Eötvös Loránd University
Institute of Mathematics
Department of Probability Theory and Statistics

Ph.D. thesis

Convergence analysis of Markovian systems

Balázs Gerencsér

Doctoral School: Mathematics
Director: Miklós Laczkovich
member of the Hungarian Academy of Sciences

Doctoral Program: Pure Mathematics
Director: András Szűcs
corresponding member of the Hungarian Academy of Sciences

Supervisor: Gábor Tusnády
member of the Hungarian Academy of Sciences

October 2012

Contents

Notations

Δ		probability of a single mutant gene not causing the disorder.
ρ		probability of a single mutant gene not causing selection.
μ		parameter of the Poisson distribution corresponding to the mutation.
λ		parameter of the Poisson distribution of mutant genes.
		in the stationary population.
p		population probability of the disorder.
q_S		conditional probability of a sibling of a malformed child being affected.
$\mathcal{X}$		the state space of the Markov chain.
n	$=$	$\lvert\mathcal{X}\rvert$, the size of the state space.
S^C	$=$	$\mathcal{X} \setminus S$, the complement of the set S.
$\mathcal{P}(\mathcal{X})$		the set of probability distributions on the finite state space $\mathcal{X}$.
$\lVert\nu\rVert_{\mathrm{TV}}$	$=$	$\max_{A\subseteq\mathcal{X}} \lvert\nu(A)\rvert$, the total variation norm.
$t_{\mathrm{mix}}(P,\varepsilon)$	$=$	$\max_{\sigma\in\mathcal{P}(\mathcal{X})} \min\left\{k : \lVert\sigma P^k - \pi\rVert_{\mathrm{TV}} \leq \varepsilon\right\}$, the mixing time.
π_i		stationary probability of state i.
π_*	$=$	$\min_i \pi_i$.
p_{ij}		transition probability from state i to state j.
Φ		the conductance of the Markov chain.
Q		the reversible part of the transition matrix.
q_i	$=$	$Q_{i-1,i} = Q_{i,i-1}$.
rR		the non-reversible part of the transition matrix.

Acknowledgments

In the first place I would like to express my sincere thanks to my supervisor, Gábor Tusnády for the great amount of time he expended on me and his continuous support throughout my PhD studies. I would also like to emphasize and acknowledge the wide variety of challenging mathematical problems he raised and we worked on.

I would also like to express my gratitude to John Tsitsiklis for hosting me at the Massachusetts Institute of Technology for nine months. I am also grateful for the Fulbright Commission and their helpful staff for making this visit possible.

I am also thankful to Balázs Ráth for the fruitful collaboration playing an important role in the research leading to this thesis.

I would like to thank Villő Csiszár for her helpful and inspiring comments during the years of my PhD research. I would also like to thank György Michaletzky for his support on performing the numerical investigations of the multifactorial inheritance models.

Also, I am thankful to Péter Csikvári and Viktor Harangi for reading the thesis and for their helpful comments.

Last but not least I am thankful for my family for the constant support and encouragement I received from them.

Chapter 1

Introduction

One could claim that Markov processes are the simplest stochastic processes in time and that we completely understand their behavior. Indeed, apart from i.i.d. processes, the Markov property governs the simplest dependence structure for a random process in time. There is a huge literature covering most of the naturally occurring cases satisfying reasonable conditions, see e.g. the detailed book of Meyn and Tweedie [43]. Still, when we try to deal with slightly unusual processes even the simplest questions become surprisingly difficult. We deal with some problems falling into this category in this thesis.

In Chapter 2, we look at biological inheritance as a Markov process. Indeed, the genetic information of a child depends on the genetic information of his ancestors only through his parents. The catch is that a child has two parents instead of one, consequently the family tree is not simply a chain. We present what one can say about the long term behavior of such processes, both in general and specially for the processes arising in the biological models. We also provide statistical investigation on fitting the model in our focus to Hungarian population data. Chapter 2 is based on the paper [27]. This is based on the joint work with Gábor Tusnády, my advisor and Balázs Ráth, who suggested the ideas for the model and the proof of Theorem 2.9. It turned out that a similar model has been previously investigated by Dawson [13].

In Chapter 3, we work on mixing time estimates. Although most of the times people search for upper bounds on mixing times of certain chains, we now look for the best chain within a class. This involves getting a universal lower bound on the mixing time for the target class. We also relax the reversibility condition which would give us technical convenience but also pose an unnecessary restriction on the chain. Chapter 3 is based on the papers [25] and [26]. This part of my research has been done with the help and supervision of John Tsitsiklis.

The papers used have been reworked and expanded for the purpose of this thesis.

Chapter 2

Convergence of Bi-Markov processes

A discrete time Markov chain can be viewed as an iteration of independent random functions $f_t : \mathcal{X} \to \mathcal{X}$ for $t = 1, 2, \ldots$. What happens if we iterate $\mathcal{X}^2 \to \mathcal{X}$ functions instead? In other words, the current state depends on two *parents* rather than one.

Such systems come up naturally in the study of genetic inheritance, hence the term *parent*. We first investigate the general properties of these so called *Bi-Markov processes*, then we show how to use them to model biological processes. Introducing these processes this way implies a discrete time version. We note that the continuous time version has been investigated by Hatvani, Toókos and Tusnády [29].

For a realization of a Markov chain, we often think of a series of random variables $X_0, X_1, \ldots, X_t, \ldots$ satisfying the proper conditional independence condition. For Bi-Markov processes, a similar realization quickly becomes cumbersome, when checking the past of X_t, we need 2^t ancestors at time 0. If t is allowed to increase to infinity, we need infinite copies of X_s at each time step s.

As an alternative, we might follow the evolution of the distribution of the states. Let us denote the probability of the system being at state i at time t by $p_i(t)$. For Markov chains, we get a recursion

$$p_i(t+1) = \sum_j \tilde{h}(j,i) p_j(t),$$

with some $\tilde{h}(j,i)$ transition probabilities. The analogous formula for Bi-Markov processes:

$$p_i(t+1) = \sum_{j,k} h(j,k,i) p_j(t) p_k(t).$$

Now the coefficients $h(j,k,i)$ denote the probability of arriving at state i after the pair of states (j,k). The straightforward conditions on h are

$$\begin{aligned} h(j,k,i) &\geq 0 && \forall j,k,i \in \mathcal{X}, \\ \sum_i h(j,k,i) &= 1 && \forall j,k \in \mathcal{X}. \end{aligned} \tag{2.1}$$

The questions for these processes are similar to the ones for Markov chains. We call a Markov or Bi-Markov process ergodic if the distribution of the states approaches a certain stationary distribution independent from the initialization. Under what conditions can we ensure the ergodicity of a Bi-Markov process? Can it happen that the distribution changes periodically? Maybe even in a chaotic way?

The existence of a stationary distribution for Markov chains is provided by an application of the Perron-Frobenius theorem:

Theorem 2.1 (Perron-Frobenius)**.** *Every finite Markov chain has a stationary distribution. It is unique for irreducible chains.*

One of the fundamental necessary conditions for ergodicity are different variants of Doeblin's condition [18]. Let us quote two versions.

Theorem 2.2 (Doeblin)**.** *Assume we have a Markov chain, a probability measure ν and some $\varepsilon > 0, k \geq 1$ such that*

$$\nu(i) < \varepsilon \quad \Longrightarrow \quad \forall j \quad \tilde{h}^k(j,i) < 1 - \varepsilon.$$

Then the structure of stationary distributions can be described precisely. For irreducible Markov chains, the stationary distribution is unique and the condition ensures ergodicity.

Theorem 2.3 (Doeblin)**.** *Assume we have a Markov chain, a probability measure ν and some $\delta > 0, k \geq 1$ such that*

$$\forall i,j \quad \tilde{h}^k(j,i) > \delta\nu(i).$$

Then the Markov chain is ergodic.

Detailed discussion of the above theorems can be found in the book of Doob [19].

Only a very small part of this general theory carries through for Bi-Markov processes. However, the existence of a stationary distribution is true in a much more general setting. We may apply Brouwer's fixed point theorem [8] as a time step is a continuous function from the convex body of probability distributions to itself.

Theorem 2.4 (Brouwer)**.** *Every finite Bi-Markov process has a stationary distribution.*

While this more general theorem provides existence, it does not tell anything about uniqueness or ergodicity. The necessary condition of Doeblin remains true only in a significantly restricted version found by László Gerencsér:

Theorem 2.5. *Assume we have a Bi-Markov process, a probability measure ν and some $\delta \geq 1/2$ such that*

$$\forall i, j, k \quad h(j, k, i) > \delta\nu(i).$$

Then the Bi-Markov process is ergodic.

Proof. We show the claim of the theorem using a coupling argument. The condition of the theorem tells us that with probability δ the process forgets the past and draws an independent new state from the distribution ν.

Let us start with two initially uncoupled copies of the process. We couple the resetting of the processes, that is, with probability δ both copies jump to the same state according to ν.

Let us denote by α_n the probability of being coupled after n steps. The state in the next step is coupled either if both of its predecessors are already coupled or if they become coupled in the current transition. This leads us to the equation

$$\alpha_{n+1} = \alpha_n^2 + \delta(1 - \alpha_n^2).$$

Starting from $\alpha_0 = 0$ and using $0 < \delta < 1$ this is a monotone increasing bounded sequence consequently it tends to a limit. The possible limit points can be obtained by solving the quadratic equation obtained by replacing α_n, α_{n+1} with α_∞ in Equation 2. We get

$$\alpha_\infty^1 = 1, \qquad \alpha_\infty^2 = \frac{\delta}{1 - \delta}.$$

Asymptotically almost sure coupling occurs when there is no limit point less than 1 which corresponds to $\delta \geq 1/2$. Any initialization of the process must lead to the same stationary distribution because we may initialize one copy of the process to be already stationary. □

What can we say when the condition holds only for some $\delta < 1/2$? Without the restriction on δ the Doeblin condition for a finite state space Bi-Markov process simply means that all $h(j, k, i)$ are positive. We do not know the answer for Bi-Markov processes, but it turns out that this condition does not imply ergodicity for closely related dynamics. We get a more general class if we omit Condition 2.1 on the sum of transition probabilities. In this setting we need to normalize after each step to get a valid probability distribution.

Thus the new dynamics of *general Bi-Markov processes* is

$$\tilde{p}_i(t+1) = \sum_{j,k} h(j,k,i) p_j(t) p_k(t),$$
$$p_i(t+1) = \frac{\tilde{p}_i(t+1)}{\sum_j \tilde{p}_j(t+1)}.$$

Tusnády [52] showed that this related model provides counterexamples when there is no restriction on δ.

Theorem 2.6 (Tusnády)**.** *There exists a general Bi-Markov process on 3 states such that every transition probability is positive and the distribution of the states changes in a periodic way.*

Conjecture 2.7 (Tusnády)**.** *There exists a general Bi-Markov process on 4 states such that every transition probability is positive and the distribution of the states shows a chaotic behavior. This is verified by computer simulations.*

Partial results on this conjecture have been presented by Dénes [14].

These results weaken the hope for a generic necessary condition providing ergodicity. At this point it is more likely that we need a unique approach for every process we work with.

To demonstrate some positive results we might expect, let us show an example of Bi-Markov processes which is reasonably understood. We sketch the problem investigated by Komlós et al. see [36] for details. At time 0, imagine an infinite series of customer service agents with $X_0(i)$ customers to serve, $i = 0, 1, \ldots$. These $X_0(i)$ are independent, identically distributed, non-negative integer variables. Each agent deals with a single customer (if any) in its queue and sends the remaining ones to the next level agents, each of which accepts customers from two of the lower level agents. In the end transitions happen according to the following rule:

$$X_j(i) = (X_{j-1}(2i) - 1)^+ + (X_{j-1}(2i+1) - 1)^+, \qquad i = 0, 1, \ldots$$

We want to decide whether the agency will be able to deal with all customers or will it collapse under the load. Note that this can be regarded as a proper Bi-Markov process, however, the only randomness is coming from the starting variables as the transitions are completely deterministic.

For this problem it is possible to find out the stationary distributions, it turns out

there are three of them:

$$X_\infty \equiv 0,$$
$$X'_\infty \equiv 2,$$
$$X''_\infty \equiv \infty.$$

On the other hand, it is unclear whether the system converges to any of these for a certain starting distribution. Partial results are available:

Theorem 2.8 (Komlós et al.)**.** *Suppose we start the process with* $X_0(i) \sim Poisson(\lambda)$, $i = 0, 1, \ldots$ *for some* $\lambda > 0$.

- *If* $\lambda < 0.999$*, the process converges to* $X_\infty \equiv 0$.
- *If* $\lambda > 1.001$*, the process blows up to* $X''_\infty \equiv \infty$.

It is natural to conjecture the critical value to be $\lambda = 1$, but this is not yet proven. Despite the neat results a natural question remains unanswered, namely whether some exotic starting distribution can lead to a periodic or chaotic process or not.

In the next section we give an overview of the biological background providing the general Bi-Markov processes we work with, followed by detailed discussion and statistical investigation of the so called *Poisson model*. This is a joint work with my advisor Gábor Tusnády and also with Balázs Ráth.

2.1 Basics of multifactorial inheritance

The concept of multifactorial inheritance goes back to Francis Galton, a contemporary of Gregor Johann Mendel (see in Karlin [33]). Instead of the case investigated by Mendel, where the appearance of a congenital malformation is controlled by a single gene, in multifactorial inheritance the number of genes involved is large or infinite. As a result their effect is concentrated in a virtual quantity, the liability having standard normal distribution. See Curnow, Smith [11] for an overview of multifactorial models. The joint distribution of the liabilities of members of a family is also normal with covariances determined by the remove degrees of relationship.

$$cov(X, Y) = \frac{h^2}{2^d},$$

where h is the heritability of the malformation and d is the degree of relationship. In the simplest case of $h = d = 1$ the conditional probability that a first order relative of a malformed person has the malformation is roughly $\sqrt{p}$, where p is the population

incidence of the malformation. This approximation is due to A. W. F. Edwards [21]. The multifactorial model was tested on Hungarian data by Czeizel and Tusnády [12] which work was criticized by Kari Sankaranarayanan because the effect of selection was neglected. He organized a group to solve the problem and some preliminary results were published by members of the group [49] while Tusnády tested the new model on original data [52]. Unfortunately a question remained unsettled: the stability of the proposed model. Here we offer a partial solution of the problem.

Let X and Y be the liabilities of the parents, then the liability of their child is

$$Z = \frac{X+Y}{2} + U,$$

where U is a normal variable with expectation zero and variance $\frac{1}{2}$. The main observation of Sankaranarayanan was that in the case of selection the bad genes causing the malformation simply flow out from the population like the water from a bathtub. It is the mutation which can supplant the bad genes. The effect in the model may be represented by changing the expectation of U to some positive number to balance the effect of selection.

Besides genetic factors there are environmental effects modifying the liability so instead of Z let us use $L = Z + V$ where V is the environmental effect with appropriate variance. Let us postulate that the appearance of the malformation is equivalent with the event $L > T$, where T is the threshold. (The random variables X, Y, U, V are independent.)

The effect of selection may be represented in the model by a second threshold $S > T$ such that if $L > S$ then there will be no descendant for the person having liability L. The stability of the model means that starting with an arbitrary distribution on parents in course of generations the distribution of the liability goes to a limit which is independent of the original distribution. This is observed for computer simulations but we have no theoretical proof. Instead we turn to the case of finitely many bad genes.

In this setting we assume that there is an infinite number of genes, but the mutations are rare, thus the number of bad genes can be treated as a non-negative integer. This concept was introduced by Kimura and Maruyama [35] and Kondrashov [37]. To avoid halving genes we let any bad gene of the parents to be given to the child independently with probability $\frac{1}{2}$. Both the mutation and the environmental effects can be represented by a Poisson variable an appropriate parameter. L may be identified in this case with a natural number coming partly from bad genes and partly from quantized environmental effects with the same habit as bad genes. In the general case let $p(L)$ be the probability that a person with liability L has the malformation. If $p(L) = 1$ iff $L \geq T$, 0 otherwise the situation is the same as in the continuous case. We call this the *discrete threshold model*. Here we also need to have mutation to avoid the bad genes from disappearing from

the population. Stability is also evident by computational results. Still, we don't have a theoretical proof to verify it.

Let us change the way the malformation appears as suggested by Balázs Ráth by choosing $p(L) = 1 - \rho^L$ with some $0 < \rho < 1$. We call this the *Poisson model*. In this setting the question of stability turns to be solvable. It is also shown by Dawson [13]. It is justifiable to consider other $p(L)$ functions, for example see Crow, Kimura [10] comparing the threshold function with linearly decreasing alternatives.

Let us give a hint on the complexity of the situation when investigating two relatives at once. We say we are thinning a Poisson variable if we represent it with balls and kill independently the balls with a certain probability. It is a well known fact that the thinning of a Poisson variable results in a Poisson variable again. Let Z be a Poisson variable with parameter λ and let it be thinned independently into random variables X_1 and Y_1 with probabilities p and q accordingly. Let the random variables X_2 and Y_2 be Poisson with parameters $(1-p)\lambda$ and $(1-q)\lambda$ and independent of the earlier random variables. The variables

$$X = X_1 + X_2, \quad Y = Y_1 + Y_2$$

may correspond to the liabilities of two relatives. Their joint distribution is somewhat cumbersome:

$$P(X = x, Y = y) = \sum_{z=0}^{\infty} \mathrm{Pois}(z, \lambda) \left[\sum_{i=0}^{z} \mathrm{Bin}(z, p, i) \mathrm{Pois}(x - i, (1-p)\lambda) \right] \times \left[\sum_{j=0}^{z} \mathrm{Bin}(z, q, j) \mathrm{Pois}(y - j, (1-q)\lambda) \right].$$

but its generating function is easily found. This observation is the driving force in our calculations on the conditional probabilities for pairs of relatives.

In Section 2.2 we present the Poisson model in detail, in Section 2.3 we prove the stability theorem, in Section 2.4 we develop the conditional probabilities for the malformation in the relatives of an affected person. In Section 2.5 the theory is applied on the Hungarian data, and in Section 2.6 the conclusions are drawn.

2.2 The working model

We consider a population with sexual reproduction, selection, synchronous generations on a short time frame in the evolutionary sense. We assume all relevant loci have the same effect in view of the birth defect, so the only thing we keep track of is the number of mutant genes one has. To get the genetic information of offsprings, we need recombination, mutation, and selection.

During recombination we assume crossovers may happen, and there is a low number of mutant genes, that is, each of them is inherited independently with probability 1/2. If the two parents have x and y mutant genes, the child will receive a random number of mutant genes from the $Binom(x+y, 1/2)$ distribution.

The child is affected by additional mutation, this is represented by adding an independent $Poisson(\mu)$ random variable to the inherited mutant gene count.

Given the number of mutant genes the child has, we have to find out two things: whether he/she is affected by the disorder and whether he/she is fertile (and viable). We assume each mutant gene may cause the disorder to appear or the loss of fertility. There is an ordering of the two symptoms, a gene causing the loss of fertility also causes the disorder to appear. The probability of a single gene *not* causing the disorder is denoted by Δ, and the probability of *not* inhibiting fertility is ρ. Clearly $\rho > \Delta$. Once again, each gene has a random effect on the individual in the following way:

- with probability Δ it has no effect,
- with probability $\rho - \Delta$ it causes the individual to be affected by the disorder, but has no effect on fertility,
- with probability $1 - \rho$ it causes the individual to be affected by the disorder and lose fertility.

We need to easily refer to the combination of these operations. For a pair of distribution of mutant genes (P_f, P_m) let us denote the female distribution of the next generation by $T_f(P_f, P_m)$. We use the analogous notation for the male counterpart. We vaguely use $T_f^k(P_f, P_m)$ for the female distribution after k generations (although we should use $T_f(T_f(P_f, P_m), T_m(P_f, P_m))$ instead of $T_f^2(P_f, P_m)$).

2.3 Stationary genotype distribution

This section deals with the long-term behavior of the genotype distribution. It is rather clear that if there is no selection, which has the role of filtering out the mutant genes, then their number will grow unboundedly. Consequently, to have a chance of stationarity, we need $\rho < 1$. We claim that in this case the distribution of mutant genes in the population stabilizes over time. We assume there is a separate set of parameters for females $(\mu_f, \rho_f, \Delta_f)$ and males $(\mu_m, \rho_m, \Delta_m)$. The model used by Dawson [13] does not use sex dependent parameters, but includes a separate modifier gene that can alter the parameters. We do not see biological evidence for μ_f and μ_m to differ but it does no harm

to include it in our study, and we get a more general result. The credit for the main ideas of this proof goes to Balázs Ráth.

