Lecture Notes MA3H2
Markov Processes and Percolation Theory

Agelos Georgakopoulos*

February 8, 2021

1 Probability theory basics

1.1 Axioms of probability theory

Let
- Ω be a set (our *sample space*),
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ a collection of subsets of Ω (the *events*), and
- $\mathbb{P} : \mathcal{F} \to [0,1]$ a function (*probability*)

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*, if

(F1) $\emptyset, \Omega \in \mathcal{F}$

(F2) If $A \in \mathcal{F}$, then $A^{\complement} := \Omega \setminus A \in \mathcal{F}$ (closed under complements)

(F3) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i \geq 1} A_i \in \mathcal{F}$ (closed under countable unions)

and

(P1) If $A_1, A_2, \ldots \in \mathcal{F}$ are pairwise disjoint, then $\mathbb{P}(\bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} \mathbb{P}(A_i)$ (countable additivity)

*Thanks to C. Panagiotis for typing bits of these notes.

(P2) $\mathbb{P}(\Omega) = 1$.

Remark. The pair $(\Omega, \mathcal{F})$ is called a *measurable space* (or measure space). $\mathcal{F}$ is called a *σ-algebra.*

All sequences $A_1, A_2, \ldots$ above are finite or countably infinite.

Exercises 1.1.

- Show that $(F1) - (F3)$ imply

  (F4) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcap_{i \geq 1} A_i \in \mathcal{F}$ (closed under countable intersections)

  Hint: use $A \cap B = \left(A^{\complement} \cup B^{\complement}\right)^{\complement}$.
  Here, $A^{\complement}$ is the complement $\Omega \backslash A$ of A.

- Show that $(P1), (P2)$ imply

  (P3) $\mathbb{P}(\emptyset) = 0$

- More generally, show that $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$.

See Durrett's book 'Elementary Probability for Applications' for examples.

1.2 Random variables

Examples:

- The height of a randomly chosen person in the room.
- The number of heads in n coin tosses.
- The temperature tomorrow at 1 pm.

Definition 1.1. A (real-valued) random variable is a *measurable function* $X : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.2. A function $X : \Omega \to \mathbb{R}$ is *measurable*, if for every interval $I \subset \mathbb{R}$ $X^{-1}(I) \in \mathcal{F}$.

Remark. If Ω is finite, and every singleton lies in $\mathcal{F}$, then every $X : \Omega \to \mathbb{R}$ is measurable. (Exercise.)

More examples:

- The indicator function $\mathbb{1}_H(\omega) := \begin{cases} 1 & \text{if } \omega = \text{heads} \\ 0 & \text{if } \omega = \text{tails} \end{cases}$ where $\Omega := \{ \text{ heads, tails } \}$ models a flip of a coin.
- More generally, for $A \subset \Omega$, $\mathbb{1}_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$
- The position of the simple random walk (on $\mathbb{Z}$) after n steps

Exercise: Consider the experiment of rolling two dice. Define a corresponding probability space, and 3 random variables on it.

1.3 Distributions

Every random variable has a (probability) distribution, telling us how likely each value is. To define it, we need to distinguish between discrete and continuous random variables.

A *continuous* random variable is a random variable that takes uncountably many possible values. If X is a continuous random variable, then its distribution can be described either by the *probability density function* f defined by the property

$$\mathbb{P}(X \in I) = \int_{\omega \in I} f(\omega) d\lambda(\omega),$$

or the *cumulative distribution function* F defined by $F(x) := \mathbb{P}(X \leq x)$.

Example: The (standard) Gaussian or Normal distribution. Its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

A *discrete* random variable is a random variable that takes only finitely or countably infinite many possible values. If X is such a random variable whose range is R, then the *probability mass function* (or distribution function) of X is the function $f : R \to [0, 1]$ such that $f(x) = \mathbb{P}(X = x)$.

Example: The indicator function $X := \mathbb{1}_A$, where A is the event of hitting the left half of the dartboard. Its probability mass function satisfies $f(0) = \frac{1}{2} = f(1)$. This is the 'same' random variable as $Y = \mathbb{1}_H$ for a coin flip, in the sense that X and Y have the same distribution. We say that X and Y are *identically distributed* (i.d.). Their distribution is called Bernoulli and is denoted by Bernoulli(1/2).

More generally, the Bernoulli distribution of parameter $p \in [0, 1]$, denoted by Bernoulli(p), is the distribution of any random variable X with $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$.

Remark: We often give names to distributions without specifying which probability space they live on!

Another example: The Geometric distribution of parameter $p \in (0, 1]$, denoted by Geometric(p), is the distribution of any random variable N with $\mathbb{P}(N = n) = p(1 - p)^{n-1}$ for every $n = 1, 2 \ldots$. The random variable N can be interpreted as the number of trials until the first 'success' in a sequence of Bernoulli(p) trials.

To summarize, the distribution of a random variable X is a complete description of the possible values of X. In some sense, Probability theory is the branch of mathematics where some numbers are replaced by probability distributions, i.e. functions on a probability space.

1.4 Expectation

Definition 1.3. For a (discrete) random variable X, define the expectation $\mathbb{E}(X) := \sum_x x\mathbb{P}(X = x)$, where the sum ranges over all possible values x that X can take.

Example: Suppose $X =$ Bernoulli(p). Then

$$\mathbb{E}(X) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

Theorem 1.1. Let $X =$ Geometric(p). Then $\mathbb{E}(X) = 1/p$.

Proof 1. Set $M := \mathbb{E}(X)$. Note that $M < +\infty$, as the exponential function decays faster than any polynomial. Moreover, M satisfies

$$M = p \cdot 1 + (1-p)(M+1). \tag{1}$$

To see this, let us consider what happens in the first trial. If the first trial is a 'success', which happens with probability p, then the process stops, and X is equal to 1. If the first trial is a 'failure', which happens with probability $1-p$, then the process starts all over again, and $X = Y + 1$, where $Y = \text{Geometric}(p)$. Solving (1) for M gives $M = 1/p$. □

Proof 2. We first express $\mathbb{E}(X)$ via the sum $\sum_{k=1}^{+\infty} kp(1-p)^{k-1}$. In order to compute the latter sum, we will use that

$$\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$$

for every $x \in (0,1)$. Differentiating both sides we obtain

$$\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$$

Setting $x = 1-p$ and multiplying by p gives

$$\sum_{k=1}^{+\infty} kp(1-p)^{k-1} = 1/p,$$

as desired. □

Exercise: Compute the *second moment* $\mathbb{E}(X^2)$ of the random variable $X = \text{Geometric}(p)$.

Exercise: Define a random variable Y with $\mathbb{E}(Y) = \infty$.

Exercise: Define a random variable Z with $\mathbb{E}(Z) < \infty$ and $\mathbb{E}(Z^2) = \infty$.

1.4.1 Linearity of expectation

We have $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ for any two random variables X, Y. This immediately extends to finite sums of random variables. It also extends to infinite sums under certain conditions: If $(X_i)_{i\in\mathbb{N}}$ is a sequence of random variables, and $\sum \mathbb{E}(|X_i|) < \infty$ or $\mathbb{E}(\sum |X_i|) < \infty$, then

$$\sum \mathbb{E}(X_i) = \mathbb{E}(\sum X_i).$$

This is implied by the Fubini-Tonelli theorem.

1.5 Independence

Suppose X_1, X_2, X_3 are Bernoulli(1/2) random variables on a common probability space. What is $\mathbb{P}(X_1 = X_2 = X_3 = 1)$? It depends!

Example: $\Omega_3 =$ 'flip 3 coins' (I) $X_i = \mathbb{1}_{\{\text{coin i heads}\}}$

(II) $X_1 = X_2 = X_3 = \mathbb{1}_{\{\text{coin 1 heads}\}}$

In case (I), the X_i are *independent*, while in case (II) they are *dependent*.

Definition 1.4. The events A_i, $1 \leq i \leq n$ are *mutually independent*, if for every $I \subset \{1, 2, \ldots, n\}$

$$\mathbb{P}(\bigcap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

Remark. This is stronger than pairwise independence!

Similarly, the random variables X_i, $1 \leq i \leq n$ are *mutually independent*, if the events $\{X_i \leq x_i\}_{1 \leq i \leq n}$ are mutually independent for every family of real numbers $\{x_i\}_{1 \leq i \leq n}$.

Examples: In example Ω_3 above, the random variables $X_i = \mathbb{1}_{\{\text{coin i heads}\}}$ are mutually independent. However, if we change X_3 to $\mathbb{1}_{\{X_1 + X_2 \text{ is even}\}}$, then the X_i are pairwise independent but not mutually independent, because

$$\mathbb{P}(X_1 = X_2 = X_3 = 0) = 0 \neq \prod_{i=1}^{3} \mathbb{P}(X_i = 0).$$

Intuitively, the standard way to produce independent random variables is by using different experiments.

1.6 Conditioning

For any two events A, B, we define the conditional probability of A subject to B by $\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Thus if A, B are independent we have $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

1.7 Coupling

Let X_1 and X_2 be two random variables defined on probability spaces $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, respectively. Then a coupling of X_1 and X_2 is a new probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a pair of random variables Y_1 and Y_2 in $(\Omega, \mathcal{F}, \mathbb{P})$ such that Y_1 has the same distribution as X_1 and Y_2 has the same distribution as X_2.

Such a coupling is particularly useful when we can define Y_1 and Y_2 so that e.g. $Y_1(\omega) \geq Y_2(\omega)$ holds for every $\omega \in \Omega$. In this case we can deduce e.g. that $\mathbb{E}(X_1) \geq \mathbb{E}(X_2)$.

Exercise 1.1. Let X_1 and X_2 be two random variables with distributions $Bernoulli(1/4)$ and $Bernoulli(1/2)$, respectively. Construct a coupling Y_1, Y_2 such that $Y_1(\omega) \leq Y_2(\omega)$ for every $\omega \in \Omega$.

Exercises

Exercise 1.2. Prove $\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \cap C \mid B \cap C)$ holds for all events A, B, C in a probability space.

Exercise 1.3. Let $(Y_i)_{i \in \mathbb{N}}$ be a sequence of mutually independent random variables with distributions $Bernoulli(p)$ for $p > 0$. Then
$\mathbb{P}(\{Y_i = 0 \forall i \in \mathbb{N}\}) = 0$.

2 Markov chains

2.1 Definition of a Markov chain

A Markov chain $X_0, X_1, \ldots$ is characterised by the property that the final position depends on the last preceding one, but not on the previous history:

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

Notice that we are actually dealing with generalized S-valued random variables by considering maps $X_n : \Omega \to S$ such that $X_n^{-1}(x) = \{\omega \in \Omega : X_n(\omega) = x\} \in \mathcal{F}$ for every $x \in S$. Also, we assume that the conditional probability

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \frac{\mathbb{P}(X_0 = x_0, \ldots X_{n+1} = x_{n+1})}{\mathbb{P}(X_0 = x_0, \ldots X_n = x_n)}$$

is well defined by assuming that $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) > 0$.

In general, a sequence of random variables $X_1, X_2, \ldots$ might have very complicated mutual dependencies. We were often discussing the case of a sequence of independent random variables (like independent coin tosses). In Markov chains we have an intermediate situation: there is some dependence, but of only restricted type–the random variable X_{n+1} depends on the preceding ones only through its last predecessor X_n.

Definition 2.1. Let S be an at most countable set and let $\Pi : S \times S \to [0, 1]$ be a mapping such that $\sum_{y \in S} \Pi(x, y) = 1$ for each $x \in S$. We say that the matrix $\Pi = (\Pi(x, y))_{x,y \in S}$ is a *stochastic matrix*.

A sequence $X_0, X_1, \ldots$ of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in S, $X_n : \Omega \to S$, $n = 0, 1, \ldots$, is called a *Markov chain* with state space S and *transition matrix* Π, if

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \Pi(x_n, x_{n+1})$$

for every $n \geq 0$ and any sequence $x_0, x_1, \ldots, x_{n+1} \in S$ such that $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) > 0$.

The distribution $\alpha = \mathbb{P} \circ X_0^{-1}$ of the values of X_0 on the state space S is the *initial distribution* of the Markov chain.

Remark. Using $\alpha = \mathbb{P} \circ X_0^{-1}$ as a shorthand for the function defined on S by $S \ni x \to \alpha(x) = P(\{\omega \in \Omega : X_0(\omega) = x\})$, we observe that it is a probability distribution since $\sum_{x \in S} \alpha(x) = \sum_{x \in S} \mathbb{P}(X_0 = x) = 1$.

Notice that definition 2.1 implies 2 statements:

- Markov property: $\mathbb{P}(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$
- Time homogeneity: $\mathbb{P}(X_{n+1} = y \mid X_n = x)$ does not depend on n.

The fact that the matrix Π is stochastic, its elements are nonnegative and each row sums to one, assures a consistence. It means that with a fixed x, the matrix element $\Pi(x, y)$ may indeed play a role of the probability of the move from x to y, $\mathbb{P}(X_{n+1} = y \mid X_n = x) = \Pi(x, y)$, with the probability of moving to anywhere in S equal to one, $\sum_{y \in S} \mathbb{P}(X_{n+1} = y \mid X_n = x) = 1$, as it should.

Notation: It is useful to introduce a special notation

$$\mathbb{P}_x(X_n = x_n, \ldots, X_1 = x_1) = \mathbb{P}(X_n = x_n, \ldots, X_1 = x_1 \mid X_0 = x)$$

and

$$\mathbb{P}_\alpha(X_n = x_n, \ldots, X_1 = x_1) = \sum_{x \in S} \mathbb{P}_x(X_n = x_n, \ldots, X_1 = x_1)\alpha(x)$$

for the probability with the initial condition x or the initial distribution α, respectively.

Lemma 2.1. We have $\mathbb{P}_x(X_n = y) = \Pi^n(x, y)$, where $\Pi^n(x, y)$ is the matrix element of the matrix Π^n, the n-th power of Π, at the intersection of the row and the column labelled by x and y, respectively. Also, for any $k < n$ we have $\mathbb{P}_x(X_n = y \mid X_k = z) = \mathbb{P}_z(X_{n-k} = y)$.

Proof. We will repeatedly rely on the equality

$$\begin{aligned}
&\mathbb{P}(X_n = y, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x) = \\
&\mathbb{P}(X_n = y \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x) \cdots \mathbb{P}(X_1 = x_1 \mid X_0 = x)\mathbb{P}(X_0 = x) = \\
&\mathbb{P}(X_n = y \mid X_{n-1} = x_{n-1}) \cdots \mathbb{P}(X_1 = x_1 \mid X_0 = x)\mathbb{P}(X_0 = x) = \\
&\mathbb{P}(X_0 = x)\Pi(x, x_1)\Pi(x_1, x_2) \cdots \Pi(x_{n-1}, x_n)
\end{aligned}$$

To verify the first equality, it suffices to use sequentially, starting at the end of the second line, the equality

$$\begin{aligned}
&\mathbb{P}(X_k = x_k \mid X_{k-1} = x_{k-1}, \ldots, X_0 = x)\mathbb{P}(X_{k-1} = x_{k-1}, \ldots, X_0 = x) = \\
&\mathbb{P}(X_k = x_k, X_{k-1} = x_{k-1}, \ldots, X_0 = x)
\end{aligned}$$

for $k = 1, 2 \ldots, n$, which holds by the definition of conditional probability.

For the first claim, we thus get

$$\begin{aligned} &\mathbb{P}(X_n = y \mid X_0 = x) = \\ &\frac{\sum_{x_1,\ldots,x_{n-1}\in S} \mathbb{P}(X_n = y, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1, X_0 = x)}{\mathbb{P}(X_0 = x)} = \\ &\sum_{x_1,\ldots,x_{n-1}\in S} \Pi(x, x_1)\Pi(x_1, x_2)\cdots\Pi(x_{n-1}, y) = \Pi^n(x, y). \end{aligned}$$

Similarly, for the second claim,

$$\begin{aligned} &\mathbb{P}(X_n = y, X_k = z, X_0 = x) = \\ &\sum_{x_{n-1},\ldots,x_{k+1},x_{k-1},\ldots,x_1\in S} \mathbb{P}(X_n = y, X_{n-1} = x_{n-1}, \ldots, X_k = z, \ldots, X_0 = x) = \\ &\sum_{x_{n-1},\ldots,x_{k+1},x_{k-1},\ldots,x_1\in S} \mathbb{P}(X_0 = x)\Pi(x, x_1)\cdots\Pi(x_{k-1}, z)\Pi(z, x_{k+1})\cdots\Pi(x_{n-1}, y) \end{aligned}$$

yielding

$$\mathbb{P}(X_n = y, X_k = z \mid X_0 = x) = \Pi^k(x, z)\Pi^{n-k}(z, y).$$

Hence,

$$\begin{aligned} \mathbb{P}_x(X_n = y \mid X_k = z) &= \frac{\mathbb{P}_x(X_n = y, X_k = z)}{\mathbb{P}_x(X_k = z)} = \frac{\mathbb{P}(X_n = y, X_k = z \mid X_0 = x)}{\mathbb{P}(X_k = z \mid X_0 = x)} = \\ &= \frac{\Pi^k(x, z)\Pi^{n-k}(z, y)}{\Pi^k(x, z)} = \Pi^{n-k}(z, y) = \mathbb{P}_z(X_{n-k} = y). \end{aligned}$$

□

In the definition above we are assuming that a suitable probability space and a sequence of random variables X_n are given so that the defining equations linking particular conditional probabilities with the transition matrix are satisfied.

