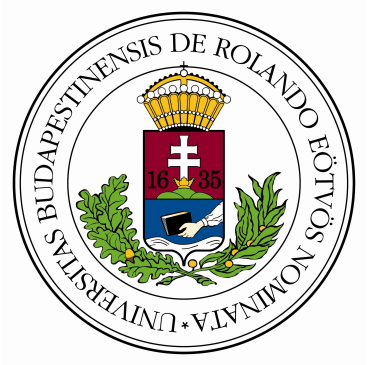
Eötvös Loránd University
Institute of Mathematics
Department of Probability Theory and Statistics



Summary of the 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

1. Introduction

In this thesis we investigate multiple questions about the long term behaviour of certain Markovian processes. Although Markovian processes can be conisdered as the simplest dependence structure one can get (apart from i.i.d. processes), even the simplest questions become surprisingly difficult when we try to deal with slightly unusual processes.

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 [6]. 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 2.1. It turned out that a similar model has been previously investigated by Dawson [2].

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 [4] and [5]. This part of my research has been done with the help and supervision of John Tsitsiklis.

2. Convergence of Bi-Markov processes

We consider a population with sexual reproduction, selection, synchronous generations on a short time frame in the evolutionary sense. We are interested in the inheritance of a certain congenital abnormality which is related to some gene mutations. 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 call this setup the *Poisson model*.

We want to deal 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 [2] 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 the proof goes to Balázs Ráth.

Theorem 2.1. *If $\rho_f, \rho_m < 1$ then for any pair of initial distributions of mutant genes, the population 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.

Let us now assume the population is 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 do this by recursively simplifying the family tree. As the simplest example, the conditional probability of the sibling of a malformed child being affected is

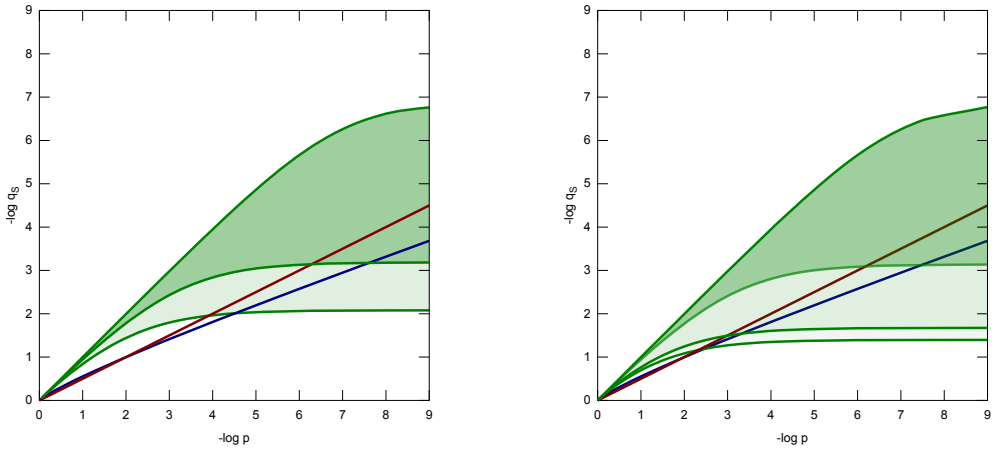
$$q_S = 1 - \frac{\exp\left(\left(\mu_y + \frac{\lambda_f+\lambda_m}{2}\right)(\Delta_y - 1)\right) - \exp\left(\mu_x(\Delta_x - 1) + \mu_y(\Delta_y - 1) + \frac{\lambda_m+\lambda_f}{4}((\Delta_x+1)(\Delta_y+1) - 4)\right)}{1 - \exp\left(\left(\mu_x + \frac{\lambda_f+\lambda_m}{2}\right)(\Delta_x - 1)\right)},$$

where x, y is the gender of the child and his/her sibling, respectively (m of f).

In order to validate our model we have to check how well does it follow biological principles and how does it fit the population. We also compare with the classical Gaussian model used by Czeizel and Tusnády in [1]. The initial requirement for a model of inheritance is to have high conditional probabilities for first order relatives, in other words $q_S \gg p$. Another guideline we use is a fundamental approximation on multifactorial disorders given by the Edwards formula [3] which states that $q_S \approx \sqrt{p}$.

We don't want to go into theoretical details, let us just present Figure 1 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. 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). Although the model does not satisfy the formula in general, we may choose the parameters to do so. Also, it is clearly visible that for rare disorders the Poisson model can achieve substantially higher conditional probabilites for siblings.



(A) Complete selection (B) Partial selection

FIGURE 1. Model probabilities and the Edwards formula

Further checks have been made for sex dependent parameters and different relatives and they are in line with the current findings.

Another way to qualify the power of the Poisson model is to check its goodness-of-fit on the Hungarian data. The population data were gathered and published by Czeizel and Tusnády [1].

In Table 1 we present the goodness-of-fit values for the fitted 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.

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 1. Goodness-of-fit of the Poisson model to Hungarian data

Finally let us present the parameter values for the best fit in Table 2.

