MA3H2 Markov Processes and Percolation theory

Stefan Adams

2010, **update 16.05.2010**

Contents

Preface

Any remarks and suggestions for improvements would help to create better notes for the next year.

Stefan Adams

Motivation

0 Basic facts of probability theory

0.1 Probability measure

0.2 Random variables

0.3 Limit theorems

1 Simple random walk

1.1 Nearest neighbour random walk on $\mathbb{Z}$

Pick $p \in (0,1)$, and suppose that $(X_n : n \in \mathbb{N})$ is a sequence (family) of $\{-1,+1\}$-valued, identically distributed Bernoulli random variables with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = 1 - p = q$ for all $i \in \mathbb{N}$. That is, for any $n \in \mathbb{N}$ and sequence $E = (e_1, \dots, e_n) \in \{-1,1\}^n$,

$$\mathbb{P}(X_1 = e_1, \dots, X_n = e_n) = p^{N(E)} q^{n-N(E)},$$

where $N(E) = \sharp\{m : e_m = 1\} = \frac{n + \sum_{m=1}^n e_m}{2}$ is the number of "1"s in the sequence E.

Imagine a walker moving randomly on the integers $\mathbb{Z}$. The walker starts at $a \in \mathbb{Z}$ and at every integer time $n \in \mathbb{N}$ the walker flips a coin and moves one step to the right if it comes up heads ($\mathbb{P}(\{\text{head}\}) = \mathbb{P}(X_n = 1) = p$) and moves one step to the left if it comes up tails. Denote the position of the walker at time n by S_n. The position S_n is a random variable, it depends on the outcome of the n flips of the coin. We set

$$S_0 = a \text{ and } S_N = S_0 + \sum_{i=1}^n X_i. \tag{1.1}$$

Then $S = (S_n)_{n \in \mathbb{N}}$ is often called a *nearest neighbour random walk on* $\mathbb{Z}$. The random walk is called *symmetric* if $p = q = \frac{1}{2}$. We may record the motion of the walker as the set $\{(n, S_n) : n \in \mathbb{N}_0\}$ of Cartesian coordinates of points in the plane (x-axis is the time and y-axis is the position S_n). We write $\mathbb{P}_a$ for the conditional probability $\mathbb{P}(\cdot | S_0 = a)$ when we set $S_0 = a$ implying $\mathbb{P}(S_0 = a) = 1$. It will be clear from the context which deterministic starting point we consider.

Lemma 1.1 *(a) The random walk is spatially homogeneous, i.e.,* $\mathbb{P}_a(S_n = j) = \mathbb{P}_{a+b}(S_n = j + b), j, b \in \mathbb{Z}$.

(b) The random walk is temporally homogeneous, i.e., $\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_{n+m} = j | S_m)$.

(c) Markov property

$$\mathbb{P}(S_{m+n} = j | S_0, S_1, \dots, S_m) = \mathbb{P}(S_{m+n} = j | S_m), n \geq 0.$$

Proof. (a) $\mathbb{P}_a(S_n = j) = \mathbb{P}_a(\sum_{i=1}^n X_i = j - a) = \mathbb{P}_{a+b}(\sum_{i=1}^n X_i = j - a)$.

(b)

$$\text{LHS } = \mathbb{P}(\sum_{i=1}^{n} X_i = j-a) = \mathbb{P}(\sum_{i=m+1}^{m+n} X_i = j-a) = \text{ RHS.}$$

(c) If one knows the value of S_m, then the distribution of S_{m+n} depends only on the jumps $X_{m+1}, \ldots, X_{m+n}$, and cannot depend on further information concerning the values of $S_0, S_1, \ldots, S_{m-1}$. □

Having that, we get the following stochastic process oriented description replacing (1.1),

$$\mathbb{P}(S_0 = a) = 1 \text{ and } \mathbb{P}(S_n - S_{n-1} = e | S_0, \ldots, S_{n-1}) = \begin{cases} p \text{, if } e = 1 \\ q \text{, if } e = -1 \end{cases} . \quad (1.2)$$

Markov property: conditional upon the present, the future does not depend on the past.

The set of realizations of the walk is the set of sequences $\boldsymbol{S} = (\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots)$ with $\boldsymbol{s}_0 = a$ and $\boldsymbol{s}_{i+1} - \boldsymbol{s}_i = \pm 1$ for all $i \in \mathbb{N}_0$, and such a sequence may be thought of as a sample path of the walk, drawn as in figure 1.

Let us assume that $S_0 = 0$ and $p = \frac{1}{2}$. The following question arise.
- How far does the walker go in n steps?
- Does the walker always return to the starting point, or more generally, is every integer visited infinitely often by the walker?

We get easily that $\mathbb{E}(S_n) = 0$ as we assume $p = \frac{1}{2}$. For the average distance from the origin we compute the squared position at time n, i.e.,

$$\mathbb{E}(S_n^2) = \mathbb{E}\big((X_1 + \cdots + X_n)^2\big) = \sum_{j=1}^{n} \mathbb{E}(X_j^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j).$$

Now $X_j^2 = 1$ and the independence of the X_i's gives $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j) = 0$ whenever $i \neq j$. Hence, $\mathbb{E}(S_n^2) = n$, and the expected distance from the origin is $\sim c\sqrt{n}$ for some constant $c > 0$. In order to get more detailed information of the random walk at a given time n we consider the set of possible sample paths. The probability that the first n steps of the walk follow a given path $\boldsymbol{S} = (\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$ is $p^r q^l$, where
$r = \sharp$ of steps of $\boldsymbol{S}$ to the right $= \sharp\{i \colon \boldsymbol{s}_{i+1} - \boldsymbol{s}_i = 1\}$
$l = \sharp$ of steps of $\boldsymbol{S}$ to the left $= \sharp\{i \colon \boldsymbol{s}_{i+1} - \boldsymbol{s}_i = -1\}$.

Hence, any event for the random walk may be expressed in terms of an appropriate set of paths.

$$\mathbb{P}(S_n = b) = \sum_r M_n^r(a,b) p^r q^{n-r},$$

where $M_n^r(a,b)$ is the number of paths $(\boldsymbol{s}_0, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$ with $\boldsymbol{s}_0 = a$ and $\boldsymbol{s}_n = b$ having exactly r rightward steps. Note that $r + l = n$ and $r - l = b - a$. Hence,

$$r = \frac{1}{2}(n + b - a) \text{ and } l = \frac{1}{2}(n - b + a).$$

If $\frac{1}{2}(n + b - a) \in \{0, 1, \ldots, n\}$, then

$$\mathbb{P}(S_n = b) = \binom{n}{\frac{1}{2}(n + b - a)} p^{\frac{1}{2}(n+b-a)} q^{\frac{1}{2}(n-b+a)}, \tag{1.3}$$

and $\mathbb{P}(S_n = b) = 0$ otherwise, since there are exactly $\binom{n}{r}$ paths with length n having r rightward steps and $n - r$ leftward steps. Thus to compute certain random walk events were are after counting corresponding set of paths. The following result is an important tool for this counting.

Notation: $N_n(a, b) = \sharp$ of possible paths from $(0, a)$ to (n, b). We denote by $N_n^0(a, b)$ the number of possible paths from $(0, a)$ to (n, b) which touch the origin, i.e., which contain some point $(k, 0), 1 \leq k < n$.

Theorem 1.2 (The reflection principle) *If $a, b > 0$ then*

$$N_n^0(a, b) = N_n(-a, b).$$

Proof. Each path from $(0, -a)$ to (n, b) intersects the x-axis at some earliest point $(k, 0)$. Reflect the segment of the path with times $0 \leq m \leq k$ in the x-axis to obtain a path joining $(0, a)$ and (n, b) which intersects/touches the x-axis, see figure 2. This operation gives a one-one correspondence between the collections of such paths, and the theorem is proved. □

Lemma 1.3

$$N_n(a, b) = \binom{n}{\frac{1}{2}(n + b - a)}.$$

Proof. Choose a path from $(0, a)$ to (n, b) and let α and β be the numbers of positive and negative steps, respectively, in this path. Then $\alpha + \beta = n$ and $\alpha - \beta = b - a$, so that $\alpha = \frac{1}{2}(n + b - a)$. Now the number of such paths is exactly the number of ways picking α positive steps out of n available steps. Hence,

$$N_n(a, b) = \binom{n}{\alpha}.$$

□

Corollary 1.4 (Ballot theorem) *If $b > 0$ then the number of paths from $(0, 0)$ to (n, b) which do not revisit the x-axis (origin) equals $\frac{b}{n} N_n(0, b)$.*

Proof. The first step of all such paths is to $(1,1)$, and so the number of such paths is

$$\begin{aligned} N_{n-1}(1,b) - N^0_{n-1}(1,b) &= N_{n-1}(1,b) - N_{n-1}(-1,b) \\ &= \binom{n-1}{\frac{1}{2}(n-1+b-1)} - \binom{n-1}{\frac{1}{2}(n-1+b+1)}. \end{aligned}$$

Elementary computations give the result. □

What can be deduced from the reflection principle? We first consider the probability that the walk does not revisit its starting point in the first n steps.

Theorem 1.5 *Let $S_0 = 0$ and $p \in (0,1)$. Then*

$$\mathbb{P}(S_1S_2\cdots S_n \neq 0, S_n = b) = \frac{|b|}{n}\mathbb{P}(S_n = b),$$

implying $\mathbb{P}(S_1S_2\cdots S_n \neq 0) = \frac{1}{n}\mathbb{E}(|S_n|)$.

Proof. Pick $b > 0$. The possible paths do not visit the x-axis in the time interval $[1,n]$, and the number of such paths is by the Ballot theorem exactly $\frac{b}{n}N_n(0,b)$, and each path has $\frac{1}{2}(n+b)$ rightward and $\frac{1}{2}(n-b)$ leftward steps.

$$\mathbb{P}(S_1S_2\cdots S_n \neq 0, S_n = b) = \frac{b}{n}N_n(0,b)p^{\frac{1}{2}(n+b)}q^{\frac{1}{2}(n-b)} = \frac{b}{n}\mathbb{P}(S_n = b).$$

The case for $b < 0$ follows similar, and $b = 0$ is obvious. □

Surprisingly, the last expression can be used to get the probability that the walk reaches a new maximum at a particular time. Denote by

$$M_n = \max\{S_i \colon 0 \leq i \leq n\}$$

the maximum value up to time n ($S_0 = 0$).

Theorem 1.6 (Maximum and hitting time theorem) *Let $S_0 = 0$ and $p \in (0,1)$.*

(a) For $r \geq 1$ it follows that

$$\mathbb{P}(M_n \geq r, S_n = b) = \begin{cases} \mathbb{P}(S_n = b) & \text{if } b \geq r \\ (\frac{q}{p})^{r-b}\mathbb{P}(S_n = 2r-b) & \text{if } b < r \end{cases}.$$

(b) The probability $f_b(n)$ that the walk hits b for the first time at the n-th step is

$$f_b(n) = \frac{|b|}{n}\mathbb{P}(S_n = b).$$

Proof. (a) The case $b \geq r$ is clear. Assume $r \geq 1$ and $b < r$. Let $N_n^r(0,b)$ denote the number of paths from $(0,0)$ to (n,b) which include some point having height r (i.e., some point (i,r) with time $0 < i < n$). Call such a path π and (i_π, r) the earliest such hitting point of the height r. Now reflect the segment with times larger than i_π in the horizontal height axis (x-axis shifted in vertical direction by r), see figure 3. The reflected path π' (with his segment up to time i_π equal to the one of π) is a path joining $(0,0)$ and $(n, 2r-b)$. Here, $2r-b$ is the result of $b+2(r-b)$ which is the terminal point of π'. There is again a one-one correspondence between paths $\pi \leftrightarrow \pi'$, and hence $N_n^r(0,b) = N_n(0, 2r-b)$. Thus,

$$\begin{aligned}\mathbb{P}(M_{n-1} \geq r, S_n = b) &= N_n^r(0,b) p^{\frac{1}{2}(n+b)} q^{\frac{1}{2}(n-b)} \\ &= (\frac{q}{p})^{r-b} N_n(0, 2r-b) p^{\frac{1}{2}(n+2r-b)} q^{\frac{1}{2}(n-2r+b)} \\ &= (\frac{q}{p})^{r-b} \mathbb{P}(S-n = b).\end{aligned}$$

(b) Pick $b > 0$ (the case for $b < 0$ follows similar). Then, using (a) we get

$$\begin{aligned}f_b(n) &= \mathbb{P}(M_{n-1} = S_{n-1} = b-1, S_n = b) \\ &= p\big(\mathbb{P}(M_{n-1} \geq b-1, S_{n-1} = b-1) - \mathbb{P}(M_{n-1} \geq b, S_{n-1} = b-1)\big) \\ &= p\big(\mathbb{P}(S_{n-1} = b-1) - (\frac{q}{p})\mathbb{P}(S_{n-1} = b+1)\big) = \frac{b}{n}\mathbb{P}(S_n = b).\end{aligned}$$

□

1.2 How often random walkers return to the origin?

We are going to discuss in an heuristic way the question how often the random walker returns to the origin. The walker always moves from an even integer to an odd integer or from an odd integer to an even integer, so we know for sure the position S_n of the walker is at an even integer if n is even or an at an odd integer if n is odd.

Example. Symmetric Bernoulli random walk, $p = \frac{1}{2}, S_0 = 0$**:**

$$\mathbb{P}(S_{2n} = 2j) = \binom{2n}{n+j} 2^{-2n} = 2^{-2n} \frac{(2n)!}{(n+j)!(n-j)!}, \quad j \in \mathbb{Z},$$

and in particular

$$\mathbb{P}(S_{2n} = 0) = 2^{-2n} \frac{(2n)!}{n!n!},$$

and with Stirling's formula

$$n! \sim n^n \mathrm{e}^{-n}\sqrt{2\pi n},$$

we finally get

$$\mathbb{P}(S_{2n}=0) = 2^{-2n}\frac{(2n)!}{n!n!} \sim 2^{-2n}\frac{2^{2n}}{\sqrt{\pi}\sqrt{n}} = \frac{1}{\sqrt{\pi n}}.$$

We know that the walker tends to go a distance about a constant times $\sqrt{n}$, and there are about $\sqrt{n}$ points that are in distance within $\sqrt{n}$ from the origin. Consider the following random variable, namely

$$\begin{aligned} R_n &= \sharp \text{ of visits to the origin up through time } 2n \\ &= Y_0 + Y_1 + \cdots + Y_n, \end{aligned}$$

where Y_j are Bernoulli variables defined by $Y_j = 1$ if $S_{2j} = 0$ and $Y_j = 0$ if $S_{2j} \neq 0$. We compute easily $\mathbb{E}(Y_j) = \mathbb{P}(Y_j = 1) + 0\mathbb{P}(Y_j = 0) = \mathbb{P}(S_{2j} = 0)$ and (invoke integral approximation for the sum)

$$\mathbb{E}(R_n) = \mathbb{E}(Y_0) + \cdots + \mathbb{E}(Y_n) = \sum_{j=0}\mathbb{P}(S_{2j}=0) \sim 1 + \sum_{j=1}^{n}\frac{1}{\sqrt{\pi}}j^{-\frac{1}{2}} \sim \frac{2n^{\frac{1}{2}}}{\sqrt{\pi}}.$$

Hence, the number of expected visits to the origin goes to infinity as $n \to \infty$. $\diamond$
What happens in higher dimensions?

Let's consider $\mathbb{Z}^d, d \geq 1$, and $x \in \mathbb{Z}^d$, $x = (x^1, \ldots, x^n)$. We study the simple random walk on $\mathbb{Z}^d$. The walker starts at the origin and at each integer time n he moves to one of the nearest neighbours with equal probability. Nearest neighbour refers here to the Euclidean distance, $|x| = \big(\sum_{i=1}^d (x^i)^2\big)^{1/2}$, and any lattice site in $\mathbb{Z}^d$ has exactly $2d$ nearest neighbours. Hence, the walkers jumps with probability $\frac{1}{2d}$ to one of its nearest neighbours. Denote the position of the walker after $n \in \mathbb{N}$ time steps by $S_n = (S_n^{(1)}, \ldots, S_n^{(d)})$ and write $S_n = X_1 + \cdots + X_n$, where $X_i = (X_i^{(1)}, \ldots, X_i^{(d)})$ are independent random vectors with

$$\mathbb{P}(X_i = y) = \frac{1}{2d}$$

for all $y \in \mathbb{Z}^d$ with $|y| = 1$, i.e., for all $y \in \mathbb{Z}^d$ that are in distance one from the origin. We compute similarly as above

$$\mathbb{E}(|S_n|^2) = \mathbb{E}((S_n^{(1)})^2 + \cdots + (S_n^{(d)})^2) = d\mathbb{E}((S_n^{(1)})^2),$$

and

$$\mathbb{E}((S_n^{(1)})^2) = \sum_{j=1}^{n}\mathbb{E}((X_j^{(1)})^2) + \sum_{i\neq j}\mathbb{E}(X_i^{(1)}X_j^{(1)}).$$

The probability that the walker moves within the first coordinate (either $+1$ or -1) is $\frac{1}{d}$, thus $\mathbb{E}((X_j^{(1)})^2) = \frac{1}{d}$ and $\mathbb{E}(|S_n|^2) = n$. Consider again an even time $2n$ and take n sufficiently large, then (law of large number, local central limit theorem) approximately $\frac{2n}{d}$ expected steps will be done by the walker in each of the d component directions. To be at the origin after $2n$ steps, the walker will have had to have an even number of steps in each of the d component directions. Now for n large the probability for this happening is about $(\frac{1}{2})^{d-1}$. Whether or not an even number of steps have been taken in each of the first $d-1$ component directions are almost independent events; however, we know that if an even number of steps have been taken in the first $d-1$ component directions then an even number of steps have been taken in the last component as well since the total number of steps taken is even. $\frac{2n}{d}$ steps in each component direction gives $\mathbb{P}(S_{2n}^{(i)} = 0) \sim \sqrt{\frac{d}{\pi}}\frac{1}{\sqrt{2n}}$, $i = 1, \ldots, d$. Hence,

$$\mathbb{P}(S_{2n} = 0) \sim 2^{1-d}\Big(\sqrt{\frac{d}{\pi}}\frac{1}{\sqrt{2n}}\Big)^d = \big(\frac{d^{d/2}}{2^{d-1}2^{d/2}\pi^{d/2}}\big)n^{-d/2}.$$

We know that the mean distance is $\sqrt{n}$ from the origin, and there are about $n^{d/2}$ points in $\mathbb{Z}^d$ that are within distance $\sqrt{n}$ from the origin. Hence, we expect that the probability of chosing a particular one would be of order $n^{-d/2}$). As in $d = 1$ the expected number of visits to the origin up to time n is

$$\mathbb{E}(R_n) = \sum_{j=0}^{n} \mathbb{P}(S_{2j} = 0) \le 1 + const \sum_{j=1}^{\infty} j^{-d/2} < \infty,$$

and it is finite as $n \to \infty$ for dimension $d \ge 3$. In the two-dimensional case one obtains (again integral approximation),

$$\mathbb{E}(R_n) = \sum_{j=0}^{n} \mathbb{P}(S_{2n} = 0) \sim 1 + const \sum_{j=1}^{n} \frac{1}{j} \sim \log n.$$

1.3 Transition function

We study random walks on $\mathbb{Z}^d$ and connect them to a particular function, the so-called transition function or transition matrix. For each pair x and y in $\mathbb{Z}^d$ we define a real number $P(x, y)$, and this function will be called transition function or transition matrix.

Definition 1.7 (Transition function/matrix) *Let $P\colon \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$ be given such that* *(i)* $0 \leq P(x,y) = P(0, y-x)$ *for all* $x, y \in \mathbb{Z}^d$*,*

(ii) $\sum_{x \in \mathbb{Z}^d} P(0,x) = 1$.

The function P *is called* **transition function** *or* **transition matrix** *on* $\mathbb{Z}^d$.

It will turn out that this function actually determines completely a random walk on $\mathbb{Z}^d$. That is, we are now finished - not in the sense that there is no need for further definitions, for there is, but in the sense that all further definitions will be given in terms of P. How is a random walk $S = (S_n)_{n \in \mathbb{N}_0}$ connected with a transition function (matrix)? We consider random walks which are homogeneous in time, that is

$$\mathbb{P}(S_{n+1} = j | S_n = i) = \mathbb{P}(S_1 = j | S_0 = i).$$

This motivates to define

$$P(x,y) = \mathbb{P}(S_{n+1} = y | S_n = x), \quad \text{for all } x, y \in \mathbb{Z}^d. \tag{1.4}$$

Hence, $P(0,x)$ corresponds to our intuitive notion of the probability of a 'one-step' transition from 0 to x. Then it is useful to define $P_n(x,y)$ as the 'n-step' transition probability, i.e., the probability that a random walker (particle) starting at the origin 0 finds itself at x after n transitions (time steps) governed by P.

Example. Bernoulli random walk: The n-step transition probability is given as

$$P_n(0,x) = p^{(n+x)/2} q^{(n-x)/2} \binom{n}{(n+x)/2}$$

when $n + x$ is even, $|x| \leq n$, and $P_n(0,x) = 0$ otherwise. ◇

Example. Simple random walk in $\mathbb{Z}^d$: Any lattice site in $\mathbb{Z}^d$ has exactly $2d$ nearest neighbours. Hence, the transition function (matrix) reads as

$$P(0,x) = \begin{cases} \frac{1}{2d}, & \text{if } |x| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

◇

Notation 1.8 *The* n*-step transition function (matrix) of the a random walk* $S = (S_n)_{n \in \mathbb{N}_0}$ *is defined by*

$$P_n(x,y) = \mathbb{P}(S_{m+n} = x | S_m = y), \quad m \in \mathbb{N}, x, y \in \mathbb{Z}^d,$$

and we write $P_1(x,y) = P(x,y)$ *and* $P_0(x,y) = \delta_{x,y}$.

