Large Random Systems

Antti Kemppainen and Kalle Kytölä

Contents

Foreword

These lecture notes are primarily intended for the regular Master's level course *Large Random Systems* at Aalto University. The notes grew out of lectures of courses given by us at the University of Helsinki in 2014 and at Aalto University in 2015, 2017, and 2019.

One of the principal aims of the course is to learn to apply probability theory to interesting probabilistic models. We thus assume familiarity with measure theoretic probability as well as various undergraduate level topics in mathematics. To facilitate the study of the main content of the lectures, we nevertheless also recall some of the needed background in the Appendices. Our preference is to include a large number of different models in the lectures, but therefore none of the models can be studied in great depth. We devote no more than two lectures to any given model, and we can therefore only establish some basic results about each model. We refer the interested reader to more specialized texts about further results. Besides treating specific models, the course contains some development of general theory, in particular related to weak convergence and tightness.

The notes are still in a very preliminary and incomplete form, and it is our goal to gradually improve and extend them. The notes will in particular be frequently updated during the current course. Please send all feedback about mistakes, misprints, needs for clarification etc. to Kalle Kytölä (`kalle.kytola@aalto.fi`).

Lecture I

Introduction

In this introductory lecture we will discuss two examples:

- Random permutations: different applications
- Random walk: recurrence and transience depending on dimension

The two examples are treated in different fashion.

For random permutations the objective is not to prove any theorems, but rather to illustrate how such apparently very simple random objects are relevant to modeling various interesting phenomena. Along as we describe the modeling, we also state a few known mathematical results relevant to the analysis of the models.

For random walks, instead, our focus is on proving Pólya's theorem: a core result which shows a qualitative difference in the long time behavior properties of the random walk depending on the dimensionality of the space. Some applications of random walks and Pólya's theorem will be encountered later on in this course, and the reader can probably come up with many other applications.

The examples in this lecture are thus intended as introductions to the topic of large random systems from the perspectives of modeling and of mathematical analysis.

1. Random permutations as large random systems

Let us first discuss random permutations, and explore the variety of situations in which they may arise.

As a preparation, we recall some facts about permutations, and fix some notation.

A *permutation* of a set $\mathfrak{X}$ is a bijective function $\sigma\colon \mathfrak{X} \to \mathfrak{X}$. The set of permutations of $\mathfrak{X}$ is denoted by $\mathfrak{S}(\mathfrak{X})$, and it naturally has the structure of a group: the group operation is the composition of functions. We are commonly interested in permutations of a finite set $\mathfrak{X}$, and if the number of elements of that set is n, it is conventional to choose $\mathfrak{X} = \{1, 2, \dots, n\}$ for simplicity. For this case, we use the special notation

$$\mathfrak{S}_n = \Big\{\sigma\colon\ \{1,\dots,n\} \to \{1,\dots,n\}\ \text{ bijective}\Big\}. \tag{I.1}$$

The group $\mathfrak{S}_n$ is called the *symmetric group* on n symbols. It is a finite group: the number of different permutations of n symbols is

$$\#\mathfrak{S}_n = n! = n \cdot (n-1) \cdots 2 \cdot 1. \tag{I.2}$$

Since $\mathfrak{S}_n$ is a finite group, there is one particularly natural probability measure on it: the uniform measure, which associates probability $\frac{1}{n!}$ to each permutation $\sigma \in \mathfrak{S}_n$. If σ is a random variable with values in $\mathfrak{S}_n$, whose law is this uniform measure on $\mathfrak{S}_n$, we say that σ is a *uniform random permutation* of n symbols.

To get a feeling for uniform random permutations, here are two problems that the reader should try to solve.

Exercise I.1 (Cycle decomposition of a random permutation)**.**
Let σ be a uniformly distributed random permutation of the set $\{1, 2, \dots, n\}$. Compute the following quantities about its cycle decomposition.[1]

(a) Let L be the length of the cycle that contains the element 1. What is the distribution of L, i.e. probabilities $\mathsf{P}[L = \ell]$? Calculate also $\mathsf{E}[L]$.
(b) Let S be the number of cycles in the cycle decomposition. Calculate $\mathsf{E}[S]$.
(c) What is the probability that elements 1 and 2 belong to the same cycle?

Exercise I.2 (Fixed point in a random permutation)**.**

(a) Let $E_1, \dots, E_n \subset \Omega$ be events. Prove the inclusion-exclusion formula:

$$\begin{aligned}
&\mathsf{P}\Big[\bigcup_{j=1}^{n} E_j\Big] \\
&= -\sum_{\substack{J\subset\{1,\dots,n\}\\ J\neq\emptyset}} (-1)^{\#J}\, \mathsf{P}\Big[\bigcap_{j\in J} E_j\Big] \\
&= \sum_{1\le j_1\le n} \mathsf{P}[E_{j_1}] - \sum_{1\le j_1<j_2\le n} \mathsf{P}[E_{j_1}\cap E_{j_2}] + \sum_{1\le j_1<j_2<j_3\le n} \mathsf{P}[E_{j_1}\cap E_{j_2}\cap E_{j_3}] - \cdots .
\end{aligned}$$

(b) What is the probability that a uniformly distributed random permutation of the set $\{1, 2, \dots, n\}$ has a fixed point, i.e., a cycle of length 1? Compute the limit of this probability as $n \to \infty$.

Hint: *In part (a), you may want to use indicator random variables and consider the complementary event. In part (b), set* $E_j = \{$*the point* j *is a fixed point*$\}$.

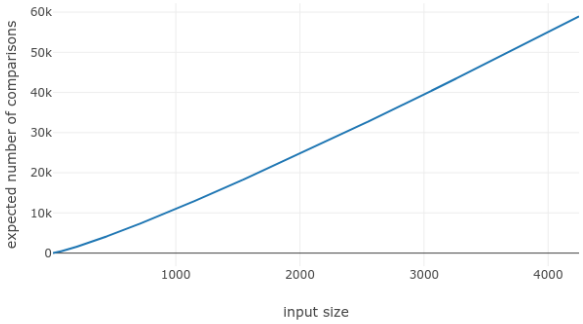
1.1. Sorting algorithms

A common task in computer programming is to sort a list of n elements, often for n very large. Computer scientists know various sorting algorithms for this purpose, among them, e.g., "quicksort", "mergesort", "insertion sort", etc. Roughly speaking, a sorting algorithm is a procedure which makes pairwise comparisons between the order of elements in the list, and then makes rearrangements of the order of the list according to the results of the comparisons, so that eventually the list becomes completely sorted, i.e., its elements appear in an increasing order.

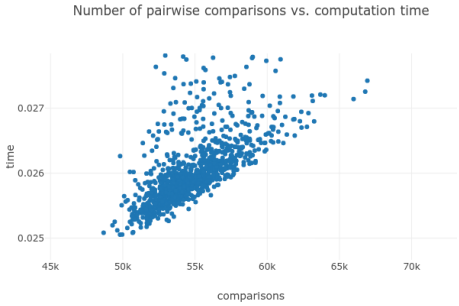
The effectiveness of an algorithm is measured by its use of computational resources, mainly by the amount of processor time needed before the algorithm outputs a final sorted list (one could also consider other aspects such as memory requirement). For sorting algorithms, the required processor time is (usually) well approximated by the number C of pairwise comparisons that were needed, see, e.g., Figure I.1(b).

The number C of comparisons depends on the input, i.e. the original list provided to the algorithm, which was to be sorted. In the absence of any further information about the input (or simply to analyze the algorithm on typical inputs), it is reasonable to assume that the input list is equally likely to be in any possible order — thus represented by a uniform random permutation of n elements. We thus model the input as being random. Although there may be nothing random about

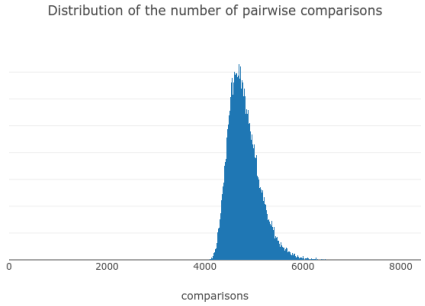
[1] *Recall:* A permutation can be written as a composition of disjoint cycles so that each element appears in exactly one cycle, and up to the order of cycles this cycle decomposition is unique.



(a) A good measure of the efficiency of the quicksort algorithm is the expected number m_n of needed pairwise comparisons as a function of the number n of elements to be sorted. Quicksort's behavior $m_n \sim 2n\log(n)$ is nearly optimal for sorting algorithms, and it is in fact not easy to distinguish the plot above from linear growth.



(b) The number of pairwise comparisons probes the actual wall clock running time of the sorting algorithm well. Shown here is data from 1000 quicksorts of lists of $n = 4000$ elements. An essentially linear relationship between number of comparisons and computation time can be seen (the outliers above the main cloud are probably due to other tasks using processor time besides just the sorting).



(c) The number of needed pairwise comparisons is quite well concentrated around its expected value m_n, and has a slightly skewed distribution. Shown here is a histogram of the number of pairwise comparisons from 40000 quicksorts of lists of $n = 500$ elements. The expected number of comparisons is $m_n \approx 4806$ and the standard deviation is $\sqrt{v_n} \approx 324$.

FIGURE I.1. Visualizatons of quicksort performance.

the behavior of the algorithm for a given input, for example the required number of comparisons needed depends on the random input, and as such becomes random, see, e.g., Figure I.1(c).

To give some concreteness to the above discussion of sorting algorithm performance, in Example I.1 below we briefly consider the average case performance of a widely used Quicksort algorithm. The interested reader will find more details both about sorting algorithms in general and also specifically on the analysis of quicksort in the excellent book [**Knu97**].

Example I.1 (Quicksort)**.**
Quicksort is a recursive algorithm, the simplest variant of which is informally described as follows:

- The input to the algorithm `QuickSort` is a list of elements $\mathbf{a} = (a_1, \dots, a_n)$ in some set $\mathfrak{X}$ with a total order relation $\leq$.
- If the input list contains no more than one element, then just output the list itself. Specifically, for a one element list (a_1) return $\texttt{QuickSort}(a_1) := (a_1)$, and for an empty list $\emptyset$ return $\texttt{QuickSort}(\emptyset) := \emptyset$.
- Otherwise the input list contains more than one element. Then choose one element a_r from the list, and compare it to all other elements: the other elements a_j with $a_j \leq a_r$ form a list $\mathbf{a}^-$ of length at most $n-1$, and those with $a_j > a_r$ form another list $\mathbf{a}^+$ of length at most $n-1$.
- Apply quicksort to the (shorter) sublists $\mathbf{a}^-$ and $\mathbf{a}^+$, to get the sorted sublists $\texttt{QuickSort}(\mathbf{a}^-)$ and $\texttt{QuickSort}(\mathbf{a}^+)$.
- Output the list constructed from the sorted sublist $\mathbf{a}^-$, the element a_r, and the sorted sublist $\mathbf{a}^-$

$$\texttt{QuickSort}(\mathbf{a}) := \big(\texttt{QuickSort}(\mathbf{a}^-);\, a_r;\, \texttt{QuickSort}(\mathbf{a}^+)\big).$$

As the comparison element a_r one could take the first element a_1, but it is often safer to actually take a randomly chosen element of the list.

Consider the performance of the quicksort algorithm above for an input of n elements in a uniformly random order, with the simplifying assumption that the list can not contain two equal elements. Denote the random number of comparisons by C_n. We can first of all ask about the expected number of comparisons needed,

$$m_n = \mathsf{E}[C_n],$$

which represents the average case performance of quicksort. The behavior of m_n as a function of the input size n is illustrated in Figure I.1(a). It is also important to know how big are the random fluctuations around the average case, and for this purpose one can compute the variance

$$v_n = \mathsf{Var}(C_n).$$

The numbers m_n and v_n can be calculated as solutions to recursive equations. The recursion for m_n, for example, is the following. To sort a list of n elements, the algorithm first needs to compare one chosen element a_r to $n-1$ other elements. Then we obtain two sublists, of lengths k and $n-1-k$, where $k \in \{0, 1, \dots, n-1\}$ is uniformly random, because the comparison element was the $k+1$:st smallest with probability $\frac{1}{n}$. The total number of comparisons is $n-1$ plus the number of comparisons needed to sort the sublists, and the expected value is

$$m_n = n - 1 + \frac{1}{n}\sum_{k=0}^{n-1} \big(m_k + m_{n-1-k}\big). \tag{I.3}$$

With initial conditions $m_0 = 0$ and $m_1 = 0$, the solution to this recursion is (verify by yourself)

$$m_n = 2(n+1)\sum_{j=1}^{n} \frac{1}{j} - 4n, \tag{I.4}$$

whose asymptotic behavior for large n is

$$m_n \sim 2n\log(n). \tag{I.5}$$

Figure I.1(a) contains a plot of m_n as a function of n.

By recursive analysis one can also show that that variance is asymptotically

$$v_n \sim c \times n^2, \qquad \text{where } c \approx 0.4202.$$

In particular, for large n, the typical random fluctuations of C_n around the expected value $m_n \sim \text{const.} \times n\log(n)$ are on a scale $\sqrt{v_n} \sim \text{const.} \times n$. Since $\sqrt{v_n}/m_n \to 0$, the random number of comparisons C_n is well concentrated around the average case value m_n. Figure I.1(c) illustrates the distribution of C_n.

Exercise I.3 (Expected number of comparisons for quicksort exactly)**.**
Check that m_n given by (I.4) satisfies the recursion (I.3)

Exercise I.4 (Expected number of comparisons for quicksort asymptotically)**.**
Check that the asymptotic behavior of m_n given by (I.4) is as stated in (I.5), or more precisely, show that

$$\lim_{n\to\infty} \frac{m_n}{n\log(n)} = 2.$$

1.2. Interface in a disordered material

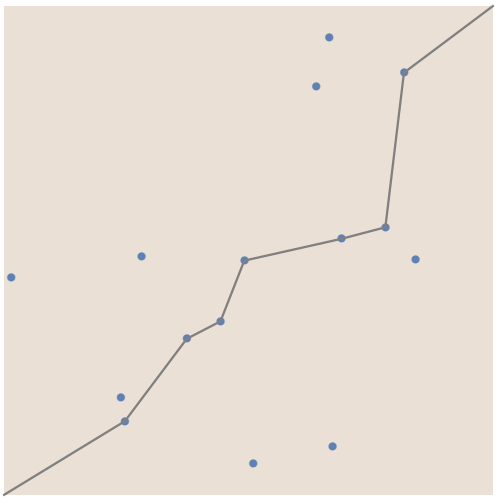
As our second example of random permutations in interesting applications, we will discuss a model of disordered material, with an interface in it.

The model we will consider is visualized in Figure I.2, and is formally described as follows. Consider the unit square $S = [0,1] \times [0,1]$ in the plane. Let $Z_j = (X_j, Y_j)$, $j = 1, \ldots, n$, be n independent random points uniformly distributed in the square S. Our object of interest will be a certain directed path γ from the bottom left corner $(0,0) \in S$ to the top right corner $(1,1) \in S$ whose both coordinates are non-decreasing functions, i.e., the direction of the path is restricted to the North-East quadrant. We want this path γ to go through as many of the random points $Z_1, \ldots, Z_n \in S$ as possible. There may not be a unique such path, but the maximal number ℓ of points on any such path is well defined, given the locations of the points. Optimal paths γ in various samples of random points are drawn in the illustrations of Figure I.2.

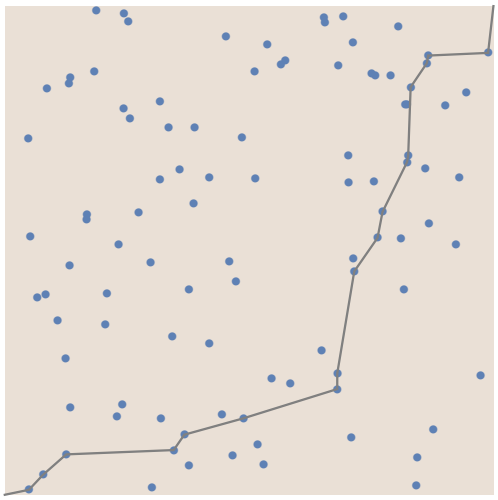
This is a simple model used in disordered materials physics. The easiest interpretation is that the square S represents a piece of some material, and the the random points $Z_1, \ldots, Z_n \in S$ represent impurities in this material. The material is considered to be weaker at the locations of the impurities, so it is easier to break at the locations of the impurities: the energy needed to break the material along a path γ could have a negative contribution from each of the impurities on γ, so that it requires less energy to break along paths with many impurities. The optimal interface γ is then the fracture that will actually be formed when the material is torn apart, and the number of impurities on it is related to the amount of energy needed to tear the material. Alternatively the interface could be modeling a directed polymer in a disordered environment or a domain wall between two phases that are enforced by boundary conditions. The reader can find more about modeling disordered materials and interfaces in them for example in [**KZ87**].

The maximal number of points that the directed path can pass through depends on the randomly chosen points $Z_1, \ldots, Z_n$, and is therefore itself a random variable, which we denote by L_n, to emphasize the underlying choice of n random points. In the disordered materials interpretation this random number represents the energy advantage that the interface can achieve by making use of the randomly located impurities. We will next argue that L_n arises in a simple way from random permutations.

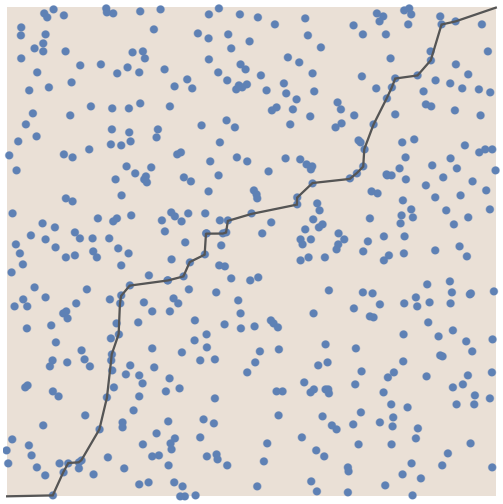
Fix the random points $Z_1 = (X_1, Y_1)$, $\ldots$, $Z_n = (X_n, Y_n)$. It is convenient to order these points by the values of their x-coordinates, so let us agree to relabel them so that $0 < X_1 < X_2 < \cdots < X_n < 1$ (note that the x-coordinate values are almost surely different, so no equalities arise with probability one). Then there is no reason for the y-coordinates to be in any specific order, and instead the rearrangement



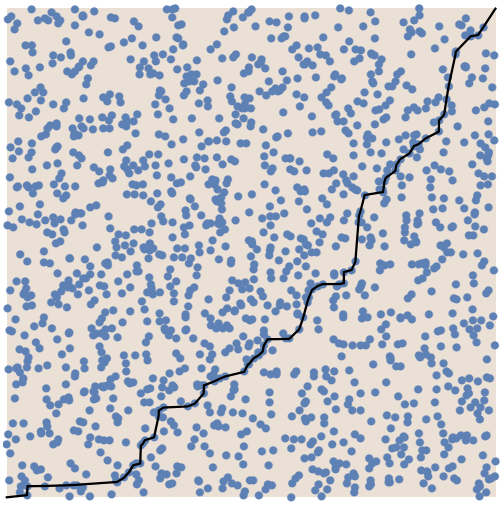
(a) A sample of $n = 15$ points in a square, and a NE path through $\ell = 7$ of them.



(b) A sample of $n = 100$ points in a square, and a NE path through $\ell = 17$ of them.



(c) A sample of $n = 500$ points in a square, and a NE path through many of them.



(d) A sample of $n = 1500$ points in a square, and a NE path through very many of them.

FIGURE I.2. North-East directed paths through maximal numbers of points in random samples of different sizes.

$\sigma \in \mathfrak{S}_n$ such that $0 < Y_{\sigma(1)} < Y_{\sigma(2)} < \cdots < Y_{\sigma(n)} < 1$, is a uniform random permutation.

An increasing subsequence in a permutation $\sigma \in \mathfrak{S}_n$ is a sequence $(i_1, i_2, \ldots, i_\ell)$ of indices such that $1 \leq i_1 < i_2 < \cdots < i_\ell \leq n$ and $\sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_\ell)$. A longest increasing subsequence in σ is such a sequence, where the length ℓ of the sequence is maximal. Again, for a given permutation σ there may not exist a unique longest increasing subsequence, but the maximal length $\ell = \ell(\sigma)$ of increasing subsequences is well defined. Now suppose that $X_1 < \cdots < X_n$ and $Y_{\sigma(1)} < \cdots < Y_{\sigma(n)}$ are as above. Then note that for any $(i_1, i_2, \ldots, i_\ell)$ such that $1 \leq i_1 < i_2 < \cdots < i_\ell \leq n$, the existence of a NE-directed path γ through points $(X_{i_1}, Y_{i_1}), \ldots, (X_{i_\ell}, Y_{i_\ell})$ is equivalent to $\sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_\ell)$. This says that $(X_{i_1}, Y_{i_1}), \ldots, (X_{i_\ell}, Y_{i_\ell})$ lie on some NE-directed path γ if and only if $(i_1, i_2, \ldots, i_\ell)$ is an increasing subsequence of the permutation σ which is needed to rearrange to

y-coordinate values to an increasing order. We conclude that the random number L_n of points that can be visited by an interface γ is the same as the length $\ell(\sigma)$ of the longest increasing subsequence in a uniform random permutation σ in $\mathfrak{S}_n$,

$$L_n = \ell(\sigma).$$

In a big piece of material, the number of impurities n becomes large, so we should study the behavior of L_n as $n \to \infty$. It can be shown that the asymptotic behaviors of the expected value and variance of L_n are

$$\mathsf{E}[L_n] \sim 2\sqrt{n} \qquad \text{and} \qquad \mathsf{Var}(L_n) \sim \text{const.} \times n^{1/3}.$$

Moreover, a celebrated result of Baik, Deift, and Johansson [**BDJ99**] says that as $n \to \infty$ the laws of the random variables

$$\frac{L_n - 2\sqrt{n}}{n^{1/6}}$$

tend to a limit known as the Tracy-Widom distribution. We refer the interested reader the original article [**BDJ99**] and a nice solitaire reinterpretation [**AD99**] of it. To get some feel for the problem, here is an exercise for the interested reader.

Exercise I.5 (Length of the longest increasing subsequence in random permutation)**.**
For $n \in \mathbb{N}$, consider a uniform random permutation σ in $\mathfrak{S}_n$. We say that k distinct indices

$$1 \le j_1 < j_2 < \cdots < j_k \le n$$

form a length k increasing subsequence of σ if

$$\sigma(j_1) < \sigma(j_2) < \cdots < \sigma(j_k).$$

We denote by $X_k^{(n)}$ the number of length k increasing subsequences of σ, and by

$$L_n = \max\Big\{k \;\Big|\; \sigma \text{ has a length } k \text{ increasing subsequence}\Big\}$$

the length of the longest increasing subsequence of σ.

(a) Show that for all $k \in \mathbb{N}$ we have $k! \ge k^k e^{-k}$.
(b) Calculate the expected number of length k increasing subsequences $\mathsf{E}\big[X_k^{(n)}\big]$, and derive the upper bound $\mathsf{E}\big[X_k^{(n)}\big] \le \big(\frac{ne^2}{k^2}\big)^k$ for it.
(c) Show that if $c > e$, then we have $\mathsf{P}\big[L_n \ge c\sqrt{n}\big] \to 0$ as $n \to \infty$.

1.3. Shufflings of a deck of cards

We finally consider a rather obvious application of random permutations, namely a shuffled deck of cards. A perfectly shuffled deck may be an idealization, and more interesting questions arise from studying how some common shuffling procedures gradually bring a deck of cards closer to being perfectly shuffled. So we will study shuffling as a process. The ideal output of a shuffling process is precisely the opposite of a sorting algorithm — sorting should bring the deck to a perfect order, whereas shuffling aims at bringing the deck to a perfect disorder.

Before studying the process, we should discuss how does one measure the distance from perfect disorder.

Total variation distance on a finite set

Let $\mathfrak{X}$ be a finite set, and let μ, ν be two probability measures on $\mathfrak{X}$. Define the *total variation distance* between the measures μ and ν as

$$\varrho_{\mathrm{TV}}(\mu,\nu) = \max_{E\subset\mathfrak{X}} \big|\mu[E] - \nu[E]\big|. \tag{I.6}$$

This can be interpreted as the largest error in the probability of any event that would be made, if the probabilities ν were used instead of the probabilities μ. It is natural to say that measures μ and ν are close if such errors are small.

Exercise I.6 (Total variation distance)**.**

(a) Show that the total variation distance between the measures μ and ν can be expressed as

$$\varrho_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\sum_{x\in\mathfrak{X}} \big|\mu[\{x\}] - \nu[\{x\}]\big|. \tag{I.7}$$

(b) Prove that the total variation distance is a metric on the space of probability measures on $\mathfrak{X}$.

Random order of cards in a deck

In a deck of n cards, the order of cards can be represented by a permutation $\pi \in \mathfrak{S}_n$. We choose the following way to do this. The n cards are given labels $1, 2, \ldots, n$, and $\pi(\ell)$ denotes the position of the card with label ℓ in the deck (counting from the top of the deck, for example).[2]

In a well shuffled deck of cards, all possible orders should be equally probable. This corresponds to the uniform probability measure ν_{unif} on $\mathfrak{S}_n$,

$$\nu_{\mathrm{unif}}[\{\pi\}] = \frac{1}{n!} \qquad \text{for all } \pi \in \mathfrak{S}_n. \tag{I.8}$$

If the random order of the deck follows a distribution ν instead, we can use the total variation distance $\varrho_{\mathrm{TV}}(\nu, \nu_{\mathrm{unif}})$ to measure how far the deck is from being well shuffled. As an extreme example, the fully ordered deck is described by the delta-measure δ_{id} at the identity permutation $\mathrm{id} \in \mathfrak{S}_n$. This (or indeed any non-random order) is very far from well shuffled since we have $\varrho_{\mathrm{TV}}(\delta_{\mathrm{id}}, \nu_{\mathrm{unif}}) = 1 - \frac{1}{n!}$.

One shuffle

The operation of shuffling a deck is modeled by applying a random permutation σ to the current order π of the deck. The card at position p is moved by the shuffle σ to the new position $\sigma(p)$, and thus the position of the card with label ℓ after the shuffle is $\sigma(\pi(\ell))$. In other words, the shuffle takes the current order π to the new order $\sigma \circ \pi$. How exactly the shuffling is performed determines the distribution of the permutation σ applied to the order of the deck.[3] For simplicity, we will only consider

[2] An alternative description would be to denote by $\lambda(p)$ the label of the card at position p. This description of the order of the deck is related to our convention by taking the inverse of the permutation, namely $\lambda = \pi^{-1}$.

[3] If the shuffle σ would itself be uniformly distributed, i.e., it would have the law ν given in (I.8), then regardless of the order π of the deck initially, the deck would be well shuffled after just one shuffle. Note that real methods used for shuffling do not have this law! In fact e.g. the most commonly used shuffle, known as "riffle shuffle" or "dovetail shuffle" starts by dividing the deck in a top part and a bottom part, and then letting the two parts interlace. In such a shuffle the order

one very simple method of shuffling described below — the interested reader can invent more realistic shuffles or find them in the literature.

Example I.2 (Top-to-middle shuffle)**.**
A very simple method of shuffling would take the top card and place it to a uniformly random position in the deck. If that position is m, then the permutation of the order of the deck is $\tau^{(m)}$, given by

$$\tau^{(m)}(p) = \begin{cases} m & \text{for } p = 1 \\ p-1 & \text{for } 1 < p \leq m \\ p & \text{for } p > m \end{cases},$$

so that the law μ of the top-to-middle shuffle σ is $\mu[\{\tau^{(m)}\}] = \frac{1}{n}$ for all m and $\mu[\{\tau\}] = 0$ for all other permutations $\tau \in \mathfrak{S}_n \setminus \{\tau^{(1)}, \dots, \tau^{(n)}\}$.

Repeated shuffles

A single shuffle does not make an ordered deck well-shuffled, but the idea is to shuffle repeatedly, i.e. apply a large number of independent shuffles one after another. Given the method of shuffling, i.e., the law μ of the shuffle σ, repeated shuffling defines a Markov chain on $\mathfrak{S}_n$, whose transition probabilities are[4]

$$P_{\pi,\pi'} = \mu[\{\pi' \circ \pi^{-1}\}] \qquad \text{for } \pi, \pi' \in \mathfrak{S}_n.$$

We want that a shuffle of an already well-shuffled deck produces a well-shuffled deck, which amounts to requiring that the uniform measure (I.8) is a stationary measure for the Markov chain. If, moreover, the Markov chain is irreducible and aperiodic, then by the basic theory of finite state space Markov chains we have that the distribution of the deck after s shuffles tends to uniform as $s \to \infty$. This, of course, is the rationale behind shuffling the deck before starting a card game.

Exercise I.7 (Stationarity of the uniform measure for top-to-middle shuffle)**.**
Check that the uniform measure (I.8) is stationary for the top-to-middle shuffle given in Example I.2.

Exercise I.8 (Irreducibility and aperiodicity of the top-to-middle shuffle Markov chain)**.**
Check that the top-to-middle shuffle given in Example I.2 defines an irreducible and aperiodic Markov chain on $\mathfrak{S}_n$.

Example I.3 (Convergence of top-to-middle shuffling)**.**
Let now $\pi_0 \in \mathfrak{S}_n$ be the initial order of the deck, and define the Markov chain as before with the top-to-middle shuffle given in Example I.2. Denote the (random) order of the deck after s shuffles by π_s, and its law by $\nu_s = \delta_{\pi_0} P^s$. By general Markov chain theory we have $\varrho_{\mathrm{TV}}(\nu_s, \nu_{\mathrm{unif}}) \to 0$ exponentially fast as $s \to \infty$.

We next analyze how the convergence happens more precisely. In particular, it turns out that there is a rather sharply defined treshold for the number of shuffles so that the deck is not well shuffled much before the treshold but is well shuffled soon after the treshold. This is a general phenomenon in shuffling, and it is of practical importance to find the treshold, because it provides an answer to the question:

---

of cards in the same part of the deck is not changed. In particular, if $i < j < k$ are three positions, then in a riffle shuffle σ it is impossible to have $\sigma(j) < \sigma(i) < \sigma(k)$. Thus, in fact, the riffle shuffle gives probability zero to a large number of permutations.

[4]We may note that the transition probabilities are invariant under the right-action of the group $\mathfrak{S}_n$ on itself, i.e., we have $P_{\pi\circ\tau,\pi'\circ\tau} = P_{\pi,\pi'}$ for any $\tau \in \mathfrak{S}_n$. These types of Markov chains on general groups G are known as random walks on groups.

How many shuffles are needed?

Our analysis is based on looking at the random time T at which the original bottom card (with label $\pi_0^{-1}(n)$) is at the top of the deck for the first time,

$$T = \min\left\{ s \in \mathbb{N} \;\middle|\; \pi_s(\pi_0^{-1}(n)) = 1 \right\}.$$

The reason this is useful is that $T+1$ is a "strong uniform time", which means that on the event $\{T = s\}$ we have that π_{s+1} follows exactly the uniform distribution (the proof is not difficult, and is left to the reader). This implies that the total variation distance between the law ν_s of π_s and the uniform measure ν_{unif} can not be more than the probability that the original bottom card has not yet been at the top of the deck,

$$\varrho_{\text{TV}}(\nu_s, \nu_{\text{unif}}) \leq \mathsf{P}\big[T \geq s\big].$$

Thus the convergence can be controlled once the distribution of T is controlled.

With a little bit of thinking, one can calculate the expected value and variance of T. Denote by $H_n = \sum_{m=1}^{n} \frac{1}{m}$ is the n:th harmonic number, and by $H_n^{(2)} = \sum_{m=1}^{n} \frac{1}{m^2}$. We then have

$$\mathsf{E}\big[T\big] = nH_{n-1} \qquad \text{and} \qquad \mathsf{Var}\big[T\big] = n^2 H_{n-1}^{(2)} - nH_{n-1}.$$

Let us just notice the following estimates: $H_n \leq 1 + \log(n)$ and $H_n^{(2)} \leq \frac{\pi^2}{6}$. With these, the Chebyshev inequality yields

$$\mathsf{P}\Big[T \geq n\log(n) + cn\Big] \leq \mathsf{P}\Big[(T - \mathsf{E}[T]) \geq (c-1)n\Big] \leq \frac{\mathsf{Var}[T]}{(cn)^2} \leq \frac{\pi^2}{6(c-1)^2}.$$

Combining with the above observation about the total variation distance, we get the conclusion that the deck becomes shuffled soon after $n\log(n)$ top-to-middle shuffles. More precisely, if $s \geq n\log(n) + cn$ for some $c > 1$, then the law ν_s of π_s satisfies

$$\varrho_{\text{TV}}(\nu_s, \nu_{\text{unif}}) \leq \frac{\pi^2}{6(c-1)^2}.$$

One can also show conversely that the deck will not be well shuffled much before $n\log(n)$ shuffles.

The phenomenon of sharp cutoff in the number of shuffles needed was made particularly well known by the work of Bayer and Diaconis [**BD92**], who studied a rather realistic "dovetail shuffle" or a "riffle shuffle". For the practical player with a deck of 52 cards, their analysis of the rate of convergence of the corresponding Markov chain could be summarized as: "seven shuffles is not enough, but eight or nine shuffles is already very good". The conclusion was quite surprising, as few players actually give more than seven shuffles in practise. The consequences to e.g. bridge hands have been considered, and there are even magic tricks making use of an insufficient shuffling.

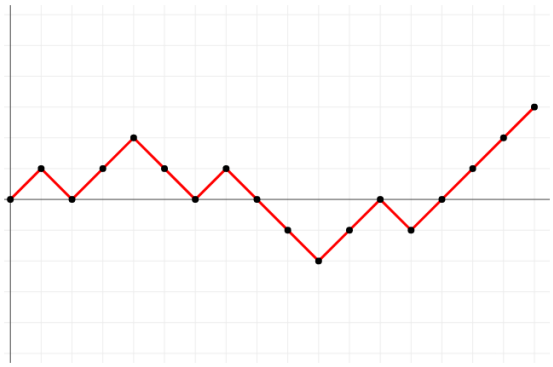
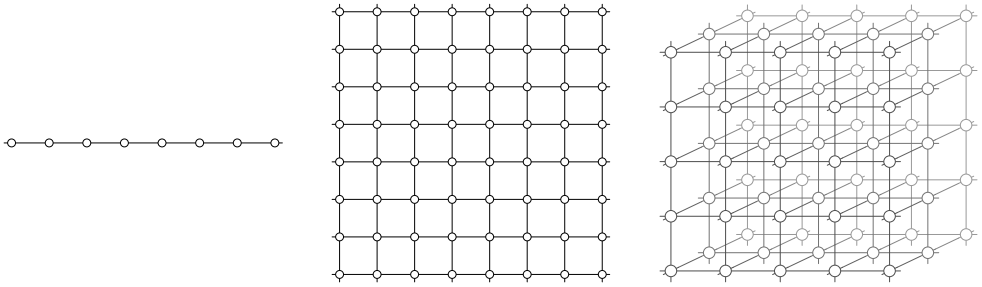


FIGURE I.3. A random walk on $\mathbb{Z}$.



(a) Integer lattice (b) Square lattice (c) Cubic lattice

FIGURE I.4. Hypercubic lattices $\mathbb{Z}^d$ in dimensions $d = 1, 2, 3$.

2. Random walk

Random walks are some of the most basic probabilistic models used in a wide variety of contexts. Here we will study the behavior of random walks on the d-dimensional *integer lattices* $\mathbb{Z}^d$, i.e., the *hypercubic lattices*. A recommended textbook about random walks is [**LL10**].

2.1. The simple random walk on d-dimensional lattice

Fix a dimension $d \in \mathbb{Z}_{>0}$. Consider the lattice

$$\mathbb{Z}^d = \Big\{ (x_1, \dots, x_d) \;\Big|\; x_1, \dots, x_d \in \mathbb{Z} \Big\}$$

of points with integer coordinates in the d-dimensional Euclidean space $\mathbb{R}^d$. These lattices are illustrated for $d = 1, 2, 3$ in Figure I.4, and they are also called hypercubic lattices. For clarity in this first lecture, we use the vector notation

$$\vec{0} = (0, 0, \dots, 0) \in \mathbb{Z}^d$$

for the origin of $\mathbb{Z}^d$ and

$$\begin{aligned}\vec{e}^{(1)} &= (1, 0, \dots, 0) \in \mathbb{Z}^d \\ \vec{e}^{(2)} &= (0, 1, \dots, 0) \in \mathbb{Z}^d \\ &\vdots \\ \vec{e}^{(d)} &= (0, 0, \dots, 1) \in \mathbb{Z}^d\end{aligned}$$

for the standard basis vectors, although later in the course we will omit the vector symbols for simplicity. We denote the Euclidean norm in $\mathbb{R}^d$ by

$$\|(x_1, \dots, x_d)\| = \sqrt{x_1^2 + \cdots + x_d^2}.$$

The set of nearest neighbors of the origin $\vec{0}$ is

$$\mathscr{N} = \left\{ \vec{v} \in \mathbb{Z}^d \;\middle|\; \|\vec{v}\| = 1 \right\}.$$

The $\#\mathscr{N} = 2d$ nearest neighbors of the origin are just the positive and negative standard basis vectors $\pm\vec{e}^{(j)}$, $j = 1, \dots, d$.

The *simple random walk* on $\mathbb{Z}^d$ is the following stochastic process. The time of the process is discrete, $t \in \mathbb{Z}_{\geq 0}$, and the position of the random walk at time t is a random point $X_t \in \mathbb{Z}^d$ of the lattice. The random walk, then, is a stochastic process, i.e., a collection of random variables $\big(X_t\big)_{t \in \mathbb{Z}_{\geq 0}}$ indexed by time t. To construct the random walk, first take a sequence $\vec{\xi}_1, \vec{\xi}_2, \dots$ of independent steps uniformly distributed on the set $\mathscr{N}$ of nearest neighbors of the origin,

$$\mathsf{P}\big[\vec{\xi}_s = \pm\vec{e}^{(j)}\big] = \frac{1}{2d}.$$

The position $X(t)$ of the random walk at time $t \in \mathbb{Z}_{\geq 0}$ is then defined as the sum of the first t steps,

$$X(t) = \sum_{s=1}^{t} \vec{\xi}_s. \tag{I.9}$$

One dimensional random walk

As a warm-up, consider the case $d = 1$. The random walk $X = \big(X(t)\big)_{t \in \mathbb{Z}_{\geq 0}}$ on $\mathbb{Z}$ is illustrated in Figure I.3. This should be a very familiar object, but even simple questions about X may require some thinking. The following easy exercise gives the law of the position $X(t)$ of the random walk at a fixed time t.

Exercise I.9 (Distribution of the position at a fixed time for a 1-dimensional random walk)**.** Show that for the random walk X on $\mathbb{Z}$, for a fixed $t \in \mathbb{Z}_{\geq 0}$ and $x \in \mathbb{Z}$ we have

$$\mathsf{P}\big[X(t) = x\big] = \begin{cases} \binom{t}{m} 2^{-t} & \text{if } x = t - 2m \text{ for some } m = 0, 1, \dots, t \\ 0 & \text{otherwise.} \end{cases}$$

The next exercise requires a more clever approach.

Exercise I.10 (Ballot problem)**.**
Show that for the random walk X on $\mathbb{Z}$, for $t \in \mathbb{Z}_{\geq 0}$ and $x \in \mathbb{Z}_{\geq 0}$ we have

$$\mathsf{P}\Big[X(t) = x \text{ and } \ X(s) \geq 0 \text{ for all } s \leq t\Big] = \binom{t}{\frac{1}{2}(t+x)} 2^{-t} - \binom{t}{\frac{1}{2}(t+x)+1} 2^{-t}.$$

2.2. Recurrence and transience of simple random walk

Our main goal for this lecture is to prove the following result about the behavior of the random walk at large times.

Theorem I.4 (Pólya's theorem)**.**
Let $X = \big(X(t)\big)_{t \in \mathbb{Z}_{\geq 0}}$ be a simple random walk on the d-dimensional hypercubic lattice $\mathbb{Z}^d$.