Theorem 2.9. *If $\rho_f, \rho_m < 1$ then for any pair P_f, P_m of initial distributions of mutant genes, the distribution of $T_f^k(P_f, P_m), T_m^k(P_f, P_m)$ will converge in distribution to a pair of limiting Poisson distributions with parameters*

$$\lambda_f = \frac{\rho_f\rho_m(\mu_m - \mu_f) + 2\rho_f\mu_f}{2 - \rho_f - \rho_m}, \qquad \lambda_m = \frac{\rho_f\rho_m(\mu_f - \mu_m) + 2\rho_m\mu_m}{2 - \rho_f - \rho_m},$$

for females and males, respectively, when $k \to \infty$.

Proof. We work with generating functions. We say that $P = (p_i)_{i=0}^{\infty}$ is a probability distribution on $\mathbb{N}$ if $p_i \geq 0$ and $\sum_{i=0}^{\infty} p_i = 1$. Denote by $\mathcal{P}$ the set of probability distributions on $\mathbb{N}$. For $P \in \mathcal{P}$ and $x \in [0, 1]$ let us define

$$G_P(x) = \sum_{i=0}^{\infty} p_i x^i.$$

The coefficients of the power series form a probability distribution, consequently $G_P(x)$ is analytic on $[0, 1]$. The operations used in our model are easy to handle with generating functions. We write out the equations for a daughter, we get the analogous equations for a son by exchanging f and m in the indices.

Convolution of distributions are reflected as multiplication of the generating functions, so adding up parental mutant genes translates to

$$G_{P'}(x) = G_{P_f}(x)G_{P_m}(x).$$

Plugging the value of the variable into a Binomial distribution with parameter $1/2$ (also known as "thinning") translates to changing the argument from x to $(1 + x)/2$. We get

$$G_{P''}(x) = G_{P'}\left(\frac{1+x}{2}\right).$$

Adding external mutation is another multiplication with the generating function of a Poisson variable with parameter μ_f:

$$G_{P'''}(x) = G_{P''}(x)e^{\mu_f(x-1)}.$$

During selection, we put weights on each p_i''', then normalize to obtain a probability distribution in the following fashion: the probability of having i mutant genes is p_i''', and the probability that a female with i mutant genes remains fertile is ρ_f^i, thus a female in the community of fertile females will have i mutant genes with probability $p_i'''\rho_f^i / \sum_{j=0}^{\infty} p_j'''\rho_f^j$.

This operation is known as the "exponential tilting" of the distribution P'''. For generating functions, the effect of selection can be computed the following way:

$$G_{P''''}(x) = \sum_{i=0}^{\infty} \frac{p_i''' \rho_f^i}{\sum_{j=0}^{\infty} p_j''' \rho_f^j} x^i = \frac{G_{P'''}(\rho_f x)}{G_{P'''}(\rho_f)}.$$

Composing the three transformations we get

$$G_{T_f(P_f,P_m)}(x) = \frac{G_{P_f}\left((1+\rho_f x)/2\right) G_{P_m}\left((1+\rho_f x)/2\right)}{G_{P_f}\left((1+\rho_f)/2\right) G_{P_m}\left((1+\rho_f)/2\right)} e^{\mu_f \rho_f (x-1)}. \tag{2.2}$$

We want to iterate T n times. Naturally we want to avoid writing down all these complicated formulas. In order to see the structure of what we get, let us write down the formula for T^2, but without arguments.

$$G_{T_f^2}(x) = \frac{G_{T_f()}()G_{T_m()}()}{G_{T_f()}()G_{T_m()}()} e^{\cdots} = \frac{\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()} e^{\cdots} \frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()} e^{\cdots}}{\frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()} e^{\cdots} \frac{G_{P_f}()G_{P_m}()}{G_{P_f}()G_{P_m}()} e^{\cdots}} e^{\cdots}. \tag{2.3}$$

From (2.2) we see that the denominator of $G_{T_f(P_f,P_m)}$ is constant in x and the constant is the normalizing factor which guarantees that $G_{P''''}(1) = 1$. Rearranging (2.3) we end up with a formula that is the product of four $G()/G()$ terms (where the denominator normalizes the numerator and the ratio takes value 1 for $x = 1$) and an exponential term. After n iterations we get that $G_{T_f^n}(x)$ is a product of the functions $\hat{G}_{T_f^n}(x)$ and $E_f^n(x)$, where $\hat{G}_{T_f^n}(x)$ is a product of 2^n terms of form $G()/G()$ and $E_f^n(x)$ is an exponential term (the generating function of some Poisson random variable).

Let us treat $\hat{G}_{T_f^n}(x)$ and $E_f^n(x)$ separately.

We first show that $\hat{G}_{T_f^n}(x) \to 1$ for all $x \in [0,1]$ as $k \to \infty$. If we put back the arguments in one of the 2^n terms of $\hat{G}_{T_f^n}(x)$, we see that it is of form

$$\frac{G(B(x))}{G(B(1))},$$

where B is an affine function, an n-fold composition of either $x \mapsto (1+\rho_f x)/2$ or $x \mapsto (1+\rho_m x)/2$, and the generating function G is either G_{P_f} or G_{P_m}. The product of all these terms look like

$$\frac{G(B(x))}{G(B(1))} \cdot \ldots \cdot \frac{G(B(x))}{G(B(1))} = \exp\left(\log \frac{G(B(x))}{G(B(1))} + \ldots + \log \frac{G(B(x))}{G(B(1))}\right), \tag{2.4}$$

with G and B changing throughout the formula. Let us make sure the use of logarithms is feasible. It is easy to see that $B(x) > 0$ for $x \geq 0$. The generating function G is a power series with non-negative (and at least one positive) coefficients, so $G(B(x)) > 0$ for $x \in [0,1]$. Now we have to estimate the terms of the form $\log(G(B(x))/G(B(1)))$. By the

mean value theorem for every $x \in [0,1]$ there is a $\xi \in [B(x), B(1)] \subseteq [1/2, (1+\rho_*)/2]$ such that

$$\log \frac{G(B(x))}{G(B(1))} = \log G(B(x)) - \log G(B(1)) = (B(x) - B(1))(\log G)'(\xi)$$

We denote $\rho_* = \max(\rho_f, \rho_m) < 1$. The coefficient of x in $B(x)$ will be at most $(\rho_*/2)^n$. Thus for any $x \in [0,1]$ we get

$$|B(x) - B(1)| \leq \left(\frac{\rho_*}{2}\right)^n.$$

The function G is continuously differentiable and bounded away from 0 on the interval $\xi \in [B(x), B(1)] \subseteq [1/2, (1+\rho_*)/2]$, consequently the derivative of the logarithm can be bounded in absolute value by some C. In the end we get

$$\left|\log \frac{G(B(x))}{G(B(1))}\right| < C\left(\frac{\rho_*}{2}\right)^n.$$

Adding up 2^n of such terms gives the bound

$$\left|\log \frac{G(B(x))}{G(B(1))} + \ldots + \log \frac{G(B(x))}{G(B(1))}\right| \leq C\rho_*^n.$$

This tends to 0 for all $x \in [0,1]$, thus the product on the left-hand side of (2.4) converges to 1 as $n \to \infty$. Observe that the exponential term in (2.2) does not depend on the initial distributions P_f, P_m. Thus we have just shown that the only part depending on the initial distributions vanishes. Consequently the convergence and the potential limit does not depend on the initial distributions.

It is now enough to show a pair of distributions satisfying

$$(P_f, P_m) = (T_f(P_f, P_m), T_f(P_f, P_m)),$$

as the previous reasoning ensures that the trivial convergence of this case implies convergence for any initial generating functions to this fixed point. We search among Poisson distributions because this family is closed for all the transformations we use. The pair (λ_f, λ_m) is invariant exactly when

$$\lambda_f = \left(\frac{\lambda_f + \lambda_m}{2} + \mu_f\right)\rho_f, \qquad \lambda_m = \left(\frac{\lambda_f + \lambda_m}{2} + \mu_m\right)\rho_m.$$

Taking the average of the two equations results in a simple expression for $(\lambda_f + \lambda_m)/2$, plugging it back gives us the parameters stated in the theorem.

To conclude we use the fact that the convergence of a sequence of generating functions to a generating function on $[0,1]$ implies the convergence of the corresponding probability distributions (see e.g. Mukherjea, Rao and Suen [45]).

□

We should note that the proof strongly relies on the specific choice of selection which we can conveniently handle using generating functions. As we mentioned in the introduction, it makes sense to consider different functions determining the risk based on the mutant gene count. However, it is unclear how one should modify the proof to resolve the alternative cases.

2.4 Theoretical disorder probabilities

From the previous section we learn that it makes sense to assume the population to be in the stationary state. It is easy to check that the number of mutant genes a newborn has follows a Poisson distribution with the following parameters depending on the gender:

$$\frac{\lambda_f + \lambda_m}{2} + \mu_f = \frac{\lambda_f}{\rho_f}, \qquad \frac{\lambda_f + \lambda_m}{2} + \mu_m = \frac{\lambda_m}{\rho_m}.$$

Consequently his/her probability for being healthy is

$$p_f = \exp\left(\lambda_f \frac{\Delta_f - 1}{\rho_f}\right), \qquad p_m = \exp\left(\lambda_m \frac{\Delta_m - 1}{\rho_m}\right).$$

Similarly, the probability of being fertile is

$$\tilde{p}_f = \exp\left(\lambda_f \frac{\rho_f - 1}{\rho_f}\right), \qquad \tilde{p}_m = \exp\left(\lambda_m \frac{\rho_m - 1}{\rho_m}\right).$$

However, if we look at a family tree at once, we see a complex multidimensional joint distribution. We want to answer simple questions like "What is the (conditional) probability of an aunt of a malformed child being affected".

We claim that we can get a closed form expression on any reasonable conditional probabilities like the one above. The resulting formulas often become enormous, but there is a way to derive them with reasonable effort.

We would like a general iterative computational scheme that can be used for most cases. The idea is to draw a graph of the family tree, transform it to simpler graphs while building the formula for the probability.

We include the possible dependence on the gender of the patient. Therefore the parameters we have are

$$\mu_f, \mu_m, \rho_f, \rho_m, \Delta_f, \Delta_m.$$

The parameters of the stationary distributions are

$$\lambda_f = \frac{\rho_f\rho_m(\mu_m - \mu_f) + 2\rho_f\mu_f}{2 - \rho_f - \rho_m}, \qquad \lambda_m = \frac{\rho_f\rho_m(\mu_f - \mu_m) + 2\rho_m\mu_m}{2 - \rho_f - \rho_m}.$$

To reduce the number of formulas, from now on we use $x, y, \ldots$ for one gender or another, thus μ_x or λ_y is the parameter corresponding to the appropriate gender. In addition we use x' for the gender opposite to x.

2.4.1 Representing graphs

First, let us visualize the situation. We may draw a family tree with some additional information.

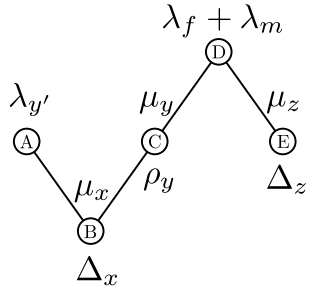


Figure 2.1: Healthy boy and aunt (or similar)

We use Figure 2.1 as an example. Suppose $x = m, y = m, z = f$ for a moment. The circles in the graph represent members or couples of the family. In this case B is the male patient we start with, A is the mother, C is the father. D represents the paternal grandparents together. We do not separate them as we use only the joint genetic information of them. The last member E is an aunt.

The genetic information moves in the following way. Each line represents a parental relation, so each gene is inherited downwards independently with probability 1/2. The values above the circles show where additional mutant genes enter the system. We always mean a Poisson random variable with the parameter being the value indicated. These are obviously μ_u for most people, and λ_u or $\lambda_f + \lambda_m$ for the people or couples we start with.

The event we want to investigate is coded in the values below the circles. They show a per-gene probability for mutant genes that the actual person complies with the event. In the figure above we have Δ_u in two positions which means we want the patient and the aunt (or uncle) to be healthy. The ρ_y under C is an implied restriction, as we need the father (or mother) to be fertile for the graph to be valid. Some places have no value indicated, we have no restriction there, we may also write 1 to these places.

This way we can only express events requiring some to be healthy, some to be fertile, but these are the one that are easy to directly compute. By basic inclusion-exclusion formulas we can also handle events about some being affected or infertile. To compute conditional probabilities we simply need to divide two of such probabilities.

Now let us get into computational details to work through our plan.

2.4.2 Processing graphs

We can handle the simplest graph possible:

Figure 2.2: Basic graph

The probability of the event described by this basic graph is

$$\sum_{i=0}^{\infty} \frac{\eta^i}{i!} e^{-\eta} \alpha^i = \exp(\eta(\alpha - 1)).$$

We introduce a few graph operations so we can transform complex graphs into simpler ones. Observe that if a final descendant receives mutant genes from multiple sources, they pose independent threats, so we can split the graph as pictured below.

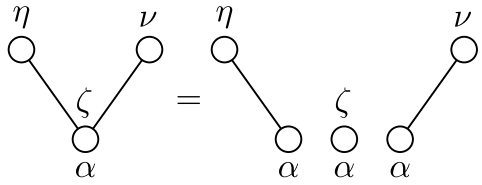


Figure 2.3: Splitting a graph

The other operation we use is to merge a child to the parent. Consider the following setting:

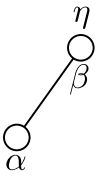


Figure 2.4: Parent and child

We condition on the number of mutant genes the parent has, suppose it is c. Then the distribution of mutant genes the child inherits follows a $Binom(c, 1/2)$ distribution. So the probability that the child behaves according to the event is

$$\sum_{i=0}^{c} \binom{c}{i} \left(\frac{1}{2}\right)^c \alpha^i = \left(\frac{1+\alpha}{2}\right)^c.$$

This is an exponential term in c, so we do not change the overall probability of the event if we omit the child but multiply the risk factor of the parent by $(\alpha + 1)/2$.

It is easy to see that any acyclic family tree can be reduced to contain only a few copies of the simplest one-node graph.

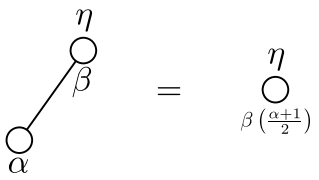


Figure 2.5: Merging a child

2.4.3 Siblings

Let us start with the simplest case, computing conditional probabilities for first order relatives. We want to find out the conditional probability of a sibling of a malformed child being affected. Figure 2.6 shows the graph for the sibling.

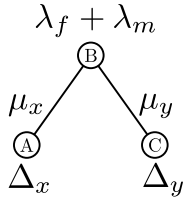


Figure 2.6: Healthy patient and sibling

Let us use the notation scheme $p_{\bar{A}C}$, this stands for the probability of A being affected by the risk and C not (and we don't count on others). This means the conditional probability q_S we need is

$$q_S = \frac{p_{\bar{A}\bar{C}}}{p_{\bar{A}}}.$$

Using inclusion-exclusion formulas we have

$$p_{\bar{A}\bar{C}} = 1 - p_A - p_C + p_{AC},$$
$$p_{\bar{A}} = 1 - p_A.$$

The method in the previous section allows us to compute these probabilities. When computing p_A, we replace the risk of C by 1. The graph decomposition is shown in Figure 2.7. By symmetry we have $p_C = p_A$. We show the graph decomposition for computing p_{AC} in Figure 2.8.

We do not aim for the simplest expressions, we rather leave it in a form that is easier to check.

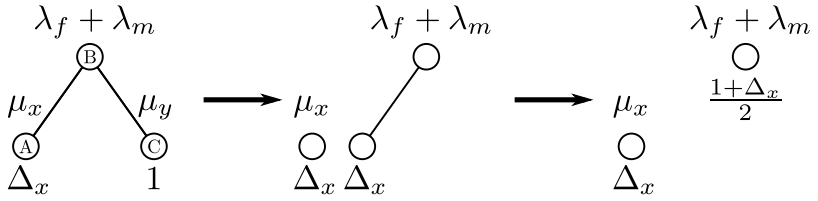


Figure 2.7: Graph decomposition to compute p_A

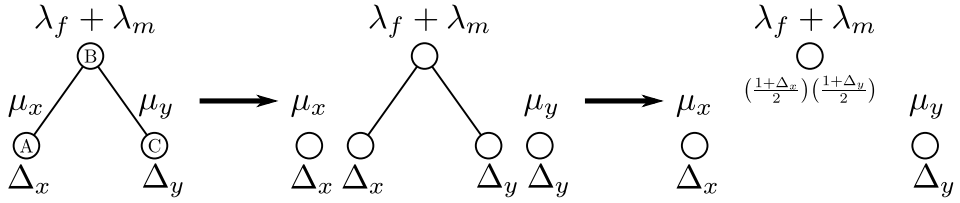


Figure 2.8: Graph decomposition to compute p_{AC}

$$p_A = \exp\left(\left(\mu_x + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_x - 1)\right),$$
$$p_C = \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_y - 1)\right),$$
$$p_{AC} = \exp\left(\mu_x(\Delta_x - 1) + \mu_y(\Delta_y - 1) + \frac{\lambda_m + \lambda_f}{4}((\Delta_x + 1)(\Delta_y + 1) - 4)\right).$$

In case of complete selection and symmetric gender roles, i.e.

$$\Delta_m = \Delta_f = \rho_m = \rho_f = \rho, \quad \lambda_f = \lambda_m = \lambda, \quad \text{and} \quad \mu_m = \mu_f = \mu,$$

the conditional probability q_S is

$$q_S = 2 - \frac{1 - e^{-t}}{1 - e^{-\mu}},$$

where

$$t = 2\mu(1 - \frac{1}{4}\rho(1 - \rho)),$$

and $\rho = \frac{\lambda}{\lambda+\mu}$. Surprisingly q_s depends on ρ through the term $\rho(1-\rho)$. In this case the population prevalence simplifies to

$$p_{\bar{A}} = 1 - \exp((\lambda + \mu)(\rho - 1)) = 1 - \exp(\lambda - (\lambda + \mu)) = 1 - \exp(-\mu).$$

thus ρ is a free parameter and q_S is a symmetric function of ρ regarding the swap $\tilde{\rho} = 1 - \rho$. We are curious whether there is a direct explanation for this symmetry. When μ is small and $\rho = \frac{1}{2}$, then $\lambda = \mu$ and a bad gene is rare. An affected child gets a bad gene fifty-fifty either from mutation or from one of his/her parents. In the second case the sibling gets the bad gene from the affected parent with half probability and the bad gene is expressed again with probability half. Accordingly q_S is close to $\frac{1}{8}$. We shall refer to this parametrization as the *standard model*.

2.4.4 Parent

Next we calculate the conditional probability for a parent being affected, which is also fairly simple. See Figure 2.9 for the describing graph. The only novelty is the Δ_y/ρ_y risk of the parent. It is easy to see that this is the risk of not being affected by the disorder conditioned on being fertile.

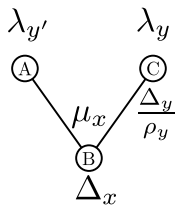


Figure 2.9: Healthy patient and parent

$$q_P = \frac{p_{\bar{B}\bar{C}}}{p_{\bar{B}}} = \frac{1 - p_B - p_C + p_{BC}}{1 - p_B}$$

$$\begin{aligned}
p_B &= \exp\left(\left(\mu_x + \frac{\lambda_f + \lambda_m}{2}\right)(\Delta_x - 1)\right), \\
p_C &= \exp\left(\lambda_y\left(\frac{\Delta_y}{\rho_y} - 1\right)\right), \\
p_{BC} &= \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \lambda_y\left(\frac{\Delta_y}{\rho_y}\left(\frac{\Delta_x + 1}{2}\right) - 1\right)\right).
\end{aligned}$$

In the standard model, when $\rho_f = \rho_m = \Delta_f = \Delta_m = 1/2$ and $\mu_f = \mu_m$ is small, we get $q_P = 0$. This is rather clear because this special case implies complete selection.

