathbf{e}_j$ and any upset U, the inequality $\mathbf{P}(\mathbf{e}_i, U) \leq \mathbf{P}(\mathbf{e}_j, U)$ is fulfilled. From the definition of update function and the property of realizable monotonicity (3.1), we have

$$\mathbf{P}(\mathbf{e}_i, U) = \int_0^1 \mathbf{1}(\phi(\mathbf{e}_i, u) \in U)du \leq \int_0^1 \mathbf{1}(\phi(\mathbf{e}_j, u) \in U)du = \mathbf{P}(\mathbf{e}_j, U)$$

as the indicator of set is equal to 1 when the indicator of its subset is equal to 1.

2. & 3. For any state $\mathbf{e}$, $\{\mathbf{e}\}^\uparrow$ is an upset and $\{\mathbf{e}\}^\downarrow$ is a downset. Thus stochastic monotonicity implies that the monotonicity is preserved for any $\{\mathbf{e}\}^\uparrow$ and $\{\mathbf{e}\}^\downarrow$, which is the definition of $^\uparrow$-weak and $^\downarrow$-weak monotonicity.

4. & 5. Möbius-$^\downarrow$ monotonicity means

$$\forall_{\mathbf{e}\preceq\mathbf{e}_k} \sum_{\mathbf{e}_i:\mathbf{e}\preceq\mathbf{e}_i} \mu(\mathbf{e}, \mathbf{e}_i)\mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^\downarrow) \geq 0,$$

thus for arbitrary $\mathbf{e}_j \in \mathbb{E}$ we have

$$\sum_{\mathbf{e}:\mathbf{e}_j\preceq\mathbf{e},\mathbf{e}_j^+\not\preceq\mathbf{e}} \sum_{\mathbf{e}_i:\mathbf{e}\preceq\mathbf{e}_i} \mu(\mathbf{e}, \mathbf{e}_i)\mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^\downarrow) \geq 0.$$

Changing the order of summation we have

$$\sum_{\mathbf{e}_i:\mathbf{e}_j\preceq\mathbf{e}_i} \sum_{\mathbf{e}:\mathbf{e}\preceq\mathbf{e}_i,\mathbf{e}_j\preceq\mathbf{e},\mathbf{e}_j^+\not\preceq\mathbf{e}} \mu(\mathbf{e}, \mathbf{e}_i)\mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^\downarrow) \geq 0,$$

$$\sum_{\mathbf{e}_i:\mathbf{e}_i\succeq\mathbf{e}_j} \mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^\downarrow) \sum_{\mathbf{e}:\mathbf{e}\preceq\mathbf{e}_i,\mathbf{e}_j\preceq\mathbf{e},\mathbf{e}_j^+\not\preceq\mathbf{e}} \mu(\mathbf{e}, \mathbf{e}_i) \geq 0.$$

Using (5.4) for each $\mathbf{e}_i$ for the (sub-)poset $(\{\mathbf{e}_i\}^{\downarrow}, \preceq)$ (where $\mathbf{e}_i$ is the maximum) with its subspaces $\{\mathbf{e}\}^{\uparrow}, \{\mathbf{e}^{+}\}^{\uparrow}$, we have

$$\mathbf{P}(\mathbf{e}_j, \{\mathbf{e}_k\}^{\downarrow}) - \mathbf{P}(\mathbf{e}_j^{+}, \{\mathbf{e}_k\}^{\downarrow}) \geq 0$$

for any $\mathbf{e}_j, \mathbf{e}_k$.
The proof that Möbius-$^{\uparrow}$ monotonicity implies weak-$^{\uparrow}$ is similar.

□

In Fig. 1, Theorem 5.1 is summarized.

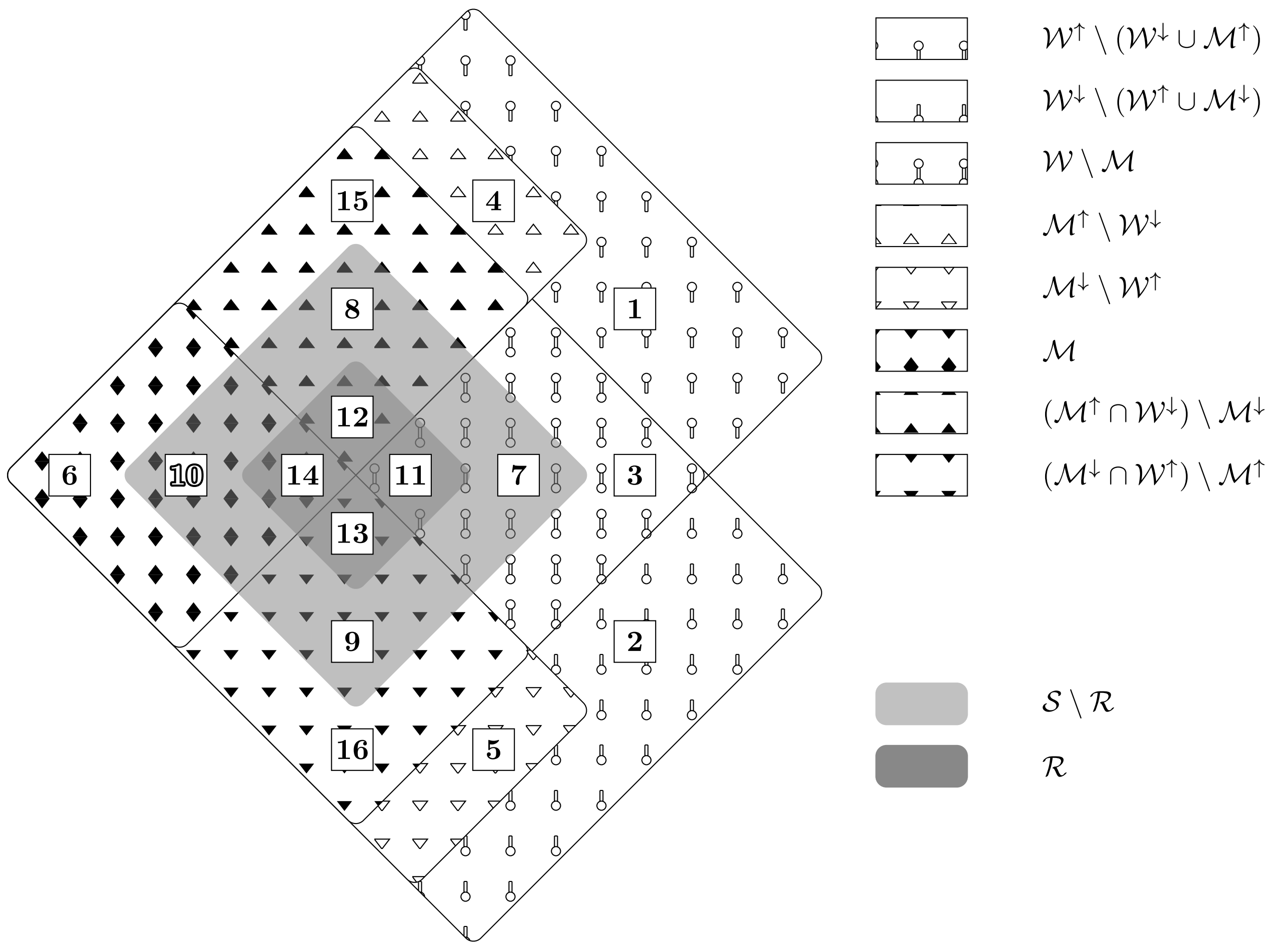


Figure 1: Relation between monotonicities. General partial ordering.

We know of no other implications involving the monotonicities we have considered. More precisely, the only one we do not know is whether Möbius-↑, Möbius-↓ and stochastic monotonicities imply realizable monotonicity (which is stated below as an open problem). The nonexistence of other implications is proven by presenting examples in Appendix A (the numbers in Fig. 1 correspond to the enumeration of these examples).

Open problem 1. *Does there exist a chain* $\mathbf{X}$*, state space* $\mathbb{E}$*, and partial ordering* $\preceq$ *such that* $\mathbf{X} \in \mathcal{M}^{\downarrow} \cap \mathcal{M}^{\uparrow} \cap \mathcal{S} \setminus \mathcal{R}$ *?*

Remark 5.1 (On Möbius monotonicity). For a total ordering (denote the states by $\mathbb{E} = \{1, \ldots, M\}$), the stochastic monotonicity of $\mathbf{X}$ can be written as

$$\forall (j, i_1 \le i_2) \quad \mathbf{P}_X(i_1, \{j\}^{\uparrow}) \le \mathbf{P}_X(i_2, \{j\}^{\uparrow}) \quad \equiv \quad \mathbf{P}_X(i_1, \{j\}^{\downarrow}) \ge \mathbf{P}_X(i_2, \{j\}^{\downarrow}).$$

In this ordering we can think of this monotonicity in two different (though equivalent) ways:

- *"Understanding 1"*. For any upset U and $\forall (i_1 \le i_2)$ we have $\mathbf{P}_X(i_1, U) \le \mathbf{P}_X(i_2, U)$ (or equivalently: for any downset D and $\forall (i_1 \le i_2)$ we have $\mathbf{P}_X(i_1, D) \ge \mathbf{P}_X(i_2, D)$).
- *"Understanding 2"*. For any upset U define $F_U(i) := \mathbf{P}_X(i, U)$. Then $\mathbf{X}$ is stochastically monotone if the function $F_U(i)$, treated as a function of i, must be "like" a distribution function, i.e., $\forall (i_1 \le i_2) F_U(i_1) \le F_U(i_2)$.

  Equivalently: for any downset D define $\bar{F}_D(i) = \mathbf{P}_X(i, D)$. Then $\mathbf{X}$ is stochastically monotone if the function $\bar{F}_D(i)$, treated as a function of i, must be "like" the tail of a distribution function, i.e., $\forall (i_1 \le i_2) \bar{F}_D(i_1) \ge \bar{F}_D(i_2)$.

Extending *"Understanding 1"* to a partial ordering $\preceq$ (we simply have different downsets and upsets, and each $i_1 \le i_2$ is replaced by $\mathbf{e} \preceq \mathbf{e}'$) leads to stochastic monotonicity as defined in Definition 3.1. Extending *"Understanding 2"* with $F_U(\cdot)$ being like a distribution function ($\bar{F}_D(\cdot)$ being like the tail of a distribution function) leads to Möbius-$^{\downarrow}$ (Möbius-$^{\uparrow}$) monotonicity, as defined in Definition 3.5.

5.1 Tree-ordering and total ordering

For a general partial ordering, we have, in Theorem 5.1, determined all the monotonicity relations. In this section we restrict our attention to some special cases: tree ordering and linear ordering.

Tree ordering. Let us start with a definition of this ordering.

Definition 5.1. A partial ordering $\preceq$ on $\mathbb{E}$ is called a *tree ordering* if there exists a maximum (which has no predecessor) and every other (non-maximal) state $\mathbf{e}$ has exactly one predecessor.

This definition affords a straightforward algorithm for inverting the matrix $\mathbf{C} = \mathbf{1}(\mathbf{e}_i \preceq \mathbf{e}_j)$. For a column corresponding to the state $\mathbf{e}$, it is enough to subtract the columns corresponding to the successors of $\mathbf{e}$. We obtain the matrix

$$\mu(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} 1 & \text{if} \quad \mathbf{e}_i = \mathbf{e}_j, \\ -1 & \text{if} \quad \mathbf{e}_i^{+} = \mathbf{e}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{5.7}$$

Theorem 5.2. *Let* $\mathbf{X}$ *be a Markov chain on* $\mathbb{E}$ *with a tree ordering* $\preceq$*. Then the following statements are equivalent.*

(i) $\mathbf{X} \in \mathcal{S}$

(ii) $\mathbf{X} \in \mathcal{R}$

(iii) $\mathbf{X} \in \mathcal{M}^{\downarrow}$

(iv) $\mathbf{X} \in \mathcal{W}^{\downarrow}$

Proof.

- $(i) \iff (ii)$

  Implication $(ii) \Rightarrow (i)$ follows from implication 1 of Theorem 5.1, whereas $(i) \Rightarrow (ii)$ follows from Theorem 4.3 in [FM01].

- $(iii) \iff (iv)$

  Implication $(iii) \Rightarrow (iv)$ follows from implication 5 of Theorem 5.1. To show $(iv) \Rightarrow (iii)$ we assume Möbius-$^{\downarrow}$ monotonicity, i.e., we have (the Möbius function given in Eq. (5.7))

$$\forall (\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}) \qquad 0 \leq \mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_j\}^{\downarrow}) - \mathbf{P}(\mathbf{e}_i^{+}, \{\mathbf{e}_j\}^{\downarrow}) = \sum_{\mathbf{e} \succeq \mathbf{e}_i} \mu(\mathbf{e}_i, \mathbf{e}) \mathbf{P}(\mathbf{e}, \{\mathbf{e}_j\}^{\downarrow}),$$

  which is exactly weak-$^{\downarrow}$ monotonicity.

- $(i) \iff (iv)$

  Implication $(i) \Rightarrow (iv)$ follows from implication 3 of Theorem 5.1. To show $(iv) \Rightarrow (i)$, note that any downset D can be written as a disjoint union of sets of the form $\{\mathbf{e}_k\}^{\downarrow}$, i.e., $D = \dot{\bigcup}_{k \in K} \{\mathbf{e}_k\}^{\downarrow}$ for some $K \subseteq \mathbb{E}$. For any $\mathbf{e}_k$, weak-$^{\downarrow}$ monotonicity implies that $\mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^{\downarrow}) - \mathbf{P}(\mathbf{e}_i^{+}, \{\mathbf{e}_k\}^{\downarrow}) \geq 0$, thus

$$\mathbf{P}(\mathbf{e}_i, D) - \mathbf{P}(\mathbf{e}_i^{+}, D) = \sum_{k \in K} \left(\mathbf{P}(\mathbf{e}_i, \{\mathbf{e}_k\}^{\downarrow}) - \mathbf{P}(\mathbf{e}_i^{+}, \{\mathbf{e}_k\}^{\downarrow}) \right) \geq 0,$$

  which implies stochastic monotonicity.

□

The monotonicity relations for tree-ordering are summarized in Fig. 2. The examples numbered 17, 18, 19 and 20 are given in Appendix A.

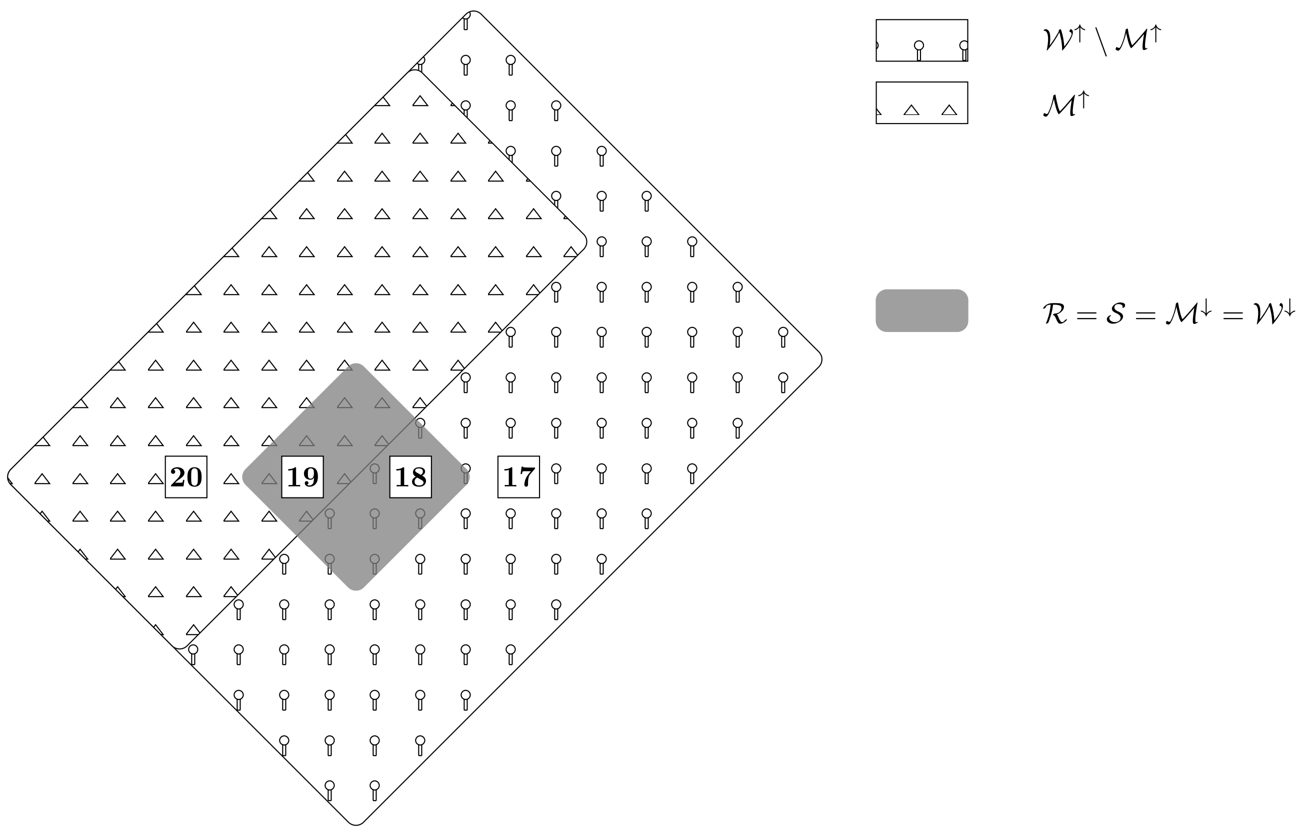


Figure 2: Relation between monotonicities. Tree-like ordering.

Total ordering. For this ordering let us denote the elements of state space $\mathbb{E}$ by $\{1, \ldots, M\}$. The Möbius function is following:

$$\mu(i,j) = \begin{cases} 1 & \text{if} \quad j = i, \\ -1 & \text{if} \quad j = i+1, \\ 0 & \text{otherwise}, \end{cases} \tag{5.8}$$

with ones on the diagonal and minus ones directly above it. For this ordering, we have the following lemma.

Lemma 5.1. *Let* $\mathbf{X}$ *be a Markov chain on* $\mathbb{E}$ *with total ordering* $\preceq := \leq$*. Then all the monotonicites* $\mathcal{S}, \mathcal{R}, \mathcal{M}^{\uparrow}, \mathcal{M}^{\downarrow}, \mathcal{W}^{\uparrow}, \mathcal{W}^{\downarrow}$ *are equivalent.*

Proof. By Theorem 5.2 it is enough to show that $\mathcal{W}^{\uparrow}$ is equivalent to $\mathcal{W}^{\downarrow}$ and that $\mathcal{W}^{\uparrow} \Rightarrow \mathcal{M}^{\uparrow}$. For a total ordering, all upsets are of the form $\{k\}^{\uparrow} = \{k, \ldots, M\}$ and all downsets are of the form $\{k\}^{\downarrow} = \{1, \ldots, k\}$. Hence (since the complement of a downset is an upset and vice versa), they are equivalent to each other and actually denote stochastic monotonicity.

Note that

$$\sum_{k \in \mathbb{E}} \mu(k,i) \mathbf{P}(k, \{j\}^{\uparrow}) = \mathbf{P}(i+1, \{j\}^{\uparrow}) - \mathbf{P}(i, \{j\}^{\uparrow}),$$

which means that $\mathcal{W}^{\uparrow}$ and $\mathcal{M}^{\uparrow}$ are equivalent.

□

6 Monotonicities and Siegmund duality

As mentioned, strong stationary duality was introduced in [DF90b]. However, a somewhat general recipe for such an SSD was only given in the case when the time reversal was stochastically monotone with respect to a total ordering. In Section 4.3 we recalled the theorem from [LS12a], that for a given ergodic chain $\mathbf{X}$ there exists a strong stationary dual chain $\mathbf{X}^*$ (with the link being a truncated stationary distribution) if and only if the time reversal of $\mathbf{X}$ is Möbius monotone. It turns out that there is a close connection between SSD and another duality. We say that the chain $\mathbf{Z}$ is a **Siegmund dual** of $\mathbf{X}$ if for any $n \geq 0, \mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}$ we have $Pr(X_n \preceq \mathbf{e}_j | X_0 = \mathbf{e}_i) = Pr(Z_n \succeq \mathbf{e}_i | Z_0 = \mathbf{e}_j)$. Siegmund [Sie76] showed that for a total ordering, such a dual exists if and only if $\mathbf{X}$ is stochastically monotone. Lorek [Lor18] gives an extension to partial orderings (the existence of the minimum and the maximum is required). The main result is that the Siegmund dual exists if and only if the chain is **Möbius-**$^{\downarrow}$ monotone. Moreover, in the latter article it is shown that the SSD from [LS12a] can be constructed in the following three steps: i) Calculate the time reversal of $\mathbf{X}$; ii) Calculate its Siegmund dual; iii) Calculate the appropriate Doob h-transform.

The results of this chapter are relevant for SSD and Siegmund duality. The general constructions of the SSD and the Siegmund dual were unknown for partial orderings. For Siegmund duality, for partially ordered state spaces, it was known that stochastic monotonicity is "not enough." Liggett, in [Lig04] a book on particle systems) writes (p. 87) "having a (reasonable) dual is a much more special property than being monotone, when the state space is not totally ordered." However, we can obtian an SSD or a Siegmund dual for a chain which is not stochastically monotone, such as is shown with the chain with the transition matrix $\mathbf{P}_6$ in Appendix A (the chain is not stochastically monotone, but is both Möbius-$^{\downarrow}$ and Möbius-$^{\uparrow}$ monotone).

A. Examples

The relations between monotonicities were given in Theorem 5.1 for a general partial ordering and in Theorem 5.2 for a tree ordering. They were summarized in Figures 1 and 2 respectively. In this section we prove (except for Open problem 1) that all the intersections in these figures are non-empty.

Given $\mathbf{P}$ and $\mathbf{C}$, checking all monotonicities except realizable monotonicity is straightforward (it only requires some matrix operations):

- Checking Möbius-$^{\downarrow}$ and Möbius-$^{\uparrow}$ monotonicity is straightforward from Definition 3.5.
- For weak monotonicities we need to precompute the offspring matrix $\mathbf{R}$. Let $\mathbb{O}(\mathbf{e}_i) = \{\mathbf{e} : \mathbf{e}_i \prec \mathbf{e} \text{ and } !\exists_{\mathbf{e}_j}(\mathbf{e}_i \prec \mathbf{e}_j \prec \mathbf{e})\}$ be the set of offspring of the state $\mathbf{e}_i$. The offspring matrix is defined as $\mathbf{R} = (\mathbf{R}^T_{\mathbf{e}_1}, \ldots, \mathbf{R}^T_{\mathbf{e}_M})^T$, where $\mathbf{R}_{\mathbf{e_i}}$ is the $|\mathbb{O}(\mathbf{e}_i)| \times |\mathbb{E}|$ matrix such that $\mathbf{R}_{\mathbf{e_i}}(\mathbf{e}_j, \mathbf{e}_i) = 1$ and $\mathbf{R}_{\mathbf{e_i}}(\mathbf{e}_j, \mathbf{e}_j) = -1$ for $\mathbf{e}_j \in \mathbb{O}(\mathbf{e}_i)$, all other entries being equal to zero. Note that $\mathbb{O}(\mathbf{e}_i)$ for all $\mathbf{e}_i \in \mathbb{E}$ and thus the matrix $\mathbf{R}$ is computed from $\mathbf{C}$. Weak-$^{\uparrow}$ monotonicity means that all entries of $\mathbf{RPC}^T$ are nonpositive, whereas weak-$^{\downarrow}$ means that all entries of $\mathbf{RPC}$ nonnegative.
- For stochastic monotonicity, we additionally need the matrix of all upsets (denoted by $\mathbf{S}$) instead of "just" the ordering matrix $\mathbf{C}$ ($\mathbf{S}$ is computed from $\mathbf{C}$). Stochastic monotonicity means that all entries of $\mathbf{RPS}$ are nonnegative.

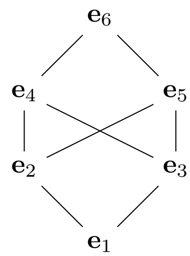
The functions checking all the above monotonicities are available in `The Julia Language` [LM17]. The functions require the transition matrix $\mathbf{P}$ and the ordering matrix $\mathbf{C}$. Also, the script checking the above monotonicities of all the examples that follow is available. The proofs concerning realizable monotonicites are given after introducing the examples. Recall that for a tree ordering, realizable monotonicity is equivalent to (among others) stochastic monotonicity, thus checking the latter is enough.

The pairing of the ordering matrices and the transition matrices:

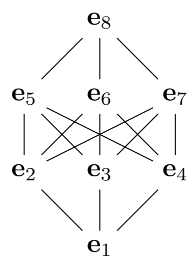
- $\mathbf{C}_1$ for processes $\mathbf{P}_3, \mathbf{P}_6, \mathbf{P}_7, \mathbf{P}_{11}, \mathbf{P}_{12}, \mathbf{P}_{13}, \mathbf{P}_{14}, \mathbf{P}_{15}, \mathbf{P}_{16}$,
- $\mathbf{C}_2$ for $\mathbf{P}_1, \mathbf{P}_2$,
- $\mathbf{C}_3$ for $\mathbf{P}_4, \mathbf{P}_5, \mathbf{P}_8, \mathbf{P}_9$ and
- $\mathbf{C}_4$ for $\mathbf{P}_{17}, \mathbf{P}_{18}, \mathbf{P}_{19}, \mathbf{P}_{20}$.

Order matrices and Hasse diagrams:

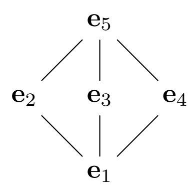
$$\mathbf{C}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



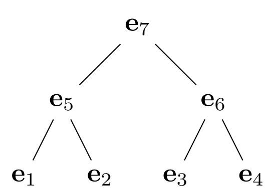
$$\mathbf{C}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



$$\mathbf{C}_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



$$\mathbf{C}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



The transition matrices

1. $\mathcal{W}^{\uparrow} \setminus (\mathcal{W}^{\downarrow} \cup \mathcal{M}^{\uparrow})$

$$\mathbf{P}_1 = \begin{pmatrix} 3/8 & 1/8 & 1/8 & 1/8 & 0 & 1/8 & 1/8 & 0 \\ 1/4 & 1/4 & 0 & 1/8 & 1/8 & 0 & 1/8 & 1/8 \\ 1/4 & 0 & 3/8 & 0 & 0 & 1/4 & 0 & 1/8 \\ 1/4 & 0 & 1/8 & 0 & 1/8 & 1/4 & 1/8 & 1/8 \\ 0 & 1/8 & 1/8 & 0 & 0 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 1/4 & 0 & 1/8 & 0 & 3/8 \\ 0 & 0 & 1/8 & 1/8 & 1/4 & 1/4 & 1/8 & 1/8 \\ 0 & 0 & 1/8 & 0 & 1/8 & 1/8 & 0 & 5/8 \end{pmatrix}$$

2. $\mathcal{W}^{\downarrow} \setminus (\mathcal{W}^{\uparrow} \cup \mathcal{M}^{\downarrow})$

$$\mathbf{P}_2 = \begin{pmatrix} 5/8 & 0 & 1/8 & 1/8 & 0 & 1/8 & 0 & 0 \\ 1/8 & 1/8 & 1/4 & 1/4 & 1/8 & 1/8 & 0 & 0 \\ 3/8 & 0 & 1/8 & 0 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 1/8 & 1/8 & 0 \\ 1/8 & 1/8 & 1/4 & 1/8 & 0 & 1/8 & 0 & 1/4 \\ 1/8 & 0 & 1/4 & 0 & 0 & 3/8 & 0 & 1/4 \\ 1/8 & 1/8 & 0 & 1/8 & 1/8 & 0 & 1/4 & 1/4 \\ 0 & 1/8 & 1/8 & 0 & 1/8 & 1/8 & 1/8 & 3/8 \end{pmatrix}$$

3. $\mathcal{W} \setminus (\mathcal{M} \cup \mathcal{S})$

$$\mathbf{P}_3 = \begin{pmatrix} 1/2 & 1/6 & 0 & 1/3 & 0 & 0 \\ 1/3 & 1/6 & 1/6 & 1/3 & 0 & 0 \\ 1/3 & 1/6 & 0 & 1/3 & 1/6 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 1/6 & 1/3 \\ 1/6 & 0 & 1/6 & 1/6 & 1/3 & 1/6 \\ 0 & 1/6 & 1/3 & 0 & 1/6 & 1/3 \end{pmatrix}$$

4. $\mathcal{M}^{\uparrow} \setminus \mathcal{W}^{\downarrow}$

$$\mathbf{P}_4 = \begin{pmatrix} 2/5 & 1/5 & 1/5 & 1/5 & 0 \\ 2/5 & 1/5 & 1/5 & 1/5 & 0 \\ 0 & 2/5 & 2/5 & 1/5 & 0 \\ 0 & 2/5 & 1/5 & 2/5 & 0 \\ 0 & 1/5 & 2/5 & 0 & 2/5 \end{pmatrix}$$

5. $\mathcal{M}^{\downarrow} \setminus \mathcal{W}^{\uparrow}$

$$\mathbf{P}_5 = \begin{pmatrix} 2/5 & 0 & 2/5 & 1/5 & 0 \\ 0 & 2/5 & 1/5 & 2/5 & 0 \\ 0 & 1/5 & 2/5 & 2/5 & 0 \\ 0 & 1/5 & 1/5 & 1/5 & 2/5 \\ 0 & 1/5 & 1/5 & 1/5 & 2/5 \end{pmatrix}$$

6. $\mathcal{M} \setminus \mathcal{S}$

$$\mathbf{P}_6 = \begin{pmatrix} 17/24 & 0 & 0 & 1/8 & 1/8 & 1/24 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/24 & 1/8 & 1/8 & 0 & 0 & 17/24 \end{pmatrix}$$

7. $\mathcal{S} \setminus (\mathcal{M}^{\uparrow} \cup \mathcal{M}^{\downarrow} \cup \mathcal{R})$

$$\mathbf{P}_7 = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/3 \\ 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/3 \end{pmatrix}$$

8. $\mathcal{S} \cap \mathcal{M}^{\uparrow} \setminus (\mathcal{M}^{\downarrow} \cup \mathcal{R})$

$$\mathbf{P}_8 = \begin{pmatrix} 2/5 & 1/5 & 1/5 & 1/5 & 0 \\ 2/5 & 1/5 & 1/5 & 1/5 & 0 \\ 2/5 & 0 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 2/5 & 1/5 & 0 \\ 0 & 2/5 & 1/5 & 0 & 2/5 \end{pmatrix}$$

9. $\mathcal{S} \cap \mathcal{M}^{\downarrow} \setminus (\mathcal{M}^{\uparrow} \cup \mathcal{R})$

$$\mathbf{P}_9 = \begin{pmatrix} 2/5 & 0 & 1/5 & 2/5 & 0 \\ 0 & 1/5 & 2/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 0 & 2/5 \\ 0 & 1/5 & 1/5 & 1/5 & 2/5 \\ 0 & 1/5 & 1/5 & 1/5 & 2/5 \end{pmatrix}$$

10. Aforementioned open problem.

11. $\mathcal{R} \setminus (\mathcal{M}^{\uparrow} \cup \mathcal{M}^{\downarrow})$

$$\mathbf{P}_{11} = \begin{pmatrix} 1/3 & 1/6 & 1/6 & 1/6 & 1/6 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/6 & 1/6 & 1/6 & 1/6 & 1/3 \end{pmatrix}$$

12. $\mathcal{R} \cap \mathcal{M}^{\uparrow} \setminus \mathcal{M}^{\downarrow}$

$$\mathbf{P}_{12} = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/3 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

13. $\mathcal{R} \cap \mathcal{M}^{\downarrow} \setminus \mathcal{M}^{\uparrow}$

$$\mathbf{P}_{13} = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/3 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

14. $\mathcal{R} \cap \mathcal{M}$

$$\mathbf{P}_{14} = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

15. $\mathcal{W}^{\downarrow} \cap \mathcal{M}^{\uparrow} \setminus (\mathcal{M}^{\downarrow} \cup \mathcal{S})$

$$\mathbf{P}_{15} = \begin{pmatrix} 17/24 & 0 & 0 & 1/8 & 1/8 & 1/24 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/24 & 1/16 & 1/16 & 1/16 & 1/16 & 17/24 \end{pmatrix}$$

16. $\mathcal{W}^{\uparrow} \cap \mathcal{M}^{\downarrow} \setminus (\mathcal{M}^{\uparrow} \cup \mathcal{S})$

$$\mathbf{P}_{16} = \begin{pmatrix} 17/24 & 1/16 & 1/16 & 1/16 & 1/16 & 1/24 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/8 & 5/16 & 5/16 & 1/12 & 1/12 & 1/12 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/12 & 1/12 & 1/12 & 5/16 & 5/16 & 1/8 \\ 1/24 & 1/8 & 1/8 & 0 & 0 & 17/24 \end{pmatrix}$$

Examples 17–20 deal with tree-ordering.