The n-step transition function can be written as

$$P_n(x,y) = \sum_{x_i\in\mathbb{Z}^d, i=1,\dots,n-1} P(x,x_1)P(x_1,x_2)\cdots P(x_{n-1},y), \quad n\ge 2. \tag{1.5}$$

This is proved in the following statement.

Theorem 1.9 *For any pair $r,s\in\mathbb{N}_0$ satisfying $r+s=n\in\mathbb{N}_0$ we have*

$$P_n(x,y)=\sum_{z\in\mathbb{Z}^d} P_r(x,z)P_s(z,y), \quad x,y\in\mathbb{Z}^d.$$

Proof. The proof for $n=0,1$ is clear. We give a proof for $n=2$. Induction will give the proof for the other cases as well. The event of going from x to y in two transitions (time steps) can be realised in the mutually exclusive ways of going to some intermediate lattice site $z\in\mathbb{Z}^d$ in the first transition and then going from site $z\in\mathbb{Z}^d$ to y in the second transition. The Markov property implies that the probability of the second transition is $P(z,y)$, and that of the first transition is clearly $P(x,z)$. Using the Markov property and the relation

$$\mathbb{P}(A\cap C|C)=\mathbb{P}(A|B\cap C)\mathbb{P}(B|C),$$

we get for any $m\in\mathbb{N}$,

$$\begin{aligned} P_2(x,y) &= \mathbb{P}(S_{m+2}=x|S_m=y) = \sum_{z\in\mathbb{Z}^d}\mathbb{P}(S_{m+2}=x, S_{m+1}=z|S_m=y) \\ &= \sum_{z\in\mathbb{Z}^d}\mathbb{P}(S_{m+2}=x|S_{m+1}=z, S_m=y)\mathbb{P}(S_{m+1}=z|S_m=y) \\ &= \sum_{z\in\mathbb{Z}^d}\mathbb{P}(S_{m+2}=x|S_{m+1}=z)\mathbb{P}(S_{m+1}=z|S_m=y) \\ &= \sum_{z\in\mathbb{Z}^d} P(x,z)P(z,y). \end{aligned}$$

□

The probability interpretation of $P_n(x,y)$ is evident, it represents the probability that a 'particle', executing a random walk and starting at the lattice site x at time 0, will be at the lattice site $y\in\mathbb{Z}^d$ at time n. We now define a function of a similar type, namely, we are asking for the probability (starting at x at time 0), that the **first** visit to lattice site y should occur at time n.

Definition 1.10 *For all $x,y\in\mathbb{Z}^d$ and $n\ge 2$ define*

$$\begin{aligned} F_0(x,y) &:= 0, \\ F_1(x,y) &:= P(x,y), \\ F_n(x,y) &:= \sum_{\substack{x_i\in\mathbb{Z}^d\setminus\{y\}\\ i=1,\dots,n-1}} P(x,x_1)P(x_1,x_2)\cdots P(x_{n-1},y). \end{aligned}$$

Important properties of the function $F_n, n \geq 2$, are summarised.

Proposition 1.11 *For all* $x, y \in \mathbb{Z}^d$*:*

(a) $F_n(x,y) = F_n(0, y-x)$.

(b) $\sum_{k=1}^n F_k(x,y) \leq 1$.

(c) $P_n(x,y) = \sum_{k=1}^n F_k(x,y) P_{n-k}(y,y)$.

Proof. (a) is clear from the definition and from the known properties of P. (b) The claim is somehow obvious. However, we shall give a proof. For $n \in \mathbb{N}$ put $\Omega_n = \{\omega = (x_0, x_1, \dots, x_n) \colon x_0 = x, x_i \in \mathbb{Z}^d, i = 1, \dots, n\}$. Clearly, Ω_n is countable, and we define a probability for any 'elementary event' $\omega \in \Omega_n$ by

$$p(\omega) := P(x, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n), \quad \omega = (x_0, x_1, \dots, x_{n-1}) \in \Omega_n.$$

Clearly, $\sum_{\omega \in \Omega \colon x_n = y} p(\omega) = P_n(x,y)$, and $\sum_{\omega \in \Omega_n} p(\omega) = \sum_{y \in \mathbb{Z}^d} P_n(x,y) = 1$. The sets A_k,

$$A_k = \{\omega \in \Omega_n \colon x_1 \neq y, x_2 \neq y, \dots, x_{k-1} \neq y, x_k = y\}, \quad 1 \leq k \leq n,$$

are disjoint subsets of Ω_n and $F_k(x,y) = \sum_{\omega \in A_k} p(\omega)$ implies that

$$\sum_{k=1}^n F_k(x,y) \leq \sum_{\omega \in \Omega_n} p(\omega) = 1.$$

(c) This can be proved in a very similar fashion, or by induction. We skip the details. □

We come up now with a third (and last) function of the type above. This time we are after the expected number of visits of a random walk to a given point within a given time. More precisely, we denote by $G_n(x,y)$ the expected number of visits of the random walk, starting at x, to the point y up to time n.

Notation 1.12

$$G_n(x,y) = \sum_{k=0}^n P_k(x,y), \quad n \in \mathbb{N}_0, x, y \in \mathbb{Z}^d.$$

One can easily convince oneself that $G_n(x,y) \leq G_n(0,0)$ for all $n \in \mathbb{N}_0, x, y \in \mathbb{Z}^d$. We can now classify the random walks according to whether they are **recurrent** or **transient** (non-recurrent). The idea is that $\sum_{k=1}^n F_k(0,0)$ represents the probability of a return to the origin before or at time n. The sequence of sums $\sum_{k=1}^n F_k(0,0)$ is non-decreasing as n increases, and by Proposition 1.11 bounded by one. Call the limit by $F \leq 1$. Further, call G the limit of the monotone sequence $(G_n(0,0))_{n \in \mathbb{N}_0}$.

Notation 1.13 *(a)* $G(x,y) = \sum_{n=0}^{\infty} P_n(x,y) \le \infty$ *for all* $x, y \in \mathbb{Z}^d$, $G := G(0,0)$.

(b) $F(x,y) = \sum_{n=1}^{\infty} F_n(x,y) \le 1$ *or all* $x, y \in \mathbb{Z}^d$, $F := F(0,0)$.

Definition 1.14 *The random walk (on* $\mathbb{Z}^d$*) defined by the transition function* P *is said to be* **recurrent** *if* $F = 1$ *and* **transient** *if* $F < 1$.

Proposition 1.15

$$G = \frac{1}{1-F} \quad \textit{with } G = +\infty \textit{ when } F = 1 \textit{ and } F = 1 \textit{ when } G = +\infty.$$

Proof. (The most convenient way is to prove it is using generating functions). We sketch a direct method.

$$P_n(0,0) = \sum_{k=0}^{n} F_k P_{n-k}(0,0), \quad n \in \mathbb{N}. \tag{1.6}$$

Summing (1.6) over $n = 1, \ldots, m$, and adding $P_0(0,0) = 1$ gives

$$G_m(0,0) = \sum_{k=0}^{m} F_k G_{m-k}(0,0) + 1, \quad m \in \mathbb{N}. \tag{1.7}$$

Letting $m \to \infty$ we get

$$G = 1 + \lim_{m\to\infty} \sum_{k=0}^{m} F_k G_{m-k} \ge 1 + G \sum_{k=0}^{N} F_k, \qquad \text{for all } N \in \mathbb{N},$$

and thus $G \ge 1 + GF$. Now (1.7) gives

$$1 = G_m - \sum_{k=0}^{m} G_k F_{m-k} \ge G_m - G_m \sum_{k=1}^{m} F_{m-k} \ge G_m(1-F),$$

and henceforth $1 \ge G(1-F)$. □

Example. Bernoulli random walk: $P(0,0) = p$, and $P(0,-1) = q = 1-p$, $p \in [0,1]$.

$$\mathbb{P}(S_{2n} = 0) = P_{2n}(0,0) = (pq)^n \binom{2n}{n} = (-1)^n (4pq)^n \binom{-\frac{1}{2}}{n},$$

where we used that

$$\binom{2n}{n} = (-1)^n 4^n \binom{-\frac{1}{2}}{n}. \tag{1.8}$$

Note that the binomial coefficients for general numbers r are defined as

$$\binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!}, \quad k \in \mathbb{N}_0.$$

We prove (1.8) by induction: For $n=1$ the LHS$= \frac{2!}{1!1!}$ and RHS$= (-1)4\frac{-1/2}{1!}$. Assumption the claim for $n \in \mathbb{N}$. Then

$$\binom{2(n+1)}{n+1} = \frac{(2n)!(2n+1)(2(n+1))}{(n+1)n!(n+1)n!} = (-1)^n 4^n \binom{-1/2}{n} 2\times$$
$$\times \frac{(2n+1)}{n+1} \frac{(-1)^n 4^n (-1/2)(-1/2-1)\cdots(-1/2-n+1)(-1)(-1/2-n)}{(n+1)!}$$
$$= (-1)^{n+1} 4^{n+1} \binom{-1/2}{n+1}.$$

Further, using Newton's generalised Binomial theorem, that is,

$$(x+y)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^{r-k} y^k, \tag{1.9}$$

we - noting that $0 \le p = 1-q$ implies that $4pq \le 1$ - get that

$$\sum_{n=0}^{\infty} t^n P_{2n}(0,0) = (1-4pqt)^{-1/2}, \quad |t| < 1.$$

Thus

$$\lim_{t\to 1, t<1} \sum_{n=0}^{\infty} t^n P_{2n}(0,0) = \sum_{n=0}^{\infty} P_{2n}(0,0) = \sum_{n=0}^{\infty} P_n(0,0) = G \le \infty,$$

henceforth

$$G = \begin{cases} (1-4pq)^{-1/2} < \infty\,, & \text{if } p \ne q, \\ +\infty\,, & \text{if } p = q. \end{cases}$$

The Bernoulli random walk (on $\mathbb{Z}$) is recurrent if and only if $p = q = \frac{1}{2}$. ◇

Example. Simple random walk in $\mathbb{Z}^d$:
The simple random walk is
$d=1$ recurrent,
$d=2$ recurrent,
$d \ge 3$ transient. ◇

1.4 Summary

The simple random walks on $\mathbb{Z}^d$ (discrete time) are examples of Markov chains on $\mathbb{Z}^d$.

Definition 1.16 *Let I be a countable set, $\lambda \in \mathcal{M}_1(I)$ be a probability measure (vector) on I, and $P = (P(i,j))_{i,j\in I}$ be a transition function (stochastic matrix). A sequence $X = (X_n)_{n\in\mathbb{N}_0}$ of random variables X_n taking values in I is called a Markov chain with state space I and transition matrix P and initial distribution λ, if*

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = P(i_n, i_{n+1}) \text{ and}$$
$$\mathbb{P}(X_0 = i) = \lambda(i), i \in I,$$

for every $n \in \mathbb{N}_0$ and every $i_0, \ldots, i_{n+1} \in I$ with $\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) > 0$. We call the family $X = (X_n)_{n\in\mathbb{N}_0}$ a (λ, P)-Markov chain.

Note that for every $n \in \mathbb{N}_0, i_0, \ldots, i_n \in I$, the probabilities are computed as

$$\mathbb{P}(X_0 = i_0, \ldots, X_n = i_n) = \lambda(i_0)P(i_0, i_1)P(i_1, i_2)\cdots P(i_{n-1}, i_n).$$

A vector $\lambda = (\lambda(i))_{i\in I}$ is called a stationary distribution of the Markov chain if the following holds:

(a) $\lambda(i) \geq 0$ for all $i \in I$, and $\sum_{i\in I} \lambda(i) = 1$.

(b) $\lambda = \lambda P$, that is, $\lambda(j) = \sum_{i\in I} \lambda(i)P(i,j)$ for all $j \in I$.

Without proof we state the following result which will we prove later in the continuous time setting.

Theorem 1.17 *Let I be a finite set and $P\colon I \times I \to \mathbb{R}_+$ be a transition function (matrix). Suppose for some $i \in I$ that*

$$P_n(i,j) \to \lambda(j) \text{ as } n \to \infty \text{ for all } j \in I.$$

Then $\lambda = (\lambda(j))_{j\in I}$ is an invariant distribution.

2 Markov processes

In this chapter we introduce continuous-time Markov processes with a countable state space. Throughout the chapter we assume that $X = (X_t)_{t\geq 0}$ is a family of random variables taking values in some countable state space I. The family $X = (X_t)_{t\geq 0}$ is called a *continuous-time random* process. We shall specify the probabilistic behaviour (or *law*) of $X = (X_t)_{t\geq 0}$. However, there are subtleties in this problem not present in the discrete-time case. They arise because, the probability of a countable disjoint union is the sum of the single probabilities, whereas for a noncountable union there is no such rule. To avoid these subtleties we shall consider only continous-time processes which are right continuous. This means that with probability one, for all $t \geq 0$, $\lim_{h\downarrow 0} X_{t+h} = X_t$. By a standard result of measure theory the probability of any event depending on a right-continuous process can be determined from its *finite-dimensional distributions*, that is, from the probabilities $\mathbb{P}(X_{t_0} = i_0, \dots, X_{t_n} = t_n)$ for $n \in \mathbb{N}_0, 0 \leq t_0 \leq \dots \leq t_n$ and $i_0, \dots, i_n \in I$. Throughout we are using both writings, X_t and $X(t)$ respectively.

Definition 2.1 *The process $X = (X_t)_{t\geq 0}$ is said to satisfy the Markov property if*

$$\mathbb{P}(X(t_n) = j | X(t_0) = i_0, \dots X(t_{n-1}) = i_{n-1}) = \mathbb{P}(X(t_n) = j | X(t_{n-1}) = i_{n-1})$$

for all $j, i_0, \dots, i_{n-1} \in I$ and any sequence $t_0 < t_1 < \dots < t_n$ of times.

We studied in the first chapter the simplest discrete time Markov process (Markov chain) having independent, identically distributed increments (Bernoulli random variables). The simplest continuous time Markov processes are those whose increments are mutually independent and homogeneous in the sense that the distribution of an increment depends only on the length of the time interval over which the increment is taken. More precisely, we are dealing with I-values stochastic processes $X = (X_t)_{t\geq 0}$ having the property that $\mathbb{P}(X_0 = x_0) = 1$ for some $X_0 \in I$ and

$$\mathbb{P}(X(t_1)-X(t_0) = i_1, \dots, X(t_n)-X(t_{n-1}) = i_n) = \prod_{m=1}^{n} \mathbb{P}(X(t_m)-X(t_{m-1}) = i_m)$$

for $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$ and all times $t_0 < t_1 < \dots < t_n$.
We introduce in the first section the Poisson process on $\mathbb{N}$. Before that we shall collect some basic facts from probability theory.

Definition 2.2 (Exponential distribution) *A random variable T having values in $[0, \infty)$ has exponential distribution of parameter $\lambda \in [0, \infty)$ if $\mathbb{P}(T > t) = \mathrm{e}^{-\lambda t}$ for all $t \geq 0$. The exponential distribution is the probability measure on $[0, \infty)$ having the (Lebesgue-) density function*

$$f_T(t) = \lambda \mathrm{e}^{-\lambda t} \mathbb{1}\{t \geq 0\}.$$

We write $T \sim E(\lambda)$ for short. The mean (expectation) of T is given by

$$\mathbb{E}(T) = \int_0^\infty \mathbb{P}(T > t)\, \mathrm{d}t = \lambda^{-1}.$$

The other important distribution is the so-called Gamma distribution. We consider random time points in the interval $(0,\infty)$ (e.g. incoming claims in an insurance company or phone calls arriving at a telephone switchboard). The heuristic reasoning is that, for every $t > 0$, the number of points in $(0,t]$ is Poisson distributed with parameter λt, where $\lambda > 0$ represents the average number of points per time. We look for a model of the r-th random point. What is the probability measure P describing the distribution of the r-th random point?
$P((0,t])$ = probability that the r-th point arrives no later than t (i.e. at least r points/arrivals in $(0,t]$). Denote by $P_{\lambda t}$ the Poisson distribution with parameter λt. We get the probability in question using the complementary event as

$$\begin{aligned} P((0,t]) &= 1 - P_{\lambda t}(\{0,\dots,r-1\}) \\ &= 1 - \mathrm{e}^{-\lambda t} \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} = \int_0^t \frac{\lambda^r}{(r-1)!} x^{r-1} \mathrm{e}^{-\lambda x} \, \mathrm{d}x. \end{aligned}$$

The last equality can be checked when differentiating with respect to t. Recall the definition of Euler's Gamma function, $\Gamma(r) = \int_0^\infty y^{r-1}\mathrm{e}^{-y}\,\mathrm{d}y, r > 0$, and $\Gamma(r) = (r-1)!$ for all $r \in \mathbb{N}$.

Definition 2.3 (Gamma distribution) *For every $\lambda, r > 0$, the probability measure $\Gamma_{\lambda,r}$ on $[0,\infty)$ with (Lebesgue-) density function*

$$\gamma_{\lambda,r}(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} \mathrm{e}^{-\lambda x}, x \geq 0,$$

is called the Gamma distribution with scale parameter λ and shape parameter r. Note that $\Gamma_{\lambda,1}$ is the exponential distribution with parameter λ.

Lemma 2.4 (Sum of exponential random variables) *If $X_i \sim E(\lambda)$, $i = 1,\dots,n$, independently, and $Z = X_1 + \cdots + X_n$ then Z is $\Gamma_{\lambda,n}$ distributed.*

Proof. Exercise of example sheet 2. □

2.1 Poisson process

In this Subsection we will introduce a basic intuitive construction of the Poisson process. The Poisson process is the backbone of the theory of Markov processes in continuous time having values in a countable state space. We will study later more general settings. Pick a parameter $\lambda > 0$ and let $(E_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. (independent identically distributed) random variables (having values in $\mathbb{R}_+$) that are exponentially distributed with parameter λ (existence of such a sequence is guaranteed - see measure theory). Now, E_i is the time gap (waiting or holding time) between the $(i-1)$-th (time) point and the i-th point. Then the sum

$$T_k = \sum_{i=1}^{k} E_i$$

is the k-th random point in time (see figure). Furthermore, let

$$N_t = \sum_{k\in\mathbb{N}} \mathbb{1}_{(0,t]}(T_k)$$

be the number of points in the interval $(0,t]$. Thus, for $s<t$, $N_t - N_s$ is the number of points in $(s,t]$. Clearly, for $t\in[T_k, T_{k+1})$ one has $N_t = k$.

Theorem 2.5 (Construction of the Poisson process) *The $N_t, t \geq 0$, are random variables having values in $\mathbb{N}_0$, and, for $0 = t_0 < t_1 < \cdots < t_n$, the increments $N_{t_i} - N_{t_{i-1}}$ are independent and Poisson distributed with parameter $\lambda(t_i - t_{i-1})$, $1 \leq i \leq n$.*

Definition 2.6 *A family $(N_t)_{t\geq 0}$ of $\mathbb{N}_0$-valued random variables satisfying the properties of Theorem 2.5 with $N_0 = N(0) = 0$ is called a* **Poisson process** *with* **intensity** *$\lambda > 0$.*

We can also write $T_k = \inf\{t > 0\colon N_t \geq k\}$, $k \geq 1$, in other words, T_k is the k-th time point at which the **sample path** $t \mapsto N_t$ of the Poisson process performs a jump of size 1. These times are therefore called **jump times** of the Poisson process, and $(N_t)_{t\geq 0}$ and $(T_k)_{k\in\mathbb{N}}$ are two manifestations of the same mathematical object.

