University of Vienna

Stochastische Prozesse

January 27, 2015

Contents

Introduction

This script is based on the lecture "Stochastische Prozesse" hold by Univ.-Prof. Dr. Josef Hofbauer in the winter semester of 2014. If you spot any mistakes, please write an email to basti.fischer.wien@gmail.com. I will upload the recent version to
https://elearning.mat.univie.ac.at/wiki/images/e/ed/Stoch_pro_14_hofb.pdf.

I want to thank Hannes Grimm-Strele and Matthias Winter for sending me the files of their script of a similar lecture held by Univ.-Prof. Dr. Reinhard Bürger in 2007.

1 Random Walks

1.1 Heads or tails

Let us assume we play a fair game of heads or tails, meaning both sides of our coin have the same probability $p = 0.5$. We play for N rounds, so there are clearly 2^N different possibilities of how our game develops, each with the same probability of $\frac{1}{2^N}$. We define the random variable

$$X_n := \begin{cases} 1, & \text{for head and} \\ -1 & \text{for tails} \end{cases}$$

as the outcome of our n'th throw and

$$S_N := \sum_{n=1}^{N} X_n.$$

So if we bet one euro on head each time (and since the game is fair, are able to win one euro each time), S_n will tell us our capital after n rounds. Mathematically speaking, S_n describes a so called random walk on the natural numbers.

Now let us look at the probability distribution of S_n. If we have k times head with N repetitions in total, we get $S_N = k - (N - k) = 2k - N$ and the probability of this event is

$$P(S_N = 2k - N) = \binom{N}{k} \frac{1}{2^N},$$

since we have to choose k out of N occasions for head and 2^N is the total number of paths. We can transform this to

$$P(S_n = j) = \begin{cases} \binom{N}{\frac{N+j}{2}} \frac{1}{2^N}, & \text{if } N + j \text{ is even and} \\ 0 & \text{if } N + j \text{ is odd,} \end{cases}$$

since this is impossible.

Exercise 1. *Compute the mean value and the variance of S_n in two ways each.*

With $A(N, j)$ we denote the number of paths from $(0, 0)$ to (N, j) and clearly it is

$$A(N, j) = \begin{cases} \binom{N}{\frac{N+j}{2}}, & \text{if } N + j \text{ is even and} \\ 0 & \text{if } N + j \text{ is odd.} \end{cases}$$

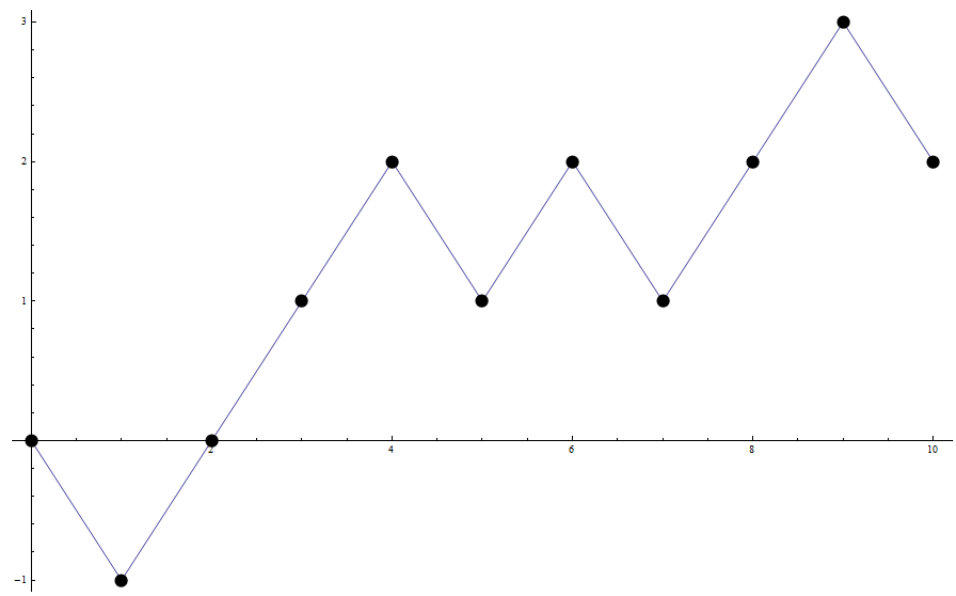


Figure 1: The random walk belonging to the event $(-1, 1, 1, 1, -1, 1, -1, 1, 1, -1)$

1.2 Probability of return

We now take a closer look on the probability of getting the same amount of heads and tails after $2N$ repetitions, so $S_{2N} = 0$. Stirlings formula

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

tells us more about the long time development, we get

$$P(S_{2N} = 0) = \binom{2N}{N}\frac{1}{2^N} = \frac{(2N)!}{2^{2N}N!^2} \sim \frac{\sqrt{4\pi N}(2Ne)^{2N}}{(\sqrt{2\pi N}N^N)^2(2e)^{2M}} = \frac{1}{\sqrt{\pi N}}$$

which tends to 0 as N grows to infinity.

Exercise 2. *Let p_n denote the probability $P(S_{2n} = 0) = \binom{2n}{n}2^{-2n}$. Prove directly that $[np_n^2, (n + \frac{1}{2})p_n^2]$ is a sequence of nested intervals.*

Exercise 3. *Show for a symmetrical random walk, that for j fixed and $N \to \infty$ one has*

$$P(S_N = j) \sim \sqrt{\frac{2}{\pi N}}.$$

1.3 Reflection principle

Lemma 1.1. *The number of paths from* $(0,0)$ *to* (N,j) *that do not hit the axis (i.e.* $S_k > 0$ *for* $k > 0$*) is given by*

$$A(N-1, j-1) - A(N-1, j+1)$$

Proof. The number of paths from $(0,0)$ to (N,j) above the axis is given by the total number of paths from $(1,1)$ to (N,j) minus the paths from $(1,1)$ to (N,j) that do hit the axis. This second number is the same as the number of paths from $(1,-1)$ to $(N,j+2)$, because we can simply reflect the part of the path before it reaches the axis for the first time. $\square$

A simple consequence of the reflection principle is the

Theorem 1.2 (Ballot theorem)**.** *The number of paths from* $(0,0)$ *to* (N,j) *that do not hit the axis is* $\frac{j}{N}$ *times the number of paths from* $(0,0)$ *to* (N,j).

We can use the Ballot theorem in daily life, imagine an election between two candidates, there are N voters, candidate A gets k votes, so B gets $k-l$ votes. Assuming B wins,what is the probability that during the counting, B is always in the lead? The theorem gives the answer by

$$\frac{\frac{k-l}{N}A(N,k-l)}{A(N,k-l)} = \frac{k-l}{N} = \frac{k-l}{k+l}.$$

Exercise 4. *Prove the Ballot theorem.*

1.4 Main lemma for symmetric random walks

We define $u_{2M} := P(S_{2M} = 0) = \binom{2M}{M}\frac{1}{2^{2M}}$, then we get

Lemma 1.3 (Main lemma)**.** *The number of paths with length* $2M$ *from* $(0,0)$ *that do not hit the axis is the same as the number of paths that end in* $(2M,0)$*. Speaking in terms of probability it is*

$$P(S_1 \neq 0, S_2 \neq 0, ..., S_{2M} \neq 0) = P(S_{2M} = 0).$$

Proof. Let us call the first number $A_{\neq 0}$ and the final point of each path $(2M, 2j)$. At first we observe simply by symmetrical reasons that $A_{\neq 0}$ is twice the number of paths that lie above the axis. So, counting all possible values of j we get

$$
\begin{aligned}
A_{\neq 0} &= 2\sum_{j=1}^{M}\big[A(2M-1,2j-1)-A(2M-1,2j+1)\big] \\
&= 2[A(2M-1,1)-\underbrace{A(2M-1,2M+1)}_{=0}] \\
&\overset{\text{reflection}}{=} A(2M-1,1)+A(2M-1,-1)=A(2M,0)
\end{aligned}
$$

□

Now it is easy to see that

Corollary 1.4. *The probability to have no tie within the first N rounds is*

$$P(S_N=0)\sim\sqrt{\frac{2}{\pi N}}\to 0 \quad (N\to\infty).$$

1.5 First return

We define the probability that the first return of a path to the axis is after $2M$ rounds as f_{2m}. Then we have

Theorem 1.5.

$$f_{2M}=u_{2M-2}-u_{2M}.$$

Proof.

$$
\begin{aligned}
&\#\text{ paths of length } 2M \text{ from } (0,0) \text{ with first return at time } 2M \\
&= \#\text{ paths of length } 2M \text{ with } S_i\neq 0 \text{ for } i=1,...,M-1 \\
&- \#\text{paths of length } 2M \text{ with } S_i\neq 0 \text{ for } i=1,...,M \\
&= 4\#\text{ paths of length } 2M-2 \text{ that do not hit the axis} \\
&- \#\text{paths of length } 2M \text{ with } S_i\neq 0 \text{ for } i=1,...,M \\
&= 4\cdot u_{2m-2}\cdot 2^{2M-2}-u_{2M}\cdot 2^{2M} \\
&= 2^{2M}[u_{2m-2}-u_{2M}].
\end{aligned}
$$

□

Corollary 1.6. $f_{2M}=\frac{1}{2M-1}u_{2m}=\frac{1}{2M-1}\binom{2M}{M}\frac{1}{2^{2M}}$

Corollary 1.7. $\sum_{M=1}^{\infty} f_{2M} = u_0 - u_2 + u_2 - u_4 + ... = u_0 = 1$

Exercise 5. *Show the following connection between the probabilities of return u_{2n} and first return f_{2n}.*

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \cdots + f_{2n} u_0.$$

Exercise 6. *Show that*

$$u_{2n} = (-1)^n \binom{-\frac{1}{2}}{n}, \quad f_{2n} = (-1)^{n-1} \binom{\frac{1}{2}}{n}.$$

Exercise 7. *From the main lemma (1.3) conclude (without calculations) that*

$$u_0 u_{2n} + u_2 u_{2n-2} + \cdots + u_{2n} u_0 = 1.$$

1.6 Last visit

Now we look at a game which lasts $2M$ rounds and we define the probability, that the last tie was at time $2k$ as $\alpha_{2k,2M}$.

Theorem 1.8 (Arcsin law of last visit)**.** $\alpha_{2k,2M} = u_{2k} \cdot u_{2M-2k}$.

Proof. The first segment of the path can be chosen in $2^{2k} u_{2k}$ ways. Setting the last tie as a new starting point the main lemma tells us, that the second segment of length $2M - 2k$ can be chosen in $2^{2M-2k} u_{2M-2k}$ ways. $\square$

Corollary 1.9.

1. *$\alpha_{2k,2M}$ is symmetric is respect of k and $M - k$.*
2. *$P($"the last tie is in the first half of the game"$) = \frac{1}{2}$.*
3. *$\alpha_{2k,2M} \overset{(1.2)}{\sim} \frac{1}{\sqrt{\pi k}} \frac{1}{\sqrt{\pi(M-k)}} = \frac{1}{\pi} \frac{1}{\sqrt{k(M-k)}}$*

The third point describes the long term development of the last tie's appearance, which is pretty non-intuitional. For example, if we play our head or tails game for one year each second, the probability that the last tie is within the first 9 days is around ten percent and within the first 2 hours and 10 minutes still around one percent. The term $\frac{1}{\pi} \frac{1}{\sqrt{k(M-k)}}$ is the density of the so-called arc-sin-distribution, because

$$\int_0^t \frac{1}{\pi} \frac{1}{\sqrt{k(M-k)}} \mathrm{d}x = \frac{2}{\pi} \arcsin(\sqrt{t}).$$

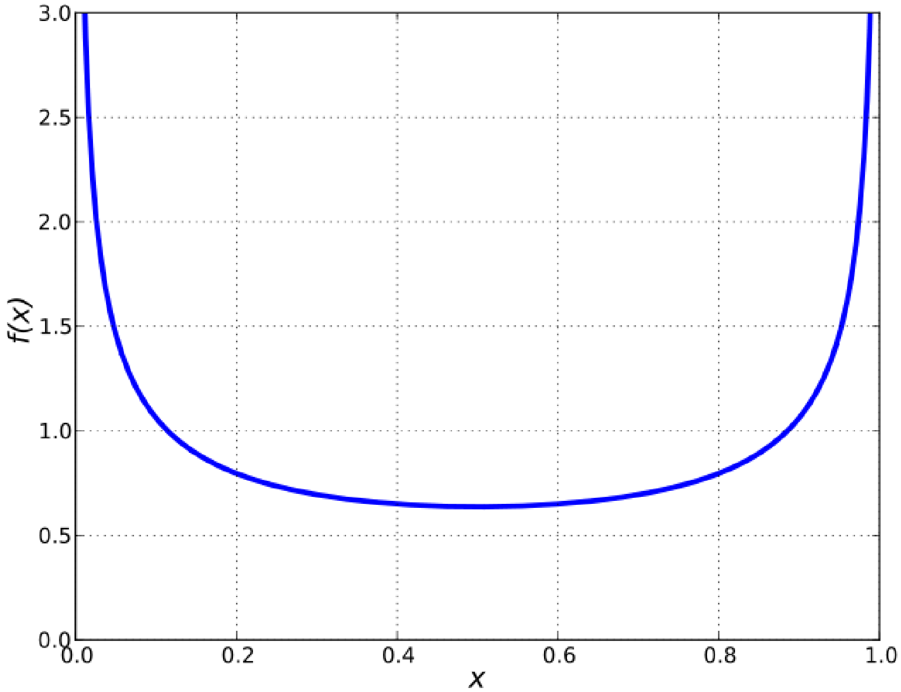


Figure 2: Arc-sin-distribution

1.7 Sojourn times

The next question is about the sojourn times. We look for which fraction of time one of the players is in the lead.

Theorem 1.10. *The probability that in the time interval from* 0 *to* $2M$ *the path spends* $2k$ *time units on the positive side and* $2M - 2k$ *units on the negative side is given by* $\alpha_{2k,2M}$.

The proof can be found in [1].

Corollary 1.11. *If* $0 < t < 1$, *the probability that less than* $t2M$ *time units are spent on the positive and more than* $(1-t)2M$ *units on the negative side tends to* $2/\pi \arcsin(\sqrt{t})$ *as* M *tends to infinity.*

If we restrict our paths to those ending on the axis, we get a different result.

Theorem 1.12. *The number of paths of length $2M$ such that $S_{2m} = 0$ and exactly $2k$ of its sides lie above the axis is independent of k and is given by*

$$\frac{1}{M+1}\binom{2M}{M},$$

which are the Catalan numbers.

Exercise 8. *Prove the previous theorem.*

The proof can also be found in [1].

1.8 Position of maxima

If we have a path of length $2M$, we say the first maximum occurs at time k if $S_i < S_k \forall i < k$ and $S_i \leq S_k \forall i > k$.

Theorem 1.13. *The probability that the first maximum occurs at time $k = 2l$ or $k = 2l + 1$ is given by*

$$\begin{cases} \frac{1}{2}u_{2l}u_{2M-2l}, & \text{if } 0 < k < 2M, \\ u_{2M} & \text{if } k = 0 \text{ and} \\ \frac{1}{2}u_{2M} & \text{if } k = 2M. \end{cases}$$

Note that for the last maxima, the probabilities are simply interchanged. If M tends to infinity and k/M tends to some fixed t, we get an arcsin-law again.

1.9 Changes of sign

We say at time k there is a change of sign if and only if $S_{k-1}S_{k+1} < 0$

Theorem 1.14. *The probability $\zeta_{r,2n+1}$ that up to time $2n + 1$ there are exactly r changes of sign is given by*

$$\zeta_{r,2n+1} = 2P(S_{2n+1} = 2r + 1) = \binom{2n+1}{n+r+1}\frac{1}{2^{2n}}$$

for $r = 0, \ldots, n$.

A proof can be found in [1] in the third chapter.

Corollary 1.15. *The following chain of inequalities holds:*

$$\zeta_{0,2n+1} \geq \zeta_{1,2n+1} > \zeta_{2,2n+1} > \dots$$

As an example, we get $\zeta_{0,99} = 0.159$, $\zeta_{1,99} = 0.153$, $\zeta_{2,99} = 0.141$ and $\zeta_{13,99} = 0.004$.

1.10 Return to the origin

Let X_k be a random variable which is 1 if $S_{2k} = 0$ and 0 else. Then we have

$$P(X_k = 1) = u_{2k} = \frac{1}{2^k}\binom{2k}{k} \sim \frac{1}{\sqrt{\pi k}}.$$

Define $X^{(n)} := \sum_{i=1}^{n} X_i$, then this random variable counts the number of returns to the origin in $2n$ steps. For the mean value we get

$$E(X^{(n)}) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} u_{2i}$$

so for n large enough

$$E(X^{(n)}) \sim \sum_{i=1}^{n} \frac{1}{\sqrt{\pi i}} = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} \frac{1}{\sqrt{i}} = 2\sqrt{\frac{n}{\pi}}$$

follows. To be more precisely, we get

$$E(X^{(n)}) = (2n+1)\binom{2n}{n}\frac{1}{2^{2n}} - 1 = 2\sqrt{\frac{n}{\pi}} - 1 + \mathrm{o}(1/n).$$

Exercise 9. *Show that*

$$\sum_{i=1}^{n} \frac{1}{\sqrt{i}} \sim 2\sqrt{n}.$$

Exercise 10. *Compute the sum*

$$\sum_{i=1}^{n} u_{2i} = \sum_{i=1}^{n} \binom{2i}{i} 2^{-2i}$$

and find an asymptotic formula as $n \to \infty$.

1.11 Random walks in the plane $\mathbb{Z}^2$

Let us look at a 2-dimensional random walk, we can go from a point $(x, y) \in \mathbb{Z}^2$ to $(x \pm 1, y \pm 1)$ with the probability $1/4$ each. After $2n$ steps we arrive at (X_{2n}, Y_{2n}). Now X_n and Y_n are independent random walks on $\mathbb{Z}$, so our previous results all work perfectly well. For example, the event $A_k :=$"after $2k$ steps the particle returns to the origin" is equal to

$$P(A_k) = P(X_{2k} = Y_{2k} = 0) = P(X_{2k} = 0)P(Y_{2k} = 0) = u_{2k}^2.$$

Now define as in the section before $U_k := \chi_{A_k}$ and $U^{(n)}$ as the sum over all U_k, which counts the number of returns to the origin. Then we get

$$E(U^{(n)}) = \sum_{k=1}^{n} u_{2k}^2 \sim \sum_{k=1}^{n} \frac{1}{\pi k} = \frac{1}{\pi} \sum_{k=1}^{n} \frac{1}{k} \sim \frac{\log(n)}{\pi},$$

so the number of returns tends to infinity if n does so.

1.12 The ruin problem

Now we look at a game where a gambler wins 1 unit with probability p and looses 1 unit with probability $q = 1 - p$. We denote his initial capital with z and the adversary's initial capital with $a - z$. The game continues until one of the two is ruined, i.e. the capital of the gambler is either 0 or a. The two things we are interested now is on one hand the probability of the gamblers ruin and on the other hand the duration of our game. We can interpret this scenario as a asymmetric (if $p \neq q$) random walk on the natural numbers $\mathbb{N}$ (with 0) with absorbing barriers. If $p < q$ we say we have a drift to the left.

We define q_z as the probability of the gamblers ruin and p_z as the probability of his winning. Our goal is to show that $q_z + p_z = 1$ and that the duration of the game is finite.

It is easy to see that

$$q_z = pq_{z+1} + qq_{z-1}$$

holds for $0 < z < a$. With the boundary conditions $q_0 = 1$ and $q_a = 0$ we get a linear recurrence equation for q_z of second order which can be solved using the ansatz $q_z = \lambda^z$. We get the two solutions $\lambda = 1$ and $\lambda = q/p$ and since the set of solutions is a vector space, our general solution is $q_z = A + B(q/p)^z$. Using the boundary conditions, our final and unique solution is

$$q_z = \frac{(\frac{q}{p})^a - (\frac{q}{p})^z}{(\frac{q}{p})^a - 1}.$$

But remark that this solution does not work for $p = q = 1/2$! In the symmetric case, we get $q_z = 1 - \frac{z}{a}$. Because of the symmetry of the problem (the gambler is now the adversary), we get

$$p_z = \begin{cases} 1 - \frac{a-z}{z} = \frac{z}{a}, & \text{if } p = q \text{ and} \\ \frac{(\frac{p}{q})^a - (\frac{p}{q})^{a-z}}{(\frac{p}{q})^a - 1} & \text{else.} \end{cases}$$

Now it is simple to check that $p_z + q_z = 1$.

What is the expected gain of our gambler? We denote this number with G and observe

$$G = \begin{cases} a - z & \text{with probability } 1 - q_z \\ -z & \text{with probability } q_z. \end{cases}$$

Now we have for the expected value in the asymmetric case

$$E(G) = a(1 - q_z) - z = a\frac{(\frac{q}{p})^a - (\frac{q}{p})^z}{(\frac{q}{p})^a - 1} - z.$$

It is easy to show that $E(G) = 0$ if $p = q$ and $E(G) < 0$ if $p < q$.