We could ask the opposite: given an initial distribution α and a transition matrix Π, do there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence of random variables X_n on it so that the defining equations are satisfied? And is this probability unique?

The answer on both is positive. Without going into technical details, let us sketch the main steps of construction of the probability measure $\mathbb{P}_\alpha$:

1) Let $\Omega := S^{\mathbb{N}}$, and let $\mathcal{F}$ be the σ-algebra generated by cylindrical sets $C(n, x_0, \ldots, x_n) := \{\omega \in \Omega : \omega(0) = x_0, \ldots, \omega(n) = x_n\}$.
2) For any cylindrical set $C(n, x_0, \ldots, x_n)$, set

$$\mathbb{P}_\alpha(C(n, x_0, \ldots, x_n)) = \alpha(x_0)\Pi(x_0, x_1), \ldots, \Pi(x_{n-1}, x_n).$$

If we were happy to consider only times n up to some large N we could restrict ourselves to the sample space S^N and stop here. But often we need to consider times that are not a priori restricted; for example, the time of first return to the origin for a random walk. As a result we have to stick with $\Omega := S^{\mathbb{N}}$. Hence the need of the next step.
3) Extend the measure $\mathbb{P}_\alpha$ to $\mathcal{F}$ with the help of the Kolmogorov consistence theorem to get a σ-additive measure on the set $\mathcal{G}$ of all cylinder sets in Ω combined with the subsequent Caratheodory construction extending the measure to the full σ-algebra $\mathcal{F}$. In our case, when Ω is given as a product of countable sets, both steps are relatively simple but will be omitted here. In any case, however, the extension, once constructed, is uniquely determined by the step 2) in view of Theorem 2.2 and the fact that the set $\mathcal{G}$ of cylindrical sets is closed under intersections.

Theorem 2.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose that $\mathcal{F}$ is generated by a set $\mathcal{G} \subset \mathcal{F}$ ($\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{G}$) and that the set $\mathcal{G}$ is closed under intersections. Then the probability measure $\mathbb{P}$ is uniquely determined by its restriction $\mathbb{P}|_{\mathcal{G}}$ to $\mathcal{G}$.

4) The random variables X_n are then simply projections $X_n : \Omega \to S$ defined by $X_n(\omega) = \omega_n$. The claim that the probability $\mathbb{P}_\alpha$ with the sequence of random variables X_n now constitutes the Markov chain with transition matrix Π can be read from the equation in step 2) combined with the definition in step 4) and with the first equation in the proof of Lemma 2.1. Notice that only the measure restricted to cylindrical sets is actually needed here.

As we saw above, applying Π n-times amounts to the use of the power Π^n which corresponds to moving by n steps forward in time. One can trace a particular path $x, x_1, x_2, \ldots, x_n$ down to a contribution to the matrix product of the form $\Pi(x, x_1)\Pi(x_1, x_2) \ldots \Pi(x_{n-1}, x_n)$. Matrix product is obtained as the sum over all paths.

Exercise 2.1. Prove that the power Π^n of a stochastic matrix Π is also a stochastic matrix.

We now explore the effect of **left and right multiplication** by the matrix Π. If the probability distribution of the random variable X_n is μ, i.e. $\mathbb{P}(X_n = x) = \mu(x)$, then the probability distribution of X_{n+1} is $\mu\Pi$ where μ is viewed as a row vector (with coordinates labeled by elements from S), and $\mu\Pi$ denotes the matrix product of μ and Π. Indeed, this follows from a straightforward computation employing the definition of a Markov chain:

$$\mathbb{P}(X_{n+1} = y) = \sum_{x\in S} \mathbb{P}(X_{n+1} = y|X_n = x)\mathbb{P}(X_n = x) = \sum_{x\in S} \Pi(x, y)\mu(x) = (\mu\Pi)(y).$$

This fact can be summarised as follows: **Multiplying by Π on the right moves today's distribution into tomorrow's distribution.**

We now turn to left multiplication by Π. Let f be a function on the state space S, considered as a column vector. Multiplying on the left by Π we have:

$$(\Pi f)(x) = \sum_{y\in S} \Pi(x, y)f(y) = \sum_{y\in S} \mathbb{P}(X_{n+1} = y|X_n = x)f(y) = \mathbb{E}(f(X_{n+1})|X_n = x).$$

Here, we are using the conditional expectation $\mathbb{E}(Y|X = x)$ of a random variable $Y : \Omega \to \mathbb{R}$ defined as

$$\mathbb{E}(Y|X = x) = \sum_{a\in Y(\Omega)} a\mathbb{P}(Y = a|X = x)$$

(assuming that $\mathbb{P}(X = x) > 0$). We also rely on the fact that the expectation (or conditional expectation) of a function $f(Y)$ of a random variable Y is

$$\mathbb{E}(f(Y)) = \sum_{b\in f(Y)(\Omega)} b\mathbb{P}(f(Y) = b) = \sum_{a\in Y(\Omega)} f(a)\mathbb{P}(Y = a),$$

using that $\mathbb{P}(f(Y) = b) = \sum_{a:f(a)=b} \mathbb{P}(Y = a)$. This fact can be summarised as follows: **Multiplying a column vector f by Π on the left takes us from a function on the state space today to the expected value of that function tomorrow in dependence on the position today.**

Examples

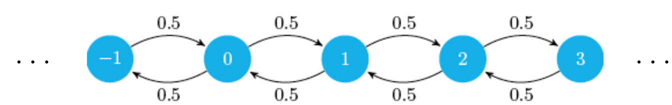
1. *Simple Random Walk on $\mathbb{Z}$.*

Walker starts at 0 and at every integer time flips a coin and moves one step right or left. (Think about a walker on a long road they stops every hundred meters and tossing a coin decides whether to proceed in the same direction for another hundred meters or to walk back for hundred meters.) Notice that here the state space $S = \mathbb{Z}$ is not finite!

This corresponds to the Markov chain with transition probabilities

$$\mathbb{P}(X_{n+1} = x \pm 1 \,|\, X_n = x) = \tfrac{1}{2}.$$

The corresponding weighted digraph is:



and the transition matrix:

$$\Pi = \begin{pmatrix} \cdots\cdots\cdots\cdots\cdots\cdots \\ \ldots, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0 \ldots \\ \ldots, 0, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, \ldots \\ \ldots, 0, 0, \frac{1}{2}, 0, \frac{1}{2}, 0, \ldots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{pmatrix} \longleftarrow \; x$$

$$\uparrow$$
$$y$$

An alternative description: if $Z_k \in \{-1, 1\}$ is the outcome of the k-th coin flip and $X_0 = 0$, then $X_n = Z_1 + Z_2 + \cdots + Z_n$.

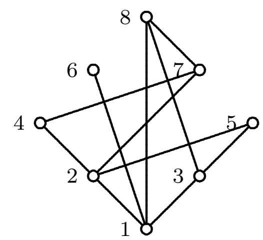
3. *Random Walk on a Graph.*

The random walk on a graph $G = (V, E)$ is the Markov chain on the state space V with the transition matrix

$$\Pi(x,y) = \begin{cases} \frac{1}{d_G(x)} \text{ if } \{x,y\} \in E, \\ 0 \text{ otherwise.} \end{cases}$$

I.e., when at x, the walker examines the neighbourhood and moves to an adjacent vertex chosen uniformly from all neighbours.

For example, the random walk on the graph



is a Markov chain with the transition matrix

$$\Pi = \begin{pmatrix} 0, \frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{4}, 0, \frac{1}{4} \\ \frac{1}{4}, 0, 0, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}, 0 \\ \frac{1}{3}, 0, 0, 0, \frac{1}{3}, 0, 0, \frac{1}{3} \\ 0, \frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}, 0 \\ 0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0 \\ 1, 0, 0, 0, 0, 0, 0, 0 \\ 0, \frac{1}{3}, 0, \frac{1}{3}, 0, 0, 0, \frac{1}{3} \\ \frac{1}{3}, 0, \frac{1}{3}, 0, 0, 0, \frac{1}{3}, 0 \end{pmatrix}.$$

2.2 Irreducibility

Definition 2.2. A transition matrix Π is **irreducible**, if for each $x, y \in S$ there exists $n = n(x,y)$ such that $\Pi^n(x,y) > 0$.

Notice that according to the definition, for any $x, y \in S$ there exists also $n(y,x)$ (not necessarily equal to $n(x,y)$) such that $\Pi^n(y,x) > 0$.

Recalling that $\Pi^n(x,y) = \mathbb{P}_x(X_n = y)$, the irreducibility of a transition matrix means that, for the corresponding Markov chain, the probability of reaching y after $n = n(x,y)$ steps if starting at x is non-vanishing.

Usually we assume irreducibility. However, this is not really a loss of generality; if Π is not irreducible and S is finite, it can be split into irreducible pieces. To show this formally, we first introduce some notation.

Definition 2.3. We say that $y \in S$ is accessible from $x \in S$, in symbols $x \to y$, if there exists $n = n(x, y) > 0$ such that $\mathbb{P}_x(X_n = y) > 0$.

We say that x communicates with y, in symbols $x \leftrightarrow y$, if either $x \to y$and $x \leftarrow y$ or $x = y$.

A communicating class is an equivalence class with respect to the equivalence relation $\leftrightarrow$.

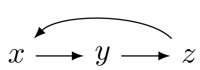
We say that x is essential if $x \to y$ implies $y \to x$ for every $y \in S$, and x is inessential if it is not essential.

Exercise 2.2. Check that $\leftrightarrow$ is an equivalence relation.

Lemma 2.3.

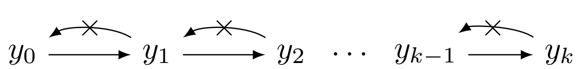
1. If Π is irreducible, then S consists of a single communicating class.
2. If x is essential and $x \to y$, then y is essential.
3. If S is finite then there exists an essential communicating class.
4. If Π is not irreducible, then it is irreducible on each essential class.

Proof. (i) If Π is irreducible, then $x \leftrightarrow y$ for all $x, y \in S$.
(ii) Let x be essential and $x \to y$. The main idea is now summarised in the following illustration.



Namely, if $y \to z$ for some z, then also $x \to z$ and thus $z \to x$ as x is essential. Since $x \to y$, we have $z \to y$ implying that y is essential.

(iii) Define inductively a sequence $(y_0, y_1, \ldots)$ of distinct states so that at each step $y_k \to y_{k+1}$ but $y_{k+1} \not\to y_k$ as illustrated here:

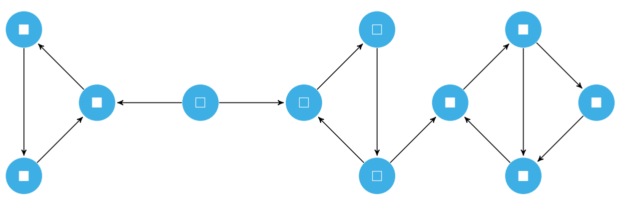


If for any y_k such y_{k+1} does not exist, then y_k is essential. Indeed, there does not exist $j < k$ such that $y_k \to y_j$ since this would mean that $y_k \to y_{k-1}$ (via the sequence $y_k \to y_j \to y_{j+1} \to \ldots y_{k-1}$) in contradiction to the construction. On the other hand, for any $y \in S \setminus \{y_0, \ldots, y_k\}$ such that $y_k \to y$

we also have $y \to y_k$ (otherwise we could have proceeded with the construction beyond y_k). If $S = y_0, \dots, y_k$, then no $y \in S \setminus \{y_k\}$ is accessible from y_k and thus y_k is essential. Given that S is finite and no repetitions are possible, the sequence necessarily terminates. We get at least one essential equivalence class $[y_k]$.

(iv) We just have to verify that Π restricted to an essential class $[x]$ is stochastic. Indeed, for all $y \notin [x]$ we have $\Pi(z, y) = 0$ for any $z \in [x]$ since $\Pi(z, y) > 0$ would imply that $z \to y$ and thus also $y \to z$ (by the essentiality of z) in contradiction with the assumption $y \notin [x]$. $\square$

Example: for the Markov chain represented by the digraph below, there are four classes, two essential and two inessential.



2.3 Periodicity

Definition 2.4. Let $\mathcal{T}(x) := \{n \geq 1 : \Pi^n(x, x) > 0\}$ be the set of possible times of returns to x.
A *period* of the state x is the greatest common divisor of $\mathcal{T}(x)$.
A Markov chain is *aperiodic*, if all its states have period 1.

For example, a simple random walk on $\mathbb{Z}^d, d \geq 1$, has period 2 since $\mathcal{T}(x) = 2k \mid k \in N$. With help of the properties of the greatest common divisor we get the following.

Lemma 2.4. If Π is irreducible, then $\gcd \mathcal{T}(x) = \gcd \mathcal{T}(y)$ for all $x, y \in S$.

Proof. Fix $x, y \in S$. There exist $m, n \in \mathbb{N}$ such that $\Pi^m(x, y) > 0$ and $\Pi^n(y, x) > 0$. Take $l := m + n$. We have $l \in \mathcal{T}(x) \cap \mathcal{T}(y)$. Indeed, $\Pi^{m+n}(x, x) = \sum_{z \in S} \Pi^m(x, z)\Pi^n(z, x) \geq \Pi^m(x, y)\Pi^n(y, x) > 0$ and thus $m + n \in \mathcal{T}(x)$. Similarly, $n + m \in \mathcal{T}(y)$. We claim that $\mathcal{T}(x) + l \subset \mathcal{T}(y)$. To see this, notice that $k \in \mathcal{T}(x)$ implies $\Pi^{n+k+m}(y, y) > 0$ and thus $k + l \in \mathcal{T}(y)$.

Hence, $\gcd \mathcal{T}(y)$ divides all elements of $\mathcal{T}(x)$ and thus $\gcd \mathcal{T}(y) \leq \gcd \mathcal{T}(x)$. Similarly, exchanging the role of x and y, we obtain $\gcd \mathcal{T}(x) \leq \gcd \mathcal{T}(y)$. □

Lemma 2.5. If Π is aperiodic and irreducible with finite state space S, then there exists r such that $\Pi^n(x,y) > 0$ for all $x, y \in S$ and any $n \in \mathbb{N}, n \geq r$.

Proof. This is based on the following proposition from number theory that we state without proof:

Proposition 2.6. Any set of non-negative integers closed under addition and with gcd equal to one contains all but finite number of integers.

The set $\mathcal{T}(x)$ is closed under addition: $s, t \in \mathcal{T}(x)$ implies $\Pi^{s+t}(x,x) \geq \Pi^s(x,x)\Pi^t(x,x) > 0$. Thus $\mathcal{T}(x)$ equals $\mathbb{N}$ up to a finite set: there exists $n(x)$ such that $n \in \mathcal{T}(x)$ for all $n \geq n(x)$. For $n \geq n(x) + n(x,y)$ (taking into account the irreducibility of Π), we have
$\Pi^n(x,y) \geq \Pi^{n-n(x,y)}(x,x)\Pi^n(x,y)(x,y) > 0$ and thus, for any $n \geq m(x) = n(x) + \max_{y \in S} n(x,y)$ we have $\Pi^n(x,y) > 0$ for all $y \in S$. Finally, for any $r \geq max_{x \in S} m(x)$ we have $\Pi^r(x,y) > 0$ for all $x, y \in S$. □

Finally, let us observe that any periodic chain may turn into aperiodic by lazyness: each time do nothing with a small probability or, with the complementary probability, proceed by applying Π. This amounts to replacing Π by a 'lazy' Markov chain $(1-\epsilon)\Pi + \epsilon\mathbb{1}$, for which

$$[(1-\epsilon)\Pi + \epsilon\mathbb{1}](x,x) \geq \epsilon > 0.$$

2.4 Asymptotic stationarity

Definition 2.5. We say that a probability distribution
$\alpha : S \to [0,1], \sum_{x \in S} \alpha(x) = 1$, is *stationary*, if

$$\alpha\Pi = \alpha$$

that is, $\sum_{x \in S} \alpha(x)\Pi(x,y) = \alpha(y)$ for every $y \in S$.