17. $\mathcal{W}^{\uparrow} \setminus \mathcal{S}$

$$\mathbf{P_{17}} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix}.$$

18. $\mathcal{S} \setminus \mathcal{M}^{\uparrow}$

$$\mathbf{P_{18}} = \begin{pmatrix} 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \end{pmatrix}.$$

19. $\mathcal{W}^{\uparrow} \cap \mathcal{S} \cap \mathcal{M}^{\uparrow}$

$$\mathbf{P_{19}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

20. $\mathcal{M}^{\uparrow} \setminus \mathcal{S}$

$$\mathbf{P_{20}} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

As already stated, checking all monotonicities except realizable monotonicity can be done automatically on a computer. Note that if the chain is not stochastically monotone, then it cannot be realizable monotone. That is why we only need to prove that

- The chains with the transition matrices $\mathbf{P}_{11}, \mathbf{P}_{12}, \mathbf{P}_{13}, \mathbf{P}_{14}$ are realizable monotone.
- The chains with the transition matrices $\mathbf{P}_7, \mathbf{P}_8, \mathbf{P}_9$ are not realizable monotone.

To prove realizable monotonicity it is enough to provide a monotone update rule, whereas showing that a given chain is not realizable monotone (i.e., that no monotone update function exists) is more challenging.

Monotone update rules for the chains with the transition matrices $\mathbf{P}_{11}, \mathbf{P}_{12}, \mathbf{P}_{13}$ and $\mathbf{P}_{14}$. For $\dot{\bigcup}_{i=1}^{6} A_i = [0,1]$ and $P(U \in A_i) = 1/6$ for $i = 1, \ldots, 6$ and $U \sim Unif[0,1]$ the following functions are monotone w.r.t. the partial ordering defined by $\mathbf{C}_1$.

- $\mathbf{P}_{11}$

$$\begin{aligned} \phi(\mathbf{e}_j, u) &= \mathbf{e}_i \quad \text{if } u \in A_i, \text{for } i = 1, \ldots, 6, j = 2, \ldots, 5, \\ \phi(\mathbf{e}_1, u) &= \begin{cases} \mathbf{e}_1 & \text{if} \quad u \in A_6, \\ \mathbf{e}_i & \text{if} \quad u \in A_i, i = 1, \ldots, 5, \end{cases} \\ \phi(\mathbf{e}_6, u) &= \begin{cases} \mathbf{e}_6 & \text{if} \quad u \in A_1, \\ \mathbf{e}_i & \text{if} \quad u \in A_i, i = 2, \ldots, 6. \end{cases} \end{aligned}$$

- $\mathbf{P}_{12}$

$$\begin{aligned} \phi(\mathbf{e}_j, u) &= \mathbf{e}_i \quad \text{if } u \in A_i, \text{for } i = 1, \ldots, 6, j = 1, \ldots, 5, \\ \phi(\mathbf{e}_6, u) &= \begin{cases} \mathbf{e}_6 & \text{if} \quad u \in A_1, \\ \mathbf{e}_i & \text{if} \quad u \in A_i, i = 2, \ldots, 6. \end{cases} \end{aligned}$$

- $\mathbf{P}_{13}$

$$\begin{aligned}\phi(\mathbf{e}_j,u) &= \mathbf{e}_i \quad \text{if } u\in A_i, \text{for } i=1,\ldots,6, j=2,\ldots,6,\\ \phi(\mathbf{e}_1,u) &= \begin{cases}\mathbf{e}_1 & \text{if} \quad u\in A_6,\\ \mathbf{e}_i & \text{if} \quad u\in A_i, i=1,\ldots,5.\end{cases}\end{aligned}$$

- $\mathbf{P}_{14}$

$$\phi(\mathbf{e}_j,u) = \mathbf{e}_i \quad \text{if } u\in A_i, \text{for } i=1,\ldots,6, j=1,\ldots,6.$$

Proofs that $\mathbf{P}_7, \mathbf{P}_8, \mathbf{P}_9$ are not realizable monotone w.r.t. the partial ordering defined by $\mathbf{C}_1$.

- The transition matrix $\mathbf{P}_7$.

  The idea of the proof is the following: we try to construct a monotone update function ϕ and deduce a contradiction. Start with defining an arbitrary update function at state $\mathbf{e}_1$:

$$\phi(\mathbf{e}_1,u)=\mathbf{e}_i \quad \text{if } u\in A_i, i=1,2,3,$$

  for $\dot\bigcup_{i=1}^3 A_i=[0,1]$, $P(U\in A_i)=1/3$, $i=1,2,3$ and $U\sim Unif[0,1]$. Since $\mathbf{e}_1\preceq\mathbf{e}_2$ we have the following requirements for $\phi(\mathbf{e}_2,\cdot)$: namely $\phi(\mathbf{e}_2,u)\succeq\mathbf{e}_i$ for $u\in A_i, i=1,2,3$. Thus the function is uniquely determined:

$$\phi(\mathbf{e}_2,u)=\begin{cases}\mathbf{e}_1 & \text{if} \quad u\in A_1,\\ \mathbf{e}_2 & \text{if} \quad u\in A_2,\\ \mathbf{e}_4 & \text{if} \quad u\in A_3.\end{cases}$$

  Similarly, since $\mathbf{e}_1\preceq\mathbf{e}_3$, we conclude that

$$\phi(\mathbf{e}_3,u)=\begin{cases}\mathbf{e}_1 & \text{if} \quad u\in A_1,\\ \mathbf{e}_3 & \text{if} \quad u\in A_3,\\ \mathbf{e}_4 & \text{if} \quad u\in A_2.\end{cases}$$

  Also, since $\mathbf{e}_2\preceq\mathbf{e}_4$, we conclude that

$$\phi(\mathbf{e}_4,u)=\begin{cases}\mathbf{e}_2 & \text{if} \quad u\in A_2,\\ \mathbf{e}_3 & \text{if} \quad u\in A_1,\\ \mathbf{e}_4 & \text{if} \quad u\in A_3.\end{cases}$$

  But this function is not monotone, since for $u\in A_2$ we have $\phi(\mathbf{e}_3,u)=\mathbf{e}_4\not\preceq\mathbf{e}_2=\phi(\mathbf{e}_4,u)$.

- The transition matrices $\mathbf{P}_8$ and $\mathbf{P}_9$.

  The idea of the proof is similar to the previous case. It will be done only for $\mathbf{P}_8$ (the proof for $\mathbf{P}_9$ is almost identical, since $\mathbf{P}_9(\mathbf{e}_i,\mathbf{e}_j)=\mathbf{P}_8(\mathbf{e}_{6-i},\mathbf{e}_{6-j}), i,j=1,\ldots,5$).

  We can start with defining an arbitrary update function at state $\mathbf{e}_1$:

$$\phi(\mathbf{e}_1,u)=\begin{cases}\mathbf{e}_1 & \text{if} \quad u\in A_0,\\ \mathbf{e}_i & \text{if} \quad u\in A_i, i=1,2,3,4\end{cases}$$

  for $\dot\bigcup_{i=0}^4 A_i=[0,1]$, $P(U\in A_i)=1/5$, $i=0,\ldots,4$ and $U\sim Unif[0,1]$. Since $\mathbf{e}_1\preceq\mathbf{e}_3$, we have the following requirements for $\phi(\mathbf{e}_3,\cdot)$: namely $\phi(\mathbf{e}_3,u)\succeq\phi(\mathbf{e}_1,u)$ for $u\in A_i, i=0,\ldots,4$. Thus the function is uniquely determined:

$$\phi(\mathbf{e}_3,u)=\begin{cases}\mathbf{e}_1 & \text{if} \quad u\in A_0\cup A_1,\\ \mathbf{e}_5 & \text{if} \quad u\in A_2,\\ \mathbf{e}_3 & \text{if} \quad u\in A_3,\\ \mathbf{e}_4 & \text{if} \quad u\in A_4.\end{cases}$$

Also, since $\mathbf{e}_3 \preceq \mathbf{e}_5$, we conclude that

$$\phi(\mathbf{e}_5, u) = \begin{cases} \mathbf{e}_2 & \text{if} \quad u \in A_0 \cup A_1, \\ \mathbf{e}_5 & \text{if} \quad u \in A_2 \cup A_4, \\ \mathbf{e}_3 & \text{if} \quad u \in A_3, \end{cases}$$

Since $\mathbf{e}_1 \preceq \mathbf{e}_4$, we conclude that there are two choices for $\phi(\mathbf{e}_4, u)$. We can have

$$\phi(\mathbf{e}_4, u) = \begin{cases} \mathbf{e}_1 & \text{if} \quad u \in A_0, \\ \mathbf{e}_3 & \text{if} \quad u \in A_1 \cup A_3, \\ \mathbf{e}_2 & \text{if} \quad u \in A_2, \\ \mathbf{e}_4 & \text{if} \quad u \in A_4, \end{cases}$$

but then for $u \in A_1$ we have $\phi(\mathbf{e}_4, u) = \mathbf{e}_3 \npreceq \mathbf{e}_2 = \phi(\mathbf{e}_5, u)$. We can also have

$$\phi(\mathbf{e}_4, u) = \begin{cases} \mathbf{e}_1 & \text{if} \quad u \in A_1, \\ \mathbf{e}_3 & \text{if} \quad u \in A_0 \cup A_3, \\ \mathbf{e}_2 & \text{if} \quad u \in A_2, \\ \mathbf{e}_4 & \text{if} \quad u \in A_4, \end{cases}$$

but then for $u \in A_0$ we have $\phi(\mathbf{e}_4, u) = \mathbf{e}_3 \npreceq \mathbf{e}_2 = \phi(\mathbf{e}_5, u)$. Thus, the function is not monotone.

Chapter 2. Absorption time and absorption probabilities for a family of multidimensional gambler models

1 Introduction

In the one-dimensional gambler's ruin problem two players start a game with the total amount of, say, N dollars and with initial values k and $N-k$. At each step they flip the coin (not necessarily unbiased) to decide who wins a dollar. The game is over when one of them goes bankrupt. There are some fundamental questions related to this process.

- Q1 Starting with i dollars, what is the probability of winning?
- Q2 Starting with i dollars, what is the distribution (or the structure) of the game duration (*i.e.,* the absorption time)? Or, what is the distribution (or the structure) of the game duration conditioned on winning/losing?

In this chapter we will answer above questions for a wide class of multidimensional generalizations of gambler's ruin problem. The proofs will be probabilistic in most cases, utilizing either Siegmund duality or intertwining between chains.

Generalized multidimensional gambler models In [Lor17] the following generalization was considered. There is one player (referred to "we") playing with $d \geq 1$ other players. Our initial assets are $(i_1, \ldots, i_d)$ and assets of consecutive players are $(N_1 - i_1, \ldots, N_d - i_d)$ ($N_j \geq 1$ is the total amount of assets with player j). Then, with probability $p_j(i_j)$ we win one dollar with player j and with probability $q_j(i_j)$ we lose it. With the remaining probability $1-\sum_{k=1}^{d}(p_k(i_j)+q_j(i_k))$ we do nothing (*i.e.,* ties are also possible). Once we win completely with player j (*i.e.,* $i_j = N_j$) we do not play with him/her anymore. We lose the whole game if we lose with at least one player, *i.e.,* when $i_j = 0$ for some $j = 1, \ldots, d$. The game can be described more formally as a Markov chain Z with two absorbing states. The state space is $\mathbb{E} = \{(i_1, \ldots, i_d) : 1 \leq i_j \leq N_j, 1 \leq j \leq d\} \cup \{-\infty\}$ (where $-\infty$ means we *lose*). For a convenience denote $p_j(N_j) = q_j(N_j) = 0, j = 1, \ldots, d$. Assume that for all $i_j \in \{1, \ldots, N_j\}, j \in \{1, \ldots, d\}$ we have $p_j(i_j) > 0, q_j(i_j) > 0$ and $\sum_{k=1}^{d}(p_k(i_k) + q_k(i_k)) \leq 1$. With some abuse of notation, we will sometimes write $(i'_1, \ldots, i'_d) = -\infty$. The transitions of the described chain are the following:

$$\mathbf{P}_Z((i_1,\ldots,i_d),(i'_1,\ldots,i'_d)) =$$

$$\begin{cases} p_j(i_j) & \text{if} \quad i'_j = i_j+1, i'_k = i_k, k \neq j, \\ q_j(i_j) & \text{if} \quad i'_j = i_j-1, i'_k = i_k, k \neq j, \\ \sum_{j:i_j=1} q_j(1) & \text{if} \quad (i'_1,\ldots,i'_d) = -\infty, \\ 1-\sum_{k=1}^d (p_k(i_k)+q_k(i_k)) & \text{if} \quad i'_j = i_j, 1 \leq j \leq d, \\ 1 & \text{if} \quad (i_1,\ldots,i_d) = (i'_1,\ldots,i'_d) = -\infty. \end{cases} \tag{1.1}$$

The chain has two absorbing states: $(N_1,\ldots,N_d)$ (we *win*) and $-\infty$ (we *lose*). Let

$$\rho((i_1,\ldots,i_d)) = P(\tau_{(N_1,\ldots,N_d)} < \tau_{-\infty} | Z_0 = (i_1,\ldots,i_d)), \tag{1.2}$$

where $\tau_{(i'_1,\ldots,i'_d)} := \inf\{n \geq 0 : Z_n = (i'_1,\ldots,i'_d)\}$. Roughly speaking, $\rho((i_1,\ldots,i_d))$ is the probability of winning starting at $(i_1,\ldots,i_d)$. In [Lor17] the formula for this probability was derived, namely

$$\rho((i_1,\ldots,i_d)) = \frac{\displaystyle\prod_{j=1}^d \left(\sum_{n_j=1}^{i_j} \prod_{r=1}^{n_j-1} \left(\frac{q_j(r)}{p_j(r)} \right) \right)}{\displaystyle\prod_{j=1}^d \left(\sum_{n_j=1}^{N_j} \prod_{r=1}^{n_j-1} \left(\frac{q_j(r)}{p_j(r)} \right) \right)}. \tag{1.3}$$

In this chapter we consider a much wider class of d-dimensional games - the chain given in (1.1) is just a special case. For example, within the class we can win/lose in one step with many players. The multidimensional chain is constructed from a variety of one-dimensional chains using Kronecker products. For this class:

- We give expressions for the winning probabilities and prove that it is a product of the winning probabilities corresponding to one-dimensional games. In particular, for a subclass of multidimensional chains, constructed from one-dimensional birth and death chains, the winning probabilities are given in (1.3). The main tool for showing winning probabilities is the Siegmund duality defined for partially ordered state spaces, exploiting the results from [Lor18].

- We give formulas for the distributions of the absorption time. In some cases a probability generating function is given, in other cases we show that the absorption time is equal, in distribution, to the absorption time of another chain, which is, in a sense, a multidimensional pure-birth chain. In many cases, the probabilistic proof is given. To show the absorption distribution, we exploit the spectral polynomials given in [Fil09b], and their variations considered in [GMZ12], [MZ17].

To have a feeling on what kind of results related to absorption time we obtain, let us have a look at Figure 1 (note that the caption can be fully understood once further sections are read).

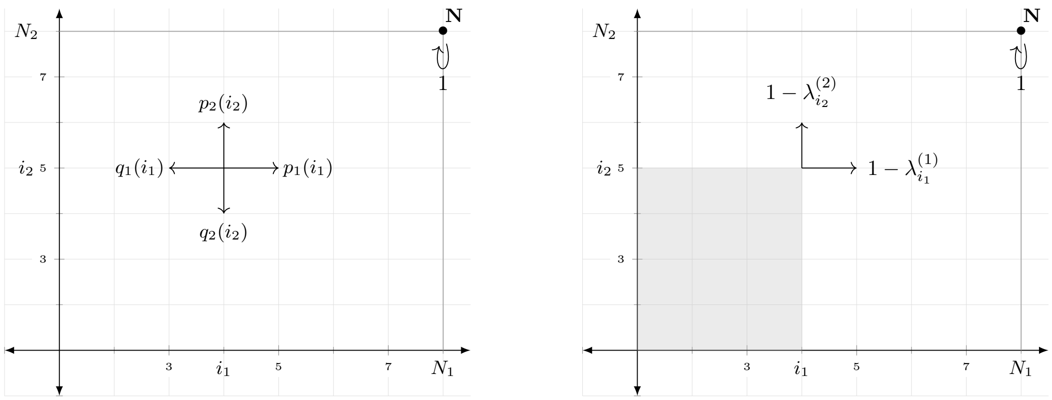


Figure 1: Sample transitions for the example from Section 6.3 with $d = 2$ and $r = 1$: X^* (left) and $\hat{X}$ (right). State $\mathbf{N} = (N_1, N_2)$ is the only absorbing one in both chains. Probabilities of staying are not depicted. If X^* starts at $(1,1)$, so does the $\hat{X}$ and $T^*_{(1,1),\mathbf{N}} = \hat{T}_{(1,1),\mathbf{N}}$ provided $q_j(1) = 0, j = 1,\ldots,d$. If, say, $\nu^*((i_1,i_2)) = 1$ then the pgf of $T^*_{(i_1,i_2),\mathbf{N}}$ is a mixture of pgfs of $\hat{T}_{(j_1,j_2),\mathbf{N}}$ for $j_1 \le i_1, j_2 \le i_2$ (shaded area).

On the left hand side of Fig. 1 a chain X^* constructed from two one-dimensional birth and death chains is presented (i-th chain has only one absorbing state $N_i, i = 1,2$). The chain is constructed in a specific way which results in the bivariate chain with independent moves (either *up, down, left* or *right*). Its transitions are consistent with transitions of a chain given in (1.1) – except there is just one absorbing state $\mathbf{N} = (N_1, N_2)$ (*i.e.,* there is no $-\infty$ state). We will show that the time to absorption of the chain X^* started at (i_1, i_2) is a mixture of times to absorption of a pure-birth chain $\hat{X}$ starting at states (i'_1, i'_2), where $i'_1 \le i_1, i'_2 \le i_2$ (shaded area on the right hand side of Fig. 1). In particular, if X^* starts in (1,1), so does the chain $\hat{X}$. The chain $\hat{X}$ has also the only absorbing state $\mathbf{N}$, it is pure-birth in the sense, that only *up* and *right* transitions are allowed. The probabilities of its transitions are related to the eigenvalues of one-dimensional birth and death chains from which X^* was constructed.

Remark 1.1. In [Lor17] the chain given in (1.1) was considered. The chain is constructed from d one-dimensional birth and death chains in a very specific way. The method from this chapter is much more general, we can construct a variety of multidimensional chains from given d one-dimensional birth and death chains. It is worth mentioning, that even for the case (1.1), the proof is quite different (from the one in [Lor17]).

Several variations (including multidimensional ones) of gambler's ruin problem have been considered. Researchers usually study absorption probabilities, absorption time, or both. In [KP02] authors consider a two-dimensional model (they consider two currencies) and study the expected game duration. In [Ros09] some multidimensional game is considered: at each step two players are randomly chosen, these players play a regular game, all till one of the players have all the coins. Author derives the probability that a specific player wins, the expected number of turns in total and between two given players. In [RS04] the following multidimensional game is considered: there are n players, at each step there is one winner who collects $n-1$ coins from other players, whereas all others lose 1 coin. An asymptotic probability for an individual ruin and dependence of ruin time are studied. In [Tzi19] the multidimensional case is considered, in which with equal probability a unit displacement in any direction is possible. Moments of leaving some ball are considered. In [CSV18] authors present a new probabilistic analysis of distributed algorithm re-considering a variation of a banker algorithm. Mathematically, it is random walk on a rectangle with specified absorbing states. The results are generalized to the case with many players and resources.

The absorption probability of a given chain may be related to the stationary distribution of some ergodic chain. This relation is given using the *Siegmund duality*, the notion introduced in [Sie76]. This is also the tool we use for showing winning probabilities. Already in [Lin52] similar duality between some random walks on integers was shown. It was also studied in financial mathematics, where the probability that a dual risk process starting at some level is ruined, is equal to the probability that the stationary queue length exceeds this level (see [AA10], [AS09]). In all these cases the Siegmund duality was defined for the linear ordering of the state space. The existence of a Siegmund dual for a linearly ordered state space requires stochastic monotonicity of the chain. In [Lor18] *if and only if* conditions for the existence of the Siegmund dual for partially ordered state spaces was derived (roughly speaking, the Möbius monotonicity is required). In this chapter, we exploit this duality defined for a coordinate-wise partial ordering.

It is worth mentioning that for one-dimensional gambler models there are several approaches to (each having its advantages and disadvantages) study the winning probability and/or game duration, including conditioning, difference equations (the most common approach to provide the formula for the winning probability in the classical – *i.e.*, the one with constant birth and death rates – gambler's ruin problem), generating functions and martingale-based methods (*e.g.*, [Len09a]), path counting (*e.g.*, [Len09b]).

Absorption time Consider a one-dimensional game corresponding to the gambler's ruin problem. Let N be the total amount of money. Being at a state $i \in \{2, \ldots, N-1\}$ we can either win one dollar with probability $p(i) > 0$ or lose it with probability $q(i) > 0$, with the remaining probability nothing happens. Assuming $p(1) > 0$ and $p(N) = q(N) = p(0) = q(0) = 0$ the transitions are following:

$$\mathbf{P}_Y(i, i') = \begin{cases} p(i) & \text{if} \quad i' = i+1, \\ q(i) & \text{if} \quad i' = i-1, \\ 1-(p(i)+q(i)) & \text{if} \quad i' = i. \end{cases} \tag{1.4}$$

States 0 and N are absorbing. Consider two cases:

Case: $q(1) = 0$ Roughly speaking, if started at $i \geq 1$ the chain never reaches 0 and this is actually a birth and death chain on $\{1, \ldots, N\}$ with N being the only absorbing state. Define $T_{a,b} = \inf\{n \geq 0 : Y_n = b \,|Y_0 = a\}$. A well known theorem attributed to [Kei79] states that the probability generating function pgf of $T_{1,N}$ is the following:

$$\mathsf{pgf}_{T_{1,N}}(u) := \mathbf{E}u^{T_{1,N}} = \prod_{k=1}^{N-1} \left[\frac{(1-\lambda_k)u}{1-\lambda_k u} \right], \tag{1.5}$$

where $-1 \leq \lambda_k < 1, k = 1, \ldots, N-1$ are $N-1$ non-unit eigenvalues of $\mathbf{P}_Y$. The proof was purely analytical. Note that (1.5) corresponds to the sum of N geometric random variables, provided that all eigenvalues are positive (which, in this case, is equivalent to the stochastic monotonicity of the chain). For this case, [Fil09b] gave a probabilistic proof of (1.5) using strong stationary duality and intertwinings between chains. Note that in this case (1.5) can be rephrased as:

Theorem 1.1. *Let X^* be an absorbing chain on $\mathbb{E} = \{1, \ldots, N\}$ starting at 1 with the transition matrix $\mathbf{P}_{X^*}$ given in (1.4) having positive eigenvalues $\lambda_k > 0, k = 1, \ldots, N$. Then $T^*_{1,N}$ has the* *same distribution as* $\hat{T}_{1,N}$*, the absorption time of* $\hat{X}$ *on* $\hat{\mathbb{E}} = \mathbb{E}$ *starting at* 1 *with the transition matrix*

$$\mathbf{P}_{\hat{X}}(i,i') = \begin{cases} 1-\lambda_i & \text{if } i' = i+1, \\ \lambda_i & \text{if } i' = i, \\ 0 & \text{otherwise.} \end{cases}$$

The chain Y on $\{1,\ldots,N\}$ is called pure-birth if $\mathbf{P}_Y(i,j) = 0$ for $j < i$. Similarly, a multidimensional chain Y on $\mathbb{E} = \{(i_1,\ldots,i_d) : 1 \le i_j \le N_j, 1 \le j \le d\}$ is said to be **pure-birth** if the probability of decreasing any set of coordinates at one step is 0.

Simply noting that for any $1 < s < N$ we have $T_{1,N} = T_{1,s} + T_{s,N}$ and that $T_{1,s}$ and $T_{s,N}$ are independent (see Cor. 2.1 [GMZ12] for a continuous time version) we have

$$\mathsf{pgf}_{T_{s,N}}(u) := \mathbf{E}u^{T_{s,N}} = \frac{\prod_{k=1}^{N-1}\left[\frac{(1-\lambda_k)u}{1-\lambda_k u}\right]}{\prod_{k=1}^{s-1}\left[\frac{(1-\lambda_k^{\lfloor s \rfloor})u}{1-\lambda_k^{\lfloor s \rfloor}u}\right]}, \tag{1.6}$$

where $\lambda_k^{\lfloor i \rfloor}$ are the eigenvalues of the substochastic $(s-1)\times(s-1)$ matrix

$$\mathbf{P}_Y^{\lfloor s \rfloor}(i,i') = \begin{cases} p(i) & \text{if} \quad i' = i+1, 1 \le i \le s-2, \\ q(i) & \text{if} \quad i' = i-1, 2 \le i \le s-1, \\ 1-(p(i)+q(i)) & \text{if} \quad i' = i, 1 \le i \le s-1. \end{cases}$$

Case: $q(1) > 0$ In this case, authors in [GMZ12] (different proof is given in [MZ17]) derived formulas for pgf of $T_{s,N}$ and $T_{s,0}$ (more precisely, they derived formulas for continuous time versions), which, in discrete case, are given by

$$\mathsf{pgf}_{T_{s,N}}(u) = \mathbf{E}u^{T_{s,N}} = \rho(s)\frac{\prod_{k=1}^{N-1}\left[\frac{(1-\lambda_k)u}{1-\lambda_k u}\right]}{\prod_{k=1}^{s-1}\left[\frac{(1-\lambda_k^{\lfloor s \rfloor})u}{1-\lambda_k^{\lfloor s \rfloor}u}\right]}, \tag{1.7}$$

$$\mathsf{pgf}_{T_{s,0}}(u) = \mathbf{E}u^{T_{s,0}} = (1-\rho(s))\frac{\prod_{k=1}^{N-1}\left[\frac{(1-\lambda_k)u}{1-\lambda_k u}\right]}{\prod_{k=1}^{N-s-1}\left[\frac{(1-\lambda_k^{\lceil s \rceil})u}{1-\lambda_k^{\lceil s \rceil}u}\right]},$$

where $\rho(s)$ is the probability of winning (*i.e.*, (1.2) with $d = 1, i_1 = s$) and $\lambda_k^{\lceil s \rceil}$ are the eigenvalues of the substochastic matrix (of the size $N-s-1$)

$$\mathbf{P}_Y^{\lceil s \rceil}(i,i') = \begin{cases} p(i) & \text{if} \quad i' = i+1, s+1 \le i \le N-2, \\ q(i) & \text{if} \quad i' = i-1, s+2 \le i \le N-1 \\ 1-(p(i)+q(i)) & \text{if} \quad i' = i, s+1 \le i \le N-1. \end{cases}$$

In this chapter we aim at presenting results similar to Theorem 1.1 and to (1.7) for a wide class of multidimensional extensions of gambler's ruin problem.

2 Kronecker product and main results

To state our main results we need to recall a notion of the Kronecker product. Let $\mathbf{A}$ be a matrix of size $n \times m$. Then, for any matrix $\mathbf{B}$ the Kronecker product of the matrices is defined as follows:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2m}\mathbf{B} \\ & \dots\dots\dots\dots & & \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \dots & a_{nm}\mathbf{B} \end{bmatrix}.$$

For square matrices $\mathbf{A}$ and $\mathbf{B}$ it is also convenient to define the Kronecker sum as:

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{A} \otimes \mathbf{I_B} + \mathbf{I_A} \otimes \mathbf{B},$$

where $\mathbf{I_A}$ ($\mathbf{I_B}$) is the identity matrix of the same size as $\mathbf{A}$ ($\mathbf{B}$).

Both, product and sum, are extended as:

$$\bigotimes_{i=1}^{n} \mathbf{A}_i = (\dots((\mathbf{A}_1 \otimes \mathbf{A}_2) \otimes \mathbf{A}_3)\dots) \otimes \mathbf{A}_n = \mathbf{A}_1 \otimes \mathbf{A}_2 \otimes \dots \otimes \mathbf{A}_n$$

and

$$\bigoplus_{i=1}^{n} \mathbf{A}_i = (\dots((\mathbf{A}_1 \oplus \mathbf{A}_2) \oplus \mathbf{A}_3)\dots) \oplus \mathbf{A}_n = \mathbf{A}_1 \oplus \mathbf{A}_2 \oplus \dots \oplus \mathbf{A}_n.$$

Notation For a convenience, for the given substochastic matrix $\mathbf{P}'_Y$ on $\mathbb{E}' = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ by $\mathbf{P}_Y = \mathcal{F}_{\mathbf{e}_0}(\mathbf{P}'_Y)$ we denote a stochastic matrix on $\mathbb{E} = \{\mathbf{e}_0\} \cup \mathbb{E}'$ constructed from $\mathbf{P}'_Y$ in the following way:

$$\mathbf{P}_Y(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} \mathbf{P}'_Y(\mathbf{e}_i, \mathbf{e}_j) & \text{if} \quad \mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}, \\ 1 - \sum_{\mathbf{e}_k \in \mathbb{E}'} \mathbf{P}'_Y(\mathbf{e}_i, \mathbf{e}_k) & \text{if} \quad \mathbf{e}_i \in \mathbb{E}', \mathbf{e}_j = \mathbf{e}_0, \\ 1 & \text{if} \quad \mathbf{e}_i = \mathbf{e}_j = \mathbf{e}_0. \\ 0 & \text{if} \quad \mathbf{e}_i = \mathbf{e}_0, \mathbf{e}_j \in \mathbb{E}. \end{cases}$$

Similarly, for a stochastic matrix $\mathbf{P}_Y$ on $\mathbb{E} = \{\mathbf{e}_0\} \cup \mathbb{E}'$ let $\mathbf{P}'_Y = \mathcal{F}^{-1}_{\mathbf{e}_0}(\mathbf{P}_Y)$ be a substochastic matrix on $\mathbb{E}'$ resulting from $\mathbf{P}_Y$ by removing the row and the column corresponding to the state $\mathbf{e}_0$.

For a Markov chain Y on $\mathbb{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ we say that $A \subseteq \mathbb{E}$ is a communication class if for all $\mathbf{e}, \mathbf{e}' \in A$ we have $\mathbf{P}^n_Y(\mathbf{e}, \mathbf{e}') > 0$ for some $n \geq 0$.

For a given chain Y we define $T_{\nu,\mathbf{e}'} := \inf\{n \geq 0 : Y_n = \mathbf{e}' | Y_0 \sim \nu\}$. Slightly abusing the notation, by $T_{\mathbf{e},\mathbf{e}'}$ we mean $T_{\nu,\mathbf{e}'}$ with $\nu = \delta_{\mathbf{e}}$. For $\mathbb{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ and for $f : \mathbb{E} \to \mathbb{R}$, we define a row vector $\boldsymbol{f} = (f(\mathbf{e}_1), \dots, f(\mathbf{e}_M))$. For $N_j > 0, j = 1, \dots, d$ we define $\mathbf{N} = (N_1, \dots, N_d)$.

2.1 Absorption probabilities

Before stating the result, let us provide some intuition behind it. Assume we play with d players (*i.e.,* we have d one-dimensional games), the winning probability playing only with player j is given by $\rho_j(i_j)$, provided we started with i_j dollars. Now, if we play with all the players independently and we define that we win *the whole game* if we win with all the players, lose *the whole game* if we lose with at least one player, then the probability of winning *the whole game* is a product of probabilities *i.e.,* $\prod_{j=1}^{d}\rho_j(i_j)$ (the formula is given in (2.3) below).

However, one can ask the following question: Can we combine the games in *some other* way, so that the resulting winning probability is still of a product form? For example: can we combine $d = 10$ games so that at one step we can play with at most $r = 5$ other players? Can the rules (for combining the games) depend on the current fortune? These type of questions (to which the answers are *yes*) were the motivation for the next theorem, where a wide class of possible combinations is allowed. The examples are provided later in Section 6.

Theorem 2.1. *Fix integers $d \geq 1, m \geq 1$. For $k = 1,\ldots,m$ let $\mathbf{A}_k \subseteq \{1,\ldots,d\}$. Assume*

- $\forall(1 \leq k \leq m)$ $\mathbf{P}_{Z_j^{(k)}} = \mathcal{F}_0(\mathbf{P'}_{Z_j^{(k)}})$ *is a stochastic matrix corresponding to a Markov chain $Z_j^{(k)}$ on $\mathbb{E}_j = \{0,1,\ldots,N_j\}$ such that for $i \in \mathbb{E}_j$ we have*

$$\rho_j^{(k)}(i) = P(\tau_{N_j} < \tau_0 | Z_j^{(k)}(0) = i) = \rho_j(i). \tag{2.1}$$

  In other words, $Z_j^{(k)}$ are m ($k = 1,\ldots,m$) chains having the same winning probability at every state i.