Exercise 11. *Consider a random walk on $\mathbb{Z}$ with probabilities p and $q = 1-p$ for moving right and left. Show that starting at 0, the probability of ever reaching the state $z > 0$ equals 1 if $p \leq q$ and $(\frac{p}{q})^z$ if $p < q$.*

1.13 How to gamble if you must

This section is named after the book of Dubbins and Savage. Assume a gambler starts with capital z and stops when he reaches $a > z$ or when he is bankrupt. For example $z = 90$ and $a = 100$, what is the best strategy, i.e. what is the right stake? It is clear that halving the stakes is the same as doubling the capitals, so we get

$$q_z = \frac{(\frac{q}{p})^{2a} - (\frac{q}{p})^{2z}}{(\frac{q}{p})^{2a} - 1} = \frac{(\frac{q}{p})^a - (\frac{q}{p})^z}{(\frac{q}{p})^a - 1} \cdot \frac{(\frac{q}{p})^a + (\frac{q}{p})^z}{(\frac{q}{p})^a + 1} \overset{\text{if } p<q}{>} q_z.$$

In the example above, if $p = 0.45$ if our stake is 10, the probability of ruin is only 0.21, while if our stake is 1 it is 0.866. As a tends to infinity, we get

$$q_z = \begin{cases} 1 & \text{if } p \leq q \\ \frac{q}{p} & \text{if } q > p. \end{cases}$$

1.14 Expected duration of the game

Let D_z be the expected duration of our game. In the chapter about Markov chains, we will prove that D_z is finite. For now we have the trivial relation

$$D_z = pD_{z+1} + qD_{z-1} + 1,$$

where the 1 is added because of the time unit that we need to get to the next condition of our game. So this time we get a non-homogeneous linear recurrence equation of second order with the boundaries $D_0 = D_a = 0$. We solve the homogeneous part as in the last section and use $D_z = Cz$ as an ansatz for the special solution. Again the symmetric case must be solved separately by the ansatz $D_z = A + Bz + Cz^2$ and so we get

$$D_z = \begin{cases} z(a-z) & \text{if } p = q \text{ and} \\ \frac{z}{q-p} - \frac{a}{q-p} \cdot \frac{1-(\frac{q}{p})^z}{1-(\frac{q}{p})^a} & \text{else.} \end{cases}$$

This result is very counterintuitive, for example, if $a = 1000$ and $z = 500$ and $p = q$ we expect to play for 250000 rounds. And for the same probabilities even if $z = 1$ and $a = 1000$ our expected duration is 999.

Exercise 12. *Consider a random walk on* $0, 1, 2, \dots$ *with only one absorbing barrier at 0 and probabilities* p *and* $q = 1 - p$ *for moving right and left. Denote again with* D_z *the expected time until the walk ends (i.e. it reaches 0) if we start at the state* z*. Show*

$$D_z = \begin{cases} \frac{z}{q-p} & \text{if } p < q \\ \infty & \text{if } p \geq q. \end{cases}$$

1.15 Generating function for the duration of the game

We now want to compute the probability, that the gambler is ruined in the n^{th} round. Of course, this depends also on the initial capital z, so we get a linear recurrence relation in two variables, namely

$$u_{z,n+1} = p \cdot u_{z+1,n} + q \cdot u_{z-1,n}, \quad 1 \leq z \leq a-1, \quad n \geq 1 \tag{1}$$

with the boundary conditions

- $u_{0,n} = u_{a,n} = 0$ for $n \geq 1$

- $u_{z,0} = 0$ for $z \geq 1$ and
- $u_{0,0} = 1$.

We define the generating function $U_z(s) := \sum_{n=0}^{\infty} u_{z,n} s^n$. Multiplying (1) with s^{n+1} and summing over all different cases of n, we get

$$\underbrace{\sum_{n=0}^{\infty} u_{z,n+1} s^{n+1}}_{U_z(s)} = ps \underbrace{\sum_{n=0}^{\infty} u_{z+1,n} s^n}_{U_{z+1}(s)} + qs \underbrace{\sum_{n=0}^{\infty} u_{z-1,n} s^n}_{U_{z-1}(s)}.$$

Therefore we get a new recurrence relation and managed to eliminate one variable. We solve

$$\begin{cases} U_z(s) = psU_{z+1}(s) + qsU_{z-1}(s) \\ U_0(s) = \sum_{n=0}^{\infty} u_{0,n} s^n = 1 \\ U_a(s) = \sum_{n=0}^{\infty} u_{a,n} s^n = 0 \end{cases}$$

with the ansatz $U_z(s) = \lambda(s)^z$ and finally compute

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps},$$

which are real solutions for $0 < s < 1$. Using the boundary values we get as a final solution

$$U_z(s) = \frac{\lambda_1(s)^a \lambda_2(s)^z - \lambda_1(s)^z \lambda_2(s)^a}{\lambda_1(s)^a - \lambda_2(s)^a}.$$

In a similar way, we find the generating function for the probability, that the gambler wins in the n^{th} round. It is given by

$$\frac{\lambda_1(s)^z - \lambda_2(s)^z}{\lambda_1(s)^a - \lambda_2(s)^a}.$$

One can also find an explicit formula for $u_{z,n}$. It is given by

$$u_{z,n} = \frac{1}{a} 2^n p^{\frac{n+z}{2}} q^{\frac{n+z}{2}} \sum_{k=1}^{n} \cos^{n-1} \frac{\pi k}{a} \sin \frac{\pi k}{a} \sin \frac{\pi z k}{a}$$

and was already found by Lagrange.

Exercise 13. *Show the previous formula for* $u_{z,n}$.

The calculation can also be found in [1] in XIV.5.

Exercise 14. *Banach's match problem: Stefan Banach got a box of matches in both of his two pockets. With probability* $\frac{1}{2}$ *he took one match out of the left respectively the right pocket. If he found one box empty, he replaced both of them with new ones containing* n *matches. What is the probability that* k *matches are left in the second box before the replacement?*

1.16 Connection with the diffusion process

We now take a look at non-symmetric random walks. We define $\sum_{k=1}^{n} X_k$ with $P(X_k = 1) = p$ and $P(X_k = -1) = 1 - p$. Simple calculation gives $E[S_n] = (p - q)n$ and $\text{Var}[S_n] = 4pqn$. Now we rescale our steps so they have length δ. Since S_n is linear, we get

- $E[\delta S_n] = (p - q)\delta n$
- $\text{Var}[\delta S_n] = 4pq\delta^2 n$.

If $p \neq q$ and n gets large, we choose δ in such a way that $E[\delta S_n]$ is bounded and therefore $\text{Var}[\delta S_n] \sim \delta$, by what the process looks like a deterministic, linear motion. From the physical point of view, this process is in connection with the Brownian motion, the random movement of a particle in a liquid. By collisions with other smaller particles, it gets displaced by $\pm\delta$ (in our situation, we only look at one dimension). If we measure the average displacement C and the variance per time unit D and assume the number of collisions per time unit is r, we should actually get $C \approx (p - q)\delta r$ and $D \approx 4pq\delta^2 r$. So as $\delta \to 0$, $r \to \infty$ and $p \to \frac{1}{2}$ we demand $(p - q)\delta r \to C$ and $4pq\delta^2 r \to D > 0$.
In an accelerated random walk, the n^{th} step ($S_n = k$) takes place at time $\frac{n}{r}$ at position $\delta S_n = \delta k$. Define $v_{k,n} := P(S_n = k)$, therefore $S_0 = 0$ and

$$v_{k,n+1} = p \cdot v_{k-1,n} + q \cdot v_{k+1,n}$$

holds. If $\frac{n}{r} \to t$ and $k\delta \to x$ we deduce

$$v(x, t + \frac{1}{r}) = p \cdot v(x - \delta, t) + q \cdot v(x + \delta, t).$$

We demand v to be smooth so that we can use the Taylor expansion, therefore

$$\begin{aligned} v(x,t) + \frac{1}{v} v_t(t,x) + \mathcal{O}(\frac{1}{r^2}) = p[v(x,t) - \delta v_x(x,t) + \frac{\delta^2}{2} v_{xx}(x,t) + \mathcal{O}(\delta^3)] \\ + q[v(x,t) + \delta v_x(x,t) + \frac{\delta^2}{2} v_{xx}(x,t) + \mathcal{O}(\delta^3)] \end{aligned}$$

and so we get

$$v_t = \underbrace{(q-p)\delta r}_{\to -c} + \frac{1}{2} \underbrace{\delta^2 r}_{\to D} v_{xx} + \mathcal{O}(\frac{1}{r}) + \mathcal{O}(r\delta^3),$$

which leads in the limit to the Focker-Planck-equation (or forward Kolmogoroff equation)

$$v_t = -cv_x + \frac{1}{2} D v_{xx},$$

where c denotes the drift and D the diffusion constant. The function $v(t,.)$ is a probability density, in fact

$$v_{k,n} \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} q^{\frac{n+k}{2}} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-n(p-q))^2}{\delta npq}} \sim \frac{2\delta}{\sqrt{2\pi Dt}} e^{-\frac{(x-ct)^2}{2Dt}}.$$

As $n \to rt$ and $k\delta \to x$ our probability $v_{k,n}$ behaves like

$$v_{k,n} \sim P(k\delta < \delta S_n < (k+2)\delta) \approx 2\delta v(x,t).$$

Notice that

$$v(x,t) = \frac{1}{\sqrt{2\pi Dt}} e^{-\frac{(x-ct)^2}{2Dt}}$$

is also a fundamental solution of the PDE above. Such a random process $(x_t)_{t\geq 0}$ whose density is $v(x,t)$ is called Brownian motion, Wiener process or diffusion process.

2 Branching processes

2.1 Extinction or survival of family names

In the 18^{th} century British scientist Francis Galton observed, that the number of family names was decreasing. Together with the mathematician Henry

William Watson, he tried to find a mathematical explanation. Today it is known that the French mathematician Irénée-Jules Bienaymé worked on the same topic around thirty years earlier and he, other than Galton and Watson, managed to solve the problem correctly. Assume we have an individual with a natural number of sons. Let p_k denote the probability that he has k sons and X_n the number of individuals with his name in the n^{th} generation. Further q is the probability that the name goes extinct (so there is some $X_n = 0$) and m is the expected number of sons.

Theorem 2.1. *1. If m does not exceed 1, the name will die out (except for the trivial case $p_1 = 1$).*

2. If m is greater than 1, q is smaller than 1, so extinction is possibly avoided.

2.2 Proof using generating functions

For the proof, we look at the conditional probability $P(X_n = i|X_m = j)$. It has two important properties, namely

- time invariance $P(X_{n+1} = i|X_{m+1} = j) = P(X_n = i|X_m = j)$ and
- independent reproducing $P(X_n = 0|X_0 = k) = P(X_n = 0|X_0 = 1)^k$,

since the individuals multiply independently. Therefore we assume $X_0 = 1$ by now. We define the generating function $F(s) = \sum_{k=0}^{\infty} p_k s^k$ which converges for $|s| \leq 1$. Furthermore we define

$$F_n(s) = \sum_{k=0}^{\infty} P(X_n = k)s^k$$

as the generating function for X_n. Clearly $F_1(s) = F(s)$ holds and also $P(X_n = 0) = F_n(0)$ if we assume $0^0 = 1$. Moreover, the sequence $(F_n(0))_{n\geq 1}$ is non-decreasing and has the limit q. Using the both properties of the conditional probability from above, we get

$$\begin{aligned} F_{n+1}(0) = P(X_{n+1} = 0) &= \sum_{k=0}^{\infty} \underbrace{P(X_{n+1} = 0|X_1 = k)}_{P(X_n=0|X_0=1)^k} \underbrace{P(X_1 = k)}_{p_k} \\ &= \sum_{k=0}^{\infty} p_k F_n(0)^k = F(F_n(0)), \end{aligned}$$

and therefore, with $n \to \infty$ we get the fixed point problem $q = F(q)$. In their paper, Watson and Gallon observed that 1 is a fixed point since $F(1) = \sum_k p_k = 1$, but they forgot to consider a smaller one.

Lemma 2.2. *q is the smallest fixed point of $F : [0,1] \to [0,1]$.*

Proof. Let $a \geq 0$, $F(a) = a$ therefore we have $F(0) \leq F(a) = a$ since $F'(s) = \sum kp_k s^{k-1} \geq 0$ hence F is increasing. Now by induction it follows

$$\begin{aligned} F_n(0) \leq a &\Rightarrow F(F_n(0)) \leq F(a) \\ &\Rightarrow F_{n+1}(0) \leq a. \end{aligned}$$

For $n \to \infty$ we get $q \leq a$. It is easy to see that $F(s)$ is convex for $s \in [0,1]$ by looking at the second derivative. Now we get two cases.

1. $F'(1) = \sum_{k=0}^{\infty} kp_k = m \leq 1$,
   therefore $F'(s) \leq F'(1) = m \leq 1 \forall s \in [0,1]$, so there can't be any fixed point but 1 (cf. fig.3), so $q = 1$

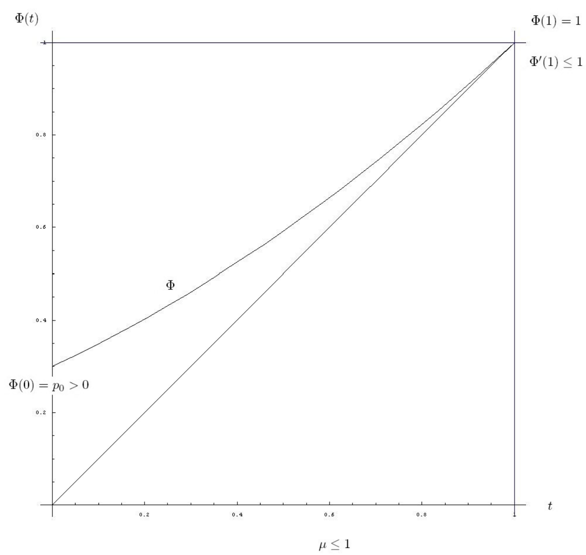


Figure 3: $F'(1) \leq 1$

2. $F'(1) = m > 1$
   therefore since $F(0) > 0$ we get with a similar argument (cf. fig.4) $q < 1$.

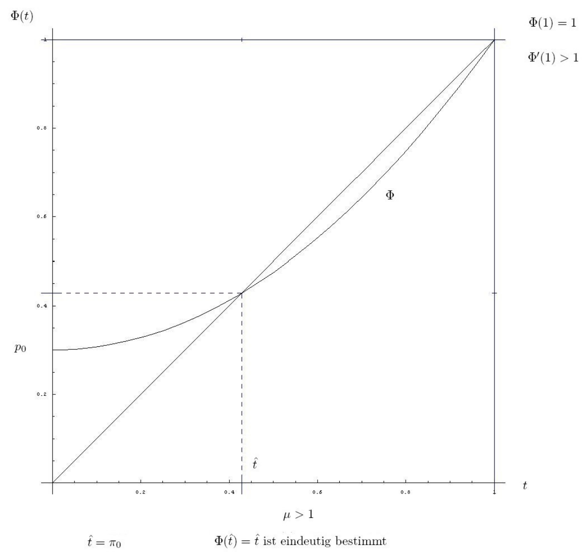


Figure 4: $F'(1) > 1$

$\square$

There are two special cases, first assume $p_0 + p_1 = 1$ then $F(s) = p_0 + sp_1$ is linear and again we have the fixed point in 1. The other case is $p_1 = 1$. Here it is trivial that extinction is impossible and therefore $q = 0$ (in the equation, every point is a fixed one). A Galton-Watson process is called

- *subcritical* if $m < 1$,
- *critical* if $m = 1$ and
- *supercritical* if $m > 1$.

As an easy example, suppose p_k is given by a geometric distribution, thus $p_k = ap^k$ with $0 < p < 1$. The number a is determined by

$$1 = \sum_{k=0}^{\infty} p_k = a \sum_{k=0}^{\infty} p^k = a \frac{1}{1-p} \qquad \Rightarrow a = 1 - p.$$

The generating function is given by

$$F(s) = (1-p) \sum_{k=0}^{\infty} p^k s^k = \frac{1-p}{1-ps}$$

and solving the fixed point equation we get the two solutions 1 and $\frac{1}{p} - 1$ and therefore

$$\frac{1}{p} - 1 \leq 1 \Leftrightarrow \frac{1}{p} \leq 2 \Leftrightarrow p \geq \frac{1}{2}.$$

So the name can only survive if and only if $p \geq \frac{1}{2}$.

Exercise 15. *Consider a Galton-Watson process with $p_0 = p_3 = \frac{1}{2}$. Find m and q, the probability of extinction.*

Exercise 16. *Consider a Galton-Watson process with an almost geometric distribution $p_k = bp^{k-1}$ for $k = 1, 2, \ldots$ and find p_0. Compute the generation function of X_n explicitly. Compute m and q, in particular for the specific choice $b = \frac{1}{5},\ p = \frac{3}{5}$.*

2.3 Some facts about generating functions

Assume X is a random variable with values $0, 1, 2, \ldots$ and denote $P(X = k)$ by p_k. Then the generating function is given by $F_x(s) = \sum_{k=0}^{\infty} p_k s^k$ and has the following properties:

- $F_x(s)$ converges for $|s| \leq 1$ and is analytic for $|s| < 1$.
- The function can also be interpreted as the expected value of s^X.
- If two random variables have the same generating function, they have the same distribution because p_k is given by
$$p_k = \frac{1}{k!} F_x^{(k)}(0).$$
- If $E[X]$ exists, it is given by $E[X] = \lim_{s \nearrow 1} F_x'(s)$.
- If $\mathrm{Var}[X]$ exists, it is given by $\mathrm{Var}[X] = F_x''(1) + F_x'(1) - F_x'(1)^2$.

As an example, we look at the Poisson distribution $\mathcal{P}(\lambda)$. The generating function is given by

$$F(s) = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{s^k \lambda^k}{k!} = e^{\lambda(s-1)}.$$

Therefore we get

$$E[X] = F'(1) = \lambda e^{\lambda(1-1)} = \lambda.$$

We end this section with two theorems which remain unproved.

Theorem 2.3. *If X and Y are independent, then $F_{x+y}(s) = F_x(s)F_y(s)$.*

Theorem 2.4. *If X_n is a sequence of random variables and X another one, then*

$$P(X_n = k) \xrightarrow{n\to\infty} P(X = k) \Leftrightarrow F_{x_n}(s) \xrightarrow{n\to\infty} F_x(s) \forall s \in [0, 1].$$

2.4 Moment and cumulant generating functions

Again X shall be a discrete random variable with non-negative values and we denote $P(X = k)$ by p_k. The moment generating function is defined by

$$M_x(t) = E[e^{tX}] = \sum_{k=0}^{\infty} p_k e^{kt} = F_x(e^t).$$

It converges for all $t \leq 0$ and dependent of the distribution also for $0 < t \leq \alpha$. It has the following properties:

- $M_x(0) = 1$.
- $M_x'(0) = E[X]$.
- $M_x''(0) = E[X^2]$.
- $M_x^{(n)}(0) = E[X^n]$, which is also called the n^{th} moment.

The cumulate generating function is defined by

$$K_x(t) = \log M_x(t) = \log F_x(e^t)$$

and has the useful properties

- $K_x'(0) = E[X]$ and
- $K_x''(0) = \mathrm{Var}[X]$.

2.5 An example from genetics by Fischer (1930)

Imagine a population of N diploid individuals, so we have $2N$ genes. Assume N large enough for later approximations, and let all individuals have initially genotype AA. By mutation, one individual develops the genotype Aa. We are now interested in the probability, that the mutant gene a will survive. We need two assumptions:

- Aa should have a selective advantage, i.e. its fitness should exceed the normal AA fitness by the factor $(1+h)$, where h is small (fitness corresponds to the number of offspring).
- As long as Aa is rare, the homozygotes aa should be too rare to be relevant.

Let the number of offspring be Poisson-distributed with parameter $\lambda = 1 + h$, then the generating function is given as in the last example in 2.3 by $F(s) = e^{\lambda(s-1)}$. To get the probability that Aa is lost, we have to solve the transcendental equation $q = F(q)$. With the approximation $q = 1 - \delta$, where δ is small, and with Taylor's theorem, we get

$$\begin{aligned}
& F(1-\delta) = 1-\delta \\
\Leftrightarrow\ & F(1) - \delta F'(1) + \frac{\delta^2}{2} F''(1) - \cdots = 1 - \delta \\
\Leftrightarrow\ & 1 - \delta\lambda + \frac{\delta^2}{2}\lambda^2 - \cdots = 1 - \delta \\
\Leftrightarrow\ & \frac{\delta}{2}\lambda^2 - \delta + 1 = 0 \\
\Leftrightarrow\ & \delta = 2\frac{\lambda - 1}{\lambda^2} = 2\frac{1+h-1}{(1+h)^2} \approx 2h.
\end{aligned}$$

Therefore, the probability of survival is given by $\delta \approx 2h$ for small h.
We now take a look at the general offspring distribution. In our fixed point equation assume $q = e^{\Theta}$, then we get

$$e^{\Theta} = F(e^{\Theta}) = M(\Theta) \Leftrightarrow \Theta = \log M(\Theta) = K(\Theta).$$

Therefore by the definition of the cumulative generating function

$$\begin{aligned}
& \Theta = m\Theta + \frac{\sigma^2}{2}\Theta^2 + \dots \\
\Leftrightarrow\ & 1 = m + \frac{\sigma^2}{2}\Theta + \cdots \Rightarrow \Theta \approx 2\frac{1-m}{\sigma^2}.
\end{aligned}$$

Since m should be larger than 1, $\Theta = \frac{-2h}{\sigma^2}$ is negative and again by Taylor's theorem

$$q = e^{\Theta} = e^{\frac{-2h}{\sigma^2}} \approx 1 - \frac{2k}{\sigma^2}.$$

Theorem 2.5. *The generating function of X_n is the n^{th} iterate*

$$\underbrace{F_x \circ F_x \circ F_x \circ \cdots \circ F_x}_{n \ times}$$

of the generating function $F_x(s) = \sum_{k=0}^{\infty} p_k s^k$.