If $d \leq 2$, then the random walk X is recurrent*, which means that the walk will almost surely return to its starting point:*

$$\mathsf{P}\Big[\textit{for some } t > 0 \textit{ one has } X(t) = \vec{0}\Big] = 1.$$

If $d > 2$, then the simple random walk is transient*, which means that the walk has a positive probability to never return to its starting point:*

$$\mathsf{P}\Big[\textit{for some } t > 0 \textit{ one has } X(t) = \vec{0}\Big] < 1.$$

Proof. Denote briefly

$$p = \mathsf{P}\Big[\text{for some } t > 0 \text{ one has } X(t) = \vec{0}\Big].$$

Consider the number of times t at which the random walk is at the origin,

$$L = \#\Big\{t \in \mathbb{Z}_{\geq 0} \;\Big|\; X(t) = \vec{0}\Big\} = \sum_{t=0}^{\infty} \mathbb{I}_{\{X(t)=\vec{0}\}},$$

and its expected value

$$\mathsf{E}\big[L\big] = \sum_{t=0}^{\infty} \mathsf{P}\big[X(t) = \vec{0}\big].$$

Let $\tau = \min\big\{t > 0 \,\big|\, X(t) = \vec{0}\big\}$ be the first time at which the walk returns to the origin. Then we can write $p = \mathsf{P}[\tau < \infty]$. Conditionally on the event $\{\tau < \infty\}$ of a return occurring, the continuation $\big(X(\tau+s)\big)_{s \in \mathbb{N}}$ of the walk after its first return to the origin has the same law as the original random walk $\big(X(s)\big)_{s \in \mathbb{N}}$, and is independent of the steps before. In particular the continuation will return to the origin with the same probability p. More generally,

$$\mathsf{P}\big[L \geq k\big] = \mathsf{P}\big[L \geq k-1\big] \times \underbrace{\mathsf{P}\Big[L \geq k \;\Big|\; L \geq k-1\Big]}_{=p}$$

which together with $\mathsf{P}[L \geq 1] = 1$ inductively shows that $\mathsf{P}[L \geq k] = p^{k-1}$ for all $k \in \mathbb{Z}_{>0}$. The expected value of L can then be calculated as

$$\mathsf{E}\big[L\big] = \sum_{k=1}^{\infty} \mathsf{P}\big[L \geq k\big] = \sum_{k=1}^{\infty} p^{k-1} = \begin{cases} \frac{1}{1-p} & \text{if } p < 1 \\ +\infty & \text{if } p = 1 \end{cases},$$

where the first equality is due to the familiar identity (B.3) from Appendix B. By this calculation we conclude that the finiteness of the expected value of L characterizes transience:

$$\begin{aligned} X \text{ is recurrent} \quad &\Longleftrightarrow \quad \mathsf{E}\big[L\big] = +\infty \\ X \text{ is transient} \quad &\Longleftrightarrow \quad \mathsf{E}\big[L\big] < +\infty. \end{aligned} \tag{I.10}$$

We thus seek to calculate $\mathsf{E}\big[L\big]$. It turns out to be convenient to first generalize slightly: for $\lambda \in (0,1)$ a parameter and $\vec{x} \in \mathbb{Z}^d$ a point on the lattice, defined the λ-weighted number of visits to $\vec{x}$ by

$$L_\lambda(\vec{x}) = \sum_{t=0}^{\infty} \lambda^t \, \mathbb{I}_{\{X(t)=\vec{x}\}},$$

and consider its expected value

$$G_\lambda(\vec{x}) = \mathsf{E}\big[L_\lambda(\vec{x})\big] = \sum_{t=0}^{\infty} \lambda^t \, \mathsf{P}\big[X(t)=\vec{x}\big].$$

The parameter $\lambda < 1$ conveniently keeps these quantities finite, because by Fubini's theorem we have

$$\sum_{\vec{x}\in\mathbb{Z}^d} G_\lambda(\vec{x}) = \sum_{t=0}^{\infty} \lambda^t \, \mathsf{E}\Big[\underbrace{\sum_{\vec{x}\in\mathbb{Z}^d} \mathbb{I}_{\{X(t)=\vec{x}\}}}_{=1} \Big] = \sum_{t=0}^{\infty} \lambda^t = \frac{1}{1-\lambda} < +\infty,$$

and in particular $G_\lambda(\vec{x}) \le \frac{1}{1-\lambda} < +\infty$ for any $\vec{x} \in \mathbb{Z}^d$. In the end we will take the limit $\lambda \to 1$, and by the Monotone Convergence Theorem (Theorem A.9 in Appendix A) then

$$\lim_{\lambda\uparrow 1} G_\lambda(\vec{0}) = \mathsf{E}\big[L\big].$$

Besides finiteness, another advantage is that $G_\lambda(\vec{x})$ can be directly calculated via its Fourier transform.

By considering separately the $2d$ possibilities $\pm\vec{e}_j$ for the first step of the random walk and then the continuation after the first step, we obtain the following difference equation for $G_\lambda(\vec{x})$

$$\begin{aligned} G_\lambda(\vec{x}) &= \sum_{t=0}^{\infty} \lambda^t \, \mathsf{P}\big[X(t)=\vec{x}\big] = \delta_{\vec{x},\vec{0}} + \sum_{t=1}^{\infty} \lambda^t \, \mathsf{P}\big[X(t)=\vec{x}\big] \\ &= \delta_{\vec{x},\vec{0}} + \frac{1}{2d}\Big(\lambda\, G_\lambda(\vec{x}+\vec{e}_1) + \lambda\, G_\lambda(\vec{x}-\vec{e}_1) + \cdots + \lambda\, G_\lambda(\vec{x}-\vec{e}_d)\Big) \\ &= \delta_{\vec{x},\vec{0}} + \frac{\lambda}{2d} \sum_{j=1}^{d} \big(G_\lambda(\vec{x}+\vec{e}_j) + G_\lambda(\vec{x}-\vec{e}_j)\big). \end{aligned}$$

The discrete Fourier transform with respect to $\vec{x}$ is

$$\widehat{G}_\lambda(\vec{\theta}) = \sum_{\vec{x}\in\mathbb{Z}^d} e^{-\mathbb{i}\,\vec{\theta}\cdot\vec{x}}\, G_\lambda(\vec{x}) \qquad (\vec{\theta} \in \mathbb{R}^d),$$

and the inverse transform expresses $G_\lambda(\vec{x})$ as

$$G_\lambda(\vec{x}) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} e^{\mathbb{i}\,\vec{\theta}\cdot\vec{x}}\, \widehat{G}_\lambda(\vec{\theta})\, \mathrm{d}^d\vec{\theta}.$$

Multiplying the difference equation by $e^{-\mathbb{i}\,\vec{\theta}\cdot\vec{x}}$ and summing over $\vec{x}$ we obtain an equation for the Fourier transform

$$\widehat{G}_\lambda(\vec{\theta}) = 1 + \frac{\lambda}{2d} \sum_{j=1}^{d} \big(e^{\mathbb{i}\,\vec{\theta}\cdot\vec{e}_j} + e^{-\mathbb{i}\,\vec{\theta}\cdot\vec{e}_j}\big)\widehat{G}_\lambda(\vec{\theta}),$$

which is easily solved to yield

$$\widehat{G}_\lambda(\vec{\theta}) = \frac{1}{1 - \frac{\lambda}{d}\sum_{j=1}^{d} \cos(\theta_j)},$$

Inverting the Fourier transform gives

$$G_\lambda(\vec{x}) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{\mathbb{i}\,\vec{\theta}\cdot\vec{x}}}{1 - \frac{\lambda}{d}\sum_{j=1}^{d} \cos(\theta_j)}\, \mathrm{d}^d\vec{\theta}.$$

We are interested the finiteness of $\mathsf{E}[L]$, which is recovered by letting $\lambda \to 1$ as

$$\mathsf{E}\big[L\big] = \lim_{\lambda\uparrow 1} G_\lambda(\vec{0}) = \lim_{\lambda\uparrow 1} \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{1}{1 - \frac{\lambda}{d}\sum_{j=1}^d \cos(\theta_j)}\, \mathrm{d}^d\vec{\theta}. \tag{I.11}$$

Recurrence and transience thus reduce to the question of whether this integral (I.11) diverges or converges. Note that outside of any neighborhood $U \subset [-\pi,\pi]^d$ of the point $\vec{\theta} = \vec{0} \in [-\pi,\pi]^d$, the expression $\big(1 - \frac{\lambda}{d}\sum_{j=1}^d \cos(\theta_j)\big)^{-1}$ is bounded uniformly in λ, so dominated convergence (Theorem A.10 in Appendix A) guarantees that the integral over $[-\pi,\pi]^d \setminus U$ tends to a finite limit as $\lambda \to 1$. Only the contribution of a neighborhood of $\vec{\theta} = \vec{0}$ could possibly make the limit (I.11) infinite. We will thus study the integral over a ball $U = B_\varepsilon(\vec{0})$ centered at $\vec{\theta} = \vec{0}$ with a small radius $\varepsilon > 0$.

Note that if $\varepsilon < \frac{\pi}{2}$ and $\vec{\theta} \in B_\varepsilon(\vec{0})$, then $\cos(\theta_j) > 0$ for all $j = 1, \dots, d$. Therefore the integrand in the following is pointwise monotone increasing, and the limit may be exchanged with the integral by monotone convergence (Theorem A.9)

$$\lim_{\lambda\uparrow 1} \int_{B_\varepsilon(\vec{0})} \frac{1}{1 - \frac{\lambda}{d}\sum_{j=1}^d \cos(\theta_j)}\, \mathrm{d}^d\vec{\theta} = \int_{B_\varepsilon(\vec{0})} \frac{1}{1 - \frac{1}{d}\sum_{j=1}^d \cos(\theta_j)}\, \mathrm{d}^d\vec{\theta}. \tag{I.12}$$

Now note also that if $|\alpha| < \frac{\pi}{4}$, then we have the "up to constant bounds estimate"[5]

$$1 - \cos(\alpha) \asymp \alpha^2.$$

Therefore for $\vec{\theta} \in B_\varepsilon(\vec{0})$ with $\varepsilon < \frac{\pi}{4}$ we have

$$\frac{1}{1 - \frac{1}{d}\sum_{j=1}^d \cos(\theta_j)} = \frac{1}{\frac{1}{d}\sum_{j=1}^d \big(1 - \cos(\theta_j)\big)} \asymp \frac{1}{\sum_{j=1}^d \theta_j^2} = \frac{1}{\|\vec{\theta}\|^2}.$$

This gives us up to constant bounds estimate for the integral (I.12):

$$\int_{B_\varepsilon(\vec{0})} \frac{1}{1 - \frac{1}{d}\sum_{j=1}^d \cos(\theta_j)}\, \mathrm{d}^d\vec{\theta} \asymp \int_{B_\varepsilon(\vec{0})} \frac{1}{\|\vec{\theta}\|^2}\, \mathrm{d}^d\vec{\theta} = \int_0^\varepsilon \frac{1}{r^2} \times A_d\, r^{d-1}\, \mathrm{d}r,$$

where the integral was performed in the latter step in radial coordinates, and $A_d > 0$ is the $d-1$-dimensional area of the unit sphere in d-dimensional Euclidean space. This integral proportional to $\int_0^\varepsilon r^{d-3}\, \mathrm{d}r$ diverges if $d \le 2$ and converges if $d > 2$.

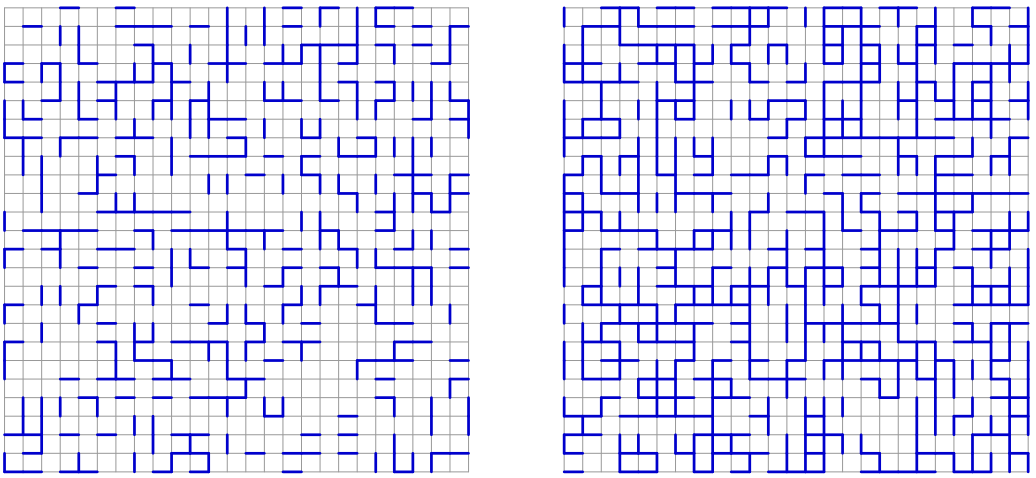
We conclude that $\mathsf{E}[L] = +\infty$ if $d \le 2$ $\mathsf{E}[L] < +\infty$ if $d > 2$. By the characterization (I.10) this says that X is recurrent if $d \le 2$ and transient if $d > 2$. □

[5]The "up to constant bounds" notation $f(x) \asymp g(x)$ means that there exists positive constants $C_1, C_2 > 0$ such that $C_1\, g(x) \le f(x) \le C_2\, g(x)$ for all x. For example here one specifically has $\frac{1}{2\sqrt{2}}\alpha^2 \le 1 - \cos(\alpha) \le \frac{1}{2}\alpha^2$ for $|\alpha| < \frac{\pi}{4}$.

Lecture II

Percolation

Percolation can be thought of as a model of porous material. In a percolation configuration, different spacial locations (of the material) are declared either open or closed, independently with some fixed probability. We may think that water can penetrate a connected set of open locations. Figure II.1 illustrates percolation in two-dimensions.



(a) Bond percolation on $\mathbb{Z}^2$ with probability of open bonds $p = 0.25$.

(b) Bond percolation on $\mathbb{Z}^2$ with probability of open bonds $p = 0.6$.

FIGURE II.1. Bond percolation on $\mathbb{Z}^2$ illustrated.

There are also various other interpretations of exactly the same mathematical model of percolation. Instead of water penetrating a porous material, we could be modeling for example the spreading of a disease (or a rumor, or a forest fire, or something else) in such a way that infected locations (or locations of individuals having heard the rumor, or burning locations in a forest) are able to infect (or spread rumor to, or transmit fire to) nearby locations provided the passage between them is open. Again, disease (or rumor, or disease) can then spread to a connected set of open locations.

In the standard setup, the model is defined on an infinite spacial domain. The simplest fundamental question is whether an infinite connected component of open locations exists. The configuration is then said to "percolate". This is interpreted as e.g. water being able to pass through the material, or forest fire (or disease) being able to spread. Formulated this way, percolation is obviously a large random system — in fact an infinite one. Our main goal is to show a phase transition result for a percolation model in dimensions at least two: the model has a percolating phase when the probability of open locations is higher than a certain critical value,

and a non-percolating phase when this probability is lower. The key technique is Kolmogorov's 0-1–law, which we recall in Appendix E, Theorem E.1.

Great textbooks on percolation are for example [**Wer09, Gri99, DC12**]. They of course develop the theory much further than we can do in just one lecture.

1. Bond percolation on hypercubic lattice

For definiteness, we consider the model of bond percolation on a d-dimensional integer lattice

$$\mathbb{Z}^d = \big\{x = (x_1, \dots, x_d) \;\big|\; x_1, \dots, x_d \in \mathbb{Z}\big\}$$

(the hypercubic lattice). By a *site*, we mean a point $x \in \mathbb{Z}^d$, and by a *bond* we mean an unordered pair of two sites at unit distance from each other.[1] We denote the set of bonds by

$$\mathrm{E}(\mathbb{Z}^d) = \Big\{\{x, y\} \;\Big|\; x, y \in \mathbb{Z}^d, \|x - y\| = 1\Big\}.$$

In bond percolation we say that each bond $e \in \mathrm{E}(\mathbb{Z}^d)$ can be open or closed, represented by a variable ω_e taking value 1 or 0, respectively. A parameter $p \in [0, 1]$ gives the probability of a bond to be open, $\mathsf{P}[\omega_e = 1] = p$, $\mathsf{P}[\omega_e = 0] = 1 - p$. The states of bonds are taken to be independent.

Thus, the sample space is

$$\Omega = \{0, 1\}^{\mathrm{E}(\mathbb{Z}^d)} = \Big\{\omega = (\omega_e)_{e \in \mathrm{E}(\mathbb{Z}^d)} \;\Big|\; \omega_e \in \{0, 1\}\ \forall e\Big\},$$

and P is the infinite product measure $\mathrm{Bernoulli}(p)^{\otimes \mathrm{E}(\mathbb{Z}^d)}$. The σ-algebra is the product σ-algebra (also called the cylinder σ-algebra), i.e., the smallest σ-algebra with respect to which all coordinate projections $\omega \mapsto \omega_e$ are measurable.

Given two sites $x, y \in \mathbb{Z}^d$, a path of length ℓ on the lattice $\mathbb{Z}^d$ from x to y is a sequence of $\ell + 1$ distinct sites $z_0, z_1, \dots, z_{\ell-1}, z_\ell \in \mathbb{Z}^d$ such that for all $j = 1, \dots, \ell$ we have $\|z_j - z_{j-1}\| = 1$. The ℓ bonds of the path are $e_j = \{z_{j-1}, z_j\}$, $j = 1, \dots, \ell$. The path is said to be an open path for the configuration $\omega = (\omega_e)_{e \in \mathrm{E}(\mathbb{Z}^d)}$ if all its bonds are open, i.e., we have $\omega_{e_j} = 1$ for all j. The sites $x, y \in \mathbb{Z}^d$ are said to be connected (in the configuration ω), if there exists an open path from x to y. This is denoted by $x \leftrightsquigarrow y$. Note that $\leftrightsquigarrow$ is a (random) equivalence relation. It divides the set of sites to (random) equivalence classes, which we call the connected components (or clusters) of the configuration. The component of $x \in \mathbb{Z}^d$ is the set of sites connected to x,

$$\mathcal{C}_x = \big\{y \in \mathbb{Z}^d \;\big|\; x \leftrightsquigarrow y\big\}.$$

Figure II.2 illustrates a connected component in a percolation configuration.

Exercise II.1 (Measurability questions in percolation)**.**
Check the following measurability properties in the cylinder σ-algebra.

- The event $\{x \leftrightsquigarrow y\}$ that x is connected to y is measurable.
- The size $\#\mathcal{C}_x$ of the connected component of x, is a measurable random variable.
- The event that there exists an infinite connected component is measurable.

[1] In different research disciplines, different terminology is used here: alternatives for sites and bonds are, e.g., vertices and edges (in a graph), or nodes and links (in a network).

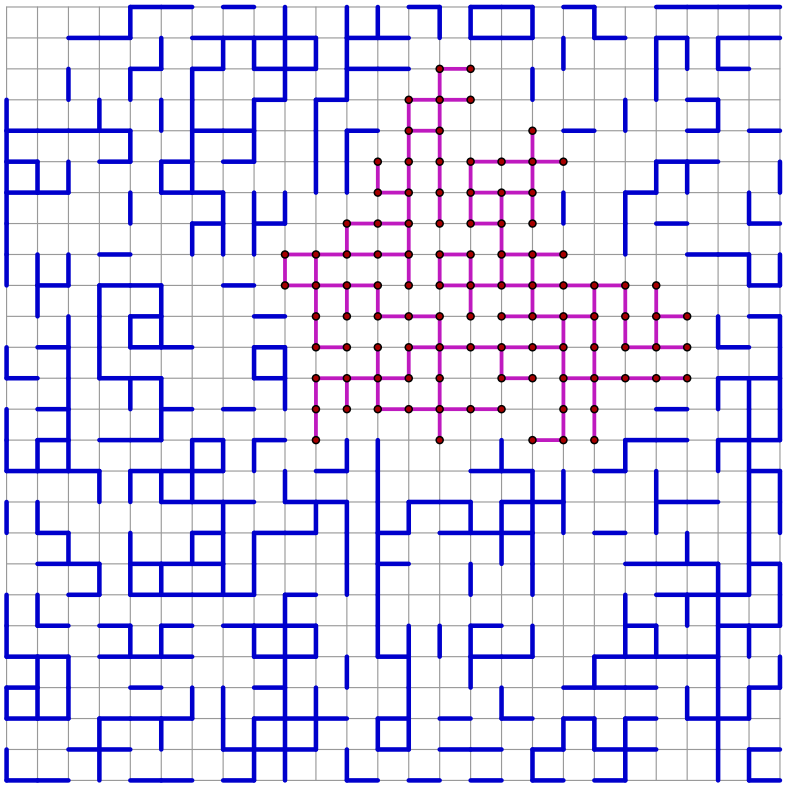


FIGURE II.2. A connected component in a bond percolation configuration on $\mathbb{Z}^2$.

As a warning against a too naive definition of what it means for a configuration to percolate, consider the following exercise.

Exercise II.2 (Too naive definition of the percolation event)**.**
For the bond percolation on $\mathbb{Z}^d$ with any parameter $p < 1$, show that

$$\mathsf{P}\Big[\text{all sites } x \in \mathbb{Z}^d \text{ belong to the same component}\Big] = 0.$$

Monotone coupling for percolation with different parameters

We introduced the probability measures P_p of percolation with a parameter $p \in [0, 1]$. Increasing p increases the probability that any given bond is open. It seems obvious that it also increases the probability of any connection $x \leftrightsquigarrow y$, and moreover that it increases the sizes of all clusters, etc. Below we make this intuition precise, in what is our first (and simplest) monotonicity result of the course.

Of course, not everything can be monotone increasing in p, so we need to abstract out a criterion which we believe should be a sufficient condition for monotonicity. The criterion below admits those events (and random variables) whose occurrence can not be undone (or whose values can not be decreased) in any configuration by opening up any additional bonds. The precise formulation involves a partial order on $\Omega = \{0, 1\}^{\mathrm{E}(\mathbb{Z}^d)}$.

The partial order $\preceq$ on the set of configurations $\Omega = \{0,1\}^{\mathrm{E}(\mathbb{Z}^d)}$ is a special case of Example D.4, and is defined as follows. A configuration $\omega \in \Omega$ is said to be smaller than a configuration $\omega' \in \Omega$, denoted $\omega \preceq \omega'$, if for all $e \in \mathrm{E}(\mathbb{Z}^d)$ we have $\omega_e \leq \omega'_e$.

A (measurable) function $f\colon \{0,1\}^{\mathrm{E}(\mathbb{Z}^d)} \to \mathbb{R}$ is said to be *increasing* (w.r.t. $\preceq$) if we have

$$f(\omega) \leq f(\omega') \qquad \text{whenever} \qquad \omega \preceq \omega'.$$

An event E is said to be *increasing* if its indicator $\mathbb{I}_E$ is an increasing function. This amounts to saying that changing any closed bonds to open could never undo the occurrence of the event E. The following monotonicity result is a corollary of a monotone coupling we construct in Proposition II.3 below.

Proposition II.1 (Monotonicity of percolation probabilities and expected values)**.** *If $f\colon \{0,1\}^{\mathrm{E}(\mathbb{Z}^d)} \to \mathbb{R}$ is an increasing function, then the expected value $\mathsf{E}_p[f]$ (w.r.t. measure P_p) is an increasing function of p. In particular, if E is an increasing event, then the probability $\mathsf{P}_p[E]$ is increasing in p.*

Example II.2 (Examples of monotonicity in percolation)**.**
All of the following are increasing functions of p:

$$p \mapsto \mathsf{P}_p\big[x \leftrightsquigarrow y\big]$$
$$p \mapsto \mathsf{P}_p\big[\text{there exists an infinite connected component}\big]$$
$$p \mapsto \mathsf{P}_p\big[\#\mathcal{C}_0 \geq s\big]$$
$$p \mapsto \mathsf{E}_p\big[\#\mathcal{C}_0\big].$$

The tool we use is a coupling of the probability measures P_p for different values of $p \in [0,1]$, which are discussed in a more general context in Appendix D. The key observation is that the components of the coupling can be made respect the partial order $\preceq$ on Ω.[2]

Proposition II.3 (Construction of monotone coupling for percolation)**.** *Let $0 \leq p_1 \leq p_2 \leq 1$. There exists a probability space $(\widetilde{\Omega}, \widetilde{\mathsf{P}}, \widetilde{\mathscr{F}})$ and $\{0,1\}^{\mathrm{E}(\mathbb{Z}^d)}$-valued random variables $\omega^{(1)}$ and $\omega^{(2)}$ on it such that the laws of $\omega^{(1)}$ and $\omega^{(2)}$ are P_{p_1} and P_{p_2}, respectively, and such that we have $\omega^{(1)} \preceq \omega^{(2)}$ pointwise on $\widetilde{\Omega}$.*

Proof. The idea is to take for each bond $e \in \mathrm{E}(\mathbb{Z}^d)$ an independent uniform random variable U_e on the unit interval $[0,1]$. Thus the sample space is $\widetilde{\Omega} = [0,1]^{\mathrm{E}(\mathbb{Z}^d)}$, a countable product of intervals. The probability measure $\widetilde{\mathsf{P}}$ is just the countable product of uniform measures on $[0,1]$, and the sigma algebra $\widetilde{\mathscr{F}}$ is the countable product sigma algebra.

We then set $\omega_e^{(1)} = \mathbb{I}_{\{U_e \leq p_1\}}$. The partial order relation $\omega^{(1)} \preceq \omega^{(2)}$ holds for these indicators pointwise on $\widetilde{\Omega}$. Clearly we have $\widetilde{\mathsf{P}}[\omega_e^{(1)} = 1] = p_1$, since U_e is uniformly distributed, and clearly $\omega_e^{(1)}$ are independent for different $e \in \mathrm{E}(\mathbb{Z}^d)$, because they are determined by the independent random variables U_e. Therefore the law of $\omega_e^{(1)}$ is P_{p_1}. Similarly, the law of $\omega^{(2)}$ is P_{p_2}. □

[2] In the sense of Definition D.8 we have that the measure with larger parameter p stochastically dominates the measure with smaller parameter, $\mathsf{P}_{p_1} \preceq_{\mathrm{st}} \mathsf{P}_{p_2}$ if $p_1 \leq p_2$.

Proposition II.1 would now in fact follow from general considerations of monotone couplings (Corollary D.7), but since this is our first encounter with the ideas of couplings and monotonicity, we spell out the short proof in detail.

Proof of Proposition II.1. Suppose that $f\colon \{0,1\}^{\mathrm{E}(\mathbb{Z}^d)} \to \mathbb{R}$ is increasing, and suppose that $0 \le p_1 < p_2 \le 1$. Use the monotone coupling of Proposition II.3 above, to write $\mathsf{E}_{p_j}[f] = \widetilde{\mathsf{E}}\big[f(\omega^{(j)})\big]$, where $\widetilde{\mathsf{E}}$ is the expected value w.r.t. $\widetilde{\mathsf{P}}$. Pointwise on $\widetilde{\Omega}$, we have $\omega^{(1)} \preceq \omega^{(2)}$ and thus $f(\omega^{(1)}) \le f(\omega^{(2)})$, since f is increasing. We get the conclusion

$$\mathsf{E}_{p_1}[f] = \widetilde{\mathsf{E}}\big[f(\omega^{(1)})\big] \le \widetilde{\mathsf{E}}\big[f(\omega^{(2)})\big] = \mathsf{E}_{p_2}[f].$$

□

Phase transition by a zero-one law

We will show that there is the following phase transition across a critical value $p_c \in [0,1]$ of the parameter p. The critical value depends on dimension d, and in more general context it would also depend on the graph which we consider, so for definiteness we sometimes write $p_c(\mathbb{Z}^d)$ for the critical value on $\mathbb{Z}^d$.

Theorem II.4 (Phase transition for the existence of infinite percolation cluster)**.** *Let $d \in \mathbb{Z}_{>0}$. Then there exists a $p_c = p_c(\mathbb{Z}^d) \in [0,1]$ such that:*

- *For $p < p_c$, almost surely all components $\mathcal{C}_x$, $x \in \mathbb{Z}^d$, are finite.*
- *For $p > p_c$, almost surely there exists an infinite connected component.*

Proof. First we remark that indeed there is a zero-one law. The event that there exists an infinite connected component does not depend on the states of any finite number of bonds. Thus this event belongs to the tail σ-algebra $\mathscr{T}_\infty$, see (E.1) (we can use any enumeration of the bonds, the tail σ-algebra does not depend on the choice). As a consequence of Kolmogorov's 0-1–law, Theorem E.1 in Appendix E, we thus have

$$\mathsf{P}_p\big[\text{there exists an infinite connected component}\big] \in \{0,1\}\,. \tag{II.1}$$

This probability is an increasing function of p by Proposition II.1, and since it can only assume values 0 or 1, it must remain 0 for p less than some p_c, and become 1 for p greater than p_c. □

Remark II.5 (What Theorem II.4 does not say)**.** The theorem above does not tell whether an infinite cluster exists at the critical point $p = p_c$ or not. This question is more subtle.

Example II.6 (The percolation critical point is trivial in one dimension)**.** Theorem II.4 also does not rule out the possibility that the transition happens at a trivial place $p_c = 0$ or $p_c = 1$. The latter triviality in fact happens in the one-dimensional case.

So consider now $d = 1$, i.e., percolation on $\mathbb{Z}$. Any n consecutive bonds are open with probability p^n. If $p < 1$, then $p^n \to 0$ as $n \to \infty$, so any given site almost surely does not belong to an infinite component. Since there is only a countable number of sites in $\mathbb{Z}$, the union bound shows that almost surely none of the components is infinite. If $p = 1$, however, almost surely all bonds are open, so the entire infinite graph $\mathbb{Z}$ forms a connected component. Thus in $d = 1$ we have $p_c(\mathbb{Z}) = 1$.

We will prove below that for $d \ge 2$ the phase transition occurs at a non-trivial critical value $0 < p_c < 1$.

Theorem II.7 (Non-triviality of percolation phase transition)**.**
Let $d \geq 2$. Then the critical point of percolation is nontrivial, $p_c(\mathbb{Z}^d) \in (0,1)$.

To prove this, we study the connected component $\mathcal{C}_0$ of the origin $0 \in \mathbb{Z}^d$, and especially its size $\#\mathcal{C}_0$. We denote by

$$\theta(p) = \mathsf{P}\big[\#\mathcal{C}_0 = \infty\big]$$

the probability that the origin belongs to an infinite open cluster. To obtain an equivalent characterization of the critical point p_c, we claim that $\theta(p)$ is positive if and only if the probability in (II.1) is one.

Lemma II.8 (Characterization of the phases by the cluster of the origin)**.**
We have the equivalence

$$\theta(p) > 0 \quad \Leftrightarrow \quad \mathsf{P}\big[\textit{there exists an infinite connected component}\big] = 1.$$

Proof. First of all, the probability (II.1) is clearly at least $\theta(p)$. Thus if $\theta(p) > 0$, then the probability can not be zero, so by the 0-1 law it has to be one. On the other hand, if $\theta(p) = 0$, i.e., the cluster of origin is almost surely finite, then by translation invariance the cluster of any point x is almost surely finite

$$\mathsf{P}\big[\#\mathcal{C}_x = \infty\big] = 0.$$

Since there are only countably many points, this allows us to conclude by the union bound that the probability of existence of an infinite cluster is

$$\mathsf{P}\big[\exists x \in \mathbb{Z}^d \text{ such that } \#\mathcal{C}_x = \infty\big] \leq \sum_{x\in\mathbb{Z}^d} \mathsf{P}\big[\#\mathcal{C}_x = \infty\big] = \sum_{x\in\mathbb{Z}^d} 0 = 0.$$

This concludes the proof of the lemma. □

The proof of II.7 will be done in two parts: we first show that $p_c > 0$ and then that $p_c < 1$.

Non-percolation for small p

Denote the set of self-avoiding lattice paths of n steps starting from the origin by

$$\Gamma_n = \Big\{\gamma = (\gamma(j))_{j=0}^n \;\Big|\; \gamma(0) = 0,\ \forall j : \gamma(j) \in \mathbb{Z}^d, \\ \forall j : \|\gamma(j) - \gamma(j-1)\| = 1,\ \forall j \neq k : \gamma(j) \neq \gamma(k)\Big\}.$$

The key observation for showing non-percolation for small p is to observe that for the cluster $\mathcal{C}_0$ of the origin to be infinite, we would need to have self-avoiding open paths of arbitrary length n.

Lemma II.9 (Non-percolation for small p)**.**
For $p < \frac{1}{2d}$ we have $\theta(p) = 0$.

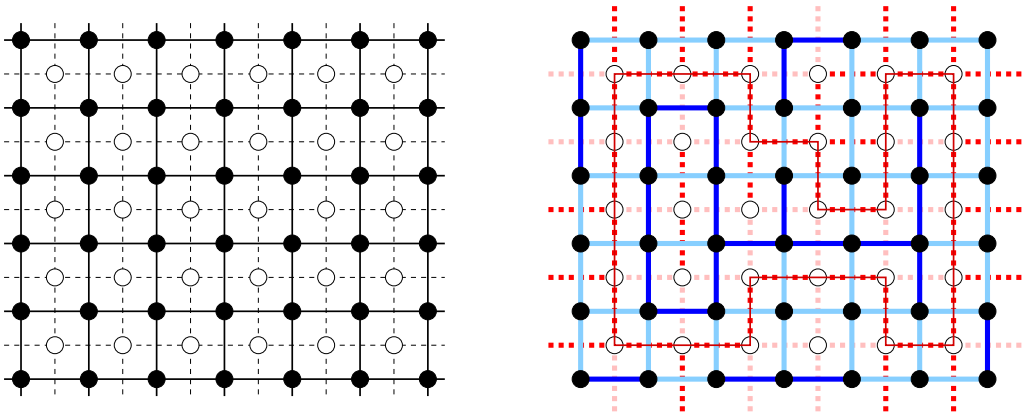
Proof. The number of self-avoiding paths of length n from the origin can obviously be bounded above by $\#\Gamma_n \le (2d)^n$. We then note that for any n

$$\begin{aligned}\theta(p) &= \mathsf{P}_p\Big[\#\mathcal{C}_0 = \infty\Big] \\ &\le \mathsf{P}_p\Big[\text{there exists } \gamma \in \Gamma_n \text{ s.t. } \omega_{\{\gamma(j-1),\gamma(j)\}} = 1 \text{ for } j = 1,\dots,n\Big] \\ &\le \sum_{\gamma\in\Gamma_n} \mathsf{P}_p\Big[\omega_{\{\gamma(j-1),\gamma(j)\}} = 1 \text{ for } j = 1,\dots,n\Big] = (\#\Gamma_n)\, p^n \le (2dp)^n.\end{aligned}$$

If $p < \frac{1}{2d}$, then this upper bound $(2dp)^n$ tends to zero as $n \to \infty$, showing $\theta(p) = 0$. □

Percolation for large p in two dimensions

We will first establish the existence of an infinite cluster for large p in the two-dimensional case, $d = 2$. In two dimensions we can exploit a duality argument. The (planar) dual lattice of $\mathbb{Z}^2$ is the lattice $(\mathbb{Z} + \frac{1}{2})^2$ of midpoints of the faces of the original square lattice $\mathbb{Z}^2$, and its bonds are $\{q, r\}$ with $q, r \in (\mathbb{Z} + \frac{1}{2})^2$ such that $\|q - r\| = 1$. Note that each such dual bond $\{q, r\}$ crosses exactly one bond $\{z, w\}$ of $\mathbb{Z}^2$ (the one with the same midpoint, $\frac{z+w}{2} = \frac{q+r}{2}$). A percolation configuration $\omega \in \{0,1\}^{\mathrm{E}(\mathbb{Z}^2)}$ on $\mathbb{Z}^2$ determines a dual configuration $\omega^* \in \{0,1\}^{\mathrm{E}((\mathbb{Z}+\frac{1}{2})^2)}$ by the rule $\omega^*_{\{q,r\}} = 1 - \omega^*_{\{z,w\}}$ — a dual edge crossing an open edge is declared closed, and vice versa.



(a) The square lattice $\mathbb{Z}^2$ and its dual $(\mathbb{Z}+\frac{1}{2})^2$.

(b) A finite open cluster is surrounded by an open circuit in the dual configuration.

FIGURE II.3. Duality of the planar lattice $\mathbb{Z}^2$ and of a percolation configuration on it.

The next key observation is illustrated in Figure II.3(b) — the finiteness of the component of the origin is equivalent to the existence of a dual open circuit surrounding the origin. By a dual circuit we mean a simple[3] closed path[4] of the dual lattice $(\mathbb{Z}+\frac{1}{2})^2$, modulo cyclic reparametrization[5] and orientation reversal[6]. Indeed,

[3] A simple path means a path with no self-intersections.

[4] A closed path means one, whose starting point and end point coincide.

[5] Any point on the closed path could be chosen as the starting point, and we consider the all such choices equal.

[6] A path traversed in the opposite direction is considered equal to the original path.

if the dual configuration ω^* contains an open circuit γ^* surrounding the origin, then the cluster $\mathcal{C}_0$ of the origin is finite (its has to be contained in the interior of the dual open circuit γ^*, because the bonds crossing the dual bonds of the circuit are all closed). Also the converse holds, if the cluster $\mathcal{C}_0$ is finite, then there is a open circuit γ^* surrounding the origin in the dual configuration ω^* (the duals of the bonds connecting the cluster to its complement form such a circuit).

The set of dual circuits of length n that surround the origin is denoted by Γ_n°. As an upper bound on their number, surround the origin, we use $\#\Gamma_n^\circ \leq \frac{n}{2}4^n$. This is seen by noting that a circuit has to cross the positive real axis at some minimal point $k-\frac{1}{2}$ with $k \in \mathbb{Z}_{>0}$, and surrounding origin requires length $n \geq 2k+2$. This minimal intersection point $k-\frac{1}{2}$ can thus be chosen in at most $\frac{n}{2}$ ways. The circuit can be viewed as an n step lattice path starting from just below $k-\frac{1}{2}$, and as such there are at most 4^n choices for it.

Lemma II.10 (Percolation in two dimensions for large p)**.**
For $d=2$ and $p > \frac{3}{4}+\frac{1}{4\sqrt{2}}$, we have $\theta(p)>0$.

Proof. We estimate

$$\begin{aligned}
1-\theta(p) &= \mathsf{P}_p\Big[\#\mathcal{C}_0 < \infty\Big] \\
&= \mathsf{P}_p\Big[\text{for some } n \text{ there exists dual open circuit } \gamma^* \in \Gamma_n^\circ\Big] \\
&\leq \sum_n \sum_{\gamma^* \in \Gamma_n^\circ} \mathsf{P}_p\Big[\gamma^* \text{ is dual open}\Big] = \sum_n (\#\Gamma_n^\circ)\,(1-p)^n \\
&\leq \frac{1}{2}\sum_n n\big(4(1-p)\big)^n.
\end{aligned}$$

If moreover $p > \frac{3}{4}$, so that $4(1-p)<1$, then we can estimate the last series by the derivative of a geometric series, and get

$$1-\theta(p) \leq \frac{1}{2}\frac{1}{\big(1-4(1-p)\big)^2}.$$

If $p > \frac{3}{4}+\frac{1}{4\sqrt{2}} \approx 0.927$, then this upper bound is strictly less than one, showing $\theta(p)>0$. □

Exercise II.3 (Numerical improvements for Lemma II.10)**.**
The above lemma shows that $\theta(p)>0$ for $p \gtrsim 0.927$ (correspondingly, one can conclude that $p_c(\mathbb{Z}^2) \lesssim 0.927$). Improve the combinatorial estimates in order to establish the same conclusion starting already from a smaller numerical value of p. (E.g., you can try to prove that $\theta(p)>0$ for $p \gtrsim 0.8156$)

Proof of non-triviality of the critical point

We now combine the above results to prove Theorem II.7.

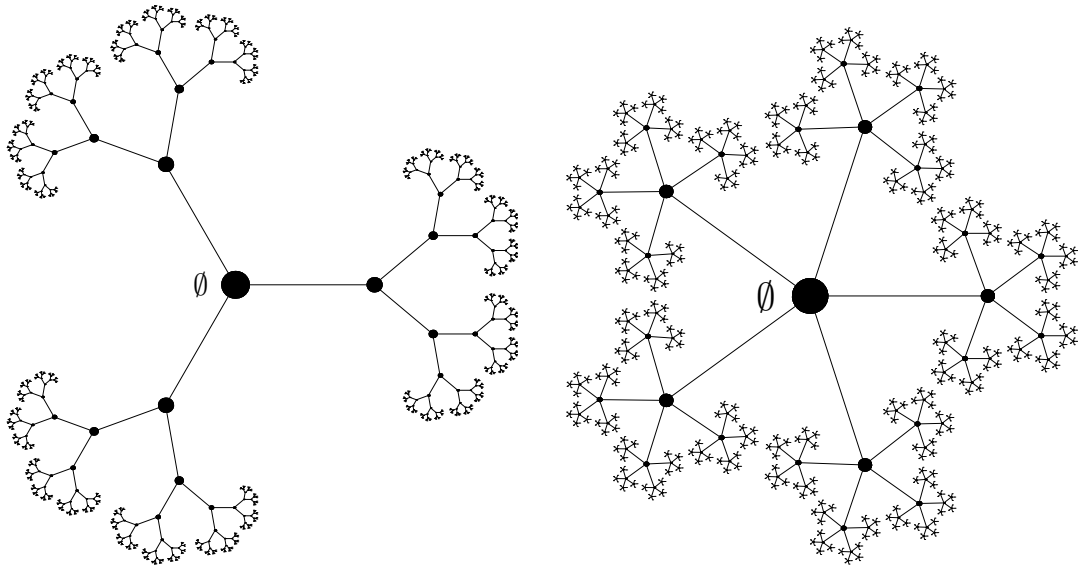
Proof of Theorem II.7. By Lemma II.9 we have $\theta(p)=0$ for $p < \frac{1}{2d}$. By Lemma II.8 this implies $p_c \geq \frac{1}{2d}$.