- *Let*

$$\mathbb{R}'_{Z_j^{(k)}} = \begin{cases} \mathbf{P'}_{Z_j^{(k)}} & \text{if } j \in \mathbf{A}_k, \\ \mathbf{I}_j & \text{if } j \notin \mathbf{A}_k, \end{cases}$$

  where $\mathbf{I}_j$ is the identity matrix of size $N_j \times N_j$.

- *Let $\mathbf{B}_i, i = 1,\ldots,m$ be either*
  - *any real numbers (i.e., $\mathbf{B}_k \in \mathbb{R}$) such that $\sum_{k=1}^{m}\mathbf{B}_k = 1$, **or***
  - *square matrices of size $\prod_{j=1}^{d} N_j \times \prod_{j=1}^{d} N_j$ such that $\sum_{k=1}^{m}\mathbf{B}_k = \mathbf{I}$ (identity matrix of the appropriate size)*

- *The matrix $\mathbf{P}_Z = \mathcal{F}_{-\infty}(\mathbf{P}_{Z'})$ with*

$$\mathbf{P'}_Z = \sum_{k=1}^{m}\mathbf{B}_k\left(\bigotimes_{j\leq d}\mathbb{R}'_{Z_j^{(k)}}\right) \tag{2.2}$$

  is stochastic on $\mathbb{E} = \{-\infty\} \cup \bigotimes_{j\leq d}\mathbb{E}'_j$, set $\mathbb{E} \setminus \{\{\mathbf{N}\} \cup \{-\infty\}\}$ is a communication class.

Then, the winning probability (i.e., the absorption at $\mathbf{N}$) of the Markov chain Z on $\mathbb{E} = \{-\infty\} \cup \{1,\ldots,N_1\} \times \ldots \times \{1,\ldots,N_d\}$ with the transition matrix $\mathbf{P}_Z = \mathcal{F}_{-\infty}(\mathbf{P'}_Z)$ is given by

$$\rho(i_1,\ldots,i_d) = \prod_{j=1}^{d}\rho_j(i_j). \tag{2.3}$$

The proof is postponed to Section 4.2.

Note that $\mathbf{P}_{Z_j^{(k)}}$ in Theorem 2.1 are general. If we only know the winning probabilities of $\mathbf{P}_{Z_j^{(k)}}$ (they cannot depend on k), then we know the winning probabilities of Z. Taking $\mathbf{P}_{Z_j^{(k)}}$ corresponding to gambler's ruin game given in (1.4) we have:

Corollary 2.2. Let $\mathbf{P}_{Z_j^{(k)}}$ for $j = 1, \ldots, d$ be the birth and death chain given in (1.4). Then, the winning probability of $\mathbf{P}_Z = \mathcal{F}_{-\infty}(\mathbf{P}'_Z)$ is given by (1.3).

Proof. For $\mathbf{P}_{Z_j^{(k)}}$ the winning probability is known (shown in (3.5)), it is

$$\rho_j(i_j) = \frac{\sum_{n_j=1}^{i_j} \prod_{r=1}^{n_j-1} \left(\frac{q_j(r)}{p_j(r)}\right)}{\sum_{n_j=1}^{N_j} \prod_{r=1}^{n_j-1} \left(\frac{q_j(r)}{p_j(r)}\right)}. \tag{2.4}$$

Assertion of Theorem 2.1 completes the proof. □

The chain Z can be interpreted as a d-dimensional game, with state $(N_1, \ldots, N_d)$ corresponding to winning and state $-\infty$ corresponding to losing.

Remark 2.1. In [Lor17, Theorem 2] we showed that the non-negativity of the resulting $\mathbf{P}'_Z$ is not required (for showing a product form formula for the winning probability of the model considered therein) – it is only required that (in our settings) for all $(i'_1, \ldots, i'_d) \in \mathbb{E}$ we have

$$\lim_{n\to\infty} \mathbf{P'}^{n}_{Z}((i'_1, \ldots, i'_d), (i_1, \ldots, i_d)) = \pi((i_1, \ldots, i_d)), \qquad \sum_{(i_1,\ldots,i_d)\in\mathbb{E}} \pi((i_1, \ldots, i_d)) = 1.$$

A one-dimensional example was provided in [Lor17, Section 4]. It is left for a future research to check if the assertion of Theorem 2.1 holds also without the assumption of the non-negativity of $\mathbf{P}'_Z$ (in this chapter we focused on stochastic proofs, whenever possible).

2.2 Absorption time

Let us start with some motivation. As recalled in the introduction, we have some expressions for the absorption time of a one-dimensional birth and death chain X^* on $1, \ldots, N$ with one absorbing state N. If we start at state 1, this time is expressed in terms of the eigenvalues of the transition matrix (formula (1.5)). Moreover, if the eigenvalues are non-negative (which corresponds to stochastic monotonicity of the chain), we have a stochastic interpretation: its absorption time is equal to the absorption time of a *pure-birth* chain $\hat{X}$, whose transitions involve the aforementioned eigenvalues (formula (1.1)).

In case when we start at $s > 1$, this absorption time of X^* can be expressed in terms of the eigenvalues of the transition matrix and of the truncated substochastic transition matrix (formula (1.6)). Using the duality-based approach given in [Fil09b], it is relatively easy to show that it can be expressed as a mixture of absorption times of a *pure-birth* chain starting at $s' \leq s$. To be more precise, the probability generating function of the absorption time of X^* is a mixture of probability generating functions of the absorption time of $\hat{X}$. Moreover, if the mixture coefficients are non-negative, we have stochastic interpretation (a sample-path construction) of this absorption time.

These observations were our motivation for a multidimensional extension. Do similar results hold then? Can we have a similar stochastic interpretation in some cases? How to construct a multidimensional chain out of many one-dimensional birth and death chains, so that the absorption time of the constructed chain can be somehow expressed in terms of pure-birth chains, whose transitions involve eigenvalues of underlying birth and death chains? In the next theorem we provide a wide class of multidimensional chains (ways of constructing such a chain from one-dimensional birth and death chains), for which we are able to express the absorption time in the aforementioned desired way.

Theorem 2.3. *Fix integers $d \geq 1, m \geq 1$. For $k = 1, \ldots, m$ let $A_k \subseteq \{1, \ldots, d\}$. Let $b_i \in \mathbb{R}, i = 1, \ldots, m$ such that $\sum_{k=1}^{m} b_i = 1$. Let, for $1 \leq j \leq d$, $\mathbf{P}_{X_j^*}$ be the stochastic matrix corresponding to a birth and death chain X_j^* on $\mathbb{E}_j = \{0, \ldots, N_j\}$ with transitions given in (1.4) with birth rates $p_j(i)$ and death rates $q_j(i)$. Let, for $1 \leq j \leq d$, $\mathbf{P}'_{X_j^*} = \mathcal{F}_0^{-1}(\mathbf{P}_{X_j^*})$ be the substochastic matrix on $\mathbb{E}'_j = \{1, \ldots, N_j\}$ and*

$$\mathbb{R}'_{X_j^{*(k)}} = \begin{cases} \mathbf{P}'_{X_j^*} & \text{if } j \in A_k, \\ \mathbf{I}_j & \text{if } j \notin A_k, \end{cases}$$

where $\mathbf{I}_j$ is the identity matrix of size $N_j \times N_j$. I.e., $\mathbb{R}'_{X_j^{*(k)}}$ is either matrix $\mathbf{P}'_{X_j^*}$ or an identity matrix. Let $\lambda_1^{(j)} \leq \ldots \leq \lambda_{N_j-1}^{(j)} < \lambda_{N_j}^{(j)} = 1$ be the eigenvalues of $\mathbf{P}'_{X_j^*}$.

Assume

- ***A1*** *The chains $\mathbf{P}_{X_j^*}, j = 1, \ldots, d$ are stochastically monotone.*
- ***A2*** *The matrix $\mathbf{P}_{X^*} = \mathcal{F}_{-\infty}(\mathbf{P}'_{X^*})$ with*

$$\mathbf{P}'_{X^*} = \sum_{k=1}^{m} b_k \left(\bigotimes_{j \leq d} \mathbb{R}'_{X_j^{*(k)}} \right) \tag{2.5}$$

  is a stochastic matrix on $\mathbb{E} = \{-\infty\} \cup \bigotimes_{j \leq d} \mathbb{E}'_j$, set $\mathbb{E} \setminus \{\{\mathbf{N}\} \cup \{-\infty\}\}$ is a communication class, $\mathbf{N} = (N_1, \ldots, N_d)$.

- ***A3*** *The matrix $\mathbf{P}_{\hat{X}}$, given below in (2.6), is non-negative.*

Let X^* be a chain with the above transition matrix $\mathbf{P}_{X^*}$. Assume its initial distribution is ν^*. The state $\mathbf{N}$ is the absorbing state, denote its absorption time by $T^*_{\nu^*,\mathbf{N}}$.
Then the time to absorption $T^*_{\nu^*,\mathbf{N}}$ has the following pgf

$$\mathsf{pgf}_{T^*_{\nu^*,\mathbf{N}}}(s) = \sum_{\hat{\mathbf{e}} \in \mathbb{E}} \hat{\nu}(\hat{\mathbf{e}}) \mathsf{pgf}_{\hat{T}_{\hat{\mathbf{e}},\mathbf{N}}}(s) \left(\prod_{j=1}^{d} \rho_j(1) \right),$$

where $\rho_j(1)$ is the winning probability of X_j^* starting at 1,

$$\hat{\nu} = \nu^* \bigotimes_{j \leq d} \Lambda_j^{-1},$$

Λ_j are given in (3.8) calculated for $\mathbf{P}'_{X_j^*}$ and $\hat{T}_{\hat{\mathbf{e}},\mathbf{N}}$ is the time to absorption for the chain $\hat{X} \sim (\delta_{\hat{\mathbf{e}}}, \mathbf{P}_{\hat{X}})$ with:

$$\mathbf{P}_{\hat{X}}((i_1,\ldots,i_d),(i'_1,\ldots,i'_d)) = \begin{cases} \prod_{j\in\mathbf{B}}\left(1-\lambda^{(j)}_{i_j}\right)\sum_{k:\mathbf{B}\subseteq\mathbf{A}_k}\left(b_k\prod_{j\in A_k\setminus B}\lambda^{(j)}_{i_j}\right) & \text{if } i'_j = i_j+1, \\ & \quad j\in\mathbf{B}\subseteq\{1,\ldots,d\}, \mathbf{B}\neq\emptyset, \\ \sum_{k=1}^{m} b_k \prod_{j\in\mathbf{A}_k}\lambda^{(j)}_{i_j} & \text{if } i'_j = i_j \text{for} j=1,\ldots,d, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

We also have

$$\forall(\mathbf{e}\in\mathbb{E})\nu^*(\mathbf{e})\neq 0 \;\Rightarrow\; \exists(\mathbf{e}'\succeq\mathbf{e})\hat{\nu}(\mathbf{e}')>0 \tag{2.7}$$

Moreover, the eigenvalues of $\mathbf{P}_{X^*}$ *and* $\mathbf{P}_{\hat{X}}$ *are the diagonal entries of* $\mathbf{P}_{\hat{X}}$*.*

Note that $\hat{X}$ is a pure-birth chain. Moreover, at one step it can increase values of coordinates by $+1$ on a set $\mathbf{B}$ such that $\mathbf{B}\subseteq\mathbf{A}_k$, for $k=1,\ldots,m$.

Remark 2.2. In case $b_i\geq 0, i=1,\ldots,m$ (*i.e.,* $(b_1,\ldots,b_m)$ is a distribution on $\{1,\ldots,m\}$), the matrix $\mathbf{P}_{X^*}$ in assumption **A2** is stochastic (thus **A2** is only about $\mathbb{E}\setminus\{\{\mathbf{N}\}\cup\{-\infty\}\}$ being a communication class) and so is the matrix $\mathbf{P}_{\hat{X}}$ given in (2.6) (*i.e.,* **A3** is fulfilled).

Remark 2.3. Note that the formula for the transitions of the resulting multidimensional chain given in (2.5) is very similar to the formula (2.2), *i.e.,* the one used in Theorem 2.1 (providing results for the winning probabilities). The main difference is that in (2.2) we have some $\mathbf{B}_k$'s which can be either numbers summing up to 1 or some matrices summing up to the identity matrix, whereas in (2.5) they must be numbers summing up to 1. Consequently, the class of the resulting multidimensional chains constructed in Theorem 2.1 is larger than the class of the chains constructed in Theorem 2.3.

Considering initial distribution having whole mass at $(1,\ldots,1)$ and/or all $q_j(1)=0, j=1,\ldots,d$ we have special cases, which we will formulate as a corollary.

Corollary 2.4. Consider the setup from Theorem 2.3.

a) Moreover, assume that $q_j(1)=0$ for all $j=1,\ldots,d$. I.e., each X^*_j has actually only one absorbing state (state 0 is not accessible). Then, $\mathbf{N}$ is the only absorbing state of X^*, $\sum_{\mathbf{e}\in\mathbb{E}}\hat{\nu}(\mathbf{e})=1$, $\rho_j(1)=1, j=1,\ldots,d$ and we have

$$\mathsf{pgf}_{T^*_{\nu^*,\mathbf{N}}}(s)=\sum_{\hat{\mathbf{e}}\in\mathbb{E}}\hat{\nu}(\hat{\mathbf{e}})\mathsf{pgf}_{\hat{T}_{\hat{\mathbf{e}},\mathbf{N}}}(s).$$

b) Moreover, assume that both $q_j(1)=0$ for all $j=1,\ldots,d$ and $\nu^*((1,\ldots,1))=1$. Then $T^*_{(1,\ldots,1),\mathbf{N}} \stackrel{d}{=} \hat{T}_{(1,\ldots,1),\mathbf{N}}$, where $\stackrel{d}{=}$ denotes the equality in the distribution.

c) Moreover, assume that $\nu^*((1,\ldots,1))=1$. Then assertions of Theorem 2.3 hold with $\hat{\nu}((1,\ldots,1))=1$ and we have

$$T^*_{(1,\ldots,1),\mathbf{N}} = \begin{cases} \hat{T}_{(1,\ldots,1),\mathbf{N}} & \text{with probability } \prod_{j=1}^d \rho_j(1), \\ -\infty & \text{with probability } 1-\prod_{j=1}^d \rho_j(1). \end{cases}$$

Sample-path construction It turns out that when $\hat{\nu}$ resulting from $\hat{\nu} = \nu^* \Lambda^{-1}$ is a distribution (which is always the case in, *e.g.*, Corollary 2.4 *b)* and *c)*), we can have a sample-path construction. *I.e.*, for X^* we can construct, sample path by sample path, a chain $\hat{X}$, so that $T^*_{\nu^*,\mathbf{N}}$ has the distribution expressed in terms of $\hat{T}_{\hat{\nu},\mathbf{N}}$ as stated in Theorem 2.3. The construction is analogous to the construction given in [DF90b] (paragraph 2.4) - note however that the construction therein was between ergodic chain and its strong stationary dual chain (*i.e.,* the chain with one absorbing state) and the link Λ was a stochastic matrix (it can be substochastic in our case). Having observed $X_0^* = \mathbf{e}_0^*$ (chosen from the distribution ν^*) we set

$$\hat{X}_0 = \hat{\mathbf{e}}_0 \text{ with probability } \frac{\hat{\nu}(\hat{\mathbf{e}}_0)\Lambda(\hat{\mathbf{e}}_0, \mathbf{e}_0^*)}{\nu^*(\mathbf{e}_0^*)}.$$

Then, after choosing $X_1^* = \mathbf{e}_1^*, \ldots, X_{n-1}^* = \mathbf{e}_{n-1}^*$ and $\hat{X}_1 = \hat{\mathbf{e}}_1, \ldots, \hat{X}_n = \hat{\mathbf{e}}_n$ we set

$$\hat{X}_n = \hat{\mathbf{e}}_n \text{ with probability } \frac{\mathbf{P}_{\hat{X}}(\hat{\mathbf{e}}_{n-1}, \hat{\mathbf{e}}_n)\Lambda(\hat{\mathbf{e}}_n, \mathbf{e}_n^*)}{(\mathbf{P}_{X^*}\Lambda)(\hat{\mathbf{e}}_{n-1}, \mathbf{e}_n^*)}.$$

This way we have constructed the chain $\hat{X}$ so that $\Lambda \mathbf{P}_{X^*} = \mathbf{P}_{\hat{X}}\Lambda$ and $\nu^* = \hat{\nu}\Lambda$ with the property that $\hat{X}_n = \hat{\mathbf{e}}_M$ if and only if $X_n^* = \mathbf{e}_M^*$.

Theorem 2.3 is actually neither an extension of (1.6) nor (1.7) to the multidimensional case, since for one-dimensional case the formula for `pgf` of $T^*_{s,N}$ has a different form, as examples given in Section 6 show.

3 Tools: dualities in Markov chains

Siegmund duality and intertwinings between chains are the key ingredients of our main theorems' proofs.

3.1 Siegmund duality

Let X be an ergodic discrete-time Markov chain with the transition matrix $\mathbf{P}_X$ and a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ partially ordered by $\preceq$. Denote its stationary distribution by π. We assume that there exists a unique minimal state, say $\mathbf{e}_1$, and a unique maximal state, say $\mathbf{e}_M$. For $A \subseteq \mathbb{E}$, define $\mathbf{P}_X(\mathbf{e}, A) := \sum_{\mathbf{e}' \in A} \mathbf{P}_X(\mathbf{e}, \mathbf{e}')$ and similarly $\pi(A) := \sum_{\mathbf{e} \in A} \pi(\mathbf{e})$. Define also $\{\mathbf{e}\}^\uparrow := \{\mathbf{e}' \in \mathbb{E} : \mathbf{e} \preceq \mathbf{e}'\}$, $\{\mathbf{e}\}^\downarrow := \{\mathbf{e}' \in \mathbb{E} : \mathbf{e}' \preceq \mathbf{e}\}$ and $\delta(\mathbf{e}, \mathbf{e}') = \mathbb{1}\{\mathbf{e} = \mathbf{e}'\}$. We say that a Markov chain Z with the transition matrix $\mathbf{P}_Z$ is the **Siegmund dual** of X if

$$\forall (\mathbf{e}_i, \mathbf{e}_j \in \mathbb{E}) \; \forall (n \geq 0) \quad \mathbf{P}_X^n(\mathbf{e}_i, \{\mathbf{e}_j\}^\downarrow) = \mathbf{P}_Z^n(\mathbf{e}_j, \{\mathbf{e}_i\}^\uparrow). \tag{3.1}$$

In all non-degenerated applications, we can find the *substochastic* matrix $\mathbf{P}'_Z$ fulfilling (3.1). Then we add one extra absorbing state, say $\mathbf{e}_0$, and define $\mathbf{P}_Z = \mathcal{F}_{\mathbf{e}_0}(\mathbf{P}'_Z)$. Note that then $\mathbf{P}_Z$ fulfills (3.1) for all states different from $\mathbb{E}$. This relation also implies that $\mathbf{e}_M$ is an absorbing state in the Siegmund dual, thus Z has two absorbing states. Taking the limits as $n \to \infty$ on both sides of (3.1), we have

$$\pi(\{\mathbf{e}_j\}^\downarrow) \;=\; \lim_{n\to\infty} \mathbf{P}_Z^n(\mathbf{e}_j, \{\mathbf{e}_i\}^\uparrow) \;=\; P(\tau_{\mathbf{e}_M} < \tau_{\mathbf{e}_0} | Z_0 = \mathbf{e}_j) = \rho(\mathbf{e}_j). \tag{3.2}$$

The stationary distribution of X is related in this way to the absorption of its Siegmund dual Z.

It is convenient to define the Siegmund duality in a matrix form. Let $\mathbf{C}(\mathbf{e}_i,\mathbf{e}_j)=\mathbb{1}(\mathbf{e}_i\preceq\mathbf{e}_j)$, then the equality (3.1) can be expressed as

$$\mathbf{P}_X^n\mathbf{C}=\mathbf{C}({\mathbf{P}'_Z}^n)^T. \tag{3.3}$$

Relation (3.2) can be rewritten in a matrix form as

$$\boldsymbol{\rho}=\boldsymbol{\pi}\mathbf{C}.$$

The inverse $\mathbf{C}^{-1}$ always exists, usually it is denoted by μ and called the *Möbius function*. To find a Siegmund dual it is enough to find $\mathbf{P}_Z$ fulfilling (3.3) with for $n=1$.

Let $\preceq:=\leq$ be a total ordering on a finite state space $\mathbb{E}=\{1,\ldots,M\}$. The chain Y is **stochastically monotone** w.r.t to the total ordering if $\forall i_1\leq i_2\ \forall j\ \mathbf{P}_Y(i_2,\{j\}^{\downarrow})\leq\mathbf{P}_Y(i_1,\{j\}^{\downarrow})$. We have

Lemma 3.1 ([Sie76]). Let X be an ergodic Markov chain on $\mathbb{E}=\{1,\ldots,M\}$ with the transition matrix $\mathbf{P}_X$. The Siegmund dual Z (w.r.t. the total ordering) exists if and only if X is stochastically monotone. In such a case $\mathbf{P}_Z=\mathcal{F}(\mathbf{P}'_Z)$, where

$$\mathbf{P}'_Z(j,i)=\mathbf{P}_X(i,\{j\}^{\downarrow})-\mathbf{P}_X(i+1,\{j\}^{\downarrow})$$

for $i,j\in\mathbb{E}$ (we mean $\mathbf{P}_X(i+1,\cdot)=0$).

Since the proof is one line long, we present it.

Proof of Lemma 3.1. The main thing is to show that (3.1) holds for $n=1$. We have

$$\mathbf{P}'_Z(j,i)\ =\ \mathbf{P}'_Z(j,\{i\}^{\uparrow})-\mathbf{P}'_Z(j,\{i+1\}^{\uparrow})=\mathbf{P}_X(i,\{j\}^{\downarrow})-\mathbf{P}_X(i+1,\{j\}^{\downarrow}).$$

The latter is non-negative if and only if X is stochastically monotone. □

Let X be an ergodic birth and death chain on $\mathbb{E}=\{1,\ldots,M\}$ with the transition matrix

$$\mathbf{P}_X(i,i')=\begin{cases}p'(i) & \text{if}\quad i'=i+1,\\ q'(i) & \text{if}\quad i'=i-1,\\ 1-(p'(i)+q'(i)) & \text{if}\quad i'=i,\end{cases} \tag{3.4}$$

where $q'(1)=p'(M)=0$ and $p'(i)>0,i=1,\ldots,M-1,q'(i)>0,i=2,\ldots,M$. Assume that $p'(i-1)+q'(i)\leq 1,i=2,\ldots,M$ (what is equivalent to stochastic monotonicity).

It is easily verifiable that when we rename transition probabilities: $p(i)=q'(i),q(i)=p'(i-1)$, then the transitions $\mathbf{P}_Y$ defined in (1.4) are the transitions of the Siegmund dual of $\mathbf{P}_X$ resulting from Lemma 3.1. From the known form of the stationary distribution of an ergodic birth and death chain, and from relation (3.2), it follows that for $\mathbf{P}_Y$ given in (1.4) we have

$$\rho(s)=\sum_{k\leq s}\pi(s)=\frac{\sum_{n=1}^{s}\prod_{r=1}^{n-1}\left(\frac{q(r)}{p(r)}\right)}{\sum_{n=1}^{M}\prod_{r=1}^{n-1}\left(\frac{q(r)}{p(r)}\right)}. \tag{3.5}$$

3.2 Intertwinings between absorbing chains

Let Λ be any nonsingular matrix of size $M \times M$. We say that matrices $\mathbf{P}_{X^*}$ and $\mathbf{P}_{\hat{X}}$ of size $M \times M$ are **intertwined by a link** Λ if

$$\Lambda \mathbf{P}_{X^*} = \mathbf{P}_{\hat{X}} \Lambda.$$

Similarly, we say that vectors $\hat{\nu}$ and ν^* of lengths M are intertwined if

$$\nu^* = \hat{\nu}\Lambda. \tag{3.6}$$

We say that a link Λ is $\mathbf{e}^*_M$-isolated if

$$\Lambda(\hat{\mathbf{e}}, \mathbf{e}^*_M) \quad \begin{cases} \neq 0 & \text{if } \hat{\mathbf{e}} = \hat{\mathbf{e}}_M, \\ \\ = 0 & \text{otherwise .} \end{cases} \tag{3.7}$$

Lemma 3.2. Let X^* and $\hat{X}$ be Markov chains on $\mathbb{E}^* = \mathbf{e}^*_0 \cup \hat{\mathbb{E}}$ and $\hat{\mathbb{E}}$ with transition matrices $\mathbf{P}_{X^*}$ and $\mathbf{P}_{\hat{X}}$ respectively. Moreover, assume X^* has the initial distribution ν^* and two absorbing states: $\mathbf{e}^*_0$ and $\mathbf{e}^*_M$, whereas $\hat{X}$ has one absorbing state $\hat{\mathbf{e}}_M$. Assume that $\mathbf{P}'_{X^*} = \mathcal{F}^{-1}_{\mathbf{e}^*_0}(\mathbf{P}_{X^*})$ and $\mathbf{P}_{\hat{X}}$ are intertwined via an $\mathbf{e}^*_M$-isolated link Λ. Let $\hat{\nu} = \nu^* \Lambda^{-1}$. Then we have

$$\mathsf{pgf}_{T^*_{\nu^*, \mathbf{e}^*_M}}(s) = \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \mathsf{pgf}_{\hat{T}_{\hat{\mathbf{e}}, \hat{\mathbf{e}}_M}}(s).$$

Proof.

$$\begin{aligned} P(T^*_{\nu^*, \mathbf{e}^*_M} \leq t) &= P(X^*(t) = \mathbf{e}^*_M) = \sum_{\mathbf{e}^* \in \mathbb{E}^* \setminus \{\mathbf{e}^*_0\}} \nu^*(\mathbf{e}^*) \mathbf{P}^t_{X^*}(\mathbf{e}^*, \mathbf{e}^*_M) \\ &= \sum_{\hat{\mathbf{e}} \in \mathbb{E}} \sum_{\mathbf{e}^* \in \mathbb{E}^* \setminus \{\mathbf{e}^*_0\}} \hat{\nu}(\hat{\mathbf{e}}) \Lambda(\hat{\mathbf{e}}, \mathbf{e}^*) \mathbf{P}^t_{X^*}(\mathbf{e}^*, \mathbf{e}^*_M) \\ &= \sum_{\hat{\mathbf{e}} \in \mathbb{E}} \sum_{\hat{\mathbf{e}}_2 \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \mathbf{P}^t_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_2) \Lambda(\hat{\mathbf{e}}_2, \mathbf{e}^*_M) \\ &= \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \mathbb{E}} \hat{\nu}(\hat{\mathbf{e}}) \mathbf{P}^t_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_M). \end{aligned}$$

Now, for pgf we have:
$\mathsf{pgf}_{T^*_{\nu^*, \mathbf{e}^*_M}}(s) =$

$$\begin{aligned} & \sum_{k=0}^{\infty} P(T^*_{\nu^*, \mathbf{e}^*_M} = k) s^k = \sum_{k=0}^{\infty} \Big(P(T^*_{\nu^*, \mathbf{e}^*_M} \leq k) - P(T^*_{\nu^*, \mathbf{e}^*_M} \leq k-1) \Big) s^k \\ = \quad & \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{k=0}^{\infty} \left(\sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \mathbf{P}^k_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_M) - \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \mathbf{P}^{k-1}_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_M) \right) s^k \\ = \quad & \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \sum_{k=0}^{\infty} \Big(\mathbf{P}^k_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_M) - \mathbf{P}^{k-1}_{\hat{X}}(\hat{\mathbf{e}}, \hat{\mathbf{e}}_M) \Big) s^k \\ = \quad & \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \sum_{k=0}^{\infty} \Big(P(\hat{T}_{\hat{\mathbf{e}}, \hat{\mathbf{e}}_M} \leq k) - P(\hat{T}_{\hat{\mathbf{e}}, \hat{\mathbf{e}}_M} \leq k-1) \Big) s^k \\ = \quad & \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \sum_{k=0}^{\infty} P(\hat{T}_{\hat{\mathbf{e}}, \hat{\mathbf{e}}_M} = k) s^k = \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M) \sum_{\hat{\mathbf{e}} \in \hat{\mathbb{E}}} \hat{\nu}(\hat{\mathbf{e}}) \mathsf{pgf}_{\hat{T}_{\hat{\mathbf{e}}, \hat{\mathbf{e}}_M}}(s). \end{aligned}$$

□

Corollary 3.3. Let assumptions of Lemma 3.2 hold and, in addition, let $\hat{\nu}$ be a distribution. Then, we have

$$T^*_{\nu^*,\mathbf{e}^*_M} = \Lambda(\hat{\mathbf{e}}_M, \mathbf{e}^*_M)\hat{T}_{\hat{\nu},\hat{\mathbf{e}}_M}.$$

From [Fil09b] we can deduce the following lemma.

Lemma 3.4. Let X^* be a birth and death chain on $\mathbb{E}^* = \{0, \ldots, M\}$ with the transition matrix $\mathbf{P}_{X^*}$ given in (1.4) with two absorbing states: 0 and M. Let $\mathbf{P}'_{X^*} = \mathcal{F}_0^{-1}(\mathbf{P}_{X^*})$. Assume the eigenvalues of $\mathbf{P}'_{X^*}$ are positive, denote them by $0 < \lambda_1 < \ldots < \lambda_M = 1$.
Define $\mathbf{Q}_1 := \mathbf{I}$ and

$$\mathbf{Q}_k := \frac{(\mathbf{P}'_{X^*} - \lambda_1 \mathbf{I}) \cdots (\mathbf{P}'_{X^*} - \lambda_{k-1}\mathbf{I})}{(1-\lambda_1)\cdots(1-\lambda_{k-1})}, k = 2, \ldots, M$$

Let Λ be the lower triangular square matrix of size $M \times M$ defined as

$$\Lambda(k, \cdot) = \mathbf{Q}_k(1, \cdot), \quad k = 1, \ldots, M. \tag{3.8}$$

Then, $\mathbf{P}'_{X^*}$ and $\mathbf{P}_{\hat{X}}$ are intertwined by the link Λ, where

$$\mathbf{P}_{\hat{X}}(i, i') = \begin{cases} 1 - \lambda_i & \text{if } i' = i+1, \\ \lambda_i & \text{if } i' = i, \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

is a matrix of size $M \times M$.

Note that Lemma 3.4 is similar to Theorem 4.2 in [Fil09b], the difference is that therein Λ is a stochastic matrix, whereas in Lemma 3.4 it can be substochastic (it is strictly substochastic if $q(0) > 0$). An almost identical Λ was considered in [GMZ12], their Proposition 3.3 yields:

Lemma 3.5.

- The matrices $\mathbf{Q}_k, 1, \ldots, M$ are non-negative and substochastic.
- The matrix Λ is non-negative and substochastic, it is lower triangular and

  $$\Lambda(1, 1) = 1, \qquad \Lambda(M, M) = \rho(1),$$

  thus Λ is nonsingular.

Remark 3.1. Note that in case X^* has no transition to 0, *i.e.,* $q(1) = 0$, it is actually a chain on $\{1, \ldots, M\}$ and $\mathbf{P}'_{X^*} = \mathcal{F}_0^{-1}(\mathbf{P}_{X^*})$ is a stochastic matrix. Then Λ is a stochastic matrix and $\Lambda(M, M) = 1$.

4 Proofs

4.1 Properties of the Kronecker product

In this section we recall some useful properties of the Kronecker product and formulate a lemma relating eigenvectors and eigenvalues of some combination of Kronecker products.

We will exploit the following properties

- bilinearity:

$$\mathbf{A}\otimes(\mathbf{B}+\mathbf{C})=\mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}, \tag{P1}$$

- mixed product:

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC})\otimes(\mathbf{BD}), \tag{P2}$$

- inverse and transposition:

$$(\mathbf{A}\otimes\mathbf{B})^{-1}=(\mathbf{A})^{-1}\otimes(\mathbf{B})^{-1}, \tag{P3}$$

$$(\mathbf{A}\otimes\mathbf{B})^{T}=(\mathbf{A})^{T}\otimes(\mathbf{B})^{T}. \tag{P4}$$

- eigenvalue and eigenvector:

having eigenvalues α_j with corresponding left eigenvectors a_j for each matrix $\mathbf{A}_j, j\le n,$ we note that $\sum_{j\le n}\alpha_j$ with $\bigotimes_{j\le n}a_j$ and $\prod_{j\le n}\alpha_j$ with $\bigotimes_{j\le n}a_j$ are the eigenvalue and the left eigenvector of $\mathbf{A}=\bigoplus_{j\le n}\mathbf{A}_j$ and $\mathbf{A'}=\bigotimes_{j\le n}\mathbf{A}_j$ respectively. (P5)

Last property leads us to the following lemma.