Proof. For notation reasons, be $F_0(s) = s$, $F_1(s) = F(s)$ and $F_{n+1}(s) = F(F_n(s))$. Further be $F_{(n)}$ the generating function of X_n. Therefore $F_{(0)}(s) = s$ and $F_{(1)}(s) = F(s)$. Under the condition $X_n = k$ the random variable X_{n+1} has the generating function $F(s)^k$, since the k individuals reproduce independently. Hence

$$\begin{aligned} F_{(n+1)} = E[s^{X_{n+1}}] &= \sum_{k=0}^{\infty} \underbrace{E[s^{X_{n+1}}|X_n = k)}_{\text{gen. fct. of } X_{n+1}|X_n=k} P(X_n = k) \\ &= \sum_{k=0}^{\infty} F(s)^k P(X_n = k) = F_{(n)}(F(s)), \end{aligned}$$

so by induction, $F_n = F_{(n)}$ for all n. □

Theorem 2.6. *In a Galton-Watson process with $X_0 = 1$ holds*

1. $E[X_n] = m^n$
2. $\mathrm{Var}[X_n] = \begin{cases} \frac{m^{n-1}(m^n-1)}{m-1}\sigma^2 & \text{if } m \neq 1 \\ n\sigma^2 & \text{if } m = 1. \end{cases}$

Proof. We only show a proof for the first claim. Using 2.5, we get by induction and the chain rule

$$\begin{aligned} E[X_n] = F'_{x_n}(1) &= (\underbrace{F_x \circ F_x \circ \cdots \circ F_x}_{n \text{ times}})'(1) := F'_n(1) \\ &= F'_{n-1}(F(1)) \cdot F'(1) = F'_{n-1}(1) \cdot m = m^{n-1}m = m^n. \end{aligned}$$

□

Exercise 17. *Prove the second part of the theorem above.*

2.6 Asymptotic behaviour

We look at the supercritical case $1 < m < \infty$ and define

$$Z_n := \frac{X_n}{m^n}$$

which tends to the random variable Z_∞ as n tends to infinity. Hence

$$P(Z_\infty = 0) = q = P(X_n = 0 \text{ for some } n).$$

Without proof, we claim

$$\text{Var} Z_n = \frac{\sigma^2}{m(m-1)}\Big(1 - \frac{1}{m^n}\Big)$$

and

$$\text{Var} Z_\infty = \frac{\sigma^2}{m(m-1)}.$$

Now for the critical case $m = 1$ we have $E[X_n] = 1$ for all n, but $q = 1$, hence X_n tends to 0 with probability 1. If $\delta^2 = \text{Var}[X_1] < \infty$, then

$$P(X_n > 0) = 1 - F_n(0) \sim \frac{2}{n\delta^2}$$

and

$$E[X_n | X_n > 0] = \frac{1 - E[X_n|X_n = 0]P(X_n = 0)}{P(X_n > 0)} = \frac{1}{P(X_n > 0)} \sim \frac{n\delta^2}{2}$$

as n tends to infinity. Finally we get

$$\lim_{n\to\infty} P(\frac{X_n}{n} > z | X_n > 0) = e^{-\frac{2z}{\delta^2}} \text{ for } z \geq 0.$$

In the subcritical case $m < 1$, we have

$$\lim_{n\to\infty} P(X_n = k | X_n > 0) = b_k \text{ with } b_0 = 0,$$

the limit law is conditional on survival. Define $B(s) = \sum_{k=0}^{\infty} b_k s^k$, then

$$B(F(s)) = mB(s) + 1 - m$$

and

$$1 - F_n(0) = P(X_n > 0) \sim \frac{m^2}{B'(1)}$$

as n tends to ∞. So the probability that the population is still alive increases geometrically. A proof for all those claims can be found in 2.

3 Markov chains

3.1 Definition

Definition 3.1.
A (stationary) Markov chain is a sequence $(X_n)_{n=1}^{\infty}$ of random variables with values in a countable state space (usually $\subseteq$ Z) such that

1. $P(X_{n+1} = j | X_n = i) =: p_{ij}$ is independent of n, that means it is time independent or stationary, and
2. $P(X_{n+1} = j | X_n = i, X_{n-1} = i_1, X_{n-2} = i_2, \dots, X_1 = i_n) = P(X_{n+1} = j | X_n = i)$, so every state only depends on the foregoing one.

This concept is based on the work of Andrei A. Markov, who started studying finite Markov chains in 1906. We call $p_{ij} = P(X_{n+1} = j | X_n = i)$ for $i, j \in S$ the transition probabilities. Then $P := (p_{ij})$ is a so-called transition matrix with the following properties.

1. $p_{ij} \geq 0 \ \forall i, j \in S$ and
2. $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$ and therefore $P \cdot \mathbf{1} = \mathbf{1}$,

where $\mathbf{1} = (1, 1, \dots, 1)^t$.

A matrix with such properties is called a *stochastic matrix*. Assume that an initial distribution for X_0 is given by $P(X_0 = k) = a_k$, then

$$P(X_0 = i \wedge X_1 = j) = P(X_0 = i) P(X_1 = j | X_0 = i) = a_i p_{ij}$$

and with this equation we compute the probability of a sample sequence as

$$P(X_0 = i_0 \wedge X_1 = i_1 \wedge \cdots \wedge X_n = i_n) = a_{i_0} \cdot p_{i_0 i_1} \cdot p_{i_1 i_2} \cdot \cdots \cdot p_{i_{n-1} i_n}.$$

3.2 Examples of Markov chains

A.
The random walk with absorbing barriers has the state space $S = \{0, 1, \dots, N\}$

and the transition matrix is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & q & 0 & p & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ & & & & & & \\ & & & & & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

where $p_{i,i+1} = p$ and $p_{i,i-1} = q$ for $0 \neq i \neq N$. For the boundary we have $p_{0i} = \delta_{0i}$ and $p_{Ni} = \delta_{Ni}$.

In fact, all random walks from section 1 are Markov chains, but without boundaries, the state space is infinite.

B.
The random walk with reflecting boundaries has the state space $S = \{1, 2, \ldots, N\}$ and the corresponding transition matrix is given by

$$\begin{pmatrix} q & p & 0 & 0 & 0 & & \cdots & 0 \\ q & 0 & p & 0 & 0 & & \cdots & 0 \\ 0 & q & 0 & p & 0 & & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & & & & \\ & & & & & 0 & p & 0 \\ & & & & & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & q & p \end{pmatrix}$$

C.
A cyclic random walk is a random walk where we can move from the state 1 to the state N and vice versa. The transition matrix is given by

$$\begin{pmatrix} 0 & p & 0 & 0 & 0 & \cdots & 0 & q \\ q & 0 & p & 0 & 0 & & \cdots & 0 \\ 0 & q & 0 & p & 0 & & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & & & & \\ & & & & & 0 & p & 0 \\ & & & & & q & 0 & p \\ p & 0 & 0 & 0 & \cdots & 0 & q & 0 \end{pmatrix}$$

D.

The Wright-Fisher model is a model from the field of genetics, the state space is $S = \{0, 1, \cdots, 2N\}$ and the probabilities are given by

$$p_{ij} = \binom{2N}{j} (\frac{i}{2N})^j (1 - \frac{i}{2N})^{2N-j}.$$

Therefore for a given N and i we have a binomial distribution with parameters $2N$ and $p = \frac{i}{2N}$. This describes a population of N individuals, on some gene locus we have two alleles, A and a. Therefore we look at $2N$ genes in total. If i is number of A's and the frequency of A is $i/2N$. Now each new generation chooses $2N$ genes randomly out of the pool of gametes. The states 0 and $2N$ are absorbing, therefore the transition matrix has the form

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ + & + & + & \cdots & + \\ \vdots & & & & \vdots \\ + & + & + & \cdots & + \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

where + denotes a positive probability.

E.

The Ehrenfest model describes two containers A and B with $N_A + N_B = N$ molecules in total. At each step we choose one molecule and move it to the other container. The random variable X_n describes the number of molecules which remained in container A, therefore the state space is again $S = \{0, 1, \ldots, N\}$. The transition probabilities are given by $p_{i,i+1} = \frac{N-i}{N}$ and $p_{i,i-1} = \frac{i}{N}$. It is pretty similar to a random walk, but the probabilities now depend on the state.

F.

The Bernoulli–Laplace model is about a compressible fluid in two containers A and B with N particles in each of them. One half of the particles is blue, the other half is white. Let X_n be the number of white particles in A, then if $X_n = k$ there are $N - k$ blue particles in A and therefore in B k blue ones and $N - k$ white ones. In every step we pick one particle from A and one from B and we swap them. This leads to the transition probabilities

$$\begin{aligned}
p_{i,i-1} &= \frac{i}{N}\frac{i}{N} = \Big(\frac{i}{N}\Big)^2, \\
p_{i,i+1} &= \Big(\frac{N-i}{N}\Big)^2 \text{ and} \\
p_{i,i} &= 2\frac{i}{N}\frac{N-i}{N}.
\end{aligned}$$

G.
In the Galton-Watson process from chapter 2, the state space is given by $S = \{0, 1, 2, \dots\}$ and 0 is absorbing, that is $p_{00} = 1$ and $p_{0j} = 0$ for $j \geq 1$.

Exercise 18. *In a Galton-Watson process with probabilities $P(X_1 = k|X_0 = 1) = p_k$, find $P(X_1 = k|X_0 = 2)$, and the transition probabilities $p_{ij} = P(X_{n+1} = j|X_n = i)$.*

3.3 Transition probabilities

A short calculation shows an important benefit of transition matrices,

$$\begin{aligned}
P(X_{n+2} = j|X_n = i) &= \sum_{k\in S} P(X_{n+2} = j \wedge X_{n+1} = k|X_n = i) \\
&= \sum_{k\in S} P(X_{n+2} = j|X_{n+1} = k \wedge X_n = i) \cdot P(X_{n+1} = k|X_{n+1} = i) \\
&= \sum_{k\in S} P(X_{n+2} = j|X_{n+1} = k) \cdot P(X_{n+1} = k|X_{n+1} = i) \\
&= \sum_{k\in S} p_{ik}p_{kj} := p_{ij}^{(2)}.
\end{aligned}$$

Here, $p_{ij}^{(2)}$ denotes the (i, j) entry of the matrix P^2. Notice that even for a countable infinite state space S, the sum over all those products converges, since

$$\sum_{k\in S} p_{ik}p_{kj} \leq \sum_{k\in S} p_{ik} = 1.$$

By induction we get $P(X_{n+m} = j|X_n = i) = p_{ij}^{(m)}$, the corresponding entry of P^m. Further it is clear that P^m is again a stochastic matrix, since $P^2\mathbf{1} = P \cdot P\mathbf{1} = P\mathbf{1} = \mathbf{1}$ and therefore again by induction $P^m\mathbf{1} = \mathbf{1}$.

3.4 Invariant distributions

We are now looking at the probability vector $u^{(n)}$ with entries $u_i^{(n)} = P(X_n = i)$ for $i \in S$, therefore it corresponds to the probability distribution at time n.

$$\triangle(S) := \{(u_i)_{i\in S} : u_i \geq 0 \ \forall i \in S, \ \sum_{i\in S} u_i = 1\}$$

is the so-called probability simplex over S. (since all the vectors together build a $|S|-$dimensional simplex in R^n). With the relation

$$P(X_{n+1} = j) = \sum_i P(X_n = i) p_{ij}$$

we get $u^{(n+1)} = u^{(n)}P$. Now (a row vector) $u \in \triangle(S)$ is called a *stationary* or *invariant probability distribution* for the Markov chain if $u = uP$, which means that u is a left eigenvector of P to the eigenvalue 1. We will show the existence of such a vector later. For the examples A,D and G, the state 0 is absorbing, i.e. $p_{00} = 1$ and $p_{0j} = 0$ for $j \geq 1$, therefore the vector $u = (1, 0, \ldots, 0)$ satisfies $uP = u$. In fact, we can show (Exercise 20) that for A and D, for $\alpha \in [0, 1]$ the vectors $u = (\alpha, 0, \ldots, 0, 1 - \alpha)$ gives all the stationary probability distributions. In example C, we get

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^T \cdot \begin{pmatrix} 0 & p & 0 & 0 & 0 & \cdots & 0 & q \\ q & 0 & p & 0 & 0 & & \cdots & 0 \\ 0 & q & 0 & p & 0 & & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ & & & & & & & \\ & & & & & 0 & p & 0 \\ & & & & & q & 0 & p \\ p & 0 & 0 & 0 & \cdots & 0 & q & 0 \end{pmatrix} = \begin{pmatrix} p+q \\ p+q \\ p+q \\ p+q \\ p+q \\ p+q \\ \vdots \\ p+q \end{pmatrix}^T = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}^T$$

therefore $u = \frac{1}{N}\mathbf{1}$ is a normalized vector for a stationary distribution.

3.5 Ergodic theorem for primitive Markov chains

We now show the existence of a stationary distribution in an important special case. A stochastic matrix is called *primitive* if

$$\exists M > 0 \ : \ \forall i, j \ : \ p_{ij}^{(M)} > 0.$$

Theorem 3.1 (Ergodic theorem)**.** *If P is a primitive stochastic matrix belonging to a Markov chain over a finite state space S with $|S| = N$, then there is a unique stationary probability distribution $u \in \triangle(S)$. Furthermore $u_j > 0$ for every $j \in S$ and $p_{ij}^{(n)}$ tends to u_j for every i and j as $n \to \infty$. Moreover, for every initial distribution $P(X_0 = i)$ the probability $P(X_n = j)$ tends to u_j.*

Examples for a primitive matrix are given by C (as long as N is odd), E and F. In A and D, the state 0 is absorbing and we get $p_{00}^{(n)} = 1$ and $p_{0i}^{(n)} = 0$ for $i > 0$. Therefore the matrices are not primitive.

Proof. 1) We first prove the theorem for a positive matrix, i.e., $p_{ij} > 0$ for all i, j, i.e., $M = 1$. Define $\delta := \min_{i,j} p_{ij}$ and since the row sum cannot exceed 1, we assume $0 < \delta < \frac{1}{2}$. Now we fix j and define

$$M_n := \max_i p_{ij}^{(n)} \text{ and } m_n := \min_i p_{ij}^{(n)}.$$

Our claim is now that M_n is a decreasing and m_n an increasing sequence and that the difference $M_n - m_n$ tends to 0. Monotonicity follows from

$$M_{n+1} = \max_i p_{ij}^{(n+1)} = \max_i \sum_l p_{il} p_{lj}^{(n)} \le \max_i \sum_l p_{il} M_n = M_n$$

and similar

$$m_{n+1} = \min_i p_{ij}^{(n+1)} = \min_i \sum_l p_{il} p_{lj}^{(n)} \ge \min_i \sum_l p_{il} m_n = m_n.$$

Let $k = k(n)$ be such that $m_n = p_{kj}^{(n)} \le p_{lj}^{(n)}$ for all l. Then

$$\begin{aligned} M_{n+1} &= \max_i \Big[p_{ik} p_{kj}^{(n)} + \sum_{l \ne k} p_{il} p_{lj}^{(n)} \Big] \\ &\le \max_i \Big[p_{ik} m_n + \sum_{l \ne k} p_{il} M_n \Big] \\ &= \max[M_n - (M_n - m_n) p_{ik} \Big] \\ &\le M_n - (M_n - m_n)\delta < M_n, \end{aligned}$$

and one can show just as well

$$M_{n+1} \ge m_n + (M_n - m_n)\delta > m_n.$$

Now

$$M_{n+1} - m_{n+1} \leq (M_n - m_n)(1 - 2\delta),$$

therefore the distance $M_n - m_n$ tends to 0. Define $u_j := \lim_{n\to\infty} M_n = \lim_{n\to\infty} m_n$ and since $m_n \leq p_{ij}^{(n)} \leq M_n$, every value in the j^{th} column tends to u_j. Since

$$\sum_j u_j = \lim_{n\to\infty} \sum_j p_{ij}^{(n)} = \lim_{n\to\infty} 1 = 1,$$

u_j is a probability vector and we also get

$$P^n \to U := \begin{pmatrix} u_1 & u_2 & \dots & u_N \\ \vdots & \vdots & & \vdots \\ & & & \\ u_1 & u_2 & \dots & u_N \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot \begin{pmatrix} u_1 & u_2 & \dots & u_N \end{pmatrix} = \mathbf{1} \cdot u$$

and therefore

$$P^n \to U \Rightarrow P^{n+1} \to UP = U \Rightarrow uP = u.$$

2) Now we look at the case that P is primitive but has entries with value 0, therefore $M > 1$. Define $Q := P^M$ then we know from the first case that Q has a stationary vector u and $q_{ij}^{(n)} \to u_j$ as n tends to infinity. Write $n = Ml + r$ for some r between 0 and M. Now for every $\epsilon > 0$ exists some $N(\epsilon)$ such that $|q_{ij}^{(n)} - u_j| < \epsilon$ for any l larger than $N(\epsilon)$. Hence

$$p_{ij}^{(n)} = p_{ij}^{(Ml+r)} = \sum_k p_{ik}^{(r)} p_{kj}^{(Ml)} = \sum_k p_{ik}^{(r)} q_{kj}^{(l)} \leq \sum_k p_{ik}^{(r)} (u_j + \epsilon) = u_j + \epsilon$$

and

$$p_{ij}^{(n)} = p_{ij}^{(Ml+r)} = \sum_k p_{ik}^{(r)} p_{kj}^{(Ml)} = \sum_k p_{ik}^{(r)} q_{kj}^{(l)} \geq \sum_k p_{ik}^{(r)} (u_j - \epsilon) = u_j - \epsilon$$

and therefore $|p_{ij}^{(n)} - u_j| \leq \epsilon$ for $n > MN(\epsilon)$. The remaining part is the same as in the first case.

3) For the uniqueness, we suppose there exists another $\tilde{u} \in \triangle(S)$ such that $\tilde{u}P = \tilde{u}$. But then we get $\tilde{u}P^n = \tilde{u}$ and therefore

$$\tilde{u} = \tilde{u}P^n \to \tilde{u}(\mathbf{1}u) = (\tilde{u}\mathbf{1})u = \sum_i (\tilde{u}_i \cdot 1)u = 1u = u.$$

□

3.6 Examples for stationary distributions

As already mentioned, the stationary vector for example C is given by $u = \frac{1}{N}(1, \ldots, 1)$.

Exercise 19. *Show for the Ehrenfest model (example E) the stationary distribution is given by $u_j = \binom{N}{j}\frac{1}{2^N}$. Hint: We only have to check that*

$$u_j = u_{j-1}p_{j-1,j} + u_{j+1}p_{j+1,j}.$$

Exercise 20. *In A and D, show that all stationary vectors are given by $(\alpha, 0, \ldots, 0, 1-\alpha)$.*

Exercise 21. *Find u for B.*

Exercise 22. *Prove that for F the stationary distribution is given by $u_k = \binom{N}{k}^2 \cdot constant$.*

3.7 Birth-death chains

Now assume the transition probabilities are given by $p_{ij} = 0$ if $|i-j| > 1$ and the state space is $S = \{0, \ldots, N\}$. If we want to find u with $uP = u$, we solve a simple system of equations,

$$\begin{aligned}
u_0 &= u_0 p_{00} + u_1 p_{10} \\
u_1 &= u_0 p_{01} + u_1 p_{11} + u_2 p_{21} \\
&\vdots \\
u_{N-1} &= u_{N-2}p_{N-2,N-1} + u_{N-1}p_{N-1,N-1} + u_N p_{N,N-1} \\
u_N &= u_{N-1}p_{N-1,N} + u_N p_{NN}.
\end{aligned}$$

We assume $p_{i-1,i} > 0$ and $p_{i,i-1} > 0$. Remembering the notation of random walks, we define $p_k := p_{k,k+1}$ and $q_k := p_{k,k-1}$. Now we simplify the system of equations by

$$\begin{aligned}
u_1 &= u_0\frac{1-p_{00}}{p_{10}} = u_0\frac{p_{01}}{p_{10}} = u_0\frac{p_0}{q_1} \\
u_2 &= u_1\frac{1-p_{10}-p_{11}}{p_{21}} = u_1\frac{p_{12}}{p_{21}} = u_1\frac{p_1}{q_2} = u_0\frac{p_0p_1}{q_1q_2} \\
&\vdots \text{ by induction} \\
u_k &= u_{k-1}\frac{p_{k-1}}{q_k} = u_0\frac{p_{k-1}p_{k-2}\cdot\dots\cdot p_0}{q_kq_{k-1}\cdot\dots\cdot q_1},
\end{aligned}$$

and since

$$\sum_k u_k = 1 = u_0\Big(1 + \frac{p_0}{q_1} + \frac{p_0 p_1}{q_1 q_2} + \cdots + \frac{p_{N-1}p_{N-2}\cdot\dots\cdot p_0}{q_N q_{N-1}\cdot\dots\cdot q_1}\Big),$$

u_0 is given as the reciprocal value of the last sum. The concept of birth-death chains covers the examples B (for which $p_0 = 0$ and $q_N = 0$),E and F.

3.8 Reversible Markov chains

A Markov chain with transition matrix P is called *reversible*, if there exists a vector π such that $\pi_i > 0$ for all i and

$$\pi_i p_{ij} = \pi_j p_{ji}$$

for every i and j. Then we automatically get

$$\sum_i \pi_i \underbrace{P(X_{n+1} = j | X_n = i)}_{p_{ij}} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

and therefore $\pi P = \pi$. Hence π is in $\triangle(S)$ if we normalize it and it is a stationary distribution for P. C with $p = q$ is a special case for a reversible Markov chain since $P = P^T$ in this example (and therefore $p_{ij} = p_{ji}$).