It remains to show that $p_c < 1$. For $d \geq 2$, the d-dimensional percolation on $\mathbb{Z}^d$ contains a two-dimensional percolation configuration on $\mathbb{Z}^2$ — if the latter has an infinite open cluster, then the former must have one as well. Thus it is sufficient to consider $d=2$. By Lemma II.10 we have $\theta(p)>0$ for $p > \frac{3}{4}+\frac{1}{4\sqrt{2}}$. By Lemma II.8 this implies $p_c \leq \frac{3}{4}+\frac{1}{4\sqrt{2}} < 1$. □

2. Percolation on a regular infinite tree

The model of percolation can be defined on any graph. On a regular infinite tree $\mathcal{T}$ of degree B, depicted in Figure II.4, percolation becomes particularly simple. In fact the analysis is essentially the the same as that of a Galton-Watson branching process.

In the case $B = 2$ the tree coincides with the integer lattice, $\mathcal{T} = \mathbb{Z}$, and the triviality of this model was discussed already. Interesting cases are when the degree is large enough, $B \geq 3$. Percolation on a regular infinite tree $\mathcal{T}$ of any degree $B \geq 3$ behaves in some sense like percolation on $\mathbb{Z}^d$ for a very high dimension d.



(a) Tree of degree three. (b) Tree of degree five.

FIGURE II.4. Regular infinite trees.

Formally, the tree $\mathcal{T}$ can be defined as follows. The set of its sites is the set of finite words $(b_1, b_2, \ldots, b_\ell)$ of any possible length $\ell \in \mathbb{Z}_{\geq 0}$, with first letter $b_1 \in \{1, 2, \ldots, B\}$ and other letters $b_j \in \{1, 2, \ldots, B-1\}$ for $j = 2, 3, \ldots, \ell$. The word of 0 letters is denoted by $\emptyset$, and it corresponds to a chosen root of the tree (e.g. the vertex in the middle of Figure II.4). Two vertices are adjacent if the word of one is the extension of the word of the other by one letter. The distance of a site from the root is the length of the corresponding word. In the percolation model, the pairs of adjacent vertices are declared open or closed, with probabilities p and $1 - p$, independently.

Exercise II.4 (Expected size of the percolation cluster of the root of the tree)**.**
Consider percolation with parameter $p \in [0, 1]$ on the regular infinite tree $\mathcal{T}$ of degree $B \geq 3$. Denote by $\mathcal{C}_\emptyset$ the connected component of the root vertex $\emptyset$, and by $\#\mathcal{C}_\emptyset$ its size. Calculate $\mathsf{E}[\#\mathcal{C}_\emptyset]$ as a function of p. At which value of p does $\mathsf{E}[\#\mathcal{C}_\emptyset]$ become infinite?

Exercise II.5 (Critical point for percolation on regular infinite tree)**.**
Consider percolation with parameter p on the regular infinite tree $\mathcal{T}$ of degree $B \geq 3$. Denote by $\mathcal{C}_\emptyset(r)$ the set of sites at distance r from the root which are connected to the root, and denote by $\mathcal{C}_\emptyset$ the set of all sites which are connected to the root. Define also $\theta(p) = \mathsf{P}_p[\#\mathcal{C}_\emptyset = \infty]$.

(a) Show that if $p < \frac{1}{B-1}$, then $\theta(p) = 0$ (use the previous exercise).
(b) For all $r \in \mathbb{N}$, define the generating functions $G_r(z) = \mathsf{E}[z^{\#\mathcal{C}(r)}]$, where $z \in (0, 1]$. Show that $G_{r+1}(z) = G_r(g(z))$, where g is the generating function $g(z) = \mathsf{E}[z^N]$ of a binomial random variable $N \sim \mathrm{Bin}(B-1, p)$.
(c) Show that we may express the probability of the root being connected to some site at distance r as follows: $\mathsf{P}[\#\mathcal{C}(r) > 0] = 1 - \lim_{z \searrow 0} G_r(z)$.
(d) Let g be as in part (b). Show that the sequence $0,\ g(0),\ g(g(0)),\ g(g(g(0))),\ \ldots$ of iterates converges to the smallest fixed point of g in $(0, 1]$.
(e) Show that g defined in part (b) has a fixed point $\xi < 1$ if $p > \frac{1}{B-1}$.
(f) Show that if $p > \frac{1}{B-1}$, then $\theta(p) > 0$.

Lecture III

Law of iterated logarithm

With Borel-Cantelli lemmas, Lemma E.3 in Appendix E, we will prove the following statement about the almost sure behavior of a random walk after a large number of steps. The same result would hold for the symmetric simple random walk with steps ± 1. but to simplify the proof we assume instead that the steps have Gaussian distribution. Thus we set

$$X_n = \sum_{k=1}^{n} \xi_k, \qquad \text{(III.1)}$$
$$\text{where } (\xi_k)_{k\in\mathbb{N}} \text{ are i.i.d. and } \xi_k \sim \mathrm{N}(0,1).$$

Theorem III.1 (Law of iterated logarithm)**.**
Set $\lambda(n) = \sqrt{n\log(\log(n))}$. Let $(X_n)_{n\in\mathbb{Z}_{\geq 0}}$ be the random walk with Gaussian steps as above. Then we have, almost surely,

$$\limsup_{n\to\infty} \frac{X_n}{\lambda(n)} = \sqrt{2}.$$

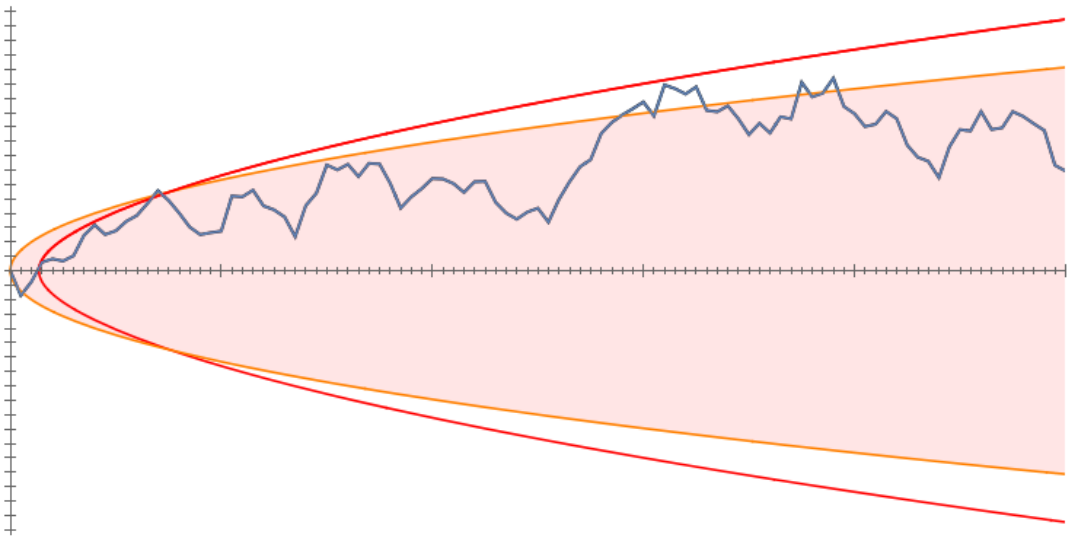


FIGURE III.1. A random walk with Gaussian steps. The shaded region bounded by the orange curves is $|x| \leq \sqrt{2n}$, and the red curves are the law of iterated logarithm tresholds $\pm\sqrt{2}\lambda(n) = \pm\sqrt{2n\log\log(n)}$.

1. Auxiliary estimates

We will need estimates about the tail of the distribution of X_n. In the Gaussian steps case, these reduce to the following simple Lemma.

Lemma III.2 (Tail estimates for the standard normal distribution)**.**
Let $\xi \sim \mathrm{N}(0,1)$ be a standard Gaussian random variable. Then we have, for any $x > 0$,

$$\mathsf{P}\big[\xi > x\big] \le e^{-\frac{1}{2}x^2} \tag{III.2}$$

Also, for any $C < \frac{1}{\sqrt{2\pi}}$ there exists x_0 such that for all $x > x_0$ we have

$$\mathsf{P}\big[\xi > x\big] \ge Cx^{-1}e^{-\frac{1}{2}x^2}. \tag{III.3}$$

Proof. For any $\theta > 0$, by Markov inequality, we get

$$\mathsf{P}\big[\xi > x\big] = \mathsf{P}\big[e^{\theta\xi} > e^{\theta x}\big] \le e^{-\theta x}\,\mathsf{E}\big[e^{\theta\xi}\big].$$

A direct calculation gives $\mathsf{E}[e^{\theta\xi}] = e^{\frac{1}{2}\theta^2}$, which we substitute above to obtain

$$\mathsf{P}\big[\xi > x\big] \le e^{\frac{1}{2}\theta^2 - \theta x}.$$

Now it is a matter of optimizing θ to get the best inequality: the minimum of $\theta \mapsto \frac{1}{2}\theta^2 - \theta x$ is at $\theta = x$, and substituting this value, we obtain the desired inequality $\mathsf{P}[\xi > x] \le e^{-\frac{1}{2}x^2}$.

The second inequality is merely calculus. Write

$$\mathsf{P}\big[\xi > x\big] = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}s^2}\,\mathrm{d}s.$$

Then note that $\frac{\mathrm{d}}{\mathrm{d}s}(s^{-1}e^{-\frac{1}{2}s^2}) = -s^{-2}e^{-\frac{1}{2}s^2} - e^{-\frac{1}{2}s^2}$, and use this to do integration by parts

$$\int_x^\infty e^{-\frac{1}{2}s^2}\,\mathrm{d}s = \frac{1}{x}e^{-\frac{1}{2}x^2} - \int_x^\infty s^{-2}e^{-\frac{1}{2}s^2}\,\mathrm{d}s.$$

We see that

$$\frac{\int_x^\infty e^{-\frac{1}{2}s^2}\,\mathrm{d}s}{x^{-1}e^{-\frac{1}{2}x^2}} \longrightarrow 1 \qquad \text{as } x \to \infty.$$

Therefore, for any $C < 1$ we have, for large enough x

$$\int_x^\infty e^{-\frac{1}{2}s^2}\,\mathrm{d}s \ge Cx^{-1}e^{-\frac{1}{2}x^2},$$

which upon dividing by $\sqrt{2\pi}$ becomes the second asserted inequality. □

Exercise III.1 (More precise tail estimates for the standard normal distribution)**.**
Use integration by parts as in the proof above, to derive the following more precise asymptotics of Gaussian tails: for any $x > 0$ show that

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\exp\left(-\frac{x^2}{2}\right) \le \int_x^\infty \exp\left(-\frac{y^2}{2}\right)\mathrm{d}y \le \frac{1}{x}\exp\left(-\frac{x^2}{2}\right).$$

With Gaussian steps as in Equation III.7, we have $X_n \sim \mathrm{N}(0,n)$. Thus we can immediately translate the above to bounds on the tail of X_n.

Lemma III.3 (Tail estimates for the random walk with Gaussian steps)**.**
For any $x > 0$ and $n \ge 1$, we have

$$\mathsf{P}\big[X_n > x\big] \le e^{-\frac{1}{2n}x^2} \tag{III.4}$$

and

$$\mathsf{P}\big[|X_n| > x\big] \;\le\; 2\,e^{-\frac{1}{2n}x^2}. \tag{III.5}$$

For any $C < \frac{1}{\sqrt{2\pi}}$ *there exists* $x_0 > 0$ *such that for all* $n \ge 1$ *and all* $x \ge \sqrt{n}x_0$ *we have*

$$\mathsf{P}\big[X_n > x\big] \;\ge\; C\frac{\sqrt{n}}{x}e^{-\frac{1}{2n}x^2}. \tag{III.6}$$

Exercise III.2 (Tail estimate for the simple random walk)**.**
Let now $(X_n)_{n\in\mathbb{Z}_{\ge 0}}$ be the symmetric simple random walk with ± 1 steps, i.e.

$$X_n = \sum_{k=1}^{n}\xi_k, \tag{III.7}$$

$$\text{where } (\xi_k)_{k\in\mathbb{N}} \text{ are i.i.d. and } \mathsf{P}\big[\xi_k = +1\big] = \frac{1}{2},\ \mathsf{P}\big[\xi_k = -1\big] = \frac{1}{2}.$$

Prove that

$$\mathsf{P}\big[X_n > x\big] \le e^{-\frac{1}{2n}x^2}.$$

We will also need auxiliary calculations about the function

$$\lambda(n) = \sqrt{n\,\log\big(\log(n)\big)},$$

which are done in the following exercise.

Exercise III.3 (Calculations with the law of iterated logarithm scale function)**.**
Let $\alpha > 1$ and for $k \in \mathbb{N}$ let $n_k = \lfloor \alpha^k \rfloor$ be the integer part of the number α^k. Calculate the limits

$$\lim_{k\to\infty}\frac{n_k}{\alpha^k}, \qquad \lim_{k\to\infty}\frac{\lambda(n_k)}{\lambda(n_{k+1})}, \qquad \lim_{k\to\infty}\frac{\lambda(n_{k+1}-n_k)}{\lambda(n_{k+1})}.$$

2. Upper bound

In Theorem III.1 we claim that almost surely,

$$\limsup_{n\to\infty}\frac{X_n}{\lambda(n)} = \sqrt{2},$$

where $\lambda(n) = \sqrt{n\log(\log(n))}$. Our first goal is to establish the upper bound,

$$\limsup_{n\to\infty}\frac{X_n}{\lambda(n)} \le \sqrt{2}.$$

We proceed in two steps — first treat an exponentially growing subsequence of values of n, and then fill in the gaps.

The result about an exponentially growing subsequence is a simple application of Borel-Cantelli lemma and the upper bound for the tail of X_n given in Lemma III.3.

Lemma III.4 (Law of iterated logarithm upper bound along subsequences)**.**
Let $\alpha > 1$*, and denote by* $n_k = \lfloor \alpha^k \rfloor$ *the integer part of* α^k*. Then we have, for any* $\beta > \sqrt{2}$*, almost surely*

$$|X_{n_k}| \le \beta\lambda(n_k)$$

except possibly for finitely many values of $k \in \mathbb{N}$. In other words, we have

$$\limsup_k \frac{|X_{n_k}|}{\lambda(n_k)} \leq \beta \qquad \textit{almost surely.}$$

Proof. By the inequality (III.5) we have

$$\mathsf{P}\Big[\frac{|X_n|}{\lambda(n)} > \beta\Big] \leq 2\exp\big(-\frac{\beta^2\lambda(n)^2}{2n}\big) = 2\log(n)^{-\frac{\beta^2}{2}}.$$

The sequence defined by $n_k = \lfloor \alpha^k \rfloor$ grows exponentially, $n_k \geq C\alpha^k$, and thus $\log(n_k)$ grows linearly, $\log(n_k) \geq ak + c$ with $a = \log(\alpha) > 0$. For $\beta > \sqrt{2}$ the following series is summable

$$\sum_{k=k_0}^{\infty} \mathsf{P}\Big[\frac{|X_{n_k}|}{\lambda(n_k)} > \beta\Big] \leq 2\sum_{k=k_0}^{\infty} \log(n_k)^{-\frac{\beta^2}{2}} \leq 2\sum_{k=k_0}^{\infty} (ak+c)^{-\frac{\beta^2}{2}} < \infty.$$

The first Borel–Cantelli Lemma (Lemma E.3) then guarantees that with probability one (i.e. almost surely) we have $\frac{|X_{n_k}|}{\lambda(n_k)} > \beta$ only for finitely many $k \in \mathbb{N}$. □

To fill in the gaps between n_k and n_{k+1}, we use the following auxiliary result.

Lemma III.5 (Controlling the random walk in the gaps of a subsequence)**.** *For any $\varepsilon > 0$ there exists an $\alpha > 1$ such that for the subsequence defined by $n_k = \lfloor \alpha^k \rfloor$ we have (almost surely)*

$$\max_{n_k \leq i \leq n_{k+1}} |X_i - X_{n_k}| \leq 2\varepsilon\lambda(n_k)$$

except possibly for finitely many values of $k \in \mathbb{N}$.

Let us postpone the proof of this key technical lemma a little, and first see how it is used to finish the proof of the upper bound in the law of iterated logarithm.

Proposition III.6 (Law of iterated logarithm upper bound)**.** *We have*

$$\mathsf{P}\Big[\limsup_{n\to\infty} \frac{X_n}{\lambda(n)} \leq \sqrt{2}\Big] = 1.$$

Proof. Let $\beta > \sqrt{2}$ and $\varepsilon > 0$. Let us choose $\alpha > 1$ as in Lemma III.5, so as to obtain for the exponentially growing subsequence $n_k = \lfloor \alpha^k \rfloor$ almost surely

$$\max_{n_k \leq i \leq n_{k+1}} |X_i - X_{n_k}| \leq 2\varepsilon\lambda(n_k),$$

except possibly for finitely many $k \in \mathbb{N}$. By Lemma III.4, we have almost surely

$$|X_{n_k}| \leq \beta\lambda(n_k)$$

except possibly for finitely many $k \in \mathbb{N}$. For $i \in \mathbb{N}$, $n_k \leq i \leq n_{k+1}$, we write

$$\frac{|X_i|}{\lambda(i)} \leq \frac{|X_i|}{\lambda(n_k)} \leq \frac{|X_{n_k}| + |X_i - X_{n_k}|}{\lambda(n_k)},$$

and unless k is one of the finitely many exceptional values, the two previous estimates give

$$\frac{|X_i|}{\lambda(i)} \leq \beta + 2\varepsilon.$$

Almost surely this holds for all except possibly finitely many values of i, since an exceptional i must be between n_k and n_{k+1} for an exceptional k. We then conclude

$$\limsup_i \frac{|X_i|}{\lambda(i)} \leq \beta + 2\varepsilon.$$

Since $\beta > \sqrt{2}$ and $\varepsilon > 0$ were arbitrary, we get, almost surely,

$$\limsup_i \frac{|X_i|}{\lambda(i)} \le \sqrt{2}.$$

□

It remains to prove Lemma III.5. For this purpose we use a general inequality about the maxima of random walks with independent steps.

Lemma III.7 (Lévy's inequality)**.**
Suppose that $(\zeta_u)_{u\in\mathbb{N}}$ are independent. Set $S_t = \sum_{u=1}^t \zeta_u$. Fix $m \in \mathbb{N}$ and consider $\max_{1\le t\le m} S_t$. Assume that for some $\sigma > 0$ and $\rho > 0$ we have, for all $t \le m$, $\mathsf{P}\big[|S_m - S_t| \le \sigma\big] \ge \rho$. Then we have

$$\mathsf{P}\Big[\max_{1\le t\le m} |S_t| > 2\sigma\Big] \le \frac{1}{\rho}\,\mathsf{P}[|S_m| > \sigma].$$

Proof. We will split according to the first time t at which $S_t > 2\sigma$. Define the events

$$\begin{aligned} A_t &= \{|S_1| \le 2\sigma, \dots, |S_{t-1}| \le 2\sigma, |S_t| > 2\sigma\} \\ B_t &= \{|S_m - S_t| \le \sigma\}. \end{aligned}$$

The events A_t and B_t are independent (the former depends on $\zeta_1, \dots, \zeta_t$ and the latter on $\zeta_{t+1}, \dots, \zeta_m$), and the events $A_1, \dots, A_n$ are disjoint. Since $\{|S_m| > \sigma\} \supset \bigcup_{t=1}^m (A_t \cap B_t)$, we get

$$\mathsf{P}\Big[|S_m| > \sigma\Big] \ge \sum_{t=1}^m \mathsf{P}[A_t]\,\mathsf{P}[B_t] \ge \rho \sum_{t=1}^m \mathsf{P}[A_t] = \rho\,\mathsf{P}\Big[\max_{1\le t\le m} |S_t| > 2\sigma\Big].$$

Dividing by ρ yields the result. □

Proof of Lemma III.5. Let $\varepsilon > 0$. It will turn out that as long as we choose $\alpha < 1 + \frac{\varepsilon^2}{2}$, we get the desired conclusion for the gaps in the subsequence $n_k = \lfloor \alpha^k \rfloor$.

To control what happens between n_k and n_{k+1}, we want to bound the probability

$$\mathsf{P}\Big[\max_{n_k < i \le n_{k+1}} |X_i - X_{n_k}| > 2\varepsilon\lambda(n_k)\Big].$$

The bound will be derived from Lévy's inequality (Lemma III.7) with

$$S_t = X_{n_k+t} - X_{n_k}, \qquad m = n_{k+1} - n_k, \qquad \sigma = \varepsilon\lambda(n_k).$$

In order to find an appropriate value of ρ, denote

$$\delta_k = \max_{n_k < i \le n_{k+1}} \mathsf{P}\big[|X_i - X_{n_k+1}| > \varepsilon\lambda(n_k)\big],$$

and observe that then $\rho = 1 - \delta_k$ will work, if it can be shown to be positive, i.e., if $\delta_k < 1$. Lévy's inequality then gives

$$\begin{aligned} &\mathsf{P}\Big[\max_{n_k < i \le n_{k+1}} |X_i - X_{n_k}| > 2\varepsilon\lambda(n_k)\Big] \\ &\le \frac{\mathsf{P}\big[|X_{n_k} - X_{n_k+1}| > \varepsilon\lambda(n_k)\big]}{\min_{n_k < i \le n_{k+1}} \mathsf{P}\big[|X_i - X_{n_k+1}| \le \varepsilon\lambda(n_k)\big]} \le \frac{\delta_k}{1-\delta_k}. \end{aligned}$$

We estimate δ_k by the inequality (III.5)

$$\begin{aligned} \delta_k &\le \max_{n_k < i \le n_{k+1}} \Big(2\exp\big(-\frac{\varepsilon^2\,\lambda(n_k)^2}{2(n_{k+1}-i)}\big)\Big) \\ &\le 2\exp\big(-\frac{\varepsilon^2\, n_k \log(\log(n_k))}{2(n_{k+1}-n_k)}\big) = 2\log(n_k)^{-\frac{\varepsilon^2 n_k}{2(n_{k+1}-n_k)}}. \end{aligned}$$

We have $\frac{n_k}{n_{k+1}-n_k} \to \frac{1}{\alpha-1}$ and $\frac{\log(n_k)}{k} \to \log(\alpha) > 0$, as $k \to \infty$. With $\alpha < 1 + \frac{\varepsilon^2}{2}$, we first of all have $\delta_k \to 0$ and moreover we again get summable probabilities

$$\sum_{k=k_0}^{\infty} \mathsf{P}\Big[\max_{n_k < i \le n_{k+1}} |X_i - X_{n_k}| > 2\varepsilon\lambda(n_k)\Big] \le \sum_{k=k_0}^{\infty} \frac{\delta_k}{1-\delta_k} < \infty.$$

The first Borel-Cantelli lemma thus implies that almost surely the undesired event

$$\max_{n_k < i \le n_{k+1}} |X_i - X_{n_k}| > 2\varepsilon\lambda(n_k)$$

can only occur for finitely many $k \in \mathbb{N}$. □

3. Lower bound

In Theorem III.1 we claim that almost surely,

$$\limsup_{n\to\infty} \frac{X_n}{\lambda(n)} = \sqrt{2},$$

where $\lambda(n) = \sqrt{n \log(\log(n))}$. We have established the upper bound in the previous section, so the remaining task is to establish the lower bound

$$\limsup_{n\to\infty} \frac{X_n}{\lambda(n)} \ge \sqrt{2}.$$

Because of the lim sup, it is enough to just find some subsequence of indices n for which the walk has large values. We will use an exponentially growing subsequence, as before.

So again for $\alpha > 1$, we set $n_k = \lfloor \alpha^k \rfloor$, the largest integer less than α^k. We then have

$$\frac{n_k}{\alpha^k} \longrightarrow 1, \qquad \text{as } k \to \infty.$$

The idea is to use the lower bound (III.6), which gives

$$\begin{aligned} \mathsf{P}\Big[\frac{X_n}{\lambda(n)} > \beta\Big] &\ge C\frac{\sqrt{n}}{\beta\lambda(n)} \exp\big(-\frac{\beta^2\lambda(n)^2}{2n}\big) \\ &= \frac{C}{\beta}\frac{1}{\sqrt{\log(\log(n))}} \log(n)^{-\beta^2/2}. \end{aligned}$$

When $\beta < \sqrt{2}$, the sum of these probabilities along the subsequence (n_k) diverges. We are almost in a position to use the converse Borel-Cantelli lemma, except that X_{n_k} are not really independent. Therefore, we consider the increments

$$R_k = X_{n_{k+1}} - X_{n_k}, \tag{III.8}$$

instead. The sequence $(R_k)_{k\in\mathbb{N}}$ is independent, and the law of R_k is the same as the law of X_{Δ_k}, where

$$\Delta_k = n_{k+1} - n_k. \tag{III.9}$$

Also these time increments Δ_k are exponentially growing,

$$\frac{\Delta_k}{\alpha^k} \longrightarrow \alpha - 1 > 0, \qquad \text{as } k \to \infty. \tag{III.10}$$

Carefully combining these observations leads to a proof of the lower bound.

Proposition III.8 (Law of iterated logarithm lower bound)**.**
We have

$$\mathsf{P}\Big[\limsup_{n\to\infty} \frac{X_n}{\lambda(n)} \geq \sqrt{2}\Big] = 1.$$

Proof. Let $\beta < \sqrt{2}$. Recall that R_k is distributed like X_{Δ_k}, and use the inequality (III.6) to estimate

$$\mathsf{P}\Big[\frac{R_k}{\lambda(\Delta_k)} > \beta\Big] \geq \frac{C}{\beta} \frac{1}{\sqrt{\log(\log(\Delta_k))}} \log(\Delta_k)^{-\beta^2/2}$$

(at least for for k large enough). By (III.10) we see that for large enough k then $\mathsf{P}\big[R_k/\lambda(\Delta_k) > \beta\big] \geq \text{const.} \times k^{-1}$, and in particular

$$\sum_{k=1}^{\infty} \mathsf{P}\Big[\frac{R_k}{\lambda(\Delta_k)} > \beta\Big] = \infty.$$

By the independence of $(R_k)_{k\in\mathbb{N}}$ and the second Borel-Cantelli lemma, we conclude that $R_k/\lambda(\Delta_k) > \beta$ for infinitely many values of k.

However, because of the upper bound established in Lemma III.4 (applied to $-X_{n_k}$), we have, for any $\beta' > \sqrt{2}$ almost surely

$$X_{n_k} \geq -\beta'\lambda(n_k)$$

for all but finitely many k. Thus there are (almost surely) infinitely many values of k for which both of the above hold, and for these values of k we see that

$$X_{n_{k+1}} = X_{n_k} + R_k \geq -\beta'\lambda(n_k) + \beta\lambda(\Delta_k).$$

Here, note that

$$\frac{-\beta'\lambda(n_k) + \beta\lambda(\Delta_k)}{\lambda(n_{k+1})} \underset{k\to\infty}{\longrightarrow} -\beta'\sqrt{\frac{1}{\alpha}} + \beta\sqrt{\frac{\alpha-1}{\alpha}} \underset{\alpha\to\infty}{\longrightarrow} \beta.$$

Therefore, for any $\varepsilon > 0$ we can choose α large enough so that for large enough k we have $\beta\lambda(\Delta_k) - \beta'\lambda(n_k) > (\beta-\varepsilon)\lambda(n_{k+1})$. From the above argument we get

$$\limsup_{k\to\infty} \frac{X_{n_{k+1}}}{\lambda(n_{k+1})} \geq \beta - \varepsilon.$$

Since $\beta < \sqrt{2}$ and $\varepsilon > 0$ were arbitrary, this concludes the proof. □

4. Proof of Theorem III.1

The combination of Propositions III.6 and III.8 yields Theorem III.1.

Exercise III.4 (Law of iterated logarithm for simple random walk)**.**
Prove Theorem III.1 for the simple symmetric random walk with ±1 steps.

Lecture IV

Weak convergence on the real line

Weak convergence[1] (of probability measures) is a notion that allows one to speak of limits of distributions of random variables — as opposed to just limits of probabilities of certain events, or limits of certain expected values.

The notion of weak convergence not only theoretically convenient, but the underlying idea is also philosophically transparent: random variables are said to converge if all sufficiently well behaved observables of them converge. In the mathematical idealization, sufficiently well behaved observables are taken to mean expected values of bounded continuous functions on the appropriate space.

We already remark that weak convergence of a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables makes sense whenever all X_n take values in the same topological space $\mathfrak{X}$, since the definition of weak convergence only requires a notion of continuous functions. In this chapter, however, we only discuss real-valued random variables. The real valued case is often important in practice, provides a more concrete introduction, and allows certain useful approaches such as characteristic functions and cumulative distribution functions that are not available generally. We return to the general theory in Lecture VI.

Recall that the *law* (or *distribution*) of a real-valued random variable X is a Borel probability measure ν_X on the real axis $\mathbb{R}$, defined by $\nu_X[B] = \mathsf{P}[X \in B]$. We make the following comment on terminology. Below the weak convergence of probability measures is defined and studied. By weak convergence of a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables we then mean the weak convergence of the sequence of measures $(\nu_{X_n})_{n\in\mathbb{N}}$ (the laws of the random variables). Note also that in such a setting, the random variables $(X_n)_{n\in\mathbb{N}}$ need not be defined on the same probability space, their laws are measures on $\mathbb{R}$ anyway.

1. The idea and definition of weak convergence

It is natural to say that a sequence of systems converge, if any observable that we can reliably measure about the systems converges. We idealize the situation by declaring that the reliably measurable observables are the expected values of sufficiently well behaved functions of the random state of the system. Note that for example probabilities of events, or means and variances of random variables can often (but not always) be expressed in terms of such expected values. For weak convergence specifically, as the idealization of sufficiently well behaved observables,

[1]Alternative terminology for the same notion is convergence in distribution, convergence in law, and in a suitable functional analysis setup also weak-$*$ convergence.

we take the bounded continuous functions (continuity guarantees in particular Borel-measurability and boundedness guarantees integrability, and thus the existence of the expected values).

Definition IV.1 (Weak convergence of probability measures on $\mathbb{R}$)**.**
A sequence $(\nu_n)_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ converges weakly to a probability measure ν_∞ if for all bounded continuous functions $f\colon \mathbb{R}\to\mathbb{R}$ we have

$$\int_{\mathbb{R}} f \,\mathrm{d}\nu_n \underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} f \,\mathrm{d}\nu_\infty. \tag{IV.1}$$

We then denote $\nu_n \overset{\mathrm{w}}{\to} \nu_\infty$.

Remark IV.2 (Uniqueness of weak limit of probability measures on $\mathbb{R}$)**.**
The condition (IV.1) uniquely characterizes the limit, a Borel probability measure ν_∞ on $\mathbb{R}$ (in the general case, this follows from Exercise H.8, but see also Theorem IV.4 below).

Remark IV.3 (Weak convergence of real valued random variables)**.**
If $(X_n)_{n=1}^\infty$ is a sequence of random variables with laws $(\nu_n)_{n=1}^\infty$, and X_∞ is a random variable with law ν_∞, then we also denote $X_n \overset{\mathrm{w}}{\to} X_\infty$ if (IV.1) holds. Unraveling the definition, this just means that for all bounded continuous functions $f\colon \mathbb{R}\to\mathbb{R}$ the expected values converge

$$\mathsf{E}[f(X_n)] \longrightarrow \mathsf{E}[f(X_\infty)].$$

This definition is meaningful even if the random variables X_n, $n\in\mathbb{N}\cup\{\infty\}$, are not defined on the same probability space — we can have $X_n\colon \Omega_n\to\mathbb{R}$ to be defined on $(\Omega_n,\mathscr{F}_n,\mathsf{P}_n)$, although then a better notation for the expected values would also be $\mathsf{E}_n[f(X_n)]\to\mathsf{E}_\infty[f(X_\infty)]$.

2. Equivalent characterizations of weak convergence

Below we give different conditions, which are equivalent with weak convergence as defined above. In practice it is often convenient to verify weak convergence on $\mathbb{R}$ using either the fifth or sixth characterization below, i.e., with cumulative distribution functions or characteristic functions.

Theorem IV.4 (Characterizations of weak convergence on the real line)**.**
Let ν_n, $n\in\mathbb{N}$, and ν be probability measures on $\mathbb{R}$, F_n and F their respective cumulative distribution functions $\mathbb{R}\to[0,1]$, and φ_n and φ their respective characteristic functions $\mathbb{R}\to\mathbb{C}$. The following conditions are equivalent:

(i) *The sequence of probability measures converges weakly*

$$\nu_n \underset{n\to\infty}{\overset{\mathrm{w}}{\longrightarrow}} \nu.$$

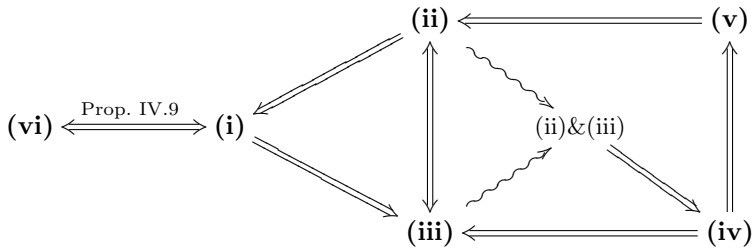
(ii) *For all open subsets $U\subset\mathbb{R}$ we have $\nu[U]\le \liminf_{n\to\infty}\nu_n[U]$.*
(iii) *For all closed subsets $A\subset\mathbb{R}$ we have $\nu[A]\ge \limsup_{n\to\infty}\nu_n[A]$.*
(iv) *For all Borel subsets $E\subset\mathbb{R}$ such that $\nu[\partial E]=0$, we have $\nu_n[E]\to\nu[E]$.*
(v) *The cumulative distribution functions converge pointwise*

$$F_n(x)\underset{n\to\infty}{\longrightarrow} F(x) \qquad \text{at all continuity points } x \text{ of } F.$$

(vi) *The characteristic functions converge pointwise*

$$\varphi_n(\theta)\underset{n\to\infty}{\longrightarrow}\varphi(\theta) \quad \text{for all } \theta\in\mathbb{R}.$$

Remark IV.5 (Road map of the proof of Theorem IV.4)**.**
Our proof of equivalence of the conditions is in principle organized according to the following road map:



However, the proof of equivalence of conditions (i), (ii), (iii) and (iv) will be postponed to Lecture VI, since the proof remains identical in the more general context of probability measures on a general metric space. Below we therefore only prove the implications (iv) ⇒ (v) and (v) ⇒ (ii), and separately in Proposition IV.9 a slightly strengthened version of the equivalence (i) ⇔ (vi).

A completely self-contained proof of the equivalence of conditions (i), (v), and (vi) can be found in, e.g., [**Kyt19**, Appendix F].

Proof of implication (iv) ⇒ (v) *in Theorem IV.4.* Assume condition (iv), i.e. $\nu_n[E] \to \nu[E]$ whenever $E \subset \mathbb{R}$ is a Borel subset such that $\nu[\partial E] = 0$. We want to prove condition (v), i.e. $F_n(x) \to F(x)$ for all continuity points x of F. Write $F_n(x) = \nu_n\big[(-\infty, x]\big]$ and note that the boundary of a semi-infinite interval is a singleton, $\partial(-\infty, x] = \{x\}$. Then note that $x \in \mathbb{R}$ is a continuity point of F if and only if $\nu[\{x\}] = 0$. For continuity points x of F, then, we have

$$F_n(x) = \nu_n\big[(-\infty, x]\big] \longrightarrow \nu\big[(-\infty, x]\big] = F(x).$$

□

Proof of implication (v) ⇒ (ii) *in Theorem IV.4.* Assume condition (v), i.e. $F_n(x) \to F(x)$ for all continuity points x of F.

First consider the case when $U = (a, b)$ is an open interval, with $a < b$ — we allow $a = -\infty$ or $b = +\infty$ but keep the simple notation. Approximate (a, b) from inside: choose real sequences $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ such that $a < a_k < b_k < b$ for all k and $a_k \downarrow a$ and $b_k \uparrow b$ as $k \to \infty$. Then $(a_k, b_k) \uparrow (a, b)$ is an increasing limit of sets, so by monotone convergence of measures we have

$$\nu\big[(a_k, b_k)\big] \to \nu\big[(a, b)\big] \qquad \text{as } k \to \infty.$$

In particular for any $\varepsilon > 0$ we can find k_ε so that

$$\nu\big[(a_{k_\varepsilon}, b_{k_\varepsilon})\big] \geq \nu\big[(a, b)\big] - \varepsilon.$$

Note also that the increasing function F can have at most countably many points of discontinuity (Exercise H.3), so that there exists some continuity points $a' \in (a, a_{k_\varepsilon})$ and $b' \in (b_{k_\varepsilon}, b)$ of F, and we still have by monotonicity of measure

$$\nu\big[(a', b')\big] \geq \nu\big[(a, b)\big] - \varepsilon.$$

Now using the fact that a', b' are continuity points of F, and the assumption (v), we get

$$\begin{aligned}\nu\big[(a, b)\big] - \varepsilon \leq \nu\big[(a', b')\big] &= \lim_{b'' \uparrow b'} \big(F(b'') - F(a')\big) = F(b') - F(a') \\ &\overset{\text{(v)}}{=} \lim_{n\to\infty} \big(F_n(b') - F_n(a')\big) = \lim_{n\to\infty} \nu_n\big[(a', b']\big] \leq \liminf_{n\to\infty} \nu_n\big[(a, b)\big].\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we have shown that

$$\liminf_{n\to\infty} \nu_n\big[(a, b)\big] \geq \nu\big[(a, b)\big],$$

i.e., condition (ii) holds in the case when $U = (a, b)$ is an open interval.

We then consider a general open subset $U \subset \mathbb{R}$. It is an easy fact (Exercise H.2) that the open set U is a disjoint union

$$U = \bigcup_j U_j$$

of at most countably many open intervals U_j. By the first part of the proof, we have

$$\liminf_{n\to\infty} \nu_n\big[U_j\big] \geq \nu\big[U_j\big]$$

for each j. Then use disjointness and Fatou's lemma (Lemma A.8 in Appendix A) to calculate

$$\nu\big[U\big] = \sum_j \nu\big[U_j\big] \leq \sum_j \liminf_{n\to\infty} \nu_n\big[U_j\big] \leq \liminf_{n\to\infty} \sum_j \nu_n\big[U_j\big] = \liminf_{n\to\infty} \nu_n\big[U\big].$$

This shows (ii). □

The following exercises illustrate situations in which it is convenient to verify weak convergence by criteria (v) or (vi).

Exercise IV.1 (Poisson approximation of binomial distribution)**.**

(a) Let $P \sim \text{Poisson}(\lambda)$, i.e., $\mathsf{P}\big[P = k\big] = \frac{\lambda^k}{k!} e^{-\lambda}$ for $k \in \mathbb{Z}_{\geq 0}$. Calculate the characteristic function $\varphi_P(\theta) = \mathsf{E}\big[e^{\mathrm{i}\,\theta P}\big]$.
(b) Let $B \sim \text{Bin}(n, p)$, i.e., $\mathsf{P}\big[B = k\big] = \binom{n}{k} p^k\,(1-p)^{n-k}$ for $k = 0, 1, \ldots, n$. Calculate the characteristic function $\varphi_B(\theta) = \mathsf{E}\big[e^{\mathrm{i}\,\theta B}\big]$.
(c) For $n \in \mathbb{N}$, let $B_n \sim \text{Bin}(n, p_n)$, and assume $np_n \to \lambda$ as $n \to \infty$. Calculate the limit $\lim_{n\to\infty} \mathsf{E}\big[e^{\mathrm{i}\,\theta B_n}\big]$.

Exercise IV.2 (Maximum of independent exponentially distributed random variables)**.**
Let X_j, $j \in \mathbb{Z}_{>0}$, be independent identically distributed random variables with $X_j \sim \text{Exp}(\lambda)$. Denote the maximum of the first n of them by $M_n = \max_{1\leq j\leq n} X_j$, and consider the shifted maxima $R_n = M_n - \frac{1}{\lambda}\log(n)$. Calculate the cumulative distribution functions $F_n(x) = \mathsf{P}[R_n \leq x]$, $n \in \mathbb{Z}_{>0}$, and show that they converge pointwise as $n \to \infty$. Calculate also the limit, and show that it is a cumulative distribution function (recall Proposition A.4).