2.4.5 Grandparent

Let us move on to higher order relatives, starting with grandparents. Figure 2.10 shows the actual graph to be processed. The conditional probability can be expressed as

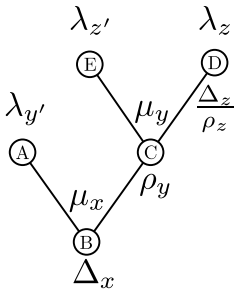


Figure 2.10: Healthy patient and grandparent

$$q_G = \frac{p_{\bar{B}C\bar{D}}}{p_{\bar{B}C}} = \frac{p_C - p_{BC} - p_{CD} + p_{BCD}}{p_C - p_{BC}}$$

$$\begin{aligned}
p_C &= \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\rho_y - 1)\right), \\
p_{BC} &= \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)\left(\rho_y \frac{\Delta_x + 1}{2} - 1\right)\right), \\
p_{CD} &= \exp\left(\left(\mu_y + \frac{\lambda_{z'}}{2}\right)(\rho_y - 1) + \lambda_z\left(\frac{\Delta_z}{\rho_z}\left(\frac{\rho_y + 1}{2}\right) - 1\right)\right), \\
p_{BCD} &= \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_{z'}}{2}\right)\left(\rho_y \frac{\Delta_x + 1}{2} - 1\right)\right. \\
&\quad \left. + \lambda_z\left(\frac{\Delta_z}{\rho_z}\left(\frac{\rho_y \frac{\Delta_x + 1}{2} + 1}{2}\right) - 1\right)\right).
\end{aligned}$$

In the standard model we get $q_G = 0$ as we expect because of the complete selection.

2.4.6 Aunt and uncle

Let us turn to investigating aunts and uncles. We use Figure 2.1 for the calculation. The conditional probability can be expressed as

$$q_A = \frac{p_{\bar{B}C\bar{E}}}{p_{\bar{B}C}} = \frac{p_C - p_{BC} - p_{CE} + p_{BCE}}{p_C - p_{BC}}.$$

We can compute the occurring probabilities as before. Without going into details, we get

$$\begin{aligned}
p_C &= \exp\left(\left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)(\rho_y - 1)\right), \\
p_{BC} &= \exp\left(\left(\frac{\lambda_{y'}}{2} + \mu_x\right)(\Delta_x - 1) + \left(\mu_y + \frac{\lambda_f + \lambda_m}{2}\right)\left(\rho_y \frac{\Delta_x + 1}{2} - 1\right)\right), \\
p_{CE} &= \exp\left(\mu_y(\rho_y - 1) + \mu_z(\Delta_z - 1) + (\lambda_f + \lambda_m)\left(\frac{(\rho_y + 1)(\Delta_z + 1)}{4} - 1\right)\right), \\
p_{BCE} &= \exp\left(\mu_z(\Delta_z - 1) + \left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \mu_y\left(\rho_y \frac{\Delta_x + 1}{2} - 1\right) + \right. \\
&\quad \left. + \frac{\lambda_f + \lambda_m}{4}\left(\left(\rho_y \frac{\Delta_x + 1}{2} + 1\right)(\Delta_z + 1) - 4\right)\right).
\end{aligned}$$

Plugging these back gives us the conditional probability we were looking for.

In the standard model the number of halving factors is 5:

– the affected child might get the bad gene by mutation

– or by the parent out of link to aunt-uncle

– the parent in the link to aunt-uncle might get the bad gene by mutation

– the grandparents need not to pass it to another child

– who needs not to express the malformation.

We get $q_A = 1/32$ as well by using the expressions above for the standard model.

2.4.7 Cousin

To compute the analogous conditional probability for cousins, we will use Figure 2.11 below.

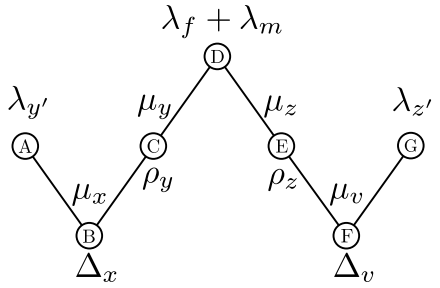


Figure 2.11: Healthy patient and cousin

Using the same method, we want to compute

$$q_C = \frac{p_{\bar{B}CE\bar{F}}}{p_{\bar{B}CE}} = \frac{p_{CE} - p_{BCE} - p_{CEF} + p_{BCEF}}{p_{CE} - p_{BCE}} = 1 - \frac{p_{CEF} - p_{BCEF}}{p_{CE} - p_{BCE}}.$$

For the individual probabilities in this setting we get

$$\begin{aligned}
p_{CE} &= \exp\left(\mu_y(\rho_y - 1) + \mu_z(\rho_z - 1) + \frac{\lambda_f + \lambda_m}{4}\left((\rho_y + 1)(\rho_z + 1) - 4\right)\right), \\
p_{BCE} &= \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \mu_z(\rho_z - 1) + \mu_y\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right) + \right. \\
&\quad \left. + \frac{\lambda_m + \lambda_f}{4}\left(\left(\rho_y\frac{\Delta_x + 1}{2} + 1\right)(\rho_z + 1) - 4\right)\right), \\
p_{CEF} &= \exp\left(\left(\mu_v + \frac{\lambda_{z'}}{2}\right)(\Delta_v - 1) + \mu_y(\rho_y - 1) + \mu_z\left(\rho_z\frac{\Delta_v + 1}{2} - 1\right) + \right. \\
&\quad \left. + \frac{\lambda_m + \lambda_f}{4}\left(\left(\rho_z\frac{\Delta_v + 1}{2} + 1\right)(\rho_y + 1) - 4\right)\right), \\
p_{BCEF} &= \exp\left(\left(\mu_x + \frac{\lambda_{y'}}{2}\right)(\Delta_x - 1) + \left(\mu_v + \frac{\lambda_{z'}}{2}\right)(\Delta_v - 1) + \right. \\
&\quad + \mu_y\left(\rho_y\frac{\Delta_x + 1}{2} - 1\right) + \mu_z\left(\rho_z\frac{\Delta_v + 1}{2} - 1\right) + \\
&\quad \left. + \frac{\lambda_f + \lambda_m}{4}\left(\left(\rho_y\frac{\Delta_x + 1}{2} + 1\right)\left(\rho_z\frac{\Delta_v + 1}{2} + 1\right) - 4\right)\right).
\end{aligned}$$

These are rather cumbersome formulas, but in the standard model, we get $q_C = 1/124$. At first this is a bit surprising, because by counting the number of halving factors as before, we get $1/2^7 = 1/128$. We should note that checking a cousin for the disorder implies he is already born, that is, his parents are fertile. Conditioning on this accounts for a division by 31/32 which brings us to the correct value.

2.5 Validation of the model

It is an important milestone to have a model which we can handle, we still have to check how well does it follow biological principles and how does it fit the population. Let us recall the notations introduced in Section 2.4:

$$p = P(\text{subject is affected}),$$
$$q_S = P(\text{sibling is affected}|\text{subject is affected}).$$

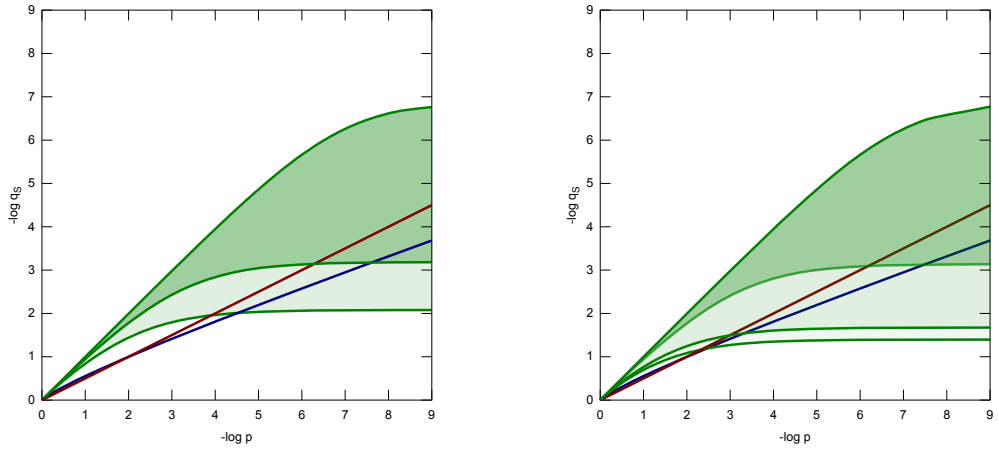
The initial requirement for a model of inheritance is to have high conditional probabilities for first order relatives, in other words $q_S \gg p$. To test this, we will try to choose the parameters to increase q_S as much as possible within the given constraints.

Another guideline we use is a fundamental approximation on multifactorial disorders given by the Edwards formula [21] which states that $q_S \approx \sqrt{p}$.

We don't want to go into theoretical details, let us just present Figure 2.12 showing the relation between $\log p$ and $\log q_S$ for $\mu \in [5 \cdot 10^{-5}, 3]$ and $\Delta \in [0.1, 1)$. On the left side, we assume complete selection, that is, $\rho = \Delta$, on the right side we consider a partial selection with $\rho = (1 + \Delta)/2$.

The upper diagonal line shows where the Edwards formula is precisely satisfied, the lower one corresponds to probabilities of the Gaussian model used by Czeizel and Tusnády in [12]. We prefer parameters where the disorder is mainly inherited, that is, $\lambda \gg \mu$. Thus we split the domain the model sweeps through into three regions, the values we can reach while $\lambda \geq 10\mu$, or just $10\mu > \lambda \geq \mu$, or only $\mu \geq \lambda$ (top to bottom). Let us note that in the case of complete selection the bottom boundary of the region of the Poisson model corresponds to the standard model. Although the model does not satisfy the formula in general, we may choose the parameters to do so.

Here is another way of comparing with the Gaussian model. Let the population frequency of a malformation be 0.00071 for males and 0.00317 for females, suppose there is no selection. The following table shows the conditional probabilities of the malformation in the first, second and third degree relatives in the Gaussian model. The rows correspond



(a) Complete selection (b) Partial selection

Figure 2.12: Model probabilities and the Edwards formula

to different genders of the malformed child, columns represent the degree of relationship and the gender of corresponding member of the family.

	I M	I F	II M	II F	III M	III F
M	0.0393	0.1149	0.0076	0.0276	0.0024	0.0195
F	0.0232	0.0749	0.0055	0.0204	0.0014	0.0087

Table 2.1: Conditional probabilities in the Gaussian model without selection

It is a remarkable property of the multifactorial threshold model that a relative with the gender of larger frequency of a malformed child with the gender of smaller frequency has the maximal conditional probability. The reason for the property is that the malformed child with smaller frequency has larger liability shifting the liability of his family upwards. The relatives with gender of larger frequency is evaluated with a smaller threshold which results in the mentioned property. The following table gives the conditional probabilities for the case with complete selection for the Gaussian model.

	I M	I F	II M	II F	III M	III F
M	0.0365	0.1025	0.0085	0.0255	0.0032	0.0113
F	0.0242	0.0739	0.0063	0.0234	0.0020	0.0088

Table 2.2: Conditional probabilities in the Gaussian model with complete selection

We compare these values with those coming from the Poisson model. We assume $\mu_f =$

μ_m, and use the remaining degree of freedom to get the highest conditional probabilities as mentioned in the beginning of this section. Having no selection means $\rho_f = \rho_m = 1$ but in this case we cannot apply Theorem 2.9. We rather choose $\rho_f = \rho_m = 1 - \varepsilon$ for some small $\varepsilon > 0$ to allow only negligible selection, but stay within the conditions of Theorem 2.9.

	I M	I F	II M	II F	III M	III F
M	0.1124	0.5015	0.0566	0.2523	0.1475	0.1722
F	0.1123	0.5012	0.0565	0.2521	0.1473	0.1721

Table 2.3: Conditional probabilities in the Poisson model with negligible selection

With complete selection:

	I M	I F	II M	II F	III M	III F
M	0.0452	0.2017	0.0135	0.0603	0.0342	0.0418
F	0.0452	0.2015	0.0135	0.0602	0.0342	0.0418

Table 2.4: Conditional probabilities in the Poisson model with complete selection

The reassuring fact we see is that we can set the conditional probabilities even higher than in the Gaussian model while leaving population probabilities unchanged.

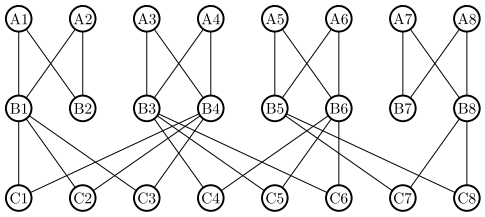


Figure 2.13: Model family

Next, we perform a Monte Carlo simulation on a model family given in Figure 2.13. We fix that A2, A4, A6, A8, B6 are women, A1, A3, A5, A7, B3 are men. The following numbers in Table 2.5 are probabilities conditioned on C5 having the malformation. We generated a large number of families starting from A1-A8 and only selected those where C5 was born and had the malformation. This explains the zeros in the first lines as they are all parents and consequently they are healthy. This does not hold for B1 as we allow him/her to be infertile thus C1 might not be born.

Gender of relative	index	A1	A2	A3	A4	A5	A6	A7	A8
M	M	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
M	F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
F	M	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
F	F	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

		B1	B2	B3	B4	B5	B6	B7	B8
M	M	0.00100	0.00071	0.00000	0.01853	0.00752	0.00000	0.00077	0.00077
M	F	0.00079	0.00064	0.00000	0.01890	0.00887	0.00000	0.00069	0.00068
F	M	0.00295	0.00305	0.00000	0.08952	0.03948	0.00000	0.00288	0.00293
F	F	0.00334	0.00290	0.00000	0.08458	0.03906	0.00000	0.00306	0.00310

		C1	C2	C3	C4	C5	C6	C7	C8
M	M	0.00701	0.00684	0.00813	0.04489	1.00000	0.04284	0.00221	0.00290
M	F	0.00706	0.00657	0.00717	0.04408	0.00000	0.04392	0.00310	0.00312
F	M	0.03072	0.03130	0.03076	0.18805	0.00000	0.19658	0.01346	0.01445
F	F	0.02944	0.02999	0.02988	0.19319	1.00000	0.19074	0.01485	0.01411

Table 2.5: Conditional probabilities in the Poisson model with complete selection

The gender of the affected child has seemingly no effect beyond randomness. One explanation for this phenomena is that in case of rare malformations the only effect that the affected child might cause is that he/she has a bad gene which is independent of gender differences. Using this setup also allows us to numerically compute more elaborate conditional and joint probabilities.

Another way to qualify the power of the Poisson model is to check its goodness-of-fit on the Hungarian data. In Table 2.6 we show the Poisson model fitted to 7 different data sets. The population data were gathered and published by Czeizel and Tusnády [12].

In Table 2.7 we present the goodness-of-fit values for the same data. We calculate the weighted average of the divergences for each relative. From another viewpoint, this is the normalized log-likelihood loss when changing real frequencies to the predicted probabilities.

table no	page	type of ICCA	Sex	pop freq 1000	No	Father			Mother			Brother			Sister		
						m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$
20	71	ASB	Boy	2.22	134	134	0	*0.0*	134	0	*0.0*	102	1	*2.3*	86	2	*2.8*
			Girl	3.59	309	309	0	*0.0*	309	0	*0.0*	210	5	*4.0*	177	4	*5.3*
30	96	CL(P)	Boy	1.33	369	304	3	*3.6*	304	5	*6.9*	143	16	*5.7*	121	0	*3.9*
			Girl	0.77	200	161	3	*2.4*	166	7	*4.9*	80	1	*4.4*	89	4	*4.1*
35	114	CHPS	Boy	2.18	112	112	1	*0.5*	112	2	*1.4*	48	2	*2.0*	38	2	*1.1*
			Girl	0.69	36	36	0	*0.5*	36	1	*1.4*	15	3	*1.3*	10	0	*0.8*
43	134	VSD	Boy	1.72	180	180	1	*0.4*	180	*2*	*1.8*	121	0	*2.6*	109	2	*2.1*
			Girl	1.41	197	197	0	*0.5*	197	3	*2.3*	96	2	*2.2*	81	1	*1.6*
56	166	CDH-BB	Boy	11.96	422	422	7	*6.3*	422	20	*32.2*	125	20	*22.6*	126	20	*32.9*
			Girl	39	1345	1345	9	*6.4*	1345	46	*32.7*	421	29	*34.3*	398	79	*72.4*
57	170	CDH-CB	Boy	8.13	75	75	1	*0.4*	75	2	*7.3*	22	2	*4.8*	21	7	*7.0*
			Girl	50.57	304	304	0	*0.4*	304	13	*7.3*	89	6	*6.8*	75	14	*17.0*
69	195	STEV	Boy	1.65	118	118	4	*2.7*	118	3	*1.5*	61	4	*2.6*	60	1	*2.2*
			Girl	0.82	56	56	1	*2.5*	56	0	*1.3*	29	2	*2.0*	30	3	*1.9*

		Paternal						Maternal						Paternal cousins						Maternal cousins					
		uncle			aunt			uncle			aunt			Boy			Girl			Boy			Girl		
		m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$	m	M	$\tilde{M}$
ASB	Boy	306	0	*4.5*	260	1	*6.6*	290	1	*4.3*	302	0	*7.6*	387	0	*5.7*	447	0	*11.3*	283	0	*4.2*	307	1	*7.8*
	Girl	505	1	*7.4*	512	0	*12.9*	497	3	*7.3*	534	0	*13.5*	850	2	*12.6*	785	2	*19.9*	606	3	*9.0*	621	3	*15.8*
CL(P)	Boy	283	3	*3.6*	307	3	*2.3*	383	0	*5.5*	366	2	*3.3*	362	1	*4.6*	374	0	*2.8*	365	1	*4.7*	387	2	*2.9*
	Girl	163	1	*2.1*	198	1	*1.5*	172	3	*2.7*	200	2	*2.0*	174	0	*2.2*	219	1	*1.6*	156	2	*2.0*	166	0	*1.3*
CHPS	Boy	153	0	*3.1*	146	0	*1.0*	180	1	*3.6*	166	1	*1.2*	130	2	*2.6*	136	0	*0.9*	154	1	*3.0*	144	1	*1.0*
	Girl	42	0	*0.9*	33	0	*0.3*	42	0	*0.9*	39	0	*0.4*	31	0	*0.6*	28	0	*0.2*	45	1	*0.9*	31	0	*0.2*
VSD	Boy	272	3	*4.3*	225	5	*3.0*	214	3	*3.4*	214	0	*2.9*	261	1	*4.1*	248	2	*3.3*	192	0	*3.0*	181	1	*2.4*
	Girl	208	3	*3.3*	168	1	*2.2*	188	4	*3.0*	209	3	*2.8*	239	0	*3.8*	168	1	*2.3*	175	0	*2.8*	174	0	*2.3*
CDH-BB	Boy	675	2	*29.7*	609	11	*92.7*	561	7	*29.4*	600	8	*95.3*	602	4	*29.3*	666	58	*103.8*	544	7	*26.6*	544	35	*84.9*
	Girl	1676	7	*71.5*	1473	25	*222.6*	1530	5	*69.4*	1607	55	*246.2*	1643	50	*79.7*	1637	128	*255.0*	1333	52	*64.8*	1282	172	*199.8*
CDH-CB	Boy	60	0	*2.4*	45	0	*9.0*	60	0	*3.1*	45	0	*9.4*	42	1	*2.0*	44	6	*9.0*	42	1	*2.0*	45	6	*9.2*
	Girl	273	2	*10.5*	263	6	*52.2*	273	1	*11.3*	263	6	*52.8*	197	4	*9.2*	203	12	*41.5*	197	4	*9.2*	204	12	*41.7*
STEV	Boy	140	0	*2.3*	161	2	*1.5*	141	0	*2.2*	151	1	*1.3*	131	1	*2.0*	130	0	*1.1*	148	2	*2.3*	121	0	*1.0*
	Girl	86	2	*1.5*	70	0	*0.7*	69	0	*1.1*	84	0	*0.8*	72	0	*1.1*	67	0	*0.6*	55	2	*0.8*	55	2	*0.5*

Table 2.6: Number of relatives (m), number of affected relatives in Hungarian data (M), expected number of affected relatives for the Poisson model ($\tilde{M}$)

disorder	GOF for all relatives	GOF for first order relatives
ASB	0.012189	0.000615
CLP	0.005341	0.008989
CHPS	0.007234	0.007099
VSD	0.005122	0.003212
CDH-BB	0.031767	0.002309
CDH-CB	0.050819	0.007456
STEV	0.007865	0.007432

Table 2.7: Goodness-of-fit of the Poisson model to Hungarian data

Finally let us present the parameter values for the best fit in Table 2.8.

disorder	μ_m	μ_f	ρ_m	ρ_f	Δ_m	Δ_f	λ_m	λ_f
ASB	0.015	0.026	0.018	0.010	0.018	0.010	0.00027	0.00026
CLP	0.012	0.0075	0.019	0.143	5.0e-14	0.085	0.00024	0.0012
CHPS	0.020	0.006	0.069	0.078	0.061	0.00052	0.0015	0.00052
VSD	0.016	0.013	0.0040	0.023	1.7e-17	1.3e-17	6.2e-5	0.00031
CDH-BB	0.036	0.175	0.028	0.142	3.4e-32	0.105	0.0014	0.027
CDH-CB	0.030	0.237	0.010	0.137	6.5e-16	0.102	0.00050	0.035
STEV	0.015	0.0073	0.091	0.048	0.047	1.2e-14	0.0015	0.00039

Table 2.8: Parameters of the Poisson model for Hungarian data

2.6 Conclusion

The form of selection investigated in this chapter is fortunate and ensures stability. The goodness-of-fit to population data is acceptable, the only problem is the extraordinarily small values for the parameter λ. This means that the number of bad genes is usually zero, and the appearance of a single bad gene causes the malformation or selection. Still, the low λ does not necessarily mean that the number of genes involved is small. As we mentioned in the introduction, we qualify our solution partial. It is a first acceptable solution for the problem resulting in a sound and practically applicable model. Still, the stability of the models with threshold remains open.