Exercise 23. *Show that birth-death chains with all $p_i, q_i > 0$ are reversible.*

But why are those Markov chains called reversible? Suppose we start from a stationary distribution π with $P(X_0 = i) = \pi_i$, then we have $P(X_n = i) = \pi_i$ for all n. Therefore

$$P(X_n = i \wedge X_{n+1} = j) = P(X_n = j \wedge X_{n+1} = i),$$

, so it does not change the probability of a certain chain if we see it the other way round. The concept of reversible Markov chains was introduced by the Russian mathematician Andrei Nikolajewitsch Kolmogorow in 1935, who also gives the following criterion.

Theorem 3.2. *A primitive matrix P describes a reversible Markov chain if and only if*

$$p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} p_{i_n i_1} = p_{i_1 i_n} p_{i_n i_{n-1}} \cdots p_{i_3 i_2} p_{i_2 i_1}$$

for all sequences $(i_1, i_2, \dots, i_n)$ in S and for every length n.

Exercise 24. *Show the first ($\Rightarrow$) direction of the proof.*

Proof. $\Leftarrow$) Fix $i = i_1$ and $j = i_n$. Then

$$p_{ii_2}p_{i_2i_3}\cdots p_{i_{n-1}j}p_{ji} = p_{ij}p_{ji_{n-1}}\cdots p_{i_3i_2}p_{i_2i}$$

holds. Summing over all states $i_2, i_3, \dots, i_{n-1}$, we get

$$p_{ij}^{(n-1)}p_{ji} = p_{ij}p_{ji}^{(n-1)},$$

where the left side tends to $u_j p_{ji}$ and the right side to $p_{ij}u_i$ for $n \to \infty$ and therefore

$$u_j p_{ji} = p_{ij}u_i.$$

□

As an example, we look at a connected graph with N vertices. Each link from i to j gets a weight and since we want an undirected graph, we assume $w_{ij} = w_{ji}$. Loops are allowed, i.e., $w_{ii} \geq 0$. We define a Markov chain via $p_{ij} := \frac{w_{ij}}{\sum_k w_{ik}}$. Now we can show that this chain is reversible. If we choose

$$\pi_i = \frac{\sum_k w_{ik}}{\sum_{l,k} w_{lk}}$$

we get

$$\pi_i p_{ij} = \frac{w_{ij}}{\sum_{l,k} w_{lk}} = \frac{w_{ji}}{\sum_{l,k} w_{lk}} = \pi_j p_{ji}.$$

3.9 The Markov chain tree formula

A Markov chain $P = (p_{ij})$ is called *irreducible* if for every i and j there is some sequence $i = i_1, i_2, \dots, i_n = j$ such that all $p_{i_k i_{k+1}}$ are positive. This is equivalent to the existence of some n such that $p_{ij}^{(n)} > 0$. Now we interpret the chain as a weighted directed graph. Then every state of the chain accords to a node and the probability p_{ij} assigns a weight to the (directed) edge from i to j. We will call the set of all directed edges $E := \{(i,j) : p_{ij} > 0\}$. A subset $t \subseteq E$ is called *directed spanning tree*, if

- there is at most one arrow out of every node,
- there are no cycles and

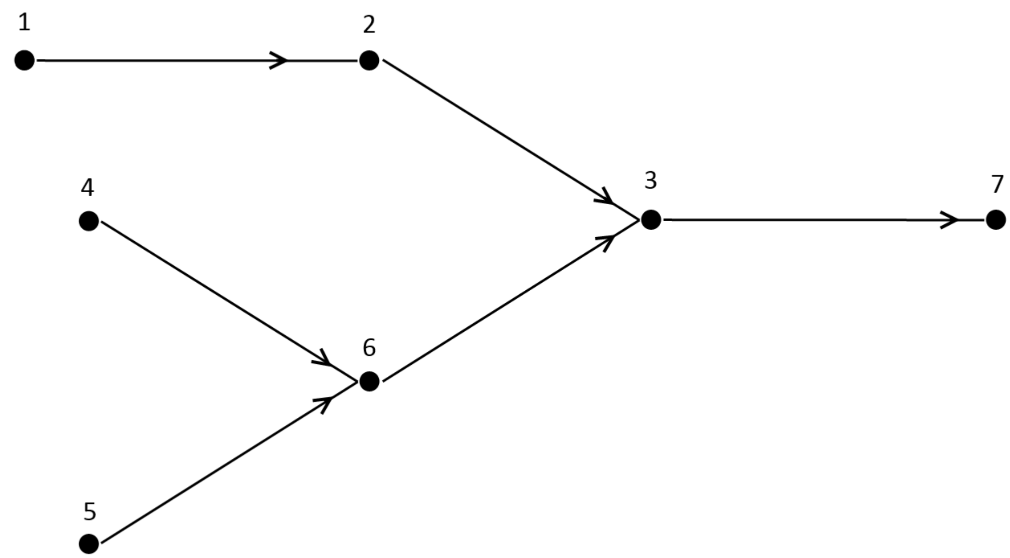


Figure 5: A directed spanning tree with root in 7

- t has maximal cardinality.

Now it is clear that since P is irreducible, the cardinality of t is $n-1$ and there exists one node with out-degree 0 called the *root* of t. We define

- the weight of t as $w(t) := \prod_{(i,j)\in t} p_{ij}$,
- the set of all directed spanning trees as $\mathfrak{T}$,
- the set of all directed spanning trees with root j as $\mathfrak{T}_j$,
- the weight of all trees with root in j as $w_j := \sum_{t\in\mathfrak{T}_j} w(t)$ and
- the total weight of $\mathfrak{T}$ as $w(\mathfrak{T}) := \sum_{t\in\mathfrak{T}} w(t)$.

The following result has been traced back to Gustav Robert Kirchhoff around 1850.

Theorem 3.3. *For finite irreducible Markov chains the stationary distribution is given by*

$$u_j = \frac{w_j}{w(\mathfrak{T})}.$$

Proof. The idea of the proof is to take a directed spanning tree with root in k and add one more arrow from k to somewhere else. For fixed k consider G_k as the set of all directed graphs on S such that

- each state i has a unique arrow out of it to some j and
- there is a unique closed loop which contains k.

Now if g is in G_k, the weight is defined as $w(g) := \prod_{(i,j)\in g} p_{ij}$ and now there are two ways to compute $w(G_k)$,

$$w(G_k) := \sum_{g\in G_k} w(g) = \begin{cases} \sum_{i\neq k} \left(\sum_{t\in\mathfrak{T}_i} w(t)\right) p_{ik} = \sum_{i\neq k} w_i p_{ik} \\ \sum_{j\neq k} \left(\sum_{t\in\mathfrak{T}_k} w(t)\right) p_{kj} = \sum_{j\neq k} w_k p_{kj}. \end{cases}$$

Adding $w_k p_{kk}$ to both outcomes, we get

$$\sum_i w_i p_{ik} = \sum_j w_k p_{kj} = w_k \sum_j p_{kj} = w_k$$

and therefore $wP = w$. If we normalize the vector we get the statement. $\square$

Remark that in the case of birth-death chains there are only edges from the state k to its neighbours.

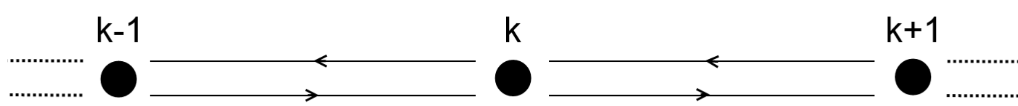
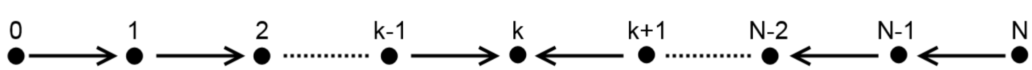


Figure 6: The graph of a birth-death chain

Therefore for every k the set of all spanning trees with root in k contains only one element, it is given by



So the weight of $\mathfrak{T}_\mathfrak{k}$ is

$$w_k = p_0 p_1 \cdots p_{k-1} q_{k+1} \cdots q_N$$

and hence

$$u_k = \frac{w_k}{w(\mathfrak{T})} = \frac{p_0 p_1 \cdots p_{k-1} q_{k+1} \cdots q_N}{p_0 p_1 \cdots p_{N-1} + \cdots + q_1 q_2 \cdots q_n} = \frac{\frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k}}{\frac{p_0 \cdots p_{N-1}}{q_1 \cdots q_N} + \cdots + 1}$$

which is the formula we already found in section 3.7.

Exercise 25. *Using the spanning trees, find the stationary distribution for $|S| = 3$ and $P > 0$.*

3.10 Mean recurrence time

Consider a finite irreducible Markov chain and define the mean time to go from i to j as m_{ij} and thus the mean recurrence time as m_{ii}. Then we get the relation

$$m_{ij} = p_{ij} + \sum_{k \neq j} p_{ik}(1 + m_{kj}) = \sum_k p_{ik} + \sum_{k \neq j} p_{ik} m_{kj} = 1 + \sum_{k \neq j} p_{ik} m_{kj}.$$

So we get $N \cdot N$ linear equations for $N \cdot N$ unknown m_{ij}.

Exercise 26. *Show that there is a unique solution for this system.*

Now let $uP = u$ be a stationary distribution, then

$$\sum_i u_i m_{ij} = \underbrace{\sum_i u_i}_{=1} + \sum_i \sum_{k \neq j} u_i p_{ik} m_{kj} = 1 + \sum_{k \neq j} m_{kj} \underbrace{\sum_i u_i p_{ik}}_{u_k} = 1 + \sum_{k \neq j} u_k m_{kj}$$

and adding $u_j m_{jj}$ to both sides gives $u_j m_{jj} = 1$ and hence the mean recurrence time is given by $m_{jj} = \frac{1}{u_j}$.

In example E, the stationary distribution is given via $u_j = \binom{N}{j} \frac{1}{2^N}$. For $j = 0$ (which means all molecules are in the first container) the mean recurrence time is given as $m_{00} = 2^N$, which is a hell of a big number if we consider one mole to have $6 \cdot 10^{23}$ molecules. For the most likely state $N/2$ we get

$$u_{N/2} = \binom{N}{\frac{N}{2}} \frac{1}{2^N} \sim \sqrt{\frac{2}{\pi N}}$$

and thus for one mole

$$m_{N/2,N/2} = \sqrt{\frac{\pi N}{2}} \sim 10^{12},$$

which is a long time but nothing compared to the other one.

3.11 Recurrence vs. transience

In this section our state space S shall be finite or countable and we fix $j \in S$ and assume $P(X_0 = j) = 1$. This has the advantage that we can disregard exactly this condition in the probabilities and simply denote $P(X_n = j) = p_{jj}^{(n)}$. The following lemma of Borel and Cantelli will help us in the next proof.

Lemma 3.4. *Let $(A_n)_{n=1}^{\infty}$ denote a sequence of events. Then the following statements hold*

1. $\sum_{k=1}^{\infty} P(A_k) < \infty \Rightarrow P(\textit{Infinitely many } A_k \textit{ occur}) = 0$

2. *If* $\sum_{k=1}^{\infty} P(A_k) = \infty$ *and the* A_k *are independent, then* $P(\textit{Infinitely many } A_k \textit{ occur}) = 1$.

Exercise 27. *Prove the first part of the Lemma of Borel-Cantelli.*

Theorem 3.5. 1. *The following chain of equivalences holds*

$$\begin{aligned} & P(\exists n \geq 1 \ : \ X_n = j) = 1 \\ \Leftrightarrow \ & P(X_n = j \textit{ for infinitely many } n) = 1 \\ \Leftrightarrow \ & \sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty. \end{aligned}$$

2. *Further we have*

$$\begin{aligned} & P(\exists n \geq 1 \ : \ X_n = j) < 1 \\ \Leftrightarrow \ & P(X_n = j \textit{ for infinitely many } n) = 0 \\ \Leftrightarrow \ & \sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty. \end{aligned}$$

Definition 3.2.
A state which fulfills one of the equivalences of 1. in the previous theorem is called *recurrent*, if it fulfills one of the equivalences of 2., it is called *transient*.

Proof. Recall $X_0 = j$, then we define

- $F_0 := \{X_n \neq j \text{ for } n = 1, 2, \ldots\}$, the set of all sequences where there is no return to j,

- $R := \{\exists n \geq 1 \ : \ X_n = j\}$, the set of all sequences where there is some return (remark $F_0 \dot\cup R = \Omega = S^{\mathbb{N}}$, the space of all sequences) and
- $F_n := \{X_n = j, \ X_k \neq j \ \forall j > n\}$, the set of all sequences with last visit to j at time n.

Then we get

$$\begin{aligned} P(F_n) &= P(X_n = j) \cdot P(X_{n+1} \neq j \wedge X_{n+2} \neq j \wedge \ldots | X_n = j) \\ &= P(X_n = j) \cdot P(X_1 \neq j \wedge X_2 \neq j \wedge \ldots | X_0 = j) \\ &= p_{jj}^{(n)} P(F_0). \end{aligned}$$

Since

$$\dot{\bigcup_{n=0}^{\infty}} F_n = \Omega \setminus \{X_n = j \text{ for infinitely many n } \}$$

holds, we get the equation

$$1 - P(X_n = j \text{ for infinitely many n }) = \sum_{n=0}^{\infty} P(F_n).$$

Remark that $p_{jj}^{(0)} = 1$. Now since $P(F_0) = 1 - P(R)$, we can conclude for the actual proof

1. If $P(R) = 1$ then $P(F_0) = 0$ and hence $P(X_n = j \text{ for infinitely many n}) = 1$. Then the Lemma of Borel and Cantelli states that $\sum_{n=0}^{\infty} P(X_n = j) = \infty$ and therefore $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$.

2. If $P(R) < 1$ then $P(F_0) > 0$ and hence $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$. Then the Lemma of Borel and Cantelli states that $P(X_n = j \text{ for infinitely many n}) = 0$.

$\square$

Remark that from the ergodic theorem for finite primitive Markov chains we already know $p_{jj}^n \to u_j > 0$. Therefore the sum $\sum_{n=1}^{\infty} p_{jj}^{(n)}$ cannot be finite, so every state is recurrent. Some simple examples for the application of the theorem: above are

- The asymmetric random walk on $\mathbb{Z}$ gives $p_{00}^{(2n+1)} = 0$ and

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n = \binom{2n}{n} \frac{1}{2^n} (4pq)^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}.$$

We can rewrite the numerator as

$$4pq = (p+q)^2 - (p-q)^2 = 1 - (p-q)^2 \begin{cases} = 1 \text{ if } p = q = \frac{1}{2} \\ < 1 := \alpha \text{ else.} \end{cases}$$

So we get for the asymmetric case

$$4pq < 1 \Rightarrow \sum_{n=1}^{\infty} p_{00}^{(2n)} \sim \frac{1}{\pi} \sum_{n=1}^{infty} \frac{\alpha^n}{\sqrt{n}} < \frac{1}{\pi} \sum_{n=1}^{\infty} \alpha^n < \infty$$

which means, as we already know, 0 is transient. For the symmetric case applies

$$p = q = \frac{1}{2} \Rightarrow \sum_{n=1}^{\infty} p_{00}^{(2n)} \sim \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty,$$

therefore 0 is recurrent.

- In example A, 0 and N are absorbing states. Therefore $p_{00} = 1$ and $p_{00}^{(n)} = 1$ for all n. Hence $\sum_n p_{00}^{(n)} = \infty$ and thus 0 is recurrent.

Exercise 28. *Show for the second example that all other states* $1, 2, \ldots, N-1$ *are transient.*

3.12 The renewal equation

We denote the probability of starting in the state i and reaching j for the first time after n steps with $f_{ij}^{(n)}$. Then

$$m_{ij} := \sum_{n=0}^{\infty} n f_{ij}^{(n)}$$

gives the mean arrival time (and as in the previous section, m_{jj} is the mean recurrence time).

Theorem 3.6. *The renewal equation is given by*

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{ij}^{(n-k)}.$$

Proof. Being at the state j after n steps with starting point i is the same as if we are in j after k steps for the first time and then again after $n-k$ steps. Summing over all possible k gives the equation. □

If we identify the probability of ever returning to some state i with $\sum_{n=1}^{\infty} f_{ii}^{(n)}$, then we get

- i is recurrent if and only if $\sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$ and
- i is transient if and only if $\sum_{n=1}^{\infty} f_{ii}^{(n)} < 1$.

The following theorem is called the renewal theorem.

Theorem 3.7. *Let (r_n) and (f_n) be any sequences. Suppose*

1. $f_n \geq 0$ *and* $\sum_{n=1}^{\infty} f_n = 1$,
2. $r_n = r_0 f_n + r_1 f_{n-1} + \cdots + r_{n-1} f_1$ *and*
3. *the set* $Q = \{n \geq 1 \ : \ f_n > 0\}$ *has greatest common divisor 1,*

then

$$r_n \to \frac{1}{\sum_{k=1}^{\infty} k f_k} \text{ for } n \to \infty,$$

where r_n is 0 if the sum is infinite.

A possible application is for a fixed j, choose $r_n = p_{jj}^{(n)}$ and $f_n = f_{jj}^{(n)}$, then

$$p_{jj}^{(n)} \to \frac{1}{\sum_{k=1}^{\infty} k f_{jj}^{(k)}} = \frac{1}{m_{jj}} \to u_j.$$

This looks like the statement of the ergodic theorem but we are not longer constricted to finite state spaces. The proof of the renewal theorem can be found in [1] or [3].

3.13 Positive vs. null-recurrence

We start this chapter with a refinement of the concept of recurrence.

Definition 3.3.
A state i is called

1. *positive recurrent* if and only if the mean recurrence time $m_{ii} = \sum_{k=1}^{\infty} k f_{ii}^{(k)}$ is finite. Then $p_{ii}^{(n)} \to \frac{1}{m_{ii}} > 0$.

2. *null-recurrent* if and only if the mean recurrence time is infinity and i is recurrent. Then $p_{ii}^{(n)} \to \frac{1}{m_{ii}} = 0$.

Remark that if i is transient, we have

$$\sum_{k=1}^{\infty} f_{ii}^{(k)} < 1 \Leftrightarrow \sum_{k=1}^{\infty} p_{ii}^{(k)} < \infty \Rightarrow p_{ii}^{(n)} \to 0 \quad (n \to \infty).$$

An example for a null-recurrent state can be found in the random walk on $\mathbb{Z}$. If we look at the state 0 (which can only be reached in an even number of steps), we get

$$m_{00} = \sum_{k=1}^{\infty} 2k f_{00}^{(k)} = \infty,$$

as we already showed in 1.10. Since we also showed that 0 is recurrent, we know that it is null-recurrent.

Exercise 29. *Compute the sum* $\sum_{k=1}^{\infty} 2k f_{00}^{(k)}$ *using the explicit formula given in 1.5.*

Another example is given by

H.
Consider a Markov chain where $S = \mathbb{N}$. In every step we either reach the next greater number (with probability p_i) or we fall back to 1 (with probability q_i. Therefore the matrix is given by

$$P = \begin{pmatrix} q_1 & p_1 & 0 & 0 & 0 & \cdots \\ q_2 & 0 & p_2 & 0 & 0 & \cdots \\ q_3 & 0 & 0 & p_3 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

where of course $p_i + q_i = 1$. We can compute the probability of first return at time n via

$$f_{11}^{(n)} = p_1 p_2 \cdots p_{n-1} q_n.$$

Furthermore if we choose any sequence $(f_n)_{n=1}^{\infty}$ with $f_n \in [0,1]$ for all n and $\sum_{n=1}^{\infty} f_n \leq 1$, we can construct p_i and q_i such that $f_{11}^{(n)} = f_n$. We choose

$$\begin{aligned}
&f_1 = q_1 \ \Rightarrow p_1 = 1 - f_1\\
&f_2 = p_1 q_2 \Rightarrow q_2 = \frac{f_2}{1 - f_1} \Rightarrow p_2 = \frac{1 - f_1 - f_2}{1 - f_1}\\
&f_3 = p_1 p_2 q_3 \Rightarrow q_3 = \frac{f_3}{1 - f_1 - f_2} \Rightarrow p_2 = \frac{1 - f_1 - f_2 - f_3}{1 - f_1 - f_2}\\
&\qquad \vdots\\
&f_n = p_1 \cdots p_{n-1} q_n \Rightarrow q_n = \frac{f_n}{1 - f_1 - \cdots - f_{n-1}} \Rightarrow p_n = \frac{1 - f_1 - \cdots - f_n}{1 - f_1 - \cdots - f_{n-1}}.
\end{aligned}$$

If i is recurrent we get the condition

$$\begin{aligned}
1 &= \sum_{n=1}^{\infty} f_{11}^{(n)}\\
&= q_1 + p_1 q_2 + p_1 p_2 q_3 + \dots\\
&= 1 - p_1 + p_1(1 - p_2) + p_1 p_2 (1 - p_3) + \dots\\
&= 1 - \prod_{n=1}^{\infty} p_n.
\end{aligned}$$

If we use the logarithm on the last product, we see that it becomes 0 if and only if

$$\sum_{n=1}^{\infty} q_n = \infty.$$

If i is positive recurrent we get the condition

$$\begin{aligned}
\infty &> \sum_{n=1}^{\infty} n f_{11}^{(n)}\\
&= 1 - p_1 + 2p_1(1 - p_2) + 3p_1 p_2 (1 - p_3) + \dots\\
&= 1 + \sum_{n=1}^{\infty} \prod_{k=1}^{n} p_n
\end{aligned}$$

which is certainly stronger.