Proof of Theorem 2.5. First note that $\{N_t = k\} = \{T_k \leq t < T_{k+1}\}$. We consider here $n = 2$ to keep the notation simple. The general case follows analogously. Pick $0 < s < t$ and $k, l \in \mathbb{N}$. It suffices to show that

$$\mathbb{P}(N_s = k, N_{t-s} = l) = \big(\mathrm{e}^{-\lambda s}\frac{(\lambda s)^k}{k!}\big)\big(\mathrm{e}^{-\lambda(t-s)}\frac{(\lambda(t-s))^l}{l!}\big). \tag{2.10}$$

Having (2.10), summing over l and k, respectively, we can then conclude that N_s and N_{t-s} are Poisson distributed (and are independent, see right hand side of (2.10)). The joint distribution of the (holding) times $(E_j)_{1\leq j\leq k+l+1}$ has the product density

$$f(x_1, \ldots, x_{k+l+1}) = \lambda^{k+l+1}\mathrm{e}^{-\lambda\tau_{k+l+1}(x)},$$

where for convenience we write $\tau_{k+l+1}(x) = x_1 + \cdots + x_{k+l+1}$. Using the equality of the events in the first line above, the left hand side of (2.10) reads as

$$
\begin{aligned}
&\mathbb{P}(N_s = k, N_{t-s} - N_s = l) = \mathbb{P}(T_k \le s < T_{k+1} \le T_{k+l} \le t < T_{k+l+1}) \\
&= \int_0^\infty \cdots \int_0^\infty \mathrm{d}x_1 \cdots \mathrm{d}x_{k+l+1} \lambda^{k+l+1} \mathrm{e}^{-\lambda \tau_{k+l+1}(x)} \\
&\times 1\!\!1\{\tau_k(x) \le s < \tau_{k+1}(x) \le t < \tau_{k+l+1}(x)\}.
\end{aligned}
$$

We integrate step by step starting from the innermost integral and moving outwards. Fix $x_1, \ldots, x_{k+l}$ and set $z = \tau_{k+l+1}(x)$,

$$
\int_0^\infty \mathrm{d}x_{k+l+1} \lambda \mathrm{e}^{-\lambda \tau_{k+l+1}(x)} 1\!\!1\{\tau_{k+l+1}(x) > t\} = \int_t^\infty \mathrm{d}z \lambda \mathrm{e}^{-\lambda z} = \mathrm{e}^{-\lambda t}.
$$

Fix $x_1, \ldots, x_k$ and make the substitution $y_1 = \tau_{k+1}(x) - s, y_2 = x_{k+2}, \ldots, y_l = x_{k+l}$ to obtain

$$
\begin{aligned}
&\int_0^\infty \cdots \int_0^\infty \mathrm{d}x_{k+1} \cdots \mathrm{d}x_{k+l} 1\!\!1\{s < \tau_{k+1}(x) \le \tau_{k+l}(x) \le t\} \\
&\quad = \int_0^\infty \cdots \int_0^\infty \mathrm{d}y_1 \cdots \mathrm{d}y_l 1\!\!1\{y_1 + \cdots + y_l \le l - s\} = \frac{(t-s)^l}{l!},
\end{aligned}
$$

which can be proved via induction on l. In a similar way one gets

$$
\int_0^\infty \cdots \int_0^\infty \mathrm{d}x_1 \cdots \mathrm{d}x_k 1\!\!1\{\tau_k(x) \le s\} = \frac{s^k}{k!}.
$$

Combing all our steps above, we obtain finally

$$
\mathbb{P}(N_s = k, N_{t-s} = l) = \mathrm{e}^{-\lambda t} \lambda^{k+l} \frac{s^k}{k!} \frac{(t-s)^l}{l!}.
$$

□

The following statement shows that the Poisson process satisfies the Markov property.

Theorem 2.7 *Let $N = (N_t)_{t \ge 0}$ be a Poisson process with intensity $\lambda > 0$, then*

$$
\mathbb{P}(N_{s+t} - N_s = k | N_\tau, \tau \in [0, s]) = \mathbb{P}(N_t = k), k \in \mathbb{N}.
$$

That is for all $s > 0$, the past $(N_\tau)_{\tau \in [0s)}$ is independent of the future $(N_{s+t} - N_s)_{t \ge 0}$. In other words, for all $s > 0$ the process after time s and counted from the level N_s remains a Poisson process with intensity λ independent of its past $(N_\tau)_{\tau \in [0s)}$.

The proof is deferred for later and the support class.

2.2 Q-matrices (generators)

In this Subsection we want to study how to construct a process $X = (X_t)_{t\geq 0}$ taking values in some countable state space I satisfying the Markov property in Definition 2.1. We will do this in several steps. First, we proceed like for Markov chains (discrete time, countable state space). In continuous time we do not have a unit length of time and hence no exact analogue of the transition function $P\colon I \times I \to [0,1]$.

Definition 2.8 (Transition probability) *Let $X = (X_t)_{t\geq 0}$ be a Markov process on a countable state space I.*

(a) The **transition probability** *$P_{s,t}(i,j)$ of the Markov process X is defined as*

$$P_{s,t}(i,j) = \mathbb{P}(X_t = j | X_s = i) \quad \text{for } s \leq t; i,j \in I.$$

(b) The Markov process X is called **homogeneous** *if*

$$P_{s,t}(i,j) = P_{0,t-s}(i,j) \quad \text{for all } i,j \in I; t \geq s \geq 0.$$

We will consider solely homogeneous Markov processes in the following, hence we write P_t for $P_{0,t}$. We write P_t for the $|I|\times|I|$-matrix given by (a) in Definition 2.8. The family $P = (P_t)_{t\geq 0}$ is called **transition semigroup** of the Markov process. For continuous time processes it can happen that rows of the transition matrix P_t do not sum up to one (see discussion below). This motivates the following definition for families of matrices on the state space I.

Definition 2.9 ((Sub-) stochastic semigroup) *A family $P = (P_t)_{t\geq 0}$ of matrices on the countable set I is called (Sub-) stochastic semigroup on I if the following conditions hold.*

(a) $P_t(i,j) \geq 0$ for all $i,j \in I$.

(b) $\sum_{j\in I} P_t(i,j) = 1$ (respectively $\sum_{j\in I} P_t(i,j) \leq 1$).

(c) **Chapman-Kolmogorov equations**

$$P_{t+s}(i,j) = \sum_{k\in I} P_t(i,k) P_s(k,j), \quad t,s \geq 0.$$

We call the family $P = (P_t)_{t\geq 0}$ **standard** *if in addition to (a)-(c)*

$$\lim_{t\downarrow 0} P_t(i,j) = \delta_{i,j} \quad \text{for all } i,j \in I$$

holds.

We show that any Markov process $X = (X_t)_{t\geq 0}$ on a finite state space I defines via Definition 2.8 a stochastic semigroup on I.

Proposition 2.10 *Let $(X_t)_{t\geq 0}$ be a right-continuous homogeneous Markov process with finite state space I. Then the following holds for $P_t(i,j) = \mathbb{P}(X_t = j|X_0 = i)$, $i,j \in I$,*

(a) $P_0 = \mathbb{1}$.

(b) The rows of P_t sum to one.

(c) $P_{s+t} = P_s P_t$ *for all* $s,t \geq 0$.

(d) $P = (P_t)_{t\geq 0}$ *is standard.*

Proof. (a) and (b) follow immediately from the definition. (d) follows from the right-continuity. We are left to show (c) which follows from the Markov property.

$$\begin{aligned} P_{s+t}(i,j) &= \mathbb{P}(X_{s+t} = j|X_0 = i) \\ &= \sum_{k\in I} \mathbb{P}(X_{s+t}|X_s = k, X_0 = i)\mathbb{P}(X_s = k|X_0 = i) \\ &= \sum_{k\in I} P_s(i,k)P_t(k,j). \end{aligned}$$

□

In a continuous time process it can happen that there are infinitely many jumps in a finite time interval. This phenomenon is called **explosion**. If explosion occurs one cannot bookmark properly the states the process visits, somehow the process is stuck. A convenient mathematical way is to add a special state, called cemetery, written as ∂, to the given state space I, i.e. to consider a new state space $I \cup \{\partial\}$. This is exactly the situation where the sub-stochastic semigroup in Definition 2.9 comes into play. Recall that if $X = (X_t)_{t\geq 0}$ is a Markov process with initial distribution ν, where ν is a probability measure on the state space I, and semigroup $(P_t)_{t\geq 0}$ the probability for times $0 = t_0 < t_1 < \cdots < t_n$ and states $i_0, \ldots, i_n \in I$ is given as

$$\mathbb{P}(X(t_0) = i_0, \ldots, X(t_n) = i_n) = \nu(i_0)P_{t_1-t_0}(i_0,i_1)\cdots P_{t_n-t_{n-1}}(i_{n-1},i_n).$$

Definition 2.11 (Markov process with explosion) *Let $P = (P_t)_{t\geq 0}$ be a sub-stocastic semigroup on a countable state space I. Further let $\{\partial\} \notin I$ and let ν be a probability measure on the augmented state space $I \cup \{\partial\}$. A $I \cup \{\partial\}$-valued family $X = (X_t)_{t\geq 0}$ is a Markov process with initial distribution ν and semigroup $(P_t)_{t\geq 0}$ if for $n \in \mathbb{N}$ and any $0 \leq t_1 < t_2 < \cdots < t_n < t$ and states $i_1, \ldots, i_n \in I$ the following holds:*

(a) $\mathbb{P}(X(t)|X(t_1) = i_1, \ldots, X(t_n) = i_n) = P_{t-t_n}(i_n|i)$ *if the left hand side is defined.*

(b) $\mathbb{P}(X_0| = i) = \nu(i)$ *for all* $i \in I \cup \{\partial\}$.

(c) $\mathbb{P}(X(t) = \partial|X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}, X(t_n) = \partial) = 1$.

We will later give conditions to ensure that a Markov process shows no explosion. The following Proposition analyses the dependence of the semigroup on the time parameter.

Proposition 2.12 *Let* $P = (P_t)_{t\geq 0}$ *be a standard stochastic semigroup on a countable state space* I*. then*

(a) $\lim_{h\downarrow 0} \sup_{t\geq 0} \sum_{j\in I} |P_{t+h}(i,j) - P_t(i,j)| = 0$ *for all* $i, j \in I$.

(b) For all $i, j \in I$ *the differential quotient*

$$q_{i,j} := \lim_{h\downarrow 0} \frac{P_h(i,j) - \delta_{i,j}}{h}$$

exists and $0 \leq q_{i,j} < \infty$ *for* $i \neq j$ *and* $0 \geq q_{i,i} \geq -\infty$*. Moreover,*

$$\sum_{j\in I, j\neq i} q_{i,j} \leq -q_{i,i} =: q_i.$$

Proof. (a) Using the semigroup property (Markov property) we get

$$\begin{aligned}
\sum_{j\in I} \big|P_{t+h}(i,j) - P_t(i,j)\big| &= \sum_{j\in I} \big|\sum_{k\in I} P_h(i,k)P_t(k,j) - \delta_{i,k}P_t(k,j)\big| \\
&\leq \sum_{k\in I}\sum_{j\in I} \big|P_h(i,k) - \delta_{i,k}\big|P_t(k,j) \\
&= \sum_{k\in I} \big|P_h(i,k) - \delta_{i,k}\big| \sum_{j\in I} P_t(k,j) \\
&\leq \sum_{k\in I} \big|P_h(i,k) - \delta_{i,k}\big| = 1 - P_h(i,i) + \sum_{k\in I\setminus\{i\}} P_h(i,k) \\
&\leq 2(1 - P_h(i,i)),
\end{aligned}$$

and hence as $h \downarrow 0$ the statement follows as semigroup is standard. (b) We skip the proof. It is similar and quite long. □

The positive entries $q_{i,j}$ are called **transition rates** if $i \neq j$, and $q_i = -q_{i,i}$ is called the rate leaving state i. The results of Theorem 2.12 (b) motivate the following definition.

Definition 2.13 *A* **Q-matrix** *or* **generator** *on a countable state space I is a matrix $Q = (q_{i,j})_{i,j \in I}$ satisfying the following conditions:*

(a) $0 \le -q_{i,i} < \infty$ *for all* $i \in I$.

(b) $q_{i,j} \ge 0$ *for all* $i \ne j, i, j \in I$.

(c) $\sum_{j \in I} q_{i,j} = 0$ *for all* $i \in I$.

A Q-matrix is also called generator because it provides a continuous time parameter semigroup of stochastic matrices and henceforth a Markov process. In this way a Q-matrix or generator is the most convenient way in construction a Markov process in particular as the non-diagonal entries are interpreted as transition rates. Unfortunately, there is some technical difficulty in defining this connection properly when the state space is infinite. However, if the state space is finite we get the following nice results. Before that recall the definition of an exponential of a finite dimensional matrix.

Theorem 2.14 *Let I be a finite state space and $Q = (q_{i,j})_{i,j \in I}$ a generator or Q-matrix. Define $P_t = P(t) := \mathrm{e}^{tQ}$ for all $t \ge 0$. Then the following holds:*

(a) $P(s+t) = P(s)P(t)$ *for all* $s, t \ge 0$.

(b) $(P(t))_{t \ge 0}$ *is the unique solution to the* **forward equation**

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t) = P(t)Q \text{ and } P(0) = \mathbb{1}.$$

(c) $(P(t))_{t \ge 0}$ *is the unique solution to the* **backward equation**

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t) = QP(t) \text{ and } P(0) = \mathbb{1}.$$

(d) For $k \in \mathbb{N}_0$

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\Big|_{t=0} P(t) = Q^k.$$

Proof. We only give a sketch of the proof as it basically amounts to well-known basic matrix algebra. For all $s, t \in \mathbb{R}_+$, the matrices sQ and tQ commute, hence $\mathrm{e}^{sQ}\mathrm{e}^{tQ} = \mathrm{e}^{(s+t)Q}$, proving the semigroup property. The matrix-valued power series

$$P(t) = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}$$

has a radius of convergence which is infinite. Hence, one can justify a term by term differentiation (we skip that) to get

$$P'(t) = \sum_{k=1}^{\infty} \frac{t^{k-1} Q^k}{(k-1)!} = P(t)Q = QP(t).$$

We are left to show that the solution to both the forward and backward equation are unique. For that let $(M(t))_{t\geq 0}$ satisfy the forward equations (case for backward equations follows similar).

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Big(M(t)\mathrm{e}^{-tQ}\Big) = \Big(\frac{\mathrm{d}}{\mathrm{d}t}M(t)\Big)\mathrm{e}^{-tQ} + M(t)\Big(\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{-tQ}\Big) = M(t)Q\mathrm{e}^{-tQ} \\ + M(t)(-Q)\mathrm{e}^{-tQ} = 0,\end{aligned}$$

henceforth $M(t)\mathrm{e}^{-tQ}$ is constant and so $M(t) = P(t)$. □

Proposition 2.15 *Let $Q = (q_{i,j})_{i,j\in I}$ be a matrix on a finite set I. Then the following equivalence holds.*

$$Q \text{ is a } Q\text{-matrix } \Leftrightarrow P(t) = \mathrm{e}^{tQ} \text{ is a stochastic matrix for all } t \geq 0.$$

Proof. Let Q be a Q-matrix. As $t \downarrow 0$ we have $P(t) = 1\!\!1 + tQ + O(t^2)$. Hence, for sufficiently small times t the positivity of $P_t(i,j), i \neq j$, follows from the positivity of $q_{i,j} \geq 0$. For larger times t we can easily use that $P(t) = P(t/n)^n$ for any $n \in \mathbb{N}$, and henceforth

$$q_{i,j} \geq 0, i \neq j \Leftrightarrow P_t(i,j) \geq 0, i \neq j \text{ for all } t \geq 0.$$

Furthermore, if Q has zero row sums then so does $Q^n = (q_{i,j}^{(n)})_{i,j\in I}$ for every $n \in \mathbb{N}$,

$$\sum_{k\in I} q_{i,j}^{(n)} = \sum_{k\in I}\sum_{j\in I} q_{i,j}^{(n-1)} q_{j,k} = \sum_{j\in I} q_{i,j}^{(n-1)} \sum_{k\in I} q_{j,k} = 0.$$

Thus

$$\sum_{j\in I} P_t(i,j) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{j\in I} q_{i,j}^{(n)} = 1,$$

and henceforth P_t is a stochastic matrix for all $t \geq 0$. Conversely, assuming that $\sum_{j\in I} P_t(i,j) = 1$ for all $t \geq 0$ gives that

$$\sum_{j\in I} q_{i,j} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \sum_{j\in I} P_t(i,j) = 0.$$

□

Example. Consider $I = \{0, 1, \dots, N\}$, $\lambda > 0$, and the following Q-matrix $Q = (q_{i,j})_{i,j \in I}$ with $q_{i,i+1} = \lambda$ and $q_{i,i} = -\lambda$ for $i \in \{0, 1, \dots, N-1\}$ and all other entries being zero. Clearly, Q is an upper-triangular matrix and so is any exponential of it. Hence, $P_t(i,j) = 0$ for $i < j$ and $t \geq 0$. The forward equation $P'(t) = P(t)Q$ reads as

$$
\begin{aligned}
P_t'(i,i) &= -\lambda P_t(i,i); P_0(i,i) = 0, i \in \{0, 1, \dots, N-1\} \\
P_t'(i,j) &= -\lambda P_t(i,j) + \lambda P_t(i,j-1); P_0(i,j) = 0, 0 \leq i < j < N, \\
P_t'(i,N) &= \lambda P_t(i,N-1); P_0(i,N) = 0, i < N.
\end{aligned}
$$

To solve these equations we first note that $P_t(i,i) = \mathrm{e}^{-\lambda t}$ for $i \in \{0, 1, \dots, N-1\}$. Using that we get for $0 \leq i < j < N$ that $\left(\mathrm{e}^{\lambda t} P_t(i,j)\right)' = \mathrm{e}^{\lambda t} P_t(i,j-1)$, and henceforth by induction

$$
\begin{aligned}
P_t(i,j) &= \mathrm{e}^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, \text{ for } 1 \leq i < j < N-1, \\
P_t(i,N) &= 1 - \sum_{l=0}^{N-i-1} \frac{(\lambda t)^l}{l!}, \text{ for } 0 \leq i < N, \\
P_t(N,N) &= 1.
\end{aligned}
$$

If $i = 0$, these are the Poisson probabilities of parameter λt. ◇

Example. A virus exists in $N+1$ strains $0, 1, \dots, N$. It keeps its strain for a random time which is exponential distributed with parameter $\lambda > 0$, then mutates to one of the remaining strains equiprobably. Find the probability that the strain at time t is the same as the initial strain. Due to symmetry, $q_i = -q_{i,i} = \lambda$ and $q_{i,j} = \frac{\lambda}{N}$ for $1 \leq i,j \leq N+1, i \neq j$. We shall compute $P_t(i,i) = \left(\mathrm{e}^{tQ}\right)_{i,i}$. Clearly, $P_t(i,i) = P_t(1,1)$ for all i,$t \geq 0$, again by symmetry. A reduced (2×2)-matrix, over states 0 and 1 is

$$
\widetilde{Q} = \begin{pmatrix} -\lambda & \lambda \\ \lambda/N & -\lambda/N \end{pmatrix}.
$$

The matrix $\widetilde{Q}$ has eigenvalues 0 and $\mu = -\lambda(N+1)/N$ with its row eigenvectors being $(1,1)$ and $(N,-1)$. Hence, we get the ansatz

$$
P_t(1,1) = A + B\mathrm{e}^{-\frac{\lambda(N+1)}{N}t}.
$$

We seek solutions of the form $A + B\mathrm{e}^{\mu t}$, and we obtain $A = 1/(N+1)$ and $B = n/(N+1)$ and

$$
P_t(1,1) = \frac{1}{N+1} + \Big(\frac{N}{N+1}\Big)\mathrm{e}^{-\frac{\lambda(N+1)}{N}t} = P_t(i,i).
$$

By symmetry,

$$P_t(i,j) = \frac{1}{N+1} - \Big(\frac{1}{N+1}\Big)\mathrm{e}^{-\frac{\lambda(N+1)}{N}t}, \quad i \neq j,$$

and we conclude

$$P_t(i,j) \to \frac{1}{N+1} \text{ as } t \to \infty.$$

$\diamond$

2.3 Jump chain and holding times

We introduce the jump chain of a Markov process, the holding times given a Q-matrix and the explosion time. Given a Markov process $X = (X_t)_{t\geq 0}$ on a countable state space there are the following cases:

(A) The process has infinitely many jumps but only finitely many in any interval $[0,t]$, $t \geq 0$.

(B) The process has only finitely many jumps, that is, there exists a $k \in \mathbb{N}$ such that the k-th waiting/holding time $E_k = \infty$.

(C) The process has infinitely many jumps in a finite time interval. After the explosion time ζ (to be defined later, see below) has passed the process starts up again.

$J_0, J_1, \ldots$ are called the **jump times** and $(E_k)_{k\in\mathbb{N}}$ are called the **holding/waiting times** of the Markov process $X = (X_t)_{t\geq 0}$.

$$J_0 := 0, J_{n+1} = \inf\{t \geq J_n \colon X_t \neq X_{J_n}\}, n \in \mathbb{N}_0,$$

where we put $\inf\{\emptyset\} = \infty$, and for $k \in \mathbb{N}$

$$E_k = \begin{cases} J_k - J_{k-1} & \text{if } J_{k-1} < \infty \\ \infty & \text{otherwise} \end{cases}.$$

The (first) **explosion time** ζ is defined by

$$\zeta = \sup_{n\in\mathbb{N}_0}\{J_n\} = \sum_{k=1}^{\infty} E_k.$$

The discrete-time process $(Y_n)_{n\in\mathbb{N}_0}$ given by $Y_n = X_{J_n}$ is called the **jump process** or the **jump chain** of the Markov process. Whenever a Markov process is satisfying that $X_t = \partial$ if $t \geq \zeta$ we call this process (realization) **minimal**.

Proposition 2.16 *Let $(E_k)_{k\in\mathbb{N}}$ be a sequence of independent random variables with $E_k \sim E(\lambda_k)$ and $0 < \lambda_k < \infty$ for all $k \in \mathbb{N}$.*

(a) *If* $\sum_{k=1}^{\infty} \frac{1}{\lambda_k} < \infty$, *then* $\mathbb{P}(\zeta = \sum_{k=1}^{\infty} E_k < \infty) = 1$.

(b) *If* $\sum_{k=1}^{\infty} \frac{1}{\lambda_k} = \infty$, *then* $\mathbb{P}(\zeta = \sum_{k=1}^{\infty} E_k = \infty) = 1$.