We will prove:
If S is finite, Π is aperiodic and irreducible, and α is stationary, then $\mathbb{P}_x(X_n = y)$ converges to $\alpha(y)$ as $n \to \infty$ for any $x, y \in S$.

First, we will show the existence and uniqueness of a stationary distribution:

Theorem 2.7. Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain with finite state space S and the transition matrix Π. Then:
i) There exists a unique probability distribution α on S such that $\alpha\Pi = \alpha$, and moreover,
ii) $\alpha(x) > 0$ for all $x \in S$, and $\alpha(x) = \mathbb{E}_x(\tau_x)$.

Let us begin by proving uniqueness assuming existence:

Lemma 2.8. For an irreducible Markov chain with finite S, there exists at most one probability distribution α such that $\alpha\Pi = \alpha$.

Proof. If α satisfies the equation $\alpha\Pi = \alpha$, then any product $\lambda\alpha, \lambda \in \mathbb{R}$, also does. Existence and uniqueness of stationary probability is thus equivalent to the claim that the row rank of the matrix $\Pi - \mathbb{1}$ is $|S| - 1$ (one-dimensional space of solutions of the equation $\alpha(\Pi - \mathbb{1}) = 0$). As the row rank equals the column rank, this is equivalent to showing that the column rank of $\Pi - \mathbb{1}$ is $|S| - 1$, i.e. the equation $(\Pi - \mathbb{1})h = 0$ has a 1-dimensional space of solutions.

Since we are trying to show that there exists at most one stationary distribution, we want to show that the space of solutions of $\Pi h = h$ is at most one-dimensional. Notice that the vector $(h(x) = 1, x \in S)$ is a solution, as $\sum_y \Pi(x, y)1 = 1$ for every $x \in S$. It thus remains to show that $\Pi h = h$ implies that h is constant on S. To prove this, let us denote $M = h(x_0) = \max_{x\in S} h(x)$. If $\Pi(x_0, z) > 0$, then $h(z) = M$, because if it were the case that $h(z) < M$, then $M = h(x_0) = \Pi(x_0, z)h(z) + \sum_{y\neq z} \Pi(x_0, y)h(y) < M$, a contradiction. Since every $y \in S$ can be reached from x_0 by a sequence $x_0, x_1, \ldots, x_n, y$ with $\Pi(x_0, x_1) > 0, \Pi(x_1, x_2) > 0, \ldots, \Pi(x_k, y) > 0$, we have $M = h(x_0) = h(x_1) = \ldots = h(y)$. It is only in the last claim that we used irreducibility. □

In the preparation for the proof of the claim ii), we first show that it makes sense:

Lemma 2.9. If Π is irreducible with finite S, then $\mathbb{E}_x(\tau_y) < \infty$ for any $x, y \in S$.

Proof. Irreducibility implies that there exists $\epsilon > 0$ and $r > 0$ such that for any $z, y \in S$ there exist $n \leq r$ such that $\Pi^n(z, y) > \epsilon$. Indeed, for each z, y there exists $n(z, y)$ such that $\Pi^{n(z,y)}(z, y) > 0$. Observing that the sets $\{n(z, y) : z, y \in S\}$ and $\{\Pi^{n(z,y)}(z, y) : z, y \in S\}$ are finite, we just choose r

as the maximum of the former and ϵ smaller than the minimum of the latter. As a consequence, for any value of X_m, the probability of hitting y for some $X_n, n \in [m, m+r]$, is at least ϵ.

We claim that $\mathbb{P}_x(X_{m+1}, \ldots, X_{m+r} \neq y \mid X_m \neq y) \leq 1 - \epsilon$.

Indeed, the LHS equals

$$\frac{\mathbb{P}_x(X_m, \ldots, X_{m+r} \neq y)}{\mathbb{P}_x(X_m \neq y)} = \text{(considering all possible positions at step } m\text{)}$$

$$\frac{\sum_{z \neq y} \mathbb{P}_x(X_{m+1}, \ldots, X_{m+r} \neq y \mid X_m = z)\mathbb{P}_x(X_m = z)}{\mathbb{P}_x(X_m \neq y)} \leq (\text{ by definition of } \epsilon)$$

$$\frac{(1-\epsilon)\sum_{z \neq y} \mathbb{P}_x(X_m = z)}{\mathbb{P}_x(X_m \neq y)} = 1 - \epsilon,$$

and so our claim is proved.

Next, we claim that

$$\mathbb{P}_x(X_{m+1}, \ldots, X_{m+r} \neq y \mid X_m \neq y) = \mathbb{P}_x(\tau_y > m + r \mid \tau_y > m).$$

Indeed, we have

$$\begin{aligned}
&\mathbb{P}_x(X_{m+1}, \ldots, X_{m+r} \neq y \mid X_m \neq y) = \text{(by the Markov Property reversed)} \\
&\mathbb{P}_x(X_{m+1}, \ldots, X_{m+r} \neq y \mid X_0, \ldots, X_m \neq y) = \text{(using Exercise 1.2)} \\
&\mathbb{P}_x(X_0, \ldots, X_{m+r} \neq y \mid X_0, \ldots, X_m \neq y) = \text{(by definition of } \tau\text{)} \\
&\mathbb{P}_x(\tau_y > m + r \mid \tau_y > m)
\end{aligned}$$

as claimed.

Combining these two claims we deduce

$$\mathbb{P}_x(\tau_y > m + r \mid \tau_y > m) \leq 1 - \epsilon.$$

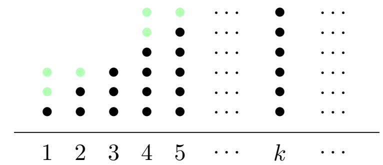
Hence, for $k > 0$:

$$\mathbb{P}_x(\tau_y > kr) = \mathbb{P}_x(\tau_y > kr \mid \tau_y > (k-1)r)\mathbb{P}_x(\tau_y > (k-1)r) \leq (1-\epsilon)\mathbb{P}_x(\tau_y > (k-1)r).$$

Iterating the right hand side, we get $\mathbb{P}_x(\tau_y > kr) \leq (1-\epsilon)^k$. Using now that the values of τ_y are nonnegative integers, and that $\mathbb{P}_x(\tau_y > n)$ is decreasing in n, we have

$$\mathbb{E}_x(\tau_y) = \sum_{n \geq 0} \mathbb{P}_x(\tau_y > n) \leq \sum_{k \geq 0} r\mathbb{P}_x(\tau_y > kr) \leq r \sum_{k \geq 0} (1-\epsilon)^k < \infty.$$

The next figure helps to clarify the last line:



Indicating the value of τ_y on the horizontal axis, each bullet above the value k represents the probability $\mathbb{P}_x(\tau_y = k)$. Hence, the k black bullets above the value k represent the contribution $k\mathbb{P}_x(\tau_y = k)$ to $\mathbb{E}_x(\tau_y)$, with a horizontal line of black bullets starting at $n+1$ representing the the probability $\mathbb{P}_x(\tau_y > n)$. The equality is just the expression of the fact that the sum over all black bullets grouped vertically equal the sum with bullets grouped horizontally. The green bullets indicate the additional contributions added in the first inequality (with $r = 3$). □

We can now complete the proof of the main result of this section:

Proof of Theorem 2.7. (We use $\mathbb{P}^x$ and $\mathbb{P}_x$ interchangeably.)

Considering an arbitrary fixed $z \in S$, let us define

$$\tilde{\alpha}^z(y) = \mathbb{E}^z(\# \text{ of visits to } y \text{ before returning to } z) = \sum_{n\geq 0} \mathbb{P}^z(X_n = y, \tau_z > n).$$

Notice that

$$\mathbb{E}^z(\tau_z) = \sum_{n\geq 0} \mathbb{P}^z(\tau_z > n) = \sum_y \tilde{\alpha}^z(y)$$

and thus $\tilde{\alpha}^z(y) \leq \mathbb{E}^z(\tau_z) < \infty$ for all $y \in S$.

The distribution $\tilde{\alpha}^z$ is stationary. Indeed, consider

$$\sum_x \tilde{\alpha}^z(x)\Pi(x,y) = \sum_x \sum_{n\geq 0} \mathbb{P}^z(X_n = x, \tau_z > n)\Pi(x,y).$$

Using that the event $\{\tau_z > n\}$ is determined by $X_1, \dots, X_n$, $\{\tau_z > n\} = \{X_1 \neq z, \dots, X_n \neq z\}$, we get

$$\mathbb{P}^z(X_{n+1} = y, X_n = x, \tau_z > n) = \mathbb{P}^z(X_n = x, \tau_z > n)\mathbb{P}^z(X_{n+1} = y \,|\, X_n = x, \tau_z > n) =$$
$$= \mathbb{P}^z(X_n = x, \tau_z > n)\mathbb{P}^z(X_{n+1} = y \,|\, X_n = x, X_1 \neq z, \dots, X_n \neq z) = \mathbb{P}^z(X_n = x, \tau_z > n)\Pi(x,y)$$

$$\sum_x \sum_{n\geq 0} \mathbb{P}^z(X_n = x, \tau_z > n)\Pi(x,y) = \sum_{n\geq 0} \sum_x \mathbb{P}^z(X_{n+1} = y, X_n = x, \tau_z > n) =$$
$$= \sum_{n\geq 0} \mathbb{P}^z(X_{n+1} = y, \tau_z > n) = \sum_{n\geq 1} \mathbb{P}^z(X_n = y, \tau_z > n-1).$$

The right hand side is almost the expression for $\tilde{\alpha}^z(y)$. Let us see, how much it differs.

$$\sum_{n\geq 1} \mathbb{P}^z(X_n = y, \tau_z > n-1) - \tilde{\alpha}^z(y) = \sum_{n\geq 1} \mathbb{P}^z(X_n = y, \tau_z > n-1) - \sum_{n\geq 0} \mathbb{P}^z(X_n = y, \tau_z > n) =$$
$$= \sum_{n\geq 1} \mathbb{P}^z(X_n = y, \tau_z = n) - \mathbb{P}^z(X_0 = y, \tau_z > 0) = \underbrace{\mathbb{P}^z(X_{\tau_z} = y)}_{\delta_{y,z}} - \underbrace{\mathbb{P}^z(X_0 = y)}_{\delta_{y,z}} = 0.$$

For the last equality, just check the cases $y = z$ and $y \neq z$, noticing that always $X_{\tau_z} = z$.

To get a probability measure, we normalize $\tilde{\alpha}^z$ by $\sum_y \tilde{\alpha}^z(y) = \mathbb{E}^z(\tau_z)$. The resulting probability distribution $\alpha^z(x) = \frac{\tilde{\alpha}^z(x)}{\mathbb{E}^z(\tau_z)}$ is also stationary, $\alpha^z\Pi = \alpha^z$. Moreover, as $\tilde{\alpha}^z(z) = 1$ (the number of visits of the chain X_n to the state z before returning to z is exactly one, $X_0 = z$) we get $\alpha^z(z) = \frac{1}{\mathbb{E}^z(\tau_z)}$.

However, in view of unicity, we can conclude that the probability distributions α^z actually do not depend on z, yielding a unique stationary α such that $\alpha(x) = \frac{\tilde{\alpha}^x(x)}{\mathbb{E}^x(\tau_x)} = \frac{1}{\mathbb{E}^x(\tau_x)}$ for any $x \in S$. □

□

2.5 Convergence of Markov chains

In this section we prove that, under mild conditions, the distribution of the state of a Markov chain at time n converges to the stationary distribution as $n \to \infty$. To formulate this statement we introduce a metric on the set of probability distributions that will allow as to define convergence:

Definition 2.6. The *total variation distance* of two probability measures μ, ν on a probability space $(S, \mathcal{F})$ is defined by

$$\|\mu - \nu\| = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

Exercise: prove that, for S finite,
$\|\mu - \nu\| = \max_{A \subset S} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|$.

Theorem 2.10 (Convergence of Markov chains)**.** Suppose that S is finite and Π is irreducible and aperiodic. Let α be the stationary distribution. Then there exists $\zeta \in (0,1)$ and $C > 0$ such that

$$\max_{x \in S} \|\mathbb{P}_x(X_n = \cdot) - \alpha\| = \max_{x \in S} \frac{1}{2} \sum_{y \in S} |\mathbb{P}_x(X_n = y) - \alpha(y)| \leq C\zeta^n.$$

Remark: Considering the stochastic matrix A consisting of $|S|$ identical rows, each equal to α, i.e. $A(x,y) = \alpha(y)$ for all $x, y \in S$, the statement implies $|\Pi^n(x,y) - A(x,y)| \to 0$. For example,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}^n \to \begin{pmatrix} 1/5 & 2/5 & 2/5 \\ 1/5 & 2/5 & 2/5 \\ 1/5 & 2/5 & 2/5 \end{pmatrix}$$

Proof. By Lemma 2.5, there exists r such that all elements of Π^r are positive. Choose δ such that $\Pi^r(x,y) \geq \delta\alpha(y)$ for any $x, y \in S$. Consider the stochastic matrix A introduced in the Remark above, and define

$$Q := \frac{1}{\xi}(\Pi^r - (1-\xi)A) \text{ where } \xi := 1 - \delta.$$

The matrix Q is stochastic:

$$\sum_{y \in S} Q(x,y) = \frac{1}{\xi}(1 - (1-\xi)1) = 1.$$

By induction on k we prove that Π^{rk} almost equals A:

$$\Pi^{rk} = (1-\xi^k)A + \xi^k Q^k.$$

Indeed, for $k = 1$, we have $\Pi^r = (1-\xi)A + \xi Q$ by the definition of Q.

Assuming the induction hypothesis is true for k, we calculate

$$\Pi^{r(k+1)} = \Pi^{rk}\Pi^r = [(1-\xi^k)A+\xi^k Q^k]\Pi^r = (1-\xi^k)A\Pi^r+\xi^k(1-\xi)Q^k A+\xi^{k+1}Q^{k+1}. \tag{2}$$

Using the stationarity of α, we can show that $A\Pi^r = A$:

$$(A\Pi^r)(x,y) = \sum_z A(x,z)\Pi^r(z,y) = \sum_z \alpha(z)\Pi^r(z,y) = \alpha(y) = A(x,y).$$

Next, we observe that $MA = A$ holds for any stochastic matrix M, (we want to use this with $M := Q^k$):

$$(MA)(x,y) = \sum_z M(x,z)A(z,y) = \sum_z M(x,z)\alpha(y) = \alpha(y).$$

Plugging the latter two facts into (2), we obtain

$$\Pi^{r(k+1)} = (1-\xi^k)A + \xi^k(1-\xi)A + \xi^{k+1}Q^{k+1},$$

which completes the induction step.

To conclude, consider $n = kr + j$ with $j \in \{0, 1, \dots, r-1\}$. Using again that $A\Pi^j = A$, we get $\Pi^{rk+j} - A = \xi^k(Q^k\Pi^j - A)$ and thus

$$\frac{1}{2}\sum_y |\Pi^{rk+j}(x,y) - \alpha(y)| \leq \frac{1}{2}\xi^k 2,$$

where the factor 2 represents the bound
$\sum_y |\mu(y) - \alpha(y)| \leq \sum_y(\mu(y) + \alpha(y)) = 2$ with $\mu(y) = (Q^k\Pi^j)(x,y)$.

Choosing $\zeta = \xi^{1/r}$ and $C = \xi^{-(r-1)/r}$, we get

$$\xi^k = \zeta^{n-j} \leq \zeta^n \zeta^{-(r-1)} = C\zeta^n.$$

□

2.6 Strong Markov Property

Let $X_0, X_1, \ldots$ be a Markov chain on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

A random variable $T : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a stopping time if the event $\{T = n\}$ depends only on $X_0, \ldots, X_n$ for every $n \in \mathbb{N}$.

Theorem 2.11 (Strong Markov Property)**.** Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix Π and state space S, and let T be a stopping time. Then conditional on $T < \infty$ and $X_T = s$, the sequence $(X_{T+n})_{n\in\mathbb{N}}$ is a Markov chain with transition matrix Π, initial state s, and independent of $X_0, \ldots, X_T$.

2.7 Reversibility

Definition 2.7. We say that a probability distribution α satisfies the detailed balance condition with respect to the transition matrix Π if

$$\alpha(x)\Pi(x, y) = \alpha(y)\Pi(y, x) \text{ for all } x, y \in S. \tag{3}$$

A Markov chain is reversible if it admits a distribution α satisfying the detailed balance condition.

Thinking about the Markov chain in terms of a big number of particles distributed over the states in S with density α and each moving independently according to the probabilities given by Π, reversibility means that, at any moment, the number $\alpha(x)\Pi(x, y)$ of particles moving from x to y equals the number $\alpha(y)\Pi(y, x)$ of particles moving from y to x. We have a very useful simple claim:

Proposition 2.12. A probability distribution α satisfying (3) is stationary.