Exercise IV.3 (Lowest dart on the dartboard)**.**
When a dart is thrown at a uniformly random position in a dartboard modeled by the unit circle, the distribution of the height or "y-coordinate" of the position has the probability density function

$$p(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2} & \text{for } y \in [-1, +1] \\ 0 & \text{for } y \in \mathbb{R} \setminus [-1, +1]. \end{cases}$$

(a) Show that as $\varepsilon \downarrow 0$ we have

$$\int_{-1}^{-1+\varepsilon} p(y)\,\mathrm{d}y = c\,\varepsilon^{3/2} + o(\varepsilon^{3/2}),$$

where c is a positive constant.

Consider repeated throwing of darts in which the heights $Y_1, Y_2, \ldots$ are independent, each having the probability density function p. Let $M_n = \min\{Y_1, \ldots, Y_n\}$ denote the lowest height among the first n throws, and let $R_n = n^{2/3}(M_n + 1)$.

(b) Using (a), prove that the sequence $(R_n)_{n\in\mathbb{N}}$ converges in distribution and that the cumulative distribution function of the limit is $F(r) = 1 - \exp(-\alpha\, r^{3/2})$ for $r \geq 0$, where α is a positive constant.

3. Tightness

Tightness is a condition which guarantees that no probability mass escapes to infinity — it states that up to an arbitrarily small error, the entire mass of each member of a family of probability measures can be found in the same compact set.

Definition IV.6 (Tightness of probability measures on $\mathbb{R}$)**.**
A collection $(\nu_i)_{i\in I}$ of probability measures on $\mathbb{R}$ is *tight*, if for any $\varepsilon > 0$ there exists some $r > 0$ such that

$$\nu_i\Big[\mathbb{R} \setminus [-r, r]\Big] < \varepsilon \qquad \forall i \in I.$$

The following lemma is an obvious reformulation of the definition of tightness in terms of cumulative distribution functions.

Lemma IV.7 (Tightness on $\mathbb{R}$ in terms of cumulative distribution functions)**.**
Let $(\nu_i)_{i\in I}$ be a collection of probability measures on $\mathbb{R}$, and let $(F_i)_{i\in I}$ be the corresponding cumulative distribution functions. The collection $(\nu_i)_{i\in I}$ is tight if and only if for any $\varepsilon > 0$ there exists some r such that

$$F_i(-r) < \varepsilon \qquad \text{and} \qquad F_i(r) > 1 - \varepsilon \qquad \forall i \in I.$$

Tightness is essentially a precompactness property for the topology of weak convergence. The following is a practical formulation.

Theorem IV.8 (Tightness implies precompactness for measures on $\mathbb{R}$)**.**
Let $(\nu_n)_{n\in\mathbb{N}}$ be a tight sequence of probability measures on $\mathbb{R}$. Then there exists a subsequence $(\nu_{n_k})_{k\in\mathbb{N}}$ which converges weakly.

Proof. In the case of probability measures on the real line, this can be proved with cumulative distribution functions, see Exercise IV.4. We later return to this in a more general case. □

Exercise IV.4 (Convergent subsequence for a tight c.d.f. sequence)**.**
Let $F_n \colon \mathbb{R} \to [0, 1]$ be cumulative distribution functions, for $n \in \mathbb{N}$.

(a) Prove that the sequence $(F_n)_{n\in\mathbb{N}}$ has a subsequence $(F_{n_k})_{k\in\mathbb{N}}$ which converges pointwise at all rational points, i.e., there exists a function $G \colon \mathbb{Q} \to [0, 1]$ s.t.

$$\forall q \in \mathbb{Q} \qquad \lim_{k\to\infty} F_{n_k}(q) = G(q).$$

Prove also that if $q, q' \in \mathbb{Q}$ and $q < q'$, then we have $G(q) \le G(q')$.

Assume now that the collection of c.d.f.'s is tight in the sense of Lemma IV.7: for all $\varepsilon > 0$ there exists $r > 0$ such that for all n we have $F_n(-r) \le \varepsilon$ and $F_n(r) \ge 1 - \varepsilon$.

(b) Let $G \colon \mathbb{Q} \to [0, 1]$ be as in part (a). Prove that

$$\lim_{q\to-\infty} G(q) = 0 \qquad \text{and} \qquad \lim_{q\to+\infty} G(q) = 1.$$

(c) Let $G \colon \mathbb{Q} \to [0, 1]$ be as before. Define $F \colon \mathbb{R} \to [0, 1]$ by

$$F(x) := \inf\Big\{G(q) \;\Big|\; q \in \mathbb{Q} \cap (x, +\infty)\Big\}.$$

Prove that F is a cumulative distribution function.

(d) Let $F \colon \mathbb{R} \to [0, 1]$ be as in part (c). Show that we have

$$\lim_{k\to\infty} F_{n_k}(x) = F(x) \qquad \text{for all continuity points } x \text{ of } F.$$

4. Weak convergence with characteristic functions

The goal of this section is to prove a characterization of weak convergence and with characteristic functions, i.e. the equivalence of conditions (i) and (vi) in Theorem IV.4. This is one of the most practical among the equivalent conditions. Note, however, that characteristic functions are defined only for real valued random variables (or slightly more generally, for vector valued random variables), and we do not have a similar characterization in the generality of Lecture VI (random variables with values in general metric spaces).

Proposition IV.9 (Weak convergence with characteristic functions)**.**
Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}$, and $\varphi_n\colon \mathbb{R}\to\mathbb{C}$ the corresponding characteristic functions $\varphi_n(\theta)=\int_{\mathbb{R}} e^{\mathrm{i}\theta x}\,\mathrm{d}\nu_n(x)$.

The sequence on probability measures $(\nu_n)_{n\in\mathbb{N}}$ converges weakly if and only if the sequence of functions $(\varphi_n)_{n\in\mathbb{N}}$ converges pointwise to a function $\varphi\colon\mathbb{R}\to\mathbb{C}$, which is continuous at $\theta=0$. Then φ is the characteristic function of the weak limit measure.

Remark IV.10 (Proposition IV.9 is a strenghtened version of the equivalence (i) $\Leftrightarrow$ (vi))**.**
From the above proposition, we obtain in particular the equivalence of conditions (i) and (vi) in Theorem IV.4: the "only if" part shows that (i) $\Rightarrow$ (vi), and the "if" part shows that (vi) $\Rightarrow$ (i). Concerning the latter, the result here is in fact stronger: we do not need to assume a priori that the pointwise limit φ is a characteristic function of some probability measure, but only that it is continuous at the origin. This condition is easier to verify in applications.

Remark IV.11 (Proof strategy: precompactness + uniqueness of limits)**.**
The non-trivial part of the claim is the "if" part: we want to deduce the convergence of the sequence $(\nu_n)_{n\in\mathbb{N}}$ of probability measures. To prove the convergence, we follow a very important and commonly used strategy — we will verify two things:

(1) We show the *precompactness* (in this case the tightness of the family $(\nu_n)_{n\in\mathbb{N}}$), which implies that *any subsequence contains a convergent further subsequence.*
(2) We concretely characterize the limit of any subsequence (in this case by showing that the characteristic function of any subsequential limit is φ), and deduce that *all convergent subsequences have a common limit.*

By usual topological arguments it then follows that the entire sequence $(\nu_n)_{n\in\mathbb{N}}$ converges (see Exercise H.12).

Proof of "only if" part of Proposition IV.9 Assume the weak convergence $\nu_n \xrightarrow{\mathrm{w}} \nu$. Since the function $x\mapsto e^{\mathrm{i}\theta x}$ is bounded and continuous (treat real and imaginary parts separately), by definition we have $\int_{\mathbb{R}} e^{\mathrm{i}\theta x}\,\mathrm{d}\nu_n(x)\to\int_{\mathbb{R}} e^{\mathrm{i}\theta x}\,\mathrm{d}\nu(x)$, i.e. the characteristic functions φ_n converge pointwise to the characteristic function φ of the measure ν. The characteristic function φ is continuous at the origin.

Proof of "if" part of Proposition IV.9 Assume that $\varphi_n(\theta)\to\varphi(\theta)$ for all $\theta\in\mathbb{R}$, where

$$\lim_{\theta\to 0}\varphi(\theta)=\varphi(0)=\lim_{n\to\infty}\varphi_n(0)=1.$$

We first show the tightness of the sequence $(\nu_n)_{n\in\mathbb{N}}$ of probability measures. For this, we use the following auxiliary calculation

$$\int_{-u}^{u}(1-e^{\mathrm{i}\theta x})\,\mathrm{d}\theta = 2u-\frac{2\sin(ux)}{x}.$$

Divide both sides by u, integrate the variable x with respect to the measure ν_n, apply Fubini's theorem on the left hand side, and estimate the right hand side from below, to get

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}(1-\varphi_n(\theta))\,\mathrm{d}\theta &= 2\int_{\mathbb{R}}\underbrace{\left(1-\frac{\sin(ux)}{ux}\right)}_{\geq 0,\text{ because }\frac{\sin(\xi)}{\xi}\leq 1}\,\mathrm{d}\nu_n(x)\\ &\geq 2\int_{\mathbb{R}\setminus(-\frac{2}{u},\frac{2}{u})}\underbrace{\left(1-\frac{\sin(ux)}{|ux|}\right)}_{\geq\frac{1}{2},\text{ when }|x|\geq 2/u}\,\mathrm{d}\nu_n(x)\\ &\geq \nu_n\left[\mathbb{R}\setminus(-\frac{2}{u},\frac{2}{u})\right].\end{aligned}$$

This inequality gives an upper bound for the measure of the complement of the interval $(-\frac{2}{u},\frac{2}{u})$ in terms of the characteristic function. Next we use the assumption $\varphi(\theta)\to 1$ as $\theta\to 0$. For any $\varepsilon>0$ we may thus choose $u>0$ small enough to guarantee

$$\frac{1}{u}\int_{-u}^{u}(1-\varphi(\theta))\,\mathrm{d}\theta\leq\varepsilon.$$

Since $\varphi_n(\theta)\to\varphi(\theta)$, it follows from the dominated convergence theorem (domination by a constant function will do) that for all n sufficiently large we have

$$\frac{1}{u}\int_{-u}^{u}(1-\varphi_n(\theta))\,\mathrm{d}\theta\leq 2\varepsilon.$$

From an earlier inequality we conclude that with u as above and for all n sufficiently large we have

$$\nu_n\left[\mathbb{R}\setminus(-\frac{2}{u},\frac{2}{u})\right]\leq 2\varepsilon.$$

The tightness of the sequence $(\nu_n)_{n\in\mathbb{N}}$ follows, because finitely many first members of the sequence may be handled separately, and ε was arbitrary.

We have shown that $(\nu_n)_{n\in\mathbb{N}}$ is tight, so it follows that there are convergent subsequences $(\nu_{n_k})_{k\in\mathbb{N}}$. We claim that for such a convergent subsequence $\nu_{n_k}\to\nu$ the characteristic function of the limit is φ. Namely, we have again

$$\varphi(\theta)=\lim_{k\to\infty}\varphi_{n_k}(\theta)=\lim_{k\to\infty}\int_{\mathbb{R}}e^{\mathrm{i}\theta x}\,\mathrm{d}\nu_{n_k}(x)=\int_{\mathbb{R}}e^{\mathrm{i}\theta x}\,\mathrm{d}\nu(x).$$

This implies that the limits of all convergent subsequences are equal. It follows that the sequence $(\nu_n)_{n\in\mathbb{N}}$ converges. □

Lecture V

Curie-Weiss model

The topic of this lecture is the Curie–Weiss model — arguably the simplest possible microscopic model for the emergence of ferromagnetism. The model has a large number N of constituent elementary magnetic units, representing for example atoms in a material. These elementary magnetic units are conventionally called *spins*. For simplicity, each spin is allowed to be in one of just two possible states, thought of as it being magnetized in one of two opposite directions. All spins "interact" with each other in the sense that the probabilities of the configurations of all spins depend on how many spins in the configuration are aligned with each other. The interaction is "ferromagnetic", in the sense that the probabilities of configurations are higher when there is more alignment among the spins, i.e. more pairs of with equal states. Our main results show that the model has a phase transition with respect to a parameter of the model that has the interpretation of temperature. In high temperature the model has "paramagnetic behavior" in the limit $N \to \infty$ of a large system: the average magnetization concentrales at zero. In low temperature the model has "ferromagnetic behavior" in the limit $N \to \infty$: the average magnetization concentrates at non-zero values. Near the critical temperature of transition between the above paramagnetic and ferromagnetic phases, the magnetization has power law type dependence on the parameters of the model characterized by certain "critical exponents".

The Curie–Weiss model disregards all spacial structure of the magnetic material — all spins are thought of as being equally near (or far from) each other, and consequently all pairs are taken to interact with each other in the same way. More realistic models, such as the Ising model studied in later lectures, might take into account the fact that interactions between nearest neighbors are stronger than interactions between components far from each other. The Curie–Weiss model is thus obviously a simplification, whose key virtue is that it becomes easier to analyze mathematically. Fortunately, even with such a simplification, the model has a qualitatively correct phase transition between the ferromagnetic and paramagnetic behaviors. Unfortunately, however, the quantitative values of the critical exponents turn out not to be exactly correct for magnetic materials in three, two, or one dimensions. The Ising model essentially differs from the Curie–Weiss model only by incorporating spacial structure (in particular dimension), and it has also the correct quantitative critical behavior to account for materials known as uniaxial ferromagnets, but it is significantly harder to analyze mathematically. In physics the idea of disregarding spatial structure is known as *mean field* approximation, and the Curie–Weiss model is thus a *mean field model of ferromagnetism* (a mean field version of the Ising model).

The analysis of the Curie–Weiss model essentially boils down to large deviations estimates (see Appendix F for a very brief introduction to large deviations). The large deviations rate functions here also have physical significance as certain thermodynamic potentials.

The Curie–Weiss model is treated in many physics textbooks, and a good mathematical treatment of it can be found for example in [**FV15**].

1. Definition and key properties of the Curie–Weiss model

The model has N *spins* (representing microscopic, elementary magnetic units), each with two possible values, $+1$ and -1. The sample space is therefore taken to be

$$\Omega_N = \{-1, +1\}^N,$$

and the possible outcomes $\vec{\sigma} = (\sigma_i)_{i=1}^N \in \Omega_N$ are thought of as states of the physical system. In the *Curie–Weiss model*, the energy of a state $\vec{\sigma}$ is taken to be

$$H_N(\vec{\sigma}) = -\frac{J}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j - B\sum_{i=1}^{N}\sigma_i, \tag{V.1}$$

where the parameters $J > 0$ and $B \in \mathbb{R}$ are respectively called *coupling strenght* and *external magnetic field*, respectively. The Boltzmann distribution with parameter $\beta > 0$ is now a probability measure $\mathsf{P}^{(N)}$ on the finite sample space Ω_N, such that the individual outcomes have probabilities

$$\mathsf{P}^{(N)}\left[\{\vec{\sigma}\}\right] = \frac{1}{Z_N}\, e^{-\beta H_N(\vec{\sigma})} \tag{V.2}$$

and the partition function Z_N normalizes the total probability to one,

$$Z_N := \sum_{\vec{\sigma}\in\Omega_N} e^{-\beta H_N(\vec{\sigma})}. \tag{V.3}$$

The parameter $\beta > 0$ in Boltzmann distributions is inversely proportional to the temperature of the material, so we follow the usual terminology and call β the *inverse temperature*. With a suitable choice of units (of energy), we may assume $J = 1$. When we wish to emphasize the dependence of the probability measure and the partition function on the remaining parameters $\beta > 0$ and $B \in \mathbb{R}$, we denote them by $\mathsf{P}^{(N)}_{\beta,B}$ and $Z_N(\beta, B)$, respectively.

In thermodynamics, it is relevant to describe the state of the system as a whole, instead of keeping track of the individual microscopic constituents. For this purpose a relevant random variable is the *empirical magnetization*

$$M_N = \frac{1}{N}\sum_{i=1}^{N}\sigma_i. \tag{V.4}$$

This is the average of the random values of the individual spins. Note, however, that unlike many common situations (laws of large numbers, central limit theorem, ...), the terms in the sum are <u>not</u> independent — this is exactly what the ferromagnetic interactions are all about!

We will prove the following results about the *thermodynamical limit* ($N \to \infty$) of the Curie–Weiss model. The first two theorems below express in slightly different ways the fact that in the absence of external magnetic field (i.e., when $B = 0$) the model has a qualitative phase transition at the critical value

$$\beta_c = \frac{1}{2}$$

of the inverse temperature β, and that the model behavior is

$$\begin{array}{ll} \text{paramagnetic} & \text{when } \beta < \beta_c \text{ (and } B = 0) \\ \text{ferromagnetic} & \text{when } \beta > \beta_c \text{ (and } B = 0) \end{array} .$$

More precisely, Theorem V.1 below states that in the thermodynamical limit without external magnetic field, in the paramagnetic phase $\beta < \beta_c$ the empirical magnetization concentrates to the value zero, whereas in the ferromagnetic phase $\beta > \beta_c$ there is a *spontaneous magnetization* $\bar{m}(\beta) > 0$: the empirical magnetization concentrates at the two non-zero values $\pm\bar{m}(\beta)$ with a random sign. Theorem V.2 states that in the thermodynamical limit with an external magnetic field $B > 0$ in the positive direction, the empirical magnetization concentrates at a positive value $\tilde{m}(\beta, B) > 0$, whose limit as the external magnetic field is removed, $B \downarrow 0$, vanishes in the paramagnetic phase and is equal to the spontaneous magnetization $\bar{m}(\beta) > 0$ in the ferromagnetic phase. To complement these two qualitative phase transition statements, Theorem V.3 addresses the critical behavior quantitatively: it gives the values of certain *critical exponents*, for which we will use the symbols $\mathfrak{b}$ and $\mathfrak{d}$.[1] According to this result, near the critical point the spontaneous magnetization has the following behavior: $\bar{m}(\beta) \sim |\beta - \beta_c|^{\mathfrak{b}}$ and $\tilde{m}(\beta_c, B) \sim |B|^{1/\mathfrak{d}}$, where $\mathfrak{b} = \frac{1}{2}$ and $\mathfrak{d} = 3$.

Theorem V.1 (Thermodynamical limit of empirical magnetization for $B = 0$)**.** *Let $\beta > 0$ and $B = 0$. As $N \to \infty$, the empirical magnetizations M_N of the Curie–Weiss model converge weakly to a random variable M_∞, whose law depends on β as follows:*

- *If $\beta < \beta_c$, then $M_\infty = 0$ almost surely.*
- *If $\beta > \beta_c$, then there exists $\bar{m} = \bar{m}(\beta) > 0$ such that*

$$\mathsf{P}\big[M_\infty = +\bar{m}\big] = \frac{1}{2} = \mathsf{P}\big[M_\infty = -\bar{m}\big].$$

Theorem V.2 (Thermodynamical limit of empirical magnetization for $B > 0$)**.** *Let $\beta > 0$ and $B > 0$. As $N \to \infty$, the empirical magnetizations M_N of the Curie–Weiss model converge weakly to a deterministic constant $\tilde{m}(\beta, B) > 0$. As $B \downarrow 0$, we have*

$$\lim_{B \downarrow 0} \tilde{m}(\beta, B) = \begin{cases} \bar{m}(\beta) > 0 & \text{if } \beta > \beta_c \\ 0 & \text{if } \beta < \beta_c \end{cases},$$

where $\bar{m}(\beta) > 0$ is the same spontaneous magnetization as in the previous theorem.

[1] The conventional symbols for these two critical exponents are β and δ, which unfortunately have other established uses already. Our non-conventional notation is an attempt to avoid confusion.

Theorem V.3 (Critical exponents of the Curie–Weiss model)**.**
The functions $\bar{m}$ and $\tilde{m}$ in the previous two theorems have the following asymptotic behavior in the vicinity of the critical point $\beta = \beta_c = \frac{1}{2}$, $B = 0$:

$$\lim_{\beta \downarrow \beta_c} \frac{\bar{m}(\beta)}{|\beta - \beta_c|^{\mathfrak{b}}} \neq 0 \qquad \text{where } \mathfrak{b} = \frac{1}{2}$$

$$\lim_{B \downarrow 0} \frac{\tilde{m}(\beta_c, B)}{B^{1/\mathfrak{d}}} \neq 0 \qquad \text{where } \mathfrak{d} = 3.$$

2. Analysis of the Curie–Weiss model

We begin by observing that the energy H_N given by (V.1) can be written in terms of the empirical magnetization M_N given by (V.4)

$$H_N(\vec{\sigma}) = -N\left(M_N(\vec{\sigma})^2 + BM_N(\vec{\sigma})\right) = N\,\Psi(M_N(\vec{\sigma})),$$
$$\text{where} \qquad \Psi(m) = -m^2 - Bm.$$

Therefore the partition function (V.3) can be written as

$$Z_N(\beta, B) = \sum_{m \in \mathcal{M}_N} z_N(m) e^{-N\beta\Psi(m)},$$

where

$$\mathcal{M}_N = \left\{ \frac{-N}{N}, \frac{2-N}{N}, \frac{4-N}{N}, \dots, \frac{N-2}{N}, \frac{N}{N} \right\}$$

$$\text{i.e.} \quad \begin{cases} \mathcal{M}_N = [-1,1] \cap \frac{2}{N}\mathbb{Z} & \text{for } N \text{ even} \\ \mathcal{M}_N = [-1,1] \cap \frac{2}{N}(\mathbb{Z} + \frac{1}{2}) & \text{for } N \text{ odd} \end{cases}$$

is the set of all possible values of the empirical magnetization, and

$$z_N(m) = \#\left\{ \vec{\sigma} \in \Omega_N \;\middle|\; \frac{1}{N}\sum_{i=1}^{N} \sigma_i = m \right\} = \binom{N}{N\frac{1+m}{2}} = \frac{N!}{(N\frac{1+m}{2})!\,(N\frac{1-m}{2})!}.$$

is a binomial coefficient, which accounts for the number of different configurations of the spins that give rise to the value m of the empirical magnetization.

To analyze the limit $N \to \infty$, we need asymptotics of the binomial coefficients. We observe that the Stirling approximation $n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$ (Theorem G.1 in Appendix G) can be used to show (Exercise F.3 in Appendix F) implies the following

$$\log\left(z_N(m)\right) = N\left(\log(2) - I(m)\right) + o(N), \tag{V.5}$$

where $I(m)$ is the Cramèr entropy

$$I(m) = \frac{1+m}{2}\log\left(1+m\right) + \frac{1-m}{2}\log\left(1-m\right). \tag{V.6}$$

Note that $I\colon [-1,1] \to \mathbb{R}$ is continuous: the limits $m \to +1$ and $m \to -1$ are obtained with the help of the familiar fact $x\log(x) \to 0$ as $x \searrow 0$.

The magnetic Gibbs free energy $F_N = -\frac{1}{\beta}\log(Z_N)$ can now be calculated in the limit $N \to \infty$ from the partition function

$$\begin{aligned} Z_N(\beta, B) &= \sum_{m \in \mathcal{M}_N} z_N(m)\, e^{-N\beta\Psi(m)} \\ &= \sum_{m \in \mathcal{M}_N} \exp\Big(- N\beta\big(g(\beta, m) - Bm\big) + o(N)\Big), \end{aligned} \tag{V.7}$$

where

$$g(\beta, m) = \frac{I(m) - \log(2)}{\beta} - m^2 \tag{V.8}$$

is called the magnetic Helmholtz free energy. The following Proposition says that the thermodynamical limit of the Gibbs free energy per spin, $f = \lim_{N\to\infty} \frac{1}{N} F_N$, is a Legendre transform of the Helmholtz free energy g.

Proposition V.4 (The Gibbs free energy as a Legendre transform)**.**
Let $\beta > 0$ and $B \in \mathbb{R}$. Then the limit

$$f(\beta, B) := \lim_{N\to\infty} \left(\frac{-1}{\beta N} \log\left(Z_N(\beta, B)\right)\right)$$

exists, and is given in terms of (V.8) *by*

$$f(\beta, B) = \inf_{m \in [-1,1]} \Big(g(\beta, m) - Bm \Big).$$

Proof. To prove the asserted equality, we prove inequalities in both directions separately.

Let us start with the lower bound for the Gibbs free energy, which will be obtained from an upper bound for the partition function

$$\begin{aligned} Z_N(\beta, B) &= \sum_{m \in \mathcal{M}_N} \exp\left[-N\beta\big(g(\beta, m) - Bm\big) + o(N)\right] \\ &\leq (N+1) \max_{m \in \mathcal{M}_N} \left(\exp\left[-N\beta\big(g(\beta, m) - Bm\big) + o(N)\right]\right). \end{aligned}$$

Taking logarithms, dividing by N, and collecting together terms that are negligible in the limit $N \to \infty$, we obtain

$$\begin{aligned} \frac{1}{N}\log\left(Z_N(\beta, B)\right) &\leq \beta \max_{m \in \mathcal{M}_N} \left(-g(\beta, m) + Bm\right) + o(1) \\ &\leq \beta \sup_{m \in [-1,1]} \left(-g(\beta, m) + Bm\right) + o(1). \end{aligned}$$

By further dividing by $-\beta$ and letting $N \to \infty$ we get the desired lower bound for $f(\beta, B)$

$$\liminf_{N\to\infty} \frac{-1}{\beta N} \log\left(Z_N(\beta, B)\right) \geq \inf_{m \in [-1,1]} \left(g(\beta, m) - Bm\right).$$

An upper bound for the Gibbs free energy is correspondingly obtained from the following lower bound on the partition function, where we only keep the largest term in the sum that defines the partition function

$$Z_N(\beta, B) \geq \max_{m \in \mathcal{M}_N} \left(\exp\left[-N\beta\big(g(\beta, m) - Bm\big) + o(N)\right]\right).$$

Again taking logarithms and dividing by N we obtain

$$\frac{1}{N}\log\left(Z_N(\beta, B)\right) \geq -\beta \max_{m \in \mathcal{M}_N} \left(g(\beta, m) - Bm\right) + o(1).$$

The function $m \mapsto g(\beta, m) - Bm$ is continuous on the interval $[-1, 1]$. By compactness, it achieves its minimum at some point. In the subset $\mathcal{M}_N \subset [-1, +1]$ there are points

at distance at most $\frac{2}{N}$ from the point at which the minimum is achieved. Therefore by continuity we have

$$\min_{m\in\mathcal{M}_N}\,(g(\beta,m)-Bm)\;\underset{N\to\infty}{\longrightarrow}\;\inf_{m\in[-1,1]}(g(\beta,m)-Bm).$$

Again dividing by $-\beta$ and letting $N\to\infty$ we get the desired upper bound for the Gibbs free energy $f(\beta,B)$

$$\limsup_{N\to\infty}\frac{-1}{\beta N}\log\left(Z_N(\beta,B)\right)\leq\inf_{m\in[-1,1]}\left(g(\beta,m)-Bm\right).$$

From the two inequalities above, we conclude the thermodynamical limit of the Gibbs free energy

$$f(\beta,B)=\lim_{N\to\infty}\left(\frac{-1}{\beta N}\log\left(Z_N(\beta,B)\right)\right)$$

exists, and is given by the asserted formula $\inf_{m\in[-1,1]}(g(\beta,m)-Bm)$. □

Let us comment on the interpretation for the quantities above. The Helmholtz free energy g expresses the rate of large deviations of the empirical magnetization: roughly speaking the probability that the empirical magnetization assumes a given value is exponentially small under the probability measure $\mathsf{P}^{(N)}_{\beta,B}$ defined in (V.2)

$$\mathsf{P}^{(N)}_{\beta,B}\big[M_N\approx m\big]\sim\,\exp\Big(-N\beta\big(g(\beta,m)-Bm-f(\beta,B)\big)\Big).$$

A precise formulation of this is given below.

Proposition V.5 (Large deviation rate function for empirical magnetization)**.** *Let $\beta>0$ and $B\in\mathbb{R}$. For any open set $A\subset[-1,1]$ we have*

$$\lim_{N\to\infty}\frac{-\log\Big(\mathsf{P}^{(N)}_{\beta,B}\big[M_N\in A\big]\Big)}{N}=\inf_{m\in A}\,\phi(m),$$

where

$$\phi(m):=\beta\Big(\big(g(\beta,m)-Bm\big)-f(\beta,h)\Big).$$

Proof. The probability of the event $M_N\in A$ is

$$\mathsf{P}^{(N)}_{\beta,B}\left[M_N\in A\right]=\frac{1}{Z_N(\beta,B)}\sum_{m\in A\cap\mathcal{M}_N}z_N(m)e^{N\beta(m^2+Bm)}.$$

By estimating the sum on the right hand side as in the proof of the previous theorem, we get

$$\frac{1}{N}\log\left(\sum_{m\in A\cap\mathcal{M}_N}z_N(m)e^{N\beta(m^2+Bm)}\right)=\,-\,\beta\inf_{m\in A}\big(g(\beta,m)-Bm\big)+o(1).$$

The asserted formula follows by taking logarithms of $\mathsf{P}^{(N)}_{\beta,B}\left[M_N\in A\right]$ and using also Proposition V.4 about the asymptotics of the logarithm of the partition function. □

We are essentially ready to proceed to the proofs our main results. Let us summarize the strategy. The minima of the Helmholz free energy (i.e. the zeroes of the large deviation rate function) are the points at which the distribution of M_N concentrates — anywhere else the probabilities are exponentially small. The minima of the Helmholz free energy will be found by straightforward calculations. We formulate the following auxiliary results for the implementation of this strategy.

Lemma V.6 (Weak limits are supported on the zeroes of the rate function)**.** *Assume that a sequence $(X_n)_{n\in\mathbb{N}}$ of real random variables converges weakly, $X_n \xrightarrow{\text{w}} X$, and that the sequence satisfies a large deviations upper bound*

$$\limsup_{n\to\infty} \frac{\log \mathsf{P}[X_n \in A]}{n} \leq -\inf_{x\in A} \phi(x) \qquad \text{for all open sets } A \subset \mathbb{R},$$

where the large deviations rate function $\phi\colon \mathbb{R} \to [0,\infty)$ is continuous. Then the limit random variable X takes values in the set $\Phi := \big\{x \in \mathbb{R} \,\big|\, \phi(x) = 0\big\}$ of zeroes of the rate function ϕ, i.e., we have $\mathsf{P}[X \in \Phi] = 1$.

Proof. Define the open sets $A_\varepsilon = \big\{x \in \mathbb{R} \,\big|\, \phi(x) > \varepsilon\big\}$ for $\varepsilon > 0$. Since $\inf_{x\in A_\varepsilon} \phi(x) \geq \varepsilon$ (by continuity), we infer from the large deviations upper bound that

$$\mathsf{P}\big[X_n \in A_\varepsilon\big] \,=\, \exp\Big(-n\,\frac{\log \mathsf{P}[X_n \in A]}{n}\Big) \,\leq\, e^{-n\varepsilon/2} \qquad \text{for } n \text{ large},$$

and in particular that $\lim_{n\to\infty} \mathsf{P}\big[X_n \in A_\varepsilon\big] = 0$. Thus by the characterization (ii) of weak convergence in Theorem IV.4, we have $\mathsf{P}[X \in A_\varepsilon] \leq \liminf_n \mathsf{P}[X_n \in A_\varepsilon] = 0$. Since the complement of the set Φ can be expressed as a countable union of sets of the form A_ε, the union bound yields the desired conclusion

$$\mathsf{P}[X \notin \Phi] = \mathsf{P}\Big[X \in \bigcup_{k=1}^{\infty} A_{1/k}\Big] \leq \sum_{k=1}^{\infty} \underbrace{\mathsf{P}[X \in A_{1/k}]}_{=0} = 0.$$

□

Figure V.1 illustrates the Helmholtz free energy function $m \mapsto g(\beta, m)$, and especially the dependence of the locations of the minima on the parameter β. The needed auxiliary results are formulated in the following.

Lemma V.7 (Properties of the large deviations rate function for $B = 0$)**.** *For fixed $\beta > 0$ and $B = 0$, the large deviations rate function*

$$\phi(m) = \beta(g(\beta, m) - f(\beta, 0))$$

determined by the Helmholtz free energy $g(\beta, m) = \frac{I(m)-\log(2)}{\beta} - m^2$ has the following properties:

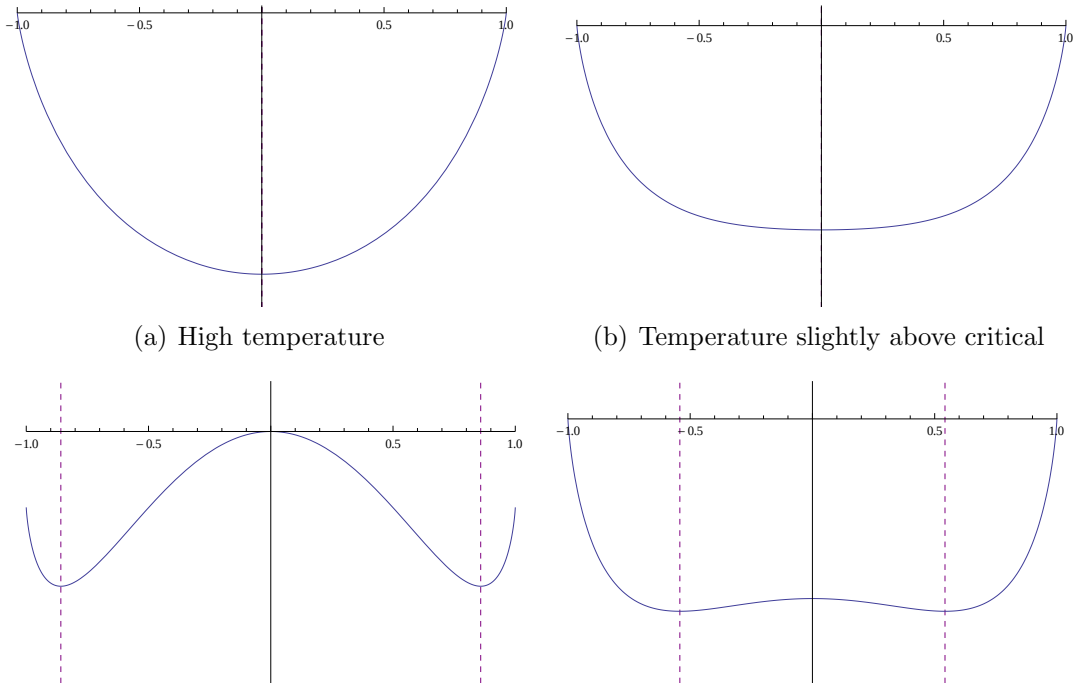
(a) *$\phi\colon [-1, 1] \to \mathbb{R}$ is continuous and its minimum value is 0.*
(b) *If $\beta \leq \frac{1}{2}$, then the only zero of the function ϕ is at $m = 0$.*
(c) *If $\beta > \frac{1}{2}$, then the equation $\phi'(m) = 0$ has a unique positive solution $m = \bar{m}(\beta)$, and the only two zeroes of ϕ are at $m = +\bar{m}(\beta)$ and $m = -\bar{m}(\beta)$.*

Proof. The continuity of the function ϕ is clear from the continuity of the function I on the interval $[-1, 1]$. Proposition V.4 implies that $f(\beta, 0)$ is the minimum of the function $m \mapsto g(\beta, m)$, so the minimum of ϕ is 0, and part (a) follows.

Moreover, ϕ is $\mathcal{C}^\infty$ on the open interval $(-1, 1)$, and it is an even function, $\phi(-m) = \phi(m)$. Thus we must have $\phi'(0) = 0$. An easy calculation gives $I''(m) = \frac{1}{1-m^2}$, and correspondingly

$$\phi''(m) = \frac{1}{1-m^2} - 2\beta.$$

We see that for $\beta \leq \frac{1}{2}$ we have $\phi''(m) \geq 0$ for all $m \in (-1, +1)$, and the unique minimum of ϕ is at $m = 0$. Correspondingly for $\beta > \frac{1}{2}$, the second derivative ϕ'' is negative on the interval $\big(-(1 - \frac{1}{2\beta})^{1/2}, +(1 - \frac{1}{2\beta})^{1/2}\big)$ around zero and non-negative elsewhere. The function ϕ therefore has a local maximum at zero, and it has two other zeroes $\pm\bar{m}(\beta)$ of the derivative ϕ' outside this interval (since $\phi'(m) \to \pm\infty$ as $m \to \pm 1$), which are the two minima of ϕ. □



(a) High temperature

(b) Temperature slightly above critical

(c) Low temperature

(d) Temperature slightly below critical

FIGURE V.1. Curie–Weiss model Helmholtz free energy $m \mapsto g(\beta, m)$ gives the rate of large deviations for the empirical magnetization: in the paramagnetic phase $\beta < \beta_c = \frac{1}{2}$ its minimum is at $m = 0$, and in the ferromagnetic phase $\beta > \beta_c = \frac{1}{2}$ there are two minima at $m = \pm\bar{m}(\beta)$. Figure V.1(a): $\beta = 0.25$, Figure V.1(b): $\beta = 0.46$, Figure V.1(d): $\beta = 0.56$, Figure V.1(c): $\beta = 0.75$.

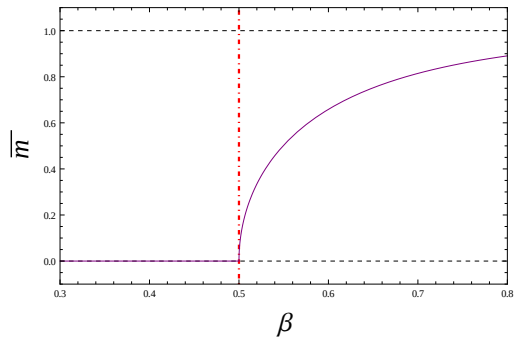
Proof of Theorem V.1. We want to show that the empirical magnetizations M_N converge weakly towards the asserted law. Note first that the sequence $(M_N)_{N\in\mathbb{N}}$ of random variables is tight (cf. Section IV.3), because the values of each M_N are on the interval $[-1, 1]$. Theorem IV.8 implies that any subsequence $(M_{N_k})_{k\in\mathbb{N}}$ has some further subsequence which is convergent. By the principle of "precompactness plus identification of the limit" (Exercise H.12), it then suffices to show that the limit of any convergent subsequence is of the asserted form.

In the case $\beta \leq \beta_c = \frac{1}{2}$, for any subsequential limit $M_{N_k} \xrightarrow{\text{w}} M_\infty$, it follows from Lemma V.6, Proposition V.5 and Lemma V.7(b) that $\mathsf{P}[M_\infty = 0] = 1$. Since any subsequential limit is of the asserted form (deterministic constant zero), we conclude $M_N \xrightarrow{\text{w}} 0$.

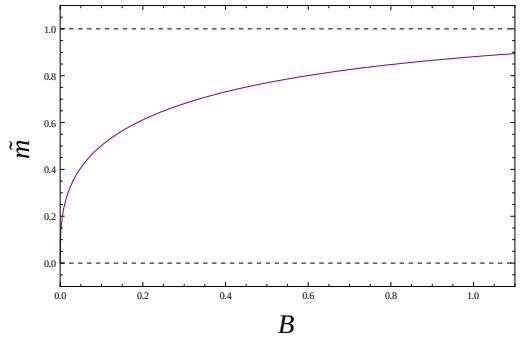
In the case $\beta > \beta_c = \frac{1}{2}$, for any subsequential limit $M_{N_k} \xrightarrow{\text{w}} M_\infty$ we have, by Lemma V.6, Proposition V.5 and Lemma V.7(c) that $\mathsf{P}\big[M_\infty \in \{-\bar{m}, +\bar{m}\}\big] = 1$. Moreover, we obviously have the symmetry $\mathsf{P}[M_N < 0] = \mathsf{P}[M_N > 0] \leq \frac{1}{2}$, since $B = 0$. By property (ii) of Theorem IV.4 this implies for the subsequential limit M_∞ that both $\mathsf{P}[M_\infty < 0] \leq \frac{1}{2}$ and $\mathsf{P}[M_\infty > 0] \leq \frac{1}{2}$ hold. Given that the support of M_∞ is just $\{-\bar{m}, +\bar{m}\}$, this is only possible if we have $\mathsf{P}[M_\infty = +\bar{m}] = \frac{1}{2} = \mathsf{P}[M_\infty = -\bar{m}]$. Again we conclude the weak convergence to the asserted limit. □

Proof of Theorem V.2. The proof is essentially the same as above. The only difference is that for $B > 0$, the minimum of the large deviations rate function ϕ is at the unique positive solution $\tilde{m}(\beta, B) > 0$ of the equation $\frac{\partial}{\partial m}\big(g(\beta, m)\big) - Bm) = 0$. □

Proof of Theorem V.3. From the considerations above, it follows that the values of $\bar{m}(\beta)$ and $\tilde{m}(\beta, B)$ are characterized as (certain) zeroes of the derivative of the large deviations rate function ϕ in Proposition V.5. We leave the remaining calculations as Exercise V.1. □



(a) Spontaneous magnetization as a function of the inverse temperature β



(b) Magnetization at critical temperature as a function of the external magnetic field B

FIGURE V.2. The critical exponents $\mathfrak{b}$ and $\mathfrak{d}$ of the Curie–Weiss model describe the power law behavior of the magnetization in the vicinity of the critical point $\beta = \beta_c = \frac{1}{2}$, $B = 0$. Figure V.2(a): $\beta \mapsto \bar{m}(\beta)$, Figure V.2(b): $B \mapsto \tilde{m}(\beta_c, B)$.