Lemma 4.1. For all $1\le j\le n$ and $1\le i\le m$, let a_j be the left eigenvectors with the corresponding eigenvalues α_j of square matrices $\mathbf{A}_j^{(i)}$ of sizes k_j respectively.
Let $\mathbf{B}_i, i=1,\ldots,m$ be square matrices of sizes $\prod_{j=1}^n k_j$ such that $\sum_{i=1}^m\mathbf{B}_i=\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of size $\prod_{j=1}^n k_j$. Then $\prod_{j\le n}\alpha_j$ with $\bigotimes_{j\le n}a_j$ are the eigenvalue and the left eigenvector of $\mathbf{A}=\sum_{i=1}^m(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})\mathbf{B}_i$.
Similarly, if $b_i, i=1,\ldots,m$ are real numbers such that $\sum_{i=1}^m b_i$ we have that $\bigotimes_{j\le n}a_j$ is the left eigenvector with the corresponding eigenvalue $\prod_{j\le n}\alpha_j$ of the matrix $\mathbf{A}=\sum_{i=1}^m(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})b_i$.

Proof. We have

$$\begin{aligned}\bigotimes_{j\le n}a_j\sum_{i=1}^m(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})\mathbf{B}_i &= \sum_{i=1}^m\bigotimes_{j\le n}a_j(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})\mathbf{B}_i=\sum_{i=1}^m\bigotimes_{j\le n}(a_j\mathbf{A}_j^{(i)})\mathbf{B}_i\\ &= \sum_{i=1}^m\bigotimes_{j\le n}(a_j\alpha_j)\mathbf{B}_i=\sum_{i=1}^m\prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_j\mathbf{B}_i\\ &= \prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_j\sum_{i=1}^m\mathbf{B}_i=\prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_j.\end{aligned}$$

Similarly,

$$\begin{aligned}\bigotimes_{j\le n}a_j\sum_{i=1}^m(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})b_i &= \sum_{i=1}^m\bigotimes_{j\le n}a_j(\bigotimes_{j\le n}\mathbf{A}_j^{(i)})b_i=\sum_{i=1}^m\bigotimes_{j\le n}(a_j\mathbf{A}_j^{(i)})b_i\\ &= \sum_{i=1}^m\bigotimes_{j\le n}(a_j\alpha_j)b_i=\sum_{i=1}^m\prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_jb_i\\ &= \prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_j\sum_{i=1}^m b_i=\prod_{j\le n}\alpha_j\bigotimes_{j\le n}a_j.\end{aligned}$$

□

Substituting stochastic matrices $\mathbf{P}_j^{(i)}$ with stationary distributions π_j (for all $1 \le j \le n$, $1 \le i \le m$) to matrices $\mathbf{A}_j^{(i)}$ with left eigenvectors α_j (for all $1 \le j \le n$, $1 \le i \le m$) gives us the following corollary (keeping in mind that 1 is the eigenvalue corresponding to the eigenvector being the stationary distribution):

Corollary 4.2. Let $\mathbf{P}_j^{(i)}$ be a stochastic matrix of size k_j with π_j being its stationary distribution for all $1 \le j \le n, 1 \le i \le m$. Let $\mathbf{B}_i$, $1 \le i \le m$ be square matrices of sizes $\prod_{j=1}^n k_j$ such that $\sum_{i=1}^m \mathbf{B}_i = \mathbf{I}$, where $\mathbf{I}$ is the identity of size $\prod_{j=1}^n k_j$. Similarly, if b_i, $1 \le i \le m$ are real numbers such that $\sum_{i=1}^m b_i = 1$, then the stochastic matrices of the form $\sum_{i=1}^m (\bigotimes_{j \le n} \mathbf{P}_j^{(i)})\mathbf{B}_i$ or $\sum_{i=1}^m (\bigotimes_{j \le n} \mathbf{P}_j^{(i)}) b_i$ have the stationary distribution of the form $\bigotimes_{j \le n} \pi_j$.

4.2 Proof of Theorem 2.1

We will find an ergodic Markov chain X with the transition matrix $\mathbf{P}_X$ and some partial ordering of the state space (expressed by the ordering matrix $\mathbf{C}$) and show that (3.2) is equivalent to (2.3).

Let $\mathbf{P}_{Z_j}^{(k)}$ (on $\mathbb{E}_j = \{0, \ldots, N_j\}$) be as in the theorem. Let $X_j^{(k)}$ be the ergodic chain on $\mathbb{E}'_j = \{1, \ldots, N_j\}$ with the transition matrix $\mathbf{P}_{X_j}^{(k)}$, such that $Z_j^{(k)}$ is its Siegmund dual w.r.t. the total ordering. *I.e.*, let $\mathbf{C}_j(s,t) = \mathbb{1}(s \le t)$, and duality means that

$$\mathbf{P}_{X_j}^{(k)} \mathbf{C}_j = \mathbf{C}_j (\mathbf{P}_{Z'_j}^{(k)})^T,$$

where $\mathbf{P}_{Z'_j}^{(k)} = \mathcal{F}_0^{-1}(\mathbf{P}_{Z_j}^{(k)})$. Assumption (2.1) and relation (3.2) imply that for fixed j, the chains $X_j^{(k)}, k = 1, \ldots, m$ have the same stationary distribution, denote it by π_j. The relation (3.2) means that $\boldsymbol{\rho}_j = \boldsymbol{\pi}_j \mathbf{C}_j$. On the state space $\mathbb{E} = \bigotimes_{j \le d} \mathbb{E}_j$ let us introduce the ordering expressed by the matrix $\mathbf{C} = \bigotimes_{j \le d} \mathbf{C}_j$. From (3.3) we can calculate the matrix $\mathbf{P}_X$:

$$\begin{aligned}
\mathbf{P}_X \quad &= \quad \mathbf{C}\mathbf{P}_Z^T\mathbf{C}^{-1} = (\bigotimes_{j \le d} \mathbf{C}_j) \left(\sum_{k=1}^m \mathbf{B}_k (\bigotimes_{j \le d} \mathbb{R}_{Z'_j}^{(k)}) \right)^T (\bigotimes_{j \le d} \mathbf{C}_j)^{-1} \\
&\overset{(P4),(P3)}{=} (\bigotimes_{j \le d} \mathbf{C}_j) \left(\sum_{k=1}^m (\bigotimes_{j \le d} (\mathbb{R}_{Z'_j}^{(k)})^T) \mathbf{B}_k^T \right) (\bigotimes_{j \le d} \mathbf{C}_j^{-1}) \\
&\overset{(P1)}{=} \sum_{k=1}^m (\bigotimes_{j \le d} \mathbf{C}_j) \left(\bigotimes_{j \le d} (\mathbb{R}_{Z'_j}^{(k)})^T \right) \mathbf{B}_k^T (\bigotimes_{j \le d} \mathbf{C}_j^{-1}) \\
&\overset{(P2)}{=} \sum_{k=1}^m (\bigotimes_{j \le d} \mathbf{C}_j (\mathbb{R}_{Z'_j}^{(k)})^T \mathbf{C}_j^{-1}) (\bigotimes_{j \le d} \mathbf{C}_j) \mathbf{B}_k^T (\bigotimes_{j \le d} \mathbf{C}_j^{-1}).
\end{aligned}$$

Let us define

$$\mathbb{R}_{X_j}^{(k)} = \mathbf{C}_j (\mathbb{R}_{Z'_j}^{(k)})^T \mathbf{C}_j^{-1} = \begin{cases} \mathbf{P}_{X_j}^{(k)} & \text{if } j \in \mathbf{A}_k, \\ \mathbf{I}_j & \text{if } j \notin \mathbf{A}_k. \end{cases}$$

In the case $j \in \mathbf{A}_k$, the distribution π_j is the unique stationary distribution. In the case $j \notin \mathbf{A}_k$, any distribution is an invariant measure, however, we fix it to be π_j. We have

$$\mathbf{P}_X = \sum_{k=1}^m (\bigotimes_{j \le d} \mathbb{R}_{X_j}^{(k)}) (\bigotimes_{j \le d} \mathbf{C}_j) \mathbf{B}_k^T (\bigotimes_{j \le d} \mathbf{C}_j^{-1}).$$

From the property (P1) we have that

$$\sum_{k=1}^{m}(\bigotimes_{j\leq n}\mathbf{C}_j)\mathbf{B}_k^T(\bigotimes_{j\leq n}\mathbf{C}_j^{-1}) = (\bigotimes_{j\leq n}\mathbf{C}_j)\sum_{k=1}^{m}\mathbf{B}_k^T(\bigotimes_{j\leq n}\mathbf{C}_j^{-1}) = (\bigotimes_{j\leq n}\mathbf{C}_j)(\bigotimes_{j\leq n}\mathbf{C}_j^{-1}) = \mathbf{I},$$

thus Corollary 4.2 implies that $\boldsymbol{\pi} = \bigotimes_{j\leq d}\boldsymbol{\pi}_j$ is the stationary distribution of $\mathbf{P}_X$, thus $\boldsymbol{\rho} = \boldsymbol{\pi}\mathbf{C}$, what is equivalent to (2.3).

□

4.3 Proof of Theorem 2.3

To prove the theorem we will construct an $\mathbf{N}$-isolated link Λ, so that $\mathbf{P}'_{X^*}$ and $\mathbf{P}_{\hat{X}}$, given in (2.5) and (2.6) respectively, are intertwined via this link.
Consider the matrix $\mathbf{P}'_{X^*_j}$. Define the stochastic matrix $\mathbf{P}_{\hat{X}_j}$ of size $N_j \times N_j$ as:

$$\mathbf{P}_{\hat{X}_j}(i,i') = \begin{cases} 1-\lambda_i^{(j)} & \text{if } i' = i+1, \\ \lambda_i^{(j)} & \text{if } i' = i, \\ 0 & \text{otherwise.} \end{cases}$$

Let Λ_j be the link intertwining matrices $\mathbf{P}'_{X^*_j}$ and $\mathbf{P}_{\hat{X}_j}$ given in (3.8). Define

$$\mathbb{R}_{\hat{X}_j^{(k)}} = \begin{cases} \mathbf{P}_{\hat{X}_j} & \text{if } j \in \mathbf{A}_k, \\ \mathbf{I}_j & \text{if } j \notin \mathbf{A}_k. \end{cases}$$

Note that matrices $\mathbf{R}'_{X^*_j}$ and $\mathbb{R}_{\hat{X}_j^{(k)}}$ are also intertwined via Λ_j for any $j = 1,\ldots,d$ and any $k = 1,\ldots,m$. Any link intertwines two identity matrices, which is the case for $j \notin \mathbf{A}_k$. *I.e.*, we have $\Lambda_j\mathbf{R}'_{X^*_j} = \mathbb{R}_{\hat{X}_j^{(k)}}\Lambda_j, j = 1,\ldots,d$. Define

$$\Lambda = \bigotimes_{j\leq d}\Lambda_j.$$

We have

$$\begin{aligned}\Lambda\mathbf{P}'_{X^*} &= \bigotimes_{j\leq d}\Lambda_j\sum_{k=1}^{m}b_k\left(\bigotimes_{j\leq d}\mathbb{R}'_{X_j^{*(k)}}\right) = \sum_{k=1}^{m}b_k\left(\bigotimes_{j\leq d}\Lambda_j\mathbb{R}'_{X_j^{*(k)}}\right) \\ &= \sum_{k=1}^{m}b_k\left(\bigotimes_{j\leq d}\mathbb{R}_{\hat{X}_j^{(k)}}\Lambda_j\right) = \sum_{k=1}^{m}b_k\left(\bigotimes_{j\leq d}\mathbb{R}_{\hat{X}_j^{(k)}}\right)\bigotimes_{j\leq d}\Lambda_j\end{aligned}$$

Simple calculations yield that $\mathbf{P}_{\hat{X}}$ given in (2.6) can be written as $\sum_{k=1}^{m}b_k\left(\bigotimes_{j\leq d}\mathbb{R}_{\hat{X}_j^{(k)}}\right)$. Thus, we have $\Lambda\mathbf{P}'_{X^*} = \mathbf{P}_{\hat{X}}\Lambda$. Now, let us calculate $\hat{\nu} = \nu^*\Lambda^{-1}$ (note that Λ is nonsingular because of the property (P3) and the fact that each $\Lambda_j, j = 1,\ldots,d$ and identity matrices $\mathbf{I}_j$ are nonsingular). In other words, we have $\nu^* = \hat{\nu}\Lambda$. The equation (2.7) holds, since Λ is lower triangular. Moreover, Λ is $(N_1,\ldots,N_d)$-isolated, since we have

$$\Lambda((i_1,\ldots,i_d),\mathbf{N}) = \prod_{j=1}^{d}\Lambda(i_j,N_j) \overset{(*)}{=} \begin{cases} \prod_{j=1}^{d}\rho_j(1) & \text{if } i_j = N_j \text{ for all j} = 1,\ldots,\text{d}, \\ 0 & \text{otherwise ,} \end{cases}$$

where in $\stackrel{(*)}{=}$ we used Lemma 3.5. Applying Lemma 3.2 completes the proof.

□

5 The outline of an alternative proof of Theorem 2.3: strong stationary duality approach

In Theorem 2.3 we related the absorption time $T^*_{\nu^*,\mathbf{N}}$ of X^* with the absorption time $\hat{T}_{\hat{\nu},\mathbf{N}}$ of $\hat{X}$. This was done by finding a specific matrix Λ, such that $\Lambda\mathbf{P}_{X^*} = \mathbf{P}_{\hat{X}}\Lambda$, exploiting the existence of such Λ for X^* and $\hat{X}$ being birth and death chains. The exploited Λ is related to spectral polynomials of the stochastic matrix $\mathbf{P}_{X^*}$. Such a link appeared first naturally as a link between an ergodic chain X and an absorbing chain X^*. The proof of Theorem 2.3 in case $q_j(1) = 0$ (*i.e.,* Corollary 2.4 a)) can be different, using intermediate ergodic chain. In this section we will describe its outline.

Strong stationary duality Let X be an ergodic Markov chain on $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ with the initial distribution ν and the transition matrix $\mathbf{P}_X$. Let $\mathbb{E}^* = \{\mathbf{e}^*_1, \ldots, \mathbf{e}^*_M\}$ be the, possibly different, state space of the absorbing Markov chain X^*, with the transition matrix $\mathbf{P}_{X^*}$ and the initial distribution ν^*, whose unique absorbing state is denoted by $\mathbf{e}^*_M$. Assume that Λ^* is a stochastic $M \times N$ matrix satisfying $\Lambda(\mathbf{e}^*_M, \mathbf{e}) = \pi(\mathbf{e})$. We say that X^* is a *strong stationary dual* (SSD) of X with link Λ^* if

$$\nu = \nu^*\Lambda \quad \text{and} \quad \Lambda^*\mathbf{P}_X = \mathbf{P}_{X^*}\Lambda^*. \tag{5.1}$$

[DF90b] prove that then the absorption time T^* of X^* is the so called *strong stationary time* for X. This is a random variable T such that X_T has the distribution π and T is independent of X_T. The main application is to study the rate of convergence of an ergodic chain to its stationary distribution, since for such a random variable we always have $d_{TV}(\nu\mathbf{P}^k_X, \pi) \leq sep(\nu\mathbf{P}^k_X, \pi) \leq P(T > k)$, where d_{TV} stands for the *total variation distance*, and *sep* stands for the *separation 'distance'*. For details, see [DF90b]. We say that SSD is **sharp** if T^* corresponds to stochastically the smallest SST, then we have $sep(\nu\mathbf{P}^k_X, \pi) = P(T^* > k)$, the corresponding SST T^* is often called the **fastest strong stationary time (FSST)**.

Strong stationary duality for birth and death chain Let X be an ergodic birth and death chain on $\mathbb{E} = \{1, \ldots, M\}$, whose time reversal is stochastically monotone with transitions given in (3.4). [DF90b] show that an absorbing birth and death chain X^* on $\mathbb{E}^* = \mathbb{E}$ with transitions given by

$$\begin{aligned} \mathbf{P}_{X^*}(i, i-1) &= \tfrac{H(i-1)}{H(i)} p'(i), \\ \mathbf{P}_{X^*}(i, i+1) &= \tfrac{H(i+1)}{H(i)} q'(i+1), \\ \mathbf{P}_{X^*}(i, i) &= 1 - (p'(i) + q'(i+1)), \end{aligned}$$

is a *sharp* SSD for X. Here we have

$$H(j) = \sum_{k \leq j} \pi(k), \qquad \Lambda^*(i, j) = \mathbb{1}\{i \leq j\} \frac{\pi(i)}{H(j)}. \tag{5.2}$$

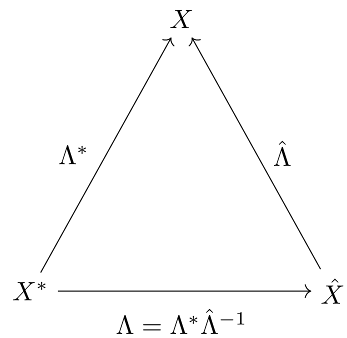


Figure 2: Intertwining between absorbing chains X^* and $\hat{X}$ via ergodic chain X.

Moreover, starting from an absorbing birth and death chain X^* on $\mathbb{E} = \{1, \ldots, M\}$, whose unique absorbing state is M, Theorem 3.1 in [Fil09b] states that we can find an ergodic chain X (and some stationary distribution π), such that X^* is its *sharp* SSD with the link given in (5.2).

Spectral pure-birth chain Again, let X be an ergodic birth and death chain on $\mathbb{E} = \{1, \ldots, M\}$. Assume its eigenvalues are non-negative, $0 \leq \lambda_1 \leq \ldots \leq \lambda_M = 1$. Then, the chain $\hat{X}$ with transitions given in (3.9) is its *sharp* SSD with the link $\hat{\Lambda}$ given in (3.8).

The outline of an alternative proof As in previous section, the main idea is to show that two absorbing birth and death chains X_j^* and $\hat{X}_j$ (pure-birth) on $\mathbb{E}_j = \{1, \ldots, N_j\}$ are intertwined by an N_j-isolated link Λ_j. Collecting above findings, we have (skipping conditions on initial distributions):

- Let X_j be an ergodic chain on $\mathbb{E}_j$, whose X_j^* is a sharp SSD, *i.e.,* we have $\Lambda_j^* \mathbf{P}_{X_j} = \mathbf{P}_{X_j^*} \Lambda_j^*$.
- Let $\hat{X}_j$ be a pure-birth sharp SSD for X_j, *i.e.,* we have $\hat{\Lambda}_j \mathbf{P}_X = \mathbf{P}_{\hat{X}_j} \hat{\Lambda}_j$.

It means that absorption times T^* and $\hat{T}$ are equal in distribution (since both X^* and $\hat{X}$ are *sharp* SSDs of X). Mathematically, we have

$$\Lambda_j \mathbf{P}_{X_j^*} = \mathbf{P}_{\hat{X}_j} \Lambda_j, \qquad \text{where} \quad \Lambda_j = \hat{\Lambda}_j \left(\Lambda_j^* \right)^{-1},$$

i.e., X_j^* and $\hat{X}_j$ are intertwined by the link Λ_j, which is N_j-isolated. Intertwining between two absorbing birth and death chains via an ergodic chain is depicted in Fig. 5. Taking $\Lambda = \bigotimes_{j \leq d} \Lambda_j$ and $\hat{\nu} = \nu^* \Lambda^{-1}$ we proceed with the proof of Theorem 2.3 as in previous section.

6 Examples

In first two subsections 6.1 and 6.2 we will present examples on the absorption time of some one-dimensional birth and death chains. Although we mainly focus on multidimensional generalizations, we consider these examples worth presenting. Next two subsections 6.3 and 6.4 contain some non-trivial multidimensional gambler models, for which we either provide results for both, the winning probability and the absorption time (Example 6.3) or only for the winning probability (Example 6.4).

6.1 A one-dimensional gambler's ruin problem with $N = 3$: calculating $T^*_{2,3}$

Here we present a simple example for calculating $T^*_{2,3}$ in a one-dimensional gambler's ruin problem using Theorem 2.3. We also check that calculations agree with the formula (1.6).

Example 6.1. Let $d = 1, N_1 = 3$ and $p_1(1) = p_1(2) = p > 0, q_1(1) = q_1(2) = q > 0$ such that $p \neq q$ and $p + q + \sqrt{pq} < 1$. The transition matrix of $\mathbf{P}_{X_1^*}$ is following

$$\mathbf{P}_{X_1^*} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ q & 1-p-q & p & 0 \\ 0 & q & 1-p-q & p \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Then, the pgf of the time to absorption starting at 2 is given by:

$$\mathsf{pgf}_{T^*_{2,3}}(s) = \frac{p(q+p+\sqrt{qp})(-q-p+\sqrt{qp})u(1-u(1-q-p))}{(p^2+qp+q^2)(1-u(1-q-p-\sqrt{qp})(-1+u(1-q-p+\sqrt{qp})} \tag{6.1}$$

Proof. We have $\mathbf{P}_{X^*} = \mathbf{P}_{X_1^*}$. The eigenvalues of $\mathbf{P}'_{X^*} = \mathcal{F}_0^{-1}(\mathbf{P}_{X^*})$ are $\lambda_1 = 1-p-q-\sqrt{pq}, \lambda_2 = 1-p-q+\sqrt{pq}, \lambda_3 = 1$. The transitions of $\mathbf{P}_{\hat{X}}$ are following

$$\mathbf{P}_{\hat{X}} = \begin{pmatrix} \lambda_1 & 1-\lambda_1 & 0 \\ 0 & \lambda_2 & 1-\lambda_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus,

$$\mathsf{pgf}_{\hat{T}_{1,3}}(s) = \frac{(1-\lambda_1)(1-\lambda_2)s^2}{(1-\lambda_1 s)(1-\lambda_2 s)}, \quad \mathsf{pgf}_{\hat{T}_{2,3}}(s) = \frac{(1-\lambda_2)s}{(1-\lambda_2 s)}.$$

Calculating Λ from (3.8) (for $\mathbf{P}_{X^*}$) we obtain

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\sqrt{pq}}{q+p+\sqrt{qp}} & \frac{p}{q+p+\sqrt{qp}} & 0 \\ 0 & 0 & \rho(1) \end{pmatrix}.$$

Calculations yield (we have $\nu^*(2) = 1$)

$$\hat{\nu} = \nu^* \Lambda^{-1} = \left(-\sqrt{\frac{q}{p}}, 1 + \frac{q}{p} + \sqrt{\frac{q}{p}}\right).$$

From (1.6) we have $\rho(i) = \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^3}, i = 1, 2, 3$. Finally,

$$\mathsf{pgf}_{T^*_{2,3}}(s) = \rho(1)\left(-\sqrt{\frac{q}{p}}\mathsf{pgf}_{\hat{T}_{1,3}}(s) + \left(1 + \frac{q}{p} + \sqrt{\frac{q}{p}}\right)\mathsf{pgf}_{\hat{T}_{2,3}}(s)\right),$$

what can be written as (6.1). On the other hand, (1.7) states that

$$\mathsf{pgf}_{T^*_{2,3}}(s) = \rho(2)\frac{\frac{(1-\lambda_1)(1-\lambda_2)s^2}{(1-\lambda_1 s)(1-\lambda_2 s)}}{\frac{(1-\lambda_1^{\lfloor 2 \rfloor})s}{1-\lambda_1^{\lfloor 2 \rfloor}s}} = \frac{1-\left(\frac{q}{p}\right)^2}{1-\left(\frac{q}{p}\right)^3} \cdot \frac{(1-\lambda_1)(1-\lambda_2)s^2(1-\lambda_1^{\lfloor 2 \rfloor}s)}{(1-\lambda_1 s)(1-\lambda_2 s)(1-\lambda_1^{\lfloor 2 \rfloor})s},$$

where $\lambda_1^{\lfloor 2 \rfloor} = 1-(p+q)$, which, as can be checked, is equivalent to (6.1). □

6.2 A one-dimensional gambler's ruin problem related to the Ehrenfest model: calculating $T^*_{m,N}$

Here we present a concrete example of a birth and death chain on $\mathbb{E} = \{1, \ldots, N\}$ with N being the only absorbing state, for which we provide pgf of the absorption time provided the chain started at an arbitrary $m \in \mathbb{E}$. We use Lemma 3.2 to calculate the link Λ. As far as we are aware, this pgf cannot be given using results from [GMZ12] *i.e.,* (1.6). This is because the eigenvalues of the transition matrix $\mathbf{P}_{X^*}$ are known, but the eigenvalues of the truncated matrix $\mathbf{P}^{\lceil m \rceil}_{X^*}$ are not known for an arbitrary $m \in \mathbb{E}$.

Example 6.2. Let X^* be a Markov chain on the state space $\mathbb{E} = \{1, \ldots, N\}$ with the transition matrix $\mathbf{P}_{X^*}$ of the form:

$$\mathbf{P}_{X^*}(i, i') = \begin{cases} \frac{N-i}{2N-2} \frac{\sum_{r=0}^{i-1} \binom{N-1}{r-1}}{\sum_{r=0}^{i} \binom{N-1}{r-1}} & \text{if } i' = i-1, i < N, \\ \frac{N-2}{2N-2} & \text{if } i' = i, i < N, \\ 1 & \text{if } i' = i = N, \\ \frac{i}{2N-2} \frac{\sum_{r=0}^{i+1} \binom{N-1}{r-1}}{\sum_{r=0}^{i} \binom{N-1}{r-1}} & \text{if } i' = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

Then the absorption time starting at $m \in \mathbb{E}$ has the following pgf:

$$\mathsf{pgf}_{T^*_{m,d}}(s) = \sum_{j \le m} \hat{\nu}(j) \mathsf{pgf}_{\hat{T}_{j,d}}(s), \tag{6.2}$$

where

$$\hat{\nu}(j) = \frac{2^{j-1}(-1)^{m+j}(m-j+1)\binom{N-1}{m}\binom{m}{j-1}}{(N-j)\sum_{k=0}^{m-1}\binom{N-1}{k}} \quad \text{and} \quad \mathsf{pgf}_{\hat{T}_{j,N}}(s) = \prod_{k=j}^{N-1} \left[\frac{(1 - \frac{k-1}{N-1})s}{1 - \frac{k-1}{N-1}s} \right]. \tag{6.3}$$

In particular, we have

$$E(T^*_{m,N}) = (N-1) \sum_{j \le m} \hat{\nu}(j) \sum_{k=j}^{N-1} \frac{1}{N-k}. \tag{6.4}$$

Proof. Let

$$\mathbf{P}_{\hat{X}}(i, i') = \begin{cases} \frac{i-1}{N-1} & \text{if } i' = i, \\ \frac{N-i}{N-1} & \text{if } i' = i+1, \\ 0 & \text{otherwise.} \end{cases}$$

To show the result using Lemma 3.2 it is enough to find Λ such that $\Lambda \mathbf{P}_{X^*} = \mathbf{P}_{\hat{X}} \Lambda$ and $\nu^* \Lambda^{-1} = \hat{\nu}$, where $\nu^*(j) = \mathbb{1}\{j = m\}$.

However, since X^* has only one absorbing state, we can – and we will – follow *the outline of an alternative proof. I.e.,* we will indicate intermediate ergodic chain X on $\mathbb{E}$ with the transition matrix $\mathbf{P}_X$ and find Λ^* and $\hat{\Lambda}$ such that $\Lambda^* \mathbf{P}_X = \mathbf{P}_{X^*} \Lambda^*$ and $\hat{\Lambda} \mathbf{P}_X = \mathbf{P}_{\hat{X}} \hat{\Lambda}$. Then, we will automatically have $\Lambda = \hat{\Lambda}(\Lambda^*)^{-1}$ and we will show that $\nu^* \Lambda^{-1} = \nu^* \Lambda^* \hat{\Lambda}^{-1} = \hat{\nu}$.

Let X be a chain on $\mathbb{E}$ with the following transition matrix:

$$\mathbf{P}_X(i,i') = \begin{cases} \frac{i-1}{2(N-1)} & \text{if } i' = i-1, \\ \frac{1}{2} & \text{if } i' = i, \\ \frac{N-i}{2(N-1)} & \text{if } i' = i+1, \\ 0 & \text{otherwise,} \end{cases}$$

i.e., X corresponds to the Ehrenfest model of $N-1$ particles with an extra probability (half) of staying (and states are enumerated $1,\ldots,N$, whereas in the classical Ehrenfest model these are $0,\ldots,N-1$). Its stationary distribution is the binomial distribution $\pi(j) = \frac{1}{2^{N-1}}\binom{N-1}{j-1}$, thus the classical link (cf. (5.2)) is given by

$$\Lambda^*(i,j) = \frac{\binom{N-1}{j-1}}{\sum_{r=0}^{i-1}\binom{N-1}{r}}\mathbb{1}\{j \leq i\},$$

i.e., we have $\Lambda^*\mathbf{P}_X = \mathbf{P}_{X^*}\Lambda$ (X^* is a sharp SSD of X). The eigenvalues of X are known, these are $\frac{i}{N-1}, i = 0,\ldots,N-1$, thus $\hat{X}$ is its pure-birth spectral dual. The link $\hat{\Lambda}$ such that $\hat{\Lambda}\mathbf{P}_X = \mathbf{P}_{\hat{X}}\hat{\Lambda}$ is known (see Eq. (4.6) in [Fil09b]), it is given by

$$\hat{\Lambda}(i,j) = \frac{\binom{i-1}{j-1}}{2^{i-1}}.$$

It can be checked that

$$\hat{\Lambda}^{-1}(i,j) = (-1)^{j-i}2^{j-1}\binom{i-1}{j-1}.$$

Note that the i-th row of $\hat{\Lambda}^{-1}$ corresponds to the coefficients[1] in expansion of $(2x-1)^{i-1}$. Thus, as outlined in Section 5 we have $\Lambda\mathbf{P}_{X^*} = \mathbf{P}_{\hat{X}}\Lambda$ with $\Lambda = \hat{\Lambda}(\Lambda^*)^{-1}$. We need only to check that $\nu^*\Lambda^{-1} = \nu^*\Lambda^*\hat{\Lambda}^{-1}$ is equal to $\hat{\nu}$ given in (6.3). We have

$$\begin{aligned} \hat{\nu}(j) &= (\nu^*\Lambda^*\hat{\Lambda}^{-1})(j) = \sum_k \Lambda^*(m,k)\hat{\Lambda}^{-1}(k,j) \\ &= \sum_{j\leq k\leq m} \frac{\binom{N-1}{k-1}}{\sum_{r=0}^{m-1}\binom{N-1}{r}}(-1)^{j-k}2^{j-1}\binom{k-1}{j-1} \\ &= \frac{2^{j-1}}{\sum_{r=0}^{m-1}\binom{N-1}{r}}\sum_{j\leq k\leq m}(-1)^{j-k}\binom{N-1}{k-1}\binom{k-1}{j-1} \\ &\overset{(*)}{=} \frac{2^{j-1}\binom{N-1}{j-1}}{\sum_{r=0}^{m-1}\binom{N-1}{r}}\sum_{j\leq k\leq m}(-1)^{j-k}\binom{N-j}{N-k}, \end{aligned}$$

where in $\overset{(*)}{=}$ we used the identity $\binom{N-1}{k-1}\binom{k-1}{j-1} = \binom{N-1}{j-1}\binom{N-j}{N-k}$. As for the last sum we have

$$\begin{aligned} \sum_{j\leq k\leq m}(-1)^{j-k}\binom{N-j}{N-k} &= \sum_{j\leq k\leq m}(-1)^{j-k}\binom{N-j}{k-j} \\ = \sum_{k=0}^{m-j}(-1)^k\binom{N-j}{k} &\overset{(**)}{=} (-1)^{m-j}\binom{N-j-1}{m-j}, \end{aligned}$$

[1]The on-line encyclopedia of integer sequences, sequence http://oeis.org/A303872.

where in $\stackrel{(**)}{=}$ we used the identity[2] $\sum_{k=0}^{M}(-1)^k\binom{n}{k} = (-1)^M\binom{n-1}{M}$. Finally,

$$\hat{\nu}(j) = \frac{2^{j-1}(-1)^{m-j}\binom{N-1}{j-1}\binom{N-j-1}{m-j}}{\sum_{r=0}^{m-j}\binom{N-1}{r}},$$

what is equal to (6.3).

Note that the pgf given in (6.3) corresponds to the distribution of $\sum_{k=j}^{N-1} Y_k$, where Y_k is a geometric random variable with parameter $\frac{k-1}{N-1}$ and $Y_1, \ldots, Y_{N-1}$ are independent. We have $EY_k = \frac{N-1}{N-k}$ thus (6.4) follows from (6.2) and (6.3).

□

Note that calculating $\hat{\nu}$ we have actually calculated the link Λ, which is given by

$$\Lambda(i,j) = \begin{cases} \frac{2^{j-1}(-1)^{i+j}(i-j+1)\binom{N-1}{i}\binom{i}{j-1}}{(N-j)\sum_{k=0}^{i-1}\binom{N-1}{k}} & \text{if } j < N, \\ 0 & j = N, i < N, \\ 1 & j = N, i = N. \end{cases}$$

Next two subsections 6.3 and 6.4 contain results for some non-trivial multidimensional gambler models.