3.14 Structure of Markov chains

Let the state space S be finite or countable. Then we define

Definition 3.4.
Two states i and j fulfill the relation $\curvearrowright$ if there is a $n \geq 0$ such that $p_{ij}^{(n)} > 0$, i.e. the state j can be reached from i. Note that since $p_{ii}^{(0)} = 1$ we always have $i \curvearrowright i$. If i and j communicate, i.e. $i \curvearrowright j$ and $j \curvearrowright i$ we simply write $i \curvearrowleft\!\!\curvearrowright j$.

Exercise 30. *Show that* $\curvearrowleft\!\!\curvearrowright$ *is a equivalence relation.*

Theorem 3.8. *Let i and j be states with $i \curvearrowleft\!\!\curvearrowright j$. Then*

1. *i recurrent $\Leftrightarrow$ j recurrent,*
2. *i transient $\Leftrightarrow$ j transient and*
3. *i null-recurrent $\Leftrightarrow$ j null-recurrent.*

Proof. If $i \curvearrowleft\!\!\curvearrowright j$ then there is some $r \geq 0$ such that $\alpha := p_{ij}^{(r)} > 0$ and some $s \geq 0$ such that $\beta := p_{ji}^{(s)} > 0$. Then

$$p_{jj}^{r+n+s} \geq p_{ij}^{(r)} p_{ii}^{(n)} p_{ji}^{(s)} = \alpha\beta p_{ii}^{(n)}.$$

From this inequality we get :

- i is recurrent $\Rightarrow \sum_n p_{ii}^{(n)} = \infty \Rightarrow \sum_n p_{jj}^{(n)} = \infty \Rightarrow j$ is recurrent.
- j is transient $\Rightarrow \sum_n p_{ii}^{(n)} < \infty \Rightarrow \sum_n p_{jj}^{(n)} < \infty \Rightarrow i$ is transient.
- j is null-recurrent $\Rightarrow p_{jj}^{(n)} \to 0 \Rightarrow p_{jj}^{r+n+s} \to 0 \Rightarrow p_{ii}^{(n)} \to 0 \Rightarrow i$ is null-recurrent.

$\square$

Definition 3.5.
The state j is periodic with period d if the greatest common divisor of $\{n : p_{jj}^{(n)} > 0\}$ is given by d.

Exercise 31. *Show that if $i \curvearrowleft\!\!\curvearrowright j$ and i has period d, than j has period d.*

As a example consider the random walk on $\mathbb{Z}$, where every state has period 2. In A, only the non-absorbing states have period 2 (for $S = |N| \geq 3$, for $N = 3$ the period of the state in the middle is not defined).

Definition 3.6.
A non-empty set $C \subseteq S$ is closed if there is no transition from C to $S \setminus C$, i.e. for every i in C and j in $S \setminus C$ we have $p_{ij} = 0$.

Remark that in a closed subset C one has $\sum_{j \in C} p_{ij} = 1$ and therefore we can restrict the Markov chain to C and get a stochastic matrix again.

Definition 3.7.
A Markov chain with state space S is reducible if there exists some $C \subsetneq S$ which is closed.

Theorem 3.9. *A Markov chain is not reducible if and only if it is irreducible.*

Exercise 32. *Proof the previous theorem.*

Theorem 3.10. *Suppose $i \in S$ is recurrent. Define $C(i) := \{j \in S : i \curvearrowright j\}$. Then $C(i)$ is closed, contains i and is irreducible, i.e. for all j and k in $C(i)$ we have $j \curvearrowleft\!\!\curvearrowright k$.*

Proof. The state i is in $C(i)$ since the relation is reflexive and closed since it is transitive. Therefore it remains to show that for each j in $C(i)$ we have $j \curvearrowright i$. We define α as the probability to reach j from i before returning to i. Then

$$\alpha = p_{ij} + \sum_{k \neq j} p_{ik} p_{kj} + \cdots > 0.$$

If we define f_{ji} as the probability of ever reaching i from j, then since i is recurrent, we get

$$\begin{aligned} 0 = 1 - f_{ii} &= P(X_n \neq i \forall n > 0 | X_0 = i) \\ &\geq \alpha \cdot P(X_n \neq i \forall n | X_0 = j) \\ &= \alpha \cdot (1 - f_{ji}) \end{aligned}$$

and therefore $f_{ij} = 1$, which means $j \curvearrowright i$. $\square$

Remark that from a recurrent state we never reach a transient state since every state in $C(i)$ is recurrent. Each recurrent state belongs to a unique closed irreducible subset. From transient states we can reach recurrent states (for example in A). The following structure theorem refines this statements.

Theorem 3.11. *The state space at a Markov chain can be divided in a unique way into disjoint sets*

$$S = T \dot\cup C_1 \dot\cup C_2 \dot\cup \dots,$$

where T is the set of all transient states and each C_i is closed, irreducible and recurrent.

I.

Assume the same stochastic matrix as in H but now all the q_i's are 0. Therefore we go surely from k to $k+1$. Hence every state is transient and there are infinitely many closed subsets $\{k, k+1, k+2, \dots\}$, but none of them is irreducible.

3.15 Limits

Theorem 3.12. *If j is transient or null-recurrent then for all i in S the probability $p_{ij}^{(n)}$ tends to 0 as n tends to infinity.*

Proof. For the first case we look at $i = j$. If j is transient, then $p_{jj}^{(n)} \to 0$ since the sum $\sum p_{jj}^{(n)}$ does converge. If j is null-recurrent, it is given by the definition and the renewal theorem. If $i \neq j$, we look at the renewal equation

$$p_{ij}^{(n)} = \sum_{k=1}^{(n)} f_{ij}^{(k)} \underbrace{p_{jj}^{(n-k)}}_{\leq 1}.$$

Since $\sum f_{ij}^{(k)} = f_{ij} \leq 1$, we can split the equation in two parts

$$p_{ij}^{(n)} = \underbrace{\sum_{k=1}^{m} f_{ij}^{(k)} p_{jj}^{(n-k)}}_{I} + \underbrace{\sum_{k=m+1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}}_{II}$$

and estimate. For the second part, we find for every $\epsilon > 0$ some m such that for the remaining terms of the sum we get $\sum_{k=m+1}^{\infty} f_{ij}^{(k)} < \epsilon$, and therefore $II < \epsilon$. Now with fixed m we choose n large enough such that $p_{jj}^{(n-k)} < \epsilon$ for $k = 1, 2, \dots, m$. Then

$$I < \sum_{k=1}^{m} f_{ij}^{(k)} \cdot \epsilon \leq \epsilon.$$

Therefore $p_{ij}^{(n)} < 2\epsilon$ and since we can choose ϵ arbitrary, we have proven the statement. $\square$

The strategy of our estimation above will be needed again later. Therefore, we formulate it as a

Lemma 3.13. *Let $x_i^{(n)}$ be a sequence with $n \in \mathbb{N}$ and $i \in S$ where S is countable. If we have*

1. $\forall i \in S : x_i^{(n)} \to x_i$ *as* $n \to \infty$,
2. $\forall i \in S \exists C_i$ *such that* $|x_i^{(n)}| \leq C_i$ *and*
3. $\sum_{i \in S} C_i < \infty$.

Then we can interchange the limit and the summation, i.e.

$$\lim_{n \to \infty} \sum_{i \in S} x_i^{(n)} = \sum_{i \in S} x_i.$$

Exercise 33. *Prove the lemma above.*

This is a special case of Lebesque's dominated convergence theorem where we choose some countable measure space S and μ as the counting measure. The theorem states that if we have a sequence of functions $f_n(x)$ that converges to some $f(x)$ for almost all $x \in S$, where S is some measurable space and

$$|f_n(x)| \leq g(x) \quad \text{and} \quad \int_S g(x) d\mu(x) < \infty$$

then

$$\lim_{n \to \infty} \int_S f_n(x) d\mu(x) = \int_S f(x) d\mu.$$

Theorem 3.14. *If j is positive recurrent and nonperiodic then $p_{ij}^{(n)}$ tends to $\frac{f_{ij}}{m_{jj}}$ (if $i \curvearrowleft\!\!\curvearrowright j$ then $f_{ij} = 1$).*

The finite case is already given by the ergodic theorem. If $i = j$ then we showed the statement with the renewal theorem.

Proof. Using the lemma, we calculate

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} \underbrace{p_{jj}^{(n-k)}}_{\to \frac{1}{m_{jj}}} \to \frac{1}{m_{jj}} \sum_{k=1}^{\infty} f_{ij}^{(k)} = \frac{f_{ij}}{m_{jj}}.$$

$\square$

Corollary 3.15. *In a finite Markov chain there is at least one recurrent state. All recurrent states are positive recurrent.*

Proof. Since $P^n\mathbf{1} = \mathbf{1}$ we get by 3.12 $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all i and j if all states are transient. But $\sum_j p_{ij}^{(n)} = 1$ for any n which is a contradiction. Therefore there exists a recurrent state. Now consider a class C of recurrent states which is irreducible and closed. Now applying theorem 3.12 to C, we get that there exists a positive recurrent state. From theorem 3.8 we now know that all states are positive recurrent. □

3.16 Stationary probability distributions

Lemma 3.16. *Let $u = (u_k) \in \triangle(s)$ be an invariant probability distribution and j a transient or null-recurrent state. Than $u_j = 0$.*

Proof. At first, note that $u = uP = uP^n$ and therefore with 3.12 and the dominated convergence theorem

$$u_j = \sum_{i\in S} u_i p_{ij} = \sum_{i\in S} u_i \underbrace{p_{ij}^{(n)}}_{\to 0} \xrightarrow{n\to\infty} 0.$$

□

In example I there may be no invariant probability distribution as long as S is infinite.

Theorem 3.17. *An irreducible positive recurrent and aperiodic Markov chain has an unique invariant probability distribution $u \in \triangle(s)$ such that $uP = u$. The entries are given by $u_i = 1/m_{ii}$ where m_{ii} is the mean recurrence time to state i.*

Proof. Since $i \curvearrowleft\!\curvearrowright j$ for all i and j, $f_{ij} = 1$ and hence by 3.14 we know that $p_{ij}^{(n)}$ tends to $\frac{1}{m_{jj}} =: u_j$. Without loss of generality we assume S is some subset of $\{0, 1, 2, \dots\}$.Then we get for all i and j

$$1 = \sum_{j=0}^{\infty} p_{ij}^{(n)} \geq \sum_{j=0}^{M} p_{ij}^{(n)} \xrightarrow{n\to\infty} \sum_{j=0}^{M} u_j \quad \forall M.$$

Now since for all M the sum $\sum_{j=0}^{M} u_j \leq 1$, this also holds for $M \to \infty$, but we have to show equality. Now if we look at the inequality

$$u_j \leftarrow p_{jj}^{(n+1)} = \sum_{i=1}^{\infty} p_{ji}^{(n)} p_{ij} \leq \sum_{i=1}^{M} p_{ji}^{(n)} p_{ij} \to \sum_{i=1}^{M} u_j p_{ij}$$

and taking the limit $M \to \infty$, we get

$$u_j \geq \sum_{i=0}^{\infty} u_i p_{ij}, \tag{2}$$

but again we need to show equality. Therefore suppose there is some j such that $u_j > \sum_{i=0}^{\infty} u_i p_{ij}$. But then we see

$$1 \geq \sum_{j=0}^{\infty} u_j > \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} u_i p_{ij} \overset{(*)}{=} \sum_{i=0}^{\infty} u_i \underbrace{\sum_{j=0}^{\infty} p_{ij}}_{=1} = \sum_{i=0}^{\infty} u_i,$$

and that is clearly a contradiction. Remark that the sum can be reordered in $(*)$ because all terms are positive. Now we know that for all j equality holds in (2) and so $u = uP$ and hence $u = uP^n$. Now because of the dominated convergence theorem we get for some $u_j > 0$

$$u_j = \sum_{i=0}^{\infty} u_i p_{ij}^{(n)} \xrightarrow{n \to \infty} \sum_{i=0}^{\infty} u_i u_j = u_j \sum_{i=0}^{\infty} u_i$$

and therefore $\sum u_i = 1$. The proof for uniqueness is exactly the same as in the finite case and can be found after theorem 3.1. □

It is also possible to show the converse theorem.

Theorem 3.18. *Consider an irreducible aperiodic Markov chain with stationary probability distribution $u \in \triangle(s)$ such that $u = uP$. Then all states $i \in S$ are positive recurrent and $u_i = \frac{1}{m_{ii}}$.*

Exercise 34. *Proof the theorem above.*

Now let us denote with $C_1, C_2, \ldots$ the different positive recurrent classes. If $\mathcal{P}$ be the set of all positive recurrent classes and $\alpha_i > 0$ for all $i \in \mathcal{P}$ and $u^{(i)}$ the unique invariant probability distribution in C_i, then

$$u = \sum_{i \in P} \alpha_i u^{(i)}$$

is an invariant probability distribution for the whole Markov chain. It can be shown that all invariant probability distributions are of this form. As an example we look at a birth-death chain on $\mathbb{N}$, given by the matrix

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \end{pmatrix}$$

where $q_i + r_i + p_i = 1$ for all i. Furthermore we want p_i and q_i be positive such that P is irreducible. We want to find an invariant distribution (we already did this for the finite case in 3.7). As in the finite case we get for $k = 0, 1, 2, \dots$

$$u_{k+1}q_{k+1} = u_k p_k$$

and therefore

$$u_k = \frac{p_{k-1}}{q_k} u_{k-1} = \dots = u_0 \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k}.$$

This defines a probability distribution ($\sum u_i = 1$) if and only if

$$\frac{1}{u_0} = \sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_1 q_2 \cdots q_k} = \infty \quad \Leftrightarrow \quad \text{all states are transient or null-recurrent.}$$

The sum is infinite if and only if all states are transient or null-recurrent. As a special case we look at the random walk with one reflecting boundary where $p_i = p$ and $q_i = q$ for all i. Then we get for the sum

$$\sum_{k=1}^{\infty} \left(\frac{p}{q}\right)^k < \infty \quad \Leftrightarrow \quad p < q.$$

Now we have

- $p < q \Leftrightarrow$ all states are positive recurrent, $u_k = u_0(\frac{p}{q})^k$ and $u_0 = 1 - \frac{p}{q}$.
- $p = q \Leftrightarrow$ all states are null-recurrent.
- $p > q \Leftrightarrow$ all states are transient.

Exercise 35. *Show the last two equivalences above.*

3.17 Periodic Markov chains

Theorem 3.19. *Let C be an equivalence class of recurrent states where one (and hence all) $i \in C$ have period $d > 1$. Then $C = C_1 \dot\cup C_2 \dot\cup \ldots \dot\cup C_d$ such that for all $i \in C_k$*

$$\sum_{j \in C_{k+1}} p_{ij} = 1 \quad (\textit{for } k \textit{ mod } d).$$

Therefore transition is only possible from C_k to C_{k+1} and therefore

$$P = \begin{pmatrix} 0 & P_{1,2} & 0 & 0 & \dots & 0 \\ 0 & 0 & P_{2,3} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & & 0 \\ & & & \ddots & \ddots & \\ 0 & 0 & \dots & & 0 & P_{d-1,d} \\ P_{d,1} & 0 & \dots & & 0 & 0 \end{pmatrix}$$

, where $P_{k,k+1}$ denotes a block matrix.

For all i and j in C we denote $q_{ij} := p_{ij^{(d)}}$, and since $\sum_{j \in C} q_{ij} = 1$ for all i in C_1, Q is an aperiodic stochastic matrix. If Q is restricted to C_1, it is irreducible and if C_1 is finite, Q restricted to C_1 is primitive. If u is recurrent, then $q_{ii}^{(n)} = p_{ii}^{(nd)}$ tends to $\frac{d}{m_{ii}}$, that is the reciprocal of the mean recurrence time for Q. Therefore $Q = P^d$ is given by the diagonal matrix

$$\begin{pmatrix} P_{12}P_{23}\cdots P_{d1} & 0 & \cdots & & 0 \\ 0 & P_{23}P_{34}\cdots P_{12} & & & \vdots \\ \vdots & & \ddots & & \\ & & & P_{d-1,d}P_{d1}\cdots P_{d-2,d-1} & 0 \\ 0 & \cdots & & 0 & P_{d1}P_{12}\cdots P_{d-1,d} \end{pmatrix}$$

The C_i's can have different sizes. As an example, if $C_1 = \{1, 2\}$, $C_2 = \{3\}$ and $C_3 = \{4, 5, 6\}$ and $d = 3$, then a possible matrix would be

If $Q = P^d$ then $q_{jj}^{(n)} = p_{jj}^{(nd)}$ and this expression tends to $\frac{d}{m_{jj}}$ ans n tends to infinity. To be more precise, we have

$$p_{ij}^{(nd+k)} \to \begin{cases} \frac{d}{m_{jj}} \text{ if } i \in C_\alpha, j \in C_\beta \text{ and } \alpha + k = \beta (\mod d) \\ 0 \text{ otherwise.} \end{cases}$$

We already know for aperiodic irreducible and positive recurrent Markov chains, then $p_{ij}^{(n)} \to u_j$. In the periodic case we get a slightly weaker result,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} p_{ij}^{(n)} = \lim_{n\to\infty} \frac{1}{d} \sum_{k=1}^{d} p_{ij}^{n+k} u_j,$$

where the left term is the time average of $p_{ij}^{(n)}$ and the right term is the average over one period. As a simple example we look at

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then $P^{2n} = \mathrm{Id}$ and $P^{2n+1} = \mathrm{P}$, hence there is no convergence. But

$$\lim_{n\to\infty} \frac{1}{N}(\mathrm{Id} + P + \cdots + P^{N-1}) = \frac{\mathrm{Id} + P}{2} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

3.18 A closer look on the Wright-Fisher Model

We recall example D, but assume $S = N$ instead of $S = 2N$ now. Then the probabilities are given by

$$p_{ij} = \binom{N}{j} \Big(\frac{i}{N}\Big)^j \Big(1 - \frac{i}{N}\Big)^{N-j},$$

which means that if $X_n = i$ then X_{n+1} has binomial distribution $\mathcal{B}(N, \frac{i}{N})$. The expected value of X_{n+1} is then given by

$$E[X_{n+1}] = N\frac{i}{N} = i = X_n.$$

Such a process is called a martingale. For X_n arbitrary we get

$$\begin{aligned}
E[X_{n+1}] &= \sum_{j=0}^{N} jP(X_{n+1} = j) = \sum_{j=0}^{N} j \sum_{i=0}^{N} P(X_n = i) p_{ij} \\
&= \sum_{i=0}^{N} P(X_n = i) \cdot \underbrace{\sum_{j=0}^{N} j p_{ij}}_{=i} = \sum_{i=0}^{N} iP(X_n = i) \\
&= E[X_n],
\end{aligned}$$

and hence $E[X_n] = E[X_0]$. We already showed with example 20 and lemma 3.16, that all states besides 0 and N are transient. For those states we have $p_{ij}^{(n)} \to 0$ and thus

$$i = E[X_0] = E[X_n] = \sum_{j=0}^{N} p_{ij}^{(n)} \cdot j \xrightarrow{n\to\infty} 0 \cdot \lim_{\to\infty} p_{i0}^{(n)} + N \cdot \lim_{n\to\infty} p_{iN}^{(n)}$$

and hence the probability of absorption in N is

$$\lim_{n\to\infty} p_{iN}^{(n)} = \frac{i}{N}$$

and the probability of absorption in 0

$$\lim_{n\to\infty} p_{i0}^{(n)} = 1 - \frac{i}{N}.$$

3.19 Absorbing Markov chains for finite state spaces

Denote $S = \{1, 2, \dots, N\} = T \dot{\cup} R$ where T is the set of all transient and R the set of all recurrent states. If we rearrange the states such that $R = \{1, \dots, r\}$ and $T = \{r+1, \dots, N\}$ and assume all recurrent states are absorbing, then the matrix is given by the block matrix

$$P = \begin{pmatrix} \mathrm{Id} & 0 \\ B & Q \end{pmatrix},$$

where B denotes a $N - r \times r$ matrix Q a $N - r \times N - r$ matrix. Now we define

- a_{ij} as the probability of absorption in j with starting point in i,
- τ_i as the expected time until absorption if we start in i and
- v_{ij} as the expected number of visits in j if we start in i.

Then we can calculate

1. $\tau_i = 1 + \sum_{k\in T} p_{ik}\tau_k$ for all $i \in T$, therefore we have $N - r$ equations for $N - r$ variables. If $\tau = (\tau_i)_{i\in T}$, then

   $$\tau = \mathbf{1} + Q\tau \iff \tau = (\mathrm{Id} - Q)^{-1}\mathbf{1}.$$

   The matrix Q is substochastic, i.e. $\sum_{j\in T} q_{ij} \le 1$ for all $i \in T$ and there is some i such that strict inequality holds if Q is irreducible.