Exercise V.1 (Critical exponents of the Curie–Weiss model)**.**
For the Curie-Weiss model, the Helmholtz free energy reads

$$g(\beta, m) = \frac{1}{\beta}\Big(\frac{1+m}{2}\log(1+m) + \frac{1-m}{2}\log(1-m) - \log(2)\Big) - m^2.$$

(a) For a fixed $\beta > \frac{1}{2}$, let $\bar{m} = \bar{m}(\beta)$ be the unique positive solution of $\frac{\partial}{\partial m} g(\beta, m) = 0$. Calculate

$$\lim_{\beta \downarrow \frac{1}{2}} \frac{\bar{m}(\beta)}{(\beta - \frac{1}{2})^{1/2}}.$$

(b) For fixed $B > 0$ and $\beta > 0$, let $\tilde{m} = \tilde{m}(\beta, B)$ be the unique positive solution of $\frac{\partial}{\partial m}\big(g(\beta, m) - Bm\big) = 0$. Set $\beta = \frac{1}{2}$, and calculate

$$\lim_{B \downarrow 0} \frac{\tilde{m}(\frac{1}{2}, B)}{B^{1/3}}.$$

Lecture VI

Weak convergence on metric spaces

We will next turn to probability theory on more general spaces.[1] We will in particular study weak convergence of probability measures. The definition of weak convergence would make sense on any topological spaces, but for concreteness we only discuss metric spaces. Relevant background in the topology of metric spaces is recalled in Appedix H.

Important applications of weak convergence on complete separable metric spaces are for example the following two, which we impressionistically illustrate in Figure VI.1 and treat in later lectures:

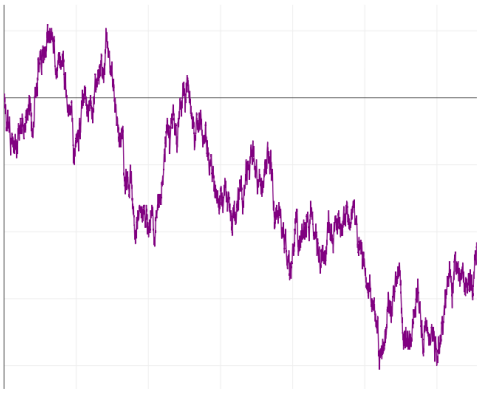
- *Donsker's theorem* (Lecture VIII):
  - The result states that appropriately rescaled random walks are well approximated by the Brownian motion.
  - This statement will be formulated in terms of weak convergence of probability measures on the space $\mathcal{C}([0,T])$ of continuous real valued functions of time.
- *Ising model thermodynamical limit* (Lecture X):
  - The idea is to define the Ising model on the infinite lattice $\mathbb{Z}^d$ by a limit of Ising models on increasingly large finite subgraphs of the lattice.
  - The precise statement is formulated in terms of weak convergence on the countable product of finite spaces $\{-1,+1\}^{\mathbb{Z}^d}$.
  - Phase transitions in the Ising model (and other models) are closely related to the (non-)uniqueness of such an infinite volume limit.

For the reader interested in studying the topic in more depth we recommend the textbook [**Bil99**], which is entirely devoted to the weak convergence of probability measures in a general setup.

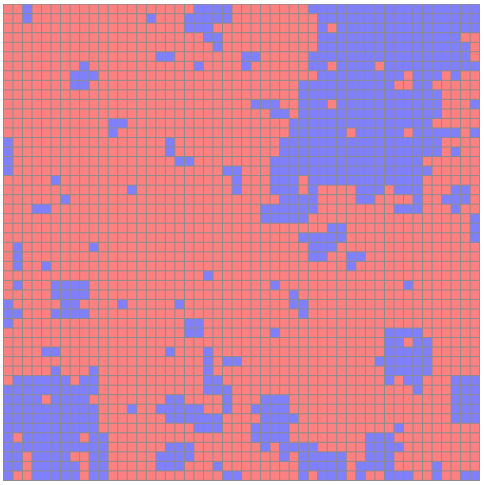
1. Weak convergence of probability measures

Recall the following idea: what is considered reliably measurable by observations about a random system (in our mathematical idealization) is the expected values of bounded continuous functions of the random state of the system. Weak convergence means exactly the convergence of all such observable quantities. For this definition,

[1] In some sense the natural generality for much of probability theory and its applications is that of Polish spaces — topological spaces which are homeomorphic to a complete separable metric space. For example, the equivalence between tightness and precompactness of probability measures (in the topology of weak convergence) will be generalized to Polish spaces in Lecture VII (the case of probability measures on the real axis was encountered in Lecture IV). In Section 3 of the present lecture, we also see the role of separability for the question of metrizability of weak convergence.



(a) The Brownian motion is a random element of a space of continuous functions.



(b) The thermodynamical limit of the Ising model is a random assignment of ± 1 spins to the sites of the lattice $\mathbb{Z}^d$.

FIGURE VI.1. Many stochastic models feature random objects, which are naturally elements of a suitable complete separable metric space.

we only need $\mathfrak{X}$ to be a topological space, so that we may talk about continuous functions on it.

Definition VI.1 (Weak convergence)**.**
A sequence $(\nu)_{n\in\mathbb{N}}$ of Borel probability measures on a topological space $\mathfrak{X}$ converges weakly to a Borel probability measure ν, if for all bounded continuous functions $f\colon \mathfrak{X} \to \mathbb{R}$ we have

$$\int_{\mathfrak{X}} f \,\mathrm{d}\nu_n \longrightarrow \int_{\mathfrak{X}} f \,\mathrm{d}\nu.$$

Although the definition makes sense without metric, we will now assume that the space $\mathfrak{X}$ is equipped with a metric ϱ which gives its topology. We denote the Borel σ-algebra on $\mathfrak{X}$ by $\mathscr{B}(\mathfrak{X})$.

Remark VI.2 (Uniqueness of weak limits)**.**
The weak limit ν is unique, if it exists. This follows directly from part (ii) of Exercise H.8 in Appendix H.

In different contexts different terminology is used for weak convergence. The alternative terminologies of convergence in distribution and convergence in law are in particular commonly used when the probability measures are the laws of some given random variables, and we prefer to emphasize the random variables themselves.

Definition VI.3 (Convergence in distribution)**.**
Let $X_1, X_2, \ldots$ be random variables with values in $\mathfrak{X}$, and let ν_n denote the law (distribution) of X_n. We say that the sequence of random variables $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* (also that it *converges in law*), if the laws ν_n converge weakly.

Equivalent characterizations of weak convergence

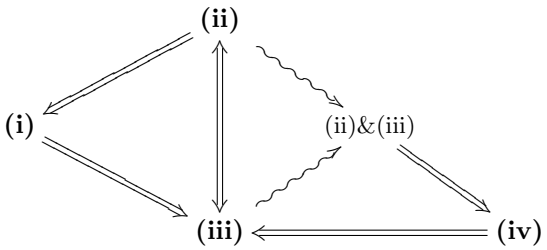
There are several equivalent characterizations of weak convergence.

Theorem VI.4 (Portmanteau theorem)**.** *Let ν_n, $n \in \mathbb{N}$, and ν be probability measures on a metric space $(\mathfrak{X}, \varrho)$. Then the following are equivalent:*

(i) *The sequence of probability measures converges weakly, $\nu_n \xrightarrow[n\to\infty]{\text{w}} \nu$.*
(ii) *For all open sets $G \subset \mathfrak{X}$ we have $\nu[G] \le \liminf_{n\to\infty} \nu_n[G]$.*
(iii) *For all closed sets $F \subset \mathfrak{X}$ we have $\nu[F] \ge \limsup_{n\to\infty} \nu_n[F]$.*
(iv) *For all Borel sets $E \subset \mathfrak{X}$, for which $\nu[\partial E] = 0$, we have $\nu_n[E] \to \nu[E]$.*

Remark VI.5 (Finishing the proof of Theorem IV.4)**.**
In the special case $\mathfrak{X} = \mathbb{R}$, the theorem above completes the proof of Theorem IV.4, by establishing the implications whose proofs were omitted in Lecture IV.

Proof of Theorem VI.4: The roadmap for the proof is a part of the roadmap we presented for the more specialized Theorem IV.4. The relevant part in the current greater generality is:



Below we establish the implications indicated in this roadmap separately.

proof of (ii) ⇔ (iii)*:* The equivalence of conditions (ii) and (iii) is clear by setting $F = \mathfrak{X} \setminus G$ and vice versa: then $\nu[F] = 1 - \nu[G]$ and similarly for ν_n.

proof of (ii)&(iii) ⇒ (iv)*:* We assume the two equivalent conditions (ii) and (iii), and prove (iv). Suppose $E \in \mathscr{B}(\mathfrak{X})$, and $\nu[\partial E] = 0$. Denote by $\overline{E}$ the closure and by E° the interior of E. We have $E^\circ \subset E \subset \overline{E}$, and since $\nu[\overline{E} \setminus E^\circ] = \nu[\partial E] = 0$, the measures of all three are equal, $\nu[E^\circ] = \nu[E] = \nu[\overline{E}]$. Assuming (ii) and (iii), we get the inequalities

$$\begin{aligned} \liminf \nu_n[E] \ \ge\ \liminf \nu_n[E^\circ] &\overset{\text{(ii)}}{\ge} \nu[E^\circ] \\ &= \nu[E] \\ &= \nu[\overline{E}] \underset{\text{(iii)}}{\ge} \limsup \nu_n[\overline{E}] \ \ge\ \limsup \nu_n[E]. \end{aligned}$$

We conclude that $\lim_{n\to\infty} \nu_n[E]$ exists and equals $\nu[E]$, establishing (iv).

proof of (iv) ⇒ (iii)*:* Suppose $F \subset \mathfrak{X}$ is closed. For $\delta > 0$, denote $F_\delta = \big\{x \in \mathfrak{X} \;\big|\; \varrho(x, F) \le \delta\big\}$. Since $\mathfrak{X} \setminus F = \bigcup_{\delta>0} \big\{x \in \mathfrak{X} \;\big|\; \varrho(x, F) = \delta\big\}$ is a disjoint union, and $\nu[\mathfrak{X} \setminus F] \le 1$, the sets $\big\{x \in \mathfrak{X} \;\big|\; \varrho(x, F) = \delta\big\}$ can have positive ν-measure for only countably many values of δ. Note that $\partial F_\delta \subset \big\{x \in \mathfrak{X} \;\big|\; \varrho(x, F) = \delta\big\}$ (see Exercise H.6), so also ∂F_δ can have positive ν-measure for only countably many values of δ. We can thus find a sequence $(\delta_k)_{k\in\mathbb{N}}$ such that $\delta_k \downarrow 0$ and $\nu[\partial F_{\delta_k}] = 0$ for all $k \in \mathbb{N}$. Assuming (iv), we then have $\nu_n[F_{\delta_k}] \to \nu[F_{\delta_k}]$ as $n \to \infty$. Therefore, for any k,

$$\limsup_{n\to\infty} \nu_n[F] \le \lim_{n\to\infty} \nu_n[F_{\delta_k}] = \nu[F_{\delta_k}].$$

As $k \to \infty$, we have $F_{\delta_k} \downarrow F$ (see Exercise H.6 again), so by monotone decreasing convergence for probability measures the right hand side above tends to $\nu[F]$, and we deduce (iii):

$$\limsup_{n\to\infty} \nu_n[F] \le \nu[F].$$

proof of (i) ⇒ (iii)*:* Suppose that $F \subset \mathfrak{X}$ is closed. For $\delta > 0$, denote $F_\delta = \{x \in \mathfrak{X} \mid \varrho(x, F) \le \delta\}$. By Lemma H.8 one can find a continuous function $f_\delta \colon \mathfrak{X} \to [0,1]$ such that $f_\delta(x) = 1$ for all $x \in F$ and $f_\delta(x) = 0$ if $\varrho(x, F) \ge \delta$, that is,

$$\mathbb{I}_F \;\le\; f_\delta \;\le\; \mathbb{I}_{F_\delta}.$$

We thus have

$$\nu_n[F] \le \int_{\mathfrak{X}} f_\delta \,\mathrm{d}\nu_n \qquad \text{and} \qquad \int_{\mathfrak{X}} f_\delta \,\mathrm{d}\nu \le \nu[F_\delta].$$

Since f_δ is continuous and bounded, assuming (i) we have

$$\limsup_{n\to\infty} \nu_n[F] \le \lim_{n\to\infty} \int_{\mathfrak{X}} f_\delta \,\mathrm{d}\nu_n \overset{\text{(i)}}{=} \int_{\mathfrak{X}} f_\delta \,\mathrm{d}\nu \le \nu[F_\delta].$$

We have $F_\delta \downarrow F$ as $\delta \downarrow 0$, so the right hand side tends to $\nu[F]$, establishing property (iii):

$$\limsup_{n\to\infty} \nu_n[F] \le \nu[F].$$

proof of (ii) ⇒ (i)*:* Suppose that $f \colon \mathfrak{X} \to \mathbb{R}$ is continuous and bounded. Assuming (ii), we will first show that

$$\liminf_{n\to\infty} \int_{\mathfrak{X}} f \,\mathrm{d}\nu_n \ge \int_{\mathfrak{X}} f \,\mathrm{d}\nu. \tag{VI.1}$$

Then applying the same to the function $-f$, we can deduce that $\limsup \int_{\mathfrak{X}} f \,\mathrm{d}\nu_n \le \int_{\mathfrak{X}} f \,\mathrm{d}\nu$. Combining the two, we get $\lim_{n\to\infty} \int_{\mathfrak{X}} f \,\mathrm{d}\nu_n = \int_{\mathfrak{X}} f \,\mathrm{d}\nu$, which will establish property (i). It thus suffices to show that (ii) implies (VI.1). Without loss of generality, we can assume that the function f is non-negative, since adding a constant to f does not change the validity of (VI.1).

Assume (ii), i.e., that for all open $G \subset \mathfrak{X}$ we have $\nu[G] \le \liminf \nu_n[G]$. Now on the probability space $(\mathfrak{X}, \mathscr{B}(\mathfrak{X}), \nu)$ consider the function $x \mapsto f(x)$ as a non-negative real-valued random variable. Apply Equation (B.2) to calculate its expected value

$$\int_{\mathfrak{X}} f \,\mathrm{d}\nu = \int_0^\infty \nu\Big[\,\{x \in \mathfrak{X} \mid f(x) > s\}\,\Big] \,\mathrm{d}s.$$

By continuity of f, the set $\{x \in \mathfrak{X} \mid f(x) > s\}$ is open, so by (ii) and Fatou's lemma (Lemma A.8 in Appendix A) we can estimate

$$\begin{aligned}\int_{\mathfrak{X}} f \,\mathrm{d}\nu &\le \int_0^\infty \liminf_{n\to\infty} \nu_n\Big[\,\{x \in \mathfrak{X} \mid f(x) > s\}\,\Big] \,\mathrm{d}s \\ &\le \liminf_{n\to\infty} \int_0^\infty \nu_n\Big[\,\{x \in \mathfrak{X} \mid f(x) > s\}\,\Big] \,\mathrm{d}s = \liminf_{n\to\infty} \int_{\mathfrak{X}} f \,\mathrm{d}\nu_n.\end{aligned}$$

This shows (VI.1), and therefore proves the implication (ii) ⇒ (i). □

2. A criterion for verifying weak convergence

Let $(\mathfrak{X}, \varrho)$ be a metric space, and $\mathscr{B}(\mathfrak{X})$ its Borel σ-algebra. The following proposition gives an occasionally useful sufficient condition for weak convergence in terms of a smaller collection than all Borel sets.

Proposition VI.6 (A sufficient condition for weak convergence)**.**
Suppose that $\mathcal{E} \subset \mathscr{B}(\mathfrak{X})$ is a collection of Borel subsets of $\mathfrak{X}$ such that the

following conditions hold:

$$\mathcal{E} \text{ is a } \pi\text{-system (stable under finite intersections)} \tag{VI.6a}$$
$$\text{i.e., if } E_1, E_2 \in \mathcal{E} \text{ then also } E_1 \cap E_2 \in \mathcal{E}$$
$$\text{any open set } G \subset \mathfrak{X} \text{ is a countable union of sets from } \mathcal{E} \tag{VI.6b}$$
$$\text{i.e., } G = \bigcup_{i=1}^{\infty} E_i \text{ with } E_i \in \mathcal{E}.$$

Then, a sequence $(\nu_n)_{n\in\mathbb{N}}$ of probability measures on $\mathfrak{X}$ converges weakly to a probability measure ν if for all $E \in \mathcal{E}$ we have $\nu_n[E] \to \nu[E]$ as $n \to \infty$.

Proof. Assume that $\nu_n[E] \to \nu[E]$ for all $E \in \mathcal{E}$ and that $\mathcal{E}$ satisfies (VI.6a) and (VI.6b).

Suppose that $E_1, \dots, E_m \in \mathcal{E}$. Recall the inclusion-exclusion formula

$$\nu_n\Big[\bigcup_{i=1}^{m} E_i\Big] = \sum_{\substack{J\subset\{1,\dots,m\}\\ J\neq\emptyset}} (-1)^{\#J-1}\, \nu_n\Big[\bigcap_{j\in J} E_j\Big]$$

By (VI.6a) also the finite intersections $\bigcap_{j\in J} E_j$ are in $\mathcal{E}$. Therefore, by the assumption of convergence, as $n \to \infty$ we have

$$\nu_n\Big[\bigcup_{i=1}^{m} E_i\Big] = \sum_{J} (-1)^{\#J-1}\, \nu_n\Big[\bigcap_{j\in J} E_j\Big] \longrightarrow \sum_{J} (-1)^{\#J-1}\, \nu\Big[\bigcap_{j\in J} E_j\Big] = \nu\Big[\bigcup_{i=1}^{m} E_i\Big].$$

If $G \subset \mathfrak{X}$ is an open set, then by (VI.6b) there exists $E_1, E_2, \dots \in \mathcal{E}$ such that $G = \bigcup_{i=1}^{\infty} E_i$. For any m, use the calculation above to get

$$\nu\Big[\bigcup_{i=1}^{m} E_i\Big] = \lim_{n\to\infty} \nu_n\Big[\bigcup_{i=1}^{m} E_i\Big] \le \liminf_{n\to\infty} \nu_n\big[G\big].$$

On the other hand $\bigcup_{i=1}^{m} E_i \uparrow G$ as $m \to \infty$, so by monotone convergence of measures the left hand side increases to $\nu[G]$. This shows that

$$\nu\big[G\big] \le \liminf_{n\to\infty} \nu_n\big[G\big],$$

and thus establishes weak convergence $\nu_n \xrightarrow{\text{w}} \nu$ by characterization (ii) of Portmanteau Theorem (Theorem VI.4). □

3. Metrizability of weak convergence

Weak convergence, Definition VI.1, says which sequences of Borel probability measures on a metric space $\mathfrak{X}$ converge. This defines a topology on the set $\mathscr{M}_1(\mathfrak{X})$ of Borel probability measures on $\mathfrak{X}$ (this is called the topology of weak convergence). In general, a topology may or may not come from a metric (and if it does, there are many different metrics giving rise to the same topology) — a topology is said to be metrizable if it does. Metrizable topologies have many good properties, and a metric makes many things more concrete.

The topology of weak convergence turns out to be metrizable if the underlying space $\mathfrak{X}$ is separable. This is obtained as a consequence of Exercises VI.1, VI.2, and VI.3 below.

Let $(\mathfrak{X}, \varrho)$ be a metric space, and let $\mathscr{M}_1(\mathfrak{X})$ be the set of Borel probability measures on $\mathfrak{X}$. For a Borel set $E \in \mathscr{B}(\mathfrak{X})$ and $\delta > 0$, denote $E^\delta = \big\{x \in \mathfrak{X} \,\big|\, \varrho(x, E) < \delta\big\}$,

where $\varrho(x,E) = \inf_{y\in E}\big(\varrho(x,y)\big)$. For $\mu,\nu \in \mathscr{M}_1(\mathfrak{X})$, define

$$\varrho_{\mathrm{LP}}(\mu,\nu) = \inf\Big\{\delta>0 \;\Big|\; \forall E\in\mathscr{B}(\mathfrak{X}):\ \mu[E]\le \nu[E^\delta]+\delta \text{ and } \nu[E]\le \mu[E^\delta]+\delta\Big\}.$$

This ϱ_{LP} is called the Lévy-Prohorov metric.

Exercise VI.1 (Lévy-Prohorov metric is a metric)**.**
Show that ϱ_{LP} is a metric on $\mathscr{M}_1(\mathfrak{X})$.

The following exercise says that the topology induced by the metric ϱ_{LP} on the space $\mathscr{M}_1(\mathfrak{X})$ of probability measures is always at least as strong as the topology of weak convergence.

Exercise VI.2 (Lévy-Prohorov convergence implies weak convergence)**.**
Assume that $\nu_n \in \mathscr{M}_1(\mathfrak{X})$, $n\in\mathbb{N}$, are such that for some $\nu\in\mathscr{M}_1(\mathfrak{X})$ we have $\varrho_{\mathrm{LP}}(\nu_n,\nu)\to 0$ as $n\to\infty$.

(a) Show that there exists a decreasing sequence $(\delta_n)_{n\in\mathbb{N}}$ of real numbers tending to zero, such that for all Borel sets $E\subset\mathfrak{X}$ we have $\nu_n[E]\le\delta_n+\nu[E^{\delta_n}]$.
(b) Show that for any Borel set $E\subset\mathfrak{X}$ we have, as $n\to\infty$, $\nu[E^{\delta_n}]\downarrow\nu[\bar{E}]$.
(c) Show that for any closed set $F\subset\mathfrak{X}$ we have $\limsup_{n\to\infty}\nu_n[F]\le\nu[F]$.
(d) Conclude that ν_n converges weakly to ν as $n\to\infty$.

Conversely, we show that under the assumption of separability of $\mathfrak{X}$, the topology of weak convergence is at least as strong as the topology induced by the metric ϱ_{LP}. In this setup, then, ϱ_{LP} metrizes the topology of weak convergence on the space $\mathscr{M}_1(\mathfrak{X})$ of probability measures.

Exercise VI.3 (In separable spaces weak convergence implies Lévy-Prohorov convergence)**.**
Assume that the metric space $(\mathfrak{X},\varrho)$ is separable. Assume that $\nu_n\in\mathscr{M}_1(\mathfrak{X})$, $n\in\mathbb{N}$, are such that ν_n converges weakly to some $\nu\in\mathscr{M}_1(\mathfrak{X})$ as $n\to\infty$. Let $\varepsilon>0$.

(a) Show that there exists a countable dense set $\{x_i \,|\, i\in\mathbb{N}\}\subset\mathfrak{X}$ and a radius $r\in(\frac{\varepsilon}{4},\frac{\varepsilon}{2})$ such that $\nu[\partial B_r(x_i)]=0$ for all $i\in\mathbb{N}$.
(b) Show that there exists some $k\in\mathbb{N}$ such that $\nu\big[\bigcup_{i=1}^k B_r(x_i)\big]\ge 1-\varepsilon$.
(c) Show that there exist finitely many disjoint Borel sets $A_1,\dots,A_k\subset\mathfrak{X}$ such that $\nu[\partial A_i]=0$ and $\operatorname{diam}(A_i)<\varepsilon$ for all $i=1,\dots,k$, and $\nu\big[\mathfrak{X}\setminus\bigcup_{i=1}^k A_i\big]\le\varepsilon$.
(d) Define the collection $\mathscr{A}=\big\{\bigcup_{i\in I}A_i \,\big|\, I\subset\{1,\dots,k\}\big\}$ of subsets of $\mathfrak{X}$. Show that for any $A\in\mathscr{A}$ we have $\nu_n[A]\to\nu[A]$ as $n\to\infty$. Conclude that there exists $N>0$ such that $\big|\nu_n[A]-\nu[A]\big|\le\varepsilon$ for all $A\in\mathscr{A}$ and $n\ge N$.
(e) Let $E\subset\mathfrak{X}$ be a Borel set. Choose $A\in\mathscr{A}$ as $A=\bigcup_{i\in I_E}A_i$, where $I_E=\big\{i \,\big|\, A_i\cap E\neq\emptyset\big\}$. Show that $A\subset E^\varepsilon$, and that $E\subset A\cup E'$, where $\nu[E']\le\varepsilon$ and $\nu_n[E']\le 2\varepsilon$ for $n\ge N$.
(f) Show that for $n\ge N$ we have $\nu_n[E]\le\nu[E^\varepsilon]+3\varepsilon$ and $\nu[E]\le\nu_n[E^\varepsilon]+2\varepsilon$.
(g) Conclude that $\varrho_{\mathrm{LP}}(\nu_n,\nu)\to 0$ as $n\to\infty$.

Lévy-Prohorov metric and close couplings

Recall that a coupling of probability measures $\mu,\nu\in\mathscr{M}_1(\mathfrak{X})$ is a probability measure λ on $\mathfrak{X}\times\mathfrak{X}$ such that for all $E\in\mathscr{B}(\mathfrak{X})$ we have $\lambda[E\times\mathfrak{X}]=\mu[E]$ and $\lambda[\mathfrak{X}\times E]=\nu[E]$. The next exercise will help get an intuition of the meaning of the above metric ϱ_{LP} on the space $\mathscr{M}_1(\mathfrak{X})$ of probability measures. It says that if

two probability measures can be coupled so that the realizations of the two components are very close with a very high probability, then the probability measures are very close.

Exercise VI.4 (Lévy-Prohorov metric and couplings)**.**

Assume that there exists a coupling λ of μ and ν in which

$$\lambda\Big[\,\big\{(x_1,x_2)\in\mathfrak{X}\times\mathfrak{X}\,\big|\,\varrho(x_1,x_2)\geq\varepsilon\big\}\,\Big]\leq\varepsilon.$$

Show that then $\varrho_{\mathrm{LP}}(\mu,\nu)\leq\varepsilon$.

Lecture VII

Tightness and Prokhorov's theorem

In this chapter we study weak convergence in metric spaces further. In particular, we define and study tightness of a family of probability measures, and we state and prove Prohorov's theorem relating tightness to weak convergence.

1. Tightness and precompactness of probability measures

Recall that an often practical strategy for proving convergence is that of Exercise H.12: show precompactness of the sequence under consideration and identify uniquely any subsequential limit. Precompactness in the topology of weak convergence is closely related to the more concrete notion of *tightness*, defined below in a general setup.

For comparison, keep in mind our results in the case of weak convergence on the real axis, in particular Definition IV.6, Theorem IV.8, and Exercise IV.4. This section and the next generalize these to the context of probability measures on complete separable metric spaces.

Definition VII.1 (Tightness of a family of probability measures)**.**
A collection $(\nu_i)_{i\in I}$ of probability measures on a metric space $\mathfrak{X}$ is *tight*, if for any $\varepsilon > 0$ there exists some compact subset $K \subset \mathfrak{X}$ such that

$$\nu_i\big[K\big] > 1 - \varepsilon \qquad \forall i \in I.$$

Definition VII.2 (Precompactness of a family of probability measures)**.**
A collection $(\nu_i)_{i\in I}$ of probability measures on a metric space $\mathfrak{X}$ is *precompact* (in the topology of weak convergence), if any sequence $(\nu_{i_n})_{n\in\mathbb{N}}$ of probability measures from the collection has a subsequence which converges weakly.

Prohorov's theorem states that in complete separable metric spaces (or in topological spaces homeomorphic to complete separable metric spaces, i.e., Polish spaces) tightness and precompactness of a family of probability measures are equivalent.

2. Prohorov's theorem

The direct half of Prohorov's theorem

By far the more commonly needed direction of Prohorov's theorem is that tightness of a family of probability measures implies its precompactness. In more concrete terms, from a tight sequence of probability measures it is always possible to extract a

weakly convergent subsequence. This implication in fact holds for all metric spaces, even without the assumptions of completeness and separability.

Theorem VII.3 (Prohorov's theorem: the direct half)**.**
If a collection $(\nu_i)_{i\in I}$ of probability measures on a metric space $\mathfrak{X}$ is tight, then it is precompact.

Proof. Assume that $(\nu_n)_{n\in\mathbb{N}}$ is a sequence from a tight collection of probability measures on $\mathfrak{X}$. We want to show that some subsequence $(\nu_{n_k})_{k\in\mathbb{N}}$ converges weakly.

For the rest of the proof, we use the following notational convention for subsets of $\mathfrak{X}$: subsets denoted by G are assumed open, subsets denoted by F are assumed closed, and subsets denoted by K or H are assumed compact (and the same convention is applied whether or not there are further subscript indices).

First, by tightness, we can choose compact subsets $K_1 \subset K_2 \subset K_3 \subset \cdots \subset \mathfrak{X}$ such that $\nu_n[K_m] > 1-\frac{1}{m}$ for all n, m. Each compact K_m is separable (Exercise H.11) so the countable union $\bigcup_m K_m$ is also separable. This union contains all the probability mass of all all members of the sequence, $\nu_n\big[\bigcup_m K_m\big] = 1$ for all n. We can therefore essentially only work on the separable subset $\bigcup_m K_m \subset \mathfrak{X}$.

Instead of working directly with all Borel sets, we will start with a certain good countable collection of subsets. First, let $S \subset \bigcup_m K_m$ be a countable dense set of points in the separable set $\bigcup_m K_m$. Let $\mathscr{A}$ consist of all open balls $B_r(s)$ with $r \in \mathbb{Q}\cap(0,\infty)$ and $s \in S$. Note that $\mathscr{A}$ is indeed countable. It has the following important property:

(a): If $G \subset \mathfrak{X}$ is an open set and $x \in G \cap \bigcup_m K_m$, then there exists some $A \in \mathscr{A}$ such that $x \in A \subset \overline{A} \subset G$.

Indeed, for such G and x we can first find some $\varepsilon > 0$ such that $B_\varepsilon(x) \subset G$, and then some $s \in S$ such that $\varrho(x,s) < \frac{\varepsilon}{2}$. Then choose $r \in \mathbb{Q}\cap(\varrho(x,s), \frac{\varepsilon}{2})$. Now $A = B_r(s) \in \mathscr{A}$ works: clearly $x \in A$ and $\overline{A} = \overline{B}_r(s) \subset \overline{B}_{r+\varepsilon/2}(x) \subset B_\varepsilon(x) \subset G$, as required for (a).

This collection $\mathscr{A}$ is a fair starting point, but even better is the collection $\mathscr{H}$ which consists of all finite unions of closures of sets from the collection $\mathscr{A}$ intersected with one of the chosen compact subsets, i.e.,

$$\mathscr{H} := \Big\{ \big(\overline{A}_1 \cup \cdots \cup \overline{A}_n\big) \cap K_m \;\Big|\; A_1, \ldots, A_n \in \mathscr{A} \Big\}$$

(the empty set $\emptyset \in \mathscr{H}$ is obtained by the union has $n = 0$ terms). Note that each $H \in \mathscr{H}$ is compact, as a closed subset of a compact set. The collection $\mathscr{H}$ is countable, and stable under finite unions. Moreover, each compact K_m has a finite covering by sets from the collection $\mathscr{A}$, and thus we have $K_m \in \mathscr{H}$.

Our task is to find a subsequence of $\big(\nu_n\big)_{n\in\mathbb{N}}$ which converges weakly. We will first make sure that the measures of the sets from the collection $\mathscr{H}$ converge. For any $H \in \mathscr{H}$, the sequence $(\nu_n[H])_{n\in\mathbb{N}}$ is bounded, so we can find a convergent subsequence, and by diagonal extraction we find a subsequence $\big(\nu_{n_k}\big)_{k\in\mathbb{N}}$ such that:

(b): For each set H in the countable collection $\mathscr{H}$, the following limit exists

$$\alpha[H] := \lim_{k\to\infty} \nu_{n_k}[H]. \tag{VII.1}$$

Note that these $\alpha[H]$ straightforwardly inherit some properties because of their construction as limits of probabilities $\nu_{n_k}[H]$. In particular, we have $0 \le \alpha[H] \le 1$ for any $H \in \mathscr{H}$, and

$$\begin{aligned}
\alpha[\emptyset] &= 0, && \\
\alpha[H_1] &\le \alpha[H_2] && \text{if } H_1 \subset H_2 \text{ and } H_1, H_2 \in \mathscr{H}, \\
\alpha[H' \cup H''] &\le \alpha[H'] + \alpha[H''] && \text{for any } H', H'' \in \mathscr{H}, \\
\alpha[H' \cup H''] &= \alpha[H'] + \alpha[H''] && \text{for any disjoint } H', H'' \in \mathscr{H}.
\end{aligned}$$

However, α is not yet a probability measure — it is not even defined on a σ-algebra, but only on the collection $\mathscr{H}$. Instead, from these $\alpha[H]$ we proceed step by step, and ultimately construct a Borel probability measure ν which is the weak limit of ν_{n_k} along the above subsequence.

For all open sets $G \subset \mathfrak{X}$ we define

$$\beta[G] := \sup_{H \subset G} \alpha[H], \tag{VII.2}$$

and for arbitrary subsets $E \subset \mathfrak{X}$ define

$$\gamma[E] := \inf_{G \supset E} \beta[G]. \tag{VII.3}$$

The strategy, carried out below, is to show that γ is an outer measure, and that its restriction to Borel sets is the probability measure we need.

Note that directly from the constructions (VII.2), and (VII.3), we have the following monotonicity properties:

$$\begin{aligned} &\beta[G_1] \leq \beta[G_2] && \text{for } G_1, G_2 \subset \mathfrak{X} \text{ s.t. } G_1 \subset G_2 \\ &\gamma[E_1] \leq \gamma[E_2] && \text{for } E_1, E_2 \subset \mathfrak{X} \text{ s.t. } E_1 \subset E_2, \end{aligned}$$

and also $0 \leq \beta[G] \leq 1$ and $0 \leq \gamma[E] \leq 1$ for any open $G \subset \mathfrak{X}$ and any $E \subset \mathfrak{X}$. The monotonicity of β and definition (VII.3) in particular imply that for any open set $G \subset \mathfrak{X}$ we have $\gamma[G] = \beta[G]$.

We first claim that:

(c): If $F \subset G$ open, and for some $H \in \mathscr{H}$ are such that $F \subset H$, then there exists a $H' \in \mathscr{H}$ such that

$$F \subset H' \subset G.$$

To see this, note first that F is compact as a closed subset of the compact $H \in \mathscr{H}$. By (a), for each $x \in F \subset G \cap \bigcup_m K_m$ there exists some open ball $A \in \mathscr{A}$ such that $x \in A \subset \overline{A} \subset G$. Such open balls A cover the compact F, so already finitely many of them suffice to cover F, and the union of closures of these intersected with one of the compacts K_m is a suitable H' for (c).

We then claim that β is finitely subadditive:

(d): For open sets $G_1, G_2 \subset \mathfrak{X}$, we have

$$\beta[G_1 \cup G_2] \leq \beta[G_1] + \beta[G_2].$$

Suppose that $H \in \mathscr{H}$ is such that $H \subset G_1 \cup G_2$. Define

$$\begin{aligned} F_1 &= \left\{ x \in H \;\middle|\; \varrho(x, G_1^c) \geq \varrho(x, G_2^c) \right\} \\ F_2 &= \left\{ x \in H \;\middle|\; \varrho(x, G_1^c) \leq \varrho(x, G_2^c) \right\}. \end{aligned}$$

From these definitions it follows that $F_1 \subset G_1$: indeed, supposing that $x \notin G_1$ we have $\varrho(x, G_1^c) = 0$, but if we would also have $x \in F_1 \subset H \subset G_1 \cup G_2$, then necessarily $x \in G_2$ and thus $\varrho(x, G_2^c) > 0$, in contradiction with the definition of F_1. Similarly we have $F_2 \subset G_2$. Using property (c) to both $F_1 \subset G_1$ and $F_2 \subset G_2$, we find sets $H_1, H_2 \in \mathscr{H}$ such that $F_1 \subset H_1 \subset G_1$ and $F_2 \subset H_2 \subset G_2$. By construction (VII.1), α is subadditive and monotone, so since $H \subset H_1 \cup H_2 \in \mathscr{H}$, we get

$$\alpha[H] \leq \alpha[H_1 \cup H_2] \leq \alpha[H_1] + \alpha[H_2] \leq \beta[G_1] + \beta[G_2].$$

Taking the supremum over $H \subset G_1 \cup G_2$ we now get (d).

We then claim that β is in fact countably subadditive:

(e): For open sets $G_1, G_2, \ldots \subset \mathfrak{X}$, we have

$$\beta\Big[\bigcup_{n=1}^{\infty} G_n \Big] \leq \sum_{n=1}^{\infty} \beta[G_n].$$

Indeed, if we have $H \subset \bigcup_{n=1}^{\infty} G_n$ for $H \in \mathscr{H}$, then by compactness of H this open cover has a finite subcover, so in fact $H \subset \bigcup_{n=1}^{n_0} G_n$ for some n_0. The definition (VII.2) of β and finite subadditivity (d) of β imply that

$$\alpha[H] \le \beta\Big[\bigcup_{n=1}^{n_0} G_n\Big] \le \sum_{n=1}^{n_0} \beta[G_n] \le \sum_{n=1}^{\infty} \beta[G_n].$$

Taking the supremum over $H \subset \bigcup_{n=1}^{\infty} G_n$ now yields (e).

Now we claim that:

(f): γ is an outer measure.

Let us verify the defining properties of outer measures for γ. We have already seen that γ is monotone. Also, since $\alpha[\emptyset] = 0$, we have $\beta[\emptyset] = 0$ and thus also $\gamma[\emptyset] = 0$ — directly from their respective definitions. It only remains to show that γ is countably subadditive. Let $E_1, E_2, \ldots \subset \mathfrak{X}$ and let $\varepsilon > 0$. For each $n \in \mathbb{N}$, choose an open subset $G_n \supset E_n$ such that $\beta[G_n] \le \gamma[E_n] + 2^{-n}\varepsilon$. Then by definition of γ and countable subadditivity of β we get

$$\gamma\Big[\bigcup_{n=1}^{\infty} E_n\Big] \le \beta\Big[\bigcup_{n=1}^{\infty} G_n\Big] \le \sum_{n=1}^{\infty} \beta[G_n] \le \sum_{n=1}^{\infty} \big(\gamma[E_n] + 2^{-n}\varepsilon\big) = \sum_{n=1}^{\infty} \gamma[E_n] + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, countable subadditivity of γ follows, establishing (f).

Our next claim is an auxiliary result, which will help us verify measurability of closed sets:

(g): For any open $G \subset \mathfrak{X}$ and closed $F \subset \mathfrak{X}$ we have

$$\beta[G] \ge \gamma[F \cap G] + \gamma[F^c \cap G].$$

To check this, let $\varepsilon > 0$. Choose $H' \subset F^c \cap G$ such that $\alpha[H'] \ge \beta[F^c \cap G] - \frac{\varepsilon}{2}$. Then choose $H'' \subset (H')^c \cap G$ such that $\alpha[H''] \ge \beta[(H')^c \cap G] - \frac{\varepsilon}{2}$. By construction H' and H'' are disjoint, and α is additive on disjoint sets, so we get

$$\begin{aligned}\alpha[H' \cup H''] = \alpha[H'] + \alpha[H''] &\ge \beta[F^c \cap G] + \beta[(H')^c \cap G] - \varepsilon \\ &\ge \gamma[F^c \cap G] + \gamma[F \cap G] - \varepsilon.\end{aligned}$$

Since $H := H' \cup H'' \subset G$ and $H \in \mathscr{H}$, by definition of β we get $\beta[G] \ge \gamma[F^c \cap G] + \gamma[F \cap G] - \varepsilon$, and since $\varepsilon > 0$ was arbitrary, we get (g).

Now we can check that there are enough measurable sets:

(h): Every closed set $F \subset \mathfrak{X}$ satisfies Carathéodory's criterion

$$\gamma[E] \ge \gamma[F \cap E] + \gamma[F^c \cap E] \qquad \forall E \subset \mathfrak{X},$$

and as such is γ-measurable.