Proof. The proof follows easily using the Monotone Convergence Theorem and independence. □

Recall the definition of the Poisson process and in particular the characterisation in Theorem 2.5. A right continuous process $(N_t)_{t\geq 0}$ with values in $\mathbb{N}_0$ is a Poisson process of rate $\lambda \in (0,\infty)$ if its holding times $E_1, E_2, \ldots$ are independent exponential random variables of parameter λ, its increments are independent, and its jump chain is given by $Y_n = n, n \in \mathbb{N}_0$. To obtain the corresponding Q-matrix we recall that the off-diagonal entries are the jump rates. Jump rates are jumps/per unit time. We 'wait' an expected time $\frac{1}{\lambda}$, then we jump by one, hence the jump rate is $\frac{1}{\frac{1}{\lambda}} = \lambda$ and the Q-matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \cdots & \cdots \\ 0 & -\lambda & \lambda & 0 & 0 & \cdots & \cdots \\ 0 & 0 & -\lambda & \lambda & 0 & \cdots & \cdots \\ 0 & \cdots & \cdots & -\lambda & \lambda & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & -\lambda & \lambda & \cdots \\ \cdots & \cdots & \cdots & & & & \end{pmatrix}.$$

The following Theorem gives a complete characterisation of the Poisson process.

Theorem 2.17 (Poisson process) *The Poisson process for parameter (intensity)* $\lambda \in (0,\infty)$ *can be characterised in three equivalent ways: a process* $(N_t)_{t\geq 0}$ *(right continuous) taking values in* $\mathbb{N}_0$ *with* $N_0 = 0$ *and:*

(a) *and for all* $0 < t_1 < t_2 < \cdots < t_n$, $n \in \mathbb{N}$, *and* $i_1, \ldots, i_n \in \mathbb{N}_0$

$$\mathbb{P}(N_{t_1} = i_1, \ldots, N_{t_n} = i_n) = P_{t_1}(0, i_1) P_{t_2 - t_1}(i_1, i_2) \cdots P_{t_n - t_{n-1}}(i_{n-1}, i_n),$$

where the matrix P_t *is defined as* $P_t = \mathrm{e}^{tQ}$ *(to be justified as the state space is not finite).*

(b) *with independent increments* $N(t_1) - N(t_0), \ldots, N(t_n) - N(t_{n-1})$, *for all* $0 = t_0 < t_1 < \cdots < t_n$, *and the infinitesimal probabilities for all* $t \geq$ *as* $h \downarrow 0$

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

where the terms $o(h)$ *do not depend on* t.

(c) spending a random time $E_k \sim E(\lambda)$ in each state $k \in \mathbb{N}_0$ independently, and then jumping to $k+1$.

Proof. We need to justify the operation $P_t = \mathrm{e}^{tQ}$ as the state space $\mathbb{N}_0$ is not finite. We are using the fact that Q is upper triangular and so is Q^k for any $k \in \mathbb{N}$ and therefore P_t is upper triangular. In order to find the entries $P_t(i, i+l)$ for any $l \in \mathbb{N}_0$ we use the forward or backward equation both with initial condition $P(0) = \mathbb{1}$. This gives $\frac{\mathrm{d}}{\mathrm{d}t} P_t(i,i) = -\lambda P_t(i,i)$ and $P_0(i,i) = 1$ and thus $P_t(i,i) = \mathrm{e}^{-\lambda t}$ for all $i \in \mathbb{N}_0$ and all $t \geq 0$. Put $l = 1$, that is consider one step above the main diagonal. Then

$$\text{(forward)}\ \frac{\mathrm{d}}{\mathrm{d}t} P_t(i, i+1) = -\lambda P_t(i, i+1) + \lambda P_t(i,i),$$
$$\text{(backward)}\ \frac{\mathrm{d}}{\mathrm{d}t} P_t(i, i+1) = -\lambda P_t(i, i+1) + \lambda P_t(i+1, i+1)$$

gives $P_t(i, i+1) = \lambda t \mathrm{e}^{-\lambda t}$ for all $i \in \mathbb{N}_0$ and $t \geq 0$. The general case (i.e. $l \in \mathbb{N}_0$) follows in the same way and henceforth

$$P_t(i, i+l) = \frac{(\lambda t)^l}{l!} \mathrm{e}^{-\lambda t}, \quad i \in \mathbb{N}_0, t \geq 0.$$

(a) $\Rightarrow$ (b): We get for $l = 0, 1$

$$\mathbb{P}(N(t+h) - N(t) = l) = \frac{(\lambda h)^l}{l!} \mathrm{e}^{-\lambda h} = \begin{cases} \mathrm{e}^{-\lambda h} = 1 - \lambda + o(h) & \text{if } l = 0 \\ \lambda h \mathrm{e}^{-\lambda h} = \lambda h + o(h) & \text{if } l = 1, \end{cases}$$

and $\mathbb{P}(N(t+h) - N(t) \geq 2) = 1 - \mathbb{P}(N(t+h) - N(t) = 0 \text{ or } 1) = 1 - (1 - \lambda h + \lambda h + o(h)) = o(h)$.

(b) $\Rightarrow$ (c): This step is more involved. We need to get around the infinitesimal probabilities, that is small times h. This is done as follows. We first check that no double jumps exists, i.e.

$$\begin{aligned}
&\mathbb{P}(\text{no jumps of size } \geq 2 \text{ in } (0,t]) \\
&= \mathbb{P}\Big(\text{no such jumps in } (\frac{k-1}{m}t, \frac{k}{m}t]\, \forall\, k = 1, \ldots, m\Big) \\
&= \prod_{k=1}^m \mathbb{P}\Big(\text{no such jumps in } (\frac{k-1}{m}t, \frac{k}{m}t]\Big) \\
&\geq \prod_{k=1}^m \mathbb{P}\Big(\text{no jump at all or single jump of size in } (\frac{k-1}{m}t, \frac{k}{m}t]\Big) \\
&= \big(1 - \lambda \frac{t}{m} + \lambda \frac{t}{m} + o(\frac{t}{m})\big)^m = (1 + o(\frac{t}{m}))^m \to 1 \text{ as } m \to \infty.
\end{aligned}$$

This is true for all $t \geq 0$ and henceforth $\mathbb{P}(\text{no jumps of size } \geq 2 \text{ ever}) = 1$. Pick $t, s > 0$ and obtain

$$
\begin{aligned}
\mathbb{P}(N(t+s)) &= \mathbb{P}(\text{no jumps in } (s, s+t]) \\
&= \mathbb{P}\Big(\text{no jumps in } (s + \frac{k-1}{m}t, s + \frac{k}{m}t] \,\forall\, k = 1, \dots, m\Big) \\
&= \prod_{k=1}^{m} \mathbb{P}\Big(\text{no jumps in } (s + \frac{k-1}{m}t, s + \frac{k}{m}t]\Big) \\
&= \big(1 - \lambda\frac{t}{m} + o(\frac{t}{m})\big)^m \to \mathrm{e}^{-\lambda t} \text{ as } m \to \infty.
\end{aligned}
$$

With some slight abuse we introduce the holding times with index starting at zero (before we started here with one):

$$
\begin{aligned}
&E_0 = \sup\{t \geq 0 \colon N(t) = 0\}, \\
&E_1 = \sup\{t \geq 0 \colon N(E_0 + t) = 1\}, \dots, E_n \quad = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

Note that the jump time J_k is also the hitting time of the state $k \in \mathbb{N}_0$. We need to show that these holding times are independent and exponential distributed with parameter λ. In order to do so we compute the probability for some given time intervals and show that it is given as a product of the corresponding densities. Pick positive $t_1, \dots, t_n$ and positive $h_1, \dots, h_n$ such that $0 < t_1 < t_1 + h_1 < \cdots < t_{n-1} + h_{n-1} < t_n$. We get

$$
\begin{aligned}
&\mathbb{P}(t_1 < J_1 \leq t_1 + h_1, \dots, t_n < J_n \leq t_n + h_n) \\
=&\mathbb{P}(N(t_1) = 0; N(t_1 + h_1) - N(t_1) = 1; \dots; N(t_n) - N(t_{n-1} + h_{n-1}) = 0; \\
&\; N(t_n + h_n) - N(t_n) = 1) \\
&= \mathbb{P}(N(t_1) = 0)\mathbb{P}(N(t_1 + h_1) - N(t_1) = 1) \cdots \\
&= \mathrm{e}^{-\lambda t_1}(\lambda h_1 + o(h_1))\mathrm{e}^{-\lambda(t_2 - t_1 - h_1)} \times \cdots \times \mathrm{e}^{-\lambda(t_n - t_{n-1} - h_{n-1})}(\lambda h_n + o(h_n)),
\end{aligned}
$$

and dividing by $h_1 \times \cdots h_n$ and taking the limit $h_i \downarrow 0$, $i = 1, \dots, n$, gives that the left hand side is the joint probability density function of the n jump times and the right hand side is the product $\big(\mathrm{e}^{-\lambda t_1}\lambda\big)\big(\mathrm{e}^{-\lambda(t_2 - t_1)}\lambda\big) \cdots \big(\mathrm{e}^{-\lambda(t_n - t_{n-1})}\lambda\big)$. Thus the joint density function reads as

$$
\begin{aligned}
f_{J_1, \dots, J_n}(t_1, \dots, t_n) &= \prod_{k=1}^{n} \big(\lambda \mathrm{e}^{-\lambda(t_k - t_{k-1})}\big) 1\!\!1\{0 < t_1 < \cdots < t_n\} \\
&= \lambda^n \mathrm{e}^{-\lambda t_n} 1\!\!1\{0 < t_1 < \cdots < t_n\}.
\end{aligned}
$$

Recall $E_0 = J_0, E_1 = J_0 + J_1 = J_1, \ldots$, hence we make a change of variables (for the n times) $e_0 = t_1, e_1 = t_2 - t_1, e_3 = t_3 - t_2, \ldots, e_{n-1} = t_n - t_{n-1}$. The determinant of the Jacobi matrix for this transformation is one and therefore

$$\begin{aligned} f_{E_0,\ldots,E_{n-1}}(e_0, e_1, \ldots, e_{n-1}) &= f_{J_1,\ldots,J_n}(e_0, e_0 + e_1, \ldots, e_0 + \cdots + e_{n-1}) \\ &= \prod_{k=0}^{n-1} \big(\lambda \mathrm{e}^{-\lambda e_k} \mathbb{1}\{e_k > 0\}\big), \end{aligned}$$

and henceforth $E_0, \ldots$ are independent and exponential distributed with parameter λ.

(c) $\Rightarrow$ (a): This is already proved in Theorem 2.5. □

We finish our discussion of the Poisson process with a final result concerning the uniform distribution of a single jump in some times interval.

Proposition 2.18 *Let $(N_t)_{t \geq 0}$ be a Poisson process. Then, conditional on $(N_t)_{t \geq 0}$ having exactly one jump in the interval $[s, s+t]$, the time at which that jump occurs is uniformly distributed in $[s, s+t]$.*

Proof. Pick $0 \leq u \leq t$.

$$\begin{aligned} \mathbb{P}(J_1 \leq u | N_t = 1) &= \mathbb{P}(J_1 \leq u \text{ and } N_t = 1)/\mathbb{P}(N_t = 1) \\ &= \mathbb{P}(N_u = 1 \text{ and } N_t - N_u = 0)/\mathbb{P}(n_t = 1) \\ &= \lambda u \mathrm{e}^{-\lambda u} \mathrm{e}^{-\lambda(t+u)}/(\lambda t \mathrm{e}^{-\lambda t}) = \frac{u}{t}. \end{aligned}$$

□

We continue setting up a general theory for Markov processes. Let I be a countable state space. The basic data for a Markov process on I is given by the Q-matrix. Having a Q-matrix one can compute the transition matrix for the corresponding jump chain of the process.

Notation 2.19 (Jump matrix Π) *The jump matrix $\Pi = (\pi_{i,j})_{i,j \in I}$ of a Q-matrix $Q = (q_{i,j})_{i,j \in I}$ is given by*

$$\pi_{i,j} = \begin{cases} \frac{q_{i,j}}{q_i} & \text{if } q_i \neq 0 \\ 0 & \text{if } q_i = 0, \end{cases} \qquad \text{if } i \neq j;$$

and

$$\pi_{i,i} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0. \end{cases}$$

A right continuous process $X = (X_t)_{t\ge 0}$ is a Markov process with initial distribution λ (probability measure on I) and Q-matrix (generator) if its jump chain $(Y_n)_{n\in\mathbb{N}_0}$ is a discrete time Markov chain with initial distribution λ and transition matrix Π (given in Notation (2.19)) and if for all $n \in \mathbb{N}$, conditional on $Y_0, \dots, Y_{n-1}$, its holding (waiting) times are independent exponential random variables of parameters $q(Y_0), \dots, q(Y_{n-1})$ (negative diagonal entries of the Q-matrix at states given by the jump chain) respectively. How we can construct a Markov process given a discrete time Markov chain? Pick a Q-matrix respectively a jump matrix Π and consider the discrete time Markov chain $(Y_n)_{n\in\mathbb{N}_0}$ having initial distribution λ and transition matrix Π. Furthermore, let $T_1, T2, \dots$ be a family of independent random variables exponential distributed with parameter 1, independent of $(Y_n)_{n\in\mathbb{N}_0}$. Put

$$E_n = \frac{T_n}{q(Y_{n-1})} \text{and } J_n = E_1 + \cdots + E_n, n \in \mathbb{N},$$

$$X_t := \begin{cases} Y_n & \text{if } J_n \le t < J_{n+1} \text{ for some } n, \\ \infty(\partial) & \text{otherwise.} \end{cases}$$

Then $(X_t)_{t\ge 0}$ has the required properties of a Markov process.

In the next result we study further properties of families of exponential random variables.

Proposition 2.20 *Let I be countable and let $T_k, k \in I$, be independent random variables with $T_k \sim E(q_k)$ with $q_k \ge 0$ and $0 < q := \sum_{k\in I} q_k < \infty$. Define $T := \inf_{kj\in I}\{T_k\}$. Then the infimum is attained at a unique random value K, with probability one, T and K are independent, and $T \sim E(q)$ and $\mathbb{P}(K = k) = \frac{q_k}{q}$.*

Proof. We skip the proof here, it's an excellent exercise in exponential random variables. □

Recall that q_i is the rate of leaving the state $i \in I$ and that $q_{i,j}$ is the rate of going from state i to state j. Hence, we shall get a criterion not having explosion of a process in terms of the q_i's.

Proposition 2.21 (Explosion) *Let $(X_t)_{t\ge 0}$ be a (λ, Q)-Markov process on some countable state space I. Then the process does not explode if any one of the following conditions holds:*

(a) I is finite.

(b) $\sup_{i\in I}\{q_i\} < \infty$.

(c) $X_0 = i$ *and the state* i *is recurrent for the jump chain (a state* i *is recurrent if* $\mathbb{P}_i(Y_n = i$ *for infinitely many* $n) = 1$*).*

Proof. Put $T_n := q(Y_{n-1})E_n$, then $T_n \sim E(1)$ and $(T_n)_{n\in\mathbb{N}}$ is independent of $(Y_n)_{n\in\mathbb{N}_0}$. (a),(b): We have $q := \sup_{i\in I}\{q_i\} < \infty$, and hence

$$q\zeta \geq \sum_{n=1}^{\infty} T_n = \infty \quad \text{with probability } 1.$$

(c) If $(Y_n)_{n\in\mathbb{N}_0}$ visits the state i infinitely often at times $N_1, \ldots$, then

$$q_i\zeta \geq \sum_{n=1}^{\infty} T_{N_n+1} = \infty \quad \text{with probability } 1.$$

□

We say a Q-matrix Q is **explosive** if $\mathbb{P}_i(\zeta < \infty) > 0$ for some state $i \in I$.

2.4 Forward and backward equations

We are heading towards a final characterisation of a continuous time Markov process on a countable state space. Recall that in case the state space I is finite there is nothing to show as the semigroup of the transition probabilities is given as an exponential of the Q-matrix. We have seen in the special case of a Poisson process that we can use the backward and forward equations to a good extent. We exploring this further in the following.

Proposition 2.22 *Let* $(P_t)_{t\geq 0}$ *be the semigroup of a Markov process* $(X_t)_{t\geq 0}$ *on a countable state space* I *with* Q*-matrix* Q*. Then* $t \mapsto P_t(i,j)$ *is for* $t \geq 0$ *and for all* $i, j \in I$ *continuously differentiable and*

$$P'(t) = QP(t), \quad t \geq 0, P(0) = \mathbb{1}.$$

Proof.

$$\begin{aligned}\frac{P_{t+h}(i,j) - P_t(i,j)}{h} &= \frac{1}{h}\Big(\sum_{k\in I} P_h(i,k)P_t(k,j) - P_t(i,j)\Big) \\ &= \frac{1}{h}\Big(\Big[P_h(i,i) - \mathbb{1}\Big]P_t(i,j) + \sum_{k\in I\setminus\{i\}} P_h(i,k)P_t(k,j)\Big).\end{aligned}$$

Interchanging limit and summation (we justify this afterwards) gives

$$\lim_{h\downarrow 0} \frac{P_{t+h}(i,j) - P_t(i,j)}{h} = q_{i,i}P_t(i,j) + \sum_{k\in I\setminus\{i\}} q_{i,k}P_t(k,j) = \sum_{k\in I} q_{i,k}P_t(k,j).$$

We need to justify the interchange of limit and summation. Pick $\varepsilon > 0$ and $J \subset I$ with $i \in J$ and $|J| < \infty$ such that $\sum_{k\in I\setminus J} q_{i,k} < \varepsilon/2$. Interchange for the finite sum in J is no problem, the remainder is estimated as follows.

$$
\begin{aligned}
\Big|\sum_{k\in I\setminus J}\Big(\frac{P_h(i,k)}{h} - q_{i,k}\Big)P_t(k,j)\Big| &\le \Big|\sum_{k\in I\setminus J}\frac{P_h(i,k)}{h}\Big| + \sum_{k\in I\setminus J} q_{i,k} \\
&< \Big|\frac{1-P_h(i,i)}{h} - \sum\sum_{k\in I\setminus\{i\}}\frac{P_h(i,k)}{h}\Big| + \varepsilon/2 \to -q_{i,i} - \sum_{k\in J\setminus\{i\}} q_{i,k} + \varepsilon/2 \\
&= \sum_{k\in I\setminus J} q_{i,k} + \varepsilon/2 < \varepsilon \text{ as } h \downarrow 0.
\end{aligned}
$$

□

If the state space is not finite we have the following characterisation of the semigroup of transition probabilities.

Proposition 2.23 (Backward/forward equation) *Let Q be a Q-matrix on a countable state space I.*

(a) Then the backward equation

$$P'(t) = QP(t), \quad P(0) = 1\!\!1,$$

has a minimal non-negative solution $(P(t))_{t\ge0}$. This solution forms a matrix semigroup $P(s)P(t) = P(s+t)$ for all $s,t \ge 0$.

(b) The minimal non-negative solution of the backward equation is also the minimal non-negative solution of the forward equation

$$P'(t) = P(t)Q, \quad P(0) = 1\!\!1.$$

Proof. The proof is rather long, and we skip it here as it goes beyond the level of the course. □

Here is now our key result for Markov processes with infinite (countable) state space I. There are just two alternative definitions left now as the infinitesimal characterisation becomes problematic for infinite state space.

Theorem 2.24 (Markov process, final characterisation) *Let $X = (X_t)_{t\ge0}$ be a minimal right continuous process having values in a countable state space I. Furthermore, let Q be a Q-matrix on I with jump matrix Π and semigroup (solution-see Proposition 2.23) $(P_t)_{t\ge0}$. Then the following conditions are equivalent:*

(a) Conditional on $X_0 = i$, the jump chain $(Y_n)_{n\in\mathbb{N}_0}$ of $(X_t)_{t\geq 0}$ is a discrete time Markov chain with initial distribution δ_i and transition matrix Π and for each $n \geq 1$, conditional on $Y_0, \ldots, Y_{n-1}$, the holding (waiting) times $E_1, \ldots, E_n$ are independent exponential random variables of paramters $q(Y_0). \ldots, q(Y_{n-1})$ respectively;

(b) for all $n \in \mathbb{N}_0$, all times $0 \leq t_0 \leq t_1 \leq \ldots \leq t_{n+1}$ and all states $i_0, \ldots, i_{n+1} \in I$

$$\mathbb{P}(X_{t_n} = i_{n+1} | X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = P_{t_{n+1}-t_n}(i_n, i_{n+1}).$$

If $(X_t)_{t\geq 0}$ satisfies any of these conditions then it is called a **Markov process** *with* **generator matrix** *Q. If λ is the distribution of X_0 it is called the initial distribution of the Markov process.*

3 Examples, hitting times, and long-time behaviour

We study the birth-and-death process, introduce hitting times and probabilities, and discuss recurrence and transience. The last Subsection is devoted to a brief introduction to queueing models.

3.1 Birth-and-death process

Birth process: This is a Markov process $X = (X_t)_{t\geq 0}$ with state space $I = \mathbb{N}_0$ which models growth of populations. We provide two alternative definitions:

Definition via 'holding times': Let a sequence $(\lambda_j)_{j\in\mathbb{N}_0}$ of positive numbers be given. Conditional on $X(0) = j, j \in \mathbb{N}_0$, the successive holding times are independent exponential random variables with parameters $\lambda_j, \lambda_{j+1}, \ldots$. The sequence $(\lambda_j)_{j\in\mathbb{N}_0}$ is thus the sequence of the **birth rates** of the process.