6.3 A multidimensional case, winning probabilities and the absorption time: changing r coordinates at one step in a d-dimensional game

We will present an example for both Theorems, 2.1 and 2.3. The chains $\mathbf{P}_{Z_j^{(k)}}$ in Theorem 2.1 are quite general, but in this example we consider birth and death chains *i.e.,* we will use $\mathbf{P}_{Z_j^{(k)}} = \mathbf{P}_{X_j^*}$ from Theorem 2.3 (birth and death chains given in (1.4)). Similarly, we have $\mathbb{R}'_{Z_j^{(k)}} = \mathbb{R}'_{X_j^{*(k)}}$ and $\mathbf{P}_Z = \mathbf{P}_{X^*}$.

Example 6.3. The idea of the example is following. We construct a d-dimensional game from one-dimensional games, in such a way, that at one step we play with r other players, where $r \in \{1, \ldots, d\}$. In other words, the multidimensional chain can change at most r coordinates in one step.

Moreover we will take, as $\mathbf{B}_i := b_i$ real numbers. In both theorems let us take $0 < r < d$, $m = \binom{d}{r} + 1$ and $b_k = 1, k = 1, \ldots, m-1, b_m = 1 - \binom{d}{r}$. Let us enumerate combinations of r positive numbers no greater than d in some way (see e.g., [Mud65]). Let $\mathbf{A}_k$ be k-th combination from this numbering, for $k = 1, \ldots, m-1$ and $\mathbf{A}_m = \emptyset$. Then we have

$$\mathbf{P}'_Z = \sum_{k=1}^{m} \mathbf{B}_k \left(\bigotimes_{j \le d} \mathbb{R}'_{Z_j^{(k)}} \right) = \sum_{k=1}^{\binom{d}{r}} \left(\bigotimes_{j \le d} \mathbb{R}'_{Z_j^{(k)}} \right) - \left(\binom{d}{r} - 1 \right) \bigotimes_{j \le d} \mathbf{I}_j. \tag{6.5}$$

We have that $\mathbb{R}'_{Z_j^{(k)}} = \mathbf{P}'_{Z_j}$ if $\{j\} \in \mathbf{A}_k$ and $\mathbb{R}'_{Z_j^{(k)}} = \mathbf{I}_j$ otherwise (for $\{j\} \notin \mathbf{A}_k$), thus $\mathbf{P}'_Z =$

$$\sum_{1 \le i_1 < i_2 < \ldots < i_r \le d} (\bigotimes_{j < i_1} \mathbf{I}_j) \otimes \mathbf{P}'_{Z_{i_1}} \otimes (\bigotimes_{i_1 < j < i_2} \mathbf{I}_j) \otimes \ldots \otimes (\bigotimes_{i_{r-1} < j < i_r} \mathbf{I}_j) \otimes \mathbf{P}'_{Z_{i_r}} \otimes (\bigotimes_{i_r < j \le d} \mathbf{I}_j) - \left(\binom{d}{r} - 1 \right) \bigotimes_{j \le d} \mathbf{I}_j.$$

[2] See *Partial sums* at https://en.wikipedia.org/wiki/Binomial_coefficient

In other words, we combine d one-dimensional birth and death chains in such a way, that the resulting d-dimensional chain can change at most r coordinates by ± 1 at one step.
We can rewrite this formula for some cases:

- $r = d$, independent games
$$\mathbf{P}'_Z = \bigotimes_{j\le d} \mathbf{P}'_{Z_j}.$$

- $r = d-1$
$$\mathbf{P}'_Z = \sum_{k=1}^{d} (\bigotimes_{j<k} \mathbf{P}'_{Z_j}) \otimes \mathbf{I}_k \otimes (\bigotimes_{j>k} \mathbf{P}'_{Z_j}) - (d-1) \bigotimes_{j\ge d} \mathbf{I}_j.$$

- $r = 2$
$$\mathbf{P}'_Z = \sum_{k=1}^{d} \sum_{i=k+1}^{d} (\bigotimes_{j<k} \mathbf{I}_j) \otimes \mathbf{P}'_{Z_k} \otimes (\bigotimes_{k<j<i} \mathbf{I}_j) \otimes \mathbf{P}'_{Z_i} \otimes (\bigotimes_{i<j\le d} \mathbf{I}_j) - \left(\binom{d}{2} - 1\right) \bigotimes_{j\ge d} \mathbf{I}_j.$$

- $r = 1$
$$\mathbf{P}'_Z = \sum_{k=1}^{d} (\bigotimes_{j<k} \mathbf{I}_j) \otimes \mathbf{P}'_{Z_k} \otimes (\bigotimes_{j>k} \mathbf{I}_j) - (d-1) \bigotimes_{j\ge d} \mathbf{I}_j.$$
This can be rewritten as
$$\mathbf{P}'_Z = \bigoplus_{j\le d} \mathbf{P}'_{Z_j} - (d-1) \bigotimes_{j\le d} \mathbf{I}_j.$$
$\mathbf{P}_Z = \mathcal{F}_{-\infty}(\mathbf{P}'_Z)$ are exactly the transition corresponding to the generalized gambler's ruin problem given in (1.1).

In all above cases, the winning probability for the chain $\mathbf{P}_Z$ is given in (1.3). This is since the winning probabilities for $\mathbf{P}_{Z_j}$ are given in (2.4), thus using (2.3) the relation is (1.3) proven.

In all above cases, if we replace $\mathbf{P}'_{Z_j}$ with $\mathbf{P}_{\hat{X}_j}$ and $\mathbf{P}'_Z$ with $\mathbf{P}_{\hat{X}}$, then we have a special cases for formula for $\mathbf{P}_{\hat{X}}$ given in (2.6). If, in addition, we assume that $\nu^*((1,\ldots,1)) = 1$, then from Corollary 2.4 c) we have

$$T^*_{(1,\ldots,1),\mathbf{N}} = \begin{cases} \hat{T}_{(1,\ldots,1),\mathbf{N}} & \text{with probability } \prod_{j=1}^{d} \rho_j(1), \\ +\infty & \text{with probability } 1 - \prod_{j=1}^{d} \rho_j(1). \end{cases}$$

For example, in case $r = 1$ (then we have $m = d+1$ and take $b_k = 1, k = 1,\ldots,d, b_{d+1} = 1-d, \mathbf{A}_k = \{k\}, k = 1,\ldots,d$ and $\mathbf{A}_{d+1} = \emptyset$) we have

$$\mathbf{P}_{\hat{X}}((i_1,\ldots,i_d),(i'_1,\ldots,i'_d)) = \begin{cases} 1-\lambda^{(j)}_{i_j} & \text{if } i'_j = i_j + 1, \\ 1 - \sum_{j:i_j<N_j} \left(1-\lambda^{(j)}_{i_j}\right) & \text{if } i'_j = i_j, j = 1,\ldots,d, \\ 0 & \text{otherwise.} \end{cases}$$

Sample transitions for case $d = 2, r = 1$ are depicted in Fig. 1.

In Figure 3 the transitions of $\hat{X}$ are presented for $d = 3$:

- When $r = 1$ only transitions along blue dotted arrows (•••►) are possible.

- When $r = 2$ only transitions along blue dotted arrows (•••••►) and green dashed arrows (- - ►) are possible.
- When $r = 3$ all transitions, along blue arrows (•••••►), green dashed arrows (- - ►) and red curly arrow (〰►) are possible.

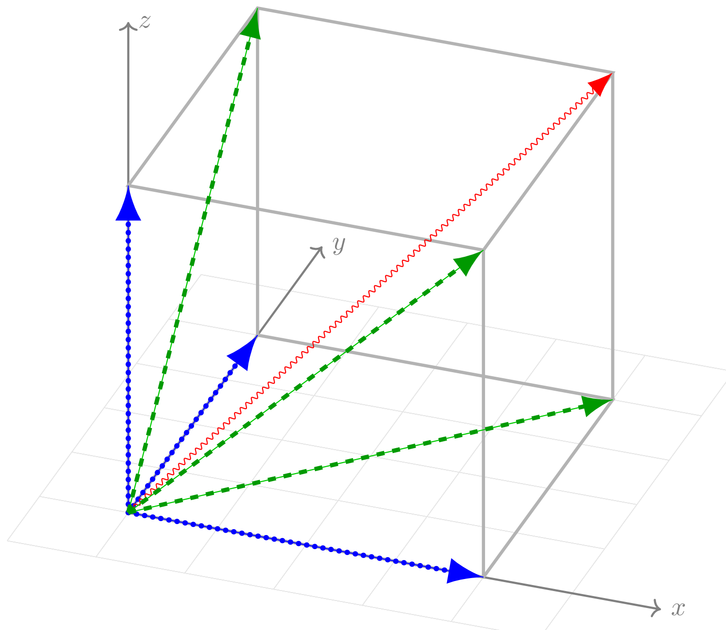


Figure 3: Sample transitions of $\hat{X}$ for the example from Section 6.3 with $d = 3$: transitions for $r = 1$ (blue •••••►), $r = 2$ (blue •••••►and green - - ►) and $r = 3$ (blue •••••►, green - - ► and red 〰►).

6.4 A multidimensional case, winning probabilities: changing r (dependent on the current fortune) coordinates at one step in a d-dimensional game

Now we will make use of the possibility that $\mathbf{B}_k$'s appearing in Theorem 2.1 can be matrices. We will provide an extension of the previous Example 6.3 – this time we will only provide the result on the winning probabilities (since the result on the absorption time provided in Theorem 2.3 requires $\mathbf{B}_k$'s to be numbers).

Example 6.4. Before providing intricate details of the example, let us clarify what we aim to achieve. In the previous Example 6.3 we showed how one may construct a multidimensional chain, so that at most $r \leq d$ coordinates can be changed in one step (in other words, we can play with at most $r \leq d$ players in one step). Now we extend this situation: we want r to be state-dependent. To be more exact: if we are in a state from a set $\mathbf{S}_r$ then we may play with at most r players. For example setting:

$$\mathbf{S}_d = \{(i_1,\ldots,i_d) : \sum_{j=1}^{d} i_j \geq 100\}, \quad \mathbf{S}_2 = \{(i_1,\ldots,i_d) : \sum_{j=1}^{d} i_j < 100\}$$

and $\mathbf{S}_r = \emptyset$ for $r \notin \{2, d\}$ models a game in which:

- If we have at least 100 dollars we may play with all the players;
- If we have less than 100 dollars we may play with at most 2 players.

Consider a partition of the state space into disjunctive sets $\mathbf{S}_r, r = 0,\ldots,d$, *i.e.*, $\mathbb{E} = \bigcup_{r=0}^{d} \mathbf{S}_r$. Each $\mathbf{S}_r$ is a set of states from which we can change at most r coordinates in one step. Let $\mathbf{I}_{\mathbf{S}_r}$ be a matrix with ones only on positions (i,i), where $i \in \mathbf{S}_r$. Let $m_r = \binom{d}{r} + 1$ and set $\mathbf{B}_i^r = \mathbf{I}_{\mathbf{S}_r}$ for $1 \leq i < m_r$ and $\mathbf{B}_{m_r}^r = \left(1 - \binom{d}{r}\right) \mathbf{I}_{\mathbf{S}_r}$. Let us enumerate combinations of r positive numbers no greater than d in some way (again, see *e.g.*, [Mud65]). Let $\mathbf{A}_i^r$ be i-th combination from this numbering, for $i = 1,\ldots,m_r - 1$ and $\mathbf{A}_{m_r}^r = \emptyset$.

Now we will renumerate everything to fit to the notation used in Theorem 2.1. For $1 \leq k \leq m$ there exists $1 \leq b \leq d, 1 \leq i \leq m_{b+1}$ such that $k = \left(\sum_{r=1}^{b-1} m_r\right) + i$, we then set $\mathbf{B}_k = \mathbf{B}_{\left(\sum_{r=1}^{b-1} m_r\right)+i} = \mathbf{B}_i^b$ and $\mathbf{A}_k = \mathbf{A}_{\left(\sum_{r=1}^{b-1} m_r\right)+i} = \mathbf{A}_i^b$. In this notation – using (2.2) – the transition matrix of the resulting chain is given by:

$$\mathbf{P}'_Z = \sum_{k=1}^{m} \mathbf{B}_k \left(\bigotimes_{j \leq d} \mathbb{R}'_{Z_j^{(k)}} \right).$$

However, we can rewrite it in a more intuitive way, letting $k_b = \sum_{r=1}^{b} m_r$ we have:

$$\begin{aligned}
\mathbf{P}'_Z &= \sum_{b=1}^{d} \sum_{i=1}^{m_b} \mathbf{B}_{k_{b-1}+i} \left(\bigotimes_{j \leq d} \mathbb{R}'_{Z_j^{(k_{b-1}+i)}} \right) \\
&= \sum_{b=1}^{d} \left\{ \sum_{i=1}^{m_b} \mathbf{I}_{\mathbf{S}_b} \left(\bigotimes_{j \leq d} \mathbb{R}'_{Z_j^{(k_{b-1}+i)}} \right) - \mathbf{I}_{\mathbf{S}_b} \left(\binom{d}{b} - 1 \right) \bigotimes_{j \leq d} \mathbf{I}_j \right\} \\
&= \sum_{b=1}^{d} \mathbf{I}_{\mathbf{S}_b} \left\{ \sum_{i=1}^{m_b} \left(\bigotimes_{j \leq d} \mathbb{R}'_{Z_j^{(k_{b-1}+i)}} \right) - \left(\binom{d}{b} - 1 \right) \bigotimes_{j \leq d} \mathbf{I}_j \right\}.
\end{aligned}$$

Thus, $\mathbf{P}'_Z$ can be rewritten as $\mathbf{P}'_Z = \sum_{b=1}^{d} \mathbf{I}_{S_b} \mathbf{P}'^{b}_Z$, where matrices $\mathbf{P}'^{b}_Z$ correspond to the previous Example 6.3, cf. (6.5). Concluding, Theorem 2.1 implies that the winning probability of the chain $\mathbf{Z}$ with the transition matrix $\mathbf{P}_Z = \mathcal{F}_{-\infty}(\mathbf{P}'_Z)$ is of the product form provided in (1.3).

Chapter 3. Conditional gambler's ruin problem with arbitrary winning and losing probabilities with applications

1 Introduction

The classical gambler's ruin problem is following. Having initially i dollars, $1 \leq i \leq N-1$, in one step we either win one dollar (*i.e.,* we move to $i+1$) with probability $p \in (0,1)$, or we lose one dollar (*i.e.,* we move to $i-1$) with probability $q = 1-p$. The game ends when the player reaches N (wins the game) or 0 (goes broke). The typical questions one can ask are:

- What is the probability of winning (*i.e.,* reaching N before 0)?
- What is the (expected) game duration?
- What is the (expected) conditional game duration (*i.e.,* game duration given we win or given we lose)?
- Is the (expected) conditional game duration symmetric in p and q?

Similarly, one can consider random walk on $\mathbf{Z}_{m+1} = \{0, \ldots, m\}$: being at state i we either move clockwise with a probability $p \in (0,1)$ (*i.e.,* from i to $i+1$ mod $(m+1)$) or we move counterclockwise with a probability $1-p$ (*i.e.,* we move from i to $i-1$ mod $(m+1)$). We will refer to this as to *the classical random walk on a polygon* (cf. [Sar06]). Assuming we start at i, the typical questions one can ask are:

- What is the probability that all vertices have been visited before the particle returns to i?
- What is the probability that the last vertex visited is j ?
- What is the expected number of moves needed to visit all the vertices?
- What is the expected additional number of moves needed to return to i after visiting all the vertices?

All above questions were answered in the classical settings. Several generalizations were studied. The probability of winning in a gambler's ruin problem with general winning and losing probabilities (*i.e.,* $p(i)$ being probability of moving from i to $i+1$ and $q(i)$ being the probability of moving from i to $i-1$, with $p(i)+q(i) \leq 1$, $i \in \{1, \ldots, N-1\}$) goes back to Parzen [Par62], revisited in [ES09]. Siegmund duality based proof is given in [Lor17] (where more general, multidimensional, game is considered). In [Len09b] the questions related to the conditional game duration are answered for the classical gambler's ruin problem with ties allowed, *i.e.,* $p+q \leq 1$

(with probability $1-(p+q)$ we can stay at a given state). In [Lef08] author considers specific generalization, namely $p(i)=q(i)=\frac{1}{2(2ci+1)}, c\geq 0$ (thus the probability of staying is $1-\frac{1}{2ci+1}$) and answers the question about the winning probability and the expected game duration (and also considers the corresponding diffusion process). In this chapter we present formulas for the expected (conditional) absorption time in terms of parameters of the system (*i.e.,* winning/losing probabilities $p(i), q(i)$). Similar problem was considered in [ES00], the recursion for the expected conditional game duration is given therein (equations (3.4) and (3.5)), however it is not solved in its general form – later on author considers only constant winning/losing probabilities. In [GMZ12] (similar results with different proofs are presented in [MZ17]) the generating function of absorption time (including a conditional one) is given in terms of eigenvalues of a transition matrix and eigenvalues of a truncated transition matrix. The questions for the classical random walk on a polygon were answered in [Sar06]. Some generalizations (rather then allowing arbitrary winning/losing probabilities, symmetric random walks on tetrahedra, octahedra, and hexahedra, are considered) are studied in [SM17].

In 1977 in [BW77] it was shown that for a classical gambler's ruin problem with $p(n)=p=1-q(n)=1-q$, the distribution of a conditional game duration is symmetric in p and q, *i.e.,* it is the same as in a game with $p'=q$ and $q'=p$. In 2009 in [Len09b] it was extended to a case $p+q<1$ (*i.e.,* the classical case with ties allowed). In this chapter we show that that the expected conditional game duration is symmetric also for non-constant winning/losing probabilities $p(n), q(n)$ as long as $q(n)/p(n)$ is constant (thus, including for example the spatially non-homogeneous case).

In Section 2 we introduce gambler's ruin problem with arbitrary winning and losing probabilities $p(i), q(i)$ together with main results. In Section 2.1 the main result is applied to constant $r(i)=r=q(i)/p(i)$, in Section 2.2 it is applied to non-homogeneous case, whereas the classical case is recalled in Section 2.3. The main example is given in Section 2.4. The results are applied to a random walk on polygon in Section 4. Last Section 5 contains proofs of main results.

2 Gambler's ruin problem

Fix an integer $N\geq 2$. Let

$$\mathbf{p}=(p(0),p(1),\dots,p(N)),\quad \mathbf{q}=(q(0),q(1),\dots,q(N)),$$

where $p(0)=q(0)=p(N)=q(N)=0$ and $p(i),q(i)>0, p(i)+q(i)\leq 1$ for $i\in\{1,2\dots,N-1\}$. Consider a Markov chain $\mathbf{X}=\{X_k\}_{k\geq 0}$ on $\mathbb{E}=\{0,1,\dots,N\}$ with transition probabilities

$$\mathbf{P}_X(i,j)=\begin{cases} p(i) & \text{if} \quad j=i+1,\\ q(i) & \text{if} \quad j=i-1,\\ 1-(p(i)+q(i)) & \text{if} \quad j=i.\end{cases}$$

We will refer to $\mathbf{X}$ starting at i as to the (gambler's ruin) *game* $G(\mathbf{p},\mathbf{q},0,i,N)$. Note that the chain will eventually end up in either in N (the *winning* state) or in 0 (the *losing* state). To simplify some notation, let $r(i)=\frac{q(i)}{p(i)}$ for $i\in\{1,\dots,N-1\}$.

Define $\tau_j=\inf\{k: X_k=j\}$. We will study the following *smaller* games $G(\mathbf{p},\mathbf{q},j,i,k)$ with k as the *winning* state and j as the *losing* state ($j\leq i\leq k$), *i.e.,* $p(j)=q(j)=p(k)=q(k)=0$.

Let us define:

$$\begin{aligned}\rho_{j:i:k} &= P(\tau_k < \tau_j | X_0 = i),\\ T_{j:i:k} &= \inf\{n \geq 0 : X_n = j \text{ or } X_n = k,\ X_0 = i\},\\ W_{j:i:k} &= T_{j:i:k} \text{ conditioned on } X_{T_{j:i:k}} = k,\\ B_{j:i:k} &= T_{j:i:k} \text{ conditioned on } X_{T_{j:i:k}} = j.\end{aligned}$$

In other words: $\rho_{j:i:k}$ is the probability that a gambler starting with i dollars wins in the smaller game; $T_{j:i:k}$ is the distribution of a game duration (time till gambler either wins or goes broke); $W_{j:i:k}$ is the distribution of $T_{j:i:k}$ conditioned on $X_{T_{j:i:k}} = k$ (winning) and similarly $B_{j:i:k}$ is the distribution of $T_{j:i:k}$ conditioned on $X_{T_{j:i:k}} = j$ (losing).

Notation. For given rates $\mathbf{p}, \mathbf{q}$ by $\mathbf{p} \leftrightarrow \mathbf{q}$ we understand new rates $\mathbf{p}' = \mathbf{q}, \mathbf{q}' = \mathbf{p}$. For some random variable R (one of ρ, T, W, B) for a game with rates $\mathbf{p}, \mathbf{q}$, by $R(\mathbf{p} \leftrightarrow \mathbf{q})$ we understand the random variable defined for a game with rates $\mathbf{p}' = \mathbf{q}, \mathbf{q}' = \mathbf{p}$ (and similarly, e.g., $ER(\mathbf{p} \leftrightarrow \mathbf{q})$ is an expectation of R defined for such a game). We say that R (ER) is *symmetric in* $\mathbf{p}$ and $\mathbf{q}$ if $R \stackrel{distr}{=} R(\mathbf{p} \leftrightarrow \mathbf{q})$ ($ER = ER(\mathbf{p} \leftrightarrow \mathbf{q})$).

By $f(n) = \Theta(g(n))$ we mean $\exists(c_1, c_2 > 0)\ \exists(n_0)\ \forall(n > n_0)\ c_1 g(n) \leq f(n) \leq c_2 g(n)$. In this section we use the convention: empty sum equals 0, empty product equals 1; however in Section 4 we use some nonstandard notation, see details on page 72.

In next theorem we provide formulas for expected game duration, for completeness (and since we will need them later) we also include known results for $\rho_{j:i:k}$.

Theorem 2.1. *Consider the gambler's ruin problem on $\mathbb{E} = \{0, 1, \ldots, N\}$ described above. We have*

$$\rho_{j:i:k} = \frac{\displaystyle\sum_{n=j+1}^{i} \prod_{s=j+1}^{n-1} \left(\frac{q(s)}{p(s)}\right)}{\displaystyle\sum_{n=j+1}^{k} \prod_{s=j+1}^{n-1} \left(\frac{q(s)}{p(s)}\right)} = \frac{\displaystyle\sum_{n=j+1}^{i} \prod_{s=j+1}^{n-1} r(s)}{\displaystyle\sum_{n=j+1}^{k} \prod_{s=j+1}^{n-1} r(s)},$$

$$ET_{j:i:k} = \frac{\sum_{n=j+1}^{k-1} [d_n \sum_{s=j+1}^{n} \frac{1}{p(s) d_s}]}{\sum_{n=j}^{k-1} d_n} \sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1} \left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s) d_s} \right], \tag{2.1}$$

where $d_s = \prod_{i=j+1}^{s} \frac{q(i)}{p(i)} = \prod_{i=j+1}^{s} r(i)$ (with convention $d_j = 1$).

The proof of Theorem 2.1 is postponed to Section 5.1.1. We will also need a formula for $ET_{j:i:k}$ in case when k is the only absorbing state.

Theorem 2.2. *Fix $j \leq i \leq k$ and consider a birth and death chain on $\{j, \ldots, k\}$ with rates $p(s), q(s), s = j, \ldots, k$ with $q(j) = p(k) = q(k) = 0$ and $q(s) > 0$ for $s = j+1, \ldots, k-1$ and $p(s) > 0$ for $s = j, \ldots, k-1$ (i.e., k is the only absorbing state). Then, the expectation of absorption time, starting from i is given by*

$$ET_{j:i:k} = \sum_{n=i}^{k-1} \left[d_n \sum_{s=j}^{n} \frac{1}{p(s) d_s} \right].$$

Now we go back to situation with two absrobing states, *i.e.,* also $p(j) = 0$. Next theorem (our main contribution) gives the formulas for $EW_{0:i:k}$ and $EB_{0:i:k}$. First, let us introduce some necessary notation. With some abuse of notation let us extend

$$\rho_{j:i:k} = \frac{\sum_{n=j+1}^{i} \prod_{s=j+1}^{n-1} \left(\frac{q(s)}{p(s)}\right)}{\sum_{n=j+1}^{k} \prod_{s=j+1}^{n-1} \left(\frac{q(s)}{p(s)}\right)} = \frac{\sum_{n=j+1}^{i} \prod_{s=j+1}^{n-1} r(s)}{\sum_{n=j+1}^{k} \prod_{s=j+1}^{n-1} r(s)}$$

for $k < i$ (but still $k > j$). Note that in such a case we may have $\rho_{j:i:k}$, thus this has no interpretation in terms of probability anymore.

For given integers n, m, k such that $n \leq m, k \in \{0, \lfloor (m-n+1)/2 \rfloor\}$ define

$$\mathbf{j}_k^{n,m} = \big\{\{j_1, j_2, \ldots, j_k\} : j_1 \geq n+1, j_k \leq m, j_i \leq j_{i+1} - 2 \text{ for } 1 \leq i \leq k-1\big\}. \tag{2.2}$$

For given $\mathbf{p}, \mathbf{q}$ and $\mathbf{j} \in \mathbf{j}_k^{n,m}$ define

$$\delta_{\mathbf{j}}^{n,m} = (-1)^k \prod_{s \in j} r(s) \prod_{s \in \{n,\ldots,m\} \setminus \mathbf{j} \cup (\mathbf{j}-1)} 1 + r(s), \tag{2.3}$$

where $\{n, \ldots, m\}$ is an empty set for $n > m$ and $\mathbf{j} - 1 = \{j_1 - 1, j_2 - 1, \ldots, j_k - 1\}$ for $\mathbf{j} = \{j_1, j_2, \ldots, j_k\}$. Finally, let

$$\xi_k^{n,m} = \sum_{\mathbf{j} \in \mathbf{j}_k^{n,m}} \delta_{\mathbf{j}}^{n,m}. \tag{2.4}$$

Now we are ready to state our main theorem.

Theorem 2.3. *Consider the gambler's ruin problem on $\mathbb{E} = \{0, 1, \ldots, N\}$ described above. We have*

$$EW_{0:i:N} = EW_{0:1:N} - EW_{0:1:i}, \quad \textit{where} \tag{2.5}$$

$$EW_{0:1:i} = \sum_{n=1}^{i-1} \frac{\rho_{0:n:i}}{p(n)} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \xi_s^{n+1,i-1}. \tag{2.6}$$

Moreover, we have

$$EB_{0:i:N} = EW'_{0:N-i:N}, \tag{2.7}$$

where $W'_{0:N-i:N}$ is defined for a gambler's ruin problem with rates $p'(i) = q(N-i)$ and $q'(i) = p(N-i)$ for $i \in \mathbb{E}$.

The proof of Theorem 2.3 is postponed to Section 5.1.2.

2.1 Constant $r(n) = r = \frac{q(n)}{p(n)}$

In this section we will apply Theorems 2.1 and 2.3 to a gambler's ruin problem with constant $r = \frac{q(i)}{p(i)}$. The winning probabilities $\rho_{0:i:N}$ are known (they are the same as in the classical formulation of the problem), we will focus on a game duration. We have

Corollary 1. Consider the gambler's ruin problem on $\mathbb{E} = \{0, \ldots, N\}$ with constant $r = \frac{q(i)}{p(i)}$. We have

$$
\begin{aligned}
r = 1: \quad ET_{j:i:k} &= \frac{i-j}{k-j} \sum_{n=j+1}^{k-1} \sum_{s=j+1}^{n} \frac{1}{p(s)} - \sum_{n=j+1}^{i-1} \sum_{s=j+1}^{n} \frac{1}{p(s)}, \\
ET_{0:i:N} &= \frac{i}{N} \sum_{n=1}^{N-1} \sum_{s=1}^{n} \frac{1}{p(s)} - \sum_{n=1}^{i-1} \sum_{s=1}^{n} \frac{1}{p(s)}, \\
r \neq 1: \quad ET_{j:i:k} &= \frac{r^j - r^i}{r^j - r^k} \sum_{n=j+1}^{k-1} \left[r^n \sum_{s=j+1}^{n} \frac{r^{-s}}{p(s)} \right] - \sum_{n=j+1}^{i-1} \left[r^n \sum_{s=j+1}^{n} \frac{r^{-s}}{p(s)} \right], \\
ET_{0:i:N} &= \frac{1 - r^i}{1 - r^N} \sum_{n=1}^{N-1} \left[r^n \sum_{s=1}^{n} \frac{r^{-s}}{p(s)} \right] - \sum_{n=1}^{i-1} \left[r^n \sum_{s=1}^{n} \frac{r^{-s}}{p(s)} \right].
\end{aligned}
$$

Proof. We have $d_k = \prod_{j=1}^{k} r = r^k$. Simple recalculations of (2.1) yield the result. □

For constant r we have that $\delta_{\mathbf{j}}^{n,m}$ (given in (2.3)) for all $i \in \{1, \ldots, N-1\}$ depends on $\mathbf{j}$ only through k, thus

$$\xi_k^{n,m} = \sum_{\mathbf{j} \in \mathbf{j}_k^{n,m}} \delta_{\mathbf{j}}^{n,m} = C_k^{n,m} (-r)^k (1+r)^{m+1-n-2k}, \tag{2.8}$$

where $C_k^{n,m} = |\mathbf{j}_k^{n,m}|$. Moreover, we have $|\mathbf{j}_k^{n,m}| = T(m+1-n, k)$, where $T(n,k) = \binom{n-k}{k}$ is the number of subsets of $\{1, 2, ..., n-1\}$ of size k containing no consecutive integers [1].

The proof of the next corollary requires the following lemma.

Lemma 2.4. Let $n \in \mathbb{N}$ and $r \geq 0$. We have

$$\sum_{k=0}^{n} \binom{n-k}{k} \left(-\frac{r}{(1+r)^2} \right)^k = \begin{cases} \dfrac{1 - r^{n+1}}{(1+r)^n (1-r)} & \text{if } r \neq 1, \\[2ex] \dfrac{n+1}{2^n} & \text{if } r = 1. \end{cases} \tag{2.9}$$

The proof of Lemma 2.4 is given in Section 5.1.2.

Remark 2.1. Note that the assertion of Lemma 2.4 can be stated in the following form (simply substituting $c = \frac{r}{(1+r)^2}$): for $n \in \mathbb{N}$ and $c \in (0, 1/4]$ we have

$$\sum_{k=0}^{n} \binom{n-k}{k} (-c)^k = \begin{cases} \dfrac{1 - \gamma^{n+1}}{(1+\gamma)^n (1-\gamma)}, & \text{where } \ \gamma = \dfrac{1 - 2c + \sqrt{1-4c}}{2c}, \text{ if } c \in (0, 1/4), \\[2ex] \dfrac{n+1}{2^n} & \text{if } c = 1/4. \end{cases}$$

[1] http://oeis.org/A011973

These sums for $c \in \{-1, 1\}$ were known ($F(n)$ is the n-th Fibonacci number):

$$\sum_{k=0}^{n} \binom{n-k}{k} = F(n+1),$$

$$\sum_{k=0}^{n} \binom{n-k}{k} (-1)^k = \begin{cases} 1 & \text{if } n \bmod 6 \in \{0,1\}, \\ 0 & \text{if } n \bmod 6 \in \{2,5\}, \\ -1 & \text{if } n \bmod 6 \in \{3,4\}. \end{cases}$$

We will give formulas for $EW_{0:1:i}$ for several cases ($EW_{0:i:N}$ can be calculated via (2.5)).