Proof. $\sum_y \alpha(y)\Pi(y, x) = \alpha(x) \sum_y \Pi(x, y) = \alpha(x)$. □

Example 1: Random walk on a graph G.
This is defined by letting $\Pi(x, y) = \mathbb{1}_{x\sim y}$. Consider the probability distribution $\alpha(x) := deg(x)/2|E|$, where the degree $deg(x)$ is the number of edges of G incident with x (notice that $\sum_x deg(x) = 2|E|$). Then α satisfies the detailed balance condition (exercise) and is thus stationary.

Example 2: Biased random walk on a cycle. The states in S form a cycle, along which we move clockwise with some probability p and anticlockwise with probability $1-p$; to make this more precise, we define

$$S = \{1, 2, \ldots, n\}, \Pi(x, y) = \begin{cases} p & \text{if } y = x+1 \pmod{n} \\ 1-p & \text{if } y = x-1 \pmod{n} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $\alpha(n) = \frac{1}{n}$ is stationary but, for $p \neq \frac{1}{2}$, it does not satisfy the detailed balance condition (3).

We will study reversible Markov chains closely in the next chapter

3 Random Walks on Graphs

Let $G = (V, E)$ be a countable connected graph, in which every vertex has at least one and at most finitely many incident edges. (See e.g. Diestel's textbook
http://diestel-graph-theory.com/index.html for graph-theoretic definitions.)

The *Simple Random Walk on* G starts at some $X_0 = o \in V$, and at step i, moves to a uniformly random neighbour X_i of the previous position X_{i-1}, chosen independently from all previous choices.

Formally, our probability space $(\Omega, \mathcal{F}, \mathbb{P}_o)$ is defined as follows.
• $\Omega = V^{\mathbb{N}}$ consists of all sequences of vertices of G.
We could alternatively define Ω as the set of all 1-way infinite walks in G, i.e. the set

$$\{v_0, v_1, \ldots \mid v_i v_{i+1} \in E \text{ for every } i \geq 0\}.$$

You can choose your favourite of these two definitions of Ω.
• $\mathcal{F}$ is generated by the sets $F_{i,v} := \{x_0, x_1, \ldots \in \Omega \mid x_i = v\}$ for all $i \in \mathbb{N}$ and $v \in V$.
Thus $\mathcal{F}$ consists of all the subsets of Ω that can be made from the $F_{i,v}$ by a countable number of complement, union and intersection operations.
• $\mathbb{P} = \mathbb{P}_o$ is defined by

$$\mathbb{P}(X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n) = \frac{1}{d(v_0)} \cdot \frac{1}{d(v_1)} \cdots \frac{1}{d(v_{n-1})} \tag{4}$$

whenever $v_0 = o$ and $v_0 v_1 \ldots v_n$ is a path in G, and

$$\mathbb{P}(X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n) = 0$$

otherwise.

Exercise 3.1. For every $\{v_0, v_1, \ldots, v_n\}$, we have

$$\{X_0, X_1, \ldots \in \Omega \mid X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n\} \in \mathcal{F}.$$

Exercise 3.2. The following events are measurable (i.e. in $\mathcal{F}$) for every $v \in V$:

- $X_i = v$ for at least 3 distinct i.
- $X_i = v$ for at most 17 distinct i.

- $X_i = v$ for infinitely many i.

It is intuitively clear, and easy to check, that

$$\mathbb{P}(X_n = v_n \mid X_0 = v_0, \dots, X_{n-1} = v_{n-1}) = \mathbb{P}(X_n = v_n \mid X_{n-1} = v_{n-1}) \quad \text{(MP)}$$

This is the *Markov property.* Thus random walk on a graph is an example of a Markov chain.

Example: Let G be a path on n edges, and q, p its end-vertices. Let $v_i, 0 \leq i \leq n$ denote the ith vertex as we move from q to p. Then

$$\mathbb{P}_{v_i}(\text{hit } p \text{ before hitting } q) = i/n.$$

The following theorem is a bit surprising, as it relates random walks with electrical networks, which we will define and study below. But it generalises the above example

Theorem 3.1. For every finite connected graph $G = (V, E)$, and every $p, q, o \in V$,

$$\mathbb{P}_o(\text{hit } p \text{ before hitting } q) = u(o),$$

where $u(x)$ denotes the voltage at $x \in V$ in the electrical network obtained from G by replacing each edge by an 1 Ohm resistor and imposing a potential of 1 Volt at p and 0 Volt at q.

Proof sketch. Both functions $f(o) := \mathbb{P}_o(\text{hit } p \text{ before hitting } q)$ and $u(o)$ are harmonic at every o except p, q, where they coincide. Only one such function can exist. □

We will now introduce the necessary terminology and lemmas in order to give a complete proof of Theorem 3.1.

Definition 3.1. A function $u : V \to \mathbb{R}$ is harmonic on $U \subset V$, if for every $x \in U$

$$u(x) = \sum_{yx \in E} u(y)/d(x).$$

Here $d(x)$ is the degree of x, i.e. the number of edges of x.

Lemma 3.2 (Maximum Principle)**.** Let $G = (V, E)$ be a finite connected graph. If $u : V \to \mathbb{R}$ is harmonic on a set of vertices $U \subsetneq V$, then it attains its maximum value at some $z \notin U$.

Proof. Notice that the average of real numbers $x_1, x_2, \ldots, x_n$ equals their maximum only if they are all equal. Now start from any $x \in U$ attaining the maximum if such x exists (if it doesn't we are done). Apply the previous remark to prove that $u(y) = u(x)$ for every $yx \in E$. Repeat the whole argument on all neighbours y of x (for which we now know that they attain the maximum of u) and then on their neighbours, their neighbours' neighbours and so on, until reaching a vertex $z \notin U$. Then u attains its maximum value (i.e. $u(x)$) at z as desired. □

Applying the Maximum Principle to $-u$ we obtain the *Minimum* Principle:

Corollary 3.3 (Minimum Principle)**.** If $u : V \to \mathbb{R}$ is harmonic on a set of vertices $U \subsetneq V$, then it attains its *minimum* value at some $z \notin U$.

As in the above sketch, we let $f(o) := \mathbb{P}_o(\text{hit } p \text{ before hitting } q)$. It is trivial that $f(p) = 1$ and $f(q) = 0$, because we count step 0 as a 'hit' by definition. We will prove that f is harmonic elsewhere.

Lemma 3.4. $f(x) = \sum_{yx \in E} f(v)/d(x)$ for every $x \in V \setminus \{p, q\}$.

Proof. Let $x \in V \setminus \{p, q\}$. The Simple Random Walk moves to each neighbour of x with probability $1/d(x)$. From there on, the probability of any event does not depend on the past by (MP). It follows that

$$\mathbb{P}_x(\text{hit } p \text{ before hitting } q) = \sum_{yx \in E} \frac{1}{d(x)} \cdot \mathbb{P}_y(\text{hit } p \text{ before hitting } q).$$

□

Let u be as in the statement of Theorem 3.1. In the next section (Proposition 3.6) we will establish

Lemma 3.5. u is harmonic on $V \setminus \{p, q\}$ by the electrical network theory (Kirchhoff's node law and Ohm's law).

Proof of Theorem 3.1. Consider the function $g : V \to \mathbb{R}$, $g(x) := f(x) - u(x)$. Then $g(p) = g(q) = 0$, and g is harmonic elsewhere. By the Maximum and Minimum Principles both its maximum and minimum are attained outside $U = V \setminus \{p, q\}$, i.e. at $\{p, q\}$. That is, both the maximum and minimum of g is 0. Hence g is identically 0, implying that f coincides with u. □

3.1 Electrical networks

Let $G = (V, E)$ be a (finite or infinite) graph, with two specified vertices $p \neq q \in V$ called the *terminals*. Let $\overrightarrow{E}$ denote the set of ordered pairs (x, y) with $xy \in E$. We write $\overrightarrow{xy}$ to denote (x, y). Note that for every $xy \in E$, both $\overrightarrow{xy}$, $\overrightarrow{yx}(:= \overleftarrow{xy})$ lie in $\overrightarrow{E}$.

We say that a function $f : \overrightarrow{E} \to \mathbb{R}$ is *antisymmetric*, and write $f : \overrightarrow{E} \hookrightarrow \mathbb{R}$, if $f(\overrightarrow{xy}) = -f(\overleftarrow{xy})$ for every $xy \in E$. All functions we consider from now on are antisymmetric.

Definition 3.2. A *flow* in $G = (V, E)$ is a function $f : \overrightarrow{E} \hookrightarrow \mathbb{R}$ such that $f^*(x) = 0$ for every $x \in V$ (KNL), where $f^*(x) := \sum_{y \sim x} f(\overrightarrow{xy})$.

Here KNL stands for Kirchhoff's Node Law. Intuitively it says that current is preserved at x.

We say that $f : \overrightarrow{E} \hookrightarrow \mathbb{R}$ is a $p - q$ flow, or a flow from p to q, if (KNL) holds for every $x \notin \{p, q\}$. Typically we have $f^*(p) \neq 0$. The *intensity* of f is $f^*(p)$. Moreover, we say that f is a flow from p if (KNL) holds for every $x \neq p$.

Exercise 3.3.

1. If f is a $p - q$ flow on a finite graph G, then $f^*(p) = -f^*(q)$.
2. If f is a flow from p and $f^*(p) \neq 0$, then G is infinite.

Definition 3.3. We say that $f : \overrightarrow{E} \hookrightarrow \mathbb{R}$ satisfies Kirchhoff's Cycle Law (KCL), if for every closed walk (equivalently for every cycle) $x_0, x_1, \ldots, x_n(= x_0)$ we have $\sum_{i=0}^{n-1} f(\overrightarrow{x_i x_{i+1}}) = 0$.

Definition 3.4. An <u>(electrical) current</u> (of intensity $I \in \mathbb{R}$ from p to q in G, is a p-q flow $i : \overrightarrow{E} \hookrightarrow \mathbb{R}$ (of intensity I) satisfying Kirchhoff's Cycle Law. If i has intensity 1, then it is called a <u>unit current</u>.

Definition 3.5. We say that a pair of functions $i : \overrightarrow{E} \hookrightarrow \mathbb{R}$, $u : V \to \mathbb{R}$ satisfies *Ohm's Law* (OL) if $i(\overrightarrow{xy}) = u(x) - u(y)$.

Note that if i, u satisfies (OL), then so does $i, u + a$ for every $a \in \mathbb{R}$. We say that i the *Ohm dual* of u and u is an Ohm dual of i, if (OL) holds.

Proposition 3.6. If i, u satisfy (OL), then u is harmonic at $x \in V$ if and only if i satisfies (KNL) at x, i.e. if $i^*(x) = 0$.

Proof. Since i, u satisfy (OL), we have

$$\sum_{y \sim x} u(y) = \sum_{y \sim x} \big(u(x) - i(\overrightarrow{xy})\big) = d(x)u(x) - \sum_{y \sim x} i(\overrightarrow{xy}),$$

which is by definition equal to $d(x)u(x) - i^*(x)$. Thus $i^*(x) = 0$ if and only if $\sum_{y \sim x} u(y)/d(x) = u(x)$, which is the definition of being harmonic. $\square$

Proposition 3.7. Let G be a connected graph, and $i : \overrightarrow{E} \hookrightarrow \mathbb{R}$. Then, there is $u : V \to \mathbb{R}$ such that the pair i, u satisfies (OL) if and only if i satisfies (KCL). This u is unique up to an additive constant.

Proof. Suppose $i : \overrightarrow{E} \hookrightarrow \mathbb{R}$ satisfies (KCL). To construct u, pick any vertex $o \in V$, and for any other vertex x an x-o walk P_x. Set $u(o) = 0$. Then, for any $x \neq o$, let $u(x) = \sum_{\overrightarrow{e} \in \overrightarrow{E}(P_x)} i(\overrightarrow{e})$, where $\overrightarrow{E}(P_x)$ is the set of edges of P_x directed from x to o.

It follows from the fact that i satisfies (KCL) that $u(x)$ does not depend on our choice of P_x; indeed, if P'_x is another x-o walk, then traversing P_x from x to o and then P'_x back from o to x we obtain a closed walk, and applying (KCL) to this walk yields $\sum_{\overrightarrow{e} \in \overrightarrow{E}(P_x)} i(\overrightarrow{e}) = \sum_{\overrightarrow{e} \in \overrightarrow{E}(P'_x)} i(\overrightarrow{e})$.

We now claim that the pair i, u satisfies (OL). To see this, let xy be any edge. Since $u(x)$ does not depend on the choice of P_x, we may assume that P_x is obtained from P_y by prefixing the edge xy. Thus, by the definition of u, we have $u(x) - u(y) = i(\overrightarrow{xy})$, in agreement with (OL).

Conversely, if a function $u : V \to \mathbb{R}$ exists such that the pair i, u satisfies (OL), then i satisfies (KCL): for given a closed walk $x_0 x_1 \ldots x_k$ we can write $\sum_{0 \leq n < k} i(\overrightarrow{x_n x_{n+1}}) = \sum_{0 \leq n < k} (u(x_n) - u(x_{n+1}))$, and the latter expression equals 0 as each $u(x_n)$ appears once with a positive and once with a negative sign in the sum.

Finally, consider the family of functions $u_r := u + r, r \in \mathbb{R}$. It is straightforward to check that all pairs i, u_r satisfy (OL). We claim that no other pair i, v does. For if it does, then the value $v(o)$ combined with (OL) uniquely determines the values $v(y)$ of all neighbours y of o. These in turn uniquely determine the values at the neighbours' neighbours, and continuing inductively like this we can prove that $v(x) = u(x) + v(o)$ for every $x \in V$. $\square$

3.2 Existence of currents

Let $G = (V, E)$ be a finite connected graph, and consider $p, q \in V$. We will construct the unit $p - q$ current using the Simple Random Walk!

Start a Simple Random Walk at p, and stop when it first reaches q. Let $f(\overrightarrow{xy})$ be the net amount of traversals of $\overrightarrow{xy}$ by this Simple Random Walk, i.e.

$$f(\overrightarrow{xy}) = (\text{\# of traversals from } x \text{ to } y) - (\text{\# of traversals from } y \text{ to } x).$$

Let $i(\overrightarrow{xy}) := \mathbb{E}f(\overrightarrow{xy})$. To see that this is well-defined (and finite), note that (# of traversals from x to y)+(# of traversals from y to x) $\leq \tau_q$ because all these traversals occur before reaching q, and that $\mathbb{E}(\tau_q) < \infty$ by Lemma 2.9. Note that f is random, but i is not.

Moreover, i is antisymmetric, and it satisfies (KNL) at all $x \notin \{p, q\}$, while $i^*(p) = 1$. We now show that it also satisfies (KCL). For this, let $C = x_0 x_1 \ldots x_k$ be a cycle. Then

$$\sum_{i=1}^{k-1} f(\overrightarrow{x_i x_{i+1}}) = \sum_{n=1}^{+\infty} X_n,$$

where

$$X_n := \begin{cases} 1, & \text{if after the } n\text{th visit to a vertex of } C\text{, we traverse an edge of } C \text{ in the forward direction,} \\ -1, & \text{if after the } n\text{th visit to a vertex of C, we traverse an edge of } C \text{ in the backward direction,} \\ 0, & \text{if after the } n\text{th visit to a vertex of } C\text{, we do not traverse an edge of } C\text{, or there is no } n\text{th visit to a vertex of } C \text{ (because we reached } q \text{ before that visit).} \end{cases}$$

Note that $\sum_{n=1}^{+\infty} |X_n| \leq \tau_q$, and therefore $\mathbb{E}(\sum_{n=1}^{+\infty} |X_n|) \leq \mathbb{E}(\tau_q) < \infty$ where we used Lemma 2.9 again. Thus by linearity of expectation (§ 1.4.1) we have

$$\sum_{i=1}^{k-1} i(\overrightarrow{x_i x_{i+1}}) = \mathbb{E}(\sum_{n=1}^{+\infty} X_n) = \sum_{n=1}^{+\infty} \mathbb{E}(X_n).$$

But each X_n has $\mathbb{E}(X_n) = 0$. Hence $\sum_{i=1}^{k-1} i(\overrightarrow{x_i x_{i+1}}) = 0$, i.e. i satisfies (KCL).

3.3 Energy

The *energy* of a flow $f : \overrightarrow{E} \hookrightarrow \mathbb{R}$ is defined by

$$E(f) := \frac{1}{2} \sum_{e \in \overrightarrow{E}} f(e)^2.$$

Similarly, the energy of a function $u : V \to \mathbb{R}$ is defined by

$$E(u) := \sum_{xy \in E} (u(x) - u(y))^2.$$

3.4 Effective Resistance

Theorem 3.8. (Without proof) Let $G = (V, E)$ be a finite connected graph, and let $a, b \in V$. Then the following values are equal.