In a certain way the Poisson setup is richer then the Gaussian one as the expression of the malformation is randomized. The situation of this model is close to dominant

Mendelian inheritance with restricted expression. If the probability of the expression depends on the gender then the situation is rather complex. In the standard model the conditional probabilities resemble the formulas of Gaussian correlations. However, when allowing gender differences in the parameters the Poisson model becomes richer: conditional probabilities (of a relative being affected when the child is affected) show stronger gender dependence in the Poisson model than in the Gaussian one. Now we are facing the question, whether the Poisson model incorporated with environmental effects offer a substantially better goodness-of-fit than the Gaussian one.

Chapter 3

Markov chain mixing time estimates

In the previous chapter we have studied the convergence of a special stochastic process. In other cases, there are well know answers to these problems. For example, we know that an irreducible aperiodic Markov chain on a finite state space approaches its stationary distribution. Still, there is another fundamental question about the long term behavior of stochastic processes. It is natural to ask for the speed the distribution converges. This is especially important for applications, where the Markov chain is allowed to run for a limited number of time steps. One possibility to quantify this speed is by the introduction of *mixing time*. Let us define this quantity together with the concepts it relies on.

We work with discrete time Markov chains on a finite state space $\mathcal{X}$ which has size $n = |\mathcal{X}|$. For the set of probability distributions on the state space $\mathcal{X}$ we use the notation $\mathcal{P}(\mathcal{X})$.

We need a metric to measure the distance of probability distributions. One of the widely used options is the *total variation norm* defined as follows:

Definition 3.1. Given a signed measure ν on $\mathcal{X}$, the total variation norm is defined as

$$\|\nu\|_{\mathrm{TV}} = \max_{A \subseteq \mathcal{X}} |\nu(A)|.$$

Alternatively, one may use an L_2 distance or the divergence for comparison, but the investigation of these options is out of the scope of this thesis.

Now let us define the core notion of this chapter.

Definition 3.2. For a Markov chain with stationary distribution π and transition matrix $P = (p_{ij})$, with p_{ij} denoting the probability of moving from state i to state j, we define the mixing time of the chain as

$$t_{\mathrm{mix}} = t_{\mathrm{mix}}(P, \varepsilon) = \max_{\sigma \in \mathcal{P}(\mathcal{X})} \min \left\{ k : \|\sigma P^k - \pi\|_{\mathrm{TV}} \leq \varepsilon \right\}.$$

Note that this quantity might be infinite. We might omit some of the arguments of t_{mix} when they are not important or when they are known from the context.

For a survey on alternative metrics see Lovász, Winkler [41]. Mixing time is an important quantity directly related to the performance of numerous applications. In Markov chain Monte Carlo simulations mixing time can be interpreted as the time needed to generate a sample, see Metropolis et al. [42], Hastings [28] and Jerrum [30].

It turns out that running a local averaging algorithm is the same as following the evolution of the distribution of a certain Markov chain. For details see Olshevsky, Tsitsiklis [47], [48] or Boyd et al. [7]. Again, the time needed to get within a certain neighborhood of a common value is quantified by the mixing time. Motivated by these applications, the estimation of mixing time is in the center of interest.

A remarkable property of certain Markov chains is *reversibility* which often makes these approximations easier, see e.g. Kelly [34].

Definition 3.3. A Markov chain is *reversible* if starting from the stationary distribution π, the probability of the consecutive pair (i, j) is the same as the probability of the consecutive pair (j, i). Formally:

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \forall i, j.$$

The necessity of the separation of reversible and non-reversible Markov chains is widely recognized in literature. Often it is easier to prove useful properties for reversible chains, and there are tighter general bounds on mixing time for them. The reason to turn to non-reversible chains is the fact that they may deliver much faster mixing than a similar reversible chain.

As we mentioned in the beginning, we want quantifiable bounds for the mixing time. The definition is quite complicated for direct approximation, but fortunately there are multiple parameters that are more accessible and can be used to bound the mixing time.

Probably the most popular of these is the spectral gap.

Definition 3.4. Let $1 = \lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of the transition matrix P. Then the spectral gap of a Markov chain is defined by

$$\gamma_P = 1 - \max_{i \geq 2} \{|\lambda_i|)\}.$$

Most of the times we simply write γ when P is obvious from the context.

Not surprisingly the results differ for reversible and non-reversible chains. From Levin, Peres, Wilmer [38] and Diaconis and Saloff-Coste [17] we learn the following bounds for the mixing time:

Theorem 3.5. *For an irreducible, aperiodic Markov chain we have*

$$\left(\frac{1}{\gamma}-1\right)\log\left(\frac{1}{2\varepsilon}\right)\leq t_{\mathrm{mix}}(P,\varepsilon)$$

Moreover, if the chain is reversible, the following bound holds as well:

$$t_{\mathrm{mix}}(P,\varepsilon)\leq\frac{1}{\gamma}\log\left(\frac{1}{\varepsilon\pi_*}\right),$$

where $\pi_*=\min_i\pi_i$.

For reversible chains, we have a rather tight bound. Apart from constants, the ratio of the upper and lower bound is $\log 1/\pi_*$. If the stationary distribution is uniform, this is just a $\log n$ factor.

To shed light on the difference between reversible and non-reversible chains in the preceding theorem let us quote a result on total variation distances from Montenegro and Tetali [44]:

Theorem 3.6. *With the notation* $d(n)=\max_{\sigma\in\mathcal{P}(\mathcal{X})}\|\sigma P^n-\pi\|_{TV}$ *we have the following two inequalities:*
For irreducible, aperiodic, reversible chains:

$$\frac{1}{2}(1-\gamma)^n\leq d(n)\leq\frac{1}{2}(1-\gamma)^n\sqrt{\frac{1-\pi_*}{\pi_*}}$$

For irreducible, aperiodic, non-reversible chains:

$$\frac{1}{2}(1-\gamma)^n\leq d(n)\leq\frac{1}{2}(1-\gamma_{PP^*})^{n/2}\sqrt{\frac{1-\pi_*}{\pi_*}}$$

The upper bound for non-reversible chains is due to Fill, Allen [23] and it uses the matrix PP^*. It is possible to have $\gamma_{PP^*}=0$. That means we do not expect a strong general spectral upper bound on mixing time for non-reversible chains.

There are other tools that also work well for non-reversible chains, the one we heavily rely on is using the *conductance* of a Markov chain introduced by Jerrum and Sinclair [31]. This is a quantity somehow measuring the worst bottleneck of the chain.

Definition 3.7. The conductance of a Markov chain is

$$\Phi=\min_{\emptyset\neq S\subsetneq\mathcal{X}}\Phi(S)=\min_{\emptyset\neq S\subsetneq\mathcal{X}}\frac{Q(S,S^C)}{\pi(S)\pi(S^C)}=\min_{\emptyset\neq S\subsetneq\mathcal{X}}\frac{\sum_{i\in S,j\in S^c}\pi_ip_{ij}}{\pi(S)\pi(S^C)},$$

where $S^C = \mathcal{X} \setminus S$, the complement of the set S. This quantity is strongly related to mixing times. This neat idea has been evolving since into different refined concepts as average conductance (see Lovász, Kannan [39]), and blocking conductance (see Kannan, Lovász, Montenegro [32]), etc. The simplest universal bound is the following, see Borgs [4] for a proof.

Proposition 3.8. *For any discrete time Markov chain we have the bound*

$$\frac{c}{\Phi} \leq t_{\text{mix}}.$$

The conductance can also be used to provide an upper bound on the mixing time. However, we need some additional condition for the Markov chain. The classical result is due to Jerrum and Sinclair [50]:

Theorem 3.9. *For an aperiodic, irreducible, reversible Markov chain the following bounds for mixing time holds:*

$$c_1 \frac{1}{\Phi} \leq t_{\text{mix}} \leq c_2 \frac{1}{\Phi^2} \log\left(\frac{1}{\pi_*}\right).$$

Let us also cite the version of Lovász and Simonovits [40]. This theorem does not require the reversibility of the chain but assumes that it is *lazy*. A Markov chain is lazy if $p_{ii} \geq 1/2$ for all i:

Theorem 3.10. *For an aperiodic, irreducible, lazy Markov chain the following bounds for mixing time holds:*

$$c_1 \frac{1}{\Phi} \leq t_{\text{mix}} \leq c_2 \frac{1}{\Phi^2} \log\left(\frac{1}{\pi_*}\right).$$

The stationary distribution is uniform for all the Markov chains we work with, so the last logarithmic factor simplifies to $\log n$.

There is a square factor between the lower and upper bounds so we will need additional tools whenever we look for the exact magnitude of the mixing time.

Besides estimating the mixing time of a certain Markov chain, we might aim for improving the algorithms mentioned in the beginning by modifying the Markov chain such that the mixing time decreases. We consider only the case when the stationary distribution is uniform. For the transition matrix this translates to the condition of being doubly stochastic.

As a starting point, let us review Example 6.6. of the survey by Montenegro and Tetali [44]. We see two quite similar Markov chains on $2n$ nodes in Figure 3.1. Both the chains

stay put with probability 1/2, other transition probabilities are indicated. On the left side, the state of the chain rotates around the cycles, on the right side, we see a symmetric random walk. This change results in rather different mixing times, cn for the left chain and cn^2 for the right one.

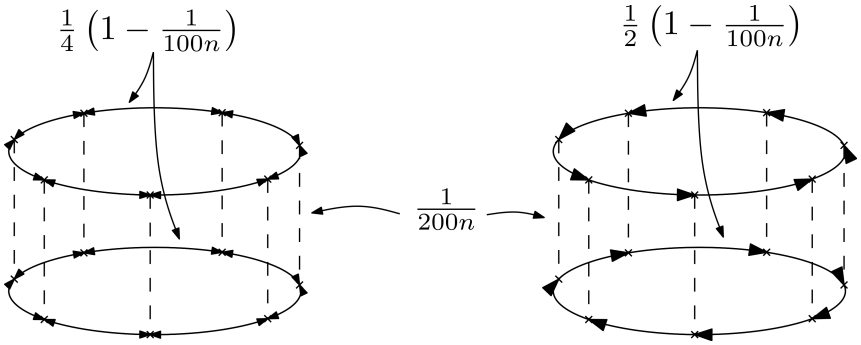


Figure 3.1: Non-reversible and reversible chains on the double cycle

We want to do something similar to speed up other Markov chains. We change transition probabilities but not the allowed transitions. For this purpose, let us define the *connectivity graph* in the following way:

Definition 3.11. The connectivity graph of a Markov chain is a graph on the states of the Markov chain. We connect nodes $i \neq j$ if either $p_{ij} > 0$ or $p_{ji} > 0$.

We shall also refer to this graph loosely as the topology of the Markov chain.

While not completely solving the problem, there are reassuring results on finding the fastest mixing reversible chains with fixed connectivity graph, see Boyd et al. [6], [5]. They can formulate this task as a semi-definite programming (SDP) problem, thus it is possible to determine the fastest reversible chain as long as there is a sufficiently fast SDP solver available.

Unfortunately there is no such result for non-reversible chains. Theorem 3.13 contributes to this topic by clarifying the situation for specific connectivity graphs.

One of the central question of this chapter is comparing the best reversible and non-reversible chains for a fixed connectivity graph. Although we do not know the amount of speedup in general, at least we have a theoretical limit, see Chen, Lovász and Pak [9] for further thoughts.

Proposition 3.12. *For some fixed connectivity graph let P and $\tilde{P}$ be the transition matrices of the best reversible and non-reversible chains, respectively. Then for the mixing times we have*

$$t_{\text{mix}}(P) \leq c t_{\text{mix}}^2(\tilde{P}) \log n.$$

Proof. Let us define $P' = (\tilde{P} + \tilde{P}^T)/2$. It is easy to see that P' is the transition matrix of a reversible Markov chain with the same connectivity graph (here we use that the stationary distribution is uniform). Moreover, observe that $\Phi_{P'}(S) = \Phi_{\tilde{P}}(S)$ for any $S \subset \mathcal{X}$ thus $\Phi_{P'} = \Phi_{\tilde{P}}$. Using Theorem 3.10 this implies

$$t_{\text{mix}}(P') \leq c_1 \frac{1}{\Phi_{P'}^2} \log n = c_1 \frac{1}{\Phi_{\tilde{P}}^2} \log n \leq c_2 t_{\text{mix}}^2(\tilde{P}) \log n.$$

The matrix P' might not be the best choice for a reversible transition matrix, but substituting it with a better P just further decreases the left hand side. □

To increase our freedom, we might allow certain changes in the connectivity graph as well. To demonstrate the possibilities, let us cite a concept by Diaconis, Holmes, Neal [16], optimized by Gade and Overton [24] and extended by Chen, Lovász and Pak [9]. This is a method to decrease the mixing time of a reversible chain up to its square root by modifying it to a non-reversible one. Here the topology of the chain changes as every node is split into multiple copies. Transition probabilities are chosen such that the marginal behaves like the original chain, but we achieve faster mixing on the new graph. The method is called *lifting*. If we look at Figure 3.1 from a different aspect, we may view the Markov chain on the left as a lifting of a symmetric random walk on a cycle.

Although this is a powerful example to show what one can achieve, in most cases the limit of how much speedup is possible is not clear. Next, in Section 3.1 we show what we can do in a specific case by relaxing the reversibility condition but not changing the connectivity graph. In Section 3.2 we keep the states of the Markov chain but allow a low number of new transitions to be introduced and we investigate the effect on the mixing time. In Section 3.3 we discuss some future research questions.

3.1 Relaxing the reversibility condition

We restrict the connectivity graph to a cycle, and allow arbitrary non-reversible transition probabilities such that the uniform distribution is invariant. Then there is a lower bound on the mixing time which has the same order of magnitude as the best lower bound for reversible chains.

For convenience, let us number the nodes according to the ordering on the cycle. We will interpret these numbers *mod* n.

We are now ready to state the main result of this section:

Theorem 3.13. *Consider a Markov chain on a cycle with n nodes having a doubly stochastic transition matrix P. Then, with some global constant $C > 0$ we have*

$$t_{\text{mix}}(P, 1/8) \geq Cn^2.$$

This has also been published by the author [25]. Note that our theorem covers all Markov chains, even non-reversible ones. Our work goes back to the basics. We search for the exact limit of what can be achieved by allowing a non-reversible chain for a given topology, in our case a cycle. It is known that the magnitude of the best mixing time of a reversible chain on a cycle scales with n^2 (we will present a proof, see Lemma 3.24). Our theorem implies that relaxing the reversibility condition does not help with this topology.

The claim of our work is simple to state, however, we did not succeed in proving using conventional methods. We had to search further and use a unique approach, presented in this section. As a result, some interesting properties of these Markov chains arise as a by-product.

The rest of the section is structured in the following way. In Subsection 3.1.1 we prepare the proof and split it into two parts. We have to work on them separately, Subsections 3.1.2 and 3.1.3 deal with these parts.

3.1.1 Preparation for the proof

To set up, let us collect some simple observations. First, let us note that in our case of finite state space

$$\|\mu - \sigma\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \sigma(x)| = \frac{1}{2} \|\mu - \sigma\|_1.$$

We should point it out that the TV distance is defined for measures, l_1 is for (real) vectors. In our case we can interpret measures as real vectors, so that this equation makes sense. This means we do not need to use the TV distance but can work with the l_1 norm instead. With this change we have to find when the l_1 distance decreases below $1/4$ to determine the appropriate mixing time.

Second, let us prove a lemma on the structure of the transition matrix.

Lemma 3.14. *The doubly stochastic transition matrix P of a Markov chain on a cycle can be decomposed as $P = Q + rR$, where Q is the transition matrix of a reversible chain on a cycle, and R is*

$$\begin{pmatrix} & 1 & & & & -1 \\ -1 & & 1 & & & \\ & -1 & & \ddots & & \\ & & \ddots & & & \\ & & & & & \\ & & & & & 1 \\ 1 & & & & -1 & \end{pmatrix}.$$

Proof. Let us start with

$$P = \frac{P + P^T}{2} + \frac{P - P^T}{2} = A + B.$$

The choice $Q = A$ clearly satisfies the conditions we have on Q. It is easy to see that B is antisymmetric, and all row and column sums are 0. Set $r = B_{12} = -B_{21}$. Then $B_{23} = -B_{32} = r$ required by the second row and column sum to be zero. Repeat this to get $B = rR$. □

For convenience, we introduce simplified indices for the elements of Q we use often: $q_i = Q_{i-1,i} = Q_{i,i-1}$.

The presence of the rR term has the heuristic effect that the chain is more likely to travel in one direction than the other. This is some sort of rotation, which will play a crucial role in our proof.

Reversing the numbering of the nodes swaps the sign of r, so without loss of generality, we may assume $r \geq 0$ Let $\mathcal{M}$ be the set of doubly stochastic transition matrices of a Markov chain on a cycle. Let $\mathcal{M}_0 \subset \mathcal{M}$ be the subset of reversible ones.

Third, let us provide a tool to simplify further discussions:

Lemma 3.15. *Let us choose any dense subset $\mathcal{N}$ of $\mathcal{M}$, some $\varepsilon > 1/8$ and K. Then*

$$\forall P \in \mathcal{N} \;\; t_{\text{mix}}(P, \varepsilon) \geq K \Longrightarrow \forall P \in \mathcal{M} \;\; t_{\text{mix}}(P, 1/8) \geq K.$$

Proof. For any matrix $P \in \mathcal{M}$ we have $\|P\|_1 \leq 1$, where the norm is the operator norm w.r.t. the l_1 norm (in fact we have $\|P\|_1 = 1$). It follows that for any two matrices $P, P' \in \mathcal{M}$,

$$\|P^K - P'^K\|_1 \leq \|P^K - P^{K-1}P'\|_1 + \|P^{K-1}P' - P^{K-2}P'^2\|_1 + \ldots +$$

$$+\|PP'^{K-1} - P'^K\|_1 \leq K\|P - P'\|_1.$$

For any $P \in \mathcal{M}$ choose $P' \in \mathcal{N}$ such that $\|P - P'\|_1 < (\varepsilon - 1/8)/K$. There is an $x \in \mathbb{R}^n$ showing $t_{\text{mix}}(P', \varepsilon) \geq K$. For this x,

$$\left\| xP^K - \frac{\mathbf{1}}{n} \right\|_1 \geq \left\| xP'^K - \frac{\mathbf{1}}{n} \right\|_1 - \left\| xP^K - xP'^K \right\|_1 > 2\varepsilon - K\frac{\varepsilon - \frac{1}{8}}{K} > \frac{1}{4}.$$

This confirms $t_{\text{mix}}(P, 1/8) \geq K$. □

We will use this lemma multiple times when we need some extra property for the matrix (such as all eigenvalues are different) which does not hold for all matrices in $\mathcal{M}$. Observe that we can use the lemma independently multiple times if $\mathcal{N}$ is residual each time.