Corollary 2. Consider the gambler's ruin problem on $\mathbb{E} = \{0, \ldots, N\}$ with constant $r = \frac{q(i)}{p(i)}$. We have:

$$r = 1: \quad EW_{0:1:i} = \sum_{n=1}^{i-1} \frac{\rho_{0:n:i}}{p(n)} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \xi_s^{n+1,i-1} = \sum_{n=1}^{i-1} \frac{n/i}{p(n)}(i-n).$$

$$r \neq 1: \quad EW_{0:1:i} = \sum_{n=1}^{i-1} \frac{\rho_{0:n:i}}{p(n)} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \xi_s^{n+1,i-1} = \sum_{n=1}^{i-1} \frac{\frac{1-r^n}{1-r^i}(1-r^{i-n})}{p(n)(1-r)}. \tag{2.10}$$

Additionally, if $p(n) = p$ is constant (so is $q(n)$ then, since $r(n)$ is constant) we have

$$r = 1: \quad EW_{0:1:i} = \frac{1}{p} \sum_{n=1}^{i-1} \frac{n}{i}(i-n) = \frac{(i-1)(i+1)}{6p}, \tag{2.11}$$

$$\begin{aligned} r \neq 1: \quad EW_{0:1:i} &= \frac{1}{p} \sum_{n=1}^{i-1} \frac{\frac{1-r^n}{1-r^i}(1-r^{i-n})}{1-r} = \frac{1}{p(1-r^i)(1-r)} \sum_{n=1}^{i-1} (1-r^n)(1-r^{i-n}) \\ &= \frac{i(1+r^i) - (1+r)\frac{1-r^i}{1-r}}{p(1-r^i)(1-r)} = \frac{1}{p(1-r)} \left(i \frac{1+r^i}{1-r^i} - \frac{1+r}{1-r} \right). \end{aligned}$$

Proof. We will only show case $r = 1$, general $p(n)$ (the proof for $r \neq 1$ is very similar). Let us calculate $\xi_s^{n+1,i-1}$ first. From (2.8) and form of $C_k^{n,m}$ for $r = 1$ we have

$$\xi_s^{n+1,i-1} = C_s^{n+1,i-1}(-1)^s 2^{i-n-1-2s} = 2^{i-n-1} \binom{i-n-1-s}{s} \left(-\frac{1}{4}\right)^s.$$

From Theorem 2.3 (eq. (2.6)) and the fact that $\rho_{0:n:i} = n/i$ (since $r = 1$) we have

$$\begin{aligned} EW_{0:1:i} &= \sum_{n=1}^{i-1} \frac{n/i}{p(n)} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \xi_s^{n+1,i-1} \\ &= \sum_{n=1}^{i-1} \frac{n/i}{p(n)} 2^{i-n-1} \sum_{s=0}^{\lfloor (i-1-n)/2 \rfloor} \binom{i-n-1-s}{s} \left(-\frac{1}{4}\right)^s \\ &\overset{\text{Lemma 2.4}}{=} \sum_{n=1}^{i-1} \frac{n/i}{p(n)} 2^{i-n-1} \frac{i-n-1+1}{2^{i-n-1}} = \sum_{n=1}^{i-1} \frac{n/i}{p(n)}(i-n), \end{aligned}$$

what finishes the proof. □

In 1977 Beyer and Waterman [BW77] showed that for a classical case *i.e.,* for constant birth $p(n) = p$ and death $q(n) = q$ rates such that $p + q = 1$, the distribution of $W_{0:i:N}$ is symmetric in p and q (*i.e.,* it has the same distribution for birth rate $p' = q$ and death rate $q' = p$). In 2009 Lengyel [Len09b] showed that this holds also for the classical case with ties allowed, *i.e.,* $p + q < 1$. In the following theorem we show that $EW_{0:i:N}$ is symmetric in $\mathbf{p}$ and $\mathbf{q}$ (*i.e.,* it is the same for case with birth deaths $p'(n) = q(n)$ and death rates $q'(n) = p(n)$) as long as $r(n) = \frac{q(n)}{p(n)}$ is constant.

Theorem 2.5. *Consider the gambler's ruin problem on* $\mathbb{E} = \{0, \ldots, N\}$ *with constant* $r = \frac{q(i)}{p(i)}$. *We have*

$$EW_{0:i:N} = EW_{0:i:N}(\mathbf{p} \leftrightarrow \mathbf{q}),$$

(i.e., $EW_{0:i:N}$ *is symmetric in* $\mathbf{p}$ *and* $\mathbf{q}$*).*

Proof. By (2.5) it is enough to show that $EW_{0:1:i} = EW_{0:1:i}(\mathbf{p} \leftrightarrow \mathbf{q})$.

Let $W_{0:1:i}$ be defined for rates $\mathbf{p}$ and $\mathbf{q}$, whereas $W'_{0:1:i}$ be defined for rates $\mathbf{p}' = \mathbf{q}$ and $\mathbf{q}' = \mathbf{p}$, thus $r' = 1/r$. Since $r = \frac{q(n)}{p(n)}$, we have $p'(n) = q(n) = rp(n)$.

$$\begin{aligned} EW'_{0:1:i} &= \sum_{n=1}^{i-1} \frac{1}{p'(n)} \frac{(1-\frac{1}{r^n})}{(1-\frac{1}{r^i})} \frac{(1-\frac{1}{r^{i-n}})}{(1-\frac{1}{r})} = \sum_{n=1}^{i-1} \frac{1}{rp(n)} \frac{r^i(1-r^n)}{r^n(1-r^i)} \frac{r(1-r^{i-n})}{r^{i-n}(1-r)} \\ &= \sum_{n=1}^{i-1} \frac{1}{p(n)} \frac{(1-r^n)}{(1-r^i)} \frac{(1-r^{i-n})}{(1-r)}, \end{aligned}$$

what is equal to (2.10). □

It is natural to state the following conjecture.

Conjecture 2.6. Consider the gambler's ruin problem on $\mathbb{E} = \{0, \ldots, N\}$ with constant $r = \frac{q(i)}{p(i)}$. Then, the distribution of $W_{0:i:N}$ is symmetric in $\mathbf{p}$ and $\mathbf{q}$.

2.2 The spatially non-homogeneous case

In this Section we consider gambler's ruin problem with birth rates $p(n) = \frac{p}{2cn+1}$ and death rates $q(n) = \frac{q}{2cn+1}$, where c is a non-negative constant. This is often called the spatially non-homogeneous gambler's ruin problem. We will thus still consider case with constant $r(n)$, but with specific rates. As far as we are aware, all results in this section, except the one for $p(n) = q(n) = 1/2$, are new.

Corollary 3. Consider the spatially non-homogeneous gambler's ruin problem. We have

$$\begin{aligned} r = 1 : ET_{0:i:N} &= \frac{1}{2p}\left(iN\left(1+\frac{2c}{3}N\right) - i^2\left(1+\frac{2c}{3}i\right)\right), \\ r \neq 1 : ET_{0:i:N} &= \frac{1}{p(r-1)}\left(\frac{1-r^i}{1-r^N}\left(-cN^2 - N\frac{(cr+c)}{r-1} - N\right) + ci^2 + i\frac{(cr+c)}{r-1} + i\right). \end{aligned}$$

Proof. Applying Corollary 1 we have:

- Case $r = 1$

$$
\begin{aligned}
ET_{0:i:N} &= \frac{i}{N}\sum_{n=1}^{N-1}\sum_{s=1}^{n}\frac{1}{p(s)} - \sum_{n=1}^{i-1}\sum_{s=1}^{n}\frac{1}{p(s)} = \frac{i}{N}\sum_{n=1}^{N-1}\sum_{s=1}^{n}\frac{2cn+1}{p} - \sum_{n=1}^{i-1}\sum_{s=1}^{n}\frac{2cn+1}{p} \\
&= \frac{1}{p}\left(\frac{i}{N}\sum_{n=1}^{N-1} n(cn+c+1) - \sum_{n=1}^{i-1} n(cn+c+1)\right) \\
&= \frac{1}{p}\left(\frac{i}{N}\frac{1}{6}(N-1)(N(2c(N+1)+3) - \frac{1}{6}(i-1)(i(2c(i+1)+3)\right) \\
&= \frac{1}{2p}\left(iN\left(1+\frac{2c}{3}N\right) - i^2\left(1+\frac{2c}{3}i\right)\right).
\end{aligned}
$$

- Case $r \neq 1$

$$
\begin{aligned}
ET_{0:i:N} &= \frac{1-r^i}{1-r^N}\sum_{n=1}^{N-1}\left[r^n\sum_{s=1}^{n}\frac{r^{-s}}{p(s)}\right] - \sum_{n=1}^{i-1}\left[r^n\sum_{s=1}^{n}\frac{r^{-s}}{p(s)}\right] \\
&= \frac{1}{p}\left(\frac{1-r^i}{1-r^N}\sum_{n=1}^{N-1}\left[r^n\sum_{s=1}^{n}r^{-s}(2cs+1)\right] - \sum_{n=1}^{i-1}\left[r^n\sum_{s=1}^{n}r^{-s}(2cs+1)\right]\right).
\end{aligned}
$$

We have

$$
\sum_{s=1}^{n} r^{-s}(2cs+1) = \frac{r^{-n}}{(r-1)^2}\left(2cr^{n+1} - 2cnr + 2cn - 2cr + r^{n+1} - r^n - r + 1\right)
$$

and

$$
\begin{aligned}
&\sum_{n=1}^{k-1}\left[r^n\frac{r^{-n}}{(r-1)^2}\left(2cr^{n+1} - 2cnr + 2cn - 2cr + r^{n+1} - r^n - r + 1\right)\right] \\
= &\frac{1}{(r-1)^2}\Bigg(-\tfrac{2cr(r-r^k)}{r-1} - c(k-1)kr + c(k-1)k - 2cr(k-1) \\
&\qquad -\frac{r(r-r^k)}{r-1} + \frac{r-r^k}{r-1} + r - kr + k - 1\Bigg) \\
= &\frac{1}{(r-1)^2}\left(-ck^2(r-1) + \frac{(2cr+r-1)(r^k-1)}{r-1} - k(cr+c+r-1)\right) .
\end{aligned}
$$

Thus,

$$\begin{aligned} ET_{0:i:N} &= \frac{1}{p(r-1)^2}\Big\{ \quad \frac{1-r^i}{1-r^N}\Big(\quad -cN^2(r-1) + \tfrac{(2cr+r-1)(r^N-1)}{r-1} - N(cr+c+r-1)\Big) \\ &\qquad - \Big(-ci^2(r-1) + \tfrac{(2cr+r-1)(r^i-1)}{r-1} - i(cr+c+r-1)\Big)\Big\} \\ &= \frac{1}{p(r-1)^2}\Big\{ \quad \frac{1-r^i}{1-r^N}\Big(\quad -cN^2(r-1) - N(cr+c+r-1)\Big) \\ &\qquad +ci^2(r-1) + i(cr+c+r-1)\Big\} \\ &= \frac{1}{p(r-1)}\left(\frac{1-r^i}{1-r^N}\left(-cN^2 - N\frac{(cr+c)}{r-1} - N\right) + ci^2 + i\frac{(cr+c)}{r-1} + i\right), \end{aligned}$$

what was to be shown.

□

Remark 2.2. Note that for $p(n) = q(n) = 1/2$ we have $ET_{0:i:N} = iN\left(1 + \frac{2c}{3}N\right) - i^2\left(1 + \frac{2c}{3}i\right)$, *i.e.,* we obtained Proposition 2.1 from [Lef08].

Concerning the conditional game duration (because of (2.7) it is enough to provide formula only for $EW_{0:i:N}$) we have

Corollary 4. Consider the spatially non-homogeneous gambler's ruin problem. We have

$$\begin{aligned} r = 1: \quad EW_{0:i:N} &= \frac{(N^2-1)(cN+1)}{6p} - \frac{(i^2-1)(ci+1)}{6p}, \\ r \neq 1: \quad EW_{0:i:N} &= \frac{cN+1}{p(1-r)}\left(\frac{r+1}{r-1} - N\frac{r^N+1}{r^N-1}\right) - \frac{ci+1}{p(1-r)}\left(\frac{r+1}{r-1} - i\frac{r^i+1}{r^i-1}\right). \end{aligned}$$

Proof. Applying Corollary 2 we have:

- $r = 1$

$$EW_{0:1:i} = \sum_{n=1}^{i-1}\frac{n/i}{p(n)}(i-n) = \frac{1}{p}\sum_{n=1}^{i-1}\frac{n}{i}(i-n)(2cn+1) = \frac{(i-1)(i+1)(ci+1)}{6p}.$$

- $r \neq 1$

$$\begin{aligned} EW_{0:1:i} &= \sum_{n=1}^{i-1}\frac{(1-r^n)(1-r^{i-n})}{p(n)(1-r^i)(1-r)} = \frac{1}{p}\sum_{n=1}^{i-1}\frac{(1-r^n)(1-r^{i-n})}{(1-r^i)(1-r)}(2cn+1) \\ &= \frac{(ci+1)((r+1)(r^i-1) - i(r-1)(r^i+1))}{p(1-r^i)(1-r)^2} \\ &= \frac{ci+1}{p(1-r)}\left(\frac{r+1}{r-1} - i\frac{r^i+1}{r^i-1}\right). \end{aligned}$$

Applying (2.5), *i.e.,* $EW_{0:i:N} = EW_{0:1:N} - EW_{0:1:i}$, completes the proof. □

2.3 The classical case.

For constant winning/losing probabilities we recover known results (all given in Sarkar [Sar06]). We state them here for completeness and will indicate how they can be derived from our more general results.

Corollary 5. Consider the gambler's ruin problem on $\mathbb{E} = \{0, 1, \ldots, N\}$ with constant winning/losing probabilities $p(i) = p, q(i) = q, i = 1, \ldots, N-1, p+q = 1$. We have

$$\rho_{0:i:N} = \begin{cases} \frac{1-r^i}{1-r^N} & \text{if r} = 1, \\ \frac{i}{N} & \text{if r} \neq 1, \end{cases}$$

$$ET_{0:i:N} = \begin{cases} i(N-i) & \text{if r} = 1, \\ \frac{r+1}{r-1}\left(i - N\frac{r^i-1}{r^N-1}\right) & \text{if r} \neq 1, \end{cases}$$

$$EW_{0:i:N} = \begin{cases} \frac{1}{3}(N-i)(N+i) & \text{if r} = 1, \\ \frac{r+1}{r-1}\left[N\frac{r^N+1}{r^N-1} - i\frac{r^i+1}{r^i-1}\right] & \text{if r} \neq 1, \end{cases}$$

$$EB_{0:i:N} = \begin{cases} \frac{1}{3}i(2N-i) & \text{if r} = 1, \\ \frac{r+1}{r-1}\left[N\frac{r^N+1}{r^N-1} - (N-i)\frac{r^{N-i}+1}{r^{N-i}-1}\right] & \text{if r} \neq 1, \end{cases}$$

Results for $ET_{0:i:N}$ follows from Corollary 1 (case $r = 1$); $EW_{0:i:N}$ from Corollary 2 eq. (2.11) followed by (2.5); $EB_{0:i:N}$ follows from results on $EW_{0:i:N}$ and Theorem 2.3 (eq. (2.7)).

2.4 Example

Fix an integer N and some $p, q > 0$. Consider a gambler's ruin problem with rates

$$p(i) = \frac{p(1+\alpha_1 i)}{2ci+1}, \quad q(i) = \frac{q(1+\alpha_2 i)}{2ci+1},$$

with fixed $\alpha_1, \alpha_2, c \geq 0$ such that $p(i), q(i) > 0, p(i) + q(i) \leq 1, i \in \{1, ..., N\}$. We want to calculate $EW_{0:1:N}$.

2.4.1 $N = 3$

We have

$$\mathbf{p} = \left(0, \frac{p(1+\alpha_1)}{2c+1}, \frac{p(1+2\alpha_1)}{2c+1}, 0\right), \quad \mathbf{q} = \left(0, \frac{q(1+\alpha_2)}{2c+1}, \frac{q(1+2\alpha_2)}{2c+1}, 0\right).$$

Note that in general (for $\alpha_1 \neq \alpha_2$) $r(n) = \frac{q(n)}{p(n)} = \frac{q}{p}\frac{(1+\alpha_2 n)}{(1+\alpha_1 n)}$ is non-constant, thus we will apply Theorem 2.3. Eq. (2.6) takes form

$$EW_{0:1:3} = \sum_{n=1}^{2} \frac{\rho_{0:n:3}}{p(n)} \sum_{s=0}^{\lfloor (2-n)/2 \rfloor} \xi_s^{n+1,2} = \frac{\rho_{0:1:3}}{p(1)}\xi_0^{2,2} + \frac{\rho_{0:2:3}}{p(2)}\xi_0^{3,2}.$$

We need winning probabilities $\rho_{0:1:3}$ and $\rho_{0:2:3}$, which can be calculated from Theorem 2.1:

$$\rho_{0:i:3} = \frac{\sum_{n=1}^{i}\prod_{s=1}^{n-1} r(s)}{\sum_{n=1}^{3}\prod_{s=1}^{n-1} r(s)} = \frac{1+(i-1)r(1)}{1+r(1)+r(1)r(2)} = \frac{1+(i-1)\frac{q}{p}\frac{1+\alpha_2}{1+\alpha_1}}{1+\frac{q}{p}\frac{1+\alpha_2}{1+\alpha_1}+\frac{q^2}{p^2}\frac{(1+\alpha_2)(1+2\alpha_2)}{(1+\alpha_1)(1+2\alpha_1)}} =: \frac{1+(i-1)\frac{q}{p}\frac{1+\alpha_2}{1+\alpha_1}}{\gamma(p,q,\alpha_1,\alpha_2)}.$$

We also need $\xi_0^{2,2}$ and $\xi_0^{3,2}$. We have $\mathbf{j}_0^{2,2} = \mathbf{j}_0^{3,2} = \{\emptyset\}$, thus

$$\xi_0^{2,2} = \delta_{\mathbf{j}}^{2,2} = 1 + r(2) = 1 + \frac{q}{p}\frac{1+2\alpha_2}{1+2\alpha_1}, \quad \xi_0^{3,2} = \delta_{\mathbf{j}}^{3,2} = 1$$

(in the latter the second product was 1, since $\{3,\dots,2\} \equiv \emptyset$).

Finally,

$$EW_{0:1:3} = \frac{1}{p\gamma(p,q,\alpha_1,\alpha_2)}\left[\frac{2c+1}{1+\alpha_1}\left(1+\frac{q}{p}\frac{(1+2\alpha_2)}{(1+2\alpha_1)}\right) + \left(1+\frac{q}{p}\frac{(1+\alpha_2)}{(1+\alpha_1)}\right)\frac{4c+1}{1+2\alpha_1}\right]. \quad (2.12)$$

Special cases:

- $\alpha_1 = \alpha_2 = \alpha$. Then (2.12) reduces to

$$EW_{0:1:3} = \frac{1+\frac{q}{p}}{p\left(1+\frac{q}{p}+\frac{q^2}{p^2}\right)}\left(\frac{2c+1}{1+\alpha}+\frac{4c+1}{1+2\alpha}\right). \quad (2.13)$$

  Note that in this case $r(n) = \frac{q}{p}$ is constant, thus (2.13) could be derived in an easier way using Corollary 2:

$$r = 1: \quad EW_{0:1:3} = \sum_{n=1}^{2}\frac{n}{3}\frac{2cn+1}{p(1+\alpha_1 n)}(3-n) = \frac{2}{3p}\left(\frac{2c+1}{1+\alpha}+\frac{4c+1}{1+2\alpha}\right),$$

$$r \neq 1: \quad EW_{0:1:3} = \sum_{n=1}^{2}\frac{\frac{1-r^n}{1-r^3}(1-r^{3-n})}{(1-r)}\frac{2cn+1}{p(1+\alpha_1 n)} = \frac{1-r^2}{p(1-r^3)}\left(\frac{2c+1}{1+\alpha}+\frac{4c+1}{1+2\alpha}\right),$$

  what is equivalent to (2.13) in both cases. *Note* also that this is not a spatially non-homogeneous case as long as $\alpha > 0$.

- $\alpha_1 = \alpha_2 = 0$. Then (2.12) (and thus (2.13)) reduces to

$$EW_{0:1:3} = \frac{2\left(1+\frac{q}{p}\right)}{p\left(1+\frac{q}{p}+\frac{q^2}{p^2}\right)}\left(\frac{3c+1}{1+\alpha}\right). \quad (2.14)$$

  Note that this is a spatially non-homogeneous case, thus (2.14) could be derived from Corollary 4 (we skip the calculations).

- $\alpha_1 = \alpha_2 = 0$ and $c = 0$, then (2.14) reduces to

$$EW_{0:1:3} = \frac{2\left(1+\frac{q}{p}\right)}{p\left(1+\frac{q}{p}+\frac{q^2}{p^2}\right)}.$$

  This situation corresponds to a gambler's ruin problem with constant birth and death rates. In particular, for $p = q = 1/2$ we have $EW_{0:1:3} = \frac{8}{3}$ what agrees with Example 1 in [Len09b].

2.4.2 General $N \geq 3, p = q$ and $\alpha_2 = \alpha_1 = 1$

We thus have $p(i) = \frac{p(1+i)}{2ci+1}, q(i) = \frac{q(1+i)}{2ci+1}$. This is constant $r(n) = \frac{q(n)}{p(n)} = \frac{q}{p} = 1$ case, which is however not spatially non-homogeneous. We skip the lengthy calculations, but we can obtain $EW_{0:1:N}$ from Corollary 2 (H_N is the N-th harmonic number):

$$
\begin{aligned}
EW_{0:1:N} &= \sum_{n=1}^{N-1} \frac{n(N-n)(2cn+1)}{pN(1+n)} \\
&= \tfrac{1}{p}\left(\tfrac{c}{3}(N-5)(N+2) + \tfrac{1}{2}(3+N)\right) + \tfrac{1}{Np}(2c-1)(1+N)H_N = \tfrac{c}{3p}N^2 + \Theta(N),
\end{aligned}
$$

which for $p(i) = p(1+i), q(i) = q(1+i)$ (*i.e.*, for $c = 0$) simplifies to

$$EW_{0:1:N} = \frac{N+3}{2p} - \frac{1}{Np}(N+1)H_N = \frac{N}{2p} + \Theta(log(N)).$$

2.4.3 General $N \geq 3, p = q$ and $\alpha_2 = \alpha_1 = \alpha$

$$
\begin{aligned}
EW_{0:1:i} &= \sum_{n=1}^{i-1} \frac{n(i-n)(2cn+1)}{pi(1+\alpha n)} \\
&= \frac{1}{6\alpha^4 pi}\left(\alpha(i-1)(\alpha^2 i(2c(i+1)+3) + \alpha(6-6ci) - 12c) + \right. \\
&\qquad \left. 6(\alpha-2c)(\alpha i+1)\left[\psi\left(1+\tfrac{1}{\alpha}\right) - \psi\left(i+\tfrac{1}{\alpha}\right)\right]\right),
\end{aligned}
$$

where ψ is a digamma function. It is known that $\psi(m) = H_{m-1} - \gamma$, where $\gamma = 0.5772156...$ is a known Euler–Mascheroni constant. Let us assume that $\alpha = \frac{1}{m}$ and m is an integer. Then $\psi\left(1+\frac{1}{\alpha}\right) - \psi\left(i+\frac{1}{\alpha}\right) = H_m - H_{i+m-1}$.

3 Random walk on a polygon

Fix an integer $m \geq 2$. Let

$$\mathbf{p} = (p(0), p(1), \ldots, p(m)), \quad \mathbf{q} = (q(0), q(1), \ldots, q(m)),$$

where $p(i), q(i) > 0, p(i) + q(i) \leq 1$ for $i \in \{0 \ldots, m\}$. Consider the following random walk $\mathbf{X} \equiv \{X_t\}_{t \in \mathbb{N}}$ on $\mathbb{E} = Z_{m+1}$. Being in state i we move to the state $i+1$ with probability $p(i)$, we move to the state $i-1$ with probability $q(i)$, and we do nothing with the remaining probability. Throughout the chapter, in the context of a random walk on a polygon, all additions and substractions are performed modulo $m+1$. We will refer to this walk as to a *random walk on a polygon*. The notation intentionally resembles that of gambler's ruin problem. Throughout the section we consider fixed $\mathbf{p}, \mathbf{q}$ and $m \geq 2$ (and omit subscripts $\mathbf{p}, \mathbf{q}$ in random variables below). We are interested in:

$$\begin{aligned}
A_i &= \{X : X_0 = i, X_n = i, \forall_{0<t<n} X_t \neq i, \forall_{k\in\mathbb{E}} \exists_{0\le t\le n} X_t = k\} \\
L_{i,j} &= \{X : X_0 = i, X_n = j, \forall_{0<t<n} X_t \neq j, \forall_{k\in\mathbb{E}} \exists_{0\le t\le n} X_t = k\} \\
V_{i,j} &= \inf\{n \ge 1 : X_0 = i, X_n = j, \forall_{k\in\mathbb{E}} \exists_{0\le t\le n} X_t = k\} \\
V_i &= \inf\{n \ge 1 : X_0 = i, \forall_{k\in\mathbb{E}} \exists_{0\le t\le n} X_t = k\} \\
R_i &= \inf\{n_2 \ge 1 : X_0 = i, X_{n_1+n_2} = i, n_1 = \inf\{n \ge 1 : \forall_{k\in\mathbb{E}} \exists_{0\le t\le n} X_t = k\}\}
\end{aligned}$$

In other words: A_i is the event that the process starting at i will return for the first time to i after all other vertices are visited; $L_{i,j}$ is the event that the process starting at i will reach for the first time state j after visiting all other vertices; $V_{i,j}$ is the number of steps of the process starting at i to reach for the first time state j after visiting all other vertices; V_i is the number of steps of the process starting at i needed to visit all vertices; R_i is the number of additional steps for the process starting at i needed to reach i after visiting all the vertices.

For $j \preceq i \preceq k$, where $\preceq$ is a cyclic order, *i.e.,* $j \le i \le k$ or $i \le k \le j$ or $k \le j \le i$, let $G(\mathbf{p}, \mathbf{q}, j, i, k)$ denote a gambler's ruin game with i being a starting state, j being a losing state and k being a winning state. Note that independently of j, i, k, winning and losing probabilities $\mathbf{p}, \mathbf{q}$ are fixed.

Notation. In contrast to a usual notation neither $\sum_{k=s}^{t} a_k = 0$ nor $\prod_{k=s}^{t} a_k = 1$ for $t < s-1$. Since we are considering operations in Z_{m+1}, we define

$$\text{For } t < s \le m, s-t > 1: \quad \sum_{k=s}^{t} a_k := a_s + a_{s+1} + \ldots + a_m + a_0 + \ldots + a_t,$$

$$\prod_{k=s}^{t} a_k := a_s \cdot a_{s+1} \cdot \ldots \cdot a_m \cdot a_0 \cdot \ldots \cdot a_t,$$

$$\text{For } s = t+1 \bmod m+1: \quad \sum_{k=s}^{t} a_k = 0 \qquad \prod_{k=s}^{t} a_k := 1.$$

In all other cases we use usual sums and products. Using this notation, we are ready to state our results.

Theorem 3.1. *Consider the random walk on a polygon described above. We have*

$$
\begin{aligned}
P(A_i) &= \frac{1}{1+r(i)}\left(\frac{1}{\sum_{n=i+1}^{i-1}\prod_{s=i+1}^{n-1} r(s)} + \frac{1}{\sum_{n=i+2}^{i}\prod_{s=n}^{i}\left(\frac{1}{r(s)}\right)}\right) && (3.1)\\
P(L_{i,j}) &= \frac{1}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)}\left(\frac{\sum_{n=i+1}^{j-1}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)} + \frac{\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=n}^{j-1}\frac{1}{r(s)}}\right) && (3.2)\\
EV_{i,j} &= \rho_{j+1:i:j-1}\left(EW_{j+1:i:j-1} + EB_{j+1:j-1:j} + ET_{j:j+1:j}\right)\\
&\quad +(1-\rho_{j+1:i:j-1})\left(EB_{j+1:i:j-1} + EW_{j:j+1:j-1} + ET_{j:j-1:j}\right) && (3.3)\\
EV_i &= \sum_{j=i+1}^{i-1} P(L_{i,j})EV_{i,j} && (3.4)\\
ER_i &= \sum_{k=i+1}^{i-1} P(L_{i,k})ET_{i:k:i} && (3.5)
\end{aligned}
$$

The proof of Theorem 3.1 is postponed to Section 5.2.1.

Constant $r(n) = r = \frac{q(n)}{p(n)}$**.**

In this case the starting point does not matter, we consider $i = 0$. Note that $P(A_i)$ and $P(L_{i,j})$ depend on $p(n)$ and $q(n)$ only through $r(n)$, thus they must reduce to known results for constant birth $p(n) = p$ and death $q(n) = q$ rates (see (3.1) and (3.3) in [Sar06]). Indeed, substituting $r(n) = r$ to (3.1) and (3.2) yields

Corollary 6. Consider the random walk on polygon with constant $r(n) = \frac{q(n)}{p(n)}$, then we have

$$
P(A_0) = \begin{cases} \frac{1}{m} & \text{if } r = 1, \\ \frac{r-1}{r+1}\frac{r^m+1}{r^m-1} & \text{if } r \neq 1, \end{cases}
$$

$$
P(L_{0,j}) = \begin{cases} \frac{1}{m} & \text{if } r = 1, \\ \frac{r^{m-j}(r-1)}{r^m-1} & \text{if } r \neq 1. \end{cases}
$$

We skip the formulas for $EV_{0,j}, EV_0$ and ER_0 in this case, noting that they can be derived from Corollaries 1 and 2.

Constant $q(n) = q, p(n) = p$

First, let us recall formulas for EV_0, ER_0 for the case $p + q = 1$.

Corollary 7. [Sar06] Consider the random walk on a polygon with constant $q(n) = q, p(n) = p, p + q = 1$. We have

$$EV_0 = \begin{cases} \frac{m(m+1)}{2} & \text{if } r=1, \\ \frac{r+1}{r-1}\left[m - \frac{1}{r-1} - \frac{m^2}{r^m-1} + \frac{(m+1)^2}{r^{m+1}-1}\right] & \text{if } r \neq 1, \end{cases}$$

$$ER_0 = \begin{cases} \frac{1}{6}(m+1)(m+2) & \text{if } r=1, \\ \frac{r+1}{r-1}\left[\frac{r}{r-1} - \frac{m(m+2)}{r^m-1} + \frac{(m+1)^2}{r^{m+1}-1}\right] & \text{if } r \neq 1, \end{cases}$$

In the case $p+q \leq 1$ note that $EB_{j:i:k} = \frac{1}{p(1+r)} EB^1_{j:i:k}$, $EW_{j:i:k} = \frac{1}{p(1+r)} EW^1_{j:i:k}$, $ET_{j:i:k} = \frac{1}{p(1+r)} ET^1_{j:i:k}$, where superscript 1 denotes the case $p+q=1$. Thus Theorem 3.1 implies $EV_0 = EV_0^1, ER_0 = ER_0^1$, *i.e.,* we have

Corollary 8. Consider the random walk on a polygon with constant $q(n) = q, p(n) = p$. We have

$$EV_0 = \begin{cases} \frac{m(m+1)}{4p} & \text{if } r=1, \\ \frac{1}{p(r-1)}\left[m - \frac{1}{r-1} - \frac{m^2}{r^m-1} + \frac{(m+1)^2}{r^{m+1}-1}\right] & \text{if } r \neq 1, \end{cases}$$

$$ER_0 = \begin{cases} \frac{1}{12p}(m+1)(m+2) & \text{if } r=1, \\ \frac{1}{p(r-1)}\left[\frac{r}{r-1} - \frac{m(m+2)}{r^m-1} + \frac{(m+1)^2}{r^{m+1}-1}\right] & \text{if } r \neq 1, \end{cases}$$

4 Fastest Strong Stationary Time for a symmetric random walk on a circle

Consider an ergodic Markov chain $\mathbf{X} = \{X_k\}_{k\geq 0} \sim (\nu, \mathbf{P}_X)$ on a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ with a stationary distribution π, initial distribution ν and a transition matrix $\mathbf{P}_X$. We are interested in measuring nonstationarity of $\mathbf{X}_k$ via **separation "distance"**

$$sep(\nu\mathbf{P}_X^k, \pi) = \max_{\mathbf{e}\in\mathbb{E}} \left(1 - \frac{\nu\mathbf{P}_X^k(\mathbf{e})}{\pi(\mathbf{e})}\right).$$

Note that it is not symmetric, that is why it is not an actual distance, however it is an upper bound on a **total variation distance** $d_{TV}(\nu\mathbf{P}_X^k, \pi) = 1/2\sum_{\mathbf{e}\in\mathbb{E}} |P(X_k = \mathbf{e}) - \pi(\mathbf{e})|$.
A random variable T is a **strong stationary time** (SST) T for $\mathbf{X}$ if it is a randomized stopping time for $\mathbf{X}$ such that

$$\forall(\mathbf{e} \in \mathbb{E}) \; P(X_k = \mathbf{e} | T = k) = \pi(\mathbf{e}).$$

The notion of separation distance fits perfectly into a notion of SST, in [AD87] it is shown that for an SST T we have

$$sep(\nu\mathbf{P}_X^k, \pi) \leq P(T > k).$$

We say that T is a **fastest strong stationary time** (FSST) if $sep(\nu\mathbf{P}_X^k, \pi) = P(T > k)$.

In this section we consider a symmetric random walk on a polygon with constant rates $p(i) = q(i) = p$ on d points (*i.e.,* $m = d-1$). Moreover, we will refer to the random walk as to a *symmetric random walk on a circle* (to be consistent with [DF90b], we will compare our result to a result from this article) on $\mathbb{Z}_d$, *i.e.,* $\{0, \ldots, d-1\}$. We will show a construction of a fastest strong stationary time for this symmetric random walk on a circle, moreover we have

Lemma 4.1. For the fastest strong stationary time T for a symmetric random walk on a circle with $d = 2N$ we have

$$ET = \begin{cases} \dfrac{2N^2+1}{12p} & \text{for } p \in (0, 1/3] \text{ and } N > 1, \\ \dfrac{1}{4p} & \text{for } p \in (0, 1/4] \text{ and } N = 1, \\ \dfrac{1}{2(1-2p)} & \text{for } p \in (1/4, 1/2) \text{ and } N = 1. \end{cases}$$

Remark 4.1. A construction of a strong stationary time for a symmetric random walk on a circle with $p = 1/3$ is presented in [DF90b]. For $d = 2^a, a > 1$ their construction yields an SST T_0 such that

$$ET_0 = \frac{3}{2}2^{2a}\left(2^{-4} + 2^{-6} + \cdots + 2^{-2(a-1)} + 2 \times 2^{-2a}\right) = \frac{1}{8}d^2 + 1$$

(see the bottom of the page 1484 in [DF90b]), whereas Lemma 4.1 states that a fastest strong stationary time T fulfills ($N = d/2$)

$$ET = \frac{1}{8}d^2 + \frac{1}{4},$$

what means that a construction from [DF90b] does not yield a *fastest* strong stationary time (authors mention this fact in their Example 3.1). Note that ET and ET_0 differ by $\frac{3}{4}$ (independently of d).