1. $R_1 := u(a) - u(b)$, for any Ohm dual u of the a-b current of intensity 1;
2. $R_2 := 1/i^*(a) = -1/i^*(b)$, where i is the Ohm dual of the unique $v : V \to \mathbb{R}$ which satisfies $v(a) = 1, v(b) = 0$ and is harmonic on $V - \{a, b\}$;
3. $R_3 := \inf\{E(j) \mid j \text{ is an } a\text{-}b \text{ flow of intensity } 1\}$;
4. $R_4 := \sup\{1/E(u) \mid u : V \to \mathbb{R} \text{ with } u(a) - u(b) = 1\}$;
5. $R_5 := \sup\{(v(a) - v(b))^2 \mid v : V \to \mathbb{R} \text{ with } E(v) = 1\}$.

We define the *effective resistance* $\mathcal{R}_{ab}$ between a and b in N to be equal to the values R_i of Theorem 3.8. The *effective conductance* $\mathcal{C}_{ab}$ between a and b is defined by $\mathcal{C}_{ab} = 1/\mathcal{R}_{ab}$.

Exercise 3.4. Check that if G consists of just one edge xy, then $\mathcal{C}_{xy} = 1$.

3.5 Series Law

Theorem 3.9. Let G, H be two finite connected graphs intersecting only in a single vertex s, and p, q be two vertices in G, H, respectively. Then $\mathcal{R}^{G \cup H}_{pq} = \mathcal{R}^G_{ps} + \mathcal{R}^H_{sq}$.

Proof. If i_G is the $p-s$ unit current in G and i_H is the $s-q$ unit current in H, then their superposition $i := i_G \cup i_H$ is a unit $p-q$ flow in $G \cup H$. Indeed, $i^*(s) = -1+1 = 0$. It is easy to see that i still satisfies (KCL), hence i is the $p-q$ unit current in $G \cup H$. Letting u_i, u_{i_G}, u_{i_H} be the corresponding Ohm duals taking value 0 at s, we easily have

$$\mathcal{R}_{pq}^{G\cup H} = u_i(p) - u_i(q) = u_i(p) - u_i(s) + u_i(s) - u_i(q),$$

$$\mathcal{R}_{ps}^{G} = u_{i_G}(p) - u_{i_G}(s) = u_i(p) - u_i(s),$$

and

$$\mathcal{R}_{sq}^{H} = u_{i_H}(s) - u_{i_H}(q) = u_i(s) - u_i(q),$$

which gives $\mathcal{R}_{pq}^{G\cup H} = \mathcal{R}_{ps}^{G} + \mathcal{R}_{sq}^{H}$, as desired. □

Corollary 3.10. Let G, H be two graphs intersecting only in a single vertex s, and p, q be two vertices in G, H, respectively. Then $\mathbb{P}_s(\text{ hit } p \text{ before } q) = \frac{\mathcal{R}_{sq}^{H}}{\mathcal{R}_{pq}^{G\cup H}}$.

Proof. Let $f(o) := \mathbb{P}_o(\text{ hit } p \text{ before } q)$. Clearly $f(p) = 1$ and $f(q) = 0$, and by Lemma 2 f is harmonic everywhere except p, q. Consider the Ohm dual u of the unit $p-q$ flow that satisfies $u(q) = 0$. Notice that $u(s) = \mathcal{R}_{sq}^{H}$ and $f(p) = u(p)/\mathcal{R}_{pq}^{G\cup H}$. By Theorem 1 f coincides with $u/\mathcal{R}_{pq}^{G\cup H}$. Hence $f(s) = \frac{\mathcal{R}_{sq}^{H}}{\mathcal{R}_{pq}^{G\cup H}} = \frac{\mathcal{R}_{sq}^{H}}{\mathcal{R}_{ps}^{G\cup H} + \mathcal{R}_{sq}^{G\cup H}}$. □

3.6 Parallel Law

Theorem 3.11. Let G, H be two finite connected graphs intersecting in two vertices p, q and sharing no edges. Then $\mathcal{C}^{G\cup H} = \mathcal{C}^{G} + \mathcal{C}^{H}$.

Proof. Let i_G be the Ohm dual of the unique harmonic function u_G in $G \setminus \{p, q\}$ with $u_G(p) = 1$ and $u_G(p) = 0$. We define i_H similarly. Then the superposition $i := i_G \cup i_H$ is the Ohm dual of $u_G \cup u_H$. We have

$$\mathcal{C}^{G\cup H} = i^*(p) = i_G^*(p) + i_H^*(p) = \mathcal{C}^{G} + \mathcal{C}^{H}.$$

□

The Series and Parallel Law can be used to compute the effective resistance of complex graphs, by decomposing them into simpler ones.

Examples 3.1. 1. If G is a single edge $\{p, q\}$, then $\mathcal{R}_{pq} = 1$.

2. If G is the path on n edges and p, q are its endpoints, then $\mathcal{R}_{pq} = n$.

3. If G is the graph with n parallel edges, then $\mathcal{R}_{pq} = 1/n$.

4. If G is a path on $n + m$ edges, p, q are its endpoints, and o is the $(m + 1)$th vertex of the path, then $\mathbb{P}_o($ hit p before $q) = \frac{n}{n+m}$.

3.7 Rayleigh's monotonicity law

Theorem 3.12 (Rayleigh's monotonicity law)**.** Let $G = (V, E)$ be a finite connected graph and let H be obtained from G by contracting an edge or adding an edge. Then $\mathcal{R}^H_{pq} \leq \mathcal{R}^G_{pq}$ for every $p, q \in V$.

Proof. Any $p - q$ flow f in G gives rise to a $p - q$ flow f' in H where $f(e) = f'(e)$ for every common edge e. The result follows by using item (3) of Theorem 3.8. □

3.8 Recurrence and transience

Definition 3.6. Let $G = (V, E)$, and $x \in V$. We define the *first hitting time* $\tau_x = \min\{n \geq 0 \mid X_n = x\}$, where X_n denotes the position of the Simple Random Walk at time n. We define the *first return time* $\tau_x^+ = \min\{n \geq 1 \mid X_n = x\}$.

Notice that $\tau_x = \tau_x^+$, unless the Simple Random Walk starts at x.

Definition 3.7. A graph $G = (V, E)$ is called *recurrent*, if $\mathbb{P}_x(\tau_x^+ < \infty) = 1$ for some $x \in V$. If G is not recurrent, then it is said to be *transient*.

Proposition 3.13. Let $G = (V, E)$ be connected, and let N_x, $x \in V$ denote the number of visits to x after time 1. Then the following are equivalent:

1. $\mathbb{P}_x(\tau_x^+ < \infty) = 1$ for some $x \in V$,

2. $\mathbb{P}_x(\tau_y^+ < \infty) = 1$ for every $x, y \in V$,

3. $\mathbb{P}_x(N_x = \infty) = 1$ for some $x \in V$,

4. $\mathbb{P}_x(N_y = \infty) = 1$ for every $x, y \in V$.

Proof. The implications $4 \to 3 \to 1$ and $2 \to 1$ are obvious.

$1 \to 3$: Let $\sigma := \sup\{n \geq 0 \mid X_n = x\}$; this is the time of the last visit to x if such a time exists, and ∞ otherwise.

For every n we have

$$\mathbb{P}_x(\sigma = n) = \mathbb{P}_x(X_n = x)\mathbb{P}_x(X_i \neq x \forall i > n) = \Pi^n(x,x)(1 - \mathbb{P}_x(\tau_x^+ < \infty)),$$

where we used the Markov property. Since $\mathbb{P}_x(\tau_x^+ < \infty) = 1$, this implies $\mathbb{P}_x(\sigma = n) = 0$ for all n, and so $\mathbb{P}_x(\sigma < \infty) = 0$ which is tantamount to $\mathbb{P}_x(N_x = \infty) = 1$.

$1 \to 2$: Let σ_n be the (random) time of the nth visit to x by random walk starting at x (thus $\sigma_0 = 0$). Since we have proved $1 \to 3$, we know that σ_n is finite for every $n \in \mathbb{N}$. Consider the random variables $Y_n := \mathbb{1}_{y \text{ is visited before } x \text{ after time } \sigma_n}$. Note that σ_n is a stopping time, and so the strong Markov property yields that each Y_n has the same distribution as Y_0, and that the Y_n are mutually independent (§ 2.6). Note that the distribution of Y_0 is *Bernoulli*(q) for some $q > 0$, because there is a path P from x to y in G and with positive probability our random walk follows P in its first $|P|$ steps. Thus Exercise 1.3 yields that some Y_i is 1 almost surely, and so $\mathbb{P}_x(\tau_y^+ < \infty) = 1$ for every $y \in V$.

Next, we want to show that $\mathbb{P}_y(\tau_x^+ < \infty) = 1$ for every $y \in V$. Suppose to the contrary that $\mathbb{P}_y(\tau_x^+ = \infty) = p > 0$. As noted above, with positive probability q' our random walk started at x follows P in its first $|P|$ steps to reach y. From there on it moves independently of its past by the Markov property, and so with probability $q'p > 0$ both these events occur, i.e. we follow P in the first $|P|$ steps and then never return to x. But this contradicts our assumption $\mathbb{P}_x(\tau_x^+ < \infty) = 1$, hence proving that $\mathbb{P}_y(\tau_x^+ < \infty) = 1$.

Combining the last statement with $\mathbb{P}_x(\tau_y^+ < \infty) = 1$ and the strong Markov property we deduce $\mathbb{P}_y(\tau_y^+ < \infty) = 1$ for every y: from y we will almost surely visit x, and from there we will almost surely re-visit y.

Thus we have proved that the assumption on x we started with holds for any $y \in V$. Repeating the first part of the above proof now yields $\mathbb{P}_x(\tau_y^+ < \infty) = 1$ for every $x, y \in V$.

$2 \to 4$: Using the implication $1 \to 3$ replacing x by y, we deduce $\mathbb{P}_y(N_y = \infty) = 1$. Combining this with $\mathbb{P}_x(\tau_y^+ < \infty) = 1$ and the Markov property we deduce $\mathbb{P}_x(N_y = \infty) = 1$: starting at x we will almost surely visit y, and from then on we will almost surely return to y infinitely many times.

□

Exercise 3.5. Suppose $\mathbb{P}_x(\tau_y < \infty) = 1$ for some $x \neq y \in V$. Does this imply that G is recurrent?

Proposition 3.14. Let $G = (V, E)$ be connected. Then the following are equivalent:

1. G is transient
2. $\mathbb{P}_x(N_y < \infty) = 1$ for every $x, y \in V$.

Exercise 3.6. Let G be a finite graph. Then G is recurrent. (Hint: Use the above characterization of transience).

Let $G = (V, E)$ be an infinite graph, and $p \in V$. Consider a sequence $G_n = (V_n, E_n)$ of finite graphs containing p such that

- $V_n \subset V_{n+1}$ for every n, and $\cup_{n=1}^{\infty} V_n = V$,
- E_n comprises those edges in E with both endpoints in V_n.

For each n, construct an auxiliary graph G_n^* in which all vertices in $V \setminus V_n$ are replaced by a new vertex z_n which is adjacent to every vertex of V_n which is adjacent to $V \setminus V_n$. Define

$$\mathcal{R}_{p\infty}^{G} := \lim_{n\to\infty} \mathcal{R}_{pz_n}^{G_n^*}.$$

Exercise 3.7. The limit exists and does not depend on the choice of G_n.

Theorem 3.15. $\mathbb{Z}^2$ is recurrent, while $\mathbb{Z}^3$ is transient. More generally $\mathbb{Z}^d$ is transient for $d > 2$.

Recall that for two graphs connected in series we have $\mathbb{P}_s(\text{ hit } p \text{ before } q) = \frac{\mathcal{R}_{sq}^{H}}{\mathcal{R}_{pq}^{G\cup H}}$. We will use this below to prove the recurrence of $\mathbb{Z}^2$: we will apply this fact to auxiliary graphs obtained from $\mathbb{Z}^2$ by contracting a certain vertex set into a single vertex that will play the role of s.

For our proof we will need the following lemma

Lemma 3.16. Let H be a connected graph and suppose G is obtained from H by contracting a finite set of vertices $U \subset V(H)$. If H is transient then so is G.

Proof sketch of recurrence of $\mathbb{Z}^2$. Let C_n be the boundary of the box $[-n, n]^2$. Let G be the graph obtained from $\mathbb{Z}^2$ by contracting C_1 into a vertex s. By Lemma 3.16, it suffices to prove that G is recurrent.

Using Corollary 3.10 we can reduce recurrence to proving that $\mathcal{R}^G_{s\infty} = \infty$ (see lecture). We define G_n to be the subgraph of $\mathbb{Z}^2$ that is surrounded by C_n. Recall the definition of G^*_n. Consider the graph H_n obtained by contracting each C_i,$i \leq n$ in G^*_n into a vertex. By Rayleigh's monotonicity law $\mathcal{R}^{G^*_n}_{0z_n} \geq \mathcal{R}^{H_n}_{0z_n}$. The series and parallel laws give $\mathcal{R}^{H_n}_{oz_n} = \sum_{k=1}^{n} 1/n_k$, where n_k is the number of edges between C_{k-1} and C_k. Notice that there is a constant $c > 0$ such that $n_k \leq ck$. Hence $\mathcal{R}^{G^*_n}_{oz_n} \geq c \sum_{k=1}^{n} 1/k$. Since the harmonic series diverges, we obtain that $\mathcal{R}^G_{s\infty} = \infty$. □

We will not prove the transience of $\mathbb{Z}^3$ here, and it will not be part of the final exam. The proof is based on similar ideas, and now one has $\mathcal{R}^{\mathbb{Z}^3}_{0\infty} < \infty$.

4 Galton-Watson trees

A Galton-Watson tree is a Markov Chain Z_n, $n \in \mathbb{N}$ with state space $S = \mathbb{N}$, and

- $Z_0 := 1$,
- $Z_{n+1} := \sum_{i=1}^{Z_n} L_i^{n+1}$,

where the L_i^j are independent, identically distributed $\mathbb{N}$-valued random variables.

We write $p_k := \mathbb{P}(Z_1 = k) = \mathbb{P}(L_1^1 = k)$. Let L be a random variable with the distribution of L_i^j. That is, $\mathbb{P}(L = k) = p_k$.

Examples 4.1. 1. If $p_0 = 1$ and $p_k = 0$ for every $k > 0$, then $Z_n = 0$ for every $n > 0$.

2. If $p_1 = 1$ and $p_k = 0$ for every $k \neq 1$, then $Z_n = 1$ for every $n > 0$.

Exercise 4.1. 1. Alice keeps having children until the first son arrives, and all her offspring repeat this strategy. Calculate p_k.

2. Delete each edge of the rooted infinite binary tree with probability p independently. Calculate p_k.

In example 2 we saw that $Z_n = 1$ for every $n > 0$ whenever $p_1 = 1$. We will show that this is the only case where a state $l > 0$ is visited infinitely many times with positive probability.

Recall that an event A occurs *almost surely*, if $\mathbb{P}(A) = 1$.

Proposition 4.1. If $p_1 \neq 1$, then almost surely, either $Z_n \to 0$ or $Z_n \to \infty$.

Proof. We first assume that $p_0 = 0$. Then every individual has almost surely at least 1 offspring, which shows that $Z_n \geq Z_{n-1}$ for every $n > 0$. We will show that $Z_n \to \infty$ almost surely.

Consider the event $A = \{Z_n \not\to \infty\}$. Notice that when A occurs, Z_n is eventually a constant sequence. Hence we have

$$\mathbb{P}(A) = \mathbb{P}(\cup_{n=0}^{\infty}\{Z_k = Z_n \text{ for every } k > n\}).$$

Moreover, for every $m, n \in \mathbb{N}$ with $m > n$ we have

$$\mathbb{P}(Z_k = Z_n \text{ for every } m \geq k > n) \leq p_1^{m-n},$$

because on that event, all individuals in the kth generation have exactly 1 offspring, and $Z_k \geq 1$. We immediately obtain that

$$\mathbb{P}(Z_k = Z_n \text{ for every } k > n) = \lim_{m \to \infty} \mathbb{P}(Z_k = Z_n \text{ for every } m \geq k > n) = 0.$$

Hence $\mathbb{P}(A) = 0$ as well, which shows that $Z_n \to \infty$ almost surely.