Indeed, for any $E \subset \mathfrak{X}$ and any open set $G \supset E$ we have by property (g) and monotonicity of γ that

$$\beta[G] \ge \gamma[F \cap G] + \gamma[F^c \cap G] \ge \gamma[F \cap E] + \gamma[F^c \cap E],$$

so by taking infimum over such G we obtain (h).

Since all closed sets are γ-measurable, it follows that all Borel sets are γ-measurable, so the restriction of the outer measure γ to Borel sets defines a Borel measure ν. This ν is in fact a probability measure: it is clear that $\nu[\mathfrak{X}] = \gamma[\mathfrak{X}] \le 1$ and on the other hand

$$\nu[\mathfrak{X}] = \gamma[\mathfrak{X}] = \beta[\mathfrak{X}] \ge \alpha[K_m] \ge 1 - \frac{1}{m} \qquad \text{for any } m \in \mathbb{N},$$

so we indeed have $\nu[\mathfrak{X}] = 1$.

We are now ready to finish the proof of Prohorov's theorem:

(*): For the subsequence indexed by n_k as in (b), we have $\nu_{n_k} \xrightarrow[k\to\infty]{\text{w}} \nu$.

Let us verify this weak convergence by criterion (ii) of Theorem VI.4. Let $G \subset \mathfrak{X}$ be open. For any $H \in \mathscr{H}$ such that $H \subset G$ we have, by (b) and monotonicity for the measures ν_{n_k},

$$\alpha[H] = \lim_{k\to\infty} \nu_{n_k}[H] \le \liminf_{k\to\infty} \nu_{n_k}[G].$$

Observe also that for the open set G we have by construction, and using the bound for $\alpha[H]$ above, that

$$\nu[G] = \gamma[G] = \beta[G] = \sup_{H \subset G} \alpha[H] \leq \liminf_{k\to\infty} \nu_{n_k}[G].$$

This is the characterizing criterion (ii) in Theorem VI.4, so we have established the weak convergence of a subsequence

$$\nu_{n_k} \xrightarrow[k\to\infty]{\text{w}} \nu,$$

and the proof is complete. □

The converse half of Prohorov's theorem

The other implication in Prohorov's theorem is that precompactness of a family of probability measures implies its tightness. This direction is less often used in applications, but it provides valuable perspective. For example, applied to a collection consisting of a single probability measure (which is obviously precompact), it implies that all probability mass except for an arbitrarily small fraction is found on some compact subset. For the implication in this direction we need the assumption that the space $\mathfrak{X}$ is complete and separable.

Theorem VII.4 (Prohorov's theorem: the converse half)**.**
If $\mathfrak{X}$ is a complete separable metric space, and if a collection $(\nu_i)_{i\in I}$ of probability measures on $\mathfrak{X}$ is precompact, then $(\nu_i)_{i\in I}$ is tight.

Proof. Assume that $(\nu_i)_{i\in I}$ is a precompact collection of probability measures on a complete separable metric space $\mathfrak{X}$.

We first claim the following.

(a) Whenever $G_1, G_2 \ldots$ are open sets such that $\bigcup_{n\in\mathbb{N}} G_n = \mathfrak{X}$, then for any $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that we have

$$\nu_i\Big[\bigcup_{n=1}^{N} G_n\Big] > 1 - \varepsilon \qquad \text{for all } i \in I.$$

Indeed, suppose by contrapositive that that this is not the case. Then for some $\varepsilon > 0$ we can find indices $i_1, i_2, \ldots \in I$ such that

$$\nu_{i_k}\Big[\bigcup_{n=1}^{k} G_n\Big] \leq 1 - \varepsilon \qquad \text{for all } k \in \mathbb{N}.$$

In particular, for any fixed $m \in \mathbb{N}$ we have

$$\liminf_k \nu_{i_k}\Big[\bigcup_{n=1}^{m} G_n\Big] \leq \liminf_k \nu_{i_k}\Big[\bigcup_{n=1}^{k} G_n\Big] \leq 1 - \varepsilon.$$

By precompactness, the sequence $(\nu_{i_k})_{k\in\mathbb{N}}$ has a subsequence $(\nu_{i_{k_\ell}})_{\ell\in\mathbb{N}}$ which converges weakly to a probability measure ν. Then by criterion (ii) for weak convergence in Theorem VI.4 we have, for any $m \in \mathbb{N}$,

$$\nu\Big[\bigcup_{n=1}^{m} G_n\Big] \leq \liminf_\ell \nu_{i_{k_\ell}}\Big[\bigcup_{n=1}^{m} G_n\Big] \leq 1 - \varepsilon.$$

But as $m \to \infty$, we have $\bigcup_{n=1}^{m} G_n \uparrow \mathfrak{X}$ by assumption, monotone convergence of measures yields

$$\lim_{m\to\infty} \nu\Big[\bigcup_{n=1}^{m} G_n\Big] = \nu[\mathfrak{X}] = 1.$$

This is a contradiction, which then proves (a).

Now let $\varepsilon > 0$.

Using separability, choose a countable dense subset $\{s_1, s_2, \ldots\} \subset \mathfrak{X}$. For each $m \in \mathbb{N}$, the open balls of radius $\frac{1}{m}$ centered at the points of this dense subset cover $\mathfrak{X}$, i.e,

$$\bigcup_{n=1}^{\infty} B_{1/m}(s_n) = \mathfrak{X}.$$

By (a), we can therefore find some $N_m \in \mathbb{N}$ such that

$$\nu_i\Big[\bigcup_{n=1}^{N_m} B_{1/m}(s_n)\Big] > 1 - \varepsilon\, 2^{-m} \qquad \text{for all } i \in I.$$

Define

$$K := \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{N_m} \overline{B}_{1/m}(s_n),$$

and note that for any $i \in I$, the ν_i-probability of the complement of K is at most

$$\begin{aligned}\nu_i\big[K^c\big] &= \nu_i\Big[\bigcup_{m=1}^{\infty} \big(\bigcup_{n=1}^{N_m} \overline{B}_{1/m}(s_n)\big)^c\Big] \\ &\le \sum_{m=1}^{\infty} \nu_i\Big[\big(\bigcup_{n=1}^{N_m} \overline{B}_{1/m}(s_n)\big)^c\Big] \le \sum_{m=1}^{\infty} \varepsilon\, 2^{-m} = \varepsilon.\end{aligned}$$

If we now show that K is compact, then tightness of $(\nu_i)_{i\in I}$ follows.

Note first that K is closed. To prove that K is compact, consider a sequence $x_1, x_2, \ldots \in K$. For every $m \in \mathbb{N}$, since $K \subset \bigcup_{n=1}^{N_1} \overline{B}_1(s_n)$, there exists an index $n_1 \le N_1$ such that infinitely many members of the sequence are in $\overline{B}_1(s_{n_1})$. We can therefore choose a subsequence which lies entirely in $B_1(s_{n_1})$. But since also $K \subset \bigcup_{n=1}^{N_2} \overline{B}_{1/2}(s_n)$, there exists an index $n_2 \le N_2$ such that infinitely many members of the subsequence are in $\overline{B}_{1/2}(s_{n_2})$, and we can choose a further subsequence which lies entirely in $\overline{B}_{1/2}(s_{n_2})$. Continuing inductively, and finally using diagonal extraction, we find a subsequence $(x_{\ell_j})_{j\in\mathbb{N}}$ such that for all $m \in \mathbb{N}$ and $j \ge m$ we have $x_{\ell_j} \in B_{1/m}(s_{n_m})$. This implies that the subsequence is Cauchy: for any $j, j' \ge m$ we have $\varrho(x_{\ell_j}, x_{\ell_{j'}}) \le \frac{2}{m}$. By completeness, the limit $x := \lim_{j\to\infty} x_{\ell_j}$ of the subsequence exists. Since K is closed, the limit x is in K, and we have proven that K is sequentially compact. Thus K is compact.

We have proved that the collection $(\nu_i)_{i\in I}$ is tight. □

3. Weak convergence in countable product of finite sets using cylinders

Let S be a finite set and I a countable index set, and consider the space $\mathfrak{X} = S^I$, which can be made a complete separable compact metric space, as in Appendix H.4. We will use this type of space both in our study of the d-dimensional Ising model, in which case $S = \{-1, +1\}$ and $I = \mathbb{Z}^d$, and for interacting particle systems, in which case S is a finite set of states depending on the system and again $I = \mathbb{Z}^d$. Combining Prokhorov's theorem and a sufficient condition for weak convergence presented in the previous lecture, we can now give a rather simple and practical characterization of weak convergence on spaces of this type.

For $i \in I$, denote by $\pi_i \colon S^I \to S$ the projection to the i:th coordinate. Subsets $C \subset S^I$ of the form

$$C = \pi_{i_1}^{-1}(A_1) \cap \cdots \cap \pi_{i_k}^{-1}(A_k) \subset S^I \tag{VII.4}$$

are called *cylinder sets*. Recall also from Appendix H.4 that the collection Π of all cylinder sets have the following properties:

- The collection Π is stable under finite intersections.
  (Proposition H.16(iv))
- The collection Π is countable.
  (Lemma H.17)
- Any open set $G \subset S^I$ is the union of all cylinder sets C contained in it.
  (Proposition H.16(iii))
- Any cylinder set C is both open and closed.
  (Proposition H.16(i)&(ii))

It follows that the collection Π of all cylinder sets satisfies the two properties of Proposition VI.6.

Theorem VII.5 (Weak convergence with cylider events)**.**
A sequence $(\nu_n)_{n\in\mathbb{N}}$ of probability measures on the space S^I converges weakly if and only if for every cylinder set C the limit $\lim_{n\to\infty}\nu_n[C]$ exists. The limit measure ν is uniquely determined by the property that $\nu[C] = \lim_{n\to\infty}\nu_n[C]$ for all C.

Proof of "only if": Suppose that $\nu_n \xrightarrow{\text{w}} \nu$, and let $C \subset S^I$ be a cylinder set. The cylinder C is both open and closed (Prop. H.16(i)&(ii)), and therefore $C^\circ = C = \overline{C}$ and $\partial C = \overline{C}\setminus C^\circ = \emptyset$. Thus obviously $\nu[\partial C] = 0$, so by condition (iv) of Portmanteau theorem (Theorem VI.4) we have $\nu_n[C] \to \nu[C]$.

Proof of "if": As remarked above, the collection $\mathcal{C}$ of all cylinder sets satisfies the two properties of Proposition VI.6. By that proposition, then, a sufficient condition for weak convergence is that for all cylinder sets C we have $\nu_n[C] \to \nu[C]$, where ν is some probability measure on S^I. We are assuming that $\alpha[C] := \lim_{n\to\infty}\nu_n[C]$ exists for all $C \in \Pi$, so it remains to show that α is a probability measure restricted to cylinder sets. Recall now that S^I is compact (Exercise H.13(d)), and therefore $(\nu_n)_{n\in\mathbb{N}}$ is automatically tight. By Prohorov's theorem (Theorem VII.4) there exists some subsequence $(\nu_{n_k})_{k\in\mathbb{N}}$ which converges weakly $\nu_{n_k} \xrightarrow{\text{w}} \nu$. Again since $\partial C = \emptyset$, we get that $\nu[C] = \lim_{k\to\infty}\nu_{n_k}[C] = \alpha[C]$ for all $C \in \Pi$.

Proof of uniqueness: By Corollary H.19, the probabilities of cylinder events $C \in \Pi$ uniquely determine a Borel probability measure on S^I. In particular, ν is determined by the condition $\nu[C] = \lim_{n\to\infty}\nu_n[C]$ for all $C \in \Pi$. □

Lecture VIII

Random walks and Brownian motion

Consider the simple random walk, defined using a sequence $(\xi_\ell)_{\ell\in\mathbb{N}}$ of steps, which are independent and indentically distributed with law $\mathsf{P}\big[\xi_\ell = -1\big] = 1/2 = \mathsf{P}\big[\xi_\ell = +1\big]$, and the sums of steps $S = (S_k)_{k\in\mathbb{Z}_{\geq 0}}$

$$S_k = \sum_{\ell=1}^{k} \xi_\ell.$$

From these values of the random walk at integer times, we extend piecewise linearly to a function defined for all times $t \in [0,\infty)$ by setting

$$S_t = S_{\lfloor t\rfloor} + \big(t - \lfloor t\rfloor\big)\, \xi_{\lfloor t\rfloor+1},$$

where $\lfloor t\rfloor \in \mathbb{Z}$ denotes the integer part of the real number $t \in \mathbb{R}$. Figure VIII.1(a) illustrates this piecewise linear interpolation $(S_t)_{t\in[0,\infty)}$ of the random walk $(S_k)_{k\in\mathbb{Z}_{\geq 0}}$.

Now fix a small scale parameter $a > 0$. The rescaled random walk $X^{(a)} = (X_t^{(a)})_{t\in[0,\infty)}$ with steps of magnitude $\sqrt{a}$ occurring at time intervals of length a is the process defined by

$$X_t^{(a)} = \sqrt{a}\, S_{t/a}. \tag{VIII.1}$$

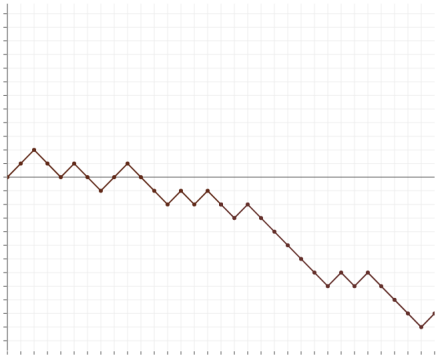
This scaling of random walks is illustrated in Figure VIII.1. Our goal is to prove the following result (the precise formulation and proof will be given after some preparations).

> **Donsker's theorem.** *The scaled random walks $X^{(a)}$ defined by* (VIII.1) *converge weakly as $a \downarrow 0$ to Brownian motion, on a space of continuous functions.*

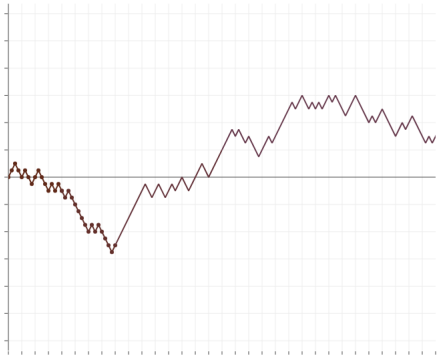
The proof strategy is the usual one, "precompactness plus identification of subsequential limits" (recall Exercise H.12 in Appendix H):

- Verify that the laws of the processes $X^{(a)}$ for different $a > 0$ form a tight family, and thus obtain by Prohorov's theorem (Theorem VII.4) that any sequence $X^{(a_m)}$ with $a_m \downarrow 0$ has convergent subsequences.
- Characterize explicitly any subsequential limit of the processes $X^{(a)}$.

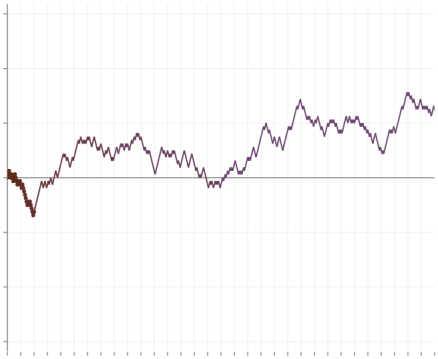
In Section 1 we give the defining properties of Brownian motion. In Section 2 we consider generalities about determining a Borel probability measure on the space of continuous functions, study compact subsets of a space of continuous functions, and establish a criterion for tightness. In Section 3 we give the precise statement of Donsker's theorem and finish the proof by proving tightness of the laws of $X^{(a)}$ for different $a > 0$ and showing that any subsequential limit is a Brownian motion.



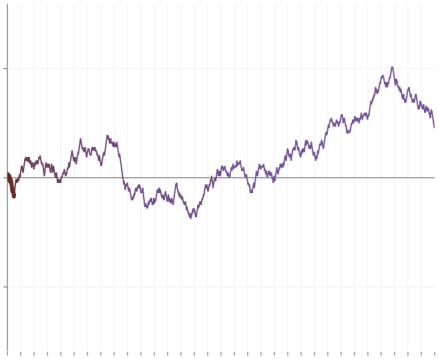
(a) The first 64 steps of a random walk.



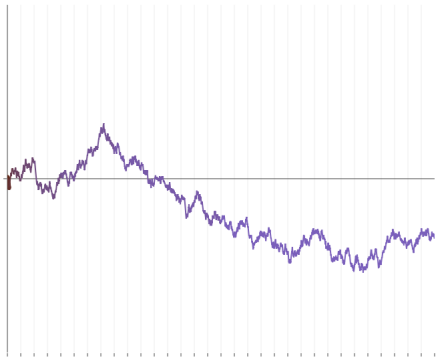
(b) The first 256 steps of the same walk.



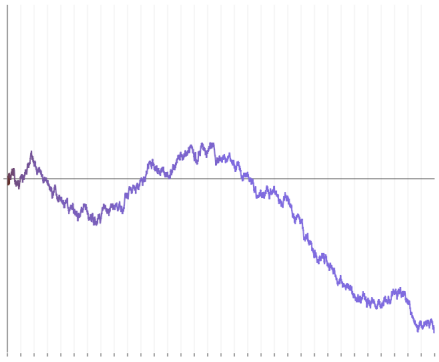
(c) The first 1024 steps of the same walk.



(d) The first 4096 steps of the same walk.



(e) The first 16384 steps of the same walk.



(f) The first 65536 steps of the same walk.

FIGURE VIII.1. A random walk in different temporal and spatial scales. The scaling used to zoom out is according to (VIII.1). The different scalings of the spatial and temporal axes is made apparent by the ticks on the two axes, which in each figure mark an equal number of original spatial and temporal units.

1. Brownian motion

In general, a (real-valued) stochastic process is a collection $(X_t)_{t\in\mathbb{T}}$ of random variables $X_t\colon \Omega \to \mathbb{R}$ (defined on the same probability space Ω) indexed by a "time parameter" t. The set $\mathbb{T} \subset \mathbb{R}$ of allowed time parameter values may be either a discrete set or an interval, and the process is correspondingly said to have discrete or continuous time. We interpret $X_t \in \mathbb{R}$ as the (random) position of the process at time t, dependent as usual on the (random) outcome $\omega \in \Omega$. An outcome $\omega \in \Omega$ determines in fact a function $t \mapsto X_t(\omega)$ of time, called the *path* (or *trajectory*) of the process. Often we have $0 \in \mathbb{T}$ and $\mathbb{T} \subset [0, +\infty)$, and we think that the process is started at time $t = 0$ from position $X_0 \in \mathbb{R}$, which could in general be random, although it is common to consider processes started deterministically from the origin, for example.

The Brownian motion $(B_t)_{t\geq 0}$ is a stochastic process, which is instrumental to a great number of applications — from finance to physics. In this section we will give defining properties of Brownian motion. The Brownian motion not only has a number of different applications, it also has a number of equivalent definitions. We first separately discuss the following properties

- Gaussianity of a process
- stationarity and independence of the increments of a process
- continuous paths of a process

and then give two equivalent definitions by requiring some combinations of such properties. It will not be immediately obvious that any process satisfying the requirements exist. There are various ways of proving the existence of Brownian motion, all of which require some amount of work. By the end of this chapter, we will be able to conclude the existence by showing that the rescaled random walks VIII.1 have a limit as $a \downarrow 0$ which satisfies the defining properties of a Brownian motion.

Gaussian processes

Recall from Appendix G.2 that a random vector X in a finite dimensional vector space $\mathbb{R}^d$ is said to have a Gaussian distribution if all one-dimension projections $a \cdot X$ (for $a \in \mathbb{R}^d$) of it are one-dimensional Gaussian random variables. In a similar vein, we define a process Gaussian process by the condition that the collection of its values on any finite set of times is a Gaussian vector.

Definition VIII.1 (Gaussian process)**.**
A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is a *Gaussian process*, if for all m and $t_1, t_2, \ldots, t_m \in \mathbb{T}$, the vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_m})$ is a Gaussian vector.

The laws of $(X_{t_1}, X_{t_2}, \ldots, X_{t_m})$ for all possible choices $t_1, t_2, \ldots, t_m \in \mathbb{T}$ are called the *finite dimensional distributions* (or *finite dimensional marginals*) of the process.

Example VIII.2 (Random walk with Gaussian steps is a Gaussian process)**.**
The random walk with i.i.d. Gaussian steps considered in Lecture III is a Gaussian process with $\mathbb{T} = \mathbb{Z}_{\geq 0}$.

Stationary and independent increments

Suppose now that for the process $(X_t)_{t\in\mathbb{T}}$ the set of allowed time parameter values $\mathbb{T} \subset \mathbb{R}$ is an additive semigroup, i.e., $0 \in \mathbb{T}$ and whenever $t, s \in \mathbb{T}$ then also $s + t \in \mathbb{T}$ (for example $\mathbb{T} = \mathbb{Z}_{\geq 0}$ or $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = [0, \infty)$). Then for a given $t, s \in \mathbb{T}$ we can consider the increment $X_{s+t} - X_s$ from time s to time $s + t$. Stationarity of increments says that such increments over all time intervals of the same duration t have the same distribution (and more generally a similar property for the joint law of several increments). Independence of increments says that the increments on time intervals that do not overlap are independent random variables.

Definition VIII.3 (Stationary increments)**.**
A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ has *stationary increments*, if for all m and $t_1, t_2, \ldots, t_m \in \mathbb{T}$ and $s \in \mathbb{T}$, the vector $(X_{s+t_1} - X_s, \ldots, X_{s+t_m} - X_s)$ has the same law as the vector $(X_{t_1} - X_0, \ldots, X_{t_m} - X_0)$.

Definition VIII.4 (Independent increments)**.**
A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ has *independent increments*, if for all m and $t_0, t_1, \ldots, t_m \in \mathbb{T}$ with $t_0 < t_1 < \cdots < t_m$ the collection $(X_{t_j} - X_{t_{j-1}})_{j=1,\ldots,m}$ of random variables is independent.

Continuity of a process

Suppose that $\mathbb{T}$ is an interval, for example $\mathbb{T} = [0, T]$, or $\mathbb{T} = [0, \infty)$, or $\mathbb{T} = \mathbb{R}$. In such a setup of continuous time, it is meaningful to consider te continuity of the paths of a process $(X_t)_{t\in\mathbb{T}}$.

Definition VIII.5 (Continuous paths of a stochastic process)**.**
A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ has *continuous paths*, if

$$\mathsf{P}\Big[\Big\{\omega \in \Omega \;\Big|\; t \mapsto X_t(\omega) \text{ is continuous}\Big\}\Big] = 1.$$

Remark VIII.6 (Measurability issues with the continuous paths property)**.**
Note that in the above definition, the event $\big\{\omega \in \Omega \;\big|\; t \mapsto X_t(\omega) \text{ is continuous}\big\}$ must first of all be measurable. Since $\mathbb{T}$ is an interval, this event depends on values X_t of the process at uncountably many times t, so the probability space $(\Omega, \mathscr{F}, \mathsf{P})$ should not be chosen carelessly!

Defining properties of Brownian motion

Proposition VIII.7 (Equivalent descriptions of finite dimensional distributions)**.**
For a stochastic process $X = (X_t)_{t\in[0,\infty)}$ the following are equivalent:

(i) *X has stationary and independent increments, and $X_t \sim \mathrm{N}(0, t)$ for all $t \geq 0$*
(ii) *X is a Gaussian process with $\mathsf{E}[X_t] = 0$ and $\mathsf{E}[X_s X_t] = \min(s, t)$, for all $s, t \geq 0$.*

Proof of (i) ⇒ (ii)*:* Suppose that X has stationary and independent increments, and $X \sim \mathrm{N}(0,t)$ for all $t \geq 0$. Let $0 = t_0 < t_1 < \cdots < t_m$ and $a_1, \ldots, a_m \in \mathbb{R}$. Then we can rewrite

$$\sum_{k=1}^{m} a_k X_{t_k} = \sum_{k=1}^{m} b_k \big(X_{t_k} - X_{t_{k-1}}\big)$$

by choosing suitable coefficients $b_1, \ldots, b_m \in \mathbb{R}$ ($b_m = a_m$, $b_{m-1} = a_{m-1} + a_m$, …). By assumption (i), the increments $X_{t_k} - X_{t_{k-1}}$, $k = 1, \ldots, m$, are independent Gaussians with zero mean. It follows that the linear combination is also Gaussian with zero mean. This shows that $(X_t)_{t\in[0,\infty)}$ is a Gaussian process and $\mathsf{E}[X_t] = 0$ for any $t \geq 0$. It remains to compute the covariance, which is done using the independence of increments: for $s \leq t$ we have

$$\begin{aligned}
\mathsf{E}\big[X_s\, X_t\big] &= \mathsf{E}\Big[X_s\big(X_s + X_t - X_s\big)\Big] \\
&= \mathsf{E}\big[X_s^2\big] + \mathsf{E}\Big[X_s\big(X_t - X_s\big)\Big] \\
&= \mathsf{E}\big[X_s^2\big] + \mathsf{E}\big[X_s\big]\,\mathsf{E}\big[X_t - X_s\big] = s + 0 \cdot 0 = s = \min(s,t).
\end{aligned}$$

Thus the property (ii) follows.

Proof of (ii) ⇒ (i)*:* Suppose that $(X_t)_{t\in[0,\infty)}$ is a Gaussian process with $\mathsf{E}[X_t] = 0$ and $\mathsf{E}[X_s X_t] = \min(s,t)$. In particular then X_t is Gaussian with mean zero and variance $\mathsf{E}[X_t^2] = t$, i.e., we have $X_t \sim \mathrm{N}(0,t)$. Also the increment $X_t - X_s$ for $s \leq t$ is Gaussian, as a linear combination of two values of the Gaussian process. The mean of the increment is

$$\mathsf{E}\big[X_t - X_s\big] = \mathsf{E}\big[X_t\big] - \mathsf{E}\big[X_s\big] = 0 - 0 = 0$$

and its variance is

$$\begin{aligned}
\mathsf{E}\big[(X_t - X_s)^2\big] &= \mathsf{E}\big[X_t^2\big] - 2\,\mathsf{E}\big[X_t\, X_s\big] + \mathsf{E}\big[X_s^2\big] \\
&= t - 2\,\min(s,t) + s = t - 2s + s = t - s.
\end{aligned}$$

We conclude that also $X_t - X_s \sim \mathrm{N}(0, t-s)$, so in particular the increment $X_t - X_s$ has the same law as X_t — a first step towards the stationarity of increments. Finally consider, for $0 = t_0 < t_1 < \cdots < t_m$, the vector

$$(X_{t_1} - X_{t_0}, \ldots, X_{t_m} - X_{t_{m-1}}),$$

which is Gaussian (by the assumption that the process X is Gaussian). The non-diagonal covariances are (assume $1 \leq j < k \leq m$)

$$\begin{aligned}
&\mathsf{E}\Big[\big(X_{t_j} - X_{t_{j-1}}\big)\,\big(X_{t_k} - X_{t_{k-1}}\big)\Big] \\
&= \min(t_j, t_k) - \min(t_{j-1}, t_k) - \min(t_j, t_{k-1}) + \min(t_{j-1}, t_{k-1}) \\
&= t_j - t_{j-1} - t_j + t_{j-1} = 0.
\end{aligned}$$

Since these non-diagonal covariances vanish, the Gaussian components (here increments) are independent by Proposition G.3. Stationarity of the increments follows as well, since the diagonal covariances are $\mathsf{E}\big[(X_{t_j} - X_{t_{j-1}})^2\big] = t_j - t_{j-1}$ by the earlier calculation. □

Definition VIII.8 (Standard Brownian motion)**.**
A stochastic process $B = (B_t)_{t\in[0,\infty)}$ is a *standard Brownian motion*, if it has the following properties

- B has continuous paths
- B satisfies the equivalent conditions of Proposition VIII.7.

Remark VIII.9 (On the existence of standard Brownian motion)**.**
It is not a priori clear that the two requirements in Definition VIII.8 can be simultaneously satisfied, i.e., it is not clear if a standard Brownian motion exists. The equivalent conditions of Proposition VIII.7 could in principle rule out the possibility of continuous paths.[1] We

[1]The finite dimensional distributions do rule out for example differentiable paths: a standard Brownian motion is everywhere non-differentiable, almost surely.

will, however, later in this lecture be able to conclude that a standard Brownian motion exists.

Remark VIII.10 (On the uniqueness of standard Brownian motion)**.**
Besides existence of a standard Brownian motion, another issue is whether the two requirements in Definition VIII.8 uniquely specify the law of the stochastic process $B = (B_t)_{t\in[0,\infty)}$ in question, since the conditions only address the "finite dimensional distributions", i.e., laws of $(B_{t_1}, \dots, B_{t_n})$ involving finitely many times $t_1, \dots, t_n \in [0,\infty)$. Our answer to this issue will be to view the law of B as a probability measure on a space of continuous functions. This is reasonable in view of the fact that by the property of continuous paths, all the probability mass is supported on the set of continuous functions. On such spaces, the finite dimensional distributions do determine a Borel probability measure, as we will discuss below in Proposition VIII.12.

Remark VIII.11 (A counterexample)**.**
Suppose that $B = (B_t)_{t\in[0,\infty)}$ is a standard Brownian motion, for which the two conditions of Definition VIII.8 hold. Let τ be a positive random variable which has a probability density (a "continuous random variable"), independent of B. Define $\widetilde{B} = (\widetilde{B}_t)_{t\in[0,\infty)}$ by

$$\widetilde{B}_t = \begin{cases} B_t & \text{if } t \neq \tau \\ B_t + 1 & \text{if } t = \tau. \end{cases}$$

Then $\widetilde{B}$ has the same finite dimensional distributions as B (since $\mathsf{P}[\tau \in \{t_1, \dots, t_n\}] = 0$), and in particular $\widetilde{B}$ satisfies the equivalent conditions of Proposition VIII.7. However, since $\mathsf{P}[t \mapsto B_t \text{ is continuous}] = 1$, we have $\mathsf{P}[t \mapsto \widetilde{B}_t \text{ is continuous}] = 0$, so $\widetilde{B}$ is not a standard Brownian motion.

2. Probability measures on a space of continuous functions

The Brownian motion $B = (B_t)_{t\in[0,\infty)}$ can be naturally considered on the semi-infinite time interval $[0,\infty)$. To simplify some topological considerations, we will however first restrict our attention to a compact time interval $\mathbb{T} = [0,T]$, and consider the space

$$\mathcal{C}([0,T]) = \{f\colon [0,T] \to \mathbb{R} \text{ continuous}\}$$

of continuous functions defined on that time interval. This space $\mathcal{C}([0,T])$ is equipped with the metric $\varrho(f_1,f_2) = \|f_1 - f_2\|_\infty$, and it is a complete separable metric space, see Appendix H.3. We equip the space $\mathcal{C}([0,T])$ with its Borel sigma algebra $\mathscr{B}\big(\mathcal{C}([0,T])\big)$, and consider probability measures on it.

Finite dimensional distributions and Borel sigma algebra

We first address the concern raised in Remark VIII.10 by showing that finite dimensional distributions uniquely specify a Borel probability measure on $\mathcal{C}([0,T])$.

Proposition VIII.12 (Knowing finite dimensional distributions is sufficient)**.**
Suppose that ν and $\widetilde{\nu}$ are two Borel probability measures on $\mathcal{C}([0,T])$ which

coincide on events of the form

$$\Phi_{t_1,\dots,t_n}(x_1,\dots,x_n;r_1,\dots,r_n) := \Big\{ f \in \mathcal{C}([0,T]) \;\Big|\; |f(t_j)-x_j| < r_j \text{ for all } j \Big\}$$

for all $0 \le t_1 < t_2 < \cdots < t_n \le T$ *and* $x_1, x_2, \dots, x_n \in \mathbb{R}$ *and* $r_1, r_2, \dots, r_n > 0$. *Then we have* $\nu = \widetilde{\nu}$.

Proof. Note first that the sets $\Phi_{t_1,\dots,t_n}(x_1,\dots,x_n;r_1,\dots,r_n)$ are open and thus indeed Borel measurable. We can write

$$\Phi_{t_1,\dots,t_n}(x_1,\dots,x_n;r_1,\dots,r_n) = \bigcap_{j=1}^{n} \Phi_{t_j}(x_j;r_j)$$

so clearly events of this form are stable under finite intersections (they form a π-system). By Dynkin's identification theorem it therefore suffices to show that the sigma algebra generated by events of this form contains the Borel σ-algebra. Recall that the Borel sigma algebra is generated by open sets, and any open set in the separable space $\mathcal{C}([0,T])$ is a countable union of closed balls by Exercise H.7. It is therefore sufficient to show that for any $g \in \mathcal{C}([0,T])$ and any $r > 0$, the closed ball $\overline{B}_r(g)$ of radius $r > 0$ around g is in the σ-algebra generated by the above collection. But we can write, using continuity of the functions f and g involved,

$$\begin{aligned}\overline{B}_r(g) &= \Big\{ f \in \mathcal{C}([0,T]) \;\Big|\; |f(t)-g(t)| \le r \text{ for all } t \in [0,T] \Big\} \\ &= \bigcap_{\substack{s\in[0,T]\cap\mathbb{Q}\\ q\in(r,\infty)\cap\mathbb{Q}}} \Big\{ f \in \mathcal{C}([0,T]) \;\Big|\; |f(s)-g(s)| < q \Big\} = \bigcap_{\substack{s\in[0,T]\cap\mathbb{Q}\\ q\in(r,\infty)\cap\mathbb{Q}}} \Phi_s(g(s);q),\end{aligned}$$

a countable intersection of events from the collection. This finishes the proof. □

Since a standard Brownian motion $B = (B_t)_{t\in[0,\infty)}$ has continuous paths (almost surely), it determines a law ν on continuous functions as follows. For all Borel subsets $\mathcal{E} \subset \mathcal{C}([0,T])$ we would like to set

$$\nu[\mathcal{E}] = \mathsf{P}\left[\Big\{ \omega \in \Omega \;\Big|\; \big(t \mapsto B_t(\omega)\big) \in \mathcal{E} \Big\} \right]. \tag{VIII.2}$$

By the almost sure continuity, we get that $\nu\big[\mathcal{C}([0,T])\big] = 1$, so ν indeed has a chance of being a probability measure on $\mathcal{C}([0,T])$. We should just make sure that for any Borel set $\mathcal{E} \subset \mathcal{C}([0,T])$ the event $\big\{ \omega \in \Omega \;\big|\; \big(t \mapsto B_t(\omega)\big) \in \mathcal{E} \big\}$ is measurable. This follows by an argument essentially identical to the proof of Proposition VIII.12, starting from the random variables B_t for $t \in [0,T]$, and the assumed measurability of the event

$$\Big\{ \omega \in \Omega \;\Big|\; t \mapsto B_t(\omega) \text{ is continuous} \Big\}.$$

In particular, Brownian motion $B = (B_t)_{t\in[0,\infty)}$ determines via (VIII.2) a unique Borel probability measure ν on the space $\mathcal{C}([0,T])$. Donsker's theorem states that the laws of rescaled random walks $X^{(a)}$ restricted to the time interval $[0,T]$ converge to this law ν as $a \downarrow 0$.

Compactness and tightness on the space of continuous functions

The first of the two main steps of the proof of Donsker's theorem is establishing tightness of the laws of the rescaled random walks $X^{(a)}$. Recall that tightness means that apart from an arbitrarily small remainder, the probability mass of all these laws are carried by the same compact subset. Arzelà-Ascoli theorem characterizes

compact subsets in the space $\mathcal{C}([0,T])$ of continuous functions. We express the conditions of Arzelà-Ascoli theorem by the modulus of continuity.

Definition VIII.13 (Modulus of continuity)**.**
For a function $f\colon [0,T] \to \mathbb{R}$, the *modulus of continuity* is the function $\mathfrak{w}_f$ defined for $\delta > 0$ as

$$\mathfrak{w}_f(\delta) = \sup_{\substack{t,s\in[0,T]\\ |s-t|\le\delta}} |f(s) - f(t)|. \tag{VIII.3}$$

Example VIII.14 (Familiar examples of modulus of continuity)**.**
The following are familiar special cases of specific forms of modulus of continuity:

- f is Lipschitz continuous if and only if $\mathfrak{w}_f(\delta) = \mathcal{O}(\delta)$, i.e., if for some $C > 0$ we have $\mathfrak{w}_f(\delta) \le C\,\delta$.
- f is Hölder continuous of exponent α if and only if $\mathfrak{w}_f(\delta) = \mathcal{O}(\delta^\alpha)$, i.e., if for some $C > 0$ we have $\mathfrak{w}_f(\delta) \le C\,\delta^\alpha$.
- f is uniformly continuous if and only if $\mathfrak{w}_f(\delta) = o(1)$, i.e., if we have $\lim_{\delta\downarrow 0} \mathfrak{w}_f(\delta) = 0$.

A family of functions $\Phi \subset \mathcal{C}([0,T])$ is said to be *equicontinuous* if

$$\lim_{\delta\downarrow 0} \sup_{f\in\Phi} \mathfrak{w}_f(\delta) = 0.$$

This condition is just a reformulation of the property (2) in Arzelà-Ascoli theorem, Theorem H.15. For the present purposes, a convenient equivalent way of stating that theorem is the following.

Theorem VIII.15 (Arzelà-Ascoli theorem, see Theorem H.15)**.**
A subset $\Phi \subset \mathcal{C}([0,T])$ is precompact if and only if the following two conditions are satisfied:

$$\sup_{f\in\Phi} |f(0)| < \infty \tag{VIII.15-i}$$

$$\lim_{\delta\downarrow 0} \sup_{f\in\Phi} \mathfrak{w}_f(\delta) = 0. \tag{VIII.15-ii}$$

The above conditions for (pre)compactness in $\mathcal{C}([0,T])$ translate to the following conditions of tightness of probability measures on $\mathcal{C}([0,T])$.

Proposition VIII.16 (Tightness in the space of continuous functions)**.**
A sequence $(\nu_n)_{n\in\mathbb{Z}_{>0}}$ of probability measures on $\mathcal{C}([0,T])$ is tight if and only if the following conditions hold:

$$\forall \varepsilon > 0 \;\; \exists M < \infty, n_0 \in \mathbb{Z}_{>0} :$$
$$n \ge n_0 \;\Rightarrow\; \nu_n\Big[\,\big\{f \;\big|\; |f(0)| \ge M\big\}\,\Big] \le \varepsilon \tag{VIII.16-i}$$
$$\forall \varepsilon, \eta > 0 \;\; \exists \delta > 0, n_0 \in \mathbb{Z}_{>0} :$$
$$n \ge n_0 \;\Rightarrow\; \nu_n\Big[\,\big\{f \;\big|\; \mathfrak{w}_f(\delta) \ge \eta\big\}\,\Big] \le \varepsilon. \tag{VIII.16-ii}$$

Proof of "only if": Assume that $(\nu_n)_{n\in\mathbb{Z}_{>0}}$ is tight. Fix $\varepsilon > 0$. Choose a compact subset $\Phi \subset \mathcal{C}([0,T])$ so that $\nu_n[\Phi] \ge 1-\varepsilon$ for all $n \in \mathbb{Z}_{>0}$. By Theorem VIII.15, for large enough $M > 0$ we have $\Phi \subset \big\{f \;\big|\; |f(0)| < M\big\}$. This shows that (VIII.16-i) holds. Fix also $\eta > 0$. By

Theorem VIII.15, for small enough $\delta > 0$ we have $\Phi \subset \{f \mid \mathfrak{w}_f(\delta) < \eta\}$. This shows that (VIII.16-ii) holds.

Proof of "if": Assume conditions (VIII.16-i) and (VIII.16-ii). Note that in both cases we can take $n_0 = 1$ by increasing $M < \infty$ and decreasing $\delta > 0$, if necessary. Fix $\varepsilon > 0$. Choose M so that the subset $\Psi_0 := \{f \mid |f(0)| \geq M\}$ satisfies $\nu_n[\Psi_0] \leq \frac{\varepsilon}{2}$ for all n. For every $k \in \mathbb{N}$ choose $\delta_k > 0$ such that $\Psi_k := \{f \mid \mathfrak{w}_f(\delta_k) \geq \frac{1}{k}\}$ satisfies $\nu_n[\Psi_k] \leq \frac{\varepsilon}{2^{k+1}}$ for all n. Then the intersection of the complements $\Phi = \Psi_0^c \cap \big(\bigcap_{k\in\mathbb{N}} \Psi_k^c\big)$ satisfies the conditions of Theorem VIII.15, so its closure $\overline{\Phi}$ is compact, and by the union bound and the choices above we see that for all n

$$\nu_n\big[\,\overline{\Phi}\,\big] \geq \nu_n\big[\Phi\big] = 1 - \nu_n\Big[\Psi_0 \cup \big(\bigcup_{k\in\mathbb{N}} \Psi_k\big)\Big] \geq 1 - \frac{\varepsilon}{2} - \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+1}} = 1 - \varepsilon.$$

This shows that $(\nu_n)_{n\in\mathbb{Z}_{>0}}$ is tight. □

3. Proof of Donsker's theorem

Recall that our goal is to show the convergence of the piecewise linearly interpolated random walks

$$X_t^{(a)} = \sqrt{a}\left(\sum_{\ell=1}^{\lfloor t/a\rfloor} \xi_\ell + \Big(\frac{t}{a} - \Big\lfloor \frac{t}{a}\Big\rfloor\Big)\xi_{\lfloor t/a\rfloor+1}\right). \tag{VIII.4}$$

with steps $(\xi_\ell)_{\ell\in\mathbb{N}}$ and scale parameter $a \downarrow 0$. The two parts of the proof are tightness and identification of subsequential limits.