Strong stationary duality For a general description of a strong stationary duality see [DF90b] (total ordering and set-valued chains) and [LS12a], [Lor18] (general partial ordering). Here we will describe this duality for chains on the same state space. Let both $\mathbf{X} \sim (\nu, \mathbf{P}_X)$ and $\mathbf{X}^* \sim (\nu^*, \mathbf{P}^*_X)$ be chains on $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$, chain $\mathbf{X}$ is ergodic with a stationary distribution π, whereas $\mathbf{X}^*$ is an absorbing chain with a unique absorbing state $\mathbf{e}_M$. We say that a stochastic matrix of size $d \times d$ is a **link** if $\Lambda(\mathbf{e}_M, \mathbf{e}) = \pi(\mathbf{e})$ for all $\mathbf{e} \in \mathbb{E}$. We say that $\mathbf{X}^*$ is a **strong stationary dual** of $\mathbf{X}$ with the link Λ if

$$\nu = \nu^*\Lambda \quad \text{and} \quad \Lambda\mathbf{P}_X = \mathbf{P}^*_X\Lambda. \tag{4.1}$$

Diaconis and Fill [DF90b] proved that the absorption time T^* of $\mathbf{X}^*$ is an SST for $\mathbf{X}$. If the corresponding T^* is an FSST for $\mathbf{X}$, then the chain $\mathbf{X}^*$ is called a **sharp SSD**.

Fix some partial ordering $\preceq$ on $\mathbb{E}$, such that $\mathbf{e}_1$ is the minimum and $\mathbf{e}_M$ is the maximum. Let $\mathbf{C}(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{1}(\mathbf{e}_i \preceq \mathbf{e}_j)$ be the corresponding *ordering matrix* (always invertible, the inverse $\mathbf{C}^{-1}$ is called the Möbius matrix). Assume that $\nu(\mathbf{e}_1) = 1$ (*i.e.,* chain $\mathbf{X}$ starts in $\mathbf{e}_1$), then (4.1) implies that also $\nu^*(\mathbf{e}_1) = 1$. Let $\overleftarrow{\mathbf{P}}_X$ be a transition matrix of a time reversed chain, *i.e.,* $\overleftarrow{\mathbf{P}}_X(\mathbf{e}_i, \mathbf{e}_j) = \frac{\pi(\mathbf{e}_j)}{\pi(\mathbf{e}_i)}\mathbf{P}_X(\mathbf{e}_j, \mathbf{e}_i)$. We have

Theorem 4.2 (Theorem 2 in [LS12a], Remark 2.2 in [LS16], simplified version)**.** *Let $\mathbf{X} \sim (\nu, \mathbf{P}_X)$ be an ergodic Markov chain on a finite state space $\mathbb{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_M\}$ starting at $\mathbf{e}_1$ (i.e., $\nu(\mathbf{e}_1) = 1$), with a stationary distribution π, partially ordered by $\preceq$ (with ordering matrix $\mathbf{C}$), with $\mathbf{e}_1$ being the minimum and $\mathbf{e}_M$ being the maximum. Assume that $\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}$ is a non-negative matrix. Then there exists a sharp SSD $\mathbf{X}^* \sim (\nu^*, \mathbf{P}^*_X)$ on $\mathbb{E}$ with $\nu^*(\mathbf{e}_1) = 1$ and transitions*

$$\mathbf{P}^*_X(\mathbf{e}_i, \mathbf{e}_j) = \frac{H(\mathbf{e}_j)}{H(\mathbf{e}_i)}\left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(\mathbf{e}_j, \mathbf{e}_i) \tag{4.2}$$

with a unique absorbing state $\mathbf{e}_M$, where $H(\mathbf{e}) = \sum_{\mathbf{e}' \preceq \mathbf{e}} \pi(\mathbf{e})$.

Remark 4.2. The condition that $\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}$ is a non-negative matrix was called $^{\downarrow}$-Möbius monotonicity in [LS12a].

Proof of Lemma 4.1. First, we will construct a sharp SSD for this symmetric random walk on a circle using Theorem 4.2. It will be more convienient to work with states numerated as $1^*, 2^*, \ldots, d^* = (2N)^*$ (instead of $0, 1, \ldots, 2N-1$). Our walk $\mathbf{X}$ moves either right or left, or it does not move, i.e., it has the transition matrix:

$$\mathbf{P}_X(i^*, j^*) = \begin{cases} 1-2p & \text{if} \quad j^* = i^*, \\ p & \text{if} \quad (j^* = (i+1)^*, i^* \neq (2N)^*) \ \vee \ (j^* = (i-1)^*, i^* \neq 1^*) \\ & \quad \vee \ (j^* = 1, i^* = (2N)^*) \vee (j^* = (2N)^*, i^* = 1). \end{cases}$$

It will be even more convienient to work with another enumeration of states. Consider a set of states $\{1, \ldots, 2N\}$ and let us define a bijection between this set and the set $\{1^*, \ldots, (2N)^*\}$ in the following way:

$$\sigma(i^*) = \begin{cases} 2i-1 & \text{if} \quad i \leq N, \\ 2(2N-i+1) & \text{if} \quad i > N. \end{cases}, \qquad \sigma^{-1}(i) = \begin{cases} \left(\frac{i+1}{2}\right)^* & \text{if} \quad \text{is odd}, \\ \left(2N - \frac{i}{2} + 1\right)^* & \text{if} \quad \text{is even}. \end{cases}$$

The bijection for $d = 2N = 8$ is following

$$\begin{aligned} \sigma((1^*, 2^*, 3^*, 4^*, 5^*, 6^*, 7^*, 8^*)) &= (1, 3, 5, 7, 8, 6, 4, 2), \\ \sigma^{-1}((1, 2, 3, 4, 5, 6, 7, 8)) &= (1^*, 8^*, 2^*, 7^*, 3^*, 6^*, 4^*, 5^*), \end{aligned}$$

it is depicted in Fig. 1 (left). The transition matrix of the chain $\mathbf{X}$ can be rewritten as:

$$\mathbf{P}_X(i,j) = \begin{cases} 1-2p & \text{if} \quad i = j, \\ p & \text{if} \quad |i-j| = 2 \ \vee \ (i=1, j=2) \ \vee \ (i=2, j=1) \ \vee \\ & \quad (i = 2N-1, j = 2N) \ \vee \ (i = 2N, j = 2N-1). \end{cases}$$

Continuing our example $d = 2N = 8$ we have (using enumeration of states $1, 2, \ldots, 2N$))

$$\mathbf{P}_X = \begin{bmatrix} 1-2p & p & p & 0 & 0 & 0 & 0 & 0 \\ p & 1-2p & 0 & p & 0 & 0 & 0 & 0 \\ p & 0 & 1-2p & 0 & p & 0 & 0 & 0 \\ 0 & p & 0 & 1-2p & 0 & p & 0 & 0 \\ 0 & 0 & p & 0 & 1-2p & 0 & p & 0 \\ 0 & 0 & 0 & p & 0 & 1-2p & 0 & p \\ 0 & 0 & 0 & 0 & p & 0 & 1-2p & p \\ 0 & 0 & 0 & 0 & 0 & p & p & 1-2p \end{bmatrix}$$

We will now compute an SSD chain using total ordering $1 < 2 < \ldots < 2N$. Mapping the total ordering $1 < 2 < \ldots 2N$ into the ordering on original states $1^*, 2^*, \ldots, (2N)^*$, we have

$$i^* \prec j^* \Leftrightarrow \sigma(i^*) < \sigma(j^*),$$

Note that $\prec$ is also a total ordering, we have

$$1^* \prec (2N)^* \prec 2^* \prec (2N-1)^* \prec \cdots \prec (N+1)^*.$$

We will thus work with total ordering $1 < 2 \ldots < 2N$ – which is equivalent (with easier notation) to working with $1^* \prec (2N)^* \prec \ldots \prec (N+1)^*$.
The ordering matrix for total ordering is $\mathbf{C}(i,j) = \mathbf{1}(i \leq j)$, the Möbius matrix (*i.e.,* the inverse of $\mathbf{C}$) is then following:

$$\mathbf{C}^{-1}(i,j) = \begin{cases} 1 & \text{if} \quad i = j, \\ -1 & \text{if} \quad i = j-1, i < 2N. \end{cases}$$

We have

$$H(i) = \sum_{j \leq i} \pi(j) = \sum_{j \leq i} \frac{1}{2N} = \frac{i}{2N}. \tag{4.3}$$

Using above derivations and the fact that the chain is reversible ($\overleftarrow{\mathbf{P}}_X = \mathbf{P}_X$), for $i < 2N$ we have:

$$\begin{aligned} \left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(i,j) &= \left(\mathbf{C}^{-1}\mathbf{P}_X\mathbf{C}\right)(i,j) = \sum_l \mathbf{C}^{-1}(i,l) \sum_{k \leq j} \mathbf{P}_X(l,k) \\ &= \sum_{k \leq j} \mathbf{P}_X(i,k) - \mathbf{P}_X(i+1,k) \\ &= \mathbf{P}_X(i,j) + \left(\sum_{k<j} \mathbf{P}_X(i,k) - \mathbf{P}_X(i+1,k+1) \right) - \mathbf{P}_X(i+1,1), \end{aligned}$$

whereas for $i = 2N$ we have

$$\left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(i,j) = \sum_{k \leq j} \mathbf{P}_X(2N,k) = \begin{cases} 0 & \text{if} \quad j < 2N-2, \\ p & \text{if} \quad j = 2N-2, \\ 2p & \text{if} \quad j = 2N-1, \\ 1 & \text{if} \quad j = 2N, \end{cases}$$

We also have:

$$\mathbf{P}_X(i,k) - \mathbf{P}_X(i+1,k+1) = \begin{cases} p & \text{if} \quad (i=1, j=2) \ \vee \ (i=2, j=1), \\ -p & \text{if} \quad (i=2N-1, j=2N-2) \ \vee \ (i=2N-2, j=2N-1). \end{cases}$$

Using above derivations we can easily calculate all the cases:

$$\left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(i,j) = \begin{cases} 1-2p & \text{if} \quad 1 < i = j < 2N-1, \\ 1-3p & \text{if} \quad i = j = 1 \ \vee \ i = j = 2N-1, \\ 1 & \text{if} \quad i = j = 2N, \\ p & \text{if} \quad |i-j| = 2, \ j \neq 2N, \\ 2p & \text{if} \quad i = 2N, \ j = 2N-1. \end{cases} \tag{4.4}$$

Continuing our example $d = 2N = 8$ we have (again, using enumeration $1, 2, \ldots, 8$)

$$\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C} = \begin{bmatrix} 1-3p & 0 & p & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-2p & 0 & p & 0 & 0 & 0 & 0 \\ p & 0 & 1-2p & 0 & p & 0 & 0 & 0 \\ 0 & p & 0 & 1-2p & 0 & p & 0 & 0 \\ 0 & 0 & p & 0 & 1-2p & 0 & p & 0 \\ 0 & 0 & 0 & p & 0 & 1-2p & 0 & 0 \\ 0 & 0 & 0 & 0 & p & 0 & 1-3p & 0 \\ 0 & 0 & 0 & 0 & 0 & p & 2p & 1 \end{bmatrix}.$$

Combining (4.4) with (4.3) and noting that $\frac{H(i)}{H(j)} = \frac{i}{j}$, Theorem 4.2 and yields the following transitions of a sharp SSD chain $\mathbf{X}^*$ (written using the original enumeration of states)

$$\mathbf{P}^*_X(i^*, j^*) = \frac{H(j^*)}{H(i^*)}\left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(j^*, i^*) = \frac{H(\sigma(j^*))}{H(\sigma(i^*))}\left(\mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C}\right)(\sigma(j^*), \sigma(i^*)),$$

thus

$$\mathbf{P}^*_X(i^*, j^*) = \begin{cases} 1-2p & \text{if} \quad 1 < \sigma(j^*) = \sigma(i^*) < 2N-1, \\ 1-3p & \text{if} \quad \sigma(j^*) = \sigma(i^*) = 1 \vee \sigma(j^*) = \sigma(i^*) = 2N-1, \\ 1 & \text{if} \quad \sigma(j^*) = \sigma(i^*) = 2N, \\ p\frac{\sigma(j^*)}{\sigma(i^*)} & \text{if} \quad |\sigma(j^*) - \sigma(i^*)| = 2, \sigma(i^*) \neq 2N, \\ 2p\frac{\sigma(j^*)}{\sigma(i^*)} & \text{if} \quad \sigma(j^*) = 2N, \sigma(i^*) = 2N-1. \end{cases}$$

We leave it to the reader to check that the condition $|\sigma(j^*) - \sigma(i^*)| = 2$ for $j, i \leq N$ or $j, i > N$ is equivalent to $|j - i| = 1$, whereas for $j \leq N, i > N$ or for $i \leq N, j > N$ the condition is never met. Thus, the transition matrix of $\mathbf{X}^*$ can rewritten in the following way, using ordering $\prec$:

$$\mathbf{P}^*_X(i^*, j^*) = \begin{cases} 1-2p & \text{if} \quad j^* = i^*, 2^* \preceq i^* \prec N \text{ or } (N+1)^* \prec i^* \preceq (2N)^*, \\ 1-3p & \text{if} \quad j^* = i^*, i^* \in \{1^*, N^*\}, \\ 1 & \text{if} \quad j^* = i^* = (N+1)^*, \\ \frac{(2i+1)p}{2i-1} & \text{if} \quad j = i+1, 1^* \preceq i^* \prec N^*, \\ \frac{(2i-3)p}{2i-1} & \text{if} \quad j = i-1, 1^* \prec i^* \preceq N^*, \\ \frac{(2N-i)p}{2N-i+1} & \text{if} \quad j = i+1, (N+2)^* \preceq i^* \prec (2N)^*, \\ \frac{(2N-i+2)p}{2N-i+1} & \text{if} \quad j = i-1, (N+2)^* \preceq i^* \preceq (2N)^*, \\ \frac{4Np}{2N-1} & \text{if} \quad i^* = N^*, j^* = (N+1)^*. \end{cases}$$

- First, let us consider case $p \in (0, 1/3]$ and $N > 1$.

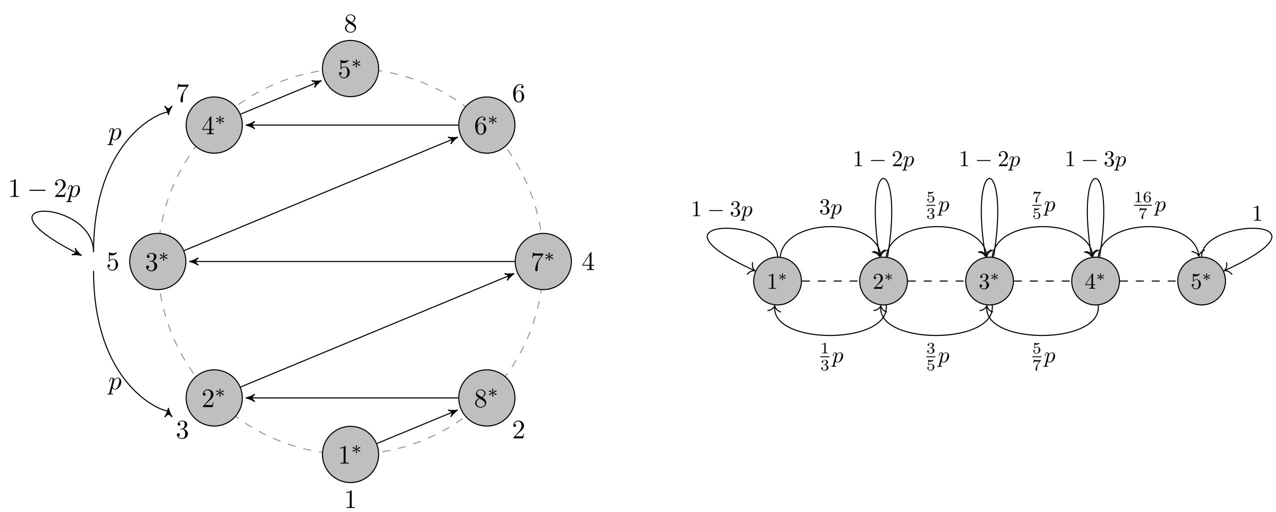
  Note that the assumption $p \in (0, 1/3]$ implies that $\mathbf{P}^*_X$ is a transition matrix. Continuing the example $d = 2N = 8$, we have (using the enumeration $1^*, 2^*, \dots, (2N)^*$)

$$\mathbf{P}^*_X = \begin{bmatrix} 1-3\,p & 3\,p & 0 & 0 & 0 & 0 & 0 & 0 \\ p/3 & 1-2\,p & 5/3\,p & 0 & 0 & 0 & 0 & 0 \\ 0 & 3/5\,p & 1-2\,p & 7/5\,p & 0 & 0 & 0 & 0 \\ 0 & 0 & 5/7\,p & 1-3\,p & \frac{16\,p}{7} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4/3\,p & 1-2\,p & 2/3\,p & 0 \\ 0 & 0 & 0 & 0 & 0 & 3/2\,p & 1-2\,p & p/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\,p & 1-2\,p \end{bmatrix}.$$

Note that the resulting chain (recall, it starts at 1^*) will never reach states $(N{+}2)^*, \dots, (2N)^*$. Denote the resulting chain on $\{1^*, \dots, (N+1)^*\}$ by $\mathbf{Y}^*$. This is a birth and death chain with a unique absorbing state $(N+1)^*$, let us write down the relevant transitions only

$$\mathbf{P}^*_Y(i^*, j^*) = \begin{cases} 1-3p & \text{if} \quad j = i, i \in \{1, N\}, \\ 1-2p & \text{if} \quad j = i, 2 \le i < N, \\ 1 & \text{if} \quad j = i = N+1, \\ \frac{(2i+1)p}{2i-1} & \text{if} \quad j = i+1, 1 \le i < N, \\ \frac{4Np}{2N-1} & \text{if} \quad i = N, j = N+1, \\ \frac{(2i-3)p}{2i-1} & \text{if} \quad j = i-1, 1 < i \le N. \end{cases}$$

For $d = 2N = 8$ the transitions are depicted in Fig. 1 (right).



X and the "Zig-zag" ordering

A sharp SSD $\mathbf{X}^*$

Figure 1: Case $d = 2N = 8$: "zig-zag" ordering and state space of $\mathbf{X}$ (left), the corresponding sharp SSD $\mathbf{Y}^*$ (right)

Since there is no confusion (in the chain $\mathbf{Y}^*$), we will identify a state i^* simply with i. Let $T \equiv T_{1:1:N+1}$ denote the absorption time (in $N+1$) of $\mathbf{Y}^*$ (starting at 1). Using Theorem

2.2 we have:

$$ET_{1:1:N+1} = \sum_{n=1}^{N} \left[d_n \sum_{s=1}^{n} \frac{1}{p(s)d_s} \right]. \tag{4.5}$$

Let us write a formula for $p(s)$ explicitly:

$$p(s) = \begin{cases} \frac{(2s+1)p}{2s-1} & \text{if} \quad i < N, \\ \frac{4Np}{2N-1} & \text{if} \quad i = N. \end{cases} \tag{4.6}$$

We need to compute $d(s)$. For $1 \leq s < N$ we have

$$d_s = \prod_{i=2}^{s} \frac{q(i)}{p(i)} = \prod_{i=2}^{s} \frac{\frac{2i-3}{2i-1}}{\frac{2i+1}{2i-1}} = \prod_{i=2}^{s} \frac{2i-3}{2i+1} = \frac{3}{(2s-1)(2s+1)}$$

and for $s = N$ we have

$$d_N = \prod_{i=2}^{N} \frac{q(i)}{p(i)} = d_{N-1} \frac{q(N)}{p(N)} = \frac{3}{(2N-3)(2N-1)} \frac{\frac{2N-3}{2N-1}}{\frac{4N}{2N-1}} = \frac{3}{4N(2N-1)}.$$

Plugging above formulas for $p(s), d_s$ in (4.5) (and using a formula $\sum_{s=1}^{n}(2s-1)^2 =$

$\frac{n(2n-1)(2n+1)}{3}$) we obtain for $N>1$: $ET_{1:1:N+1} =$

$$
\begin{aligned}
& \sum_{n=1}^{N}\left[d_n\sum_{s=1}^{n}\frac{1}{p(s)d_s}\right] \\
= & \sum_{n=1}^{N-1}\left[d_n\sum_{s=1}^{n}\frac{1}{p(s)d_s}\right]+d_N\sum_{s=1}^{N}\frac{1}{p(s)d_s} \\
= & \sum_{n=1}^{N-1}\left[d_n\sum_{s=1}^{n}\frac{1}{p(s)d_s}\right]+d_N\sum_{s=1}^{N-1}\frac{1}{p(s)d_s}+\frac{d_N}{p(N)d_N} \\
= & \sum_{n=1}^{N-1}\left[\frac{3}{(2n-1)(2n+1)}\sum_{s=1}^{n}\frac{1}{p\frac{2s+1}{2s-1}\frac{3}{(2s-1)(2s+1)}}\right]+\frac{3}{4N(2N-1)}\sum_{s=1}^{N-1}\frac{1}{p\frac{2s+1}{2s-1}\frac{3}{(2s-1)(2s+1)}}+\frac{1}{p\frac{4N}{2N-1}} \\
= & \sum_{n=1}^{N-1}\left[\frac{1}{p(2n-1)(2n+1)}\sum_{s=1}^{n}(2s-1)^2\right]+\frac{1}{p4N(2N-1)}\sum_{s=1}^{N-1}(2s-1)^2+\frac{2N-1}{p4N} \\
= & \sum_{n=1}^{N-1}\left[\frac{1}{p(2n-1)(2n+1)}\frac{n(2n-1)(2n+1)}{3}\right]+\frac{1}{p4N(2N-1)}\frac{(N-1)(2N-3)(2N-1)}{3}+\frac{2N-1}{p4N} \\
= & \frac{1}{3p}\sum_{n=1}^{N-1}n+\frac{(N-1)(2N-3)}{p12N}+\frac{2N-1}{p4N} \\
= & \frac{4N}{12pN}\frac{N(N-1)}{2}+\frac{(N-1)(2N-3)}{p12N}+\frac{3(2N-1)}{p12N} \\
= & \frac{2N^2(N-1)+(N-1)(2N-3)+3(2N-1)}{12pN} \\
= & \frac{2N^3-2N^2+2N^2-5N+3+6N-3}{12pN}=\frac{2N^3+N}{12pN}=\frac{2N^2+1}{12p}.
\end{aligned}
$$

- Now consider case $N=1$.

  We can directly compute a separation distance $sep(\nu\mathbf{P}_X^k,\pi)$ for $\mathbf{X}$ starting at 1 (i.e., $\nu=(1,0)$). We have

$$sep(\nu\mathbf{P}_X^k,\pi)=\max_{i\in\{1,2\}}\left(1-\frac{\mathbf{P}_X^k(1,i)}{\frac{1}{2}}\right)=1-2\min_{i\in\{1,2\}}\mathbf{P}_X^k(1,i). \tag{4.7}$$

  Spectral decomposition yields

$$\mathbf{P}_X^k=\frac{1}{2}\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}1&0\\0&(1-4p)^k\end{pmatrix}\begin{pmatrix}1&1\\-1&1\end{pmatrix}=\frac{1}{2}\begin{pmatrix}1+(1-4p)^k&1-(1-4p)^k\\1-(1-4p)^k&1+(1-4p)^k\end{pmatrix} \tag{4.8}$$

  and thus

$$sep(\nu\mathbf{P}_X^k,\pi)=1-\min\{1+(1-4p)^k,1-(1-4p)^k\}=\begin{cases}(1-4p)^k & \text{if} \quad p\in(0,1/4),\\ (4p-1)^k & \text{if} \quad p\in(1/4,1/2).\end{cases}$$

On the other hand we know that there always exists a fastest strong stationary time T (see Proposition 1.10 (b) in [DF90b]), i.e., $sep(\nu\mathbf{P}_X^k, \pi) = P(T > k)$. For $p \in (0, 1/4)$ we have that T has distribution $Geo(4p)$, whereas for $p \in (1/4, 1/2)$ we have $P(T > k) = (4p-1)^k = (1-2(1-2p))^k$, thus T has distribution $Geo(2(1-2p))$. It implies that

$$ET = \begin{cases} \dfrac{1}{4p} & \text{if} \quad p \in (0, 1/4), \\[2ex] \dfrac{1}{2(1-2p)} & \text{if} \quad p \in (1/4, 1/2). \end{cases}$$

□

Remark 4.3. For a case $N = 1$ and $p \leq 1/4$ we can have a duality-based proof, similar to the one we had for $N > 1$. From equation (4.6) we have $p(1) = p(N) = 4p$, using Theorem 2.2 we directly have

$$ET_{1:1:2} = d_1 \frac{1}{p(1)d_1} = \frac{1}{p(1)} = \frac{1}{4p}.$$

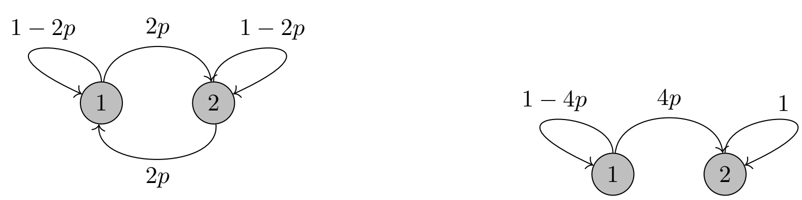
Let us have a closer look at this case. Note that both, a random walk on a circle and a resulting strong stationary dual, are the chains on two points. The ordering matrix is given by $\mathbf{C} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and we directly have

$$\mathbf{P}_X = \begin{pmatrix} 1-2p & 2p \\ 2p & 1-2p \end{pmatrix}, \quad \mathbf{C}^{-1}\overleftarrow{\mathbf{P}}_X\mathbf{C} = \begin{pmatrix} 1-4p & 2p \\ 0 & 1 \end{pmatrix}.$$

From (4.2) we obtain (with $\pi(1) = \pi(2) = 1/2$)

$$\mathbf{P}_X^* = \begin{pmatrix} 1-4p & 4p \\ 0 & 1 \end{pmatrix}.$$

The transitions of $\mathbf{X}$ and $\mathbf{X}^*$ are depicted in Figure 2.



Random walk $\mathbf{X}^*$ A sharp SSD $\mathbf{X}^*$

Figure 2: Case $d = 2N = 2$: Original random walk on a circle $\mathbf{X}$ (left), the corresponding sharp SSD $\mathbf{X}^*$ (right)

Of course, time to absorption in $\mathbf{X}^*$ has $Geo(4p)$ distribution, thus $ET = \frac{1}{4p}$.

Remark 4.4. Note that the assumptions on p in Lemma 4.1 (i.e., $p \leq 1/3$ for $N > 1$ and casee $p \leq 1/4, p \in (1/4, 1/2)$ for $N = 1$) are equivalent to non-negativity of the resulting matrix $\mathbf{P}_X^*$. In other words the assumption implies that $\mathbf{X}$ is $^{\uparrow}$-Möbius monotne (it is *if and only if* condition).

5 Proofs

5.1 Gambler's ruin problem, absorbing birth and death chain

5.1.1 Proof of Theorems 2.1 and 2.2

Proof of Theorem 2.1. Consider the birth and death chain with j and k ($j < k$) as recurrent absorbing states ($p(j) = q(j) = p(k) = q(k) = 0$). First step analysis yields (for $j < i < k$)

$$ET_{j:i:k} = p(i)(1 + ET_{j:i+1:k}) + q(i)(1 + ET_{j:i-1:k}) + (1 - q(i) - p(i))(1 + ET_{j:i:k}), \tag{5.1}$$

thus

$$ET_{j:i+1:k} = ET_{j:i:k} + \frac{q(i)}{p(i)}\left(ET_{j:i:k} - ET_{j:i-1:k} - \frac{1}{q(i)}\right). \tag{5.2}$$

Since $ET_{j:j:k} = 0$, we have:

$$ET_{j:j+2:k} = ET_{j:j+1:k}\left(1 + \frac{q(j+1)}{p(j+1)}\right) - \frac{q(j+1)}{p(j+1)}\frac{1}{q(j+1)}.$$

Recall that $d_s = \prod_{i=j+1}^{s} \frac{q(i)}{p(i)}$ (where $d_j = 1$), iterating the above equations yields:

$$ET_{j:i:k} = ET_{j:j+1:k}\sum_{s=j}^{i-1} d_s - \sum_{s=j+1}^{i-1}\left[d_s \sum_{m=j+1}^{s} \frac{1}{p(m)d_m}\right], \tag{5.3}$$

what can be checked by induction. Plugging (5.3) into (5.2) we have:

$$
\begin{aligned}
&ET_{j:i+1:k} = \\
&\quad ET_{j:j+1:k}\sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] \\
&\quad + \frac{q(i)}{p(i)} \quad \left(ET_{j:j+1:k}\sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] \right. \\
&\qquad\qquad \left. -ET_{j:j+1:k}\sum_{n=j}^{i-2} d_n - \sum_{n=j+1}^{i-2}[d_n \sum_{m=j+1}^{n} \frac{1}{p(s)d_s}] - \frac{1}{q(i)} \right) \\
&= ET_{j:j+1:k}\sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] \\
&\quad + \frac{q(i)}{p(i)} \left(ET_{j:j+1:k} d_{i-1} - d_{i-1}\sum_{s=j+1}^{i-1} \frac{1}{p(s)d_s} - d_i \frac{1}{d_i q(i)} \right) \\
&= ET_{j:j+1:k}\sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] + ET_{j:j+1:k} d_i - d_i \sum_{s=j+1}^{i-1} \frac{1}{p(s)d_s} - d_i \frac{1}{d_i p(i)} \\
&= ET_{j:j+1:k}\sum_{n=j}^{i} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] - d_i \sum_{s=j+1}^{i} \frac{1}{p(s)d_s} \\
&= ET_{j:j+1:k}\sum_{n=j}^{i} d_n - \sum_{n=j+1}^{i}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] .
\end{aligned}
$$

Since $ET_{j:k:k} = 0$, we have:

$$
0 = ET_{j:j+1:k}\sum_{n=j}^{k-1} d_n - \sum_{n=j+1}^{k-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right] \Rightarrow ET_{j:j+1:k} = \frac{\sum\limits_{n=j+1}^{k-1}\left[d_n \sum\limits_{s=j+1}^{n} \frac{1}{p(s)d_s}\right]}{\sum\limits_{n=j}^{k-1} d_n},
$$

thus

$$
ET_{j:i:k} = \frac{\sum_{n=j+1}^{k-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right]}{\sum_{n=j}^{k-1} d_n} \sum_{n=j}^{i-1} d_n - \sum_{n=j+1}^{i-1}\left[d_n \sum_{s=j+1}^{n} \frac{1}{p(s)d_s}\right],
$$

what was to be shown.

□

Proof of Theorem 2.2. Similarly as to proof of the Theorem 2.1 we consider birth and death chain on $\{j, \ldots, k\}$ $(j < k)$, however now only k is absorbing (*i.e.*, $p(k) = q(k) = q(j)0$, but $p(j) > 0$). For $i : j < i < k$ we can rewrite Eq. (5.1):

$$ET_{j:i:k} = p(i)(1 + ET_{j:i+1:k}) + q(i)(1 + ET_{j:i-1:k}) + (1 - q(i) - p(i))(1 + ET_{j:i:k}),$$

we have

$$ET_{j:i:k} = ET_{j:i+1:k} - \frac{q(i)}{p(i)}\left(ET_{j:i:k} - ET_{j:i-1:k} - \frac{1}{q(i)}\right). \tag{5.4}$$

However, for $i = j$ we have

$$ET_{j:j:k} = (1 - p(j))(1 + ET_{j:j:k}) + p(j)(1 + ET_{j:j+1:k}),$$

i.e.,

$$ET_{j:j:k} = \frac{1}{p(j)} + ET_{j:j+1:k}.$$

Recall that $d_s = \prod_{i=j+1}^{s} \frac{q(i)}{p(i)}$ (where $d_j = j$), iterating the above equations yields:

$$ET_{j:i:k} = ET_{j:i+1:k} + \sum_{s=j}^{i} \frac{d_i}{p(s)d_s}, \tag{5.5}$$

what can be checked by induction. Plugging (5.5) (for $i := i-1$) into (5.4) we have:

$$\begin{aligned} ET_{j:i:k} &= ET_{j:i+1:k} - \frac{q(i)}{p(i)}\left(ET_{j:i:k} - \left(ET_{j:i:k} + \sum_{s=j}^{i-1} \frac{d_{i-1}}{p(s)d_s}\right) - \frac{1}{q(i)}\right) \\ &= ET_{j:i+1:k} + \frac{d_i}{d_{i-1}}\left(\left(\sum_{s=j}^{i-1} \frac{d_{i-1}}{p(s)d_s}\right) + \frac{d_{i-1}}{p(i)d_i}\right) \\ &= ET_{j:i+1:k} + \sum_{s=j}^{i} \frac{d_i}{p(s)d_s}. \end{aligned}$$

Since $ET_{j:k:k} = 0$, we have:

$$ET_{j:k-1:k} = \sum_{s=1}^{k-1} \frac{d_{k-1}}{p(s)d_s}.$$

Iterating the above equations yields:

$$ET_{j:i:k} = \sum_{n=i}^{k-1}\left[d_n \sum_{s=1}^{n} \frac{1}{p(s)d_s}\right],$$

what was to be shown.