We now assume that $p_0 > 0$. We will show that $Z_n = l$ occurs finitely often for every $l > 0$ almost surely. Indeed, if $Z_n = l > 0$, we have $\mathbb{P}(Z_{n+1} = 0) = p_0^l$. This implies that the probability of returning to state l is at most $1 - p_0^l$. Thus returning or not to state l is a Bernoulli random variable with parameter at most $1 - p_0^l$. It is now easy to see that

$$\mathbb{E}(\# \text{ returns to l}) \leq \mathbb{E}(X) = \mathbb{E}(\text{Geometric}(1 - p_0^l)) < \infty,$$

where X is the number of successes in a sequence of Bernoulli$(1 - p_0^l)$ until the first failure. On the other hand, if $\mathbb{P}(Z_n = l \text{ for infinitely many } n) > 0$, then $\mathbb{E}(\# \text{ returns to l}) = \infty$, which is a contradiction (for $l = 0$ this does not apply). Since each $l > 0$ is visited finitely often almost surely, either $Z_n = 0$ eventually or $Z_n \to \infty$. □

How do we decide which of the two ($Z_n \to 0$ or $Z_n \to \infty$) will happen with what probability?

Definition 4.1. The probability generating function (pdf) of L is defined by

$$f(s) := \sum_{k \geq 0} p_k s^k$$

for $s \in [0, 1]$.

Notice that $f(s) = \mathbb{E}(s^L)$ and $f(0) = p_0$. The following proposition generalises the latter fact.

Proposition 4.2. $\mathbb{E}(s^{Z_n}) = f^{(n)}(s) := (\underbrace{f \circ f \circ \ldots \circ f}_{n \text{ times}})(s)$ for every $s \in [0, 1]$.

Proof. Let $g_n(s) := \mathbb{E}(s^{Z_n})$. We will prove the assertion inductively. Assume that $g_{n-1}(s) = f^{(n-1)}(s)$ for some n. Conditioning on Z_{n-1} we have

$$g_n(s) = \sum_{l=1}^{\infty} \mathbb{P}(Z_{n-1} = l)\mathbb{E}(s^{\sum_{i=1}^{l} L_i^n}) = \sum_{l=1}^{\infty} \mathbb{P}(Z_{n-1} = l)\mathbb{E}(\prod_{i=1}^{l} s^{L_i^n}).$$

Recall that $\mathbb{E}(A \cdot B) = \mathbb{E}(A) \cdot \mathbb{E}(B)$ when A, B are independent random variables. Hence we have

$$\mathbb{E}(\prod_{i=1}^{l} s^{L_i^n}) = \prod_{i=1}^{l} \mathbb{E}(s^{L_i^n}) = \prod_{i=1}^{l} \mathbb{E}(s^L) = f(s)^l,$$

which gives

$$g_n(s) = \sum_{l=1}^{\infty} \mathbb{P}(Z_{n-1} = l) f(s)^l = \mathbb{E}(f(s)^{Z_{n-1}}) = g_{n-1}(f(s)).$$

By our inductive hypothesis we have

$$g_n(s) = g_{n-1}(f(s)) = f^{(n-1)}(f(s)) = f^{(n)}(s),$$

as desired. □

The above proposition is important because $f^{(n)}(0) = \mathbb{P}(Z_n = 0)$.

Definition 4.2. Let $q := \mathbb{P}(Z_n \to 0) = \mathbb{P}(\text{ there is } n \text{ such that } Z_n = 0)$ be the extinction probability.

Proposition 4.3. $q = \lim_{n\to\infty} f^{(n)}(0)$.

Proof. Let A_n be the event $\{Z_n = 0\}$. Notice that

$$\{\text{there is } n \text{ such that } Z_n = 0\} = \cup_{n\in\mathbb{N}} A_n,$$

and $A_n \subset A_{n+1}$. This implies that $q = \mathbb{P}(\cup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} f^{(n)}(0)$. □

Theorem 4.4. Suppose that $p_1 < 1$. Then

1. q is the smallest root of the equation $f(s) = s$,
2. $q = 1$ if and only if $f'(1) \leq 1$.

Proof. Notice that f satisfies the following properties:

1. $f(1) = 1$,
2. $0 \leq f(s) \leq 1$ for every $s \in [0,1]$,
3. f and f' are continuous functions on $[0,1]$,
4. $f'(s) = \sum_{k\geq 1} kp_k s^{k-1} \geq 0$ for every $s \in [0,1]$ (f is increasing).

Let r be the smallest root of the equation $f(s) = s$. Let $s_0 = 0$ and $s_n = f^{(n)}(0)$. By the previous proposition, s_n converges to q, and by the continuity of f, $f(s_n)$ converges to $f(q)$. We have $s_{n+1} = f(s_n)$, hence $f(q) = q$, i.e. q is a root of the equation $f(s) = s$. To prove that q is the smallest root, observe that $f^{(n)}$ is an increasing function as a composition of increasing functions. Hence $f^{(n)}(0) \leq f^{(n)}(r) = r$, and taking the limit as n goes to infinity we obtain that $q \leq r$, implying that $q = r$.

For the second item of the theorem, let us assume that $f'(1) > 1$. Then there is some $\delta < 1$ such that $f'(s) > 1$ for every $s \in [\delta, 1]$. If $f(x) \geq x$ for some $x \in [\delta, 1)$, then $f(s) > s$ for every $s \in (x, 1]$, which contradicts the fact that $f(1) = 1$. Therefore, $f(s) < s$ for every $s \in [\delta, 1)$, and using the intermediate value theorem we obtain that $q < 1 - \delta < 1$.

Assume now that $f'(1) \leq 1$. If $p_0 + p_1 < 1$, then

$$f''(s) = \sum_{k\geq 1} k(k-1)p_k s^{k-2} > 0$$

for every $s > 0$, hence f' is strictly increasing for every $s > 0$. This shows that $f'(s) < f'(1) \leq 1$ for every $s \in [0,1)$. If $p_0 + p_1 = 1$, then $f(s) = p_0 + p_1 s$ and $f'(s) = p_1 < 1$ for every s. In both cases, $f(s) - s$ is a strictly decreasing function on the interval $[0,1]$. If there is some $x \in [0,1)$ such that $f(x) = x$, then $f(s) < s$ for every $s \in [x,1]$, contradicting that $f(1) = 1$. Hence no such x exists, which gives that $q = 1$. □

We remark that $f'(1) = \sum_{k=1}^{\infty} kp_k = \mathbb{E}(L)$. We will denote $\mathbb{E}(L)$ by μ. As a corollary of the previous theorem we obtain

Corollary 4.5. Assume that $p_1 < 1$. Then $q = 1$ if and only if $\mu \leq 1$.

Examples 4.2.
1. Flip a fair coin to decide whether to ever have another baby, and suppose all your offspring follow that strategy. Then $\mu = 1$, hence your genealogical tree eventually dies out almost surely.

2. If $p_1 = 1$, then $\mu = 1$ but $q = 0$. This shows that the condition of the theorem is important.

3. Consider percolation on a d-ary tree with parameter p. Then $\mu = d \cdot p$, hence $p_c = 1/d$.

Exercises

1. The **Poisson branching process** with intensity λ is defined as the Galton-Watson tree with the offspring distribution L defined by

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}.$$

   Find the generating function of L and the equation determining the extinction probability of the process.

5 Percolation

5.1 Definitions

Let $G = (V, E)$ be an infinite graph, and $p \in [0, 1]$. Let $\{\omega(e), e \in E\}$ be a family of i.i.d. Bernoulli(p) random variables. Intuitively, we construct a random subgraph of G by keeping an edge e if $\omega(e) = 1$ and deleting e if $\omega(e) = 0$.

For each p, this defines a probability space: $\Omega := \{0, 1\}^E$, $\mathcal{F}$ is generated by the sets $\Omega_{e=i} := \{\omega \in \Omega, \omega(e) = i\}$, where $i \in \{0, 1\}$. We have $\mathbb{P}_p(\Omega_{e=1}) = p$.

We will mostly concentrate on the case where $G = (\mathbb{Z}^d, \{(x, x \pm e_i)\})$ is the standard lattice in $\mathbb{R}^d$; that is, $V(G) := \{(x_1, \dots, x_d)\}$ is the set of d-dimensional vectors with integer coordinates, and $E(G)$ contains an edge xy whenever x and y differ in at most one coordinate and the differ by exactly 1 in that coordinate.

Theorem 5.1. For every $p \in [0, 1]$, there is a unique probability measure $\mathbb{P}_p$ on $(\Omega, \mathcal{F})$ such that for every two finite disjoint subsets $O, C \subset E$,

$$\mathbb{P}_p(\{\omega(e) = 1 \text{ for every } e \in O\} \text{ and } \{\omega(f) = 0 \text{ for every } f \in C\}) = p^{|O|}(1-p)^{|C|}.$$

Proof. Exercise using Caratheodory's Extension Theorem and Dynkin's Lemma. □

We call $(\Omega, \mathcal{F}, \mathbb{P}_p)$ the Bernoulli bond percolation model on G, with parameter p.

Site percolation: similar, except we delete vertices (and their incident edges) instead of edges.

Definition 5.1. Any $\omega \in \Omega$ is called a percolation instance, or configuration.

We say that e is open in ω if $\omega(e) = 1$, and closed if $\omega(e) = 0$.

The clusters of ω are the connected components spanned by its open edges. The cluster containing a vertex o is denoted by $C_o = C_o(\omega)$.

Theorem 5.2. For every graph $G = (V, E)$, and any $o \in V$, $\mathbb{P}_p(\{C_o \text{ is infinite}\})$ is monotone increasing in p.

Proof. We consider the following 'realisation' of $\mathbb{P}_p$. Let $\{u(e), e \in E\}$ be i.i.d. uniform random variables in $[0, 1]$. Set $\omega(e) = \omega_p(e) = 1$ whenever $u(e) \leq p$. Note that $\omega(e) = Bernoulli(p)$. The resulting random element ω of Ω has distribution $\mathbb{P}_p$ by Theorem 5.1. So we may as well consider it as the definition of $\mathbb{P}_p$. Let $p_1 < p_2$ in $[0, 1]$. Then $\omega_{p_1}(e) = 1$ implies that $\omega_{p_2}(e) = 1$ because $u(e) \leq p_1$ implies that $u(e) \leq p_2$. Thus whenever $\{C_o(\omega_{p_1})$ is infinite$\}$ happens, $\{C_o(\omega_{p_2})$ is infinite$\}$ happens as well, hence $\mathbb{P}_{p_1}(\{C_o \text{ is infinite}\}) \leq \mathbb{P}_{p_2}(\{C_o \text{ is infinite}\})$. □

Proposition 5.3. Let G be a graph, and $p \in [0, 1]$. The following are equivalent:

1. $\mathbb{P}_p(\{C_o \text{ is infinite}\}) = 0$ for every $v \in V$,
2. $\mathbb{P}_p(\{\text{there is an infinite cluster}\}) = 0$.

Proof. The backward direction is obvious. For the forward direction, notice that

$$\begin{aligned}\mathbb{P}_p(\{\text{there is an infinite cluster}\}) &= \mathbb{P}_p(\cup_{v \in V}\{C_v \text{ is infinite}\})\\ &\leq \sum_{v \in V} \mathbb{P}_p(\{C_v \text{ is infinite}\}) = 0\end{aligned}$$

by the union bound. □

5.2 The percolation threshold

Definition 5.2. We define the percolation threshold

$$p_c := \inf_{p \in [0,1]} \{\mathbb{P}_p(\exists \text{ an infinite cluster}) > 0\} = \{\sup_{p \in [0,1]} \mathbb{P}_p(\exists \text{ an infinite cluster}) = 0\}.$$

As we will see, $\mathbb{P}_p(\exists \text{ an infinite cluster})$ is either 0 or 1.

Definition 5.3. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, and let G_k be the σ-algebra generated by the subsequence $X_k, X_{k+1}, \ldots$ If $A \in \cap_{k=1}^{\infty} G_k$, then we say that A is a tail event.

Theorem 5.4. (Kolmogorov's 0-1 law) Let $X_1, X_2, \ldots$ be a sequence of independent random variables, and let A be a tail event. Then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Lemma 5.5. (Konig's lemma) Let G be an infinite, connected, locally finite graph. Then G contains an infinite path.

Proposition 5.6. $\mathbb{P}_p(\exists$ an infinite cluster$)$ is either 0 or 1.

Proof. We claim that $\{\exists$ an infinite cluster$\}$ is a tail event with respect to the sequence $\{\omega(e), e \in E\}$. Indeed, we can use Konig's lemma to deduce that the event $\{\exists$ an infinite cluster$\}$ happens in ω if and only if there is an infinite open path in ω. The existence of an infinite open path is independent of the state of any finite collection of edges (exercise), which shows that $\{\exists$ an infinite cluster$\}$ is a tail event. Applying Kolmogorov's 0-1 law we conclude that $\mathbb{P}_p(\exists$ an infinite cluster$)$ is either 0 or 1. □

Remark: We assumed G to be locally finite here but it is not really needed.

Another important concept is the percolation density

$$\theta(p) := \mathbb{P}_p(C_o \text{ is infinite}).$$

From the monotonicity of $\mathbb{P}_p(\exists$ an infinite cluster$)$ we have that $\theta(p) = 0$ for every $p < p_c$, and $\theta(p) > 0$ for every $p > p_c$. Understanding the behaviour of θ at p_c is a notoriously hard problem. The following is still open despite the efforts of many experts:

Conjecture. θ is continuous at p_c on $\mathbb{Z}^3$.

5.3 Bounding p_c

We will now focus on $\mathbb{Z}^2$, and prove that $p_c \in (0, 1)$, i.e. a phase transition occurs.

Definition 5.4. A self-avoiding walk (SAW) is a path, i.e. a walk where no vertex is visited twice. (Which of the two words is used is a cultural matter; in statistical mechanics people prefer *SAW*, in graph theory *path* is standard.)

Proposition 5.7. $p_c(\mathbb{Z}^2) \geq 1/3$.

Proof. Let N_n denote the number of open SAWs with n edges in $\mathbb{Z}^2$ containing o. The number of SAWs of size n in $\mathbb{Z}^2$ containing o is at most $4 \cdot 3^{n-1}$ because there are four choices for the first step and at most three choices for every other step. Hence

$$\mathbb{E}_p(N_n) \leq 4 \cdot 3^{n-1} p^n.$$

On the event $\{C_o \text{ is infinite}\}$ there is an infinite SAW starting from o, and so for every n, there is a SAW of size n starting from o. Hence

$$\mathbb{P}_p(C_o \text{ is infinite}) \leq \mathbb{P}_p(\exists \text{ an open SAW of size n}) \leq \mathbb{E}(N_n)$$

for every positive integer n. If $p < 1/3$, then $\mathbb{E}(N_n)$ converges to 0, implying that $\mathbb{P}_p(C_o \text{ is infinite}) = 0$. Thus, $p_c(\mathbb{Z}^2) \geq 1/3$. □

Our next aim is to prove an upper bound for $p_c(\mathbb{Z}^2)$ that is smaller than 1. First, we need to introduce some new notions.

A planar graph G is a graph that can be embedded in the plane $\mathbb{R}^2$, i.e. it can be drawn in such a way that no edges cross each other. Such an embedding is called a planar embedding of the graph. A plane graph is a (planar) graph endowed with a fixed planar embedding.

A plane graph $G \subset \mathbb{R}^2$ divides the plane into regions called faces. More precisely, the faces of G are the connected components (as defined in any topology textbook) of $\mathbb{R}^2 \backslash G$; here the fixed embedding of G has allowed as to think of G as a subset of the plane $\mathbb{R}^2$. Using the faces of G we define its dual graph G^* as follows. The vertices of G^* are the faces of G, and we connect two vertices of G^* with an edge whenever the corresponding faces of G share an edge. Thus there is a bijection $e \mapsto e^*$ from $E(G)$ to $E(G^*)$. Notice that G^* can be embedded in the plane in such a way that an edge $f \in E(G^*)$ intersects an edge $e \in E(G)$ if and only if $f = e^*$.

Consider a plane graph G and a finite subgraph H of G. Let ∂H be the set of edges in $E(G) \setminus E(H)$, at least one endvertex of which lies in $V(H)$. The minimal cut of H is the minimal set of edges in ∂H, the removal of which disconnects H from infinity.

Lemma 5.8. For every finite minimal cut F of G, there is a cycle in G^* the edge set of which is F^*.

Using the above lemma we can easily find an upper bound for $p_c(\mathbb{Z}^2)$.

Proposition 5.9. $p_c(\mathbb{Z}^2) \leq 2/3$.