2. For the expected number of visits we get

$$v_{ij} = \delta_{ij} + \sum_{k \in T} p_{ik} v_{kj}$$

for i and j out of T. Writing this equation with matrices, we get

$$V = \mathrm{Id} + QV \;\Leftrightarrow\; V = (\mathrm{Id} - Q)^{-1}.$$

3. For the absorption probabilities we get

$$a_{ij} = p_{ij} + \sum_{k \in T} p_{ik} a_{kj}$$

for $i \in T$ and $j \in R$, or writing $A = (a_{ij})$

$$A = B + QA \Leftrightarrow A = (\mathrm{Id} - Q)^{-1} B.$$

Exercise 36. *Show that* $\mathrm{Id} - Q$ *is invertible.*

Exercise 37. *Compute* τ_i *for example D.*

3.20 Birth-death chains with absorbing states

Assume the state space of a birth-death chain is given by $S = \{0, 1, \ldots, N\}$ and let the state 0 be absorbing. Then the general matrix is given by

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ q_1 & r_1 & p_1 & 0 & \ldots & 0 \\ 0 & q_2 & r_2 & p_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & \cdots & 0 & 0 & q_N & r_N \end{pmatrix}.$$

We assume all q_i and p_i positive, then all states besides 0 are transient and the unique stationary probability distribution is given by $u = (1, 0, \ldots, 0)^t$, i.e. $p_{i0}^{(n)} \to 1$ and $p_{ij}^{(n)} \to 0$ for $j \geq 1$. Now τ_k describes the time until absorption in 0 when we start in k. We get the equations

$$\tau_k = 1 + q_k \tau_{k-1} + r_k \tau_k + p_k \tau_{k+1} \text{ for } k = 1, \ldots, N-1 \text{ and}$$
$$\tau_N = 1 + q_N \tau_{N-1} + r_N \tau_N = 1 + q_N \tau_{N-1} + (1 - q_N) \tau_N.$$

From the second equation we get $q_N(\tau_N - \tau_{N-1}) = 1$ and the first can be written as

$$0 = 1 + q_k(\tau_{k-1} - \tau_l) + p_k(\tau_{k+1} - \tau_k)$$

and therefore we get a system of equations for the differences of two entries of τ

$$\begin{aligned}
\tau_2 - \tau_1 &= -\frac{1}{p_1} + \frac{q_1}{p_1}\tau_1 \\
\tau_3 - \tau_2 &= \frac{1}{p_2}(-1 + q_2(\tau_2 - \tau_1)) = -\frac{1}{p_2} - \frac{q_2}{p_1 p_2} + \frac{q_1 q_2}{p_1 p_2}\tau_1 \\
\tau_4 - \tau_3 &= \frac{1}{p_3}(-1 + q_3(\tau_3 - \tau_2)) = -\frac{1}{p_3} - \frac{q_3}{p_2 p_3} - \frac{q_2 q_3}{p_1 p_2 p_3} + \frac{q_1 q_2 q_3}{p_1 p_2 p_3}\tau_1 \\
&\vdots \\
\tau_N - \tau_{N-1} &= -\frac{1}{p_{N-1}} - \frac{q_{N-1}}{p_{N-2} p_{N-1}} - \cdots - \frac{q_2 q_3 \cdots q_{N-1}}{p_1 p_2 \cdots p_{N-1}} + \frac{q_1 q_2 \cdots q_{N-1}}{p_1 p_2 \cdots p_{N-1}}\tau_1
\end{aligned}$$

and since we already know $\tau_N - \tau_{N-1} = \frac{1}{q_n}$ we can compute

$$\tau_1 = \frac{1}{q_1} + \frac{p_1}{q_1 q_2} + \frac{p_1 p_2}{q_1 q_2 q_3} + \cdots + \frac{p_1 p_2 \cdots p_{N-1}}{q_1 q_2 \cdots q_N}$$

and therefore all other τ_k.

3.21 (Infinite) transient Markov chains

Consider our state space S as finite or countable and remember the division into the set of transient states T and the set of recurrent states R. Then we already know that Q, the restriction of P to T is a substochastic matrix again. Define $Q^n = (q_{ij}^{(n)})$, then we have

$$q_{ij}^{(n+1)} = \sum_{k \in T} q_{ik} q_{kj}^{(n)}$$

again. The row sum of Q_n we denote by $\sigma_i^{(n)}$ and therefore

$$\sigma_i^{(n)} = \sum_{j \in T} q_{ij}^{(n)} = P(X_n \in T | X_0 = i).$$

Since Q is substochastic we have $\sigma_i^{(1)} \leq 1$ and now we can calculate

$$\sigma_i^{(2)} = \sum_{j\in T} q_{ij}^{(2)} = \sum_{j\in T}\sum_{k\in T} q_{ik}q_{kj}^{(1)} = \sum_{k\in T}\sum_{j\in T} q_{ik}q_{kj}^{(1)} = \sum_{k\in T} q_{ik}\sigma_k^{(1)} \leq \sum_{k\in T} q_{ik} = \sigma_i^{(1)}.$$

By induction we get $\sigma_i^{(n+1)} \leq \sigma_i^{(n)}$. The probability to stay in T forever provided that we start in i is given by $\lim_{n\to\infty} \sigma_i^{(n)} := \sigma_i$. For $n \to \infty$ we get

$$\sigma_i = \sum_{k\in T} q_{ik}\sigma_k,$$

i.e. the vector σ is a eigenvector of Q. If $x = (x_i)_{i\in T}$ is a solution for $\sigma = Q\sigma$ with $0 \leq x_i \leq 1$ then $0 \leq x_i \leq \sigma_i^{(n)}$ and by induction we get $0 \leq x_i \leq \sigma_i$ hence σ is the maximal solution with $0 \leq \sigma_i \leq 1$.

Theorem 3.20. *The probabilities x_i that starting from state i the Markov chain stays forever in T are given by the maximal solution σ with $0 \leq \sigma_i \leq 1$.*

Exercise 38. *Given $S = \{1, \dots, N\}$ and $x = Px$, show*

- *if $0 \leq x_i \leq 1$ holds for all i then $\{i \ : \ x_i = 1\}$ is closed,*
- *if $i \leftrightarrow j$ then $x_i = x_j$ and*
- *if P is irreducible then $x = \lambda\mathbf{1}$.*

As an example we look at a birth-death with infinity state space chain given by

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots \\ & \ddots & \ddots & \ddots & \end{pmatrix}$$

then $R = \{0\}$ and $T = \{1, 2, \dots\}$. If we restrict P to the transient state, i.e. we cancel the first row and column. Then we try to solve $x = Qx$, we get the equations

$$\begin{aligned} x_1 &= r_1x_1 + p_1x_2 \\ x_2 &= q_2x_1 + r_2x_2 + p_2x_3 \\ &\vdots \end{aligned}$$

By replacing r_i with $1 - p_i - q_i$ the first line gives $x_2 > x_1$ and we get a system for the differences

$$\begin{aligned} q_2(x_2 - x_1) &= p_2(x_3 - x_2) \\ q_3(x_3 - x_2) &= p_3(x_4 - x_3) \\ &\vdots \\ q_k(x_k - x_{k-1}) &= p_k(x_{k+1} - x_k). \end{aligned}$$

This gives the equations

$$x_{k+1} - x_k = \frac{q_k}{p_k}(x_k - x_{k-1}) \frac{q_2 \dots q_k}{p_2 \dots p_k}(x_2 - x_1)$$

and

$$x_{k+1} - x_1 = (x_2 - x_1) \sum_{i=2}^{k} \frac{q_2 \cdots q_i}{p_2 \cdots p_i}.$$

Then we get (x_n) is bounded if and only if $\sum_{i=2}^{k} \frac{q_2 \cdots q_i}{p_2 \cdots p_i} < \infty$. In this case the process remains in T forever with positive probability.

3.22 A criterion for recurrence

Theorem 3.21. *In an irreducible Markov chain on $S = \{0, 1, 2, \dots\}$ the state 0 is recurrent if and only if the only solution of $x = Px$ with $0 \le x_i \le 1$ is given by $x_i = 0$ for $i = 1, 2, \dots$.*

Proof. $\Rightarrow$) Define $Q := (P_{ij})_{i,j=1}^{\infty}$ and consider $\sigma = Q\sigma$ as in the last chapter with $0 \le \sigma_i \le 1$. Then σ_i is the probability that X_n is non-zero for all positive n if we start in i. If 0 is recurrent then the probability of reaching 0 from i given by $f_{i0} = \sum f_{i0}^{(n)} = 1$ for all i and therefore $\sigma_i = 0$. $\Leftarrow$) Since $x_i = 0$ is the only solution, $\sigma_i = 0$ for $i = 1, 2, \dots$ and therefore $f_{i0} = 1$ and therefore i is recurrent. $\square$

Remember the example from the last chapter with p_i and q_i positive, then we get with the previous theorem

$$\begin{aligned}
&x = Qx \text{ has a bounded solution with } 0 \le x_i \le 1 \\
\Leftrightarrow &\sum_{i=2}^{\infty} \frac{q_2 \cdots q_i}{p_2 \cdots p_i} < \infty \\
\Leftrightarrow &\sum_{k=1}^{\infty} \frac{q_1 \cdots q_k}{p_1 \cdots p_k} < \infty \\
\Leftrightarrow &\ P \text{ is transient.}
\end{aligned}$$

We also get that P is recurrent if and only if the sum above is infinite. In the special case $p_i = p$ and $q_i = q$ we get

$$P \text{ is recurrent} \Leftrightarrow \sum_{k=1}^{\infty} \left(\frac{q}{p}\right)^k = \infty \Leftrightarrow q \ge p \Leftrightarrow p \le \frac{1}{2}$$

which we already stated in 3.16. If we denote

$$\pi_k = \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k}$$

and $\pi_0 = 1$ we get as a summary of 3.16 and 3.21 for birth-death chains

- $\sum_{k=1}^{\infty} \pi_k < \infty \;\Leftrightarrow\;$ positive recurrent
- $\sum_{k=1}^{\infty} \pi_k = \infty$ and $\sum_{k=1}^{\infty} \frac{1}{p_k \pi_k} = \infty \;\Leftrightarrow\;$ null recurrent
- $\sum_{k=1}^{\infty} \frac{1}{p_k \pi_k} < \infty \;\Leftrightarrow\;$ transience

3.23 Mean absorption times in the Wright-Fisher model

We will not give an explicit formula for τ_i in the model but discuss a heuristic method for an approximating formula. This formula has already been found by Wright in 1931. We assume N large and define $x = \frac{i}{N}$ and for the transition from i from j we denote $\frac{j}{N} = x + \delta_x$. Then we have

$$E[\delta_x] = E\Big[\frac{j-i}{N}\Big] = \frac{E[j] - E[i]}{N} = 0$$

because of the Martingale property stated in 3.18 and since for every fixed state we have a Binomial distribution, we get

$$\begin{aligned} E[(\delta_x)^2] &= \mathrm{Var}[\delta_x] = \mathrm{Var}[x+\delta_x] = \mathrm{Var}\Big[\frac{j}{N}\Big] = \frac{1}{N^2}\mathrm{Var}[j] \\ &= \frac{1}{N^2} N \frac{i}{N}\Big(1-\frac{i}{N}\Big) = \frac{x(1-x)}{N}. \end{aligned}$$

We write $\tau(x)$ for τ_i for $0 \le x \le 1$ and assume that τ is twice differentiable, then the recurrence relation for τ_i translates to

$$\begin{aligned} \tau(x) = 1 + E[\tau(x+\delta_x)] &\overset{\text{Taylor}}{=} 1 + E[\tau(x) + \delta_x \tau'(x) + \frac{1}{2}\delta_x^2 \tau''(x) + \dots] \\ &= 1 + \tau(x) + E[\delta_x]\tau'(x) + \frac{1}{2}E[\delta_x^2]\tau''(x) + \dots. \end{aligned}$$

Stopping the Taylor expansion after the second term, we get the differential equation

$$\tau(x) = 1 + \tau(x) + \frac{1}{2N} x(1-x)\tau''(x)$$

which we solve be simple integration. Hence

$$\begin{aligned} \tau''(x) &= \frac{-2N}{x(1-x)} \Rightarrow \tau'(x) = -2N(\log x - \log(1-x) + C) \\ &\Rightarrow \tau(x) = -2N[x\log x - x + (1-x)\log(1-x) - (1-x) + Cx + D], \end{aligned}$$

and with the boundary conditions $\tau(0) = \tau(1) = 0$ we get the entropy function

$$\tau(x) = -2N(x\log x + (1-x)\log(1-x).$$

In our model we get for $i = 1$, i.e. one single newly arising allele,

$$\tau_1 \approx \tau\Big(\frac{1}{N}\Big) = -2\log\frac{1}{N} - 2N\Big(1-\frac{1}{N}\Big)\log\Big(1-\frac{1}{N}\Big) \sim 2\log N + 2$$

and for $x = \frac{1}{2}$ we get

$$\tau\Big(\frac{1}{2}\Big) = -2N\frac{1}{2}2\log\frac{1}{2} = 2N\log 2 \approx 2.8N$$

3.24 The Moran model

The following model is related to the Wright-Fisher model and was described by Patrick Moran in 1958. We look at a population with N individuals of two types A and B and denote with i the number of individuals of type A. In each step we choose one individual for reproduction and one for death. Therefore the states 0 and N are absorbing and the probabilities are clearly given by

$$\begin{aligned} p_{i,i-1} &= \frac{N-i}{N}\frac{i}{N} \\ p_{i,i+1} &= \frac{i}{N}\frac{N-i}{N}, \end{aligned}$$

so they are symmetric but state dependent. Thus the model is essentially a birth-death chain with $p_i = q_i$ and with the same calculations as in 3.18 we get the martingale property. There we also have already shown that in that case

$$\lim_{n\to\infty} p_{iN}^{(n)} = \frac{i}{N} \text{ and } \lim_{n\to\infty} p_{i0}^{(n)} = 1 - \frac{i}{N}$$

holds.

Exercise 39. *Prove that the mean time for absorption is given by*

$$\tau_i = N\Big(\sum_{j=1}^{i} \frac{N-i}{N-j} + \sum_{j=i+1}^{N-i} \frac{i}{j}\Big)$$

if we start in state i. Hint: The proof is similar to that one in 3.20.

Exercise 40. *If we try the same approximation as in the previous chapter, show that*

$$\tau(x) = -N^2\Big((1-x)\log(1-x) + x\log x\Big).$$

Hint: Use the formula for the harmonic series $H_n \approx \log n + \gamma$ where γ is the Euler-Mascheroni constant.

One can improve the model by a selection process, so the A individuals get the fitness f_i and the B individuals the fitness g_i. The number of gametes are hence given by

$$if_i + (N-i)g_i$$

and therefore the new probabilities are given by

$$p_{i,i+1} = \frac{if_i}{if_i + (N-i)g_i} \frac{N-i}{N} \text{ and}$$
$$p_{i,i-1} = \frac{(N-i)g_i}{if_i + (N-i)g_i} \frac{i}{N}.$$

In that case one can get τ_i to grow like N, N^2 or even e^{cN}.

3.25 Birth-death chains with two absorbing states

Consider a birth-death chain where 0 and N are absorbing and denote

$$\gamma_i = \frac{q_1 q_2 \cdots q_i}{p_1 p_2 \cdots p_i}$$

and declare $\gamma_0 = 1$. We state for the absorption probabilities

$$a_{kN} = \frac{\sum_{i=0}^{k-1} \gamma_i}{\sum_{i=0}^{N-1} \gamma_i} \text{ and}$$
$$a_{k0} = \frac{\sum_{i=k}^{N-1} \gamma_i}{\sum_{i=0}^{N-1} \gamma_i}.$$

Exercise 41. *Show the formulas for a_{kN} and a_{k0}. Hint: The proof is similar to the one in 3.20. One have to solve for $x_k = a_{k0}$, $x_0 = 1$ and $x_N = 0$*

$$x_i = q_i x_{i-1} + r_i x_i + p_i x_{i+1}.$$

If we want to calculate the expected time to absorption τ_k, we have to solve

$$\tau_k = 1 + q_k \tau_{k-1} + r_k \tau_k + p_k \tau_{k+1}$$

with $\tau_0 = \tau_N = 0$.

Exercise 42. *Show that the solution for τ_i is given by*

$$\tau_1 = \frac{1}{1 + \gamma_1 + \cdots + \gamma_{N-1}} \sum_{k=1}^{N-1} \sum_{l=1}^{k} \frac{\gamma_k}{p_l \gamma_l}$$
$$\tau_j = -\tau_1 \sum_{k=j}^{N-1} \gamma_k + \sum_{k=j}^{N-1} \sum_{l=1}^{k} \frac{\gamma_k}{p_l \gamma_l}.$$

Hint: This implies the solution of exercise 39.

3.26 Perron-Frobenius theorem

The Perron-Frobenius is a general result about matrices, we will state different versions, but will not proof them.

Theorem 3.22 (Perron-Frobenius A). *Let $A = (a_{ij})$ denote a strictly positive or primitive $n \times n$ square matrix, i.e.*

$$\exists n : \ \forall i, j : \ a_{ij}^{(n)} > 0,$$

then

1. *there is an eigenvalue $r > 0$ with some eigenvector $w > 0$ such that $Aw = rw$ and r is an algebraically simple eigenvalue.*
2. *Furthermore we have for all eigenvalues $\lambda \neq r$ that $|\lambda| < r$.*

One consequence of the theorem is, that if we apply it to A^T, we get the existence of $v > 0$ with $v^T A = r v^T$. Furthermore we have for the limit case

$$\frac{A^n}{r^n} \xrightarrow{n \to \infty} \frac{w v^T}{v^T w}.$$

Exercise 43. *Given all statements from the theorem above, proof that*

$$\frac{A^n}{r^n} \xrightarrow{n \to \infty} \frac{w v^T}{v^T w}$$

assuming A is diagonalizable. Hint: Use a basis of eigenvalues and use $|\lambda| < r$.

If A is a stochastic matrix and primitive, then we already know $A\mathbf{1} = \mathbf{1}$, so $r = 1$ and $w = \mathbf{1}$. Therefore there is some positive vector u with $u^T A = u^T$ with $\sum u_i = 1$. Furthermore A^n tends to $\mathbf{1} u^T$, which is the ergodic theorem for primitive Markov chains. A second version of the theorem is

Theorem 3.23 (Perron-Frobenius B). *Let $A = (a_{ij})$ denote a non-negative irreducible $n \times n$ square matrix, i.e.*

$$\forall i, j \exists n : \ a_{ij}^{(n)} > 0,$$

then

1. *there is an eigenvalue $r > 0$ with some eigenvector $w > 0$ such that $Aw = rw$ and r is an algebraically simple eigenvalue.*

2. *Furthermore we have for all eigenvalues $\lambda \neq r$ that $|\lambda| \leq r$.*

In this case we still get for $r > 0$ and $w \geq 0$ such that $Aw = rw$ that none of the entries of w is 0. If we weaken the conditions for the theorem even more, we get

Theorem 3.24 (Perron-Frobenius C)**.** *Let $A = (a_{ij})$ denote a non-negative irreducible $n \times n$ square matrix, then*

1. *there is some eigenvalue $r \geq 0$ and some $w > 0$ such that $Aw = rw$.*

2. *For all eigenvalues λ we have $|\lambda| \leq r$.*

Finally the last statement is given by

Theorem 3.25 (Perron-Frobenius D)**.** *For two matrices B and A with $0 \leq B \leq A$ for every entry, we have for the spectral radius r*

$$r(B) \leq r(A).$$

If A or B is irreducible, then $r(B) = r(A)$ if and only if $A = B$.

3.27 Quasi stationary distributions

We look at some finite Markov chain with transient states T and recurrent states R where all recurrent states are absorbing. If we reorder the states, the chain is given by the block matrix

$$P = \begin{pmatrix} \mathrm{Id} & 0 \\ A & Q \end{pmatrix}.$$

The study of this question has already been started by Wright in 1931 and Yaglom in 1947 for the Galton-Watson process and Ewens and Seneta continued it in 1965 for general cases. For simplicity we combine all recurrent states into the first such that $R = \{0\}$ and we assume Q is irreducible. With

$$\pi(n) = (\pi_0(n), \pi_1(n), \ldots, \pi_N(n))$$

we denote the probability distribution at time n. Then

$$\pi(n+1) = \pi(n)P$$

holds. With $q(n)$ we denote the conditional distribution on $T = \{1, \dots, N\}$ in time n if we are not yet absorbed. This distribution is given by

$$q(n) = \frac{(\pi_1(n), \dots, \pi_N(n))}{\sum_{i=1}^N \pi_i(n)} = \frac{(\pi_1(n), \dots, \pi_N(n))}{1 - \pi_0(n)}.$$

We are looking for a stationary conditional distribution such that $q(n+1) = q(n) = q$. Using the block structure of P we get with

$$\Big(\pi_0(n+1), \tilde{\pi}(n+1)\Big) = \Big(\pi_0(n), \tilde{\pi}(n)\Big)P$$

the system of equations

$$\begin{cases} \pi_0(n+1) = \pi_0(n) + \tilde{\pi}(n)A \\ \tilde{\pi}(n+1) = \tilde{\pi}(n)Q. \end{cases}$$

Since $\tilde{\pi}(n) = (1 - \pi_0(n))q(n)$ holds, the second equation gives for the stationary distribution

$$c_n q := \frac{1 - \pi_0(n+1)}{1 - \pi_0(n)} q = qQ$$

with $q \geq 0$ and $q \in \triangle(T)$. For some irreducible Q the c_n are independent of n since q is a left eigenvector (which we denoted with v in previous sections). The c_n is simply the range of Q, denoted by r. Therefore we know there are v and w such that

$$\begin{aligned} v^T Q &= r v^T \\ Qw &= rw \end{aligned}$$

with both vectors greater 0. Therefore exists a stationary conditional distribution $q \in \triangle(T)$. Now we normalize v and w such that

$$\sum_i v_i = 1 = \sum_i v_i w_i,$$

choose Q to be primitive and write $p_{i0}^{(n)} = P(X_n = 0|X_0 = i)$. Therefore we get for $i, j \in T$

$$P(X_n = j|X_0 = i, X_n \neq 0) = \frac{p_{ij}^{(n)}}{1 - p_{i0}^{(n)}}$$
$$= \frac{q_{ij}^{(n)}}{\sum_{k\in T} q_{ik}^{(n)}} \xrightarrow{n\to\infty} \frac{r^n w_i v_j}{\sum_{k\in T} r^n w_i v_k} = v_j.$$

For the limit we used

$$\frac{A^n}{r^n} \to \frac{wv^T}{v^T w},$$

and thus approximated Q^n with $r^n wv^T$ and $q_{ij}^{(n)}$ with $r^n w_i vj$. We see that the v_j are independent from our starting point i, hence v is the limiting conditional distribution on T.