□

5.1.2 Proof of Lemma 2.4 and Theorem 2.3

Proof of Lemma 2.4. Denote by $f(n)$ lhs of (2.9) and by $h(n)$ its rhs. We will show that generating functions of f and h are equal. Let us start with $\mathfrak{g}_f(x)$, the generating function of f at x:

$\mathfrak{g}_f(x) =$

$$\begin{aligned}
\sum_{n=0}^{\infty} f(n)x^n &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n-k}{k}\left(-\frac{r}{(1+r)^2}\right)^k x^n = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\binom{n-k}{k}\left(-\frac{r}{(1+r)^2}\right)^k x^n \\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\binom{n-k}{k}\left(-\frac{r}{(1+r)^2}\right)^k x^n \\
&= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\binom{n}{k}\left(-\frac{r}{(1+r)^2}\right)^k x^{n+k} = \sum_{k=0}^{\infty}\left(-\frac{r}{(1+r)^2}\right)^k x^k \sum_{n=0}^{\infty}\binom{n}{k}x^n
\end{aligned}$$

Applying $\sum_{n=0}^{\infty}\binom{n}{k}x^n = \frac{x^k}{(1-x)^{k+1}}$ we have

$$\begin{aligned}
\mathfrak{g}_f(x) &= \sum_{k=0}^{\infty}\left(-\frac{r}{(1+r)^2}\right)^k x^k \frac{x^k}{(1-x)^{k+1}} = \frac{1}{1-x}\sum_{k=0}^{\infty}\left(\frac{-rx^2}{(1+r)^2(1-x)}\right)^k \\
&= \frac{1}{(1-x)}\frac{(1+r)^2(1-x)}{(1+r)^2(1-x)+rx^2} = \frac{(1+r)^2}{(1+r)^2(1-x)+rx^2}.
\end{aligned}$$

On the other hand, the generating function of h is following:

$$\begin{aligned}
\mathfrak{g}_h(x) &= \sum_{n=0}^{\infty} h(n)x^n = \sum_{n=0}^{\infty}\frac{1-r^{n+1}}{(1+r)^n(1-r)}x^n = \frac{1}{(1-r)}\left(\sum_{n=0}^{\infty}\frac{1}{(1+r)^n}x^n - \sum_{n=0}^{\infty}\frac{r^n}{(1+r)^n}x^n\right) \\
&= \frac{1}{(1-r)}\left(\sum_{n=0}^{\infty}\frac{1}{(1+r)^n}x^n - \sum_{n=0}^{\infty}\frac{r^n}{(1+r)^n}x^n\right) = \frac{1}{(1-r)}\left(\frac{1+r}{1+r-x} - r\frac{1+r}{1+r-xr}\right) \\
&= \frac{1+r}{(1-r)}\frac{1+r-xr-r-r^2-xr}{(1+r-x)(1+r-xr)} = \frac{1+r}{(1-r)}\frac{(1+r)(1-r)}{(1+r)^2-(1+r)(x+xr)+x^2r} \\
&= \frac{(1+r)^2}{(1+r)^2(1-x)+rx^2},
\end{aligned}$$

thus $\mathfrak{g}_h(x) = \mathfrak{g}_f(x)$, what finishes the proof.

□

The following lemma will be needed in the proof of Theorem 2.3.

Lemma 5.1. Consider the gambler's ruin problem with general rates $\mathbf{p}, \mathbf{q}$. Define

$$\begin{aligned}
a_i &= -\frac{\rho_{0:i:i+1}}{p(i)}, \\
b_i &= \frac{(p(i)+q(i))\rho_{0:i:i+1}}{p(i)}, \\
c_i &= -\frac{q(i)}{p(i)}\rho_{0:i-1:i+1}.
\end{aligned}$$

Then, for all $N \geq 1$ we have

$$\prod_{j=2}^{N}\begin{pmatrix} b_j & c_j & a_j \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & a_1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & A_N \\ 1 & 0 & A_{N-1} \\ 0 & 0 & 1 \end{pmatrix},$$

where

$$A_M = -\sum_{n=1}^{M} \frac{1}{p(n)} \rho_{0:n:M+1} \sum_{k=0}^{\lfloor (M-n)/2 \rfloor} \xi_k^{n+1,M},$$

$\xi_k^{n+1,M}$ was defined in (2.4).

Proof. Recall that $\mathbf{j}_k^{n,m}$ was defined in (2.2) as

$$\mathbf{j}_k^{n,m} = \big\{\{j_1, j_2, \ldots, j_k\} : j_1 \geq n+1, j_k \leq m, j_i \leq j_{i+1} - 2 \text{ for } i \in \{1, k-1\}\big\}.$$

For given $\mathbf{p}, \mathbf{q}$, b_n, c_n and $\mathbf{j} \in \mathbf{j}_k^{n,m}$ define

$$D_{\mathbf{j}}^{n,m} = b_n b_{n+1} \ldots b_{j_1-2} c_{j_1} b_{j_1+1} b_{j_1+2} \ldots b_{j_2-2} c_{j_2} \ldots b_{j_{k-1}+1} b_{j_{k-1}+2} \ldots b_{j_k-2} c_{j_k} b_{j_k+1} b_{j_k+2} \ldots b_m$$

and let

$$S_k^{n,m} = \sum_{\mathbf{j} \in \mathbf{j}_k^{n,m}} D_{\mathbf{j}}^{n,m}.$$

Let

$$\begin{aligned} \alpha_i &= -\frac{1}{p(i)}, \\ \beta_i &= \frac{(p(i)+q(i))}{p(i)} = 1 + r(i), \\ \gamma_i &= -\frac{q(i)}{p(i)} = -r(i). \end{aligned}$$

$D_{\mathbf{j}}^{n,m}$ can be rewritten as

$$\begin{aligned} D_{\mathbf{j}}^{n,m} &= \rho_{0:n:m+1} \beta_n \beta_{n+1} \cdots \beta_{j_1-2} \gamma_{j_1} \beta_{j_1+1} \beta_{j_1+2} \cdots \beta_{j_2-2} \gamma_{j_2} \cdots \\ &\quad \cdot \beta_{j_{k-1}+1} \beta_{j_{k-1}+2} \ldots \beta_{j_k-2} \gamma_{j_k} \beta_{j_k+1} \beta_{j_k+2} \cdots \beta_m \\ &= (-1)^k \prod_{s \in \mathbf{j}} r(s) \prod_{s \in \{n,\ldots,m\} \setminus \mathbf{j} \cup \mathbf{j}-1} 1 + r(s) = \rho_{0:n:m+1} \delta_{\mathbf{j}}^{n,m}. \end{aligned}$$

Thus $S_k^{n,m} = \sum_{\mathbf{j} \in \mathbf{j}_k^{n,m}} D_{\mathbf{j}}^{n,m} = \rho_{0:n:m+1} \sum_{\mathbf{j} \in \mathbf{j}_k^{n,m}} \delta_{\mathbf{j}}^{n,m} =: \rho_{0:n:m+1} \xi_k^{n,m}$ and A_M can be rewritten as

$$A_M = \sum_{n=1}^{M} a_n \sum_{k=0}^{\lfloor (M-n)/2 \rfloor} S_k^{n+1,M}.$$

We will show this by induction.

- For $M = 1$ we have

$$A_1 = \sum_{n=1}^{1} a_n \sum_{k=0}^{\lfloor (1-n)/2 \rfloor} S_k^{n+1,1} = a_1 \sum_{k=0}^{\lfloor 0/2 \rfloor} S_k^{2,1} = a_1 S_0^{2,1} = a_1.$$

- For $N \geq M \geq 2$ assuming $A_M = \sum_{n=1}^{M} a_n \sum_{k=0}^{\lfloor (M-n)/2 \rfloor} S_k^{n+1,M}$ we shall prove that $A_{N+1} = b_{N+1}A_N + c_{N+1}A_{N-1} + a_{N+1}$. We have
  $b_{N+1}A_N + c_{N+1}A_{N-1} + a_{N+1} =$

$$\begin{aligned}
&= b_{N+1}\sum_{n=1}^{N} a_n \sum_{k=0}^{\lfloor (N-n)/2 \rfloor} S_k^{n+1,N} + c_{N+1}\sum_{n=1}^{N-1} a_n \sum_{k=0}^{\lfloor (N-n-1)/2 \rfloor} S_k^{n+1,N-1} + a_{N+1} \\
&= \sum_{n=1}^{N} a_n \sum_{k=0}^{\lfloor (N-n)/2 \rfloor} b_{N+1} \sum_{\mathbf{j}_k^{n+1,N}} D^{n+1,N}_{\mathbf{j}_k^{n+1,N}} + \sum_{n=1}^{N-1} a_n \sum_{k=0}^{\lfloor (N-n-1)/2 \rfloor} c_{N+1} \sum_{\mathbf{j}_k^{n+1,N-1}} D^{n+1,N-1}_{\mathbf{j}_k^{n+1,N-1}} + a_{N+1} \\
&= \sum_{n=1}^{N} a_n \sum_{k=0}^{\lfloor (N+1-n)/2 \rfloor} \sum_{\mathbf{j}_k^{n+1,N+1}:j_k\neq N+1} D^{n+1,N+1}_{\mathbf{j}_k^{n+1,N+1}} \\
&\quad + \sum_{n=1}^{N} a_n \sum_{k=0}^{\lfloor (N+1-n)/2 \rfloor} \sum_{\mathbf{j}_k^{n+1,N+1}:j_k= N+1} D^{n+1,N+1}_{\mathbf{j}_k^{n+1,N+1}} + a_{N+1} \\
&= \sum_{n=1}^{N+1} a_n \sum_{k=0}^{\lfloor (N+1-n)/2 \rfloor} \sum_{\mathbf{j}_k^{n+1,N+1}} D^{n+1,N+1}_{\mathbf{j}_k^{n+1,N+1}} = \sum_{n=1}^{N+1} a_n \sum_{k=0}^{\lfloor (N+1-n)/2 \rfloor} S_k^{n+1,N+1} = A_{N+1}
\end{aligned}$$

what finishes the proof.

□

Proof of Theorem 2.3. First step analysis yields (for $N > i > 1$):

$$\begin{aligned}
EW_{0:i:N} &= (1+EW_{0:i-1:N})P(X_1 = i-1|X_0 = i, X_T = N) \\
&\quad +(1+EW_{0:i:N})P(X_1 = i|X_0 = i, X_T = N) \\
&\quad +(1+EW_{0:i+1:N})P(X_1 = i+1|X_0 = i, X_T = N).
\end{aligned}$$

We have $EW_{0:N:N} = 0$ and for simplicity we also set $EW_{0:0:N} = 0$. We have

$$\begin{aligned}
P(X_1 = i-1|X_0 = i, X_T = N) &= \tfrac{P(X_1=i-1|X_0=i)P(X_T=N|X_1=i-1)}{P(X_T=N|X_0=i)} = \frac{q(i)\rho_{0:i-1:N}}{\rho_{0:i:N}} = q(i)\rho_{0:i-1:i}, \\
P(X_1 = i|X_0 = i, X_T = N) &= \frac{(1-p(i)-q(i))\rho_{0:i:N}}{\rho_{0:i:N}} = 1-p(i)-q(i), \\
P(X_1 = i+1|X_0 = i, X_T = N) &= \frac{p(i)\rho_{0:i+1:N}}{\rho_{0:i:N}} = p(i)\rho_{0:i+1:i}.
\end{aligned}$$

For $i = 1$ we have

$$EW_{0:1:N} = [1+EW_{0:1:N}](1-p(1)-q(1)) + [1+EW_{0:2:N}]p(1)\rho_{0:2:1},$$

thus

$$EW_{0:2:N} = \frac{(p(1)+q(1)-1)\rho_{0:1:2}}{p(1)} - 1 + \frac{(p(1)+q(1))\rho_{0:1:2}}{p(1)}EW_{0:1:N}.$$

For $1 \leq i \leq N$ we have

$$EW_{0:i:N} = (1+EW_{0:i-1:N})q(i)\rho_{0:i-1:i} + (1+EW_{0:i:N})(1-p(i)-q(i)) + (1+EW_{0:i+1:N})p(i)\rho_{0:i+1:i} \tag{5.6}$$

and

$$\begin{aligned}
EW_{0:i+1:N} &= \frac{(p(i)+q(i))\rho_{0:i:i+1}}{p(i)} - \frac{q(i)}{p(i)}\rho_{0:i-1:i+1} - 1 - \frac{\rho_{0:i:i+1}}{p(i)} \\
&\quad + \frac{(p(i)+q(i))\rho_{0:i:i+1}}{p(i)} EW_{0:i:N} - \frac{q(i)}{p(i)}\rho_{0:i-1:i+1} EW_{0:i-1:N}, \\
&= b_i + c_i - 1 + a_i + b_i EW_{0:i:N} + c_i EW_{0:i-1:N} \\
&\overset{(*)}{=} a_i + b_i EW_{0:i:N} + c_i EW_{0:i-1:N}, \qquad (5.7)
\end{aligned}$$

where a_i, b_i, c_i were defined in Lemma 5.1 and in $(*)$ we used the fact that

$$\begin{aligned}
b_i + c_i &= \frac{(p(i)+q(i))\rho_{0:i:i+1}}{p(i)} - \frac{q(i)}{p(i)}\rho_{0:i:i+1} \\
&= \frac{p(i)+q(i)}{p(i)} \frac{\sum_{n=1}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} - \frac{q(i)}{p(i)} \frac{\sum_{n=1}^{i-1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} \\
&= \frac{\sum_{n=1}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right) + \frac{q(i)}{p(i)}\sum_{n=1}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right) - \frac{q(i)}{p(i)}\sum_{n=1}^{i-1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} \\
&= \frac{\sum_{n=1}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right) + \frac{q(i)}{p(i)}\sum_{n=i}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} \\
&= \frac{\sum_{n=1}^{i}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right) + \prod_{k=1}^{i}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} = \frac{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{i+1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} = 1.
\end{aligned}$$

Equations (5.6) and (5.7) can be written in a matrix form:

$$\begin{pmatrix} EW_{0:i+1:N} \\ EW_{0:i:N} \\ 1 \end{pmatrix} = \begin{pmatrix} b_i & c_i & a_i \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} EW_{0:i:N} \\ EW_{0:i-1:N} \\ 1 \end{pmatrix}. \qquad (5.8)$$

Note that $c_1 = -\frac{q_1}{p_1}W_0^2 = -\frac{q_1}{p_1}0 = 0$ and

$$b_1 = \frac{(p(1)+q(1))\rho_{0:1:2}}{p(1)} = \frac{p(1)+q(1)}{p(1)} \frac{\sum_{n=1}^{1}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)}{\sum_{n=1}^{2}\prod_{k=1}^{n-1}\left(\frac{q(k)}{p(k)}\right)} = \frac{1+\frac{q(1)}{p(1)}}{1}\frac{1}{1+\frac{q(1)}{p(1)}} = 1,$$

thus using (5.8) recursively we obtain

$$\begin{pmatrix} 0 \\ EW_{0:N-1:N} \\ 1 \end{pmatrix} = \begin{pmatrix} EW_{0:N:N} \\ EW_{0:N-1:N} \\ 1 \end{pmatrix} = \prod_{j=2}^{N-1} \begin{pmatrix} b_j & c_j & a_j \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & a_1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} EW_{0:1:N} \\ EW_{0:0:N} \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & A_{N-1} \\ 1 & 0 & A_{N-2} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} EW_{0:1:N} \\ 0 \\ 1 \end{pmatrix},$$

where A_N is given in Lemma 5.1, what implies

$$EW_{0:1:N} = -A_{N-1}$$

and thus proves (2.6). Equation (2.5) follows from the fact that $W_{0:1:N} \stackrel{(distr)}{=} W_{0:1:i} + W_{0:i:N}$ (Markov property, $W_{0:1:i}$ and $W_{0:i:N}$ are independent).

□

5.2 Random walk on a polygon

5.2.1 Proof of Theorem 3.1

Proof of eq. (3.1) . Let F_i denote the event that at the first time we leave state i (recall, ties are allowed) we move clockwise. Similarly, let F_i^c denotes the event that at the first time we leave state i we move counterclockwise. We have

$$\begin{aligned} P(F_i) &= \frac{p(i)}{p(i)+q(i)} = \frac{1}{1+r(i)}, \\ P(F_i^c) &= \frac{q(i)}{p(i)+q(i)} = \frac{r(i)}{1+r(i)} \end{aligned}$$

and

$$P(A_i) = P(F_i)P(A_i|F_i) + P(F_i^c)P(A_i|F_i^c) = \frac{1}{1+r(i)}P(A_i|F_i) + \frac{r(i)}{1+r(i)}P(A_i|F_i^c).$$

- For $P(A_i|F_i)$ we have: we start at $i+1$ and we have to reach $i-1$ before reaching i. This is the probability of winning in the game $G(\mathbf{p}, \mathbf{q}, i, i+1, i-1)$. We thus have

$$P(A_i|F_i) = \rho_{i:i+1:i-1} = \frac{1}{\sum_{n=i+1}^{i-1} \prod_{s=i+1}^{n-1} r(s)}.$$

- Similarly for $P(A_i|F_i^c)$ we have: we start at $i-1$, and we have to reach $i+1$ before reaching i which corresponds to losing in the game $G(\mathbf{p}, \mathbf{q}, i+1, i-1, i)$. We thus have

$$P(A_i|F_i^c) = 1 - \rho_{i+1:i-1:i} = 1 - \frac{\sum_{n=i+2}^{i-1} \prod_{s=i+2}^{n-1} r(s)}{\sum_{n=i+2}^{i} \prod_{s=i+2}^{n-1} r(s)} = \frac{\prod_{s=i+2}^{i-1} r(s)}{\sum_{n=i+2}^{i} \prod_{s=i+2}^{n-1} r(s)} = \frac{1}{\sum_{n=i+2}^{i} \prod_{s=n}^{i-1} \left(\frac{1}{r(s)}\right)}.$$

Finally

$$\begin{aligned} P(A_i) &= \frac{1}{(1+r(i))\sum_{n=i+1}^{i-1}\prod_{s=i+1}^{n-1} r(s)} + \frac{r(i)}{(1+r(i))\sum_{n=i+2}^{i}\prod_{s=n}^{i-1}\left(\frac{1}{r(s)}\right)} \\ &= \frac{1}{(1+r(i))\sum_{n=i+1}^{i-1}\prod_{s=i+1}^{n-1} r(s)} + \frac{1}{(1+r(i))\sum_{n=i+2}^{i}\prod_{s=n}^{i}\left(\frac{1}{r(s)}\right)}. \end{aligned}$$

□

Proof of eq. (3.2). Let us define $T_1 = \inf\{t : X_t = j-1 \vee X_t = j+1 | X_0 = i\}$ and consider separately two cases when at T_1 we are at $j-1$ or $j+1$. The first one corresponds to winning, whereas the second one corresponds to losing in the game $G(\mathbf{p}, \mathbf{q}, j+1, i, j-1)$. The winning probability is

$$\rho_{j+1:i:j-1}.$$

In the first case (when we get to the $j-1$ before $j+1$) vertex j will be the last one if we reach $j+1$ earlier - this can be interpreted as losing in the game $G(\mathbf{p}, \mathbf{q}, j+1, j-1, j)$, what happens with probability:

$$1-\rho_{j+1:j-1:j} = 1 - \frac{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=j+2}^{n-1} r(s)} = \frac{\prod_{s=j+2}^{j-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=j+2}^{n-1} r(s)}.$$

In the second case (when we get to the $j+1$ before $j-1$) vertex j will be the last one if we reach $j-1$ earlier - this can be interpreted as winning in the game $G(\mathbf{p}, \mathbf{q}, j, j+1, j-1)$, what happens with probability:

$$\rho_{j:j+1:j-1} = \frac{1}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)}.$$

Finally:

$$
\begin{aligned}
P(L_{i,j}) &= (1-\rho_{j+1:i:j-1})\rho_{j:j+1:j-1} + \rho_{j+1:i:j-1}(1-\rho_{j+1:j-1:j}) \\
&= \left(1 - \frac{\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)}\right) \frac{1}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)} + \frac{\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)} \frac{\prod_{s=j+2}^{j-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=j+2}^{n-1} r(s)} \\
&= \frac{\sum_{n=i+1}^{j-1}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)} \frac{1}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)} + \frac{\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)} \frac{\prod_{s=j+2}^{j-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=j+2}^{n-1} r(s)} \\
&= \frac{1}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)} \left(\frac{\sum_{n=i+1}^{j-1}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)} + \frac{\left(\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)\right)\left(\prod_{s=j+2}^{j-1} r(s)\right)}{\sum_{n=j+2}^{j}\prod_{s=j+2}^{n-1} r(s)} \right) \\
&= \frac{1}{\sum_{n=j+2}^{j-1}\prod_{s=j+2}^{n-1} r(s)} \left(\frac{\sum_{n=i+1}^{j-1}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+1}^{j-1}\prod_{s=j+1}^{n-1} r(s)} + \frac{\sum_{n=j+2}^{i}\prod_{s=j+2}^{n-1} r(s)}{\sum_{n=j+2}^{j}\prod_{s=n}^{j-1} \frac{1}{r(s)}} \right).
\end{aligned}
$$

□

Proof of eqs. (3.3), (3.4) and (3.5) . Let us start with the expectation of $V_{i,j}$ – number of steps to visit all vertices starting at i when j is the last visited vertex. As earlier, let $T_1 = \inf\{t : X_t = j-1 \vee X_t = j+1\}$. We have two cases:

- If $X_{T_1} = j-1$ (and j was the last visited vertex) then the expected game time consists of: expected time to win in $G(\mathbf{p}, \mathbf{q}, j+1, i, j-1)$, expected time to lose in $G(\mathbf{p}, \mathbf{q}, j+1, j-1, j)$ and expected duration of the game $G(\mathbf{p}, \mathbf{q}, j, j+1, j)$. That is:
$$EW_{j+1:i:j-1} + EB_{j+1:j-1:j} + ET_{j:j+1:j}$$
- If $X_{T_1} = j+1$ (and j was last visited vertex) then the expected game time consists of: expected time to lose in $G(\mathbf{p}, \mathbf{q}, j+1, i, j-1)$, expected time to win in $G(\mathbf{p}, \mathbf{q}, j, j+1, j-1)$ and expected duration of the game $G(\mathbf{p}, \mathbf{q}, j, j-1, j)$. That is:
$$EB_{j+1:i:j-1} + EW_{j:j+1:j-1} + ET_{j:j-1:j}$$

Now, conditioning on X_{T_1}, we obtain:

$$
\begin{aligned}
EV_{i,j} = \rho_{j+1:i:j-1}\left(EW_{j+1:i:j-1} + EB_{j+1:j-1:j} + ET_{j:j+1:j}\right) \\
+(1-\rho_{j+1:i:j-1})\left(EB_{j+1:i:j-1} + EW_{j:j+1:j-1} + ET_{j:j-1:j}\right).
\end{aligned}
$$

Equations (3.4) and (3.5) are simply obtained by conditioning on the states.

□

Bibliography

[AA10] Søren Asmussen and Hansjörg Albrecher. *Ruin Probabilities.* World Scientific Pubishing, Singapore, 2010.

[AD86] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5):333–348, 1986.

[AD87] David Aldous and Persi Diaconis. Strong Uniform Times and Finite Random Walks. *Advances in Applied Mathematics*, 97:69–97, 1987.

[AS09] Søren Asmussen and Karl Sigman. Monotone Stochastic Recursions and their Duals. *Probability in the Engineering and Informational Sciences*, 10(01):1, jul 2009.

[BW77] W. A. Beyer and M. S. Waterman. Symmetries for Conditioned Ruin Problems. *Mathematics Magazine*, 50(1):42–45, 1977.

[CLR01] George Casella, Michael Lavine, and Christian P. Robert. Explaining the Perfect Sampler. *The American Statistician*, 55(4):299–305, 2001.

[Com70] Louis Comtet. *Analyse combinatoire.* Presses universitaires de France, Paris, 1970.

[CSV18] Nicolas Champagnat, René Schott, and Denis Villemonais. Probabilistic non-asymptotic analysis of distributed algorithms. *arXiv e-prints*, page arXiv:1802.02644, 2018.

[DF90a] Persi Diaconis and James Allen Fill. Examples for the Theory of Strong Stationary Duality with Countable State Spaces. *Probability in the Engineering and Informational Sciences*, 4(2):157–180, 1990.

[DF90b] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18(4):1483–1522, 1990.

[Dim01] Xeni K. Dimakos. A Guide to Exact Simulation. *International Statistical Review*, 69(1):27–48, 2001.

[DSC06] Persi Diaconis and L Saloff-Coste. Separation cut-offs for birth and death chains. *The Annals of Applied Probability*, 16(4):2098–2122, 2006.

[ES00] Mohamed A. El-Shehawey. Absorption probabilities for a random walk between two partially absorbing boundaries: I. *Journal of Physics A: Mathematical and General*, 33(49):9005–9013, 2000.

[ES09] Mohamed A. El-Shehawey. On the gambler's ruin problem for a finite Markov chain. *Statistics & Probability Letters*, 79(14):1590–1595, jul 2009.

[Fil98] James Allen Fill. An interruptible algorithm for perfect sampling via Markov chains. *The Annals of Applied Probability*, 8(1):131–162, 1998.

[Fil09a] James Allen Fill. On hitting times and fastest strong stationary times for skip-free and more general chains. *Journal of Theoretical Probability*, 22(3):587–600, 2009.

[Fil09b] James Allen Fill. The Passage Time Distribution for a Birth-and-Death Chain: Strong Stationary Duality Gives a First Stochastic Proof. *Journal of Theoretical Probability*, 22(3):543–557, 2009.

[FL14] James Allen Fill and Vince Lyzinski. Hitting times and interlacing eigenvalues: a stochastic approach using intertwinings. *Journal of Theoretical Probability*, 27(3):954–981, 2014.

[FM01] James Allen Fill and Motoya Machida. Stochastic monotonicity and realizable monotonicity. *The Annals of Probability*, 29(2):938–978, 2001.

[FMMR00] James Allen Fill, Motoya Machida, Duncan J. Murdoch, and Jeffrey S. Rosenthal. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms*, 17(3-4):290–316, 2000.

[GMZ12] Yu Gong, Yong Hua Mao, and Chi Zhang. Hitting Time Distributions for Denumerable Birth and Death Processes. *Journal of Theoretical Probability*, 25(4):950–980, 2012.

[HM16] Thierry Huillet and Servet Martínez. On Möbius Duality and Coarse-Graining. *Journal of Theoretical Probability*, 29(1):143–179, 2016.

[HN98] O Häggström and K Nelander. Exact sampling from antimonotone systems. *Statistica Neerlandica*, 52(3):360–380, 1998.

[Hub03] Mark Huber. A Bounding Chain for Swendsen-Wang. *Random Structures and Algorithms*, 22(1):43–59, 2003.

[Kei79] Julian Keilson. *Rarity and Exponentiality.* In: Keilson J. (eds) Markov Chain Models – Rarity and Exponentiality. Applied Mathematical Sciences, vol 28. Springer, New York, NY, 1979.

[Ken98] Wilfrid S. Kendall. Perfect Simulation for the Area-Interaction Point Process. In L. Accardi and C. C. Heyde, editors, *Probability towards 2000*, volume 128 of *Lecture Notes in Statistics*, pages 218–234. Springer-Verlag, Berlin, 1998.

[KKO77] T. Kamae, U. Krengel, and G. L. O'Brien. Stochastic Inequalities on Partially Ordered Spaces. *The Annals of Probability*, 5(6):899–912, 1977.

[KM00] Wilfrid S. Kendall and Jesper Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 2000.

[KP02] Andrej Kmet and Marko Petkovšek. Gambler's Ruin Problem in Several Dimensions. *Advances in Applied Mathematics*, 28(2):107–118, feb 2002.

[Lef08] Mario Lefebvre. The gambler's ruin problem for a Markov chain related to the Bessel process. *Statistics & Probability Letters*, 78(15):2314–2320, 2008.

[Len09a] Tamás Lengyel. Gambler's ruin and winning a series by m games. *Annals of the Institute of Statistical Mathematics*, 63(1):181–195, jan 2009.

[Len09b] Tamás Lengyel. The conditional gambler's ruin problem with ties allowed. *Applied Mathematics Letters*, 22(3):351–355, 2009.

[Lig04] Thomas M. Liggett. *Interacting Particle Systems.* Springer-Verlag, Berlin, Heidelberg, 2004.

[Lin52] D. V. Lindley. The theory of queues with a single server. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(02):277–289, oct 1952.

[LM17] Paweł Lorek and Piotr Markowski. `The Julia Language` module for checking monotonicities in Markov chains. *GitHub repository*, https://github.com/lorek/MarkovChains_monotonicities/. 2017.

[Lor17] Paweł Lorek. Generalized Gambler's Ruin Problem: Explicit Formulas via Siegmund Duality. *Methodology and Computing in Applied Probability*, 19(2):603–613, 2017.

[Lor18] Paweł Lorek. Siegmund duality for markov chains on partially ordered state spaces. *Probability in the Engineering and Informational Sciences*, 32(4):495–521, 2018.

[LS12a] Paweł Lorek and Ryszard Szekli. Strong stationary duality for Möbius monotone Markov chains. *Queueing Systems*, 71(1–2):79–95, 2012.

[LS12b] Paweł Lorek and Ryszard Szekli. Strong stationary duality for Möbius monotone Markov chains. *Queueing Systems*, 71(1-2):79–95, 2012.

[LS16] Paweł Lorek and Ryszard Szekli. Strong stationary duality for Möbius monotone Markov chains: examples. *Probability and Mathematical Statistics*, 36(1):75–97, 2016.

[Mac01] Motoya Machida. *Stochastic monotonicity and realizable monotonicity*. Phd Thesis, The Johns Hopkins University, 2001.

[Mud65] V.I. Mudrov. An algorithm for numbering combinations. *USSR Computational Mathematics and Mathematical Physics*, 5(4):280–282, jan 1965.

[MZ17] Yong-Hua Mao and Chi Zhang. Hitting Time Distributions for Birth–Death Processes With Bilateral Absorbing Boundaries. *Probability in the Engineering and Informational Sciences*, 31(3):345–356, 2017.

[Par62] Emanuel Parzen. *Stochastic Processes*. Holden-Day, Inc., 1962.

[PW96] James Gary Propp and David Bruce Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.

[Ros09] Sheldon M. Ross. A simple solution to a multiple player gambler's ruin problem. *American Mathematical Monthly*, 116(1):77–81, 2009.

[Rot64] Gian-Carlo Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Probability Theory and Related Fields*, 368:340–368, 1964.

[RS04] A.L. Rocha and F. Stern. The asymmetric n-player gambler's ruin problem with equal initial fortunes. *Advances in Applied Mathematics*, 33(3):512–530, oct 2004.

[Sar06] Jyotirmoy Sarkar. Random walk on a polygon. In Connie Sun, Jiayang and DasGupta, Anirban and Melfi, Vince and Page, editor, *Lecture Notes–Monograph Series*, volume 50, pages 31–43. Institute of Mathematical Statistics, 2006.

[Sie76] David Siegmund. The Equivalence of Absorbing and Reflecting Barrier Problems for Stochastically Monotone Markov Processes. *The Annals of Probability*, 4(6):914–924, 1976.

[SM17] Jyotirmoy Sarkar and Saran Ishika Maiti. Symmetric Random Walks on Regular Tetrahedra, Octahedra, and Hexahedra. *Calcutta Statistical Association Bulletin*, 69(1):110–128, may 2017.

[Ste75] Frederick Stern. Conditional Expectation of the Duration in the Classical Ruin Problem. *Mathematics Magazine*, 48(48):200–203, 1975.

[Str65] V. Strassen. The Existence of Probability Measures with Given Marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[Tzi19] Achillefs Tzioufas. The several dimensional gambler's ruin problem. *Markov Processes And Related Fields*, 25(1):101–123, 2019.