Proof. Let ω be a percolation configuration such that $C_o(\omega)$ is finite. Then there is a minimal cut F separating $C_o(\omega)$ from infinity, i.e. $C_o(\omega)$ is contained in a finite component of $\mathbb{Z}^2 \setminus F$. The union bound implies that

$$\mathbb{P}_p(C_o \text{ finite}) \leq \sum \mathbb{P}_p(\text{F is closed}),$$

where the sum ranges over all minimal edge cuts F of o. Notice that the dual $\mathbb{Z}^{2*}$ of $\mathbb{Z}^2$ can be identified with $\mathbb{Z}^2 + (1/2, 1/2)$, the translation of $\mathbb{Z}^2$ by the vector $(1/2, 1/2)$. By the above lemma, the latter sum is equal to $\sum \mathbb{P}_p(\text{C is closed})$, where now we are summing over all cycles in $\mathbb{Z}^{2*}$ that contain o in their interior. We can now express $\sum \mathbb{P}_p(\text{C is closed})$ via $\sum_{n=1}^{\infty} a_n (1-p)^n$, where a_n is the number of cycles in $\mathbb{Z}^{2*}$ of size n that contain o in their interior. Any such cycle contains some vertex of the form $(k + 1/2, 1/2)$ for some k satisfying $0 \leq k < n$. The number of cycles of size n in $\mathbb{Z}^{2*}$ containing a fixed vertex, is at most $4 \cdot 3^{n-1}$, hence $a_n \leq 4n3^{n-1}$. We can now conclude that

$$\mathbb{P}_p(C_o \text{ finite}) \leq \sum_{n=1}^{\infty} 4n3^{n-1}(1-p)^n.$$

This implies that $\mathbb{P}_p(C_o \text{ finite}) < 1$ when p is large enough, hence $p_c < 1$.

The above argument is called *Peierls' argument*. Next, we will refine it a bit to obtain the desired $p_c \leq 2/3$.

Let $p > 2/3$ and consider some n_0 such that

$$q := \sum_{n=n_0}^{\infty} 4n3^{n-1}(1-p)^n < 1.$$

Let $\Lambda = \Lambda(n_0)$ be the box $o + \{-n_0, \dots, n_0\}^2$. Notice that any minimal cut outside of Λ contains at least n_0 edges. Define the events
$A := \{\text{all edges in } \Lambda \text{ are open}\}$ and
$B := \{\text{no minimal cut outside } \Lambda \text{ is closed}\}$.

The event B is independent of the state of the edges in $E(\Lambda)$, hence A and B are independent. Moreover, $\mathbb{P}_p(B) \geq 1 - q > 0$ by Peierls' argument. Thus

$$\mathbb{P}_p(A \cap B) = \mathbb{P}_p(A)\mathbb{P}_p(B) > 0.$$

On the other hand, when both A and B occur in ω, then $C_o(\omega)$ is infinite. Therefore,

$$\mathbb{P}_p(C_o \text{ is infinite}) \geq \mathbb{P}_p(A \cap B) > 0.$$

Since this holds for every $p > 2/3$, we conclude that $p_c(\mathbb{Z}^2) \leq 2/3$. □

The following propositions show that not every graph G undergoes a genuine phase transition, i.e. $p_c(G)$ may equal 0 or 1.

Proposition 5.10. $p_c(\mathbb{Z}) = 1$.

Proof. Percolation on $\mathbb{Z}$ is identically distributed with the union of two Galton-Watson trees with $L = Bernoulli(p)$, which die out when $p < 1$. □

We will now show that there are graphs with $p_c = 0$.

Proposition 5.11. There is a graph G with $p_c(G) = 0$.

Proof. Let G be a tree with root o such that every vertex at distance n from o has $n + 1$ children. Then the d-ary tree T_d is a subgraph of G for every d. Since $p_c(T_d) = \frac{1}{d-1}$, we have $p_c(G) \leq p_c(T_d) = \frac{1}{d-1}$ for every d. Thus, $p_c(G) = 0$. □

5.4 The exponential decay threshold p_D

In this section we consider percolation on $\mathbb{Z}^d$. Let $o = (0, \ldots, 0)$ denote the origin. The *box* Λ_n is the subgraph of $\mathbb{Z}^d$ spanned by the vertices all coordinates of which lie in the interval $[-n, n]$. We let $B_o(n)$ denote the ball of radius n around o in $\mathbb{Z}^d$, i.e. the subgraph spanned by the vertices at graph-distance at most n from o. For a subgraph X of $\mathbb{Z}^d$, the boundary ∂X of X comprises the edges of $\mathbb{Z}^d$ with exactly one end-vertex in X.

Definition 5.5. The susceptibility $\chi(p)$ is defined by

$$\chi(p) := \mathbb{E}_p(|C(o)|).$$

The susceptibility threshold is defined by

$$p_\infty := \sup_p \{\chi(p) < \infty\} = \inf_p \{\chi(p) = \infty\}.$$

We write $\{o \leftrightarrow Y\}$ for the event that the cluster of o contains an element of a set Y.

Definition 5.6. The exponential decay threshold is defined by

$$p_D := \sup_p\{\exists c_p < 1 \text{ such that } \mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) < c_p^n\} =$$
$$\sup_p\{\exists c_p < 1 \text{ such that } \mathbb{P}_p(o \leftrightarrow \partial B_o(n)) < c_p^n\}.$$

Proposition 5.12. $p_D \leq p_\infty \leq p_c$.

Proof. If $p < p_D$ then
$\chi(p) = \sum_n \sum_{x\in V, d(x,o)=n} \mathbb{P}_p(x \in C(o)) \leq \sum_n Cn^{d-1}c_p^n < \infty$. Thus $p \leq p_\infty$. Here, we used the fact that $\chi(p) = \sum_{x\in V} \mathbb{E}_p(\mathbb{1}_{x\in C(o)}) = \sum_{x\in V} \mathbb{P}_p(x \in C(o))$ (Exercise), and that $\partial B_o(n)$ has at most Cn^{d-1} elements for some constant C.

If $p > p_c$ then $\mathbb{P}_p(|C(o)| = \infty) = \epsilon > 0$. Hence
$\mathbb{E}_p(|C(o)|) \geq \epsilon\infty = \infty$, and so $p_c \geq p_\infty$.

□

In fact, we will later prove that $p_D = p_\infty = p_c$, but the proof is much more involved. Using this statement, we now tighten our upper bound on $p_c(\mathbb{Z}^2)$:

Theorem 5.13 (Kesten's theorem)**.** $p_c(\mathbb{Z}^2) \leq 1/2$.

Proof. Sample a percolation instance ω with $p = 1/2$. Let B denote the set of open edges, and R the set of closed edges. Note that the dual R^* of R has the same law $\mathbb{P}_{1/2}$ as B.

Suppose that $p_c(\mathbb{Z}^2) > 1/2$. Then both B, R^* are subcritical, and for every $x \in \mathbb{Z}^2$, $\mathbb{P}_{1/2}(|C(x)| < \infty) = 1$. Since the dual of the minimal cut of the cluster of o is a cycle, there is a cycle D_0 in R^* surrounding o. But the cluster in $(\mathbb{Z}^2)^*$ containing D_0 is finite, hence there is a cycle D_1 in B surrounding D_0, and thus o. Continuing in this manner, we find almost surely a sequence of cycles (D_i) surrounding o, such that D_{2i} lies in R^* and D_{2i+1} lies in B.

Let H be the infinite horizontal path $o + \{0, 1, \ldots\} \times \{0\}$ in $\mathbb{Z}^2$, and let h_n denote the nth vertex of H. Let $X := |\{n : h_n \text{ lies in an open cycle surrounding } o\}|$. Every cycle D_{2i+1} intersects H and surrounds o, hence

$$\mathbb{E}_{1/2}(X) = \infty.$$

On the other hand, assuming $p_D = p_c$ (which is proved below), there is a constant $0 < c = c_{1/2} < 1$ such that for every vertex $x \in \mathbb{Z}^2$, $\mathbb{P}_{1/2}(x \leftrightarrow x + \Lambda_n) \leq c^n$. This implies that

$$\mathbb{P}_{1/2}(h_n \text{ lies in an open cycle surrounding o}) \leq c^n.$$

Therefore,

$$\mathbb{E}_{1/2}(X) < \infty,$$

which leads to a contradiction. □

5.5 An Ergodic Theorem

For $z \in \mathbb{Z}^d$, let $\tau_z(\omega) := \omega(e + z)$ be the shift of ω by z.

Definition 5.7. An event A is shift-invariant, if $\omega \in A$ implies $\tau_z(\omega) \in A$ for every $z \in \mathbb{Z}^d$. In other words, $\tau_z(A) = A$.

Examples 5.1. Both {there is an infinite cluster} and {# of infinite clusters is k} are shift-invariant events. On the other hand, {$C(o)$ is infinite} (or any event talking about $C(o)$) is not shift-invariant.

Theorem 5.14. If A is shift-invariant then for every $p \in [0, 1]$, $\mathbb{P}_p(A)$ is either 0 or 1.

Corollary 5.15. For Bernoulli bond percolation on $\mathbb{Z}^d$ and for every $p \in [0, 1]$, there is $k = k(p) \in \mathbb{N} \cup \{\infty\}$ such that $\mathbb{P}_p(N = k) = 1$, where $N = N(\omega) := \#\{\text{infinite clusters of } \omega\}$.

Proof. For every $k \in \mathbb{N} \cup \{\infty\}$, the event $\{N = k\}$ is a shift-invariant event. By the above theorem $\mathbb{P}_p(N = k) \in \{0, 1\}$. As $\sum_{k\in\mathbb{N}\cup\{\infty\}} \mathbb{P}_p(N = k) = 1$, we have $\mathbb{P}_p(N = k) = 1$ for a unique $k = k(p)$. □

Proof of Theorem. Let $p \in [0, 1]$ and $\varepsilon > 0$. We claim that there is $n \in \mathbb{N}$ and an event B depending on the edges in Λ_n only, such that $\mathbb{P}_p(A \triangle B) < \varepsilon$, where $A \triangle B := (A \cap B^c) \cup (A^c \cap B)$ denotes the symmetric difference of A and B. This is because the events B of this form generate our σ-algebra.

Pick such an event B and $z \in \mathbb{Z}^d \setminus \Lambda_{3n}$. Since A is shift invariant, $\mathbb{P}_p\big(\tau_z(A)\big) = \mathbb{P}_p(A) = \mathbb{P}_p\big(A \cap \tau_z(A)\big)$. The boxes $\Lambda_n(z)$ and $\Lambda_n(o)$ are disjoint, which implies that B and $\tau_z(B)$ are independent events. We will estimate the difference $\mathbb{P}_p(A) - \mathbb{P}_p(A)^2$. We have

$$
\begin{aligned}
|\mathbb{P}_p(A) - \mathbb{P}_p(A)^2| &= |\mathbb{P}_p\big(A \cap \tau_z(A)\big) - \mathbb{P}_p(A)^2| \\
\leq |\mathbb{P}_p\big(A \cap \tau_z(A)\big) - \mathbb{P}_p\big(B \cap \tau_z(A)\big)| &+ |\mathbb{P}_p\big(B \cap \tau_z(A)\big) - \mathbb{P}_p\big(B \cap \tau_z(B)\big)| \\
&+ |\mathbb{P}_p\big(B \cap \tau_z(B)\big) - \mathbb{P}_p(B)^2| + |\mathbb{P}_p(B)^2 - \mathbb{P}_p(A)^2|
\end{aligned}
$$

The independence of B and $\tau_z(B)$ gives that

$$|\mathbb{P}_p\big(B \cap \tau_z(B)\big) - \mathbb{P}_p(B)^2| = 0.$$

Moreover,

$$|\mathbb{P}_p(B)^2 - \mathbb{P}_p(A)^2| = |\big(\mathbb{P}_p(B) - \mathbb{P}_p(A)\big)\big(\mathbb{P}_p(B) + \mathbb{P}_p(A)\big) < 2\varepsilon.$$

Notice that

$$|\mathbb{P}_p\big(A \cap \tau_z(A)\big) - \mathbb{P}_p\big(B \cap \tau_z(A)\big)| \leq \mathbb{P}_p(A \triangle B) < \varepsilon$$

and

$$|\mathbb{P}_p\big(B \cap \tau_z(A)\big) - \mathbb{P}_p\big(B \cap \tau_z(B)\big)| \leq \mathbb{P}_p\big(\tau(A) \triangle \tau(B)\big) < \varepsilon,$$

which follow from the fact that for every events C_1, C_2 and D

$$|\mathbb{P}_p(C_1 \cap D) - \mathbb{P}_p(C_2 \cap D)| \leq \mathbb{P}_p\big((C_1 \cap D) \triangle (C_2 \cap D)\big) \leq \mathbb{P}_p(C_1 \triangle C_2).$$

Therefore,

$$|\mathbb{P}_p(A) - \mathbb{P}_p(A)^2| < 4\varepsilon.$$

Since ε was arbitrary, we get that $\mathbb{P}_p(A) = \mathbb{P}_p(A)^2$, showing that $\mathbb{P}_p(A)$ is either 0 or 1. $\square$

Recall that for every $p \in [0,1]$ there exists a unique $k \in \mathbb{N} \cup \{\infty\}$ such that $\mathbb{P}_p(N = k) = 1$.

Lemma 5.16. For percolation on $\mathbb{Z}^d$, if $\mathbb{P}_p(N = k) > 0$ for some $2 \leq k < \infty$, then $\mathbb{P}_p(N = 1) > 0$.

Proof. Let

$$A_n = \{N > 0 \text{ and } \Lambda_n \text{ intersects all infinite clusters}\}, \text{ and}$$

$$B_n = \{\text{all edges in } \Lambda_n \text{ are open}\}.$$

(Exercise: Show that A_n is a measurable event.) There is $n > 0$ such that $\mathbb{P}_p(A_n) > 0$ because

$$\{0 < N < \infty\} = \cup_{n=1}^{\infty} A_n$$

and $\mathbb{P}_p(0 < N < \infty) > 0$. Notice that if A_n occurs and we open all edges of Λ_n, then all infinite components are merged into one, hence

$$\mathbb{P}_p(N = 1) \geq \mathbb{P}_p(A_n \cap B_n).$$

But A_n and B_n are independent events. Thus

$$\mathbb{P}_p(N = 1) \geq \mathbb{P}_p(A_n)\mathbb{P}_p(B_n) > 0.$$

□

The above lemma combined with Corollary 5.15 implies that

Corollary 5.17. For every $p \in [0, 1]$, exactly one of

$$\{N = 0\}, \{N = 1\}, \{N = \infty\}$$

occurs almost surely.

5.6 Finiteness of the number of infinite clusters

We will now exclude the possibility $\{N = \infty\}$ of Corollary 5.17.

Definition 5.8. We say that $x \in \mathbb{Z}^d$ is a trifurcation point in ω, if $|C(x)| = \infty$, and $C(x) \setminus \{x\}$ has at least 3 infinite clusters. Let $\mathcal{T}(\omega)$ be the set of trifurcation points in ω.

Theorem 5.18 (Burton-Keane)**.** For percolation on $\mathbb{Z}^d$, $\mathbb{P}_p(N = \infty) = 0$ for every $p \in [0, 1]$.

Proof. Assume that $p \in (0, 1)$ (otherwise trivial). We claim that if $\mathbb{P}_p(N = \infty) = 1$, then $\mathbb{P}_p(o \in \mathcal{T}(\omega)) > 0$. Indeed, choose n large enough that

$$\mathbb{P}_p(\exists \text{ a,b,c } \in \partial\Lambda_n \text{ belonging to distinct infinite clusters}) > 1/2.$$

This implies that

$\mathbb{P}_p(\exists$ a',b',c' $\in \partial\Lambda_n$ belonging to distinct infinite clusters in $\mathbb{Z}^d \setminus \Lambda_n) > 1/2$.

The size of $\partial\Lambda_n$ is of order n^{d-1}, hence the is a constant $c > 0$ such that the number of triples (a', b, c') as above is at most $cn^{3(d-1)}$. The union bound implies that there exist $a', b', c' \in \partial\Lambda_n$ such that

$$\mathbb{P}_p(a', b', c' \text{ belong to distinct infinite clusters in } \mathbb{Z}^d \setminus \Lambda_n) > \frac{1}{2cn^{3(d-1)}}.$$

(the estimate $\frac{1}{2cn^{3(d-1)}}$ is more precise than we actually need; what matters is that it is strictly positive, and that a', b', c' are now fixed vertices of Λ_n.)

Fixing these $a', b', c' \in \partial\Lambda_n$, it is not hard to see that there are 3 edge-disjoint paths in Λ_n connecting o to each of a', b', c'. If we open the edges of all these paths and close every other edge of Λ_n, then o becomes a trifurcation point. But the state of the edges inside Λ_n is independent of the event

$$a', b', c' \text{ belong to distinct infinite clusters in } \mathbb{Z}^d \setminus \Lambda_n.$$

Hence $\mathbb{P}_p(o \in \mathcal{T}(\omega)) > \epsilon$ for some $\epsilon = \epsilon(p) > 0$, which proves our claim.