3.28 How to compute quasi-invariant distributions

We are now interested in explicit formulas for the v_j. Again our state space is finite and given by $S = \{0, 1, \ldots, N\}$, where 0 is the only absorbing state. From the q_j we get

$$\begin{aligned} q_j(n+1) &= \frac{p_j(n+1)}{1 - p_0(n+1)} = \frac{\sum_i p_i(n)p_{ij}}{1 - p_0(n)} \cdot \frac{1 - p_0(n)}{1 - p_0(n+1)} \\ &= \sum_{i\in T} q_i(n)p_{ij} \frac{1 - p_0(n)}{1 - p_0(n+1)} \\ &= \sum_{i\in T} q_i(n)p_{ij} \frac{1 - p_0(n)}{1 - p_0(n) - \sum_{k\in T} p_k(n)p_{k0}} \\ &= \sum_{i\in T} q_i(n)p_{ij} \frac{1}{1 - \sum_{k\in T} q_k(n)p_{k0}}, \end{aligned}$$

where we used $q_i(n) = p_i(n)/1 - p_0(n)$ twice. Since $q_i(n)$ tends to v_i, we get the system of equations

$$v_j = \frac{\sum_{i\in T} v_i p_{ij}}{1 - \sum_{k\in T} v_k p_{k0}} \;\Leftrightarrow\; v_j\Big(1 - \sum_{k\in T} v_k p_{k0}\Big) = \sum_{i\in T} v_i p_{ij},$$

consisting of N equations, which are not linear. In the simpler case of birth-death chains with only one absorbing state in 0 the system simplifies quite a lot. The matrix is given by

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q_1 & r_1 & p_1 & 0 & \dots & 0 \\ 0 & q_2 & r_2 & p_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & q_{N-1} & r_{N-1} & p_{N-1} \\ 0 & \cdots & 0 & 0 & q_N & r_N \end{pmatrix}$$

which gives the equations

$$v_j(1 - q_1 v_1) = v_{j-1}p_{j-1} + v_j(1 - p_j - q_j) + v_{j+1}q_{j+1}$$

and for the case $j = 1$

$$v_1(1 - q_1 v_1) = v_1(1 - p_1 - q_1) + v_2 q_2.$$

This can be solved for v_2 and continuing this patters we can express the v_j in terms of v_1 as a polynomial of degree j. The fact that all v_j sum up to 1 gives the last equation.

If we let q_1 tend to 0, then $Q1 = 1$ implies that v is a stationary distribution for Q. The equation $v^T = v^T Q$ gives nice approximate formulas for v. The assumption of choosing q_1 quite small is reasonable, because in this case the time to absorption is long and therefore the quasi-stationary distribution becomes relevant at all.

Exercise 44. *Modify the setting from above for $S = \{0, 1, \dots, N\}$ with absorbing states 0 and N. Apply it to the Moran model and show that v is approximately uniformly distributed (i.e. $v_i = 1/(N-1)$) for large N.*

Exercise 45. *Find a procedure to compute w and $v_i w_i$.*

With the Banach fixed-point theorem it is fairly easy to compute a numerical approximation of v. The function

$$x \mapsto \frac{xQ}{||xQ||_1}$$

maps $\triangle(T)$ to $\triangle(T)$ and has the unique fixed point v. If we assume the probability of being absorbed (which is in our example equal to the probability of leaving the transient states) is given by ε, we can reduce the Markov chain to

$$\begin{pmatrix} 1 & 0 \\ \varepsilon & 1-\varepsilon \end{pmatrix}.$$

This matrix only distinguishes between absorption and remain in the transient states. For birth-death chains, we have the simple relation

$$\varepsilon = \sum_{k \in T} v_k p_{k0} = v_1 q_1.$$

The random variable T_v, defined as the time to absorption if $P(X_0 = j) = v_j$ is geometrically distributed,

$$P(T_v = j) = (1-\varepsilon)^{j-1}\varepsilon$$

and therefore the mean time to absorption τ_v is given by $\frac{1}{\varepsilon}$.

Exercise 46. *Show that the expected value of a geometrical distribution with parameter ε is given by $1/\varepsilon$.*

As mentioned above for birth-death chains we get

$$\tau_v = \frac{1}{\sum_{k \in T} v_k p_{k0}} = \frac{1}{v_1 q_1},$$

so using this in the Moran model together with exercise 45 and our previous results, the mean time to absorption is given by

$$\tau_v = \frac{1}{v_1 q_1} \approx \frac{1}{\frac{1}{N-1}\frac{(N-1)}{N^2}} = N^2.$$

4 Poisson process

Our next big topic is about a stochastic process with countable state space S, but the time will be continuous now. The simplest example for such a process is the so called Poisson process.

4.1 Definition

To get a better feeling about this kind of process, we start with an example. If we go fishing and let $N(t)$ count the number of fish we already got, then $N(t)$ is a step function. We can now state some legitimate postulates.

(P0) The function $t \mapsto N(t)$ is a random variable from $\mathbb{R} \to \mathbb{N}$. It fulfills $N(0) = 0$, is increasing and continuous from the right.

(P1) If $[t, s)$ and $[u, v)$ are disjoint, then $N(s) - N(t)$ is independent from $N(u) - N(v)$, therefore the events in one time interval do not affect the events in another disjoint time interval.

(P2) The stationary increments distribution of $N(s) - N(t)$ depends on $s - t$ but not on t, i.e. $N(t + s) - N(s)$ and $N(t) - N(0)$ are identically distributed.

(P3) We have the two limits

$$\frac{1}{h} P(N(t+h) - N(t) \geq 1) \xrightarrow{h \downarrow 0} \lambda > 0$$
$$\frac{1}{h} P(N(t+h) - N(t) \geq 2) \xrightarrow{h \downarrow 0} 0.$$

Definition 4.1.
A stochastic process $N(t)$ with $t > 0$ satisfying the four postulates above is called a Poisson process with rate λ.

The last postulate is often written as

$$P(N(t+h) - N(t) \geq 1) = \lambda h + o(h)$$
$$P(N(t+h) - N(t) \geq 2) = o(h).$$

4.2 Characterization

Theorem 4.1. *The four postulates $(P0) - (P3)$ imply that $N(t)$ is Poisson distributed with parameter λ*

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

and hence has the expected value $E[N(t)] = \lambda t$.

Exercise 47. *Proof a kind of converse result. If (P0) and (P1) is given and we know*

$$\forall s, t > 0 \;:\; P(N(t+s) - N(s) = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

then we already get (P2) and (P3).

Proof. We define $P_m(t) := P(N(t) = m)$. Then we get from (P1) and (P2)

$$P_0(t+k) = P_0(t) \cdot P_0(k)$$

and therefore we know

$$\frac{P_0(t+h) - P_0(t)}{h} = \frac{1}{h}P_0(t)\underbrace{[P_0(h) - 1]}_{-(1-P_0(h))} = -P_0(t)\frac{P[N(h) - N(0) \geq 1]}{h}.$$

Pushing h to zero in this equation gives

$$P_0'(t) = -\lambda P_0(t)$$

and therefore

$$P_0(t) = e^{-\lambda t}P_0(0) = e^{-\lambda t}.$$

Furthermore we have

$$P_m(t+h) = \sum_{k=0}^{m} P_k(t)P_{m-k}(h) = P_m(t)P_0(h) + P_{m-1}(t)P_1(h) + \dots,$$

which we can divide by h to get

$$\frac{P_m(t+h) - P_m(t)}{h} = P_m(t)\underbrace{\frac{P_0(h) - 1}{h}}_{\to\lambda} + P_{m-1}(t)\frac{P_1(h)}{h} + \dots.$$

In this equation we can estimate some expressions with (P3)

$$\frac{P_i(h)}{h} = \frac{P(N(h) \geq i) - P(N(h) \geq 2)}{h} \xrightarrow{h\downarrow 0} \begin{cases} \lambda - 0 = \lambda \text{ for } i = 1 \\ 0 - 0 = 0 \text{ for } i = 2, 3, \dots, \end{cases}$$

and therefore it simplifies in the limit to

$$P_m'(t) = -\lambda P_m(t) + \lambda P_{m-1}(t).$$

To solve this, we define $Q_m(t) := e^{\lambda t}P_m(t)$. Then we get

$$Q_m'(t) = \lambda e^{\lambda t}P_m(t) + e^{\lambda t}P_m'(t) = \lambda Q_{m-1}(t).$$

For the case $m = 1$ we get $Q_1'(t) = \lambda Q_0(t) = \lambda$ and therefore $Q_1(t) = Q_1(0) + \lambda t = \lambda t$. Hence $P_1(t) = \lambda t e^{-\lambda t}$, the remaining cases can be shown by induction. □

4.3 Waiting times

Let $N(t)$ be a Poisson process and define T_1 as the time until the first jump. Then we have

$$P(T_1 > 0) = P(N(t) = 0) = P_0(t) = e^{-\lambda t} = \int_t^\infty \lambda e^{-\lambda\tau} \mathrm{d}\tau.$$

Using the Taylor expansion as an approximation, we get for small t

$$P(T_1 \leq t) = 1 - P_0(t) = 1 - e^{\lambda t} = 1 - (1 - \lambda t + \frac{\lambda^2 t^2}{2} + \dots) = \lambda t + o(t).$$

The waiting time is therefore exponentially distributed with parameter λ.

Exercise 48. *Show that the expected value of T_1 is given by $\frac{1}{\lambda}$.*

If we in addition define T_n as the time between the n^{th} and the foregoing jump of $N(t)$, we conclude for the conditional probability

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(\text{no jumps in } (s, s+t] | T_1 = s) \\ &= P(N(t+s) - N(s) = 0 | T_1 = s) \\ &\overset{(P2)}{=} P(N(t) - N(0) = 0) = e^{-\lambda t}. \end{aligned}$$

Thus T_2 is also exponentially distributed.

Theorem 4.2. *The random variables T_n are i.i.d. (independent, identically distributed) with exponential distribution with expectation $1/\lambda$.*

This result can be interpreted as the memorylessness of the Poisson process, which is also a way to characterize it. Let T_n be a sequence of independent identically exponentially distributed random variables with expectation $1/\lambda$ given and define

$$S_n := \sum_{k=1}^n T_k.$$

Furthermore define

$$\begin{aligned} N(t) &= \begin{cases} 0 \text{ for } 0 \leq t < S_1 \\ 1 \text{ for } S_1 \leq t < S_2 \\ \quad \vdots \\ n \text{ for } S_n \leq t < S_{n+1} \end{cases} \\ &= \max\{n : S_n \leq t\}, \end{aligned}$$

then $N(t)$ is a Poisson process with parameter λ. We show this in two steps. For the beginning

Exercise 49. *Show that S_n is $\Gamma-$distributed with density function*

$$f_n(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

With the result of the foregoing exercise we know

$$P(S_n \leq t) = 1 - \sum_{m=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^m}{m!}$$

and hence

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) \\ &= e^{-\lambda t} \frac{(\lambda t)^n}{n!}. \end{aligned}$$

4.4 Memorylessness of the exponential distribution

As mentioned at the end of the previous chapter, the exponential distribution can be characterized by its memorylessness. If we look at the conditional probability that the waiting time T is larger than some $s+t$ if we already waited for t, we see

$$P(T > s+t | T > t) = \frac{P(T > s+t)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s),$$

so the distribution does not recognize what happened in the past. If we rewrite the computation from above to the well-known functional equation

$$f(t+s) = f(s)f(t).$$

One can show that the exponential distributions are the only functions which satisfy this equation and are bounded at once.

Exercise 50. *Show that the exponential distributions are besides the zero function the only functions which satisfy*

$$\begin{cases} f(t+s) = f(t)f(s) \ \forall s,t > 0 \\ 0 \leq f(t) \leq 1 \ \forall t > 0. \end{cases}$$

4.5 Waiting time paradox

The waiting time paradox describes a counter-intuitive day-to-day situation. We try to get our bus to the university. Suppose buses arrive on average every τ minutes. We arrive at a random time at the bus stop. We expect a waiting time of $\tau/2$ on the average. But if the buses arrive according to let us say an exponential distribution, then our expected waiting time is in fact even τ. Actually no matter what the distribution is, if it has mean τ and standard deviation σ, then our average waiting time is given by

$$\frac{\tau}{2} + \frac{\sigma^2}{2\tau},$$

which is clearly larger than what we would expect intuitively. A simple explanation for this situation is, the longer the interval between two buses, the more probable is it for us to arrive in this particular interval. Therefore it is more probable to wait for some bus that is already late than to catch one which is too early. To be more precise, if $f(t)$ is the density function of the length of the intervals between two consecutive buses, then the density function of the random time interval till arrival of the next bus is not $f(t)$ but proportional to $tf(t)$, since the probability that we arrive during a certain interval is proportional to the length of this interval. Parts missing!!!

4.6 Conditional waiting time

Suppose we are at time t and $N(t) = 1$. Our next question is, when did this jump occur? We calculate

$$\begin{aligned}
P(T_1 < s | N(t) = 1) &= \frac{P(T_1 < s \wedge N(t) = 1)}{P(N(t) = 1)} \\
&= \frac{P(\text{ one jump in } [0, s) \wedge \text{ no jump in } [s, t))}{\lambda t e^{-\lambda t}} \\
&\overset{(\text{P1})}{=} \frac{P(\text{ one jump in } [0, s)) \cdot P(\text{ no jump in } [s, t))}{\lambda t e^{-\lambda t}} \\
&= \frac{\lambda s e^{-\lambda s} \cdot e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} = \frac{s}{t}.
\end{aligned}$$

Hence the appearance of the jump is uniformly distributed in the interval $[0, t]$. If $N(t) = n$, we write as in the previous chapter $S_n = \sum_k T_k$ and choose

$$0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = t$$

and h_i such that $t_i + h_i < t_{i+1}$. Then an similar calculation as above gives

$$
\begin{aligned}
&P(t_i \leq S_i \leq t_i + h_i \text{ for } i = 1, 2, \ldots, n | N(t) = n) \\
&= \frac{P(\text{ one jump in } [t_i, t_i + h_i] \text{ for } i = 1, \ldots, n \text{ and no other jump })}{P(N(t) = n)} \\
&\overset{(P1)}{=} \frac{\lambda h_1 e^{-\lambda h_1} \cdots \lambda h_n e^{-\lambda h_n} e^{-\lambda(t - h_1 - h_2 - \cdots - h_n)}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \\
&= \frac{n!}{t^n} h_1 \cdots h_n.
\end{aligned}
$$

This result is again independent of t_i and therefore we have again a uniform distribution.

4.7 Non-stationary Poisson process

We now look at some stochastic process with $t \in [0, \infty]$, $N(t) \in \mathbb{N}_0$ and some intensity function (at least integrable) $\lambda(t) \geq 0$. For this process, we keep the postulates (P0) and (P1), but we omit (P2) and modify (P3) to

(P3)'

$$
\begin{aligned}
P(N(t+h) - N(t) \geq 1) &\overset{h \downarrow 0}{=} \lambda(t) h + o(h) \\
P(N(t+h) - N(t) \geq 2) &\overset{h \downarrow 0}{=} o(h).
\end{aligned}
$$

Then analogous to theorem 4.1 we get

$$
P(N(t+s) - N(t) = k) = e^{-m(t+s)+m(t)} \frac{(m(t+s) - m(t))^k}{k!}
$$

where $m(t)$ is given by

$$
m(t) = \int_0^t \lambda(s) \mathrm{d}s.
$$

Suppose we have a Poisson process with rate λ. We count events (e.g. radioactive emissions), but we miss some of them. We count with probability $\frac{\lambda(t)}{\lambda}$ with $0 \leq \lambda(t) \leq 1$ at time t. In this case the number of counted events follows a non-stationary Poisson process with intensity function $\lambda(t)$. We only have to check (P3') and consider

$$\begin{aligned}
&P(\text{ Count an event in } [t, t+h]) \\
&= P(\text{ there is an event in } [t, t+h]) \cdot P(\text{ the event is counted }) \\
&= (\lambda h + o(h)) \frac{\lambda(t)}{\lambda} = \lambda(t) h + o(h).
\end{aligned}$$

5 Markov processes

5.1 Continuous-time Markov process

As the title adumbrates, we will now look at processes with the Markov property but will use continuous time. The state space in contrary remains discrete. This type of process is also called Markov-jump-process. We start with a family of random variables $X(t) : \Omega \to S$ with $t \geq 0$ which fulfills for every $0 \leq t_1 < t_2 < \cdots < t_n$ and for every $i_1, i_2, \ldots, i_{n-1}, j \in S$

$$P\Big(X(t_n) = j | \bigwedge_{k=1}^{n-1} X(t_k) = i_k\Big) = P\Big(X(t_n) = j | X(t_{n-1}) = i_{n-1}\Big).$$

If we interpret t_n as future and t_{n-1} as present, then the process is independent from the past. Similar to the third section we write

$$p_{ij}(t_1, t_2) = P(X(t_2) = j | X(t_1) = i)$$

for $i, j \in S$ and $t_1 < t_2$. But this definition is too much, since we want the Markov process to be homogeneous, i.e.

$$p_{ij}(t_1, t_2) = p_{ij}(t_2 - t_1).$$

Hence the process does not recognize when two events take place but only in which interval. Since $p_{ij}(t)$ should still describe probabilities, we further demand

$$\sum_{j \in S} p_{ij}(t) = 1$$

for all $i \in S$ and for every time $t \geq 0$. Our third demand on the process is the so-called Chapman-Kolmogorov equation

$$p_{ij}(s+t) = \sum_{k \in S} p_{ik}(s) p_{kj}(t) \ \forall s, t \geq 0$$

or in Matrix notation

$$P(s+t) = P(s) \cdot P(t) \ \forall s, t \geq 0.$$

Therefore the Markov processes are represented by a semi group of stochastic matrices with $P(0) = \mathrm{Id}$. Our last assumption, justified by the practical appearances of Markov processes, is that the map $t \mapsto P(t)$ is continuous.

One example for Markov processes is the Poisson process. We get

$$\begin{aligned} p_{ij}(t) = P(X(t+s) = j | X(s) = i) &= \frac{P(X(s) = i, X(t+s) - X(s) = j - i)}{P(X(s) = i)} \\ = P(X(t+s) - X(s) = j - i) &= \begin{cases} \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t} \text{ for } j \geq i \\ 0 \text{ for } j < i. \end{cases} \end{aligned}$$

5.2 Transition rates

We define the transition rates q_{ij} assuming the limit exists as

$$q_{ij} := \lim_{t \downarrow 0} \frac{p_{ij(t)} - p_{ij(0)}}{t}.$$

The transition rates describe the rate at which the process moves between two states. For finite state spaces we conclude

$$\sum_j p_{ij}(t) = 1 \Rightarrow \frac{d}{dt} \sum_j p_{ij}(t)|_{t=0} = 0,$$

where in the countable case we need suitable assumptions to interchange derivation and summation. Therefore the row sums of the transition rate matrix $Q = (p_{ij})$ are 0. Using q_{ij} as the linearization of p_{ij}, we get for $h \downarrow 0$

$$p_{ij}(h) = \delta_{ij} + q_{ij} h + o(h).$$

In the Poisson process, this gives

$$q_{ij} = \begin{cases} \lambda \text{ for } j = i+1 \\ 0 \text{ for } j > i+1 \text{ or } j < i \\ -\lambda \text{ for } j = i. \end{cases}$$

5.3 Pure birth process

A pure birth process is a growth process where we assume no death (this is in fact a generalization of Poisson process). We start with the transition rates

$$q_{ij} = \begin{cases} \lambda_i \text{ for } j = i+1 \\ -\lambda_i \text{ for } j = i \\ 0 \text{ otherwise.} \end{cases}$$

We write $P_n(t) = P(X(t) = n)$ and get

$$P_n(t+h) = P_n(t)[1 - \lambda_n h] + \lambda_{n-1} h P_{n-1}(t) + o(h).$$

Dividing this by h and going over to the limit gives the differential equation

$$P_n'(t) = \lim_{h \downarrow 0} \frac{P_n(t+h) + P_n(t)}{h} = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$$

and for the special case $n = 0$

$$P_0'(t) = -\lambda_0 P_0(t).$$

Now this can be solved recursively starting with $P_0(t) = P_0(0)e^{-\lambda_0 t}$. This solution is unique for every initial condition $P_n(0)$. A first example is the so-called Yule process, found by George Udny Yule in 1924. The random variable $X(t)$ counts the number of individuals in a population at time t, where each individual can split into two with rate λ. Hence the growth in a short time interval of length h is given by $\lambda h + o(h)$. The individuals split independently from each other, therefore $\lambda_n = n\lambda$. If we solve the system of differential equations in the case $P_n(0) = \delta_{ni}$, we get the solution

$$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i},$$

which is a negative binomial distribution.

Exercise 51. *Proof that the given solution of the Yule process example is correct (for example by induction).*

5.4 Divergent birth process

Again we grow with rate λ_n from n to $n+1$. Then the following statement holds.