Definition via 'infinitesimal probabilities': Pick $s, t \geq 0, t > s$, conditional on $X(s)$, the increment $X(t) - X(s)$ is positive and independent of $(X(u))_{0\leq u\leq s}$. Furthermore, as $h \downarrow 0$ uniformly in $t \geq 0$, it holds for $j, m \in \mathbb{N}_0$ that

$$\mathbb{P}(X(t+h) = j+m | X(t) = j) = \begin{cases} \lambda_j h + o(h) & \text{if } m = 1, \\ o(h) & \text{if } m > 1, \\ 1 - \lambda_j h + o(h) & \text{if } m = 0. \end{cases}$$

From the latter definition we get the difference of the transition probabilities as

$$P_{t+h}(j,k) - P_t(j,k) = P_t(j,k-1)\lambda_{k-1}h - P_t(j,k)\lambda_k h + o(h), \quad j \in \mathbb{N}_0, k \in \mathbb{N}$$
$$P_t(j,j-1) = 0,$$

hence the forward equations read as

$$P_t'(j,k) = \lambda_{k-1}P_t(j,k-1) - \lambda_k P_t(j,k), \quad j \in \mathbb{N}_0, k \in \mathbb{N}, k \geq j.$$

Alternatively, conditioning on the time of the first jump yields the following relation

$$P_t(j,k) = \delta_{j,k}\mathrm{e}^{-\lambda_j t} + \int_0^t \lambda_j \mathrm{e}^{-\lambda_j s} P_{t-s}(j+1,k)\,\mathrm{d}s,$$

and the backward equations read

$$P_t'(j,k) = \lambda_j P_t(j+1,k) - \lambda_j P_t(j,k).$$

Theorem 3.1 (Birth process) *(a) With the initial condition $P_0(j,k) = \delta_{j,k}$ the forward equation has a unique solution which satisfies the backward equation.*

(b) If $(P_t)_{t\geq 0}$ is a unique solution to the forward equation and $(B_t)_{t\geq 0}$ any solution of the backward equation, then $P_t(j,k) \leq B_t(j,k)$ for all $j,k \in \mathbb{N}_0$.

Proof. We give only a brief sketch. We get easily from the definition

$$\begin{aligned} P_t(j,k) &= 0, \quad k < j, \\ P_t(j,j) &= \mathrm{e}^{-\lambda_j t}, \\ P_t(j,j+1) &= \mathrm{e}^{-\lambda_{j+1}t}\int_0^t \lambda_j \mathrm{e}^{-(\lambda_j - \lambda_{j+1})s}\,\mathrm{d}s \\ &= \frac{\lambda_j}{\lambda_j - \lambda_{j+1}}\Big(\mathrm{e}^{-\lambda_{j+1}t} - \mathrm{e}^{-\lambda_j t}\Big). \end{aligned}$$

□

Examples. (a) Simple birth process where the birth rates are linear, i.e. $\lambda_j = \lambda j, \lambda > 0, j \in \mathbb{N}_0$. (b) Simple birth process with immigration where $\lambda_j = \lambda j + \nu, \nu \in \mathbb{R}$. ◇

Birth-and-death process Let two sequences $(\lambda_k)_{k\in\mathbb{N}_0}$ and $(\mu_k)_{k\in\mathbb{N}_0}$ of positive numbers be given. At state $k \in \mathbb{N}_0$ we have a birth rate λ_k and a **death rate**

μ_k, and we only allow 1-step transitions, that is either one birth or one death. The Q-matrix reads as

$$Q = \begin{pmatrix} \lambda_0 & -\lambda_0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & 0 & 0 & \dots \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0\dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

We obtain the infinitesimal probabilities

$$\begin{aligned} \mathbb{P}(\text{exactly 1 birth in } (t,t+h]|k) &= \lambda_k h + o(h), \\ \mathbb{P}(\text{exactly 1 death in } (t,t+h]|k) &= \mu_k h + o(h), \\ \mathbb{P}(\text{no birth in}(t,t+h]|k) &= 1 - \lambda_k h + o(h), \\ \mathbb{P}(\text{no death in}(t,t+h]|k) &= 1 - \mu_k h + o(h). \end{aligned}$$

In the following figure (see) we have three potential transitions to the state k at time $t+h$, namely if we have at time t the state $k+1$ we have one death, if we have $k-1$ at time t we do have exactly one birth, and if at time t we have already state k then we do not have a birth or a death. This is expressed in the following relation for the transition probabilities.

$$\begin{aligned} P_{t+h}(0,k) &= P_t(0,k)P_h(k,k) + P_t(0,k-1)P_t(k-1,k) \\ &\quad + P_t(0,k+1)P_h(k+1,k) \\ P_{t+h}(0,0) &= P_t(0,0)P_h(0,0) + P_t(0,1)P_h(1,0). \end{aligned}$$

Combining these facts yields

$$\begin{aligned} \frac{\mathrm{d}P_t(0,k)}{\mathrm{d}t} &= -(\lambda_k+\mu_k)P_t(0,k) + \lambda_{k+1}P_t(0,k-1) + \mu_{k+1}P_t(0,k+1) \\ \frac{\mathrm{d}P_t(0,0)}{\mathrm{d}t} &= -\lambda_0 P_t(0,0) + \mu_1 P_t(0,1). \end{aligned}$$

This can be seen as a probability flow. Pick a state k, then the probability flow rate into state k is given as $\lambda_{k-1}P_t(0,k-1) + \mu_{k+1}P_t(0,k+1)$, whereas the probability flow rate out of the state k is given as $(\lambda_k+\mu_k)P_t(0,k)$, henceforth the probability flow rate is the difference of the flow into and out of a state.

3.2 Hitting times and probabilities. Recurrence and transience

In this section we study properties of the single states of a continuous time Markov process. Let a countable state space I and a Markov process $X = (X_t)_{t\geq 0}$ with state space I be given. If $i, j \in I$ we say that i **leads to** j and write $i \longrightarrow j$ if $\mathbb{P}_i(X_t = j \text{ for some } t \geq 0) > 0$. We say i **communicates with** j and write $i \longleftrightarrow$ if both $i \longrightarrow j$ and $j \longrightarrow i$ hold.

Theorem 3.2 *Let $X = (X_t)_{t\geq 0}$ be a Markov process with state space I and Q-matrix $Q = (q_{i,j})_{i,j\in I}$ and jump matrix $\Pi = (\pi_{i,j})_{i,j\in I}$. The following statements for $i, j \in I, i \neq j$ are equivalent.*

(a) $i \longrightarrow j$.

(b) $i \longrightarrow j$ *for the corresponding jump chain* $(Y_n)_{n\in\mathbb{N}_0}$.

(c) $q_{i_0,i_1}q_{i_1,i_2}\cdots q_{i_{n-1},i_n} > 0$ *for some states* $i_0, \dots, i_n \in I$ *with* $i_0 = i$ *and* $i_n = j$.

(d) $P_t(i,j) > 0$ *for all* $t > 0$.

(e) $P_t(i,j) > 0$ *for some* $t > 0$.

Proof. All implications are clear, we only show (c) $\Rightarrow$ (d). If $q_{i,j} > 0$, then

$$P_t(i,j) \geq \mathbb{P}_i(J_1 \leq t, Y_1 = j, E_2 > t) = (1 - \mathrm{e}^{-q_i t})\pi_{i,j}\mathrm{e}^{-q_j t} > 0,$$

because $q_{i,j} > 0$ implies $pi_{i,j} > 0$. □

Let a subset $A \subset I$ be given. The **hitting time** of the set A is the random variable D^A defined by

$$D^A = \inf\{t \geq 0 \colon X_t \in A\}.$$

Note that this random time can be infinite. It is therefore of great interest if the probability of ever hitting the set A is strictly positive, that is the probability that D^A is finite. The **hitting probability** h_i^A of the set A for the Markov process $(X_t)_{t\geq 0}$ starting from state $i \in I$ is defined as

$$h_i^A := \mathbb{P}_i(D^A < \infty).$$

Before we state and prove general properties let us study the following example concerning the expectations of hitting probabilities. The average time, starting from state i, for the Markov process $(X_t)_{t\geq 0}$ to reach the set A is given by $k_i^A := \mathbb{E}_i(D^A)$.

Example. Let be given four states $1, 2, 3, 4$ with the following transition rates (see figure). $1 \to 2 = 1; 1 \to 3 = 1; 2 \to 1 = 2; 2 \to 3 = 2; 2 \to 4 = 2; 3 \to 1 = 3; 3 \to 2 = 3; 3 \to 4 = 3$. How long does it take to get from state 1 to state 4? Note that once the process arrives in state 4 he will be trapped. Write $k_i := \mathbb{E}_i(\text{time to get to state } 4)$. Starting in state 1 we spend an average time $q_1^{-1} = \frac{1}{2}$ in state 1, then we jump with equal probability to state 2 or state 3, i.e.

$$k_1 = \frac{1}{2} + \frac{1}{2}k_2 + \frac{1}{2}k_3,$$

and similarly

$$k_2 = \frac{1}{6} + \frac{1}{3}k_1 + \frac{1}{3}k_3,$$
$$k_3 = \frac{1}{9} + \frac{1}{3}k_1 + \frac{1}{3}k_2.$$

Solving these linear equations gives $k_1 = \frac{17}{12}$. ◇

Proposition 3.3 *Let $X = (X_t)_{t\geq 0}$ be a Markov process with state space I and Q-matrix $Q = (q_{i,j})_{i,j\in I}$ and $A \subset I$.*

(a) The vector $h^A = (h_i^A)_{i\in I}$ is the minimal non-negative solution to the system of linear equations

$$\begin{cases} h_i^A = 1 & \text{if } i \in A, \\ \sum_{j\in I} q_{i,j}h_j^A = 0 & \text{if } i \notin A. \end{cases}$$

(b) Assume that $q_i > 0$ for all $i \notin A$. The vector $k^A = (k_i^A)_{i\in I}$ of the expected hitting times is the minimal non-negative solution to the system of linear equations

$$\begin{cases} k_i^A = 0 & \text{if } i \in A, \\ -\sum_{j\in I} q_{i,j}k_j^A = 1 & \text{if } i \notin A. \end{cases}$$

Proof. (a) is left as an exercise. (b) $X_0 = i \in A$ implies $D^A = 0$, so $k_i^A = 0$ for $i \in A$. If $X_0 = i \notin A$ we get that $D^A \geq J_1$. By the Markov property of the corresponding jump chain $(Y_n)_{n\in\mathbb{N}_0}$ it follows that

$$\mathbb{E}_i(D^A - J_1 | J_1 = j) = \mathbb{E}_j(D^A).$$

Using this we get

$$\begin{aligned} k_i^A = \mathbb{E}_i(D^A) &= \mathbb{E}_i(J_1) + \sum_{j\in I\setminus\{i\}} \mathbb{E}_i(D^A - J_1 | Y_1 = j)\mathbb{P}_i(Y_1 = j) \\ &= q_i^{-1} + \sum_{j\in I\setminus\{i\}} \pi_{i,j}k_j^A, \end{aligned}$$

and therefore $-\sum_{j\in I} q_{i,j}k_j^A = 1$. We skip the details for proving that this is the minimal non-negative solution. □

Example. Birth-and-death process: Recall that a birth-and-death process with birth rates $(\lambda_k)_{k\in\mathbb{N}_0}$ and death rates $(\mu_k)_{k\in\mathbb{N}_0}$ with $\mu_0 = 0$ has the Q-matrix $Q = (q_{i,j})_{i,j\in\mathbb{N}_0}$ given by

$$q_{j,j+1} = \lambda_j, q_{j,j-1} = \mu_j, j > 0, q_j = \lambda_j + \mu_j.$$

Let $k_{j,j+1}$ be the expected time it takes to reach state $j+1$ when starting in state j. The holding (waiting) time in state $j > 0$ has mean (expectation) $\frac{1}{\lambda_j+\mu_j}$. Hence,

$$\mathbb{E}_j(D^{\{j+1\}}) = k_{j,j+1} = (\lambda_j + \mu_j)^{-1} + \frac{\mu_j}{\lambda_j + \mu_j}\big(k_{j-1,j} + k_{j,j+1}\big),$$

and therefore $k_{j,j+1} = \lambda_j^{-1} + \big(\frac{\mu_j}{\lambda_j}\big)k_{j-1,j}$ for $j \geq 1$ and $k_{0,1} = \lambda_0^{-1}$. The solution follows by iteration. ◇

Let a Markov process $(X_t)_{t\geq 0}$ on a countable state space I be given. A state $i \in I$ is called **recurrent** if $\mathbb{P}_i(\{t \geq 0 \colon X_t = i\}$ is unbounded$) = 1$, and the state $i \in I$ is called **transient** if $\mathbb{P}_i(\{t \geq 0 \colon X_t = i\}$ is unbounded$) = 0$. The first passage (or hitting) time of the process to the state $i \in I$ when starting in state k is defined as

$$T_{k,i} = \inf\{t \geq J_1 \colon X_t = i\}.$$

(we write T_i if $X_0 = k$ is clear from the context). The Q-matrix of the process $(X_t)_{t\geq 0}$ is called **irreducible** if the whole state space I is a single class with respect to the $\longleftrightarrow$ - equivalence relation defined above.

Notation 3.4 (Invariant distribution) *Let $(X_t)_{t\geq 0}$ be a Markov process on a countable state space I.*

(a) A vector $\lambda = (\lambda(i))_{i\in I}$, $\lambda \in \mathcal{M}_1(I)$ (set of probability measures on I), is called an **invariant** *distribution, or a* **stationary***, or an* **equilibrium probability measure** *if for all $t \geq 0$ and for all states $j \in I$ it holds that*

$$\mathbb{P}(X_t = j) = \lambda(j), \text{ i.e. } \lambda P_t = \lambda.$$

(b) A vector $(\lambda(j))_{j\in I}$ with $\lambda(j) \geq -0$, $\sum_{j\in I} \lambda \neq 1$, and $\lambda P_t = \lambda$ for all $t \geq 0$ is called an **invariant measure***. If in addition $\sum_{j\in I} \lambda(j) < \infty$ holds, an invariant distribution (equilibrium probability measure) is given via*

$$\widetilde{\lambda}(j) = \lambda_j\big(\sum_{i\in I} \lambda(i)\big)^{-1}.$$

Proposition 3.5 *Assume that $(X_t)_{t\geq 0}$ is a non-explosive Markov process on a countable state space I with Q-matrix $Q = (q_{i,j})_{i,j\in I}$.*

(a) Then $\lambda = (\lambda(i))_{i\in I}$ is an invariant measure for the process $(X_t)_{t\geq 0}$ if and only if for all states $j \in I$

$$\sum_{i\in I} \lambda(i) q_{i,j} = 0, \quad \textit{i.e.} \quad \lambda Q = 0.$$

(b) Assume in addition that $q_i > 0$ for all $i \in I$ and let Π be the transition matrix of the jump chain $(Y_n)_{n\in\mathbb{N}_0}$ and $\lambda = (\lambda(i))_{i\in I}$. Then

$$\lambda \textit{ invariant measure for } (X_t)_{t\geq 0} \Leftrightarrow \mu\Pi = \mu \textit{ with } \mu(i) = \lambda(i) q_i, i \in I.$$

Proof. (a) λ is an invariant measure if $\mathbb{P}(X_t = j) = \lambda(j)$ for all $t \geq 0$ and all $j \in I$, i.e. $\lambda P_t = \lambda$. Thus the row vector λ is annihilated by the matrix Q:

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}t} \lambda P_t = \lambda \frac{\mathrm{d}}{\mathrm{d}t} P_t = \lambda P_t Q, \\ 0 &= \lambda P_t Q\Big|_{t=0} = \lambda Q. \end{aligned}$$

This argument cannot work for an explosive chain as in this case one cannot guarantee that $\frac{\mathrm{d}}{\mathrm{d}t}(\lambda P_t) = 0$.

(b) Write $\mu\Pi = \mu$ as $\mu\Pi - \mu\mathbb{1} = 0$, or

$$\begin{aligned} (\mu\Pi - \mu\mathbb{1})_j &= \sum_{i\in I\setminus\{j\}} \mu_i \frac{q_{i,j}}{q_i} - \mu_j = \sum_{i\in I} \mu_i \Big(\big(1 - \delta_{i,j}\big) \frac{q_{i,j}}{q_i} - \delta_{i,j} \Big) \\ &= \sum_{i\in I} \Big(\frac{q_{i,j}}{q_i} - \delta_{i,j} \Big(1 + \frac{q_{i,j}}{q_i} \Big) \Big) = \sum_{i\in I} \lambda(i) q_{i,j} = (\lambda Q)_j. \end{aligned}$$

Now the LHS is zero if and only if the RHS is. □

Example. Birth-and-death process: Assume that $\lambda_n >$ for all $n \in \mathbb{N}_0$ and $\mu_n > 0$ for all $n \in \mathbb{N}$,$\mu_0 = 0$, that is all states communicate. The corresponding jump chain $(Y_n)_{n\in\mathbb{N}_0}$ has transition matrix Π defined as

$$\pi_{n,n-1} = \frac{\mu_n}{\lambda_n + \mu_n}, \quad \pi_{n,n+1} = \frac{\lambda_n}{\lambda_n + \mu_n}.$$

It is easy to show that the following equivalence holds:

$$(X_t)_{t\geq 0} \text{ recurrent } \Leftrightarrow (Y_n)_{n\in\mathbb{N}_0} \text{ recurrent.}$$

◇

Example. Irreducible Birth-and-death process (BDP): Assume that $\lambda_n >$ for all $n \in \mathbb{N}_0$ and $\mu_n > 0$ for all $n \in \mathbb{N}$,$\mu_0 = 0$, that is all states communicate, i.e. the Q-matrix is irreducible. The corresponding jump chain $(Y_n)_{n\in\mathbb{N}_0}$ has transition matrix Π defined as

$$\pi_{n,n-1} = \frac{\mu_n}{\lambda_n + \mu_n}, \quad \pi_{n,n+1} = \frac{\lambda_n}{\lambda_n + \mu_n}.$$

If the jump chain $(Y_n)_{n\in\mathbb{N}_0}$ is transient then

$$p(n) = \mathbb{P}_n(\text{chain ever reaches } 0) \to 0 \text{ as } n \to \infty.$$

Clearly,

$$\begin{aligned} p(0) &= 1, \\ p(n)(\mu_n + \lambda_n) &= p(n-1)\mu_n + p(n+1)\lambda_n. \end{aligned}$$

We shall find the function $p(n)$. We get by iteration

$$\begin{aligned} p(n) - p(n+1) &= \frac{\mu_n}{\lambda_n}\Big(p(n-1) - p(n)\Big), \quad n \geq 1, \\ p(n) - p(n+1) &= \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n}\Big(p(0) - p(1)\Big), \end{aligned}$$

and thus

$$\begin{aligned} p(n+1) &= \Big(p(n+1) - p(0)\Big) + p(0) \\ &= \sum_{j=0}^{n}\Big(p(j+1) - p(j)\Big) + 1 = \Big(p(1) - 1\Big)\sum_{j=0}^{n}\frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j} + 1, \end{aligned}$$

where by convention the term for $j = 0$ is equal to one. We can find a nontrivial solution for the function $p(n)$ if the sum converges. Hence, we can derive the following fact:

Fact I: BDP is transient $\Leftrightarrow \sum_{n=1}^{\infty} \frac{\mu_1 \cdots \mu_n}{\lambda_1 \cdots \lambda_n} < \infty$. ◇

Recall that a state $i \in I$ is recurrent if $\mathbb{P}_i(T_i < \infty) = 1$, so this state is visited for indefinitely large times. As in the previous example consider a Markov chain $(Y_n)_{n\in\mathbb{N}_0}$ on a countable state space I such that a limiting probability measure $\nu \in \mathcal{M}_1(I)$ exists, that is

$$\lim_{n\to\infty} P_n(x, y) = \nu(x) \quad \text{for all } x, y \in I,$$

where P_n is the n-step transition function (entry of the n-th power of the transition matrix Π). However, if the chain $(Y_n)_{n\in\mathbb{N}_0}$ is transient then we have seen that $\lim_{n\to\infty} P_n(x,y) = 0$ for all $x, y \in I$. Hence, in this case no limiting probability measure exists. However, $\lim_{n\to\infty} P_n(x,y) = 0$ can hold for a recurrent chain. This shows the example of the simple random walk for which we proved that $P_{2n}(0,0) \to 0$ as $n \to \infty$. This motivates to define two types of recurrence:

null-recurrent is recurrent but $\lim_{n\to\infty} P_n(x,y) = 0$ for all $x, y \in I$, otherwise **positive recurrent**.

Definition 3.6 *A state $i \in I$ is called* **positive recurrent (PR)** *if $m_i := \mathbb{E}_i(T_i) < \infty$, and it is called* **null-recurrent (NR)** *if $\mathbb{P}_i(T_i < \infty) = 1$ but $m_i = \infty$.*

It turns out that positive recurrent processes behave very similar to Markov chains with finite state spaces. The following Theorem is important as it connects the mean return time m_i to an invariant distribution.