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. Parameters of the Poisson model for Hungarian data

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, 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 based selection (like the Gaussian model) 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. 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.

3. Markov chain mixing time estimates

We know that an irreducible aperiodic Markov chain on a finite state space approaches its stationary distribution. To formulate this 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

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

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 the *mixing time*:

Definition 3.2. For a Markovchain 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

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

We might omit some of the arguments when they are obvious or unimportant. Note that this might be infinite.

A remarkable property of certain Markov chains is *reversibility* which often makes these approximations easier, see e.g. Kelly [7].

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:

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

Our primary goal is to determine the best possible mixing time we can achieve by changing the transition probabilites but not the set of allowed transitions. We also require the stationary distribution to always remain uniform. Our first result is for Markov chains where the graph of allowed transitions is a cycle.

Theorem 3.4. *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*

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

It is well known that the best mixing time of a reversible Markov-chain on a cycle with n nodes is of the order of n^2. Consequently this theorem tells us that allowing non-reversible chains does not help in this case.

In the hope of achieving some speedup we increase the number of allowed transitions by adding some random edges to the cycle. The target edge density of the added edges is $cn^{-\alpha}$ for some parameter $\alpha \in (1, 2)$. Thus we expect $cn^{2-\alpha}$ extra edges. Let us introduce three slightly different models for choosing them.

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

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 $n^{-\alpha}$.

In the beginning we only consider the simple case of *homogeneous* chains when there are three common transition probabilities: q_c+r for clockwise and 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 we can avoid this issue.

Theorem 3.5. *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 asymptotically almost surely (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.*

Using Theorem 3.4 we can show the following bound.

Proposition 3.6. *For M1, let us assume the nodes with long range edges are equidistant from each other. Then for any homogeneous chain,*

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

All the other mixing time bounds are based on estimating the *conductance*:

Definition 3.7. The conductance of a Markov chain on the set $\mathcal{X}$ is

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

where S^C denotes the complement of S.

Theorem 3.8. *For M1 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

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

Theorem 3.9. *For M2 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

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

Theorem 3.10. *For M3 the conductance of the homogeneous chain satisfies the following inequality a.a.s.:*

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

Connecting the conductance with the mixing time is done using the theorem of Lovász and Simonovits [8]. 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 3.11. *For an aperiodic, irreducible, lazy Markov chain the following bounds for mixing time holds:*

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

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

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

Corollary 3.12. *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 homogeneous 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.

At this stage, it is easy to transform this result as follows:

Corollary 3.13. *The bounds of Corollary 3.12 also hold for the fastest M1, M2, M3 chains.*

There is a big gap between the lower and upper bounds for non-reversible chains. Still, we hope there is a considerable gain as shown in Figure 2. This is a plot of mixing times of homogeneous reversible and non-reversible chains on several graphs coming from M2, for $\alpha = 1.5$.

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.

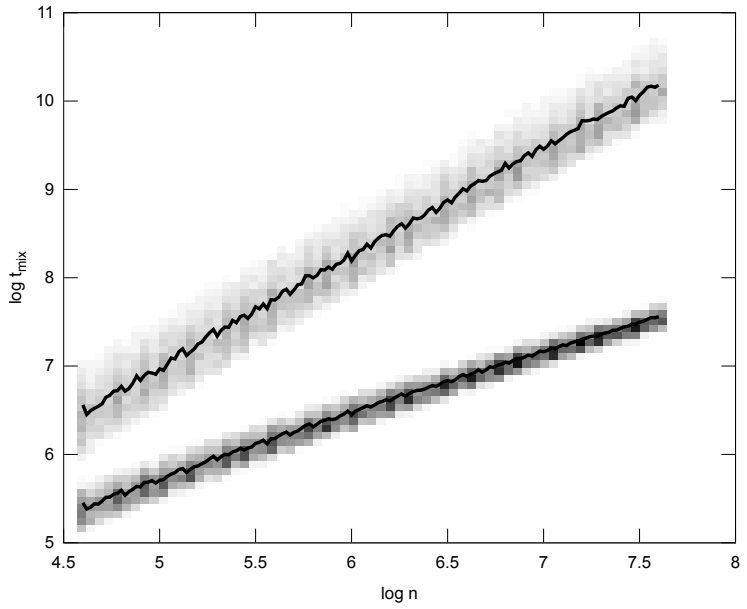


FIGURE 2. Log-log plot for mixing times of homogeneous M2 chains

On the other hand we may guess the mixing time for reversible chains is roughly $n \log^{\delta} n$ based on Corollary 3.12. 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.

REFERENCES

[1] E. CZEIZEL AND G. TUSNÁDY, *Aetiological studies of isolated common congenital abnormalities in Hungary*, Akadémiai Kiadó Budapest, 1984.

[2] 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.

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

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

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

[6] B. GERENCSÉR, B. RÁTH, AND G. TUSNÁDY, *Multifactorial inheritance with selection*, Acta Math. Hung., (2012). Online First.

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

[8] 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.