From now on, we have to continue on two tracks. The interesting thing is that we can not prove Theorem 3.13 by a single method. In the following two subsections we introduce two arguments, one works in the "general" case, when $r > c/n$ and the other works where the chain is almost reversible in the sense that $0 \leq r \leq c/n$. None of the two arguments can be naturally carried over to the other domain.

The status of c/n is also different in the two parts. In the first part, the value of c is obtained from the proof and is not convenient to change. However, the second argument works for arbitrary c. Of course the resulting bound on the mixing time depends on the choice of c. Using this flexibility it is enough to prove these two parts as they can be stitched together to cover all possible chains.

3.1.2 General non-reversible chains

In this subsection we deal with the case when $r > c/n$, in other words when the chain is "far from reversible".

Theorem 3.16. *Given a Markov chain on an n node cycle consider the doubly stochastic transition matrix $P = Q + rR$ as in Lemma 3.14. If $r > 2^{11}/n$, then*

$$t_{\text{mix}}(P, 1/8) \geq \frac{1}{2^{12}} n^2.$$

First we give a very short outline of the proof. We use variables not yet defined and relations not yet shown, the point is to sketch the formal structure of the proof.

As a start let us look at a series of vectors x^l approximately following the evolution of the chain:

$$x^l P = x^{l+1} + e^l,$$

with x^1 being a probability distribution. Observe that P does not increase the l_1 norm, this confirms the following:

$$\begin{aligned} \left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 &\geq \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \left\| x^{k-1} P - x^k \right\|_1 \\ &\quad - \left\| x^{k-2} P^2 - x^{k-1} P \right\|_1 - \ldots - \left\| x^1 P^{k-1} - x^2 P^{k-2} \right\|_1 \qquad (3.1) \\ &\geq \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \sum_{l=1}^{k-1} \left\| e^l \right\|_1 . \end{aligned}$$

The left hand side is the quantity we need to keep above $1/4$ as long as possible to ensure a large mixing time. For all l we may use the bound $\|e^l\|_1 < B$, and for an

appropriate k we have $\|x^k - \mathbf{1}/n\|_1 > A$. Now using $k \geq n^2/2^{12}$ and $A - kB > 1/4$, we get the bound on the mixing time we are aiming for.

The following things are left.

We have to construct the series x^l. It needs to approximately follow the effect of P so that e^l is small. We also want to easily access elements with high indices in order to have a lower bound of the type $\|x^k - \mathbf{1}/n\|_1 > A$. In the end the structure that will give us these vectors will be completely different from a Markov chain, but with the proper tuning it will coincide with it in some sense.

Then we need to prove the lower and upper estimates we used above.

The construction

The main idea is to find x^l in such a way that x^{l+1} is obtained from x^l by a kind of rotation. To define the rotation of a vector we proceed as follows. We consider the unit circle and we fix a function f defined on the circle. We will fix a set of n "observation points" $Z_0, Z_1, \ldots, Z_{n-1}$, and define

$$y_i^0 = f(Z_i).$$

The rotation of the vector $y^0 = (y_i^0)$ is constructed via the rotation of f, defined as

$$f^\alpha\left((\cos(u+\alpha), \sin(u+\alpha))\right) = f\left((\cos(u), \sin(u))\right).$$

Then define

$$y^\alpha = (f^\alpha(Z_0), f^\alpha(Z_1), \ldots, f^\alpha(Z_{n-1})).$$

When we use angles we mean them as *mod* 2π numbers. Obviously, the vectors y^α need not to be probability vectors, so they will have to be normalized. This will be much easier to describe later, let us leave this for now.

Now, let us specify the functions and variables introduced, starting with f^α. This is piecewise linear in the angle:

$$f^\alpha\left((\cos(u+\alpha), \sin(u+\alpha))\right) = \left|\frac{u}{2\pi}\right|, \ u \in [-\pi, \pi).$$

This implies that a rotation by a small angle φ would entail a change in y_i^α by an amount of $\pm\varphi/(2\pi)$, except perhaps for the indices corresponding to observation points near α and $\pi + \alpha$.

To achieve a similar effect as this rotation by the Markov dynamics, we need

$$(y^\alpha P)_i - y_i^\alpha = \pm\lambda$$

for some constant λ, and as many (α, i) pairs as possible. We won't solve this right away, but use it as a motivation. Let us write out the left side:

$$(yP)_i - y_i = y_{i-1}(q_i + r) + y_i(1 - q_i - q_{i+1}) + y_{i+1}(q_{i+1} - r) - y_i =$$
$$= -(y_i - y_{i-1})(q_i + r) + (y_{i+1} - y_i)(q_{i+1} - r).$$

Roughly speaking, the use of the functions f^α implies that most $y_i - y_{i-1}$ are proportional to the angular difference of Z_i and Z_{i-1} (neglecting the sign). Let us replace $y_i - y_{i-1}$ with δ_i in the equation above, and think of δ_i as this angular difference. We will properly explain this $y - Z - \delta$ relation later.

For the right hand side, let us choose $\lambda = -2r$ (this is a convenient, but arbitrary choice) and drop the sign so that we end up with the system of equations:

$$-\delta_i(q_i + r) + \delta_{i+1}(q_{i+1} - r) = -2r, \quad i = 0, 1, \ldots, n-1. \tag{3.2}$$

Now the key point is that this system has a positive solution in δ_i. The following lemma ensures this positive solution exists. Once we have δ_i at our hands, we will properly specify Z_i and thus y_i.

Lemma 3.17. *Consider the system of equations*

$$-u_i a_i + u_{i+1} b_{i+1} = -c_i, \quad i = 0, 1, \ldots, n-1.$$

Suppose $a_i > b_i > 0$, $c_i > 0$ for $i = 0, 1, \ldots, n-1$. The indices are taken mod n. Then the system has a unique, positive solution in u_i, $i = 0, 1, \ldots, n-1$.

Proof. We may rearrange the equation to

$$u_{i+1} = u_i \frac{a_i}{b_{i+1}} - \frac{c_i}{b_{i+1}}.$$

This is a linear equation where u_i has a positive coefficient, and a positive constant is subtracted. We can start with $i = 0$ to get an expression for u_1 in terms of u_0. Then we plug this into $i = 1$, and so on. After going through the full cycle, we end up at

$$u_0 = Au_0 - C.$$

Here $C > 0$ because it is the sum of positive numbers, and $A = \frac{\prod_{i=0}^{n-1} a_i}{\prod_{i=0}^{n-1} b_i} > 1$. So the solution $u_0 = C/(A-1)$ is positive. Plugging this back allows us to compute all other u_i and we just made it sure that it will be consistent when we arrive back to u_0.

Suppose $u_i \leq 0$ for some i. From the equation it follows that $u_{i+1} < 0$. If we continue this we find $u_0 < 0$ which is impossible, so indeed $u_i > 0$.

Uniqueness is clear by the method we described. □

Let us add up Equation (3.2) for all i. Lot of terms cancel out and we get

$$\sum_{i=0}^{n-1} \delta_i = n.$$

As we said before, we want these δ_i to be proportional to the angles between Z_i's. In order to fit these on the circle, we have to scale them down. Let Z_0 be the point at angle 0, and Z_i be the point at angle $2\pi \sum_{j=1}^{i} \delta_j/n$.

Let us check the construction. On the half circle where f increases with the angle $y_i^\alpha - y_{i-1}^\alpha = \delta_i/n$, so by Equation (3.2),

$$(y^\alpha P)_i - y_i^\alpha = -2r/n.$$

The same happens on the other half but with opposite signs. Naturally the nodes near α and $\pi + \alpha$ may behave differently, and we have to make sure they stay under control. This change of $\pm 2r/n$ corresponds to a $4\pi r/n$ angle rotation of f^α. This justifies the definition of the error term

$$d^\alpha = y^\alpha P - y^{\alpha + \frac{4\pi r}{n}}. \tag{3.3}$$

After describing our variables we need to prove the bounds used in the outline of the proof.

Bounds on errors

First we prove a bound on δ_i.

Lemma 3.18. *For every i,*

$$\delta_i \leq \frac{2}{q_i}.$$

Proof. Let us start from equation (3.2) on δ_i. We can write it in the following way:

$$(q_i + r)\delta_i - (q_{i+1} - r)\delta_{i+1} = 2r.$$

If $\delta_i > 2$ (or $\delta_{i+1} > 2$) it follows that

$$q_i\delta_i - q_{i+1}\delta_{i+1} < 0.$$

Now suppose $\delta_i > \frac{2}{q_i}$. This is clearly more than 2, so we have

$$\delta_{i+1} > \frac{q_i}{q_{i+1}}\delta_i > \frac{2}{q_{i+1}}.$$

We can continue this argument for the next index:

$$\delta_{i+2} > \frac{q_{i+1}}{q_{i+2}}\delta_{i+1} > \frac{q_{i+1}}{q_{i+2}}\frac{q_i}{q_{i+1}}\delta_i = \frac{q_i}{q_{i+2}}\delta_i > \frac{2}{q_{i+2}}.$$

After doing this n times, we end up with $\delta_i > \delta_i$ which is a contradiction, so the claim of the lemma is indeed true.

□

The previous lemma helps to bound d^α. This d^α will become the error term e^l used in the outline of the proof after proper scaling.

Lemma 3.19. *For every α,*

$$\|d^\alpha\|_1 \leq \frac{24}{n}.$$

Proof. If we pick a node i, and f^α is linear on the joint arc between Z_{i-1} and Z_{i+1}, things work as we designed them, and $(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}} = 0$. There are two irregular arcs, those containing α and $\pi+\alpha$, this effects at most four nodes. Let us focus on these nodes.

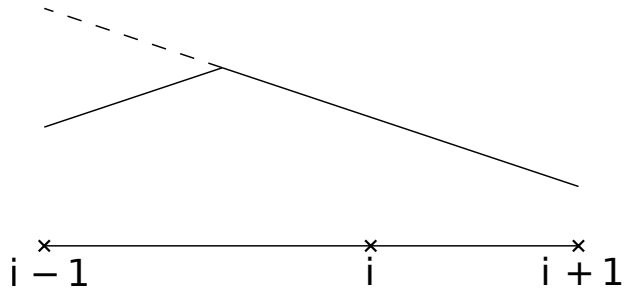


Figure 3.2: Node near the peak

There would be no error at node i if we used the dashed line, so we have to measure the difference caused by switching to the real, solid line.

The slope of the line is $1/(2\pi)$ so the difference at y_{i-1}^α is at most δ_i/n. During the rotation, the peak might reach Z_i so the value of $y_i^{\alpha+\frac{4\pi r}{n}}$ might deviate at most $4r/n$ from the dashed line. Adding up these two sources of error we get

$$\left|(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}}\right| \leq \frac{\delta_i}{n}(q_i + r) + \frac{4r}{n}.$$

Let us note $q_i + r$ and $q_i - r$ are both transition probabilities, thus $r \leq q_i$ and $r \leq 1/2$.

$$\left|(y^\alpha P)_i - y_i^{\alpha+\frac{4\pi r}{n}}\right| \leq 2\frac{\delta_i}{n}q_i + \frac{2}{n} \leq \frac{6}{n}.$$

The last inequality follows from Lemma 3.18. The same bound is true if the peak is between Z_i and Z_{i+1}. Adding four of these and a few zeros proves the lemma.

□

Bounds on initial distance

Although we want to use y^α for x^1, it is generally not a probability distribution, so we have to figure out how to scale it. Observe that

$$y^\alpha + y^{\pi+\alpha} = \frac{\mathbf{1}}{2}.$$

Consequently $\|y^\alpha\|_1 + \|y^{\pi+\alpha}\|_1 = \frac{n}{2}$. The value $\|y^\beta\|_1$ is continuous in β, so we can choose β such that

$$\left\|y^\beta\right\|_1 = \frac{n}{4}. \tag{3.4}$$

This β will be fixed from now on, and also $x^1 = \frac{4}{n}y^\beta$, which is now a valid probability distribution.

The last building block of the proof can be summarized in the following lemma:

Lemma 3.20. *Suppose the assumptions of Theorem 3.16 holds. Then there exist a* $k \in [\frac{1}{2^{12}}n^2 + 1, \frac{1}{2^{11}}n^2 + 1]$ *such that*

$$\left\|\frac{4}{n}y^{\beta + k\frac{4\pi r}{n}} - \frac{\mathbf{1}}{n}\right\|_1 > \frac{1}{3}.$$

We need some simple lemmas to prove this. Let us introduce the notation

$$s(\alpha) = \left\|\frac{4}{n}y^\alpha - \frac{\mathbf{1}}{n}\right\|_1.$$

Lemma 3.21. *The function* s *cannot change too fast:*

$$|\overline{s}'(\alpha)| \le \frac{2}{\pi},\ |\underline{s}'(\alpha)| \le \frac{2}{\pi} \quad \forall \alpha \in [0, 2\pi).$$

Here $\overline{s}'$ *and* $\underline{s}'$ *are the upper and lower derivatives, respectively. On the other hand, for the average value:*

$$\frac{1}{2\pi}\int_0^{2\pi} s(\alpha)d\alpha = \frac{1}{2}.$$

Proof. The derivative of $\frac{4}{n}y_i^\alpha$ is in $[-\frac{2}{n\pi}, \frac{2}{n\pi}]$. This also holds if we subtract a constant and take absolute value. If we add up n of these, we get exactly what we stated.

For the second claim,

$$\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{4}{n}y_i^\alpha - \frac{1}{n}\right|d\alpha = \frac{1}{2}\int_0^2\left|\frac{u}{n} - \frac{1}{n}\right|du = \frac{1}{2n}.$$

Adding these up gives the second formula. □

Lemma 3.22. *The function* $s(\alpha)$ *is continuous and piecewise linear with at most* $4n$ *segments on* $[0, 2\pi]$*, assuming 0 and* 2π *are stitched together.*

Proof. Again,

$$s_i(\alpha) = \left| \frac{4}{n} y_i^\alpha - \frac{1}{n} \right|$$

is piecewise linear with four segments (four points of nonlinearity). If we add up n of such functions, it will have at most $4n$ segments. □

Now we turn back to the lemma we left over.

Proof of Lemma 3.20: Suppose the claim doesn't hold. Let us mark the set G of "good" points in the following sense:

$$G = \left\{ \alpha \in [0, 2\pi],\ s(\alpha) > \frac{1}{3} \right\}.$$

Let's look at Lemma 3.22. While we go around the circle on each segment we might step in or out of G, but at most once. This means G is the union of at most $2n$ intervals.

On the range of k we are working on, we are rotating

$$\frac{n^2}{2^{12}} \frac{4\pi r}{n} = 2\pi \frac{rn}{2^{11}} > 2\pi.$$

In other words, we rotate through the whole circle. This is the point where we use the lower bound on r. If the claim does not hold, it means we never hit G as k sweeps its range. This means we jump over every interval when we reach it. Consequently each interval is at most $\frac{4\pi r}{n}$ long.

At both ends of such an interval $s(\alpha) = 1/3$, so by the bound in Lemma 3.21 $s(\alpha)$ can increase up to at most $\frac{1}{3} + \frac{2\pi r}{n} \frac{2}{\pi}$ on such a short interval. We can construct an upper estimate on the average distance using different bounds on G and outside G. Using Lemma 3.21 again for the average value gives us the following:

$$2\pi \cdot \frac{1}{2} \leq \frac{4\pi r}{n} 2n \cdot \left(\frac{1}{3} + \frac{2\pi r}{n} \frac{2}{\pi} \right) + \left(2\pi - \frac{4\pi r}{n} 2n \right) \cdot \frac{1}{3}.$$

Rearranging this gives

$$r \geq \sqrt{\frac{n}{96}}.$$

We know r is at most $1/2$. By the condition $r > 2^{11}/n$ we also have $n > 2^{12}$. But then the right hand side becomes more than $1/2$ and this leads us to contradiction. □

The proof of Theorem 3.16

It only remained to put things together. Based on Equation (3.4) we defined $x^1 = \frac{4}{n} y^\beta$ as the starting probability distribution. The scaled versions of the rotated vectors are

$$x^l = \frac{4}{n} y^{\beta + (l-1)\frac{4\pi r}{n}}.$$

The error terms bounded in Lemma 3.19 scale by the same factor thus we may define

$$e^l = \frac{4}{n} d^{\beta+(l-1)\frac{4\pi r}{n}}.$$

Using these notations, Equation (3.3) defining d^α becomes

$$e^l = x^l P - x^{l+1}.$$

Recall we started from Equation (3.1):

$$\left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 \geq \left\| x^k - \frac{\mathbf{1}}{n} \right\|_1 - \sum_{l=1}^{k-1} \|e^l\|_1.$$

Let us choose k from Lemma 3.20, then the first term in the right hand side is more than $1/3$. By the definition of e^l and Lemma 3.19 we have $\|e^l\|_1 < 96/n^2$. Plugging these in and using the bound on k gives

$$\left\| x^1 P^{k-1} - \frac{\mathbf{1}}{n} \right\|_1 > \frac{1}{3} - (k-1)\frac{96}{n^2} \geq \frac{1}{3} - \frac{3}{64} > \frac{1}{4}.$$

Consequently

$$t_{\text{mix}}(P, 1/8) \geq k - 1 \geq \frac{1}{2^{12}} n^2.$$

3.1.3 Almost reversible chains

In this subsection we will prove the following:

Theorem 3.23. *Given a Markov chain on an n node cycle consider the doubly stochastic transition matrix $P = Q + rR$. Suppose $0 \leq r \leq c/n$ for some fixed $c > 0$. Then there is a $c' > 0$ such that*

$$t_{\text{mix}}(P, 1/8) \geq c' n^2,$$

and c' depends only on c.

The idea is to compare our chain to a reversible one. We try to estimate the errors when r is small enough. We do this first with an additional condition on the chain, but we will be able to relax it later.

The reversible case

Let us see how does the proof go if the transition matrix is symmetric. Our argument will be slightly different and more constructive than the usual eigenvalue estimation.

To reduce complexity, we state and prove Lemma 3.24 only if n is even. The same argument works for the odd case, we only have to do trivial adjustments.

Lemma 3.24. *Suppose Q is as before, n is even. Then for the initial distribution*

$$x = \frac{4}{n^2}\left(0, 1, \ldots, \frac{n}{2} - 1, \frac{n}{2}, \frac{n}{2} - 1, \ldots, 2, 1\right),$$

some global $c_1 > 0$ and for any $k \leq c_1 n^2$ we have the bound

$$\left\| xQ^k - \frac{\mathbf{1}}{n} \right\|_1 > \frac{5}{12}.$$

Consequently,

$$t_{\text{mix}}(Q, 1/8) > c_1 n^2.$$

Proof. Let us consider the vector

$$x_0 = \frac{4}{n^2}\left(0, 1, \ldots, \frac{n}{2} - 1, \frac{n}{2}, \frac{n}{2} - 1, \ldots, 2, 1\right) - \frac{\mathbf{1}}{n}.$$

This is almost the same as x, where $4/n^2$ is chosen to normalize the vector in parentheses to a probability distribution. Then we subtract the uniform distribution to make x_0 orthogonal to it. (If n was odd, the maximal coordinate would be $(n+1)/2$ and we would have an extra 0 in the end.)

We will split $x_0 Q^k$ into two components. One pointing in the x_0 direction, providing the vector is far from uniform, and another perturbing this. We want the first to be large, the second to be small. Let's start estimating the first.

It is well known that the Laplacian of the chain is $I - Q$ and that

$$x_0(I - Q)x_0^T = \frac{1}{2}\sum_{i,j}(x_{0,i} - x_{0,j})^2 Q_{ij}.$$

The nonzero terms of this sum are $16/n^4 \cdot Q_{ij}$. If we add these up, we get

$$x_0(I - Q)x_0^T = \frac{8}{n^3}.$$

On the other hand $x_0 x_0^T = 1/(3n) + 8/(3n^2) > 1/(3n)$, so it follows that

$$\frac{x_0(I - Q)x_0^T}{x_0 x_0^T} < \frac{24}{n^2}.$$

Using Lemma 3.15 we may assume all eigenvalues of Q are different. Moreover, the matrix Q is symmetric so its eigenvectors e_i form an orthonormal basis. Let the corresponding real eigenvalues be λ_i. We can express x_0 in this base as $x_0 = \sum_i \alpha_i e_i$ for some α_i. Using these notations we may rewrite the previous equation as

$$\frac{\sum_i (1 - \lambda_i)\alpha_i^2}{\sum_i \alpha_i^2} < \frac{24}{n^2}.$$

It is clear that $1-\lambda_i^k < k(1-\lambda_i)$ for any $\lambda_i \in [-1,1]$, so it follows that

$$\frac{\sum_i (1-\lambda_i^k)\alpha_i^2}{\sum_i \alpha_i^2} < \frac{24k}{n^2},$$

or with the original matrix notation

$$\frac{x_0(I-Q^k)x_0^T}{x_0 x_0^T} < \frac{24k}{n^2},$$

$$\frac{x_0 Q^k x_0^T}{x_0 x_0^T} > 1 - \frac{24k}{n^2}.$$

This is what we need for the part pointing in the x_0 direction, so let us now focus on the remainder.