Theorem 3.7 *Let $(X_t)_{t\geq 0}$ be an irreducible and recurrent Markov process on a countable state space I and Q-matrix $Q = (q_{i,j})_{i,j\in I}$. Then:*

(a) either every state $i \in I$ is PR or every state $i \in I$ is NR;

(b) or Q is PR if and only if it has a (unique) invariant distribution $\pi = (\pi(i))_{i\in I}$, in which case

$$\pi(i) > 0 \text{ and } m_i = \frac{1}{\pi(i) q_i} \text{ for all } i \in I.$$

Proof. We give a brief sketch.

$$\begin{aligned} m_i &= \text{mean return time to } i \\ &= \text{mean holding time at } i + \sum_{j\in I\setminus\{i\}} (\text{mean time spent at } j \text{ before return to } i). \end{aligned}$$

The first term on the right hand side is clearly $q_i^{-1} =: \gamma_i$ and for $j \neq i$ we write

$$\begin{aligned} \gamma_j &= \mathbb{E}_i(\text{mean time spent at } j \text{ before return to } i) \\ &= \mathbb{E}_i\Big(\int_{J_1}^{T_i} 1\!\!1\{X(t) = j\}\,\mathrm{d}t\Big) \\ &= \int_0^\infty \mathbb{E}_i\Big(1\!\!1\{X(t) = j, J_1 < t < T_i\}\Big)\,\mathrm{d}t. \end{aligned}$$

Then

$$m_i = \sum_{j\in I}\gamma_j = \frac{1}{q_i} + \sum_{j\in I\setminus\{i\}}\gamma_j \begin{cases} <\infty & \text{if state } i \text{ is PR,}\\ =\infty & \text{if state } i \text{ is NR.}\end{cases}$$

This defines the vector $\gamma^{(i)} = (\gamma_j^{(i)})_{j\in I}$ where we put the index to stress the dependence on the state $i\in I$. If T_i^Y is the return time to state i of the jump chain $(Y_n)_{n\in\mathbb{N}_0}$ then we get

$$\begin{aligned}\gamma_j^{(i)} &= \mathbb{E}_i\Big(\sum_{n\in\mathbb{N}_0}(J_{n+1}-J_n)1\!\!1\{Y_n=j, n<T_i^Y\}\Big)\\ &= \sum_{n\in\mathbb{N}_0}\mathbb{E}_i\Big((J_{n+1}-J_n)|Y_n=j\Big)\mathbb{P}_i(Y_n=j, 1\le n<T_i^Y)\\ &= \frac{1}{q_j}\mathbb{E}_i\Big(\sum_{n=1}^{Y_i^Y-1}1\!\!1\{Y_n=j\}\Big) =: \frac{\widetilde{\gamma}_j^{(i)}}{q_j},\end{aligned}$$

where we set $\widetilde{\gamma}_i^{(i)}=1$, and for $j\neq i$,

$$\begin{aligned}\widetilde{\gamma}_i^{(i)} &= \mathbb{E}_i\Big(\sum_{n=1}^{T_i^Y-1}1\!\!1\{Y_n=j\}\Big)\\ &= \mathbb{E}_i\Big(\text{time spent at } j \text{ in } (Y_n)_{n\in\mathbb{N})} \text{ before returning to } i\Big)\\ &= \mathbb{E}_i\Big(\text{number of visits to } j \text{ before returning to } i\Big).\end{aligned}$$

If the process $(X_t)_{t\ge0}$ is recurrent then so does the jump chain $(Y_n)_{n\in\mathbb{N}_0}$. Then the vector $\widetilde{\gamma}^{(i)}$ gives an invariant measure with $\widetilde{\gamma}_j^{(i)}<\infty$, and all invariant measures are proportional to $\widetilde{\gamma}^{(i)}$. Then the vector $\gamma^{(i)}$ with $\gamma_j^{(i)} = \frac{1}{q_j}\widetilde{\gamma}_j^{(i)}$ gives an invariant measure for the process $(X_t)_{t\ge0}$. Furthermore, all invariant measures are proportional to $\gamma^{(i)}$. If the state i is positive recurrent, then

$$m_i = \sum_{j\in I}\gamma_j^{(j)} < \infty.$$

But then $m_k = \sum_{j\in I}\gamma_j^{(k)}<\infty$, for all k, i.e. all states become positive recurrent. Similarly, if i is null-recurrent, then that applies to all states as well. Hence (a). If Q is PR, then

$$\pi_j = \frac{\gamma_j^{(i)}}{\sum_{j\in I}\gamma_j^{(i)}} = \frac{1}{q_i m_i},\quad j\in I,$$

yields a (unique) invariant distribution π. Clearly, $\pi_i>0$ and

$$\mathbb{E}_k(\text{time spent at } j \text{ before returning to } k) = \frac{\pi_j}{\pi_k q_k}.$$

Conversely, if $(X_t)_{t\geq 0}$ has an invariant distribution π then all invariant measures have finite sum. This implies that $m_i = \sum_{j\in I}\gamma_j^{(i)} < \infty$, henceforth i is positive recurrent. □

Let us give some summary:

(I) Irreducible Markov processes $(X_t)_{t\geq 0}$ with $|I| > 1$ have rates $q_i > 0$ for all $i \in I$.

(II) Non-explosive Markov processes can be transient or recurrent.

(III) Irreducible Markov processes can be

  (a) null-recurrent, i.e. $m_i = \infty$, no invariant measure λ with $\sum_{i\in I}\lambda(i) < \infty$ exists.

  (b) positive recurrent, i.e. $m_i < \infty$, unique invariant distribution $\lambda = (\lambda(i)_{i\in I}$ with $\mathbb{E}_i(T_i) = \frac{1}{\lambda(i)q_i}$.

(IV) Explosive Markov processes are always transient.

We state the following large time results without proof as this goes beyond the level of the course. It is, however, important to realise how this is linked to invariant distributions.

Theorem 3.8 *Let $(X_t)_{t\geq 0}$ be a Markov process on a countable state space I with initial distribution $\lambda \in \mathcal{M}_1(I)$ and Q-matrix $Q = (q_{ii,j})_{i,j\in I}$ and with invariant distribution $\pi = (\pi(i))_{i\in I}$. Then for all states $i \in$ we get, as $t \to \infty$,*

(I)

$$\frac{1}{t}\int_0^t \mathbb{1}\{X_s = i\}\,\mathrm{d}s = \textit{fraction of time at } i \textit{ in } (0,t) \longrightarrow \pi_i$$

$$= \frac{1}{m_i q_i} = \frac{\textit{mean holding time at } i}{\textit{mean return time to } i}.$$

(II)

$$\frac{1}{t}\mathbb{E}\Big(\int_0^t \mathbb{1}\{X_s = i\}\,\mathrm{d}s\Big) = \frac{1}{t}\int_0^t \mathbb{P}(X_s = i)\,\mathrm{d}s \longrightarrow \pi_i.$$

Proposition 3.9 (Convergence to quilibrium) *Let Q be an irreducible and non-explosive Q-matrix with semigroup $(P_t)_{t\geq 0}$ and invariant distribution $\pi = (\pi(i))_{i\in I}$. Then for all $i,j \in I$*

$$P_t(i,j) \longrightarrow \pi(j) \textit{ as } t \to \infty.$$

Example. Recurrent BDP: Assume that $\lambda_n >$ for all $n \in \mathbb{N}_0$ and $\mu_n > 0$ for all $n \in \mathbb{N}$,$\mu_0 = 0$, that is all states communicate. Positive recurrence of the BDP implies that $\lim_{t\to\infty} \mathbb{P}(X_t = n|X_0 = m) = \pi(n)$ for all $m \in \mathbb{N}_0$. If the process is in the limiting probability, i.e., if $\mathbb{P}(X_t = n) = \pi(n)$, then $P_t'(n) = 0$. Recall that $P_t(n) = \mathbb{P}(X_t = n)$ and that

$$P_t'(n) = \mu_{n+1}P_t(n+1) + \lambda_{n-1}P_t(n-1) - (\mu_n + \lambda_n)P_t(n).$$

Then the limiting probability $\pi = (\pi(n))_{n\in\mathbb{N}_0}$ should solve

$$0 = \lambda_{n+1}\pi)n-1) + \mu_{n+1}\pi(n+1) - (\lambda_n + \mu_n)\pi(n).$$

We solve this directly: $n = 0$ gives $\pi(1) = \frac{\lambda_0}{\mu_1}\pi(0)$ and for $n \geq 1$ we get $\mu_{n+1}\pi(n+1) - \lambda_n\pi(n) = \mu_n\pi(n) - \lambda_{n-1}\pi(n-1)$. Iterating this yields

$$\mu_{n+1}\pi(n+1) - \lambda_n\pi(n) = \mu_1\pi(1) - \lambda_0\pi(0).$$

Hence, $\pi(n+1) = \big(\frac{\lambda_n}{\mu_{n+1}}\big)\pi(n)$ and thus $\pi(n) = \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}\pi(0)$.

Fact II: BDP

$$\text{positive recurrent } \Leftrightarrow q := \sum_{n=0}^{\infty} \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n} < \infty,$$

in which case $\pi(n) = \frac{\lambda_0\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}q^{-1}$.

Definition 3.10 (Reversible process) *A non-explosive Markov process $(X_t)_{t\geq 0}$ with state space I and Q-matrix Q is called* **reversible** *if for all $i_0, \dots i_n \in I, n \in \mathbb{N}$, and times $0 = t_0 < t_1 < \cdots < t_n = T, T > 0$,*

$$\mathbb{P}(X_0 = i_0, \dots, X_T = i_n) = \mathbb{P}(X_0 = i_n, \dots, X_{T-t_1} = i_1, X_T = i_0).$$

Equivalently,

$$(X_t \colon 0 \leq t \leq T) \sim (X_{T-t} \colon 0 \leq t \leq T) \quad \textit{for all } T > 0,$$

where $\sim$ stands for equal in distribution. Note that in order to define the reversed process one has to fix a time $T > 0$.

Theorem 3.11 (Detailed balance equations) *A non-explosive Markov process $(X_t)_{t\geq 0}$ with state space I and Q-matrix $Q = (q_{i,j})_{i,j\in I}$ and initial distribution $\lambda = (\lambda(i))_{i\in I}$ is reversible if and only if the* **detailed balance equations (DBEs)**

$$\lambda(i)q_{i,j} = \lambda(j)q_{j,i} \quad \textit{for all } i, j \in I, i \neq j,$$

hold.

Proof. Suppose the detailed balance equations hold. Hence,

$$(\lambda Q)_j = \sum_{i\in I} \lambda(i) q_{i,j} = \sum_{i\in I} \lambda_j q_{j,i} = 0.$$

By induction, the DBEs hold for all powers of Q,

$$\lambda(i) q^{(k)}_{i,j} = \lambda(i) \sum_{l\in I} q_{i,l} q^{(k-1)}_{l,j} = \sum_{l\in I} q_{l,i} \lambda(l) q^{(k-1)}_{l,j} = \sum_{l\in I} q^{(k-1)}_{j,l} q_{l,i} \lambda(j) = \lambda(j) q^{(k)}_{j,i}.$$

Henceforth

$$\lambda(i) P_t(i,j) = \lambda(j) P_t(j,i), \quad t \ge 0. \tag{3.11}$$

We shall check that

$$\mathbb{P}(X_{t_k} = i_k, 0 \le k \le n) = \mathbb{P}(X_{T-t_k} = i_k, 0 \le k \le n). \tag{3.12}$$

Using (3.11) several times we get

$$\begin{aligned} \text{L.H.S. of } (3.12) &= \lambda(i_0) P_{t_1-t_0}(i_0,i_1) \cdots P_{t_n-t_{n-1}}(i_{n-1},i_n) \\ &= P_{t_1-t_0}(i_1,i_0) \lambda(i_1) P_{t_2-t_1}(i_1,i_2) \cdots P_{t_n-t_{n-1}}(i_{n-1},i_n) \\ &= \cdots \\ &= P_{t_1-t_0}(i_0,i_1) P_{t_2-t_1}(i_1,i_2) \cdots P_{t_n-t_{n-1}}(i_n,i_{n-1}) \lambda(i_n). \end{aligned}$$

We rearrange this to obtain the right hand side of (3.12) as

$$\lambda(i_n) P_{t_n-t_{n-1}}(i_n,i_{n-1}) \cdots = \mathbb{P}(X_{T-t_k} = i_k, 0 \le k \le n).$$

Conversely, suppose now that the process is reversible and put $n = 1, i_0 = i \in I, j_0 = j \in I$ and let $T > 0$. Then reversibility gives

$$\lambda(i) P_T(i,j) = \lambda(j) P_t(j,i).$$

We differentiate this with respect to the parameter T and set $T = 0$ to obtain the DBEs using that $\frac{\mathrm{d}}{\mathrm{d}t} P_t(i,j) = q_{i,j}$. □

Notation 3.12 (Time reversed process) *We denote the time reversed process (reversed about $T > 0$) by $(X^{(\mathrm{tr})}_t)_{0\le t\le T}$ which is defined by*

$$\mathbb{P}(X^{(\mathrm{tr})}_0 = i_0, \dots, X^{(\mathrm{tr})}_T = i_n) = \mathbb{P}(X_0 = i_n, \dots, X_{T-t_1} = i_1, X_T = i_0).$$

3.3 Application to queuing theory

We give a very brief introduction to queueing theory. Let the state space $I = \mathbb{N}_0$. Furthermore, consider the jump rates $q_{i,i+1} = \lambda > 0$ and $q_{i,i-1} = \mu > 0$. The most simple model is the M/M/1 queue. Here, the first two M's are standing for memoryless inter-arrival times and memoryless service times, and the '1' indicate that there is only one server. Customers are joining the queue, and we denote this arrival process by $(A(t))_{t\geq 0}$. The arrival process is a Poisson process with parameter $\lambda > 0$. The service times are independent identical distributed exponential random variables with parameter $\mu > 0$. The queueing rule here is FCFS (first comes first served) also called FIFO (first in first out). The process $(D(t))_{t\geq 0}$ denote the number $D(t)$ of customers served up to time time $t \geq$. The queue length $L(t)$ at time t is the following composed process

$$L(t) = A(t) - D(t), \quad t \geq 0,$$

having the Q-matrix $Q = (q_{i,j})_{i,j\in\mathbb{N}_0}$ with $q_{0,0} = -\lambda, q_{0,1} = \lambda, q_{i,i-1} = \mu, q_{i,i+1} = \lambda, q_{i,i} = -(\mu+\lambda)$ for all $i \in \mathbb{N}$. We call the process $(L(t))_{t\geq 0}$ the M/M/1-queue. We have to convince ourself that this process is a Markov process. For that assume that at time $t = 0$ there are $i > 0$ customers joining the queue. Let T be the time to serve the first customer of this queue and A be the time of the next arrival. The first jump time $J_1 = A \wedge T$ which is exponential distributed with parameter $\lambda + \mu$ (see Proposition 2.20). $L_{J-1} = i-1$ if $T < A$ and $L_{J_1} = i+1$ if $T > A$, both events are independent of J_1, with probabilities $\mu/(\mu+\lambda)$ and $\lambda/(\mu+\lambda)$ respectively. Conditional on $J_1 = T$, then $A - J_1$ is exponential with parameter λ independent of J_1, and conditional on $J_1 = A, T - J_1$ is exponential with parameter μ independent of J_1. Hence, conditional on $L(J_1) = j$ the process $(L(t))_{t\geq 0}$ begins afresh from j at time J_1.

Hitting: We are in particular interested when there is no queue, i.e. $h_i = \mathbb{P}_i(\text{hit } 0), i \geq 0$. Clearly, $h_0 = 1$, and Proposition 3.3 gives $\sum_{j\in\mathbb{N}_0} q_{i,j}h_j = 0$, i.e.

$$-(\lambda+\mu)h_i + q_{i,i+1}h_{i+1} + q_{i,i-1}h_{i-1} = 0,$$

and thus

$$(\lambda+\mu)h_i = \lambda h_{i+1} + \mu h_{i-1}, i \geq 1. \tag{3.13}$$

The general solution to (3.13) is given as

$$h_i = \begin{cases} A + B\left(\frac{\mu}{\lambda}\right) & \text{if } \lambda \neq \mu, \\ A + Bi & \text{if } \lambda = \mu. \end{cases}$$

$\lambda \leq \mu$: the minimal non-negative solution must have $B = 0, A = 1$, giving $h_i = 1$, and the M/M/1-queue is recurrent.

$\lambda > \mu$: then $A = 0$ and $B = 1$ giving $h_i = (\mu/\lambda)^i, i \geq 0$, and the M/M/1-queue is transient.

Invariant distribution: We try first the DBEs: $\pi_i \lambda = \pi_{i+1}\mu, i \geq 1$, which gives

$$\pi_{i+1} = \big(\frac{\lambda}{\mu}\big)\pi_i = \cdots = \big(\frac{\lambda}{\mu}\big)^i \pi_0$$

Normalising gives the value for π_0, i.e.

$$1 = \sum_{i \in \mathbb{N}_0} \pi_i = \pi_0 \sum_{i \in \mathbb{N}_0} \big(\frac{\lambda}{\mu}\big)^i = \pi_0 \big(1 - \frac{\lambda}{\mu}\big)^{-1}.$$

Finally $\pi_i = (1-\rho)\rho^i$ with $\rho = \frac{\lambda}{\mu}$. From the facts (I and II) of BDPs we get that $\lim_{t\to\infty} \mathbb{P}(L(t) = j) = (1-\rho)\rho^j$ for $\mu > \lambda$.

Mean return time to 0: $m_0 = \frac{1}{\pi_0 q_0}$, and the mean busy period is therefore $m_0 - \frac{1}{\lambda} = \frac{1}{\mu - \lambda}$ $(\mu > \lambda)$.

Mean waiting time W of a customer (in equilibrium):

$$\mathbb{E}(W) = \sum_{i \in \mathbb{N}_0} \mathbb{E}(W|L = i)\pi_i = \sum_{i \in \mathbb{N}_0} \frac{i}{\mu}(1-\rho)\rho^i = \frac{\lambda}{\mu(\mu - \lambda)},$$

and hence the mean sojourn time (waiting time plus service time) is $1/(\mu - \lambda)$ which equals the mean busy period.

Theorem 3.13 (Burke's theorem) *Let $\rho = \lambda/\mu < 1$ and $(L(t))_{t\geq 0}$ be a M/M/1-queue. Then*

(a) With $L(t) = L(0) + A(t) - D(t)$ it holds that

$$(D(t))_{t\geq 0} \sim (A(t))_{t\geq 0}.$$

(b) For all $T > 0$, the process $(L(t+T))_{t\geq 0}$ is independent of $(D(t))_{0\leq t\leq T}$.

Proof. For $\lambda < \mu$ the process $(L(t))_{t\geq 0}$ is reversible. Jumps up of the trajectory become jumps down when we reverse the time. Pick $T > 0$ and times $0 = t_0 < t_1 < \cdots < t_n = T$ and states $i_0, \ldots, i_n \in \mathbb{N}_0$, then

$$\mathbb{P}(L_0^{(\mathrm{tr})} = i_0, \ldots, L_T^{(\mathrm{tr})} = i_n) = \mathbb{P}(L_0 = i_n, \ldots, L_{T-t_1} = i_1, L_T = i_0).$$

Hence, writing $L^{(\mathrm{tr})}(t) = A^{(\mathrm{tr})}(t) - D^{(\mathrm{tr})}(t)$,

$$(A(t))_{t\geq 0} \sim (D^{(\mathrm{tr})}(t))_{t\geq 0}; (D(t))_{t\geq 0} \sim (A^{(\mathrm{tr})}(t))_{t\geq 0}$$

But $(D^{(\mathrm{tr})}(t)) \sim (D(t))$ and $(A^{(\mathrm{tr})}(t)) \sim (A(t))$ by reversibility, henceforth

$$(A(t)) \sim (D^{(\mathrm{tr})}(t)) \sim (D(t)) \sim (A^{(\mathrm{tr})}(t)) \sim (A(t)).$$

□

If we set $\lambda = \mu$, i.e. $\rho = 1$ we get that the M/M/1-queue is null-recurrent. If $\lambda > \mu$ we get that the M/M/1-queue is transient, i.e.

$$\mathbb{P}(\lim_{t\to\infty} L(t) = \infty) = 1.$$

4 Percolation theory

We give a brief introduction in percolation theory in Section 4.1, provide important basic tools in Section 4.2, and study and prove the important Kesten Theorem for bond percolation in Section 4.3.

4.1 Introduction

Percolation theory was founded by Broadbent and Hammersley 1957, in order to model the flow of a fluid in a porous medium with randomly blocked channels. Percolation is a simple probabilistic model which exhibits a phase transition (as we explain below). The simplest version takes place on $\mathbb{Z}^2$, which we view as a graph with edges between neighboring vertices. All edges of $\mathbb{Z}^2$ are, independently of each other, chosen to be open with probability p and closed with probability $1-p$. A basic question in this model is "What is the probability that there exists an open path, i.e., a path all of whose edges are open, from the origin to the exterior of the square $\Lambda_n := [-n, n]^2$?" This question was raised by Broadbent in 1954 at a symposium on Monte Carlo methods. It was then taken up by Broadbent and Hammersley, who regarded percolation as a model for a random medium. They interpreted the edges of $\mathbb{Z}^2$ as channels through which fluid or gas could flow if the channel was wide enough (an open edge) and not if the channel was too narrow (a closed edge). It was assumed that the fluid would move wherever it could go, so that there is no randomness in the behavior of the fluid, but all randomness in this model is associated with the medium. We shall use 0 to denote the origin. A limit as $n \to \infty$ of the question raised above is "What is the probability that there exists an open path from 0 to infinity?" This probability is called the percolation probability and denoted by $\theta(p)$. Clearly $\theta(0) = 0$ and $\theta(1) = 1$, since there are no open edges at all when $p = 0$ and all edges are open when $p = 1$. It is also intuitively clear that the function $p \mapsto \theta(p)$ is nondecreasing. Thus the graph of θ as a function of p should have the form indicated in Figure (XX), and one can define the critical probability by $p_{\mathrm{c}} = \sup\{p \in [0,1] \colon \theta(p) = 0\}$. Why is this model interesting? In order to answer this we define the (open) cluster $C(x)$ of the vertex $x \in \mathbb{Z}^2$ as the collection of points connected to x by an open path. The clusters $C(x)$ are the maximal connected components of the collection of open edges of $\mathbb{Z}^2$, and $\theta(p)$ is the probability that $C(0)$ is infinite. If $p < p_{\rm c}$, then $\theta(p) = 0$ by definition, so that $C(0)$ is finite with probability 1. It is not hard to see that in this case all open clusters are finite. If $p > p_{\rm c}$, then $\theta(p) > 0$ and there is a strictly positive probability that $C(0)$ is infinite. An application of Kolmogorov's zero-one law shows that there is then with probability 1 some infinite cluster. In fact, it turns out that there is a unique infinite cluster. Thus, the global behavior of the system is quite different for $p < p_{\rm c}$ and for $p > p_{\rm c}$. Such a sharp transition in global behavior of a system at some parameter value is called a phase transition or a critical phenomenon by statistical physicists, and the parameter value at which the transition takes place is called a critical value. There is an extensive physics literature on such phenomena. Broadbent and Hammersley proved that $0 < p_{\rm c} < 1$ for percolation on Z^2, so that there is indeed a nontrivial phase transition. Much of the interest in percolation comes from the hope that one will be better able to analyze the behavior of various functions near the critical point for the simple model of percolation, with all its built-in independence properties, than for other, more complicated models for disordered media. Indeed, percolation is the simplest one in the family of the so-called random cluster or Fortuin-Kasteleyn models, which also includes the celebrated Ising model for magnetism. The studies of percolation and random cluster models have influenced each other.