Tightness of scaled random walks

We now present lemmas which help us verify tightness of the laws of $X^{(a)}$ in $\mathcal{C}([0,T])$ by the conditions of Proposition VIII.16.

We will obtain bounds for the modulus of continuity ultimately by the following.

Lemma VIII.17 (A bound for the modulus of continuity)**.** *Suppose that $0 = t_0 < t_1 < t_2 < \cdots < t_k = T$ are such that $\min_{1\leq j\leq k}(t_j - t_{j-1}) \geq \delta$. Then for any $f \in \mathcal{C}([0,T])$ we have*

$$\mathfrak{w}_f(\delta) \leq 3 \max_{1\leq j\leq k} \sup_{s\in[t_{j-1},t_j]} |f(s) - f(t_{j-1})|.$$

Proof. Denote, for brevity, the maximum on the right hand side by

$$M := \max_{1\leq j\leq k} \sup_{s\in[t_{j-1},t_j]} |f(s) - f(t_{j-1})|.$$

Suppose that $s,t \in [0,T]$ are such that $|s-t| \leq \delta$, so they need to be considered in the definition (VIII.3) of $\mathfrak{w}_f(\delta)$. Assume without loss of generality that $s < t$. There are two different cases: either $s,t \in [t_{j-1},t_j]$ for some j, or $s \in [t_{j-1},t_j]$ and $t \in [t_j,t_{j+1}]$ for some j.

If $s,t \in [t_{j-1},t_j]$, then by triangle inequality

$$|f(s) - f(t)| \leq |f(s) - f(t_{j-1})| + |f(t_{j-1}) - f(t)| \leq 2M.$$

If $s \in [t_{j-1},t_j]$ and $t \in [t_j,t_{j+1}]$, then by triangle inequalities

$$|f(s) - f(t)| \leq |f(s) - f(t_{j-1})| + |f(t_{j-1}) - f(t_j)| + |f(t_j) - f(t)| \leq 3M.$$

In either case $|f(s)-f(t)| \le 3M$ and thus $\mathfrak{w}_f(\delta) \le 3M$. □

To verify the important condition (VIII.16-ii) for the random walks (VIII.4), we use two more lemmas, formulated below with slightly different assumptions about the sequence $(\xi_\ell)_{\ell\in\mathbb{N}}$ of steps.

Lemma VIII.18 (A sufficient condition for tightness)**.**
Let $(\xi_\ell)_{\ell\in\mathbb{N}}$ be a sequence of random variables that is stationary in the sense that for any $\ell' \in \mathbb{N}$ and $L \in \mathbb{N}$, $(\xi_{L+1}, \dots, \xi_{L+\ell'})$ has the same law as $(\xi_1, \dots, \xi_{\ell'})$. Let also $(a_n)_{n\in\mathbb{N}}$ be a sequence of scales $a_n > 0$ tending to zero, $a_n \downarrow 0$. Suppose that

$$\lambda^2 \limsup_{h\to\infty} \mathsf{P}\Big[\max_{1\le\ell'\le h} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \ge \lambda\sqrt{h}\Big] \underset{\lambda\to+\infty}{\longrightarrow} 0. \tag{VIII.5}$$

Then the laws of $X^{(a_n)}$, for $n \in \mathbb{N}$, as defined by (VIII.4) *for $t \in [0,T]$, form a tight family in $\mathcal{C}([0,T])$.*

Proof. Condition (VIII.16-i) holds trivially for the laws of $\big(X^{(a_n)}\big)_{n\in\mathbb{N}}$ given by (VIII.4) since $X_0^{(a_n)} = 0$ for all n. We only need to verify condition (VIII.16-ii), i.e. that for all $\eta > 0$

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge \eta\Big] = 0.$$

Therefore we fix $\eta > 0$ once and for all. For now fix also a small $\delta > 0$ (we will in particular assume $\delta \le T$). For any time interval division points $0 = t_0 < t_1 < \cdots < t_k \ge T$ such that $t_j - t_{j-1} \ge \delta$ for all j, we could use Lemma VIII.17 and the union bound to estimate

$$\begin{aligned}\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge \eta\Big] &\le \mathsf{P}\Big[\sup_{s\in[t_{j-1},t_j]} |X_s^{(a_n)} - X_{t_{j-1}}^{(a_n)}| \ge \frac{\eta}{3} \text{ for some } j \in \{1,\dots,k\}\Big] \\ &\le \sum_{j=1}^{k} \mathsf{P}\Big[\sup_{s\in[t_{j-1},t_j]} |X_s^{(a_n)} - X_{t_{j-1}}^{(a_n)}| \ge \frac{\eta}{3}\Big].\end{aligned}$$

Now for the given δ and $n \in \mathbb{N}$, we define the integers $h = h_n = \lceil \delta/a_n \rceil$ and $k = \lceil T/\delta \rceil$, and choose time division points $t_j = j h_n a_n$ for $j = 0, 1, \dots, k$. Note that $h_n a_n \to \delta$ as $n \to \infty$, so for large enough n we have $a_n \le \frac{2\delta}{h_n}$.

The function $s \mapsto |X_s^{(a_n)} - X_{t_{j-1}}^{(a_n)}|$ on $s \in [t_{j-1}, t_j]$ attains its maximum at some point $s \in a_n\mathbb{Z} \cap [t_{j-1}, t_j]$. The estimate based on Lemma VIII.17 then reads

$$\begin{aligned}\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge \eta\Big] &\le \sum_{j=1}^{k} \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ (j-1)h_n<\ell'\le jh_n}} \Big|\sqrt{a_n} \sum_{\ell=(j-1)h_n+1}^{\ell'} \xi_\ell\Big| \ge \frac{\eta}{3}\Bigg] \\ &= k\, \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ 0<\ell'\le h_n}} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \ge \frac{\eta}{3\sqrt{a_n}}\Bigg]\end{aligned}$$

where the last equality uses the assumed stationarity of $(\xi_\ell)_{\ell\in\mathbb{N}}$. Recall that for large enough n we have $a_n \le \frac{2\delta}{h_n}$ and also $k \le \frac{2T}{\delta}$ (recall that $\delta \le T$). We can thus further estimate the last expression to get the upper bound

$$\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge \eta\Big] \le \frac{2T}{\delta}\, \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ 0<\ell'\le h_n}} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \ge \frac{\eta\sqrt{h_n}}{3\sqrt{2\delta}}\Bigg].$$

Now denote $\lambda = \lambda(\delta) = \frac{\eta}{3\sqrt{2\,\delta}}$, so that $\delta \downarrow 0$ corresponds to $\lambda \uparrow \infty$. Then rewrite the estimate as

$$\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \geq \eta\Big] \leq \frac{36T}{\eta^2}\,\lambda^2\,\mathsf{P}\bigg[\max_{\substack{\ell' \in \mathbb{Z} \\ 0<\ell' \leq h_n}} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \geq \lambda\sqrt{h_n}\bigg].$$

As $n \to \infty$, also $h_n \to \infty$, so we have

$$\limsup_{n\to\infty} \mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \geq \eta\Big] \leq \frac{36T}{\eta^2}\,\lambda^2\,\limsup_{h\to\infty} \mathsf{P}\bigg[\max_{\substack{\ell' \in \mathbb{Z} \\ 0<\ell' \leq h}} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \geq \lambda\sqrt{h}\bigg].$$

Thus by the assumption (VIII.5) we get the desired validity of the condition (VIII.16-ii):

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \geq \eta\Big] \;=\; 0.$$

This finishes the proof. □

We also use the following Etemadi's inequality, very similar to Lévy's inequality (Lemma III.7) in Lecture III.

Lemma VIII.19 (Etemadi's inequality)**.**
Suppose that $(\xi_\ell)_{\ell\in\mathbb{N}}$ are independent. Denote $S_k = \sum_{\ell=1}^k \xi_\ell$. Then for all $h \in \mathbb{Z}_{>0}$ and $\Lambda > 0$ we have

$$\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq 3\Lambda\Big] \leq 3 \max_{1\leq k\leq h} \mathsf{P}\Big[\big|S_k\big| \geq \Lambda\Big].$$

Proof. Denote

$$p = \max_{1\leq k\leq h} \mathsf{P}\Big[\big|S_k\big| \geq \Lambda\Big]$$

for brevity. For $k = 1, 2, \ldots, h$, define the event

$$A_k = \big\{\big|S_k\big| \geq 3\Lambda \;\text{ and }\; \big|S_m\big| < 3\Lambda \text{ for } m = 1, 2, \ldots, k-1\big\}$$

and also the event

$$B = \big\{\big|S_h\big| \geq \Lambda\big\}.$$

The events $(A_k)_{k=1,\ldots,h}$ are disjoint. We can now estimate

$$\begin{aligned}\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq 3\Lambda\Big] \;&=\; \mathsf{P}\Big[\bigcup_{k=1}^h A_k\Big] = \mathsf{P}\big[B\big] + \sum_{k=1}^h \mathsf{P}\big[A_k \cap B^c\big] \\ &\leq p + \sum_{k=1}^h \mathsf{P}\Big[A_k \cap \big\{|S_h - S_k| > 2\Lambda\big\}\Big].\end{aligned}$$

Now A_k is independent of $S_h - S_k$, so we have

$$\begin{aligned}\mathsf{P}\Big[A_k \cap \big\{|S_h - S_k| > 2\Lambda\big\}\Big] \;&= \mathsf{P}\big[A_k\big]\;\mathsf{P}\big[|S_h - S_k| > 2\Lambda\big] \\ &\leq \mathsf{P}\big[A_k\big]\Big(\mathsf{P}\big[|S_k| > \Lambda\big] + \mathsf{P}\big[|S_h| > \Lambda\big]\Big) \;\leq\; 2p\,\mathsf{P}\big[A_k\big].\end{aligned}$$

Substituting this to the eariler estimate gives

$$\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq 3\Lambda\Big] \leq p + 2p\sum_{k=1}^h \mathsf{P}\big[A_k\big] \;\leq\; 3p,$$

which concludes the proof. □

With the auxiliary results above we can show that the laws of the scaled random walks $X^{(a)}$ are tight on $\mathcal{C}([0,T])$.

Proposition VIII.20 (Tightness of rescaled simple random walks)**.**
Suppose that $(a_n)_{n\in\mathbb{N}}$ is a sequence of scales tending to zero, $a_n \downarrow 0$, and consider the scaled random walks $X^{(a_n)}$ with i.i.d. steps $(\xi_\ell)_{\ell\in\mathbb{N}}$ such that $\mathsf{P}\big[\xi_\ell = +1\big] = \frac{1}{2} = \mathsf{P}\big[\xi_\ell = -1\big]$. Then the laws of $X^{(a_n)}$ form a tight family of probability measures on $\mathcal{C}([0,T])$.

Proof. The sequence $(\xi_\ell)_{\ell\in\mathbb{N}}$ of steps is i.i.d., and thus in particular stationary. We can therefore verify tightness using Lemma VIII.18.

Denote again $S_k = \sum_{\ell=1}^k \xi_\ell$. Note that since $\mathsf{E}[\xi_\ell] = 0$ and (ξ_ℓ) are i.i.d., we have

$$\begin{aligned}\mathsf{E}\big[S_k^4\big] &= \sum_{\ell=1}^k \mathsf{E}\big[\xi_\ell^4\big] + 6 \sum_{1\le \ell < m \le k} \mathsf{E}\big[\xi_\ell^2\big]\, \mathsf{E}\big[\xi_m^2\big] \\ &= k\, \mathsf{E}\big[\xi_1^4\big] + 3k(k-1)\, \mathsf{E}\big[\xi_1^2\big]^2.\end{aligned}$$

Noting that $|\xi_1| = 1$, the expected values are trivial to compute, and we can in particular deduce the upper bound $\mathsf{E}\big[S_k^4\big] \le 3k^2$.

Then set $\Lambda = \frac{1}{3}\lambda\sqrt{h}$ in Lemma VIII.19 to get

$$\mathsf{P}\Big[\max_{1\le k\le h} \big|S_k\big| \ge \lambda\sqrt{h}\Big] \le 3 \max_{1\le k\le h} \mathsf{P}\Big[\big|S_k\big| \ge \frac{1}{3}\lambda\sqrt{h}\Big]$$

Markov's inequality with the fourth moment of S_k gives

$$\mathsf{P}\Big[\big|S_k\big| \ge \frac{1}{3}\lambda\sqrt{h}\Big] = \mathsf{P}\Big[\frac{S_k^4}{(\frac{1}{3}\lambda\sqrt{h})^4} \ge 1\Big] \le \frac{\mathsf{E}[S_k^4]}{(\frac{1}{3}\lambda\sqrt{h})^4} \le \frac{3^5\, k^2}{\lambda^4\, h^2}.$$

Combining with the above, we get

$$\mathsf{P}\Big[\max_{1\le k\le h} \big|S_k\big| \ge \lambda\sqrt{h}\Big] \le 3 \max_{1\le k\le h} \frac{3^5\, k^2}{\lambda^4\, h^2} \;=\; 3^6\, \lambda^{-4}.$$

From here we deduce the estimate

$$\lambda^2 \limsup_{h\to\infty} \mathsf{P}\Big[\max_{1\le k\le h} \big|S_k\big| \ge \lambda\sqrt{h}\Big] \le 3^6\, \lambda^{-2} \underset{\lambda\to\infty}{\longrightarrow} 0,$$

so Lemma VIII.18 implies that the laws of $X^{(a_n)}$ are tight on $\mathcal{C}([0,T])$. □

Brownian motion as the limit of scaled random walks

Having established tightness of the laws of the scaled random walks $(X^{(a)})$, the remaining task in the proof of Donsker's theorem is to identify any subsequential limit of these as the standard Brownian motion. With the definition of Brownian motion based on Gaussianity and independent increments, this is a relatively straightforward consequence of the central limit theorem. We still state precisely Donsker's theorem, and then proceed directly to its proof.

Theorem VIII.21 (Donsker's theorem)**.**
For $a > 0$, let $X^{(a)}$ be the scaled random walk given by (VIII.4) *with i.i.d. steps $(\xi_\ell)_{\ell\in\mathbb{N}}$ such that $\mathsf{P}\big[\xi_\ell = +1\big] = \frac{1}{2} = \mathsf{P}\big[\xi_\ell = -1\big]$. Then for any $T > 0$, as $a \downarrow 0$, the laws of the walks $X^{(a)}|_{[0,T]}$ restricted to the time interval $[0,T]$ converge weakly on $\mathcal{C}([0,T])$ to the law of the standard Brownian motion $B|_{[0,T]}$ restricted to the same time interval.*

Proof. By Proposition VIII.20 and Prohorov's theorem, from any sequence $a_n \downarrow 0$ of scales, we may extract some subsequence $(a_{n_k})_{k\in\mathbb{N}}$ so that the laws of $X^{(a_{n_k})}$ converge weakly to a limit. Let X^* denote a random process with such a limit law. We must only show that X^* is a

standard Brownian motion, the convergence then follows by the usual argument combining precompactness and uniqueness of subsequential limits (Exercise H.12). To avoid cumbersome notation, denote the convergent subsequence again by $(a_n)_{n\in\mathbb{N}}$, to that $X^{(a_n)} \xrightarrow{\text{w}} X^*$ as $n \to \infty$.

To show that the subsequential limit X^* is the standard Brownian motion, we will calculate its finite dimensional distributions — by Proposition VIII.12 these characterize the law uniquely. So fix $0 = t_0 < t_1 < t_2 < \cdots < t_m \le T$ and consider the limit as $n \to \infty$ of laws of $(X^{(a_n)}_{t_1}, \dots, X^{(a_n)}_{t_m})$. Note that the piecewise linear interpolation (VIII.4) is of no significance here: if we would consider instead the piecewise constant random walk $\bar{X}^{(a_n)}_t = \sqrt{a_n}\sum_{\ell=1}^{\lfloor t/a_n\rfloor}\xi_\ell$, then $|X^{(a_n)}_{t_j} - \bar{X}^{(a_n)}_{t_j}| \le \sqrt{a_n} \to 0$ so the finite dimensional distributions

$$(X^{(a_n)}_{t_1}, \dots, X^{(a_n)}_{t_m}) \qquad \text{and} \qquad (\bar{X}^{(a_n)}_{t_1}, \dots, \bar{X}^{(a_n)}_{t_m})$$

tend to the same limit.

Now we can write the increments of $\bar{X}^{(a_n)}$ as

$$\bar{X}^{(a_n)}_{t_j} - \bar{X}^{(a_n)}_{t_{j-1}} = \sqrt{a_n} \sum_{\ell \,\in\, \mathbb{Z}\cap\left(\frac{t_{j-1}}{a_n}, \frac{t_j}{a_n}\right]} \xi_\ell.$$

By the central limit theorem, this expression tends weakly to a centered Gaussian of variance $t_j - t_{j-1}$, i.e., the law $\mathrm{N}(0, t_j - t_{j-1})$. Moreover, the increments

$$(\bar{X}^{(a_n)}_{t_1} - \bar{X}^{(a_n)}_{t_0}, \dots, \bar{X}^{(a_n)}_{t_m} - \bar{X}^{(a_n)}_{t_{m-1}})$$

are independent (no two of them involve the same step ξ_ℓ, and the steps are independent), so the joint law of the increments converges weakly to m independent centered Gaussians with variances $t_1 - t_0, \dots, t_m - t_{m-1}$. This is the joint law of the increments of the standard Brownian motion, and obviously the joint law of the increments specifies the finite dimensional distribution. We conclude that any subsequential limit X^* has the finite dimensional distributions of a standard Brownian motion. This finishes the proof of Donsker's theorem. □

Exercise VIII.1 (Shifting, scaling, and inverting Brownian motion)**.**
Suppose that $B = (B_t)_{t\ge0}$ is a standard Brownian motion and define three other stochastic processes $W^{(k)} = (W^{(k)}_t)_{t\ge0}$, $k = 1, 2, 3$, by setting

$$W^{(1)}_t = B_{s+t} - B_s$$
$$W^{(2)}_t = \lambda^{-1/2} B_{\lambda t}$$
$$W^{(3)}_t = \begin{cases} tB_{1/t} & , \text{ when } t > 0 \\ 0 & , \text{ when } t = 0 \end{cases}$$

where $s \ge 0$ and $\lambda > 0$ are constants. Show that all these three stochastic processes $W^{(k)} = (W^{(k)}_t)_{t\ge0}$, $k = 1, 2, 3$, are also standard Brownian motions.

Exercise VIII.2 (From Brownian motion to Brownian bridge)**.**
Suppose that $B = (B_t)_{t\ge0}$ is a standard Brownian motion. Define a stochastic process $X = (X_t)_{t\in[0,1]}$ by setting $X_t = B_t - tB_1$.

(a) Show that X is a Gaussian process. Calculate the mean function and the covariance function of X, i.e., $t \mapsto \mathsf{E}[X_t]$ and $(s,t) \mapsto \mathsf{Cov}[X_s, X_t] = \mathsf{E}[X_sX_t] - \mathsf{E}[X_s]\,\mathsf{E}[X_t]$.

(b) Show that X and B_1 are independent.

Hint: *The distribution of X is determined by the finite dimensional marginals $\mathsf{P}[X_{t_1} \in A_1, \dots, X_{t_n} \in A_n]$, for $0 \le t_1 < \cdots < t_n \le 1$ and $A_1, \dots, A_n \subset \mathbb{R}$ Borel. Consider also using the Gaussianity of the processes (recall: Gaussians are independent iff they are uncorrelated).*

Exercise VIII.3 (From Brownian bridge to Brownian motion)**.**
Suppose that $X = (X_t)_{t\in[0,1]}$ is a continuous and Gaussian process, for which $\mathsf{E}[X_t] = 0$ for all $t \in [0,1]$ and $\mathsf{Cov}[X_s, X_t] = s(1-t)$ for all $0 \le s \le t \le 1$. Let $Y \sim N(0,1)$ be a

random variable independent of X. Define a stochastic process $W = (W_t)_{t\in[0,1]}$ by setting $W_t = X_t + tY$.

(a) Show that W has the same finite dimensional distributions as a standard Brownian motion on the time interval $[0, 1]$.

The event $\{W_1 = y\}$ has zero probability, but it is natural to define conditioning on this event by the following limiting procedure. For any $\varepsilon > 0$, the event $\{|W_1 - y| < \varepsilon\}$ has positive probability, so we can consider the conditional distribution of the process W given the event $\{|W_1 - y| < \varepsilon\}$, and then take the (weak) limit of this conditional distribution as $\varepsilon \searrow 0$.

(b) Show that the conditional distribution of the process W given $\{W_1 = y\}$ is Gaussian, and calculate their mean and covariance functions.
Note: *Interpret the distribution of the process as the collection of its finite dimensional distributions, and conditioning as defined by the limiting procedure above.*

Exercise VIII.4 (Law of iterated logarithm for Brownian motion)**.**
Let $B = (B_t)_{t\in[0,\infty)}$ be a standard Brownian motion. For $t \geq e$ define

$$\lambda(t) = \sqrt{t \log(\log(t))}.$$

(a) Show that almost surely $\limsup_{n\to\infty} \frac{|B_n|}{\lambda(n)} = \sqrt{2}$, where the lim sup is taken along $n \in \mathbb{N}$.
Hint: *Observe that the restriction to Brownian motion to integer times, $(B_n)_{n\in\mathbb{N}}$, is a random walk with Gaussian steps. Recall the law of iterated logarithm proven earlier.*

(b) Find a function $\tilde{\lambda}\colon \mathbb{N} \to [0, \infty)$ with the following properties: $\frac{\tilde{\lambda}(n)}{\lambda(n)} \to 0$ as $n \to \infty$, and almost surely $\max_{s\in[n,n+1)} |B_s - B_n| \leq \tilde{\lambda}(n)$ except for finitely many values of $n \in \mathbb{N}$.
Hint: *You may use the fact that* $\mathsf{P}\big[\max_{s\in[0,1]} B_s > r\big] = \frac{2}{\sqrt{2\pi}} \int_r^\infty e^{-\frac{1}{2}v^2}\, \mathrm{d}v$ *for* $r \geq 0$.

(c) Show that almost surely

$$\limsup_{t\to\infty} \frac{|B_t|}{\lambda(t)} = \sqrt{2},$$

where the lim sup is taken along $t \in [0, \infty)$.
Hint: *Use (a) and (b).*

(d) Show that almost surely

$$\limsup_{t\to 0} \frac{|B_t|}{\sqrt{t \log\big(|\log(1/t)|\big)}} = \sqrt{2}.$$

Hint: *Use (c) and the last part of Exercise VIII.1.*

Lecture IX

Ising model: correlation inequalities

We will now introduce and study one of the most fundamental models of statistical physics — the Ising model. The model can be defined directly on any finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$. Its degrees of freedom are "spins" σ_x located at the sites $x \in \mathrm{V}$ of the graph, taking two possible values: up or down ($\sigma_x = +1$ or $\sigma_x = -1$). The probability measure is a Boltzmann measure associated to an energy function which essentially counts the number of neighboring vertices on the graph whose spins do not agree with each other. Therefore the model favors alignment of spins by assigning a greater probability to configurations with more alignment of neighbors. The strength of this preference to alignment is determined by a parameter β, the inverse temperature of the Boltzmann distribution.

We will mostly be interested in the Ising model model on finite subgraphs of the d-dimensional lattice $\mathbb{Z}^d$, and in the next lecture (Lecture X) we will consider how the model can be defined in "thermodynamical limit", on the infinite lattice $\mathbb{Z}^d$ via a weak limit.

The definition of the Ising model closely resembles that of the Curie–Weiss model studied in Lecture V, and it is also a model of a ferromagnetic material. In particular, in dimensions $d \geq 2$, the thermodynamical limit of the Ising model can be shown to have a phase transition between paramagnetic and ferromagnetic phases, qualitatively similar to the Curie-Weiss model. The crucial refinement is that the Ising model incorporates spacial structure also, via the graph $\mathcal{G}$ on which the model is defined. This refinement is enough to make the quantitative critical behavior of the Ising model correctly dependent on the dimension d of the lattice $\mathbb{Z}^d$, in the

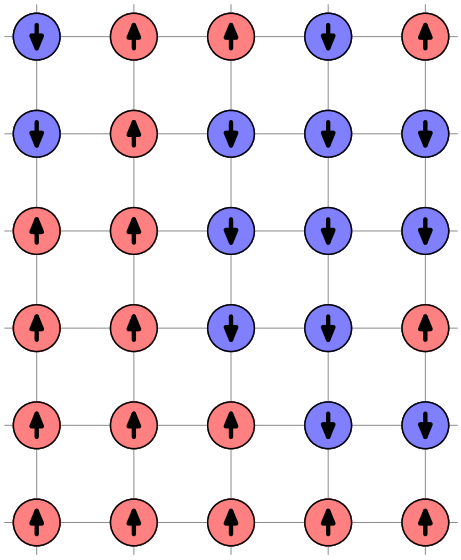


FIGURE IX.1. The Ising model is a random assignment of ± 1-valued spins σ_x to the vertices x of a (finite) graph.

sense that the critical exponents (in $d = 3$ and $d = 2$) match with those observed in uniaxial ferromagnetic materials. Note that the Ising model is essentially the simplest possible model for the preference for local alignment, and as such it is also applicable to various other phenomena besides ferromagnetism.

The goal of the present lecture is to define the Ising model on finite graphs, and to establish general correlation inequalities for the model. In the next lecture we will use monotonicity properties based on the correlation inequalities to establish the existence of thermodynamical limits.

1. Ising model on finite graphs

Definition of the Ising model

Let $\mathcal{G} = (\mathrm{V}, \mathrm{E})$ be a finite graph: V is a finite set of *sites* and E a collection of unordered pairs of sites called the *bonds* of the graph. Let $\beta > 0$ and $B \in \mathbb{R}$ be parameters, interpreted respectively as the *inverse temperature* and the *external magnetic field*, like in the Curie-Weiss model in Lecture V.

Each site $x \in \mathrm{V}$ has a spin $\sigma_x \in \{-1, +1\}$ (representing an elementary magnetic unit at the location x) and the configuration of all spins $\sigma = (\sigma_x)_{x\in\mathrm{V}}$ is called the spin configuration. The sample space of the Ising model on the graph $\mathcal{G}$ is the set

$$\Omega = \Omega_{\mathcal{G}} = \{-1, +1\}^{\mathrm{V}}$$

of all possible spin configurations. The energy of a configuration $\sigma \in \{-1, +1\}^{\mathrm{V}}$ is defined to be

$$H_{\mathcal{G}}(\sigma) = -\sum_{\{x,y\}\in\mathrm{E}} \sigma_x\sigma_y - B\sum_{x\in\mathrm{V}} \sigma_x. \tag{IX.1}$$

The Boltzmann distribution (with parameter $\beta > 0$) is the probability measure $\mathsf{P} = \mathsf{P}^{(\mathcal{G})}_{\beta,B}$ on the finite sample space $\Omega_{\mathcal{G}}$, such that the individual outcomes have probabilities

$$\mathsf{P}^{(\mathcal{G})}_{\beta,B}\left[\{\sigma\}\right] = \frac{1}{Z_{\mathcal{G}}(\beta, B)} e^{-\beta H_{\mathcal{G}}(\sigma)} \tag{IX.2}$$

and the partition function $Z_{\mathcal{G}}(\beta, B)$ normalizes $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ to be a probability measure

$$Z_{\mathcal{G}}(\beta, B) = \sum_{\sigma\in\Omega_{\mathcal{G}}} e^{-\beta H_{\mathcal{G}}(\sigma)}. \tag{IX.3}$$

The partition function seen as a function of the parameters of the model, however, is more than just a constant that normalizes the total probability to one. The following exercise illustrates the general principle that the partition function in fact contains information about interesting expected values in the model.

Exercise IX.1 (Average magnetization in the Ising model)**.**
Consider the Ising model on a finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$, and let $\#\mathrm{V}$ denote the number of its sites. Let $\mathsf{E}^{(\mathcal{G})}_{\beta,B}$ denote the expected value with respect to the probability measure $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$. Define the average magnetization as

$$M := \mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\frac{1}{\#\mathrm{V}}\sum_{x\in\mathrm{V}} \sigma_x\Big].$$

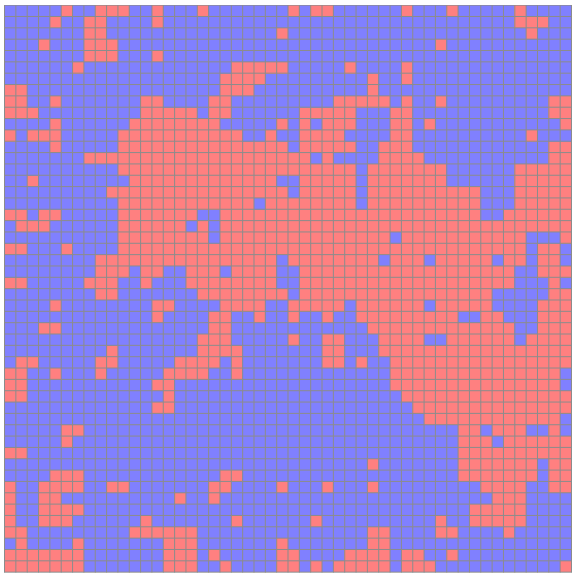


FIGURE IX.2. Monte-Carlo Markov chain methods are the standard technique for producing sample spin configurations of the Ising model. Sampling from the Ising probability measure becomes non-trivial when the graph is even moderately large.

Show that

$$M = \frac{1}{\beta\,\#\mathrm{V}}\,\frac{\partial}{\partial B}\Big(\log Z_{\mathcal{G}}(\beta, B)\Big),$$

where the partition function $Z_{\mathcal{G}}(\beta, B)$ is viewed as a function of β and B, as in (IX.3).

Monte Carlo Markov chain sampling using Glauber dynamics

The following exercise describes the Glauber dynamics for Ising model. The Glauber dynamics is a stochastic process on the Ising model state space, which can be thought of as modeling the thermal motion that mixes the configuration in the material. The Ising model probability measure is stationary for the process, and from any initial configuration the state of the process approaches this stationary Ising model distribution as time increases.

The Glauber dynamics is also commonly used for sampling from the Ising model measure. To appreciate that generating a sample $\sigma = (\sigma_x)_{x\in\mathrm{V}}$ with the desired distribution $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ may be a nontrivial task, note that if the graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$ is relatively large, then the state space $\Omega_{\mathcal{G}} = \{-1, +1\}^{\mathrm{V}}$ is enormous and it is in practice completely infeasible to even list all the possible configurations — or to compute the partition function (IX.3) which is needed in the formula (IX.2) for their probabilities. For example, in a very modest simulation we might consider the graph which consists of a three dimensional $10 \times 10 \times 10$ box: the number of sites is $\#\mathrm{V} = 1000$, and no computer will carry out the sum over the $\#\Omega = 2^{\#\mathrm{V}} = 2^{1000} \approx 1.07 \times 10^{301}$ terms in the partition function. In a realistic small physical magnet there could be of the order of magnitude 10^{23} elementary magnets (molecules or atoms): the appropriate graph has $\#\mathrm{V} \approx 10^{23}$ sites and there would be $\#\Omega = 2^{\#\mathrm{V}} \approx 2^{10^{23}} \approx 10^{30102999566398110000000}$ possible spin configurations.

Exercise IX.2 (Glauber dynamics for Ising model)**.**

Consider the Ising model on a finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$. For $\tau \in \{-1,+1\}^{\mathrm{V}}$, $x \in \mathrm{V}$, and $\epsilon \in \{-1,+1\}$, denote

$$c_\epsilon^{(x)}(\tau) := \Big(1 + \exp\big(-2\epsilon\beta B - 2\epsilon\beta \sum_{y:\{x,y\}\in \mathrm{E}} \tau_y\big)\Big)^{-1}.$$

(a) Let $x \in \mathrm{V}$ and $\tau \in \{-1,+1\}^{\mathrm{V}}$. Show that the conditional distribution of σ_x, given that the configuration σ coincides with τ outside x, is given by

$$\mathsf{P}_{\beta,B}^{(\mathcal{G})}\Big[\sigma_x = \epsilon \;\Big|\; \sigma_y = \tau_y \;\forall y \neq x\Big] = c_\epsilon^{(x)}(\tau).$$

(b) Let $X = (X_t)_{t\in[0,\infty)}$ be a continuous time Markov process on the state space $\{-1,+1\}^{\mathrm{V}}$ with jump rates

$$\lambda(\sigma,\tau) = \begin{cases} c_{\tau_x}^{(x)}(\sigma) & \text{if } \exists x \in \mathrm{V} \text{ s.t. } \tau_x \neq \sigma_x \text{ and } \tau_y = \sigma_y \;\forall y \neq x \\ 0 & \text{if } \#\big\{x \in \mathrm{V} \;\big|\; \sigma_x \neq \tau_x\big\} \neq 1 \end{cases}$$

Show that the Ising model probability measure $\mathsf{P}_{\beta,B}^{(\mathcal{G})}$ is the unique stationary measure of the process X.

It is relatively straightforward to sample the Markov process $X = (X_t)_{t\in[0,\infty)}$ (or even easier to sample some discrete time variant of it) starting from some chosen initial state. In the long run the Markov chain approaches its stationary measure, which by the above exercise is the Ising probability measure $\mathsf{P}_{\beta,B}^{(\mathcal{G})}$. We can thus produce an approximately correctly distributed Ising sample by taking the random state X_t of the process for some very large time t.

2. Griffiths' correlation inequality

Besides the spin σ_x at a single site $x \in \mathrm{V}$, among the simplest random variables in Ising model is a product

$$\prod_{x\in A} \sigma_x$$

of spins in some finite subset $A \subset \mathrm{V}$ of sites. For example, for a bond $e = \{x,y\}$ the product $\sigma_x\sigma_y$ represents, up to sign, the contribution of that bond to the total energy (IX.1). With this in mind, the energy density at the bond $e = \{x,y\}$ is defined as the random variable

$$\epsilon_e = -\sigma_x\sigma_y.$$

We will establish a correlation inequality known as Griffiths' inequality, which concerns the behavior of random variables of the above product form $\prod_{x\in A}\sigma_x$. Griffiths' inequality has two statements, both of which should appear quite intuitive in ferromagnets. The first statement says that the product $\prod_{x\in A}\sigma_x$ of spins can not on average align against the external magnetic field. The second statement says that any products of spins are positively correlated with each other. The more precise statements are given below.

Proposition IX.1 (Griffiths' correlation inequality for Ising model)**.**

Consider the Ising model on a finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$, with non-negative external magnetic field $B \geq 0$.

(i) *For any subset* $A \subset \mathrm{V}$ *we have*

$$\mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{x\in A}\sigma_x\Big] \;\geq\; 0. \tag{IX.4}$$

(ii) *For any subsets* $A_1, A_2 \subset \mathrm{V}$ *we have*

$$\mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{x\in A_1}\sigma_x \prod_{x\in A_2}\sigma_x\Big] - \mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{x\in A_1}\sigma_x\Big]\,\mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{x\in A_2}\sigma_x\Big] \;\geq\; 0. \tag{IX.5}$$

A similar statement holds more generally than just for the Ising model — we only need ferromagnetic interactions that make alignment of spins energetically favorable. We prove the statement in such generalized form, which will also be useful later.

Proposition IX.2 (General Griffiths' correlation inequality)**.**
Let V *be a non-empty finite set and* P *a probability measure on* $\{-1,+1\}^V$ *given by*

$$\mathsf{P}\big[\,\{\sigma\}\,\big] = e^{-h(\sigma)}\Big/\sum_{\tau\in\{-1,+1\}^V} e^{-h(\tau)},$$

where

$$h(\sigma) = -\sum_{C\subset V} J_C \prod_{z\in C}\sigma_z \qquad \text{with } J_C \geq 0 \text{ for every } C \subset V.$$

(i) *For any subset* $A \subset V$ *we have*

$$\mathsf{E}\Big[\prod_{x\in A}\sigma_x\Big] \;\geq\; 0.$$

(ii) *For any subsets* $A_1, A_2 \subset V$ *we have*

$$\mathsf{E}\Big[\prod_{x\in A_1}\sigma_x \prod_{x\in A_2}\sigma_x\Big] - \mathsf{E}\Big[\prod_{x\in A_1}\sigma_x\Big]\,\mathsf{E}\Big[\prod_{x\in A_2}\sigma_x\Big] \;\geq\; 0.$$

Remark IX.3 (Specializing the general Griffiths' inequality to Ising model)**.**
If we take V above to be the set V of sites our finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$, and we choose the coupling constants J_C so that $J_{\{x\}} = \beta B$ for every $x \in \mathrm{V}$ and $J_{\{x,y\}} = \beta$ for every $\{x, y\} \in \mathrm{E}$ and for all other subsets $C \subset \mathrm{V}$ we let $J_C = 0$, then the model above is exactly the Ising model on $\mathcal{G} = (\mathrm{V}, \mathrm{E})$. Proposition IX.1 is thus a special case of Proposition IX.2.

Proof of part (i) of Proposition IX.2: Denote

$$Z = \sum_{\tau\in\{-1,+1\}^V} e^{-h(\tau)},$$

so that $\mathsf{P}\big[\,\{\sigma\}\,\big] = \frac{1}{Z}e^{-h(\sigma)}$ and

$$\mathsf{E}\Big[\prod_{x\in A}\sigma_x\Big] = \frac{1}{Z}\sum_{\sigma\in\{-1,+1\}^V}\Big(\prod_{x\in A}\sigma_x\Big)e^{-h(\sigma)}.$$

We clearly have $Z > 0$, so it suffices to show the non-negativity of the numerator, which we can expand as

$$\begin{aligned}
&\sum_{\sigma\in\{-1,+1\}^V} \Big(\prod_{x\in A}\sigma_x\Big) e^{-h(\sigma)} \\
&= \sum_{\sigma\in\{-1,+1\}^V} \Big(\prod_{x\in A}\sigma_x\Big)\, \exp\Big(\sum_{C\subset V} J_C \prod_{z\in C}\sigma_z\Big) \\
&= \sum_{\sigma\in\{-1,+1\}^V} \Big(\prod_{x\in A}\sigma_x\Big) \sum_{n=0}^{\infty}\frac{1}{n!}\Big(\sum_{C\subset V} J_C \prod_{z\in C}\sigma_z\Big)^n \\
&= \sum_{\sigma\in\{-1,+1\}^V} \sum_{n=0}^{\infty}\frac{1}{n!} \sum_{C_1,\dots,C_n\subset V} J_{C_1}\cdots J_{C_n}\Big(\prod_{x\in A}\sigma_x\Big)\Big(\prod_{z_1\in C_1}\sigma_{z_1}\Big)\cdots\Big(\prod_{z_n\in C_n}\sigma_{z_n}\Big).
\end{aligned}$$

Here we rewrite

$$\Big(\prod_{x\in A}\sigma_x\Big)\Big(\prod_{z_1\in C_1}\sigma_{z_1}\Big)\cdots\Big(\prod_{z_n\in C_n}\sigma_{z_n}\Big) = \prod_{y\in V}\sigma_y^{m_y(A,C_1,\dots,C_n)},$$

where $m_y = m_y(A, C_1, \dots, C_n)$ denotes the number of times that the site y appears in the subsets $A, C_1, \dots, C_n$. Then note that the sum over a single spin value takes the form

$$\sum_{\sigma_y\in\{-1,+1\}} \sigma_y^{m_y} = \begin{cases} 2 & \text{if } m_y \text{ is even} \\ 0 & \text{if } m_y \text{ is odd} \end{cases}$$

and consequently the sum over spin configurations becomes

$$\sum_{\sigma\in\{-1,+1\}^V}\prod_{y\in V}\sigma_y^{m_y} = \prod_{y\in V}\sum_{\sigma_y\in\{-1,+1\}}\sigma_y^{m_y} = \begin{cases} 2^{\#V} & \text{if } m_y \text{ is even for all } y\in V \\ 0 & \text{otherwise.} \end{cases}$$

Thus we have simplified the numerator to a sum of non-negative terms

$$\begin{aligned}
&\sum_{\sigma\in\{-1,+1\}^V} \Big(\prod_{x\in A}\sigma_x\Big) e^{-h(\sigma)} \\
&= \sum_{n=0}^{\infty}\frac{1}{n!} \sum_{C_1,\dots,C_n\subset V} J_{C_1}\cdots J_{C_n}\; 2^{\#V} \prod_{y\in V} \mathbb{I}_{m_y(A,C_1,\dots,C_n) \text{ even}}
\end{aligned}$$

and the assertion $\mathsf{E}\big[\prod_{x\in A}\sigma_x\big] \ge 0$ follows.