Let us look at the orthogonal decomposition $x_0 Q^k = \alpha x_0 + y$, where $\alpha > 1 - 24k/n^2$ according to the previous estimate. The matrix Q is non-expanding w.r.t the $\|.\|_2$ norm, so we have

$$\alpha^2 \|x_0\|_2^2 + \|y\|_2^2 \leq \|x_0\|_2^2.$$

We need to transform this inequality to bound $\|y\|_1$. We can do this using the inequality of arithmetic and quadratic means:

$$\frac{\|y\|_1^2}{n} \leq \|y\|_2^2 \leq \|x_0\|_2^2(1-\alpha^2).$$

Here $\|x_0\|_2^2 = 1/(3n) + 8/(3n^2) < 2/n$ for $n \geq 2$. The final estimate is

$$\left\| \left(x_0 + \frac{\mathbf{1}}{n} \right) Q^k - \frac{\mathbf{1}}{n} \right\|_1 \geq \alpha \|x_0\|_1 - \|y\|_1 \geq \alpha \frac{1}{2} - \sqrt{2(1-\alpha^2)}.$$

It is easy to verify that this is more than $5/12$ if $\alpha > 599/600$. We can ensure this whenever $k < n^2/15000$, so in the end we get that $c_1 = 1/15000$ is a sufficient choice for the lemma to be true.

□

Non-reversible, but lazy chains

As we outlined before, we want to relate our generic chain to a reversible one. We use the vector x previously defined. Let us look at the following decomposition:

$$x(Q+rR)^k = xQ^k + \sum_{l=1}^{k} xQ^{l-1}rR(Q+rR)^{k-l}. \tag{3.5}$$

We know how the first term behaves, so we need to see that the other term is small. $Q+rR$ is non-expanding w.r.t. $\|.\|_1$, so estimating $xQ^{l-1}rR$ is enough. If $r < c/n$ then

there is some hope, as for $l = 1$ this vector has elements of size $8c/n^3$, so $\|xrR\|_1 \leq 8c/n^2$, which is acceptable if we want to add up an order of n^2 of these.

We want a similar inequality for other l, but for this we need the chain to be very lazy, which means $q_i \leq 1/4$ for all i. We can ensure this by replacing Q with $(3I + Q)/4$, but later we will have to deal with the problem to get back to the original Q.

To get a different view on this error term we may use the estimate

$$\|yR\|_1 = \sum_{i=0}^{n-1} |y_{i+1} - y_{i-1}| \leq 2 \sum_{i=0}^{n-1} |y_{i+1} - y_i| =: 2V(y). \tag{3.6}$$

In other words we are measuring how much the coordinates of a vector vary as we go around the cycle. The following lemma is what we need to bound this.

Lemma 3.25. *Suppose Q is as before and $q_i \leq 1/4$ for all i. Using the previously defined x and any $k \geq 0$,*

$$V(xQ^k) \leq \frac{4}{n}.$$

Proof. The proof is cleaner if we assume that the coordinates of xQ^k are different for each k. This is allowed by using Lemma 3.15.

If we go around the cycle we see that the coordinates of x consist of two monotone series, so there are only two local extrema. We call these *peaks*. The key thing is to show that this property remains as we multiply by Q. During the proof we will look at a few consecutive nodes and a single time step at once and find out how their ordering can change. We will do this until we cover all possibilities which can occur.

We mostly work by modifying weighted sums of some y_i by exchanging one y_i to a larger y_j. This way we maintain a sequence of inequalities to find out the new ordering.

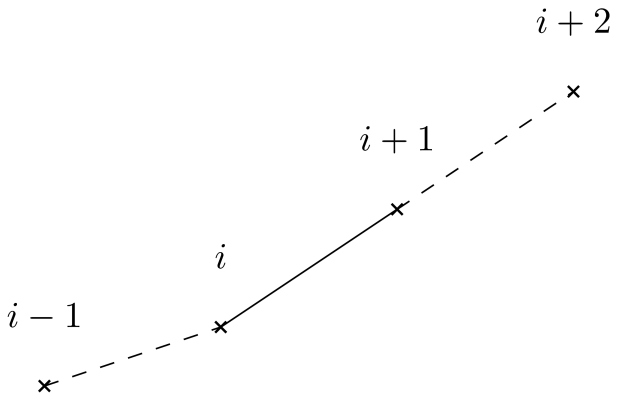


Figure 3.3: Two non-peak nodes

One possibility is if there are two non-peak nodes after each other. This means the 4 nodes form a monotone sequence, e.g. $y_{i-1} < y_i < y_{i+1} < y_{i+2}$. In this case, we have the following:

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1 - q_i - q_{i+1}) + y_{i+1}q_{i+1} \\ &\leq y_i(1 - q_{i+1}) + y_{i+1}q_{i+1}\end{aligned}$$

Here is the only other type of step we use. This time we change the weights instead of the values. We increase the weight of the larger y_{i+1} and decrease the weight of the smaller y_i. We use the assumption $q_{i+1} < 1/4$.

$$\begin{aligned}\ldots &< y_i q_{i+1} + y_{i+1}(1 - q_{i+1}) \\ &\leq y_i q_{i+1} + y_{i+1}(1 - q_{i+1} - q_{i+2}) + y_{i+2}q_{i+2} \\ &= (yQ)_{i+1}.\end{aligned}$$

Consequently the ordering of the values at nodes i and $i+1$ will remain the same.

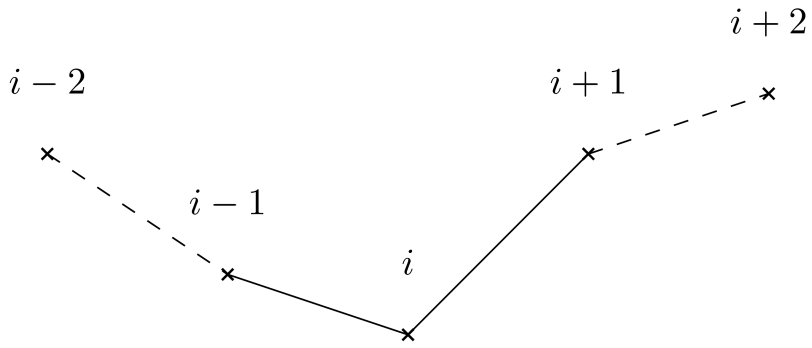


Figure 3.4: Single peak node

The only other setting that occurs initially if there is a peak node between two non-peak nodes. Without the loss of generality we may assume they are ordered as $y_{i-2} > y_{i-1} > y_i < y_{i+1} < y_{i+2}$, and $y_{i-1} < y_{i+1}$. A similar claim works as in the previous case:

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1 - q_{i+1} - q_i) + y_{i+1}q_{i+1} \\ &\leq y_i(1 - q_{i+1} - q_i) + y_{i+1}(q_{i+1} + q_i)\end{aligned}$$

Here we use $q_i < 1/4$ as in the previous case.

$$\begin{aligned}\ldots &< y_i q_{i+1} + y_{i+1}(1 - q_{i+1}) \\ &\leq y_i q_{i+1} + y_{i+1}(1 - q_{i+1} - q_{i+2}) + y_{i+2}q_{i+2} \\ &= (yQ)_{i+1}.\end{aligned}$$

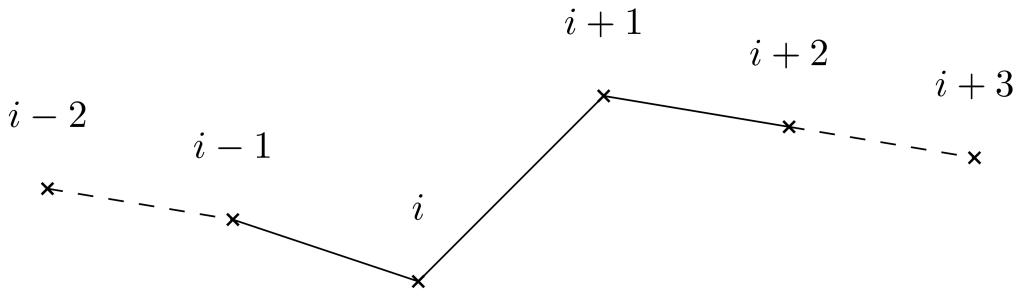


Figure 3.5: Two peak nodes

The ordering between node i and $i+1$ remains the same, but it might change between node $i-1$ and i. In either case, the number of peak nodes will not increase, although their position might change.

So far the only thing that could have happened that these peaks moved around. After a few steps we might find a setting different from the previous two, namely when two peak nodes appear next to each other. We may assume they are ordered as $y_{i-1} > y_i < y_{i+1} > y_{i+2}$. As there are only 2 peak nodes, the sequence $y_i, y_{i-1}, \ldots, y_{i+1}$ is increasing, therefore $y_{i-1} < y_{i+1}$ and $y_i < y_{i+2}$. Now we have

$$\begin{aligned}(yQ)_i &= y_{i-1}q_i + y_i(1 - q_{i+1} - q_i) + y_{i+1}q_{i+1} \\ &\leq y_i(1 - q_{i+1} - q_i) + y_{i+1}(q_{i+1} + q_i)\end{aligned}$$

We need $q_i < 1/4$ again, this time the condition is sharp.

$$\begin{aligned}&\leq y_i(q_{i+1} + q_{i+2}) + y_{i+1}(1 - q_{i+1} - q_{i+2}) \\ &\leq y_i q_{i+1} + y_{i+1}(1 - q_{i+1} - q_{i+2}) + y_{i+2}q_{i+2} \\ &= (yQ)_{i+1}.\end{aligned}$$

This shows that at least the ordering in the middle will remain as it was. If any of the other two changes, it has the same effect as in the previous case, namely a peak node will become non-peak, and maybe the non-peak node after will become a peak node. So the number of peak nodes does not increase, therefore no other setting can occur.

We covered all possibilities, and the bottom line is that there are only two peaks for all xQ^k. Clearly one is a maximum, the other is a minimum, and for such vectors

$$V(y) = 2\left(\max_i y_i - \min_i y_i\right).$$

This difference does not increase in our case due to the fact that Q is non-expanding w.r.t. $\|.\|_\infty$. In the end $V(xQ^k)$ is at most its initial value, $4/n$. □

Now we are ready to solve the lazy case.

Lemma 3.26. *Given a Markov chain on an n node cycle consider the transition matrix $P = Q + rR$. Suppose $q_i \le 1/4$ for all i and $0 \le r \le c/n$ for some fixed $c > 0$. Then there is a $c_2 > 0$ depending only on c such that for any $k \le c_2 n^2$ we have the bound*

$$\left\| xP^k - \frac{\mathbf{1}}{n} \right\|_1 > \frac{4}{12}.$$

Consequently,

$$t_{\text{mix}}(P, 1/8) > c_2 n^2.$$

Proof. Consider the error introduced by the rR terms in Equation (3.5), use Equation (3.6) and the previous lemma:

$$\begin{aligned}\left\| \sum_{l=1}^{k} xQ^{l-1}rR(Q+rR)^{k-l} \right\|_1 &\le \sum_{l=1}^{k} \left\| xQ^{l-1}rR(Q+rR)^{k-l} \right\|_1 \\ &\le r \sum_{l=1}^{k} \left\| xQ^{l-1}R \right\|_1 \le 2r \sum_{l=1}^{k} V(xQ^{l-1}) \le \frac{8rk}{n} \le \frac{8ck}{n^2}.\end{aligned}$$

If $k \le n^2/(100c)$, this error is at most $1/12$. We want to use Lemma 3.24 so fix $c_2 = \min(1/(100c), c_1)$, and choose $k \le c_2 n^2$. For such a k we have

$$\begin{aligned}\left\| x(Q+rR)^k - \frac{\mathbf{1}}{n} \right\|_1 &\ge \left\| xQ^k - \frac{\mathbf{1}}{n} \right\|_1 - \left\| \sum_{l=1}^{k} xQ^{l-1}rR(Q+rR)^{k-l} \right\|_1 \\ &\ge \frac{5}{12} - \frac{1}{12} = \frac{4}{12}.\end{aligned}$$

□

Relaxing laziness

We need to transfer our conclusion to non-lazy chains. We use a binomial expansion to go back to the original Q.

$$x\left(\frac{3I}{4} + \frac{Q+rR}{4}\right)^k = \sum_{l=0}^{k} x\binom{k}{l}\frac{3^{k-l}}{4^k}(Q+rR)^l.$$

This allows us to form an inequality for the l_1 distances:

$$\left\| x\left(\frac{3I}{4} + \frac{Q+rR}{4}\right)^k - \frac{\mathbf{1}}{n} \right\|_1 \le \sum_{l=0}^{k} \binom{k}{l}\frac{3^{k-l}}{4^k} \left\| x(Q+rR)^l - \frac{\mathbf{1}}{n} \right\|_1.$$

We will carry through the following idea. Start with $k = \lfloor c_2 n^2 \rfloor$ as in Lemma 3.26. The right side is a weighted average of some l_1 distances. If the mixing time was very small for

$Q + rR$, then these distances would be small for most of the terms. Then the average will be less than 4/12 which we previously proved for the left hand side. This contradiction will prove our claim, and complete the theorem.

Suppose $t_{\text{mix}}(Q + rR, 1/8) < k/8$. The l_1 distance is nonincreasing in l, so the terms are at most 1 and 1/4 before and after $t_{\text{mix}}(Q + rR, 1/8)$, respectively.

$$\sum_{l=0}^{k} \binom{k}{l} \frac{3^{k-l}}{4^k} \left\| x(Q+rR)^l - \frac{\mathbf{1}}{n} \right\|_1 \leq \sum_{l=0}^{[k/8]-1} \binom{k}{l} \frac{3^{k-l}}{4^k} 1 + \sum_{l=[k/8]}^{k} \binom{k}{l} \frac{3^{k-l}}{4^k} \frac{1}{4}$$
$$\leq \frac{1}{4} + \frac{3}{4} P\left(Binom\left(k, \frac{1}{4}\right) < \frac{k}{8} \right).$$

This probability can be easily bounded e.g. by Chebyshev's inequality:

$$P\left(Binom\left(k, \frac{1}{4}\right) < \frac{k}{8} \right) \leq P\left(\left| Binom\left(k, \frac{1}{4}\right) - \frac{k}{4} \right| > \frac{k}{8} \right) \leq \frac{\frac{3k}{16}}{\frac{k^2}{64}} = \frac{12}{k}.$$

We can find an n_0 such that $n > n_0$ implies $k > 108$. In this case the probability is less than 1/9, and the right hand side is strictly less than

$$\frac{1}{4} + \frac{3}{4 \cdot 9} = \frac{4}{12}.$$

This is the contradiction we were looking for.

In the end, let us choose $c' = \min(c_2, 1/n_0^2)$ so that the statement is also true for small n. This concludes the proof for the almost reversible case and thus for the whole theorem.

It turned out that there is no real speedup in the case of a cycle by introducing non-reversible chains. Let us see what happens if we slightly alter the connectivity graph.

3.2 Adding long range connections

During the struggle to decrease the mixing time of a Markov chain, we may consider slightly changing the connectivity graph hoping for some improvement.

We have already mentioned the concept of lifting at the beginning of the current chapter, which has the potential to decrease the mixing time to its square root. However, we have to note there is no general algorithm to construct the proper lifting of a chain, one has to work on each problem separately.

In this section we choose another route. We keep the nodes as they are and add some new connections among them and find out what we can achieve. As an initial attempt, we tried to come up with promising designs based on some heuristics. These trials failed miserably which is in line with the fact that the construction of expander graphs is extremely hard task. The only viable option is to use random graphs.

Consequently we form the following plan: We take a cycle with n nodes, add certain number of random edges according to some rule, then estimate the mixing time of reasonable chains over this connectivity graph. Naturally the result strongly depends on the number of added edges.

If this number is around cn for some $c > 0$ constant we suddenly arrive at a model of Small World Networks (SWN). Namely if we add an Erdős-Rényi random graph $G(n, c/n)$ to the cycle we get the model of Newman et al. [46]. This and other similar models were built to model large real networks which have small diameter but tend to show clustering, see Watts, Strogatz [53] for details. There is an intensive research activity on SWNs, the mixing time of random walks on them has also been investigated. The following result is due to Durrett [20], Addario-Berry and Lei [1] as it is roughly quoted here:

Theorem 3.27. *Consider an n node graph from the model of Newman et al. For the symmetric random walk we have*

$$c_1 \log^2 n < t_{\text{mix}} < c_2 \log^2 n$$

asymptotically almost surely (a.a.s.) for some global constants $c_1, c_2 > 0$.

This is a huge speed gain compared to the mixing time of n^2 for the cycle alone what we have seen in the previous section. Similar results for alternative models have been shown by Tahbaz-Salehi and Jadbabaie [2]. At this point our goal is to investigate the options and effects of adding a lower, $o(n)$ number of extra edges.

The target edge density of the added edges is $n^{-\alpha}$ for some parameter $\alpha \in (1, 2)$. Thus we expect $cn^{2-\alpha}$ extra edges. We are interested in the order of the mixing time as n increases but we do not care about constant factors. Accordingly, from now on c denotes a positive constant which may change from line to line (unless indicated otherwise). Let us introduce three slightly different models for choosing the random edges:

M1: We add a random matching on the almost equidistant $[n^{2-\alpha}]$ nodes $\{[in^{\alpha-1}],\ 0 \le i < n^{2-\alpha}\}$.

M2: From all possible long range edges we draw $[n^{2-\alpha}]$ randomly uniformly.

M3: For all possible long range edge we randomly decide to include it or not. Each edge is included independently with probability $n^{-\alpha}$.

The models differ only slightly, but the results are different and depend on the techniques we can apply for them. In each case our goal is to choose the transition probabilities to achieve the fastest mixing while keeping the stationary distribution uniform.

On the technical side, let us omit all the integer rounding operations. It is always clear what we mean, and those little errors do not affect the asymptotic properties we look for. Also, in the M2 and M3 models we allow edges of the cycle to be included as long range edges to simplify our discussion.

In the beginning we only consider the simple case of *homogeneous* chains when there are three common transition probabilities: $q_c + r$ for clockwise, $q_c - r$ for counter-clockwise transitions and $q_l/d(\alpha)$ for long range edges. These $q_c > r > 0,\ q_l > 0$ are some global constants. There might be a problem if a node has a lot of long range edges causing the sum of transition probabilities to go above 1. The following theorem ensures that this is not an issue.

Theorem 3.28. *There is a function $d(\alpha) : (1,2) \to \mathbb{N}$ such that there is no node with more than $d(\alpha)$ long range edges a.a.s. for M1, M2, M3 graphs.*
Consequently, assuming $2q_c + q_l \leq 1$ and using the current $d(\alpha)$, homogeneous chains will be feasible Markov chains a.a.s.

Proof. The graphs from M1 do not pose a strong restriction, every node has 0 or 1 long range edge.

Let us now check a single node of an M3 graph. Denote the number of its long range edges by X which follows a $Binom(n-1, n^{-\alpha})$ distribution. We want some upper bound on $P(X > d(\alpha))$. Let us use a Chernoff-type estimate:

$$P(X > d(\alpha)) = P\left(e^{tX} > e^{td(\alpha)}\right) \leq \frac{\mathbb{E}(e^{tX})}{e^{td(\alpha)}},$$

with arbitrary $t > 0$. The moment generating function of X is

$$\mathbb{E}(e^{tX}) = \left(1 + n^{-\alpha}(e^t - 1)\right)^{n-1}.$$

Let us choose $t = (\alpha - 1)\log n$ to get the following:

$$\mathbb{E}(e^{tX}) = \left(1 + n^{-\alpha}(n^{\alpha-1} - 1)\right)^{n-1} \to c$$

as $n \to \infty$. Let us use this t for the tail probability estimate and choose $d(\alpha) = 2/(\alpha - 1)$ to get

$$P(X > d(\alpha)) \leq \frac{c}{e^{(\alpha-1)\log n d(\alpha)}} \leq \frac{c}{n^2}.$$

The probability of any node having more than $d(\alpha)$ edges can be bounded above by the sum of the probabilities for every node. This is still at most c/n thus it does not happen a.a.s.