Theorem 5.1. *The map $P_n(t)$ remains a probability distribution for every positive time t if and only if the sum of the reciprocal growth rates is infinite, i.e.*

$$\sum_{n=0}^{\infty} P_n(t) = 1 \ \forall t \geq 0 \ \Leftrightarrow \ \sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \infty.$$

Remark that otherwise there would be some t such that the sum of probabilities is less than 1. Therefore with probability $1 - \sum_{n=0}^{\infty} P_n(t)$ the population would have reached infinity.

Proof. We define

$$S_h(t) := P_0(t) + P_1(t) + \cdots + P_h(t).$$

Since the map $h \mapsto S_h(t)$ is increasing and bounded we can define

$$\mu(t) := \lim_{h \to \infty} (1 - S_h(t)) = 1 - \sum_{n=0}^{\infty} P_n(t).$$

Now we remember the differential equations from the previous chapter with initial condition $P_n(0) = \delta_{ni}$

$$\begin{cases} P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t) \\ P_0'(t) = -\lambda_0 P_0 \end{cases}$$

and sum them up from $n = 0$ to k. Then we have

$$S_k'(t) = -\lambda_k P_k$$

which we integrate from 0 to t to get

$$S_k(t) - S_k(0) = -\lambda_k \int_0^t P(\tau) \mathrm{d}\tau.$$

Now we have for $k \geq i$

$$\mu(t) \leq 1 - S_k(t) = \lambda_k \int_0^t P_k(\tau) \mathrm{d}\tau \leq 1,$$

which we divide by λ_k to get

$$\frac{\mu(t)}{\lambda_k} \leq \int_0^t P_k(\tau)\mathrm{d}\tau \leq \frac{1}{\lambda_k}.$$

Summing those inequalities up from $k = i$ to n we get

$$\mu(t)\Big[\frac{1}{\lambda_i} + \cdots + \frac{1}{\lambda_n}\Big] \leq \int_0^t S_n(\tau)\mathrm{d}\tau \leq \frac{1}{\lambda_i} + \cdots + \frac{1}{\lambda_n}.$$

Remark that we added a few terms in the middle part of the inequality, but they are 0 because of the initial condition anyway. Now suppose $S_n(t)$ tends to 1 for N to infinity for all t. Then the integral

$$\int_0^t S_n(\tau)\mathrm{d}\tau$$

would tend to t by monotone convergence theorem and therefore we get

$$t \leq \frac{1}{\lambda_i} + \cdots + \frac{1}{\lambda_n} \ \forall n$$

for an arbitrary t, hence the right side is infinite. On the other hand, if the sum of the reciprocal growth rates is infinite we get

$$\mu(t)\Big[\frac{1}{\lambda_i} + \cdots + \frac{1}{\lambda_n}\Big] \leq t \ \forall n$$

and therefore $\mu(t)$ has to be 0 for every fixed t. But this means

$$\sum_{k=0}^{\infty} P_k(t) = 1$$

for every t. □

What is the intuitive interpretation of this theorem? If we are in some state n, then we move to the next state with probability λ_n. Since the expected time to stay in some fixed state n is exponentially distributed, it is given by $\frac{1}{\lambda_n}$, and therefore the sum $\sum_{k=0}^{n} \frac{1}{\lambda_k}$ can be interpreted as the expected time spent at states 0 to n. An example for an explosive process is given by the growth rate $\lambda_n = n^2\lambda$, which is similar to the deterministic growth process given by the differential equation $x'(t) = \lambda x(t)^2$.

5.5 The Kolmogorov differential equations

If we use the Markov property $P(s+t) = P(s)P(t)$ to find the derivative of P, we get

$$P'(t) = \lim_{h \downarrow 0} \frac{P(t+h) - P(t)}{h} = \lim_{h \downarrow 0} \frac{P(h) - \mathrm{Id}}{h} P(t) = QP(t).$$

Here we used $P(t+h) = P(h+t) = P(h)P(t)$. The equation $P'(t) = QP(t)$ is called **Kolmogorov backward equation**. Without the matrix form, it is given entry-wise by

$$p'_{ij}(t) = \sum_{k \in S} q_{ik} p_{kj}(t).$$

If we use the Markov property the other way round, we get

$$P'(t) = \lim_{h \downarrow 0} \frac{P(t+h) - P(t)}{h} = P(t) \lim_{h \downarrow 0} \frac{P(h) - \mathrm{Id}}{h} = P(t)Q$$

or again entry-wise

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) q_{kj}.$$

As one can expect, this equation is called **Kolmogorov forward equation**. The derivation of P can be done without problems for finite state spaces, remember that for infinite state spaces, we need more suitable conditions. For the finite case, the unique solution is given by $P(t) = e^{Qt}$.

5.6 Stationary distributions

In this section we will show a very useful condition for stationary distributions. But first we have to define the concept of stationary distributions for continuous time.

Definition 5.1.
We call $u \in \triangle(S)$ a stationary or invariant probability distribution of $P(t)$ if $uP(t) = u$ holds for all $t \geq 0$.

Theorem 5.2. *A vector u is a stationary distribution of $P(t)$ if and only if $uQ = 0$ holds, where Q is the transition rate matrix of $P(t)$.*

Proof. Assume u is a stationary probability distribution. If we differentiate $u = uP(t)$, we get with Kolmogorov's forward equation

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}(uP(t)) = u\frac{\mathrm{d}}{\mathrm{d}t}P(t) = uP(t)Q = uQ.$$

For the other direction, we use a similar trick. Since $u\frac{\mathrm{d}}{\mathrm{d}t}P(t) = uQP(t) = 0$, we see that $P(t)$ has to be constant, and therefore we get

$$uP(t) = uP(0) = u.$$

$\square$

5.7 Birth-death process

In this section we generalize the concept of birth-death chains for continuous time, the state space is again $S = \{0, 1, 2, \dots\}$ and the associated matrix is given by

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -\mu_1 - \lambda_1 & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -\mu_2 - \lambda_2 & \lambda_2 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \end{pmatrix}.$$

Therefore the transition probabilities are given by

$$\begin{aligned} &p_{i,i+1}(h) = \lambda_i h + o(h) \text{ for } i \geq 0 \\ &p_{i,i-1}(h) = \mu_i h + o(h) \text{ for } i \geq 1 \\ &p_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h) \text{ for } i \geq 0. \end{aligned}$$

Now we look at the Kolmogorov backward equation assuming the initial condition $p_{ij}(0) = \delta_{ij}$ and all λ_i and μ_i positive. For $i = 0$ we get

$$p'_{0,j}(t) = -\lambda_0 p_{0,j}(t) + \lambda_0 p_{1,j}(t)$$

and for any $i > 0$

$$p'_{i,j}(t) = \mu_i p_{i-1,j}(t) - (\mu_i + \lambda_i)p_{i,j}(t) + \lambda_i p_{i+1,j}(t).$$

For now there is no clear way to solve this system. For the forward equation we have

$$p'_{i,0}(t) = -\lambda_0 p_{i,0}(t) + \mu_1 p_{i,1}(t) \text{ for } j = 0 \text{ and}$$
$$p'_{i,j}(t) = \lambda_{j-1} p_{i,j-1}(t) - (\mu_j + \lambda_j) p_{i,j}(t) + \mu_{j+1} p_{i,j+1}(t) \text{ for } j \geq 1.$$

This could be solvable of $\mu_i = 0$ for all i, i.e. there is no death. Forgetting about the first index (it does not matter in the system above), we get for $p_k(t) = P(X(t) = k)$ the recursively solvable system

$$\begin{cases} p'_0 = -\lambda_0 p_0 + \mu_1 p_1 \\ p'_j = \lambda_{j-1} p_{j-1} - (\lambda_j + \mu_j) p_j + \mu_{j+1} p_{j+1} \\ p_k(0) = \delta_{ki} \end{cases}$$

which we already considered in section 5.3.

Exercise 52. *Solve the pure death process ($\lambda_i = 0$) with initial value $p_N(0) = 1$, given by the equations*

$$\begin{cases} p'_j = -\mu_j p_j + \mu_{j+1} p_{j+1} \\ p'_N = \mu_N p_N. \end{cases}$$

It is not hard to find the stationary distribution for the birth-death process. Using the theorem from the last chapter, we try to find the solution for the forward equations with $p'_j = 0$ for all j. We get for the first equation $p_1 = \frac{\lambda_0}{\mu_1} p_2$ and similar as for the chain in section 3.7, we get by induction

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} p_0 := \pi_n p_0.$$

If $\sum_{k_0}^{\infty} \pi_k < \infty$ holds, then there exists a stationary distribution $p \in \triangle(S)$.

5.8 Linear growth with immigration

In this section we consider an example for a birth-death process with birth rate $\lambda_n = n\lambda + a$, where a is an immigration rate so that 0 is not absorbing in the model. The death rate is given by $\mu_n = n\mu$. With the notation from the previous chapter we get

$$\pi_n = \frac{a(\lambda + a)(2\lambda + a) \cdots ((n-1)\lambda + a)}{\mu^n n!}.$$

The quotient criterion gives

$$\frac{\pi_{n+1}}{\pi_n} = \frac{n\lambda + a}{(n-1)\mu} \xrightarrow{n\to\infty} \frac{\lambda}{\mu},$$

so the sum over all π_i converges if and only if λ is strictly smaller than μ. In this case there exists a stationary distribution.

Exercise 53. *Compute the expected value of individuals by looking at the derivative*

$$M'(t) = \sum_{k=0}^{\infty} k p_k'(t)$$

and deduce the differential equation

$$M'(t) = a + (\lambda - \mu)M(t).$$

Solve this equation and look at the limit cases for $\mu > \lambda$ and $\mu \leq \lambda$ if t tends to infinity.

5.9 The Moran process

As in the Moran chain, we look at a population with N individuals, each of type a or A. The random variable $X(t)$ counts the number of type a individuals. The state changes in the time interval $(t, t+h)$ for each individual with rate λ. For some t with $X(t) = j$ we choose an a individual with probability $\frac{j}{N}$ and an A individual with probability $1 - \frac{j}{N}$. Furthermore we add a mutation rate, i.e. a chosen a individual mutates to type A with probability γ_1 and a chosen A individual mutates to type a with probability γ_2. Now we get for the probability that some a replaces an A individual

$$\underbrace{(1 - \frac{j}{N})}_{A \text{ selected}} \underbrace{\Big(\frac{j}{N}(1-\gamma_1) + (1 - \frac{j}{N})\gamma_2\Big)}_{\text{replaced by } a}.$$

The Moran process can be considered as a birth-death process, where the rates are given by

$$\begin{aligned}
\lambda_j &= \lambda(1 - \frac{j}{N})\Big(\frac{j}{N}(1-\gamma_1) + (1 - \frac{j}{N})\gamma_2\Big) \\
\mu_j &= \lambda\frac{j}{N}\Big((1 - \frac{j}{N})(1-\gamma_2) + \frac{j}{N}\gamma_1\Big).
\end{aligned}$$

If we want to find a stationary distribution we have to look at the π_k as in the foregoing section. Since these terms are pretty hard to compute, we only try to solve the limit case as $N \to \infty$ with the additional condition that $N\gamma_i$ tends to some small $\varepsilon_i > 0$. We look at the random variable $\frac{1}{N}X(t)$ with state space $S = \{0, 1/n, 2/n, \ldots, 1\}$. As we evaluate π_k as $N \to \infty$, the fraction k/N should go to some $x \in [0,1]$. Taking the logarithm if the product, we get

$$\log \pi_k = \sum_{i=0}^{k-1} \log \lambda_i - \sum_{i=1}^{k} \log \mu_i.$$

Rewriting the terms of λ_j and μ_j we get

$$\begin{aligned}
\lambda_j &= \lambda \frac{j}{N}(1 - \frac{j}{N})\Big(1 - \gamma_1 + (\frac{N}{j} - 1)\gamma_2\Big) \\
&= \lambda \frac{j}{N}(1 - \frac{j}{N})(1 - \gamma_1 - \gamma_2)\Big(1 + \frac{N\gamma_2}{(1 - \gamma_1 - \gamma_2)j}\Big) \\
&:= \lambda \frac{j}{N}(1 - \frac{j}{N})(1 - \gamma_1 - \gamma_2)\Big(1 + \frac{a}{j}\Big)
\end{aligned}$$

and similar

$$\begin{aligned}
\mu_j &= \lambda \frac{j}{N}(1 - \frac{j}{N})(1 - \gamma_1 - \gamma_2)\Big(1 + \frac{N\gamma_1}{(1 - \gamma_1 - \gamma_2)(N - j)}\Big) \\
&= \lambda \frac{j}{N}(1 - \frac{j}{N})(1 - \gamma_1 - \gamma_2)\Big(1 + \frac{b}{N - j}\Big).
\end{aligned}$$

Therefore the representation for π_k simplifies to

$$\log \pi_k = \underbrace{\log \lambda\gamma_2}_{I} + \underbrace{\sum_{j=1}^{k-1} \log(1 + \frac{a}{j}}_{II} - \underbrace{\sum_{j=1}^{k} \log \frac{b}{N - j}}_{III} - \underbrace{\log \lambda \frac{k(N-k)}{N^2}(1 - \gamma_1 - \gamma_2)}_{IV}.$$

Now calculating the difference $I - IV$ we get $\log \frac{aN}{k(N-k)}$ and for the parts II and III we use the Taylor expansion of $\log(1 + x)$

$$II = \sum_{j=1}^{k-1} \log(1 + \frac{a}{j}) = a \sum_{j=1}^{k-1} \frac{1}{j} + c'_k \approx a \log_k + c_k,$$

where c'_k and c_k are converging sequences. Similar we get

$$
\begin{aligned}
III &= \sum_{j=1}^{k} \log(1 + \frac{b}{N-j}) \\
&= b\sum_{j=1}^{k} \frac{1}{N-j} + d'_k = b\Big(\frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{N-k}\Big) + d'_k \\
&= b(\log N - \log(N-k)) + d_k
\end{aligned}
$$

and therefore

$$
\log \pi_k = a \log k - b \log \frac{N}{N-k} + \log \frac{aN}{k(N-k)} + c_k - d_k.
$$

Now for the limit $N \to \infty$ we have $a \to \varepsilon_2$ and $b \to \varepsilon_1$ and hence with $C_k = e^{c_k - d_k}$ we conclude

$$
\pi_k = C_k a k^{a-1} (1 - \frac{k}{N})^{b-1} = C_k a N^{a-1} \Big(\frac{k}{N}\Big)^{a-1} \Big(1 - \frac{k}{N}\Big)^{b-1}.
$$

Therefore we have

$$
\frac{\pi_k}{N^{a-1}} \xrightarrow{N\to\infty} \varepsilon_2 C_k x^{\varepsilon_2 - 1}(1-x)^{\varepsilon_1 - 1}
$$

and for the sum over all entries of π we get

$$
\frac{1}{N^a} \sum_{k=1}^{N-1} \pi_k = \underbrace{\frac{a}{N} \sum_{k=1}^{N-1} C_k \Big(\frac{k}{N}\Big)^{a-1} \Big(1 - \frac{k}{N}\Big)^{b-1}}_{\text{Riemann sum}} \xrightarrow{N\to\infty} aC \int_0^1 x^{\varepsilon_2 - 1}(1-x)^{\varepsilon_1 - 1} \mathrm{d}x,
$$

which is also known as Euler's beta integral.

5.10 Queuing (waiting lines)

We look at the random arrival of custommers at some type of counter (a taxi stand, a post office,...) and are interested in the waiting time for being served. The random variable $X(t)$ counts the length of the queue. We assume $\lambda_i = \lambda$ and $\mu_i = \mu$, so neither the arrival nor the departure depends on the length of the queue. The handling time for one customer shall be exponentially distributed with mean $1/\mu$. If λ is smaller than μ, there exists a stationary distribution. We have

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n,$$

so the number of customers waiting, given by

$$p_n = \frac{\pi_n}{\sum_i \pi_i} = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n$$

is geometrically distributed.

Exercise 54. *As in exercise 54 find a differential equation for the mean $M(t)$ and solve it.*

Exercise 55. *Assume $\lambda = \mu$ and compute $P_n(t) = P(X(t) = n)$.*

In our next example we consider an infinite server queue where each customer is served immediately. The rates are given by $\lambda_n = \lambda$ and $\mu_n = n\mu$. Looking for the stationary distribution we get $\pi_n = \frac{\lambda^n}{n!\mu^n}$ and if we normalize it we have

$$p_n = \frac{\frac{\lambda^n}{n!\mu^n}}{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!\mu^k}} = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n e^{-\frac{\lambda}{\mu}},$$

which is a Poisson distribution.

Exercise 56. *Show that the mean for the infinite server queue fulfills the differential equation $M'(t) = \lambda - \mu M(t)$ and solve it.*

If we assume we have a fixed number of N servers, our rate is given by $\lambda_n = \lambda$ and

$$\mu_n = \begin{cases} n\mu \text{ if } n \leq N \\ N\mu \text{ if } n \geq N. \end{cases}$$

Therefore we get

$$\pi_n = \begin{cases} \frac{\lambda^n}{n!\mu^n} \text{ if } n \leq N \\ \frac{\lambda^n}{N!N^{n-N}\mu^n} \text{ if } n \leq N. \end{cases}$$

Now the quotient criterion tells us that the sum over all π_k is finite if and only if the quotient $\frac{\lambda}{N\mu}$ is smaller than 1, i.e.

$$N > \frac{\lambda}{\mu}.$$

Only in this case there exists an stationary distribution.

In our last example we look at N machines working independently. With rate λ one of them brakes down. The repair time is exponentially distributed with parameter μ. For the number of broken machines we have $\lambda_n = (N-n)\lambda$ since $N-n$ working machines remain. Furthermore we have

$$\mu_n = \begin{cases} 0 \text{ if } n = 0 \\ \mu \text{ else.} \end{cases}$$

The stationary distribution is now given by

$$\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} = \frac{\lambda_0 \cdots \lambda_{N-1}}{\lambda_n \cdots \lambda_{N-1}\mu_1 \cdots \mu_n} = \frac{N!\lambda^N}{(N-n)!\lambda^{N-n}\mu^n} = \frac{N!\lambda^n}{(N-n)!\mu^n}.$$

The sum over the π_n is given by

$$\begin{aligned} \sum_{n=0}^{N} \pi_n &= N! \sum_{k=0}^{N} \frac{1}{(N-n)!}\Big(\frac{\lambda}{\mu}\Big)^n = N!\Big(\frac{\lambda}{\mu}\Big)^N \sum_{n=0}^{N} \frac{1}{(N-n)!}\Big(\frac{\lambda}{\mu}\Big)^{n-N} \\ &= N!\Big(\frac{\lambda}{\mu}\Big)^N \sum_{k=0}^{N} \frac{1}{k!}\Big(\frac{\mu}{\lambda}\Big)^k. \end{aligned}$$

At least the probability that all machines break down gives a nice result known as **Erlang's loss formula**

$$p_N = \frac{1}{1 + \frac{\mu}{\lambda} + \frac{1}{2!}(\frac{\mu}{\lambda})^2 + \dots}.$$

5.11 Irreducible Markov process with finite state space

In this section we want to show a analogue statement to the ergodic theorem 3.1 for finite Markov chains but now with continuous time. Let $Q = (q_{ij})$ be an $N \times N$ matrix of transition rates with

(a) all q_{ij} non-negative if $i \neq j$ and

(b) all rows sum up to 0.

Definition 5.2.
The matrix Q is called irreducible if for all i and j there exists some k and $i_1, i_2, \dots, i_{k-1}$ all different such that

$$q_{i,i_1} q_{i_1,i_2} \cdots q_{i_{k-1},j} > 0.$$

Theorem 5.3. *Let Q be an irreducible $N \times N$ matrix fulfilling (a) and (b). Then*

1. *$P(t) = e^{Qt} > 0$ for $t > 0$ is a positive semigroup of stochastic matrices.*
2. *There is a unique invariant probability distribution $u \in \triangle_N$ with $u_i > 0$ for all i and $\sum u_i = 1$ such that $u^T Q = 0$.*
3. *The probability $p_{ij}(t)$ tends to u_j for $t \to \infty$ for all i and j.*

Remark that in contrast to the discrete time case we have no problems with periodic behavior in continuous time.

Proof. 1.) Because of $Q\mathbf{1} = 0$ we have

$$P(t)\mathbf{1} = e^{Qt}\mathbf{1} = (\mathrm{Id} + Qt + \frac{1}{2!}Q^2t^2 + \dots)\mathbf{1} = \mathrm{Id}\mathbf{1} + Q\mathbf{1}t + \frac{1}{2!}Q^2\mathbf{1}t^2 + \dots = \mathbf{1},$$

therefore the row sums of $P(t)$ are 1. Furthermore we have

$$P(t+s) = e^{Q(t+s)} = e^{Qt}e^{Qs} = P(t)P(s).$$

To show that the entries are positive let c be a lower bound such that $q_{i,i} > -c$ for all i . Hence $Q + c\mathrm{Id}$ is positive. Now

$$P(t) = e^{Qt} = e^{-ct}e^{(Q+c\mathrm{Id})t} = e^{-ct}\sum_{n=0}^{\infty}\frac{(Q+c\mathrm{Id})^n t^n}{n!} \geq 0.$$

In addition we know that since Q is irreducible, $Q+c\mathrm{Id}$ is just as well (because the diagonal entries are not used in the definition). Therefore we know for every i and j there is some n such that $(Q+c\mathrm{Id})^n_{ij}$ is positive.

□

References

[1] William Feller, *An Introduction to Probability Theory an Its Applications.* John Wiley & Sons Inc, New York, Second Edition, 1950.

[2] K. B. Athreya and P. E. Ney, *Branching Processes.*

[3] Samuel Karlin and Howard M. Taylor, *A First Course in Stochastic Processes.*