Let us begin and collect some obvious notations. $\mathbb{Z}^d, d \geq 1$, is the set of all vectors $x = (x_1, \dots, x_d)$ with integral coordinates. The (graph-theoretic) distance $\delta(x, y)$ from x to y is defined by

$$\delta(x, y) = \sum_{i=1}^{d} |x_i - y_i|,$$

and we write $|x|$ for $\delta(0, x)$. We turn $\mathbb{Z}^d$ into a graph, called the d-dimensional cubic lattice, by adding edges between all pairs x, y of points of $\mathbb{Z}^d$ with $\delta(x, y) = 1$. We write for this graph $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ where $\mathbb{E}^d$ is the set of edges. If $\delta(x, y) = 1$ we say that x and y are adjacent, and we write in this case $x \sim y$ and represent the edges from x to y as $\langle x, y \rangle$. We shall introduce now some probability. Denote by

$$\Omega = \prod_{e \in \mathbb{E}^d} \{0, 1\} = \{0, 1\}^{\mathbb{Z}^d} = \{\omega \colon \mathbb{E}^d \to \{0, 1\}\}$$

the set of configurations $\omega = (\omega(e))_{e \in \mathbb{E}^d}$ (set of all mappings $\mathbb{E}^d \to \{0, 1\}$)with the interpretation that the edge $e \in \mathbb{E}^d$ is **closed** for the configuration ω if

$\omega(e) = 0$ and the edge e is **open** for the configuration ω if $\omega(e) = 1$. The set Ω will be our sample or probability space. We need further a σ-algebra and a measure for this sample space. An obvious choice for the σ-algebra of events is the one which is generated by all **cyclinder** events $\{\omega \in \Omega \colon \omega(e) = a_e, a_e \in \{0,1\}, e \in \Delta, \Delta \subset \mathbb{E}^d \text{ finite } \}$, and we call it $\mathcal{F}$. For every $e \in \mathbb{E}^d$ let μ_e be the Bernoulli (probability) measure on $\{0,1\}$ defined by

$$\mu_e(\omega(e) = 0) = 1 - p \text{ and } \mu_e(\omega(e) = 1) = p, \quad p \in [0,1].$$

Then the product of these measures defines a probability measure on the space of configurations Ω, denoted by

$$\mathbb{P}_p = \prod_{e \in \mathbb{E}^d} \mu_e.$$

In the following we are going to consider only the measure $\mathbb{P}_p \in \mathcal{M}_1(\Omega)$ for different parameters $p \in [0,1]$. As the probability measure $\mathbb{P}_p$ is a product measure (over all edges) it is a model for the situation where each edges is open (or closed) independently of all other edges with probability p (respectively with probability $1-p$). If one considers a probability measure on Ω which is not a product of probability measures on the single edges, one calls the corresponding percolation model dependent. In this course we only study independent (hence the product measure $\mathbb{P}_p$) percolation models.

A **path** in $\mathbb{L}^d$ is an alternating sequence $x_0, e_0, x_1, e_1, \ldots, e_{n-1}, x_n$ of distinct vertices x_i and edges $e_i = \langle x_i, x_{i+1} \rangle$; such a path has length n and is said to connect x_0 to x_n. A **circuit** is a closed path. Consider the random subgraph of $\mathbb{L}^d$ containing the vertex set $\mathbb{Z}^d$ and the open edges (bonds) only. The connected components of the graph are called **open clusters**. We write $C(x)$ for the open cluster containing the vertex x. If A and B are set of vertices we write $A \longleftrightarrow B$ if there exists an open path joining some vertex in A to some vertex in B. Hence,

$$C(x) = \{y \in \mathbb{Z}^d \colon x \longleftrightarrow y\},$$

and we denote by $|C(x)|$ the number of vertices in $C(x)$. As above we write $C = C(0)$ for the open cluster containing the origin.

$$\theta(p) = \mathbb{P}_p(|C| = \infty) = 1 - \sum_{n=1}^{\infty} \mathbb{P}_p(|C| = n).$$

It is fundamental to percolation theory that there exists a critical value $p_{\rm c} = p_{\rm c}(d)$ of p such that

$$\theta(p) \begin{cases} = 0 & \text{if } p < p_{\rm c}, \\ > 0 & \text{if } p > p_{\rm c}; \end{cases}$$

$p_c(d)$ is called the **critical probability**. As above $p_c(d) = \sup\{p \in [0,1] \colon \theta(p) = 0\}$. In dimension $d = 1$ for any $p < 1$ there exist infinitely many closed edges to the left and to the right of the origin almost surely, implying $\theta(p) = 0$ for $p < 1$, and thus $p_c(1) = 1$. The situation is quite different for higher dimensions. Note that the d-dimensional lattice $\mathbb{L}^d$ may be embedded in $\mathbb{L}^{d+1}$ in a natural way as the projection of $\mathbb{L}^{d+1}$ onto the subspace generated by the first d coordinates; with this embedding, the origin of $\mathbb{L}^{d+1}$ belongs to an infinite open cluster for a particular value of p whenever it belongs to an infinite open cluster of the sublattice $\mathbb{L}^d$. Thus

$$p_c(d+1) \leq p_c(d), \quad d \geq 1.$$

Theorem 4.1 *If* $d \geq 2$ *the* $0 < p_c(d) < 1$.

This means that in two or more dimension there are two phases of the process. In the **subcritical phase** $p < p_c(d)$, every vertex is almost surely in a finite open cluster. In the **supercritical phase** when $p > p_c(d)$, each vertex has a strictly positive probability of being in an infinite open cluster.

Theorem 4.2 *The probability* $\Psi(p)$ *that there exists an infinite open cluster satisfies*

$$\Psi(p) = \begin{cases} 0 & \text{if } \theta(p) = 0, \\ 1 & \text{if } \theta(p) > 0. \end{cases}$$

We shall prove both theorems in the following. For that we derive the following non-trivial upper and lower bounds for $p_c(d)$ when $d \geq 2$.

$$\frac{1}{\lambda(2)} \leq p_c(2) \leq 1 - \frac{1}{\lambda(2)}, \tag{4.14}$$

and

$$\frac{1}{\lambda(d)} \leq p_c(d) \quad \text{for } d \geq 3; \tag{4.15}$$

where $\lambda(d)$ is the connective constant of $\mathbb{L}^d$, defined as

$$\lambda(d) = \lim_{n \to \infty} \sqrt[n]{\sigma(n)},$$

with $\sigma(n)$ being the number of paths (or 'self-avoiding walks') of $\mathbb{L}^d$ having length n and beginning at the origin. It is obvious that $\lambda(d) \leq 2d - 1$; to see this, note that each new step in a self-avoiding walk has at most $2d - 1$ choices since it must avoid the current position. Henceforth $\sigma(n) \leq 2d(2d-1)^{n-1}$. Inequality (4.15) implies that $(2d-1)p_c(d) \geq 1$, and it is known that further $p_c(d) \sim (2d)^{-1}$ as $d \to \infty$.

Proof of Theorem 4.1 and (4.14)**.** As $p_c(d+1) \le p_c(d)$ it suffices to show that $p_c(d) > 0$ for $d \ge 2$ and that $p_c(2) < 1$.
We show that $p_c(d) > 0$ for $d \ge 2$: We consider bond percolation on $\mathbb{L}^d$ when $d \ge 2$. It suffices to show that $\theta(p) = 0$ whenever p is sufficiently close to 0. As above denote by $\sigma(n)$ the number of paths ('self-avoiding walks') of length n starting at the origin and denote by $N(n)$ the number of those paths which are open. Clearly, $\mathbb{E}_p(N(n)) = p^n\sigma(n)$. If the origin belongs to an infinite open cluster then there exist open paths of all lengths beginning at the origin, so that

$$\theta(p) \le \mathbb{P}_p(N(n) \ge 1) \le \mathbb{E}_p(N(n)) = p^n\sigma(n) \text{ for all } n.$$

We have that $\sigma(n) = (\lambda(d) + o(1))^n$ as $n \to \infty$, hence,

$$\theta(p) \le \big(p\lambda(d) + o(1)\big)^n \to 0 \text{ as } n \to \infty \text{ if } p\lambda(d) < 1.$$

Thus we have shown that $p_c(d) \ge \lambda(d)^{-1}$ where $\lambda(d) \le 2d - 1 < \infty$ and henceforth $p_c(d) > 0$.

We show that $p_c(2) < 1$: We use the famous 'Peierls argument' in honour of Rudolf Peierls and his 1936 article on the Ising model. We consider bond percolation on $\mathbb{L}^2$. We shall show that $\theta(p) > 0$ if p is sufficiently close to 1. Let $(\mathbb{Z}^2)^*$ be the dual lattice, i.e. $(\mathbb{Z}^2)^* = \mathbb{Z}^2 + (-1/2, 1/2)$, see Figure 1 where the dotted edges are the ones for the dual lattice.

There is a one-one correspondence between the edges of $\mathbb{L}^2$ and the edges of the dual, since each edge of $\mathbb{L}^2$ is crossed by a unique edge of the dual. We declare an edge of the dual to be open or closed depending respectively on whether it crosses an open or closed edge of $\mathbb{L}^2$. We thus obtain a bond percolation process on the dual with the same edge-probability. Suppose that the open cluster at the origin of $\mathbb{L}^2$ is finite, see Figure 2. We see that the origin is surrounded by a necklace of closed edges which are blocking off all possible routes from the origin to infinity. Clearly, this is satisfied when the corresponding edges of the dual contain a closed circuit in the dual having the origin of $\mathbb{L}^2$ in its interior. If the origin is in the interior of a closed circuit in the dual then the open cluster at the origin is finite

$$|C| < \infty \Leftrightarrow 0 \in \text{ interior of a closed circuit in dual.}$$

Similarly to the first part we now count the number of such closed circuits in the dual. Let $\rho(n)$ be the number of circuits of length n in the dual which contain the origin of $\mathbb{L}^2$. We get an upper bound for this number as follows. Each circuits passes through some vertex (lattice site) of the form $(k+1/2, 1/2)$ for some integer $0 \le k < n$. Furthermore, a circuit contains a self-avoiding walk

Figure Dual lattice

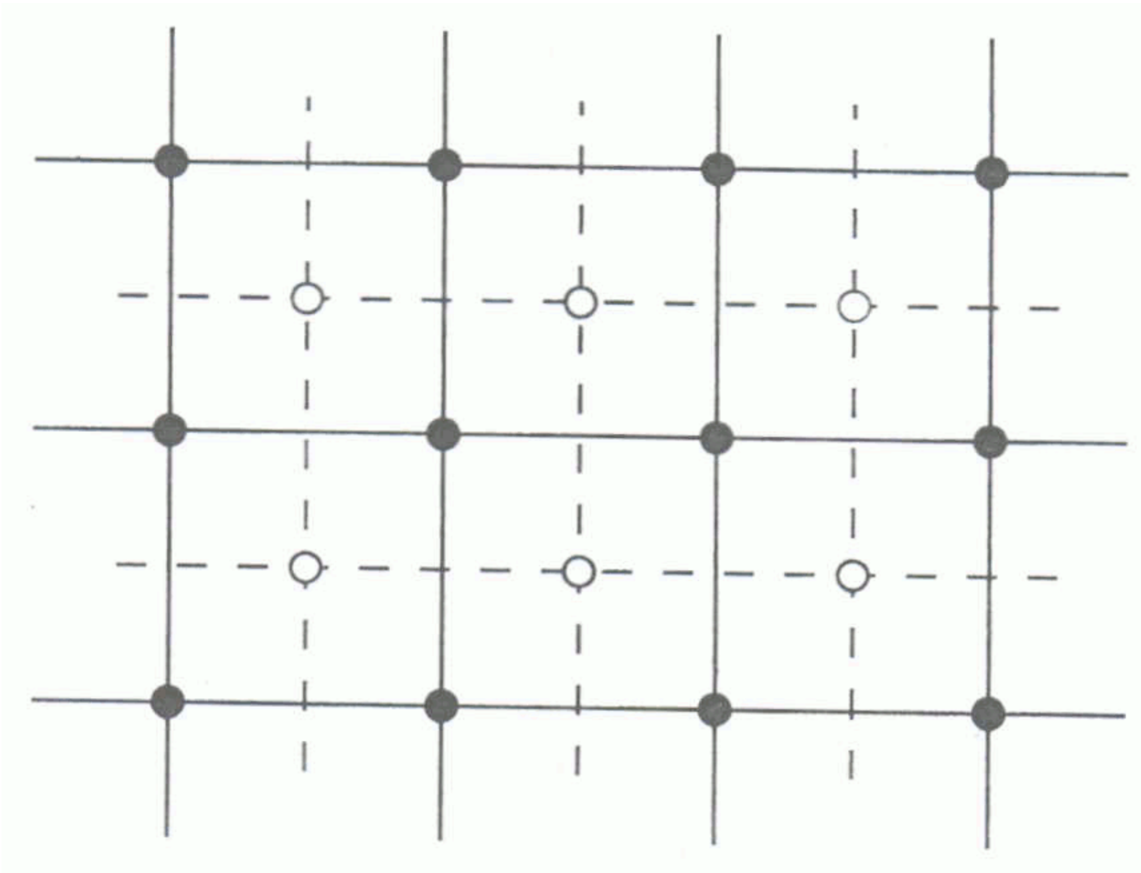

Figure 1: Dual lattice

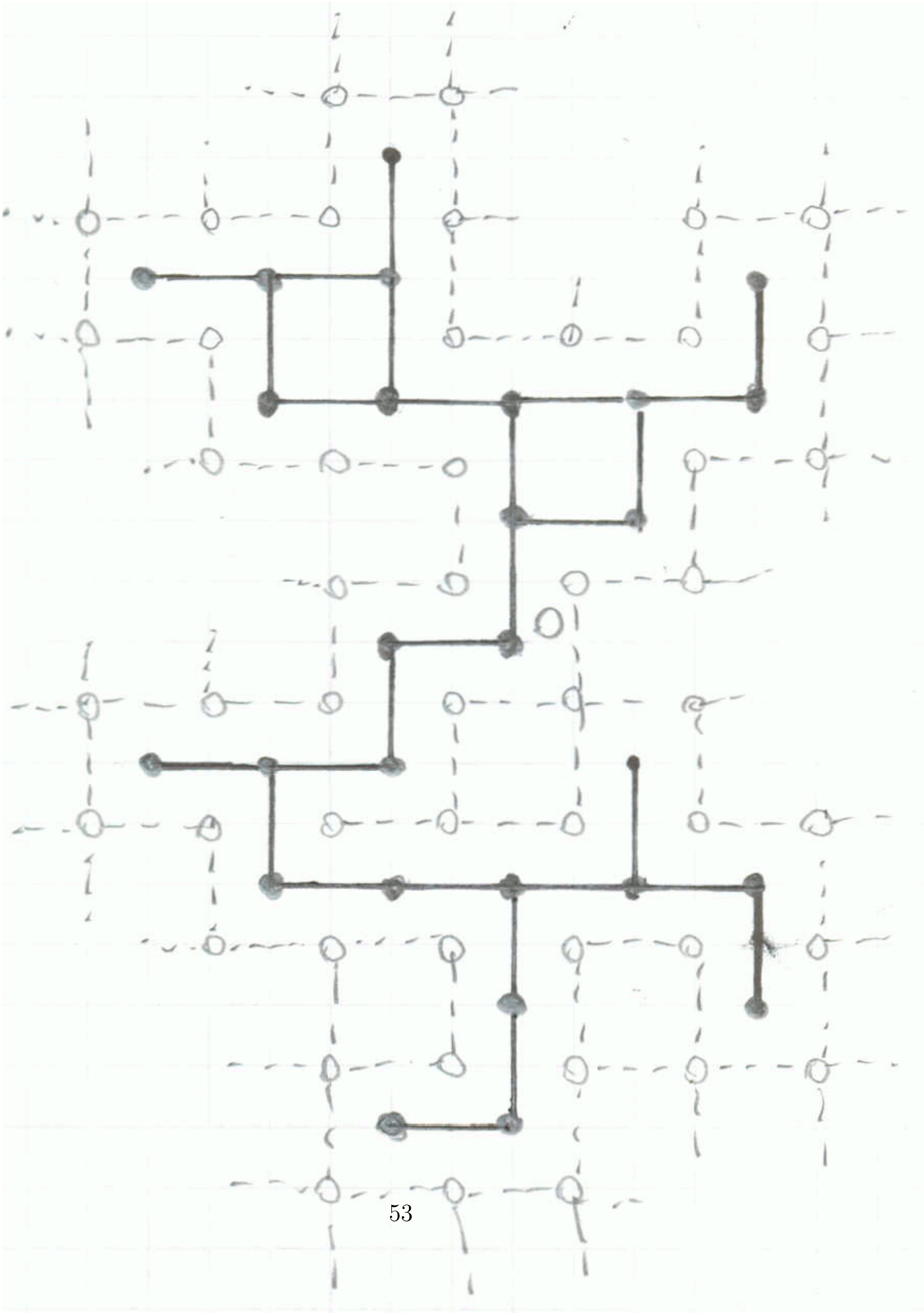

Figure 2: Closed circuit in dual lattice

of length from a vertex of the form $(k + 1/2, 1/2)$ for some integer $0 \leq k < n$. The number of such self-avoiding walks is at most $n\sigma(n-1)$. Hence, the upper bound follows

$$\rho(n) \leq n\sigma(n-1).$$

In the following denote by $\mathcal{C}_0^*$ the set of circuits in the dual containing the origin of $\mathbb{L}^2$. We estimate (we write $|\gamma|$ for the length of any path/circuit), recalling that $q = 1 - p$ is the probability of an edge to be closed,

$$\sum_{\gamma \in \mathcal{C}_0^*} \mathbb{P}_p(\gamma \text{ is closed}) = \sum_{n=1}^{\infty} \sum_{\gamma \in \mathcal{C}_0^*, |\gamma| = n} \mathbb{P}_p(\gamma \text{ is closed}) \leq \sum_{n=1}^{\infty} q^n \sigma(n-1)$$
$$\leq \sum_{n=1}^{\infty} qn\big(q\lambda(2) + o(1)\big)^{n-1} < \infty,$$

if $q\lambda(2) < 1$. Furthermore, $\sum_{\gamma \in \mathcal{C}_0^*} \mathbb{P}_p(\gamma \text{ is closed}) \to 0$ as $q = 1 - p \to 0$. Hence, there exists $\widetilde{p} \in (0, 1)$ such that

$$\sum_{\gamma \in \mathcal{C}_0^*} \mathbb{P}_p(\gamma \text{ is closed}) \leq \frac{1}{2} \quad \text{for } p > \widetilde{p}.$$

Let $M(n)$ be the number of circuits of $\mathcal{C}_0^*$ having length n. Then

$$\mathbb{P}_p(|C| = \infty) = \mathbb{P}_p(M(n) = 0 \text{ for all } n) = 1 - \mathbb{P}_p(M(n) \geq 1 \text{ for some } n)$$
$$\geq 1 - \sum_{\gamma \in \mathcal{C}_0^*} \mathbb{P}_p(\gamma \text{ is closed}) \geq \frac{1}{2}$$

if we pick $p > \widetilde{p}$. This gives $p_c(2) \leq \widetilde{p} < 1$. We need to improve the estimates to obtain that $p_c(2) \leq 1 - \lambda(2)^{-1}$. We skip these details and refer to the book by Grimmett for example. □

Proof of Theorem 4.2. The event

$$\big\{\mathbb{L}^d \text{ contains an infinite open cluster}\big\}$$

does not depend upon the states of any finite collection of edges. Hence, we know by the Zero-one law (Kolmogorov) that the probability $\Psi(p)$ can only take the values 0 or 1. If $\theta(p) = 0$ then

$$\Psi(p) \leq \sum_{x \in \mathbb{Z}^d} \mathbb{P}_p(|C(x)| = \infty) = 0.$$

If $\theta(p) > 0$ then

$$\Psi(p) \geq \mathbb{P}_p(|C) = \infty) > 0,$$

so that $\Psi(p)$ by the zero-one law. $\square$

Another 'macroscopic' quantity such as $\theta(p)$ and $\Psi(p)$ is the mean (or expected) size of the open cluster at the origin, $\chi_p = \mathbb{E}_p(|C|)$.

$$\chi_p = \infty \mathbb{P}_p(|C| = \infty) + \sum_{n=1}^{\infty} n\mathbb{P}_p(|C| = n).$$

Clearly, $\chi_p = \infty$ if $p > p_{\rm c}(d)$.