For M2 graphs the number of long rang edges of a single node follows a hypergeometric distribution which is less convenient than the binomial distribution before. We want to use our bounds for M3 graphs so we show a special way of generating an M2 graph.

We start with a modified M3 graph where the edge probability is $4n^{-\alpha}$. Let the number of long range edges be m. Depending on whether this is more or less than $[n^{2-\alpha}]$, we either discard some uniformly from the selected ones or add some uniformly from the unselected ones. This way we get the prescribed number of edges and the symmetry ensures everything happens uniformly.

We know there is a $d(\alpha)$ such that the initial M3 graph has at most $d(\alpha)$ long range edges at every node a.a.s. If we have to discard edges from this graph then this remains true. The only problem is in the case when we have to add edges, but the probability of this to happen is:

$$P\left(m < n^{2-\alpha}\right) < P\left(\left|m - 4n^{-\alpha}\frac{n(n-1)}{2}\right| > \frac{1}{2}n^{2-\alpha}\right) < c\frac{n^2 4n^{-\alpha}(1-4n^{-\alpha})}{n^{4-2\alpha}} < cn^{\alpha-2}.$$

Here we used Chebyshev's inequality. In the end, this probability also vanishes as $n \to \infty$, consequently the $d(\alpha)$ we got for modified M3 graphs also works for M2 graphs. □

Let us now try to get some estimates on the mixing time. The first result is a simple lower bound based on our original result for cycles.

Proposition 3.29. *For M1, let us assume the nodes with long range edges are equidistant from each other. Then for any homogeneous chain,*

$$t_{\text{mix}} \geq Cn^{2\alpha-2}.$$

Proof. Observe that we can "wind up" the chain around a cycle of $n^{\alpha-1}$ nodes so that long range edges become loop edges, see Figure 3.6.

For a certain starting distribution, mixing on the original chain implies mixing on the small cycle. Thus the lower bound from Theorem 3.13 for the small cycle also applies for the mixing time of the original chain. □

In the following two subsections we perform a systematic study on the three models. In Subsection 3.2.1 we determine the exact magnitude of the conductance for each case. In Subsection 3.2.2 we apply these results to bound the mixing times and also present simulation results where there is no sharp bound available.

3.2.1 Conductance estimates

For these Markov chains on random graphs we can estimate the mixing times using the conductance. First we show a technical tool to simplify the minimization for calculating the conductance.

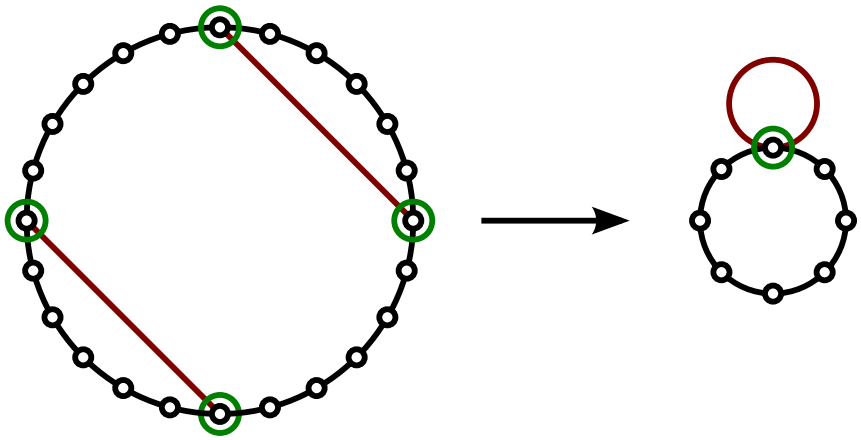


Figure 3.6: Reducing M1 graphs

Lemma 3.30. *Suppose that $S_1, S_2 \subset \mathcal{X}$, $S_1 \cap S_2 = \emptyset$ and there is no edge between them. Then*

$$\Phi(S_1 \cup S_2) \geq \min(\Phi(S_1), \Phi(S_2)).$$

Proof.

$$\Phi(S_1 \cup S_2) = \frac{Q(S_1 \cup S_2, (S_1 \cup S_2)^C)}{\pi(S_1 \cup S_2)\pi((S_1 \cup S_2)^C)} = \frac{Q(S_1, S_1^C) + Q(S_2, S_2^C)}{\pi(S_1) + \pi(S_2)} \cdot \frac{1}{\pi((S_1 \cup S_2)^C)}.$$

The first term is between $Q(S_1, S_1^C)/\pi(S_1)$ and $Q(S_2, S_2^C)/\pi(S_2)$. The second term is greater than both $1/\pi(S_1^C)$ and $1/\pi(S_2^C)$, thus the lemma follows. □

This means if S has at least 2 connected components, we may discard all but one to get closer to Φ:

Corollary 3.31. *The set S which attains the minimum in the definition of Φ must be connected.*

Let us now present three theorems to determine the exact order of the conductance for all three models.

Theorem 3.32. *For M1 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

$$c_1 d(\alpha)^{-1} n^{1-\alpha} < \Phi < c_2 n^{1-\alpha}.$$

Proof. The upper bound is simple: Let A be one of the $cn^{\alpha-1}$ long arcs without a long range edge. We can use this to bound the conductance:

$$\Phi = \min_{\emptyset \neq S \subsetneq V} \frac{Q(S, S^C)}{\pi(S)\pi(S^C)} \leq \frac{Q(A, A^C)}{\pi(A)\pi(A^C)} \leq c\frac{2n^{-1}}{n^{\alpha-2}} = cn^{1-\alpha}.$$

The lower bound is a bit more intricate. Using Corollary 3.31 we have to minimize over connected subsets to find Φ. Connected subgraphs are composed of a collection of arcs which are connected by long range edges. Let us define the following new chain with nodes $\tilde{\mathcal{X}}$ as shown in Figure 3.7. For every node of $\mathcal{X}$ with a long range edge there is one node in $\tilde{\mathcal{X}}$. Two nodes of $\tilde{\mathcal{X}}$ are connected if they are connected in $\mathcal{X}$ or if they follow each other on the cycle. In other words, we reduce all long empty arcs to single edges. Clearly the new chain has $cn^{2-\alpha}$ nodes.

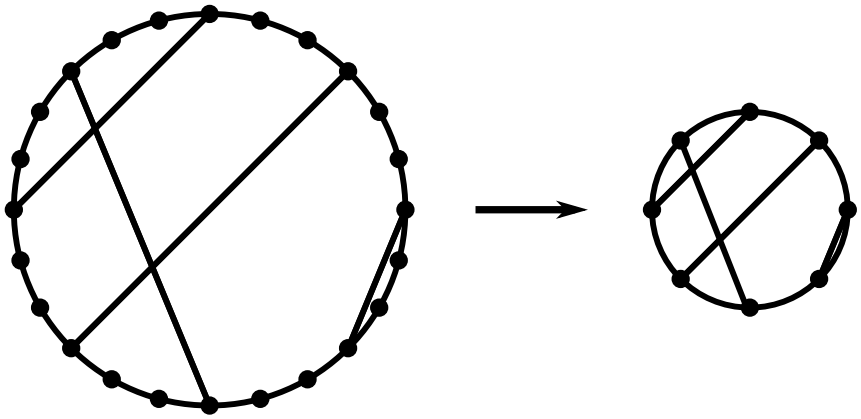


Figure 3.7: Reducing M1 graphs

We want to compare the conductance Φ of the original chain with the conductance $\tilde{\Phi}$ of the new one. For any connected $S \subset \mathcal{X}$ we may naturally define $\tilde{S} \subset \tilde{\mathcal{X}}$ by keeping only the nodes in $\tilde{\mathcal{X}}$. We want to bound $\Phi(S)$, but we do not need it for all $S \subset \mathcal{X}$. It makes no difference to work with S or S^C because $\Phi_S = \Phi_{S^C}$. If $|\tilde{S}| > |\tilde{\mathcal{X}}|/2$, let us swap S for S^C (and pick one of its connected components if needed). This way we can ensure $|\tilde{S}| \leq |\tilde{\mathcal{X}}|/2$. We need to estimate the expressions appearing in Φ_S. The transition probabilities are the same, the stationary measure changes, thus

$$Q(S, S^C) > c\frac{n^{2-\alpha}}{n} Q(\tilde{S}, \tilde{S}^C).$$

For any node in $\tilde{S}$ there are at most the two adjacent empty arcs present in S, consequently

$$\pi(S) < 2\tilde{\pi}(\tilde{S}).$$

We need a similar bound for $\pi(S^C)$. We made sure $\tilde{S}$ is "small" before so we have

$$\pi(S^C) < 1 \leq 2\tilde{\pi}(\tilde{S}^C).$$

Putting these together we arrive at

$$\begin{aligned} \Phi_S &\geq 4cn^{1-\alpha}\tilde{\Phi}_{\tilde{S}}, \\ \Phi &\geq 4cn^{1-\alpha}\tilde{\Phi}. \end{aligned} \tag{3.7}$$

The reduced graph is a cycle with $n^{2-\alpha}$ nodes with a random matching added, which is in fact the Bollobás-Chung small world model [3]. Our reduced chain is slightly different as the long range edges have transition probabilities $q_l/d(\alpha)$ instead of a global constant. The conductance of the Bollobás-Chung model is already known see e.g. Durrett [20] p. 163-164., where it is shown that it is bounded below by a positive constant. The conductance scales with the transition probabilities, hence for our reduced chain we have

$$\tilde{\Phi} \geq \frac{c}{d(\alpha)}. \tag{3.8}$$

Using this bound in the previous inequality 3.7 finishes the proof. □

Theorem 3.33. *For M2 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

$$c_1 d(\alpha)^{-1} \frac{n^{1-\alpha}}{\log n} < \Phi < c_2 \frac{n^{1-\alpha}}{\log n}.$$

Proof. For an upper bound, we search again for a long arc A without a long range edge. In this context, adding $n^{2-\alpha}$ random edges means we cut the cycle into arcs at $k = 2n^{2-\alpha}$ random points. It is rather clear that asymptotically this is equivalent to splitting the unit interval by k i.i.d. uniform variables. The largest gap is known to be asymptotically $\log k/k$ long, see Devroye [15] or Slud [51]. This tells us the number of nodes in the longest empty arc A is

$$n\frac{\log k}{k} = n\frac{(2-\alpha)\log n + \log 2}{2n^{2-\alpha}} = cn^{\alpha-1}\log n + O(n^{\alpha-1}).$$

Consequently we can use a similar estimate like before:

$$\Phi = \min_{\emptyset \neq S \subsetneq V} \frac{Q(S, S^C)}{\pi(S)\pi(S^C)} \leq \frac{Q(A, A^C)}{\pi(A)\pi(A^C)} \leq c\frac{2n^{-1}}{n^{\alpha-2}\log n} = c\frac{n^{1-\alpha}}{\log n}.$$

For the proof of the lower bound we want to follow the same idea as for Theorem 3.32. Let us check what we have to change. First of all, there might be nodes which have multiple long range edges. For the graph on $\tilde{\mathcal{X}}$ we want the long range edges to form a random matching. Thus we include multiple copies of such a node, each retaining one long range edge, see Figure 3.8.

The other difference is that empty arcs have different lengths now. Still, we can use $cn^{\alpha-1}\log n$ as an upper bound for these lengths as we have shown during the first part of the proof.

We use similar inequalities to those in the proof of Theorem 3.32:

$$\begin{aligned} Q(S, S^C) &> 2cn^{1-\alpha} Q(\tilde{S}, \tilde{S}^C), \\ \pi(S) &< 2\log n \tilde{\pi}(\tilde{S}), \\ \pi(S^C) &< 2\tilde{\pi}(\tilde{S}^C). \end{aligned}$$

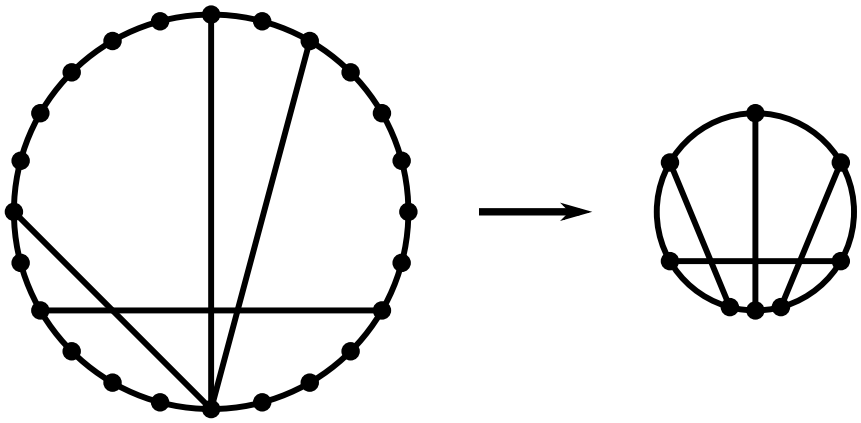


Figure 3.8: Reducing M2 graphs

The second inequality is slightly different because the omitted arcs might be longer than in the previous case. For the third inequality we use the same trick as before to ensure $\tilde{S}^C$ is large. This time we get the inequalities

$$\Phi_S \geq 8c\frac{n^{1-\alpha}}{\log n}\tilde{\Phi}_{\tilde{S}},$$
$$\Phi \geq 8c\frac{n^{1-\alpha}}{\log n}\tilde{\Phi}.$$

We use the the bound 3.8 on $\tilde{\Phi}$ again to conclude the proof. □

Theorem 3.34. *For M3 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

$$c_1 d(\alpha)^{-1}\frac{n^{1-\alpha}}{\log n} < \Phi < c_2\frac{n^{1-\alpha}}{\log n}.$$

Proof. Let us start with the lower bound. For any $S \subset \mathcal{X}$, $|S| \leq n/2$ we have

$$\frac{Q(S, S^C)}{\pi(S)} \geq \frac{c|\partial S|/n}{d(\alpha)|S|/n} = cd(\alpha)^{-1}\frac{|\partial S|}{|S|},$$

where ∂S is the set of edges between S and S^C. We have to ensure this is large enough for all possible subsets S. Let us fix $s = |S| \leq n/2$ and the number of disjoint intervals l it consists of. We look at only these subsets at once.

We can estimate the number k of possible subsets in the following way:

$$k \leq \binom{n}{l}\binom{s}{l}.$$

The first binomial coefficient counts how we can choose the starting points of the intervals, the second distributes the total length of s among them. To continue we use the following inequality:

$$\binom{m}{t} \leq \left(\frac{me}{t}\right)^t.$$

For k this gives us

$$k \leq \left(\frac{ne}{l}\right)^l \left(\frac{se}{l}\right)^l \leq \exp\left(l\left(\log\frac{n}{l} + \log\frac{s}{l} + 2\right)\right) < \exp(4l\log n).$$

The outgoing edges from S are partially from the interval boundaries and partially from the random long-range edges. We have $2l$ edges at the interval boundaries and the number of long-range edges L follow a $Binom(s(n-s), n^{-\alpha})$ distribution. A subset violates the conductance we proposed if

$$\underline{c}\frac{n^{1-\alpha}}{\log n} > \frac{|\partial S|}{|S|} = \frac{L+2l}{s},$$

for some $\underline{c}$ constant. We introduce this new notation because the choice of $\underline{c}$ is important as we will see. The probability of this happening can be written in the following way:

$$p = P\left(Binom(s(n-s), n^{-\alpha}) < s\underline{c}\frac{n^{1-\alpha}}{\log n} - 2l\right).$$

Let us introduce the temporary notation $r = s\underline{c}n^{1-\alpha}/\log n - 2l$. If $r \leq 0$, then this probability is 0, we can't expect anything better. If not, then we have the implied inequality

$$s\underline{c}n^{1-\alpha} > 2l\log n. \tag{3.9}$$

In this case we have to find an upper bound on p.

$$\begin{aligned} p &\leq rP(Binom(sn/2, n^{-\alpha}) = r) = \\ &= r\binom{sn/2}{r} n^{-\alpha r}\left(1-n^{-\alpha}\right)^{sn/2-r} \leq r\left(\frac{sne}{2r}\right)^r n^{-\alpha r}\left(1-n^{-\alpha}\right)^{sn/2-r} \leq \\ &\leq \exp\left(\log r + r\log\frac{sne}{2r} - \alpha r\log n - \frac{sn}{2n^\alpha} + \frac{r}{n^\alpha}\right). \end{aligned}$$

We want to find out the asymptotic behavior of this expression. We have $\alpha r\log n \gg \log r$ and $\alpha r\log n \gg r/n^\alpha$ so the corresponding three terms add up to some negative number. The second term can be bounded the following way:

$$r\log\frac{sne}{2r} < r\log(2n^2) < \frac{s\underline{c}n^{1-\alpha}}{\log n}3\log n = (3\underline{c})sn^{1-\alpha}.$$

This is dominated by the fourth term $-sn^{1-\alpha}/2$ if $\underline{c}$ is small enough.

Now let us look at all subsets S of s nodes and l intervals. The probability that there is one which violates the conductance is at most kp. Using Equation 3.9 we have an upper bound for k,

$$\log k < 4l\log n < (2\underline{c})sn^{1-\alpha}.$$

Let us join our previous estimates. For n large enough we have

$$\log(kp) < (2\underline{c})sn^{1-\alpha} + (3\underline{c})sn^{1-\alpha} - \frac{1}{2}sn^{1-\alpha} = \left(5\underline{c} - \frac{1}{2}\right)sn^{1-\alpha}.$$

For $\underline{c} \leq 1/20$ we get a coefficient at most $-1/4$. From 3.9 again,

$$\frac{1}{4} s n^{1-\alpha} > \frac{l}{2\underline{c}} \log n.$$

Here we need $\underline{c} \leq 1/6$ to get at least $3 \log n$. After all, we end up with

$$kp < \frac{1}{n^3}.$$

The only thing left is to sum over all possible s and l values. This introduces an extra n^2 term, but the sum remains asymptotically 0. In the end we see the lower bound on the conductance is false only with asymptotically vanishing probability.

Now let us turn our attention to the upper bound. If we find an arc A that is at least $\overline{c} n^{\alpha-1} \log n$ long and has no long range edges then we can use the same estimate as before:

$$\Phi = \min_{\emptyset \neq S \subsetneq V} \frac{Q(S, S^C)}{\pi(S)\pi(S^C)} \leq \frac{Q(A, A^C)}{\pi(A)\pi(A^C)} \leq \overline{c} \frac{2n^{-1}}{n^{\alpha-2} \log n} = \overline{c} \frac{n^{1-\alpha}}{\log n}.$$

Again, the choice of $\overline{c}$ plays an important role, this is the reason for the distinct notation. To find such an arc, let us split the cycle into arcs of length $b = \overline{c} n^{\alpha-1} \log n$. We define a graph on these arcs, we connect two of them if there is any long range edge between them. This is in fact an Erdős-Rényi random graph. Our goal translates to finding an isolated node in it.

It is known for Erdős-Rényi graphs [22] with k nodes and edge probability p that there are some isolated nodes a.a.s. if $p < (1 - \varepsilon) \log k / k$. In our case the number of nodes is

$$k = \frac{n}{b} = \frac{n^{2-\alpha}}{\overline{c} \log n}.$$

We can bound the edge probability by adding up the probabilities of all the possible long range edges between two arcs:

$$p \leq b^2 n^{-\alpha} = \overline{c}^2 n^{\alpha-2} \log^2 n.$$

We have to compare this quantity with the following:

$$\frac{\log k}{k} = \overline{c} n^{\alpha-2} \log n ((2 - \alpha) \log n - \log \overline{c} - \log \log n).$$

The major term is the first one, which is fortunately of the same order as p. In order to have an isolated node we simply need

$$\overline{c}^2 < \overline{c}(2 - \alpha),$$

$$\overline{c} < 2 - \alpha.$$

There was no other restriction on $\overline{c}$ apart from being positive so we can choose it to satisfy this last inequality. This concludes the proof. □

3.2.2 Mixing time bounds

Using Theorem 3.10 on the previous conductance bounds we arrive at the following corollary about the mixing times:

Corollary 3.35. *For M1 the mixing time of the reversible homogeneous chain satisfies the following inequality a.a.s.:*

$$c_1 n^{2\alpha-2} < t_{\text{mix}} < c_2 d(\alpha)^2 n^{2\alpha-2} \log n.$$

Similarly, for the M2 and M3 reversible homogeneous chains we have

$$c_1 n^{2\alpha-2} \log^2 n < t_{\text{mix}} < c_2 d(\alpha)^2 n^{2\alpha-2} \log^3 n.$$

For non-reversible chains, the asymptotic bounds become

$$c_1 n^{\alpha-1} < t_{\text{mix}} < c_2 d(\alpha)^2 n^{2\alpha-2} \log n,$$
$$c_1 n^{\alpha-1} \log n < t_{\text{mix}} < c_2 d(\alpha)^2 n^{2\alpha-2} \log^3 n.$$

for homogeneous M1 and homogeneous M2 or M3 chains, respectively.