Let from now on $n \in \mathbb{N}$ be arbitrary, and let $T_n := |\mathcal{T}(\omega) \cap \Lambda_n|$. We have $\mathbb{P}_p(x \in \mathcal{T}(\omega)) = \mathbb{P}_p(o \in \mathcal{T}(\omega))$ for every $x \in \mathbb{Z}^d$. By the linearity of expectation,

$$\mathbb{E}_p(T_n) = \mathbb{P}_p(o \in \mathcal{T}(\omega))|\Lambda_n|. \tag{5}$$

Our second claim is that

$$T_n \leq |\partial\Lambda_n|.$$

This will imply that

$$\mathbb{E}_p(T_n) \leq |\partial\Lambda_n|,$$

hence $\mathbb{E}_p(T_n)$ is of order at most n^{d-1}. But by (5), $\mathbb{E}_p(T_n)$ is of order n^d. Choosing n large enough we derive a contradiction. In order to prove our claim we will use the following lemma, the proof of which is an easy exercise in graph theory.

Lemma 5.19. Let T be a finite tree (or forest), let $L = L(T)$ be the number of leaves of T, and let $K = K(T)$ be the number of vertices of degree at least 3. Then $L \geq K + 2$.

Let F be a forest comprising a spanning tree of each cluster of Λ_n. Let $L \subset F$ be maximal such that each leaf of L lies in $\partial\Lambda$. Let $Y \subset L$ comprise the components of L having at least 3 leaves in $\partial\Lambda_n$. Note that all trifurcation points $\mathcal{T}(\omega)$ remain trifurcation points when we change ω inside Λ_n into Y, i.e. letting

$$\omega'(e) := \begin{cases} 1 & e \in E(Y) \\ 0 & e \in E(\Lambda_n) \setminus E(Y) \\ \omega(e) & e \notin E(\Lambda_n) \end{cases}$$

we have

$$\mathcal{T}(\omega') \cap \Lambda_n = \mathcal{T}(\omega) \cap \Lambda_n.$$

Notice that every trifurcation point of $\mathcal{T}(\omega') \cap \Lambda_n$ has degree at least 3 in Y, hence $K(Y) \geq T_n$. Moreover, all the leaves of Y lie in $\partial\Lambda_n$. Using Lemma 5.19 we obtain

$$K(Y) < L(Y) \leq |\partial\Lambda_n|.$$

Therefore

$$T_n < |\partial\Lambda_n|,$$

as desired. □

5.7 Harris' theorem

In this section we will complete the proof that $p_c(\mathbb{Z}^2) = 1/2$.

We start with a nice general lemma that we will use without proof (see e.g. the book of Lyons & Peres if you are curious). A percolation event A is said to be increasing, if $\omega \in A$ and $\omega \subseteq \omega'$ imply $\omega' \in A$. (Intuitively, adding more open edges can only help A occur.)

Lemma 5.20 (Harris' inequality)**.** If A, B are increasing events of percolation on a graph then, for every $p \in [0, 1]$, we have

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

Using this we can prove

Lemma 5.21 (The square root trick)**.** Let A_1 and A_2 be increasing events. If $A = A_1 \cup A_2$ and $\mathbb{P}_p(A_1) = \mathbb{P}_p(A_2)$, then $\mathbb{P}_p(A_1) \geq 1 - (1 - \mathbb{P}_p(A))^{1/2}$.

Proof. We have $(1-\mathbb{P}_p(A_1))^2 = \mathbb{P}_p(A_1^c)^2 = \mathbb{P}_p(A_1^c)\mathbb{P}_p(A_2^c)$. By Harris' inequality, this is at most $\mathbb{P}_p(A_1^c \cap A_2^c) = \mathbb{P}_p(A^c) = 1-\mathbb{P}_p(A)$. Taking the square root of both sides and rearranging we deduce $\mathbb{P}_p(A_1) \geq 1-(1-\mathbb{P}_p(A))^{1/2}$. □

Theorem 5.22 (Harris' theorem)**.** On $\mathbb{Z}^2$, $\mathbb{P}_{1/2}(\exists \text{ an infinite cluster}) = 0$.

Proof. Suppose not. Then

$$\mathbb{P}_{1/2}(\exists \text{ an infinite cluster}) = 1. \tag{6}$$

Define the event

$$A_n^l := \{\text{the left side of } \partial\Lambda_n \text{ meets an infinite cluster in } \mathbb{Z}^d \setminus \Lambda_n\}.$$

Similarly, we define A_n^r (for right), A_n^t (for top), A_n^b (for bottom). Then (6) implies that $\mathbb{P}_p(A_n^l \cup A_n^r \cup A_n^t \cup A_n^b)$ converges to 1 as n goes to ∞. Note that $\mathbb{P}_p(A_n^l) = \mathbb{P}_p(A_n^r) = \mathbb{P}_p(A_n^t) = \mathbb{P}_p(A_n^b)$ by symmetry, and A_n^l is an increasing event. Using the square root trick from above (adapted to 4 events), we can deduce that $\mathbb{P}_p(A_n^u)$ converges to 1 as n goes to ∞ for every $u = l, r, t, b$.

Thus we can choose m large enough that $\mathbb{P}_{1/2}(A_m^u) > 7/8$. Define A_n^{u*} like A_n^u but in the dual $\mathbb{Z}^{2*}$. Since $\mathbb{Z}^{2*}$ is isomorphic to $\mathbb{Z}^2$, we have $\mathbb{P}_{1/2}({A_m^u}^*) > 7/8$ as well. We couple percolation on the primal and percolation on the dual by defining $\omega^*(e^*) = 1-\omega(e)$. Let $A := A_m^l \cap A_m^r \cap {A_m^t}^* \cap {A_m^b}^*$. The union bound shows that

$$\mathbb{P}_{1/2}(A^c) < 1/2,$$

hence

$$\mathbb{P}_{1/2}(A) > 1/2.$$

If A occurs, then both the left and the right side of Λ_n meet infinite clusters of ω, and the top and bottom side meet infinite clusters of ω^*. But the Burton-Keane theorem tells us that both ω, ω^* have a unique infinite cluster. Thus there is an open path in ω from the left side of $\partial\Lambda_n$ to the right, and an open path in ω^* from the top side of $\partial\Lambda_n$ to the bottom. These two paths must intersect at an edge e due to topological reasons. This leads to a contradiction, as e must then be open in both ω, ω^*, which is impossible as we have defined $\omega^* := 1-\omega$. □

Combining Theorems 5.13 and 5.22, we obtain

Theorem 5.23 (Harris–Kesten theorem)**.** $p_c(\mathbb{Z}^2) = 1/2$.

Exercises

1. Let L denote the cubic lattice, i.e. the graph with vertex set $\mathbb{Z}^3$ in which two vertices (x, y, z) and (x', y', z') are joined with an edge when $|x - x'| + |y - y'| + |z - z'| = 1$. Thus every vertex has degree 6. Using a coupling with an appropriate Galton-Watson tree, prove that $p_c(L) \geq 1/5$.

2. Prove that if the events A, B are both increasing, then so is $A \cap B$. Use this to extend Harris' inequality to more than 2 increasing events.

3. Extend the square root trick to more than two events.

4. Let T be a Galton–Watson tree with offspring distribution given by a random variable L with expectation $\mu := \mathbb{E}(L)$. Colour each edge of T blue with probability p, and red with probability $1 - p$, independently of all other experiments. For which values of p do we have

$$\mathbb{P}(\text{ the blue subtree of } T \text{ containing the origin is infinite }) > 0?$$

THE END?

The remainder of these lecture notes is optional material, not examinable.

5.8 Proof of $p_D = p_c$

Our final goal is to show that $p_D = p_c$ on $\mathbb{Z}^d$, completing the proof of Kesten's theorem. Let us start by fixing some notation.

Given a finite set of vertices on $\mathbb{Z}^d$, we denote its edge-boundary by ΔS, defined by all the edges xy with $x \in S$ and $y \notin S$. Two vertices x and y are *connected in* $S \subset \mathbb{Z}^d$ if there exists an open path from x to y with vertices in S. We denote this event by $x \overset{S}{\leftrightarrow} y$.

For $p \in [0,1]$ and finite $o \in S \subset \mathbb{Z}^d$, define

$$\varphi_p(S) := p \sum_{xy \in \Delta S} \mathbb{P}_p[o \overset{S}{\leftrightarrow} x]$$

and

$$\tilde{p}_c := \sup \big\{ p \in [0,1] \text{ s.t. there exists a finite set } o \subset S \subset \mathbb{Z}^d \text{ with } \varphi_p(S) < 1 \big\}.$$

Theorem 5.24 (Duminil-Copin & Tassion)**.** For any $d \geq 2$, $\tilde{p}_c = p_c$. Moreover,

1. For $p < \tilde{p}_c$, there exists $0 < c = c(p) < 1$ such that for every $n \geq 1$,

$$\mathbb{P}_p[o \leftrightarrow \partial \Lambda_n] \leq c^n.$$

2. For $p > \tilde{p}_c$,

$$\mathbb{P}_p[C(o) = \infty] \geq \frac{p - p_c}{p(1 - p_c)} > 0.$$

Proof. Note that (1) and (2) imply that $\tilde{p}_c = p_c$.

Let $p < \tilde{p}_c$. Fix a finite set S containing o, such that $\varphi_p(S) < 1$. Let $L > 0$ such that $S \subset \Lambda_{L-1}$.

Since $S \cap \partial \Lambda_{kL} = \emptyset$, if the event $o \leftrightarrow \partial \Lambda_{kL}$ occurs for some $k \geq 1$, then there is an edge $xy \in \Delta S$ such that:

- o is connected to x in S,
- xy is open,

- y is connected to $\partial\Lambda_{kL}$ in $\mathcal{C}^c$.

Indeed, pick an open path from o to $\partial\Lambda_{kL}$ and let xy be the last edge in ∂S it traverses.

We will bound the probability that such xy exists. Let

$$\mathcal{C} = \{z \in S : o \overset{S}{\leftrightarrow} z\}.$$

Using first the union bound, and then a decomposition with respect to possible values of $\mathcal{C}$, we find

$$\begin{aligned}&\mathbb{P}_p(o \leftrightarrow \partial\Lambda_{kL})\\ &\le \sum_{xy\in\Delta S}\sum_{C\subset S}\mathbb{P}_p\big(\{o \overset{S}{\leftrightarrow} x, \mathcal{C} = C\} \cap \{xy \text{ is open}\} \cap \{y \overset{\mathbb{Z}^d\setminus C}{\leftrightarrow} \partial\Lambda_{kL}\}\big)\end{aligned}$$

Using the fact that the three events depend on different sets of edges and are therefore independent we have that

$$\mathbb{P}_p(o \leftrightarrow \partial\Lambda_{kL}) \le p\sum_{xy\in\Delta S}\sum_{C\subset S}\mathbb{P}_p\big(o \overset{S}{\leftrightarrow} x, \mathcal{C} = C\big)\mathbb{P}_p\big(y \overset{\mathbb{Z}^d\setminus C}{\leftrightarrow} \partial\Lambda_{kL}\big).$$

Since $y \in \Lambda_L$, we can bound $\mathbb{P}_p(y \overset{\mathbb{Z}^d\setminus C}{\leftrightarrow} \partial\Lambda_{kL})$ by $\mathbb{P}_p(o \leftrightarrow \partial\Lambda_{(k-1)L})$ in the last expression. Moreover,

$$\sum_{C\subset S}\mathbb{P}_p\big(o \overset{S}{\leftrightarrow} x, \mathcal{C} = C\big) = \mathbb{P}_p(o \overset{S}{\leftrightarrow} x).$$

Hence, we get

$$\mathbb{P}_p(o \leftrightarrow \partial\Lambda_{kL}) \le \varphi_p(S)\mathbb{P}_p\big(y \leftrightarrow \partial\Lambda_{(k-1)L}\big).$$

which by induction gives

$$\mathbb{P}_p(o \leftrightarrow \partial\Lambda_{kL}) \le \varphi_p(S)^{k-1}.$$

This proves the desired exponential decay.

We will prove the second item of Theorem 5.24 by providing a differential inequality valid for every p.

Lemma 5.25. Let $p \in [0, 1]$ and $n \ge 1$,

$$\frac{d}{dp}\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) \ge \frac{1}{p(1-p)} \cdot \inf_{\substack{S\subset\Lambda_n\\ o\in S}} \varphi_p(S) \cdot \big(1 - \mathbb{P}_p(o \leftrightarrow \partial\Lambda_n)\big). \qquad (7)$$

Remark. As Λ_n is finite, $\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n)$ is a polynomial in p, hence differentiable.

Remark. The above lemma is reminiscent of

Lemma 5.26 (Russo's formula)**.** For every finite percolation model and for every increasing event A,

$$\frac{d}{dp}\mathbb{P}_p(A) = \mathbb{E}_p(\# \text{ of pivotal edges for A}) = \sum_e \mathbb{P}_p(e \text{ is pivotal for a}),$$

where an edge xy is pivotal for the event A if $\omega_{xy} \notin A$ and $\omega^{xy} \in A$. (The configuration ω_{xy}, respectively ω^{xy}, coincides with ω except that the edge xy is closed, respectively open.)

Let us first see how (7) implies the second item of Theorem 5.24. Setting $g(p) = \mathbb{P}_p(o \leftrightarrow \partial\Lambda_n)$ and using that $\inf_{\substack{S\subset\Lambda_n \\ o\in S}} \varphi_p(S) \geq 1$ for every $p \geq \tilde{p}_c$, we get that

$$\frac{g'(p)}{1-g(p)} \geq \frac{1}{p(1-p)}.$$

Notice that

$$\frac{g'(p)}{1-g(p)}$$

is the derivative of

$$\log\big(\frac{1}{1-g(p)}\big),$$

while

$$\frac{1}{p(1-p)}$$

is the derivative of

$$\log(\frac{p}{1-p}).$$

Integrating the differential inequality between $\tilde{p}_c$ and $p > \tilde{p}_c$ implies that for every $n \geq 1$,

$$\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) \geq \frac{p-\tilde{p}_c}{p(1-\tilde{p}_c)}.$$

Letting n go to infinity, we obtain the desired lower bound on $\mathbb{P}_p(o \leftrightarrow \infty) = \mathbb{P}_p(|C(o)| = \infty)$.

We now prove (7). By Russo's formula and the fact that the state of a pivotal edge for an event is independent of the event, we have

$$\begin{aligned}\frac{d}{dp}\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) &= \sum_{e\subset\Lambda_n} \mathbb{P}_p\big(e \text{ is pivotal}\big)\\ &= \frac{1}{1-p}\sum_{e\subset\Lambda_n} \mathbb{P}_p(e \text{ is pivotal}, e \text{ is closed})\\ &= \frac{1}{1-p}\sum_{e\subset\Lambda_n} \mathbb{P}_p\big(e \text{ is pivotal}, o \not\leftrightarrow \partial\Lambda_n\big).\end{aligned}$$

Let

$$\mathcal{S} := \{x \in \Lambda_n \text{ such that } x \not\leftrightarrow \partial\Lambda_n\}.$$

Summing over the possible values for $\mathcal{S}$ we obtain

$$\frac{d}{dp}\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) = \frac{1}{1-p}\sum_{\substack{S\subset\Lambda_n\\ o\in S}}\sum_{e\subset\Lambda_n} \mathbb{P}_p\big(e \text{ is pivotal}, \mathcal{S} = S\big)$$

Notice that on the event $\mathcal{S} = S$, the pivotal edges are the edges $xy \in \Delta S$ such that o is connected to x in S. This implies that

$$\frac{d}{dp}\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) = \frac{1}{1-p}\sum_{\substack{S\subset\Lambda_n\\ o\in S}}\sum_{xy\in\Delta S} \mathbb{P}_p\big(o \overset{S}{\leftrightarrow} x, \mathcal{S} = S\big).$$

The event $\{\mathcal{S} = S\}$ depends only on the state of the edges outside S and is therefore independent of $\{o \overset{S}{\leftrightarrow} x\}$. We obtain

$$\begin{aligned}\frac{d}{dp}\mathbb{P}_p(o \leftrightarrow \partial\Lambda_n) &= \frac{1}{1-p}\sum_{\substack{S\subset\Lambda_n\\ o\in S}}\sum_{xy\in\Delta S} \mathbb{P}_p\big(o \overset{S}{\leftrightarrow} x\big)\mathbb{P}_p\big(\mathcal{S} = S\big)\\ &= \frac{1}{p(1-p)}\sum_{\substack{S\subset\Lambda_n\\ o\in S}} \varphi_p(S)\mathbb{P}_p\big(\mathcal{S} = S\big)\\ &\geq \frac{1}{p(1-p)}\inf_{\substack{S\subset\Lambda_n\\ o\in S}} \varphi_p(S)\cdot\mathbb{P}_p\big(o \not\leftrightarrow \partial\Lambda_n\big),\end{aligned}$$

as desired. □

THE END