4.2 Some basic techniques

We study some basic techniques for percolation theory. We begin with events which are increasing. What we mean with an increasing event. E.g., consider the event $\{|C| = \infty\}$ that the open cluster at the origin is infinite. A configuration $\omega \in \Omega$ is element of this event if it contains an open cluster at the origin which is infinite. Therefore, $\omega \in \{|C| = \infty\}$ implies that a configuration $\omega' \in \{|C| = \infty\}$ whenever $\omega \le \omega'$ (i.e. whenever $\omega(e) \le \omega'(e)$ for all $e \in \mathbb{E}^d$). This is easy to see as the configuration ω' has more open edges as the configuration ω and as ω has an infinite open cluster at the origin more open edges in ω' do not destroy the infinite open cluster. An event $A \in \mathcal{F}$ is called **increasing** if

$$1\!\!1_A(\omega) \le 1\!\!1_A(\omega') \text{ whenever } \omega \le \omega'.$$

We call the event $A \in \mathcal{F}$ decreasing if the complement $A^{\rm c}$ is increasing. A (real valued) random variable N on $(\Omega, \mathcal{F})$ is called increasing if $N(\omega) \le N(\omega')$ whenever $\omega \le \omega'$. An example for an increasing event is $A(x, y) = \{\exists$ open path $x \leftrightarrow y\}$, and an increasing random variable is for example $N(x, y)$ the number of open paths from x to y.

Theorem 4.3 *Let N be an increasing random variable, then*

$$\mathbb{E}_{p_1}(N) \le \mathbb{E}_{p_2}(N) \text{ whenever } p_1 \le p_2$$

as long as both expectations are finite. If the event $A \in \mathcal{F}$ is increasing, then $\mathbb{P}_{p_1}(A) \le \mathbb{P}_{p_2}(A)$.

Proof. We construct a random configuration in Ω as follows. Pick a family $(X(e))_{e \in \mathbb{E}^d}$ of independent and uniformly distributed random variable on $[0, 1]$ (i.e. the random variable take values in $[0, 1]$ with uniform distribution), and denote the distribution of the whole family by $\mathbb{P}$. For a fixed parameter $p \in [0,1]$ define a random configuration $\eta_p \in \Omega$ by

$$\eta_p(e) = \begin{cases} 1 & \text{if } X(e) < p, \\ 0 & \text{if } X(e) \geq p. \end{cases}$$

Clearly this is a random configuration and an edge e is called p-open if $\eta_p(e) = 1$. Furthermore, $\mathbb{P}(\eta_p(e) = 0) = 1 - p$ and $\mathbb{P}(\eta_p(e) = 1) = p$. The random configuration η_p is the random outcome of the bond percolation process on $\mathbb{L}^d$ and clearly $\eta_{p_1} \leq \eta_{p_2}$ if $p_1 \leq p_2$. Thus $N(\eta_{p_1}) \leq N(\eta_{p_2})$ for any increasing random variable N on $(\Omega, \mathcal{F})$. Taking expectations gives the desired result (note that putting $N = 1\!\!1_A$ gives the result for events). □

Going back to Harris 1960 we call two increasing events A and B positively correlated if $\mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A)\mathbb{P}_p(B)$.

Theorem 4.4 (FKP inequality (Fortuin, Kasteleyn, Ginibre 1971))

(a) If X and Y are increasing random variables such that their second moments are finite, i.e. $\mathbb{E}_p(X^2) < \infty, \mathbb{E}_p(Y^2) < \infty$, then

$$\mathbb{E}_p(XY) \geq \mathbb{E}_p(X)\mathbb{E}_p(Y).$$

(b) If A and B are increasing events then

$$(\textit{Harris}) \qquad \mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

Proof. We give a short proof only as we later introduce a more convenient and abstract setting. Here we restrict to random variables X and Y which depend only on the states of finitely many edges, say $e_1, \ldots, e_n$. The general case can be obtained later via limit $n \to \infty$. We proceed with induction by n:
$n = 1$: X and Y depend on $\omega(e_1) \in \{0,1\}$ with probability $1-p$ and p respectively. It is then easy to see that (check briefly all possible cases)

$$\big(X(\omega_1) - X(\omega_2)\big)\big(Y(\omega_1) - Y(\omega_2)\big) \geq 0, \quad \omega_1, \omega_2 \in \{0,1\}.$$

Thus

$$\begin{aligned} &2\big(\mathbb{E}_p(XY) - \mathbb{E}_p(X)\mathbb{E}_p(Y)\big) \\ &\quad = \sum_{\omega_1,\omega_2 \in \{0,1\}} \big(X(\omega_1) - X(\omega_2)\big)\big(Y(\omega_1) - Y(\omega_2)\big)\mathbb{P}_p(\omega(e_1) = \omega_1)\mathbb{P}_p(\omega(e_1) = \omega_2) \\ &\quad \geq 0. \end{aligned}$$

Assume now that the result is valid for $n < k$. Then X and Y are increasing functions of the states $\omega(e_1), \dots, \omega(e_k)$. Then using that X are Y are increasing in $\omega(e_k)$ and that the conditional expectations are increasing functions of the states of the $k-1$ edges we get

$$
\begin{aligned}
\mathbb{E}_p(XY) &= \mathbb{E}\big(\mathbb{E}_p(XY|\omega(e_1),\dots,\omega(e_{k-1}))\big) \\
&\geq \mathbb{E}_p\Big[\mathbb{E}_p\big(X|\omega(e_1),\dots,\omega(e_{k-1})\big)\mathbb{E}_p\big(|\omega(e_1),\dots,\omega(e_{k-1}))\big)\Big] \\
&\geq \mathbb{E}_p\big(X|\omega(e_1),\dots,\omega(e_{k-1})\big)\mathbb{E}_p\big(|\omega(e_1),\dots,\omega(e_{k-1}))\big) \\
&= \mathbb{E}_p(X)\mathbb{E}_p(Y).
\end{aligned}
$$

□

We introduce now an easy abstract setting to ease proofs later and to gain more insight. For that assume we have a finite set Λ say $\Lambda = \{1, \dots, n\}$ and we consider the corresponding power set $\mathcal{P}(\Lambda)$ as the set of possible events. This power set can be identified with the cube $Q^n := \{0,1\}^n$ as follows. A subset $A \subset \Lambda$ corresponds to a sequence $(a_i)_{i=1}^n$ such that a_i when $i \in A$ and $a_i = 0$ when $i \notin A$. Further it is natural to consider Q^n as a graph with vertex set $\mathcal{P}(\Lambda)$, in which two sets A and B are joined if their symmetric difference is one, i.e.

$$A \sim B :\Leftrightarrow |A\Delta B| = 1.$$

We introduce an order via $A \leq B$ if $A \subset B$ or for $a, b \in Q^n$ we define $a \leq b$ if their is an oriented path from a to b in the graph Q^n, that is if $a_i \leq b_i$ for all i. A vertex of Q^n is naturally identified with the outcome of a sequence of n coin tosses: for $a \in Q^n$ we have $a_i = 1$ if the toss lands as heads. This puts a probability measure on the vertex set of Q^n as follows where p_i is the probability that the ith coin lands as heads

$$\mathbb{P}(A) = \sum_{a\in A}\prod_{a_i=1} p_i \prod_{a_i=0}(1-p_i), \quad A \subset Q^n. \tag{4.16}$$

Formally, letting $x_1, \dots, x_n$ be independent Bernoulli random variables with $\mathbb{P}(x_i = 1) = p_i$ and $x = (x_i)_{i=1}^n$ be the random outcome then $X = \{i \colon x_i = 1\}$ is a random subset of $\mathcal{P}(\Lambda)$. We call a set $U \subset Q^n$ **increasing (up-set)** if $a, b \in Q^n, a \in U$ and $a \leq b$ imply that $b \in U$. A set $D \subset Q^n$ is called **decreasing (down-set)** if $a, b \in Q^n, a \in D$ and $a \geq b$ imply that $b \in D$. The above setting enables us to state and prove the fundamental correlation inequality in the cube, proved by Harris 1960.

Given a set $A \subset Q^n$ and $t - 0, 1$, define

$$A_t := \{(a_i)_{i=1}^{n-1} \colon (a_1, \dots, a_{n-1}, t) \in A\} \subset Q^{n-1} \text{ for } t = 0, 1. \tag{4.17}$$

If A is increasing we have $A_0 \subset A_1$ whereas $A_1 \subset A_0$ if A is decreasing. We now put $p_i = p$ for $i = 1, \ldots, n$, and denote the measure in (4.16) as $\mathbb{P}_p$ (corresponds to our bond percolation measure). Both sets, A_0 and A_1 are subset in Q^{n-1}, however we get the probability for the corresponding $A \subset Q^n$ just with two choices for the nth entry (one shall be zero with probability $1-p$ whereas the other entry shall be one with probability 1). Hence,

$$\mathbb{P}_p(A) = (1-p)\mathbb{P}_p(A_0) + p\mathbb{P}_p(A_1) \quad \text{for all } A \subset Q^n. \tag{4.18}$$

We are in the position to state and prove Harris's Lemma.

Lemma 4.5 (Harris's Lemma) *Let $A, B \subset Q^n$ then*

(a) If both are up-sets or both are down-sets then

$$\mathbb{P}_p(A \cap B) \geq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

(b) If A is an up-set and B a down-set then

$$\mathbb{P}_p(A \cap B) \leq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

Proof. (a) We use induction by $n \geq 2$. Suppose (a) holds for $n-1$ (statement for $n = 1$ is obvious). For $A \in Q^n$ and $t = 0, 1$ write

$$A_t = \{(a_i)_{i=1}^{n-1} \colon (a_1, a_2, \ldots, a_{n-1}, t) \in A\} \in Q^{n-1}$$

and note that for an event A

$$\mathbb{P}_p(A) = (1-p)\mathbb{P}_p(A_0) + p\mathbb{P}_p(A_1), \quad \forall A \subset Q^n. \tag{4.19}$$

As both are either increasing or decreasing, that is $A_0 \subset A_1$ and $B_0 \subset B_1$ or $A_1 \subset A_0$ and $B_1 \subset B_0$ we get that

$$\Big(\mathbb{P}_p(A_0) - \mathbb{P}_p(A_1)\Big)\Big(\mathbb{P}_p(B_0) - \mathbb{P}_p(B_1)\Big) \geq 0.$$

By this, (4.18) and our assumption we get

$$\begin{aligned}
\mathbb{P}_p(A \cap B) &= (1-p)\mathbb{P}_p(A_0 \cap B_0) + p\mathbb{P}_p(A_1 \cap B_1) \\
&\geq (1-p)\mathbb{P}_p(A_0)\mathbb{P}_p(B_0) + p\mathbb{P}_p(A_1)\mathbb{P}_p(B_1) \\
&\geq \Big((1-p)\mathbb{P}_p(A_0) + p\mathbb{P}_p(A_1)\Big)\Big((1-p)\mathbb{P}_p(B_0) + p\mathbb{P}_p(B_1)\Big) \\
&= \mathbb{P}_p(A)\mathbb{P}_p(B),
\end{aligned}$$

where we that $p^2 \leq p$ and $(1-p)^2 \leq (1-p)$. (b) A increasing, B decreasing. Apply (i) to A and B^c. B^c is increasing. Hence,

$$\mathbb{P}_p(A \cap B) \leq \mathbb{P}_p(A) - \mathbb{P}_p(A)\mathbb{P}_p(B^c) = \mathbb{P}_p(A)(1-(1-\mathbb{P}_p(B)) = \mathbb{P}_p(A)\mathbb{P}_p(B).$$

$\square$

If two sets A and B are increasing it is difficult to get an upper on the probability of their intersection. But what about a different set operation which provides an upper bounds. Consider $A \subset Q^n$ be increasing, e.g. having at least one run of three heads, and $B \subset Q^n$ increasing, e.g. having at least one run of five heads. Then we define the new operation by $\square$ and define $A\square B$ as the event that there is a run of three heads and a run of at least five other heads. We can formalise this definition in our abstract setting, however, here we return directly to our percolation models.

Notation 4.6 *An event $A \in \mathcal{F}$ is said to occur on the subset $\Lambda \subset \mathbb{E}^d$ of bonds/edges in the configuration $\omega \in \Omega$ if A occurs using only bonds in Λ, independent of the values of the bonds in Λ^c. The collection of such ω is denoted by*

$$A|_\Lambda = \{\omega \colon \forall \widetilde{\omega}, \widetilde{\omega} = \omega \text{ on } \Lambda \, (\text{i.e. } \widetilde{\omega}(e) = \omega(e) \, \forall e \in \Lambda) \Rightarrow \widetilde{\omega} \in A\}.$$

The $\square$-operation (also called BK operation due to van den Berg and Kesten) is defined as

$$A\square B := \{\omega \colon \exists \Lambda_1, \Lambda_2 \subset \mathbb{E}^d, \Lambda_1 \cap \Lambda_2 = \emptyset, \omega \in A|_{\Lambda_1} \cap B|_{\Lambda_2}\}.$$

An example for $A\square B$ is indicated in the figure 3 below.

An example for $A \cap B \setminus (A\square B)$ s indicated in the figure 4 below.

With this new $\square$-operation one can show the following upper bound which goes back to van den Berg and Kesten.

Theorem 4.7 (van den Berg-Kesten inequality) *If A and B are increasing events then*

$$\mathbb{P}_p(A\square B) \leq \mathbb{P}_p(A)\mathbb{P}_p(B).$$

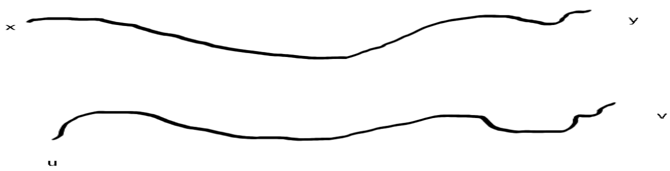

Figure 3: $A \square B$

Figure 4: $A \square B \setminus (A \square B)$

Proof. We skip a detailed proof here. An easy adaption of the proof methods for Harris's Lemma gives the proof for this inequality. □

For an increasing event we know that the probability $\mathbb{P}_p(A)$ is a non-decreasing function of the parameter p. We shall estimate therefore the rate of change of $\mathbb{P}_p(A)$ as a function of p. For that we denote an edge e to be **pivotal** for an event A in a configuration ω if precisely one of ω^+ and ω^- is in A, where $\omega^\pm$ are the configurations that agree with ω on all bonds other than e, with e open in ω^+ and closed in ω^-. We state we no proof the following result for the change of the parameter p.

Proposition 4.8 (Russo's Lemma) *Let A be an increasing event and denote by $N(A)$ the (random) number of edges which are pivotal for A. Then*

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbb{P}_p(A) = \mathbb{E}_p(N(A)).$$

Finally we shall address an important property of percolation models. That is role played by large scale behaviour . What we mean with that? Recall that we showed in the previous section that $0 < p_c(d) < 1$. We proved this inequality in two so-called non-critical regimes, also called Peierls regimes, one for small p (close to 0) and the other one for large p (close to 1). Viewed on a large scale (far away where rectangles look like single bonds) any non-critical systems acts if it is in a regime of extremely high or extremely low density of open bonds. In other words, away from the critical point, the large scale analogue of an open bond or open dual bond is very probable or very improbable. What is the large scale analogue of a bond? A possible choice is a nL by L horizontal rectangle where $L > 0$ is a scale and $n \in \mathbb{N}$. An open bond corresponds than to an open horizontal (the long way) crossing of a nL by L rectangle. We denote the probability of an open horizontal crossing of a nL by L rectangle by $R_{n,L}(p)$. i.e. the probability of the vent pictured in figure 5.

We are interested to see how the probability $R_{n,L}(p)$ changes if we change the size of the rectangle. It would be of great help later in the proof of the critical value for bond percolation if we can show that e.g. doubling of the horizontal length does not change too much the corresponding probability. This is the context of the following important lemma which we will prove in detail using all the preceding new techniques. In the proof it is useful to use solely graphical representations of the corresponding events as has been done in the lecture. However, here we cannot offer this graphical representation so you the reader is advised to make his own graphical representation by hand.

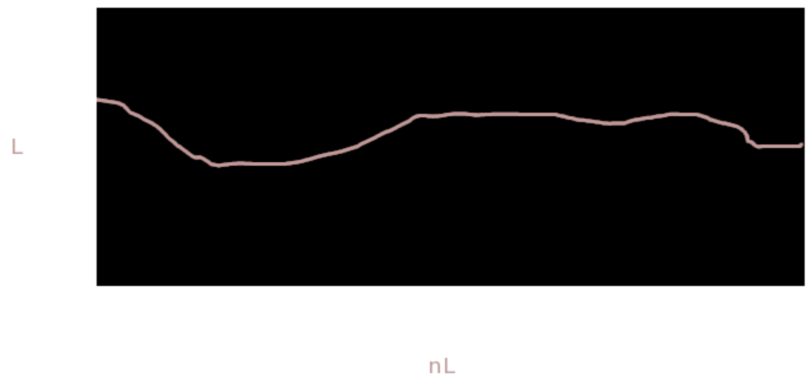

Figure 5: crossing of nL by L rectangle

Lemma 4.9 (Aizenman-Chayes-Chayes-Fröhlich-Russo) *Pick* $c = \frac{1}{16}$ *and* $\lambda \in (0,1)$. *If* $R_{2,L}(p) \geq 1 - c\lambda$ *then* $R_{2,2L}(p) \geq 1 - c\lambda^2$.

Proof. See figure 6. □

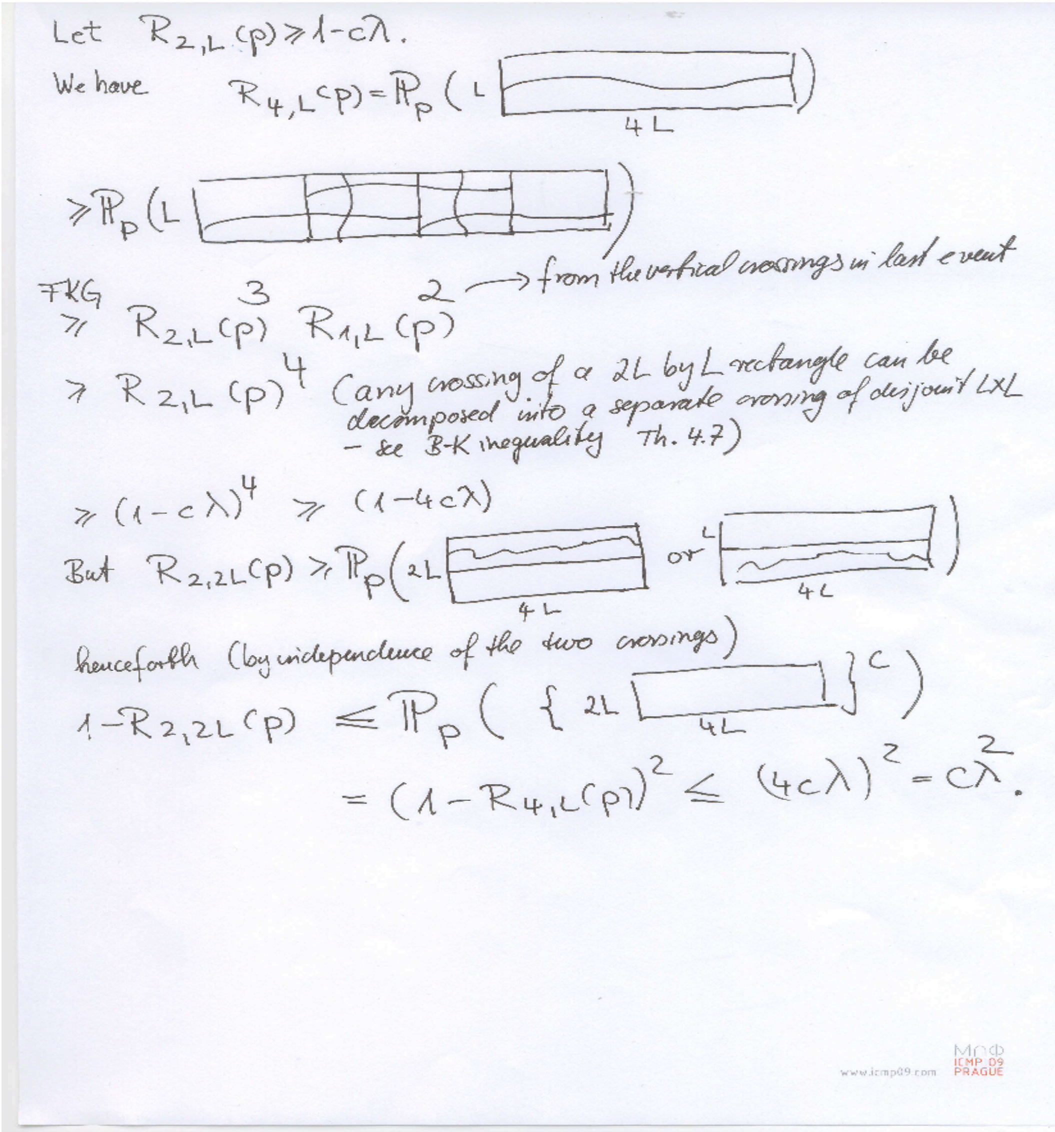

Figure 6: Proof of Lemma 4.9

4.3 Bond percolation in $\mathbb{Z}^2$