Proof. The general bounds follow simply by combining Theorem 3.10 with the conductance bounds Theorem 3.32, Theorem 3.33 and Theorem 3.34.

We only need to show the sharper lower bounds for reversible chains. Observe that during the proofs of the previous conductance bounds we always used the fact that there is a long arc without a long range edge a.a.s. To be precise, for M3 chains we only found an arc $cn^{\alpha-1} \log n$ long that has no long range edge going out of it. Still, the probability of having no long range edge within the arc is

$$(1-n^{-\alpha})^{(cn^{\alpha-1}\log n)^2} = (1-n^{-\alpha})^{n^\alpha c \frac{\log^2 n}{n^{2-\alpha}}} > e^{-2c\frac{\log^2 n}{n^{2-\alpha}}}$$

This is 1 in the limit, consequently a.a.s. the arc we have chosen does not have any long range edge at all. Let us now focus only on this arc. Without going into details, mixing only within this part needs at least $cn^{2\alpha-2}$ steps for reversible M1 chains or $cn^{2\alpha-2}\log^2 n$ steps for reversible M2 and M3 chains. This provides the missing bound. □

All the previous results are about homogeneous chains. However, we are more interested in the best mixing time we can achieve on the same graphs using different chains. The following Corollary provides the answer.

Corollary 3.36. *The bounds of Corollary 3.35 also hold for the fastest M1, M2, M3 chains.*

Proof. The mixing time of the fastest chain can not be higher then the mixing time of any specific chain on the same graph. This ensures the upper bound.

To get the lower bound observe that the long empty arc is still there. It poses the same upper bound on the conductance (the constant might be different) also giving the general lower bound and provides the same lower bound for mixing time in the reversible case. □

For reversible chains the gap is reasonably tight. For non-reversible chains it is still unclear where the truth lies in between these bounds.

Still, we hope there is a considerable gain for non-reversible chains as shown in Figure 3.9. This is a plot of mixing times of homogeneous reversible and non-reversible chains on several graphs coming from M2, for $\alpha = 1.5$.

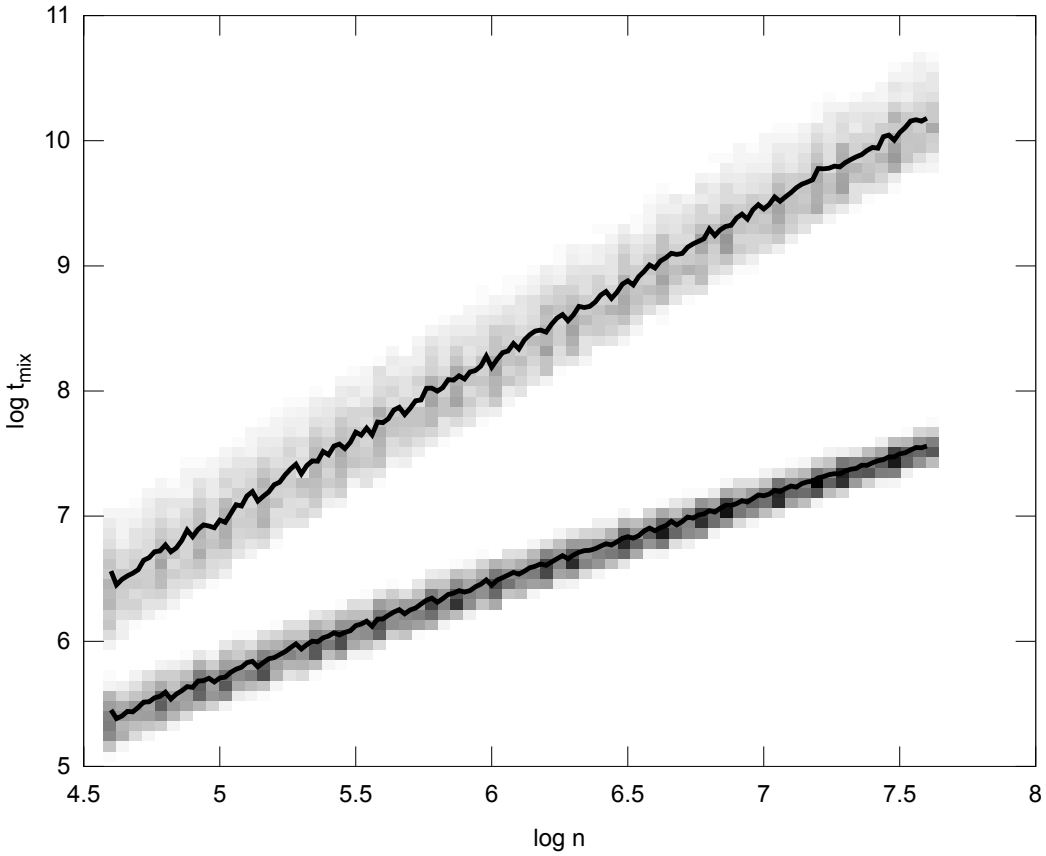


Figure 3.9: Log-log plot for mixing times of homogeneous M2 chains

The sizes of the graphs change between 100 and 2000 exponentially. We generated 222 random graphs for each size and calculated the mixing times for a reversible and a non-reversible chain on them. As we are not interested in the extremes, we discarded the lowest and highest 5% of mixing times for each size, leaving us with a total number of 30000 graphs. The results for these graphs are plotted as a histogram on a log-log

scale. The upper cluster contains the data for the reversible chains, the lower for the non-reversible ones. The two noisy diagonal lines are simply the averages.

It is clearly visible that non-reversible chains offer a significant speedup over reversible ones in this setting. We hope to quantify this gain in the future but at this point, we do not aim for a bold guess as $\log n$ and n^δ factors can be easily mistaken for each other on this scale.

On the other hand we may guess the mixing time for reversible chains is $n \log^\delta n$ based on Corollary 3.35. By looking for the best fit on the data we arrive at the estimate $\delta = 2.02$, which suggests that the lower bound is the one that is sharp.

3.3 Open questions

Let us begin the list with the last topic, where our knowledge ends. In the light of the measurements shown in Figure 3.9 we are eager to find theoretical estimates about the speedup that is clearly visible.

Another natural question is to ask for a lower bound of the mixing time for other connectivity graphs. It is easy to answer this problem in the extreme cases. When the graph is a tree, all chains will be reversible (assuming uniform stationary distribution), and known theory applies. For a complete graph mixing in a single step is possible even without violating reversibility.

We might also try to extend our results for time-inhomogeneous Markov chains. In the case of the cycle, there is one thing that is clear. If we do not require all the transition matrices to be doubly stochastic, we can easily construct a chain that mixes in the order of n steps.

Another direction to look forward is the problem of graph design: here one may want to find the fastest mixing chain satisfying specific constraints such as an upper bound on the edges of the connectivity graph, or a locality constraint. Note that general methods, such as the Metropolis-Hastings algorithm ([28]) do not give the fastest mixing chain for specific problems. Namely, if we want to sample from the uniform distribution, then it necessarily produces a reversible chain, and it does not exploit the possibility of using a non-reversible one. This alone shows that the problem of design deserves a closer look.

Bibliography

[1] L. Addario-Berry and T. Lei, *The mixing time of the Newman–Watts small world*, 18 Jan 2012, arXiv:1201.3795v1.

[2] A. J. Alireza Tahbaz-Salehi, *Small world phenomenon, rapidly mixing Markov chains, and average consensus algorithms*, in Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 276–281.

[3] B. Bollobás and F. Chung, *The diameter of a cycle plus a random matching*, SIAM J. Discrete Math., 1 (1988), pp. 328–333.

[4] C. Borgs, *Statistical physics expansion methods in combinatorics and computer science*, CBMS lecture notes (in preparation), (2004).

[5] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Fastest mixing Markov chain on graphs with symmetries*, SIAM J. Optim., 20 (2009), pp. 792–819.

[6] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM Rev., 46 (2004), pp. 667–689 (electronic).

[7] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.

[8] L. Brouwer, *Über abbildung von mannigfaltigkeiten*, Mathematische Annalen, 71 (1911), pp. 97–115.

[9] F. Chen, L. Lovász, and I. Pak, *Lifting Markov chains to speed up mixing*, in Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999), ACM, New York, 1999, pp. 275–281 (electronic).

[10] J. Crow and M. Kimura, *Efficiency of truncation selection*, P. Nat. Acad. Sci., 76 (1979), pp. 396–399.

[11] R. Curnow and C. Smith, *Multifactorial models for familial diseases in man*, J. Roy. Stat. Soc. A, (1975), pp. 131–169.

[12] E. Czeizel and G. Tusnády, *Aetiological studies of isolated common congenital abnormalities in Hungary*, Akadémiai Kiadó Budapest, 1984.

[13] K. Dawson, *The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations*, Theor. Popul. Biol., 55 (1999), pp. 1–22.

[14] A. Dénes, *Theoretical and computer-aided stability examination of population dynamical systems (in Hungarian)*, PhD thesis, University of Szeged, 2011.

[15] L. Devroye, *Laws of the iterated logarithm for order statistics of uniform spacings*, Ann. Prob., (1981), pp. 860–867.

[16] P. Diaconis, S. Holmes, and R. M. Neal, *Analysis of a nonreversible Markov chain sampler*, Ann. Appl. Probab., 10 (2000), pp. 726–752.

[17] P. Diaconis and L. Saloff-Coste, *What do we know about the Metropolis algorithm?*, J. Comupt. Syst. Sci., 57 (1998), pp. 20–36.

[18] W. Doeblin, *Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples*, B. Math. Soc. Sci. Math., 39 (1937), pp. 57–115.

[19] J. Doob, *Stochastic processes*, Wiley, 1953.

[20] R. Durrett, *Random Graph Dynamics*, Cambridge University Press, 2006.

[21] J. Edwards, *The genetic basis of common disease*, Am. J. Med., 34 (1963), pp. 627–638.

[22] P. Erdős and A. Rényi, *On random graphs*, Publ. Math. Debrecen., 6 (1959), pp. 290–297.

[23] J. A. Fill, *Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process*, Ann. Appl. Probab., 1 (1991), pp. 62–87.

[24] K. K. Gade and M. L. Overton, *Optimizing the asymptotic convergence rate of the Diaconis-Holmes-Neal sampler*, Adv. in Appl. Math., 38 (2007), pp. 382–403.

[25] B. Gerencsér, *Markov chain mixing time on cycles*, Stoch. Proc. Appl., 121 (2011), pp. 2553 – 2570.

[26] ——, *Effects of long range connections on Markov chain mixing times*, in Proceedings of the MTNS 2012, Melbourne, Australia, 2012.

[27] B. Gerencsér, B. Ráth, and G. Tusnády, *Multifactorial inheritance with selection*, Acta Math. Hung., (2012). Online First.

[28] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57 (1970), pp. 97–109, http://biomet.oxfordjournals.org/cgi/reprint/57/1/97.pdf.

[29] L. Hatvani, F. Toókos, and G. Tusnády, *A mutation-selection-recombination model in population genetics*, Dyn. Syst. Appl., 18 (2009), pp. 335–362.

[30] M. Jerrum, *Mathematical foundations of the Markov chain Monte Carlo method*, in Probabilistic methods for algorithmic discrete mathematics, vol. 16 of Algorithms and Combinatorics, Springer, Berlin, 1998, pp. 116–165.

[31] M. Jerrum and A. Sinclair, *Conductance and the rapid mixing property for Markov chains: the approximation of permanent resolved*, in Proceedings of the twentieth annual ACM symposium on Theory of computing, ACM, 1988, pp. 235–244.

[32] R. Kannan, L. Lovász, and R. Montenegro, *Blocking conductance and mixing in random walks*, Combin. Probab. Comput., 15 (2006), pp. 541–570.

[33] S. Karlin, *Models of multifactorial inheritance: I, multivariate formulations and basic convergence results*, Theor. Popul. Biol., 15 (1979), pp. 308–355.

[34] F. Kelly, *Reversibility and stochastic networks*, vol. 40, Wiley New York, 1979.

[35] M. Kimura and T. Maruyama, *The mutational load with epistatic gene interactions in fitness*, Genetics, 54 (1966), pp. 1337–1351.

[36] J. Komlós, A. Odlyzko, L. Ozarow, and L. Shepp, *On the properties of a tree-structured server process*, Ann. Appl. Probab., 1 (1991), pp. 118–125.

[37] A. Kondrashov, *Selection against harmful mutations in large sexual and asexual populations*, Genetical Research, 40 (1982), pp. 325–332.

[38] D. Levin, Y. Peres, and E. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2009.

[39] L. Lovász and R. Kannan, *Faster mixing via average conductance*, in Proceedings of the thirty-first annual ACM symposium on Theory of computing, ACM, 1999, pp. 282–287.

[40] L. Lovász and M. Simonovits, *The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume*, in 31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990), IEEE Comput. Soc. Press, Los Alamitos, CA, 1990, pp. 346–354.

[41] L. Lovász and P. Winkler, *Mixing times*, Microsurveys in Discrete Probability, 41 (1998), pp. 85–134.

[42] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1092.

[43] S. Meyn and R. Tweedie, *Markov chains and stochastic stability*, Springer-Verlag, London, 1993.

[44] R. Montenegro and P. Tetali, *Mathematical aspects of mixing times in Markov chains*, Theor. Comput. Sci., 1 (2006), pp. 237–354.

[45] A. Mukherjea, M. Rao, and S. Suen, *A note on moment generating functions*, Stat. Probabil. Lett., 76 (2006), pp. 1185 – 1189.

[46] M. Newman, C. Moore, and D. Watts, *Mean-field solution of the small-world network model*, Phys. Rev. Lett., 84 (2000), pp. 3201–3204.

[47] A. Olshevsky and J. N. Tsitsiklis, *Convergence speed in distributed consensus and averaging*, SIAM J. Control Optim., 48 (2009), pp. 33–55.

[48] A. Olshevsky and J. N. Tsitsiklis, *Convergence speed in distributed consensus and averaging*, SIAM Review, 53 (2011), pp. 747–772.

[49] K. Sankaranarayanan, N. Yasuda, R. Chakraborty, G. Tusnady, and A. Czeizel, *Ionizing radiation and genetic risks. V. Multifactorial diseases: A review of epidemiological and genetic aspects of congenital abnormalities in man and of models on maintenance of quantitative traits in populations*, Mutat. Res./Rev. Genet., 317 (1994), pp. 1–23.

[50] A. Sinclair and M. Jerrum, *Approximate counting, uniform generation and rapidly mixing Markov chains*, Inform. and Comput., 82 (1989), pp. 93–133.

[51] E. Slud, *Entropy and maximal spacings for random partitions*, Probab. Theory Related Fields, 41 (1978), pp. 341–352.

[52] G. Tusnády, *Mutation and selection (in Hungarian)*, Magyar Tudomány, 7 (1997), pp. 792–805.

[53] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440–442.

Abstract

In this thesis we investigate the long term behavior of random processes. The dynamics we consider are all unusual in some sense, the standard techniques need to be accompanied by unique approaches.

In the first part we work on problems with biological motivation. Our goal is to model the inheritance of congenital abnormalities. The main difficulty comes from the underlying process - the genetic information - which evolves in a way that two copies generate a single new one. Studying the stability of such processes is often hopeless. However, in this case it was possible to find the so called *Poisson model* where we could show that the process approaches a well defined stationary distribution independent of the starting population.

We could determine the risk of the appearance of the malformation in different scenarios. For example, we know the conditional probability of an uncle of a malformed child developing the disease. The significance of this comes from the fact that these are the quantities we can acquire from Hungarian population data. This way we get the opportunity to check how realistic our model is. It turns out that the Poisson model fits to most of the Hungarian datasets pretty well, partially because the Poisson model is richer than the classical Gaussian model.

In the second part of the thesis we work with Markov chain mixing times. The fundamental goal is to modify the transition probabilities of a Markov chain in order to decrease the mixing time as much as possible. This has to happen while retaining the set of allowed transitions and the stationary distribution. The problem is manageable whenever the Markov chain is restricted to be reversible, in other words when the transition probability is the same in the two directions for any allowed transition. In this case there are known algorithms to find the fastest chain and also algebraic tools work well to deal with the problem.

The situation is much more complicated for non-reversible chains. Still, when the allowed transition form a cycle of n nodes we could show that the mixing time is still at least of the order of n^2, as in the reversible case. To achieve some speedup, we allow a few more transitions to occur. It is known that with the inclusion of cn new transitions, the mixing time drops to $\log^2 n$, but we want a lower number of new transitions. In the reversible case we could determine the magnitude of the mixing time up to a $\log n$ factor. In the non-reversible case the resulting bounds are not sharp, but with the help of computer simulations we think it is possible to achieve a notable speedup compared to reversible chains.

Összefoglalás

Az értekezésben különböző véletlen folyamatok hosszútávú viselkedését vizsgáljuk. Az előkerülő dinamikák mind valamilyen szempontból szokatlanok, ezért a standard módszereken túl egyedi megközelítésekre is szükség van.

Az első részben biológiai indíttatású kérdésekkel foglalkozunk. Célunk születési rendellenességek öröklődésének modellezése. A fő nehézséget az okozza, hogy a háttérben húzódó folyamat - a genetikai információ - úgy terjed, hogy két példányból keletkezik az egy újabb. Az ilyen folyamatok stabilitásának vizsgálata sokszor reménytelen, szerencsére azonban sikerül egy olyan modellt, az ún. *Poisson modellt* találni, ahol be lehet látni, hogy a kezdeti populációtól függetlenül a folyamat egy jól meghatározott stacionárius eloszláshoz tart.

Sikerült a modellben a betegség különböző előfordulásainak kockázatát meghatározni, például megmondani egy beteg gyerek esetén a nagybácsi betegségének feltételes valószínűségét. Ennek azért van jelentősége, mert magyar orvosi adatokból ezeknek a mennyiségeknek kapjuk meg a mért értékét. Ezáltal lehetőség nyílik arra, hogy a modellünk realitását teszteljük. A legtöbb adatsorra kifejezetten jól illeszkedik a Poisson modell, ez annak is köszönhető, hogy gazdagabb, mint a klasszikusan használt Gaussi modell.

Az értekezés második részében Markov-lánc keverési időkkel dolgozunk. Szeretnénk az átmenetvalószínűségeket úgy meghatározni, hogy a keverési idő a lehető legkisebb legyen. Mindezt úgy, hogy a lehetséges átmenetek halmaz a rögzített és a stacionárius eloszlás nem változhat. Egyszerűbb a helyzet, ha a Markov-lánc reverzibilis, ami esetünkben úgy is fogalmazható, hogy minden átmenetnél ugyanannyi a két irányba az átmenetvalószínűség. Ekkor léteznek algoritmikus módszerek a leggyorsabb lánc megtalálására, és hatékony algebrai eszközök is rendelkezésre állnak.

Lényegesen nehezebb a helyzet, ha a reverzibilitást nem tesszük fel. Mégis, sikerül belátni, hogy amennyiben a lehetséges átmenetek egy n csúcsú kört alkotnak, a legjobb keverési idő továbbra is n^2 nagyságrendű, ahogy a reverzibilis esetben is. A gyorsítás elérése érdekében néhány plusz lehetséges átmenetet adunk a körhöz. Ismert, hogy cn átmenet hozzáadásával a keverési idő $\log^2 n$-re csökken, azonban mi ennél kevesebb új átmenetet szeretnénk megengedni. Reverzibilis esetben egy $\log n$ faktor erejéig pontosan meg tudjuk határozni a keverési idő nagyságrendjét. Nem-reverzibilis esetben a korlátaink nem élesek, azonban számítógépes szimuláció eredményeképpen látszik, hogy lényeges gyorsítás érhető el a reverzibilis esethez képest.