Proof of part (ii) of Proposition IX.2: To prove the second statement, we consider a duplicate system, i.e., the product measure $\mathsf{P}^2 = \mathsf{P}\otimes\mathsf{P}$ on $\{-1,+1\}^V \times \{-1,+1\}^V$. The product measure is determined by

$$\mathsf{P}^2\big[\,\{(\sigma,\widetilde\sigma)\}\,\big] = \frac{1}{Z^2}\, e^{-h(\sigma)-h(\widetilde\sigma)}.$$

To simplify notation, let us use the following shorthands for products of spins:

$$\sigma_A := \prod_{x\in A}\sigma_x, \qquad \widetilde\sigma_A := \prod_{x\in A}\widetilde\sigma_x, \qquad \text{for } A\subset V.$$

With expected values E^2 under the product measure, we can write the expression in part (ii) as

$$\mathsf{E}\Big[\sigma_{A_1}\sigma_{A_2}\Big] - \mathsf{E}\Big[\sigma_{A_1}\Big]\,\mathsf{E}\Big[\sigma_{A_2}\Big] = \mathsf{E}^2\Big[\sigma_{A_1}\Big(\sigma_{A_2}-\widetilde\sigma_{A_2}\Big)\Big].$$

It is thus sufficient to prove the non-negativity of

$$\begin{aligned}
&Z^2\,\mathsf{E}^2\Big[\sigma_{A_1}\Big(\sigma_{A_2}-\widetilde\sigma_{A_2}\Big)\Big] \\
&= \sum_{\sigma,\widetilde\sigma\in\{-1,+1\}^V} \sigma_{A_1}\big(\sigma_{A_2}-\widetilde\sigma_{A_2}\big)\,\exp\Big(\sum_{C\subset V} J_C(\sigma_C+\widetilde\sigma_C)\Big).
\end{aligned}$$

Change variables in the second summation to $\widetilde{\tau}_x = \widetilde{\sigma}_x/\sigma_x$

$$\begin{aligned}
&\sum_{\sigma,\widetilde{\sigma}\in\{-1,+1\}^V} \sigma_{A_1}\left(\sigma_{A_2}-\widetilde{\sigma}_{A_2}\right)\,\exp\Big(\sum_{C\subset V} J_C(\sigma_C+\widetilde{\sigma}_C)\Big)\\
&= \sum_{\sigma,\widetilde{\tau}\in\{-1,+1\}^V} \sigma_{A_1}\sigma_{A_2}\left(1-\widetilde{\tau}_{A_2}\right)\,\exp\Big(\sum_{C\subset V} J_C(1+\widetilde{\tau}_C)\,\sigma_C\Big)\\
&= \sum_{\widetilde{\tau}\in\{-1,+1\}^V} \left(1-\widetilde{\tau}_{A_2}\right) \sum_{\sigma\in\{-1,+1\}^V} \sigma_{A_1}\sigma_{A_2}\,\exp\Big(\sum_{C\subset V} J_C(1+\widetilde{\tau}_C)\,\sigma_C\Big).
\end{aligned}$$

Consider the inner summation over σ, with fixed $\widetilde{\tau}$. By part (i), applied with coupling constants $J'_C = J_C(1+\widetilde{\tau}_C) \geq 0$, we have that the inner sum is non-negative,

$$\sum_{\sigma\in\{-1,+1\}^V} \sigma_{A_1}\sigma_{A_2}\,\exp\Big(\sum_{C\subset V} J_C(1+\widetilde{\tau}_C)\,\sigma_C\Big) \geq 0.$$

Since moreover $\left(1-\widetilde{\tau}_{A_2}\right) \geq 0$, also the outer summation over $\widetilde{\tau}$ yields a non-negative result, and we can conclude that

$$Z^2\;\mathsf{E}^2\Big[\sigma_{A_1}\left(\sigma_{A_2}-\widetilde{\sigma}_{A_2}\right)\Big]\;\geq 0,$$

which proves the assertion (ii). □

The Griffiths' inequality is often useful as such, but it also has less direct important consequences. As an example we next deduce from it that enlarging the graph $\mathcal{G}$ or adding bonds to it makes correlations of Ising spins stronger. This will be instrumental in the next lecture when proving that the infinite volume limit of the Ising model with free boundary conditions exists.

Proposition IX.4 (Enlarging the graph strengthens Ising correlations)**.** *Let $\mathcal{G} = (\mathrm{V},\mathrm{E})$ and $\widetilde{\mathcal{G}} = (\widetilde{\mathrm{V}},\widetilde{\mathrm{E}})$ be two finite graphs such that $\mathrm{V}\subset\widetilde{\mathrm{V}}$ and $\mathrm{E}\subset\widetilde{\mathrm{E}}$. Fix an inverse temperature $\beta\geq 0$ and a non-negative external magnetic field $B\geq 0$. Consider the Ising models $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ and $\mathsf{P}^{(\widetilde{\mathcal{G}})}_{\beta,B}$ on these two graphs. Then, for any $A\subset\mathrm{V}$ we have*

$$\mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{z\in A}\sigma_z\Big]\;\leq\;\mathsf{E}^{(\widetilde{\mathcal{G}})}_{\beta,B}\Big[\prod_{z\in A}\sigma_z\Big].$$

Proof. Let us first observe that both sides of the asserted inequality can be seen as expected values of $\prod_{z\in A}\sigma_z$ under some probability measures P and $\widetilde{P}$ on the sample space $\{-1,+1\}^{\widetilde{\mathrm{V}}}$ corresponding to the larger graph. Indeed, the right hand side is exactly the expected value under $\widetilde{P} := \mathsf{P}^{(\widetilde{\mathcal{G}})}_{\beta,B}$, the Ising measure on the larger graph $\widetilde{\mathcal{G}}$. This measure is of the form

$$\widetilde{P}[\{\sigma\}] = \frac{1}{\widetilde{Z}}\,e^{-\widetilde{h}(\sigma)}, \qquad \text{where} \qquad \widetilde{h}(\sigma) = -\sum_{x\in\widetilde{\mathrm{V}}}\beta B\,\sigma_x - \sum_{\{x,y\}\in\widetilde{\mathrm{E}}}\beta\,\sigma_x\sigma_y.$$

We observe that also the left hand side is the expected value under the measure P of the form

$$P[\{\sigma\}] = \frac{1}{Z}\,e^{-h(\sigma)}, \qquad \text{where} \qquad h(\sigma) = -\sum_{x\in\mathrm{V}}\beta B\,\sigma_x - \sum_{\{x,y\}\in\mathrm{E}}\beta\,\sigma_x\sigma_y,$$

since under this measure the restriction $\sigma|_{\mathrm{V}}$ of the spin configuration $\sigma\in\{-1,+1\}^{\widetilde{\mathrm{V}}}$ to the subset $\mathrm{V}\subset\widetilde{\mathrm{V}}$ has the same distribution as the Ising model on the smaller graph $\mathcal{G}$ — the other spins $\sigma|_{\widetilde{V}\setminus V}$ are independent of $\sigma|_V$.

We will more generally, for $s,t\in[0,1]$, introduce the measure

$$P_{s,t}[\{\sigma\}] = \frac{1}{Z_{s,t}}\,e^{-h_{s,t}(\sigma)},$$

where

$$h_{s,t}(\sigma) = -\sum_{x\in \mathrm{V}} \beta B\,\sigma_x - s\sum_{x\in \widetilde{\mathrm{V}}\setminus \mathrm{V}} \beta B\,\sigma_x - \sum_{\{x,y\}\in \mathrm{E}} \beta\,\sigma_x\sigma_y - t\sum_{\{x,y\}\in \widetilde{\mathrm{E}}\setminus \mathrm{E}} \beta\,\sigma_x\sigma_y.$$

Note that the measures $P_{s,t}$ are of the form considered in Proposition IX.2, and that $P_{0,0} = P$ and $P_{1,1} = \widetilde{P}$. Therefore, in order to prove the asserted inequality concerning expected values under P and $\widetilde{P}$, it suffices to show that

$$\mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big] = \Big(\sum_{\sigma\in\{-1,+1\}^{\widetilde{\mathrm{V}}}} (\prod_{z\in A}\sigma_z)e^{-h_{s,t}(\sigma)}\Big)\Big/\Big(\sum_{\sigma\in\{-1,+1\}^{\widetilde{\mathrm{V}}}} e^{-h_{s,t}(\sigma)}\Big)$$

is increasing in both s and t.

To prove that $\mathsf{E}_{s,t}\big[\prod_A \sigma_z\big]$ is increasing in s, calculate the derivative

$$\begin{aligned}
&\frac{\partial}{\partial s}\,\mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big] \\
&= \frac{\partial}{\partial s}\Big(\frac{\sum_\sigma \big(\prod_A \sigma_z\big)e^{-h_{s,t}(\sigma)}}{\sum_\sigma e^{-h_{s,t}(\sigma)}}\Big) \\
&= \frac{\sum_\sigma(\prod_A \sigma_z)(-\frac{\partial h_{s,t}(\sigma)}{\partial s})e^{-h_{s,t}(\sigma)}}{\sum_\sigma e^{-h_{s,t}(\sigma)}} - \frac{\big(\sum_\sigma(\prod_A \sigma_z)e^{-h_{s,t}(\sigma)}\big)\big(\sum_\sigma(-\frac{\partial h_{s,t}(\sigma)}{\partial s})e^{-h_{s,t}(\sigma)}\big)}{\big(\sum_\sigma e^{-h_{s,t}(\sigma)}\big)^2}.
\end{aligned}$$

Noting that

$$-\frac{\partial h_{s,t}(\sigma)}{\partial s} = \beta B\sum_{x\in \widetilde{\mathrm{V}}\setminus \mathrm{V}} \sigma_x,$$

we recognize in these two terms certain expected values, and we get

$$\frac{\partial}{\partial s}\,\mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big] = \beta B\sum_{x\in \widetilde{\mathrm{V}}\setminus \mathrm{V}}\Big(\mathsf{E}_{s,t}\Big[(\prod_{z\in A}\sigma_z)\,\sigma_x\Big] - \mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big]\,\mathsf{E}_{s,t}\Big[\sigma_x\Big]\Big).$$

By Proposition IX.2 (ii), the quantity inside the parentheses is nonnegative, so we see that

$$\frac{\partial}{\partial s}\,\mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big] \;\geq\; 0,$$

and we conclude that $\mathsf{E}_{s,t}\big[\prod_A \sigma_z\big]$ is increasing as a function of s.

By a similar calculation based on the observation

$$-\frac{\partial h_{s,t}(\sigma)}{\partial t} = \beta\sum_{\{x,y\}\in \widetilde{\mathrm{E}}\setminus \mathrm{E}} \sigma_x\sigma_y,$$

and Proposition IX.2 (ii) again, one gets that

$$\frac{\partial}{\partial t}\,\mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big] = \beta\sum_{\{x,y\}\in \widetilde{\mathrm{E}}\setminus \mathrm{E}}\Big(\mathsf{E}_{s,t}\Big[(\prod_{z\in A}\sigma_z)\,\sigma_x\sigma_y\Big] - \mathsf{E}_{s,t}\Big[\prod_{z\in A}\sigma_z\Big]\,\mathsf{E}_{s,t}\Big[\sigma_x\sigma_y\Big]\Big) \;\geq\; 0.$$

This shows that $\mathsf{E}_{s,t}\big[\prod_A \sigma_z\big]$ is also increasing as a function of t. This concludes the proof, since

$$\mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\prod_{z\in A}\sigma_z\Big] = \mathsf{E}_{0,0}\Big[\prod_{z\in A}\sigma_z\Big] \leq \mathsf{E}_{1,1}\Big[\prod_{z\in A}\sigma_z\Big] = \mathsf{E}^{(\widetilde{\mathcal{G}})}_{\beta,B}\Big[\prod_{z\in A}\sigma_z\Big].$$

□

3. FKG inequality for Ising model

A key idea in the proof of existence of the thermodynamical limit is monotonicity. To make sense of monotonicity, recall from Appendix D.2 that the sample space

$\Omega_{\mathcal{G}} = \{-1,+1\}^{\mathrm{V}}$ has a natural partial order $\preceq$, and this gives a notion of increasing functions f defined on $\Omega_{\mathcal{G}}$.

Specifically, for the proof of existence of the infinite volume limit with plus boundary conditions in the next lecture, we will use the following monotonicity result known as the *FKG inequality*, named after the work of Fortuin, Kasteleyn, and Ginibre [**FKG71**].

Let $\mathcal{G} = (\mathrm{V}, \mathrm{E})$ be a finite graph and let $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ be the probability measure (IX.2) on $\Omega_{\mathcal{G}} = \{-1,+1\}^{\mathrm{V}}$. Denote the expected values with respect to this measure by E.

Theorem IX.5 (FKG inequality for Ising model)**.**
Consider the Ising model $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ on a finite graph $\mathcal{G}$. If f, g are increasing functions $\Omega_{\mathcal{G}} \to \mathbb{R}$, then the following FKG inequality holds

$$\mathsf{E}\big[f\,g\big] \;\geq\; \mathsf{E}\big[f\big]\,\mathsf{E}\big[g\big]. \tag{IX.6}$$

Proof. We will prove this using Holley's criterion, Proposition D.9 in Appendix D.

The asserted inequality (IX.6) is not affected by adding a constant to g, so we may assume that g is positive. Moreover, (IX.6) is not affected by multiplying g by a positive constant, so we may assume that the expected value of g is one. Consider thus, without loss of generality, the case

$$g(\sigma) > 0 \quad \forall \sigma \in \{-1,+1\}^{\mathrm{V}} \qquad \text{and} \qquad \mathsf{E}\big[g\big] = 1.$$

Denote $P = \mathsf{P}^{(\mathcal{G})}_{\beta,B}$ the Ising probability measure and

$$p(\sigma) = P[\{\sigma\}] = \frac{e^{-\beta H_{\mathcal{G}}(\sigma)}}{Z_{\mathcal{G}}(\beta, B)} > 0.$$

Define, for all $\sigma \in \{-1,+1\}^{\mathrm{V}}$,

$$q(\sigma) = g(\sigma)\, p(\sigma),$$

and note that

$$q(\sigma) > 0 \qquad \text{and} \qquad \sum_{\sigma} q(\sigma) = \sum_{\sigma} g(\sigma)\, p(\sigma) = \mathsf{E}\big[g\big] = 1,$$

so that q defines another probability measure Q on $\sigma \in \{-1,+1\}^{\mathrm{V}}$ by $Q[\{\sigma\}] = q(\sigma)$. Note also that for any $f\colon \{-1,+1\}^{\mathrm{V}} \to \mathbb{R}$ we have

$$\int f \,\mathrm{d}Q = \sum_{\sigma} f(\sigma)\, q(\sigma) = \sum_{\sigma} f(\sigma)\, g(\sigma)\, p(\sigma) = \int f g \,\mathrm{d}P.$$

In view of the equations $\mathsf{E}[f] = \int f \,\mathrm{d}P$ and $\mathsf{E}[fg] = \int fg \,\mathrm{d}P = \int f \,\mathrm{d}Q$ and the assumption $\mathsf{E}[g] = 1$, the assertion (IX.6) amounts to proving that for all increasing functions $f\colon \{-1,+1\}^{\mathrm{V}} \to \mathbb{R}$ we have

$$\int f \,\mathrm{d}P \leq \int f \,\mathrm{d}Q.$$

This requirement is by definition equivalent to the stochastic domination $P \preceq_{\mathrm{st}} Q$. We will verify Holley's criterion (D.6) for p and q, and the stochastic domination then follows from Proposition D.9.

Let $\sigma, \tau \in \{-1,+1\}^{\mathrm{V}}$ be two spin configurations such that $\sigma \preceq \tau$, and let $z \in \mathrm{V}$. Denote by $\sigma^{\uparrow z}, \sigma^{\downarrow z}, \tau^{\uparrow z}, \tau^{\downarrow z} \in \{-1,+1\}^{\mathrm{V}}$ the configurations modified at the site z as in (D.5). Then

calculate

$$\begin{aligned}\frac{q(\tau^{\uparrow z})}{q(\tau^{\downarrow z})}\,\frac{p(\sigma^{\downarrow z})}{p(\sigma^{\uparrow z})} &= \frac{g(\tau^{\uparrow z})\,p(\tau^{\uparrow z})}{g(\tau^{\downarrow z})\,p(\tau^{\downarrow z})}\,\frac{p(\sigma^{\downarrow z})}{p(\sigma^{\uparrow z})}\\ &= \frac{g(\tau^{\uparrow z})}{g(\tau^{\downarrow z})}\,\exp\Big(-\beta\big(H_{\mathcal{G}}(\tau^{\uparrow z})-H_{\mathcal{G}}(\tau^{\downarrow z})+H_{\mathcal{G}}(\sigma^{\downarrow z})-H_{\mathcal{G}}(\sigma^{\uparrow z})\big)\Big)\\ &= \frac{g(\tau^{\uparrow z})}{g(\tau^{\downarrow z})}\,\exp\Big(2\beta\sum_{\substack{w\in\mathcal{G}\\ \{z,w\}\in \mathrm{E}(\mathcal{G})}}(\tau_w-\sigma_w)\Big).\end{aligned}$$

Now since $\tau^{\downarrow z}\preceq\tau^{\uparrow z}$ and g is increasing, we have $\frac{g(\tau^{\uparrow z})}{g(\tau^{\downarrow z})}\geq 1$. Moreover, since $\sigma\preceq\tau$, we have $\tau_w-\sigma_w\geq 0$ and thus $\exp\big(2\beta\sum(\tau_w-\sigma_w)\big)\geq 1$. We conclude that

$$\frac{q(\tau^{\uparrow z})}{q(\tau^{\downarrow z})}\,\frac{p(\sigma^{\downarrow z})}{p(\sigma^{\uparrow z})}\geq 1,$$

which implies Holley's criterion (D.6), and thus finishes the proof. □

As a corollary, we obtain that conditioning any subset of spins to be positive increases the probability that other spins are positive, too. This should not be surprising in view of the fact that Ising model is intended to describe a ferromagnet. For a subset $A\subset \mathrm{V}$, we denote $\sigma_{|A}\equiv+1$ if $\sigma_x=+1$ for all $x\in A$.

Corollary IX.6 (Positive association property for Ising model)**.**
Consider the Ising model on a finite graph $\mathcal{G}=(\mathrm{V},\mathrm{E})$. Then for any subsets $A,B\subset\mathrm{V}$ we have the following positive association property

$$\mathsf{P}^{(\mathcal{G})}_{\beta,B}\Big[\sigma_{|A}\equiv+1\,\Big|\,\sigma_{|B}\equiv+1\Big]\geq\mathsf{P}^{(\mathcal{G})}_{\beta,B}\Big[\sigma_{|A}\equiv+1\Big].$$

Proof. Define f and g to be the indicators of the events $\{\sigma_{|A}\equiv+1\}$ and $\{\sigma_{|B}\equiv+1\}$, respectively. Then f and g are increasing functions, and

$$\begin{aligned}\mathsf{E}[f] &= \mathsf{P}\big[\sigma_{|A}\equiv+1\big]\\ \frac{\mathsf{E}[fg]}{\mathsf{E}[g]} &= \frac{\mathsf{P}\big[\sigma_{|A}\equiv+1\text{ and }\sigma_{|B}\equiv+1\big]}{\mathsf{P}\big[\sigma_{|B}\equiv+1\big]}=\mathsf{P}\Big[\sigma_{|A}\equiv+1\,\Big|\,\sigma_{|B}\equiv+1\Big].\end{aligned}$$

By the FKG inequality, Theorem IX.5, the left hand side of the second line is greater than or equal to the left hand side of the first line. The assertion follows. □

Lecture X

Ising model: thermodynamical limit

In this chapter we continue to study the Ising model, now specifically the d-dimensional version of it. We show that Ising model on the infinite lattice $\mathbb{Z}^d$ can be defined as the weak limit of the Ising models on increasingly large finite subgraphs of $\mathbb{Z}^d$. This is the "thermodynamical limit" or the "infinite volume limit" of the Ising model. More specifically, we impose certain boundary conditions for the Ising model on finite subgraphs, and for each choice show the existence of an infinite volume limit.

Such infinite volume limits can be used to give a mathematically precise meaning to phase transitions — in the Ising model and more generally. Specifically for the Ising model in dimensions $d \geq 2$ it can be shown that in temperatures above a certain critical temperature, the infinite volume limits with different boundary conditions all coincide, whereas in temperatures below critical, there exist different infinite volume limits which in some way remember something about boundary conditions arbitrarily far away (through a long range ferromagnetic order). In this sense the infinite volume limit of the Ising model in dimensions $d \geq 2$ exhibits a ferromagnetic low temperature phase and a paramagnetic high temperature phase, with a phase transition at the critical Curie temperature.

In addition to the correct qualitative phase transition, the Ising model appears to have the quantitatively correct critical exponents as well as other universal critical behavior for d-dimensional uniaxial ferromagnetic materials. In this sense it is a significant improvement on the Curie–Weiss model.

1. Preparations for infinite volume limits

The main goal of the present lecture is to construct the infinite volume limits of the Ising model. The idea is to obtain the Ising model on the infinite graph $\mathbb{Z}^d$, a probability measure on $\{-1,+1\}^{\mathbb{Z}^d}$, as a weak limit of the Ising model probability measures on an increasing sequence of finite subgraphs of $\mathbb{Z}^d$.

Increasing sequence of finite subgraphs

Consider a sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of finite induced subgraphs of the lattice $\mathbb{Z}^d$, i.e., for each $n \in \mathbb{N}$ let $\mathcal{G}_n = (\mathrm{V}_n, \mathrm{E}_n)$, where

$$\mathrm{V}_n \subset \mathbb{Z}^d$$

is a finite subset of sites and

$$\mathrm{E}_n = \big\{\{x,y\} \;\big|\; x,y \in \mathrm{V}_n,\ \|x-y\| = 1\big\}$$

is the set of bonds of the lattice connecting pairs of these sites. Assume that the sequence of graphs exhausts the whole infinite lattice, $\mathrm{V}_n \uparrow \mathbb{Z}^d$, in the sense that

$$\mathrm{V}_1 \subset \mathrm{V}_2 \subset \cdots \qquad \text{and} \qquad \bigcup_{n\in\mathbb{N}} \mathrm{V}_n = \mathbb{Z}^d.$$

The space for weak convergence

A spin configuration $\sigma^{(n)} = (\sigma_x^{(n)})_{x\in\mathrm{V}_n}$ of the Ising model on $\mathcal{G}_n$ is an element of

$$\Omega_{\mathcal{G}_n} = \{-1,+1\}^{\mathrm{V}_n}.$$

To speak of weak convergence of the laws of $\sigma^{(n)}$, we will need to make all of these live in the same space — namely the sample space

$$\Omega = \{-1,+1\}^{\mathbb{Z}^d}$$

of the Ising model on the infinite lattice $\mathbb{Z}^d$.[1] We choose to extend a spin configuration $\sigma^{(n)} = (\sigma_x^{(n)})_{x\in\mathrm{V}_n}$ as constant $+1$ outside of the finite subset $\mathrm{V}_n \subset \mathbb{Z}^d$ (this choice is particularly natural for the plus boundary conditions), i.e., to define the corresponding $\sigma = (\sigma_x)_{x\in\mathbb{Z}^d}$ by

$$\sigma_x = \begin{cases} \sigma_x^{(n)} & \text{if } x \in \mathrm{V}_n \\ +1 & \text{if } x \in \mathbb{Z}^d \setminus \mathrm{V}_n \end{cases} \tag{X.1}$$

We can thus interpret $\Omega_{\mathcal{G}_n} \subset \Omega$ as a subset, and any probability measure ν_n on $\Omega_{\mathcal{G}_n}$ also as a probability measure on Ω.

Note that the space $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$ is of the form considered in Appendix H.4: a countably infinite Cartesian product of finite spaces. We recall in particular that it is compact, and its topology comes from a complete separable metric. In Section VII.3 we have obtained a convenient criterion for weak convergence on such spaces in terms of probabilities of cylinder events.

2. Infinite volume limit with free boundary conditions

We first treat the case of free boundary conditions. Fix the inverse temperature $\beta > 0$ and external magnetic field $B \in \mathbb{R}$. Let us assume $B \geq 0$.[2]

Let $\mathcal{G}_n = (\mathrm{V}_n, \mathrm{E}_n)$, $n \in \mathbb{N}$, form an increasing sequence of finite subgraphs of the lattice $\mathbb{Z}^d$ as in Section 1 above. Let $\mathsf{P}_n = \mathsf{P}_{\beta,B}^{(\mathcal{G}_n)}$ be the corresponsing Ising model probability measures, interpreted as measures on $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$, i.e.,

$$\mathsf{P}_n\left[\{\sigma\}\right] = \frac{1}{Z_n} e^{-\beta H_n(\sigma)} \qquad \text{for } \sigma \in \Omega_{\mathcal{G}_n} \subset \Omega, \tag{X.2}$$

[1] Extending the spin configuration from V_n to $\mathbb{Z}^d$ is analogous to what we did to random walks to prove Donsker's theorem — we linearly interpolated the walks in between the discrete times of their steps to obtain continuous functions. In this way we could discuss weak convergence of their laws in a space of continuous functions.

[2] The assumption $B \geq 0$ of non-negative external magnetic field does not entail any essential loss of generality. Indeed, the case of negative external magnetic field $B < 0$ can be reduced to it by considering $-\sigma$ instead of σ.

where

$$H_n(\sigma) = -\sum_{\{x,y\}\in \mathrm{E}_n} \sigma_x\sigma_y - B\sum_{x\in \mathrm{V}_n}\sigma_x \qquad \text{for } \sigma\in\Omega_{\mathcal{G}_n}\subset\Omega. \tag{X.3}$$

The terminology "free boundary conditions" is used for this case, since under the measure P_n, the spins living in $\mathrm{V}_n \subset \mathbb{Z}^d$ do not in any way interact with the spins outside the finite subgraph, i.e., in the complement $\mathbb{Z}^d \setminus \mathrm{V}_n$.

We will denote expected values with respect to the measure P_n above by E_n.

Our statement about the infinite volume limit with free boundary conditions is the following.

Theorem X.1 (Infinite volume Ising model with free boundary conditions)**.**
There exists a probability measure $\mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;\mathbf{free})}$ on $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$ such that we have

$$\mathsf{P}_n \xrightarrow{\mathrm{w}} \mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;\mathbf{free})} \qquad \text{as } n\to\infty.$$

The probability measure $\mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;\mathbf{free})}$ does not depend on the chosen sequence of subgraphs $(\mathcal{G}_n)_{n\in\mathbb{N}}$.

The key to the existence of the weak limit is the following limit of correlations, which relies on monotonicity results of the previous lecture.

Lemma X.2 (Spin correlations with free boundary conditions converge)**.**
Let $A\subset\mathbb{Z}^d$ be a finite subset. Then the following limit

$$c_A := \lim_{n\to\infty} \mathsf{E}_n\Big[\prod_{z\in A}\sigma_z\Big]$$

exists, and does not depend on the chosen sequence of subgraphs $(\mathcal{G}_n)_{n\in\mathbb{N}}$.

Proof. Note first that since $\mathrm{V}_n \uparrow \mathbb{Z}^d$ then for large enough n we have $A\subset \mathrm{V}_n$. To study the limits as $n\to\infty$, it is sufficient to consider such large n.

Note also that $\mathcal{G}_n$ is a subgraph of $\mathcal{G}_{n+1}$ in the sense that $\mathrm{V}_n\subset\mathrm{V}_{n+1}$ and $\mathrm{E}_n\subset\mathrm{E}_{n+1}$. Therefore it follows from Proposition IX.4 (a corollary of Griffiths' inequality) that

$$\mathsf{E}_n\Big[\prod_{z\in A}\sigma_z\Big] \le \mathsf{E}_{n+1}\Big[\prod_{z\in A}\sigma_z\Big]$$

for all sufficiently large $n\in\mathbb{N}$. Also since the product $\prod_{z\in A}\sigma_z$ is either $+1$ or -1, the expected values are bounded, $-1\le \mathsf{E}_n\big[\prod_{z\in A}\sigma_z\big]\le 1$ for all $n\in\mathbb{N}$. The increasing bounded sequence has a limit

$$c_A := \lim_{n\to\infty}\mathsf{E}_n\Big[\prod_{z\in A}\sigma_z\Big].$$

In order to show that the limit c_A does not depend of the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of graphs, let us consider another sequence $(\widetilde{\mathcal{G}}_n)_{n\in\mathbb{N}}$ of graphs $\widetilde{\mathcal{G}}_n = (\widetilde{\mathrm{V}}_n, \widetilde{\mathrm{E}}_n)$. We also assume these to be induced subgraphs of $\mathbb{Z}^d$ which exhaust the infinite lattice, $\widetilde{\mathrm{V}}_n\uparrow\mathbb{Z}^d$. Let us denote by $\widetilde{\mathsf{P}}_n = \mathsf{P}_{\beta,B}^{(\widetilde{\mathcal{G}}_n)}$ the corresponding Ising probability measures, $\widetilde{\mathsf{E}}_n$ the corresponding expected values, and

$$\widetilde{c}_A := \lim_{n\to\infty}\widetilde{\mathsf{E}}_n\Big[\prod_{z\in A}\sigma_z\Big]$$

the corresponding limit of spin correlations. We claim that $c_A = \widetilde{c}_A$. For $\varepsilon>0$, choose n_0 large enough so that $\mathsf{E}_{n_0}\big[\prod_{z\in A}\sigma_z\big] > c_A - \varepsilon$. Then observe that the finite set $\mathrm{V}_{n_0}\subset\mathbb{Z}^d$

is contained in $\widetilde{\mathrm{V}}_n$ for large enough n, since $\widetilde{\mathrm{V}}_n \uparrow \mathbb{Z}^d$. Thus for large enough n we have by Proposition IX.4 again that

$$\widetilde{\mathsf{E}}_n\Big[\prod_{z\in A}\sigma_z\Big] \geq \mathsf{E}_{n_0}\Big[\prod_{z\in A}\sigma_z\Big] > c_A - \varepsilon,$$

and in the limit $n\to\infty$ we obtain $\widetilde{c}_A \geq c_A - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we get $\widetilde{c}_A \geq c_A$. Similarly one shows that that $c_A \geq \widetilde{c}_A$. It follows that $c_A = \widetilde{c}_A$, i.e., the limit c_A did not depend on the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$. □

Proof of Theorem X.1: We will verify the existence of the weak limit of P_n by Theorem VII.5. For this, we must establish the convergence of probabilities of cylinder events. In the present case, a cylinder event is an event of the form

$$C^{\epsilon_1,\ldots,\epsilon_m}_{x_1,\ldots,x_m} := \Big\{\sigma\in\{-1,1\}^{\mathbb{Z}^d} \;\Big|\; \sigma_{x_1}=\epsilon_1,\ldots,\sigma_{x_m}=\epsilon_m\Big\},$$

where $x_1,\ldots,x_m\in\mathbb{Z}^d$ and $\epsilon_1,\ldots,\epsilon_m\in\{-1,+1\}$. Observe that we have

$$\prod_{j=1}^m\big(1+\epsilon_j\sigma_{x_j}\big) = \begin{cases} 0 & \text{if } \sigma_{x_j}\neq\epsilon_j \text{ for some } j\\ 2^m & \text{if } \sigma_{x_j}=\epsilon_j \text{ for all } j,\end{cases}$$

because each factor $1+\epsilon_j\sigma_{x_j}$ is either 0 or 2, depending on whether σ_{x_j} coincides with ϵ_j or not. This allows us to write the indicator of the cylinder event as

$$\begin{aligned}\mathbb{I}_{C^{\epsilon_1,\ldots,\epsilon_m}_{x_1,\ldots,x_m}} &= 2^{-m}\prod_{j=1}^m\big(1+\epsilon_j\sigma_{x_j}\big)\\ &= 2^{-m}\sum_{J\subset\{1,\ldots,m\}}\prod_{j\in J}\epsilon_j\sigma_{x_j},\end{aligned}$$

where the second equality is seen by expanding the product with the binomial formula. For $J\subset\{1,\ldots,m\}$, let us denote $A_J := \big\{x_j\,\big|\,j\in J\big\}\subset\mathbb{Z}^d$. Then by taking expected values under the Ising probability measures P_n and using linearity, we get

$$\mathsf{P}_n\big[C^{\epsilon_1,\ldots,\epsilon_m}_{x_1,\ldots,x_m}\big] = 2^{-m}\sum_{J\subset\{1,\ldots,m\}}\Big(\prod_{j\in J}\epsilon_j\Big)\,\mathsf{E}_n\Big[\prod_{x\in A_J}\sigma_x\Big].$$

It now follows from Lemma X.2 that we have

$$\lim_{n\to\infty}\mathsf{P}_n\big[C^{\epsilon_1,\ldots,\epsilon_m}_{x_1,\ldots,x_m}\big] = 2^{-m}\sum_{J\subset\{1,\ldots,m\}}\Big(\prod_{j\in J}\epsilon_j\Big)\,c_{A_J}.$$

This establishes the convergence of the probabilities of cylinder events, and it follows from Theorem VII.5 that the sequence $(\mathsf{P}_n)_{n\in\mathbb{N}}$ of probability measures has a weak limit. Moreover, since c_{A_J} above does not depend on the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of subgraphs by Lemma X.2, it follows that the limits of probabilities of cylinder events do not depend on $(\mathcal{G}_n)_{n\in\mathbb{N}}$, either. Therefore by the uniqueness part in Theorem VII.5 the weak limit does not depend on $(\mathcal{G}_n)_{n\in\mathbb{N}}$. □

3. Infinite volume limit with plus boundary conditions

We then treat the case of plus boundary conditions. Fix the inverse temperature $\beta > 0$ and external magnetic field $B\in\mathbb{R}$.

Let $\mathcal{G}_n = (\mathrm{V}_n,\mathrm{E}_n)$, $n\in\mathbb{N}$, form an increasing sequence of finite subgraphs of the lattice $\mathbb{Z}^d$ as in Section 1 above. Besides the set $\mathrm{V}_n\subset\mathbb{Z}^d$ of sites and the set E_n of bonds of the graph, let us denote by

$$\partial\mathrm{E}_n := \Big\{\{x,y\}\;\Big|\; x\in\mathrm{V}_n,\ y\in\mathbb{Z}^d\setminus\mathrm{V}_n,\ \|x-y\|=1\Big\}$$

the set of those bonds of $\mathbb{Z}^d$ which connect a site of the subgraph to a site of the complement. The set $\partial \mathrm{E}_n$ of such bonds is interpreted as the boundary of the finite subgraph $\mathcal{G}_n$, and is used below in formulating the boundary conditions.

Recall that we extend a spin configuration $\sigma^{(n)} \in \{-1,+1\}^{\mathrm{V}_n} = \Omega_{\mathcal{G}_n}$ from $\mathrm{V}_n \subset \mathbb{Z}^d$ according to (X.1), i.e., we identify it with the configuration $\sigma \in \{-1,+1\}^{\mathbb{Z}^d} = \Omega$ on the infinite lattice, which coincides with $\sigma^{(n)}$ in V_n and is constant $+1$ outside V_n. We thus interpret $\Omega_{\mathcal{G}_n} \subset \Omega$. The Ising model on $\mathcal{G}_n$ with plus boundary conditions is the probability measure on $\Omega_{\mathcal{G}_n} \subset \Omega$ defined by setting

$$\mathsf{P}_n^+\big[\,\{\sigma\}\,\big] = \frac{1}{Z_n^+}\, e^{-\beta H_n^+(\sigma)} \qquad \text{for } \sigma \in \Omega_{\mathcal{G}_n} \subset \Omega, \tag{X.4}$$

where the energy with plus boundary conditions is defined by

$$H_+^{(\mathcal{G})}(\sigma) = -\sum_{\{x,y\}\in \mathrm{E}\cup\partial\mathrm{E}} \sigma_x\sigma_y - B\sum_{x\in\mathrm{V}}\sigma_x \qquad \text{for } \sigma \in \Omega_{\mathcal{G}_n} \subset \Omega, \tag{X.5}$$

and the partition function is

$$Z_+^{(\mathcal{G}_n)} = \sum_{\sigma\in\Omega_{\mathcal{G}_n}} e^{-\beta H_+^{(\mathcal{G}_n)}(\sigma)}. \tag{X.6}$$

The only difference in the energy (X.5) with plus boundary conditions and the earlier (X.3) is that the sum of interaction terms include also interactions across the bonds on the boundary. Through this interaction, the random spins in V_n feel the fixed plus spins outside V_n.

We will denote expected values with respect to the measure P_n^+ above by E_n^+.

Our statement about the infinite volume limit with plus boundary conditions is the following.

Theorem X.3 (Infinite volume Ising model with plus boundary conditions)**.**
There exists a probability measure $\mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;+)}$ on $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$ such that we have

$$\mathsf{P}_n^+ \xrightarrow{\mathrm{w}} \mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;+)} \qquad \text{as } n\to\infty.$$

The probability measure $\mathsf{P}_{\beta,B}^{(\mathbb{Z}^d;+)}$ does not depend on the chosen sequence of subgraphs $(\mathcal{G}_n)_{n\in\mathbb{N}}$.

The key to the existence of the weak limit is again a monotonicity result from the previous lecture.

Let us first, however, note that the Ising model with plus boundary conditions on a smaller finite subgraph is obtained from the Ising model with plus boundary conditions on a larger finite subgraph by conditioning the spins outside the smaller subgraphs to be plus.

Lemma X.4 (Domain Markov property for Ising model)**.**
Let $\mathcal{G}_n = (\mathrm{V}_n, \mathrm{E}_n)$, $n\in\mathbb{N}$, be a sequence of increasing induced subgraphs of $\mathbb{Z}^d$, and let P_n^+, $n\in\mathbb{N}$, be the corresponding Ising model probability measures with plus boundary conditions, as above. Then for any $n\in\mathbb{N}$ we have the equality

of probability measures

$$\mathsf{P}_n^+[\,\cdot\,] \;=\; \mathsf{P}_{n+1}^+\Big[\;\cdot\;\Big|\;\sigma_{|\mathrm{V}_{n+1}\setminus\mathrm{V}_n}\equiv+1\Big].$$

The proof is done by observing that both probability measures are supported on the same set $\Omega_{\mathcal{G}_n}\subset\Omega$ of spin configurations, and that one can verify directly from the definition that the probabilities of such spin configurations under the two measures are proportional to each other. We leave the details of the verification as an exercise.

Exercise X.1 (Domain Markov property for Ising model)**.**
Prove Lemma X.4.

We can then observe that the effect of plus boundary conditions to favor plus spins becomes weaker when the finite subgraph becomes larger, i.e., when the boundary is further away.

Proposition X.5 (Magnetizations are smaller when plus boundary is further)**.**
Let $\mathcal{G}_n=(\mathrm{V}_n,\mathrm{E}_n)$, $n\in\mathbb{N}$, be a sequence of increasing induced subgraphs of $\mathbb{Z}^d$, and let P_n^+, $n\in\mathbb{N}$, be the corresponding Ising model probability measures with plus boundary conditions, as above. Then for any $n\in\mathbb{N}$ and any subset $A\subset\mathrm{V}_n\subset\mathrm{V}_{n+1}$ we have

$$\mathsf{P}_n^+\Big[\sigma_{|A}\equiv+1\Big]\geq\mathsf{P}_{n+1}^+\Big[\sigma_{|A}\equiv+1\Big].$$

Proof. By the domain Markov property of Lemma X.4 above, we have

$$\mathsf{P}_n^+\Big[\sigma_{|A}\equiv+1\Big] \;=\; \mathsf{P}_{n+1}^+\Big[\sigma_{|A}\equiv+1\;\Big|\;\sigma_{|\mathrm{V}_{n+1}\setminus\mathrm{V}_n}\equiv+1\Big].$$

Observe then, that the FKG inequality holds for P_{n+1}^+ — the proof is exactly the same as in Theorem IX.5. It follows that the positive association of Corollary IX.6 also holds for P_{n+1}^+, and we therefore get

$$\mathsf{P}_{n+1}^+\Big[\sigma_{|A}\equiv+1\;\Big|\;\sigma_{|\mathrm{V}_{n+1}\setminus\mathrm{V}_n}\equiv+1\Big] \;\geq\; \mathsf{P}_{n+1}^+\Big[\sigma_{|A}\equiv+1\Big].$$

Combining the conclusion from the domain Markov property with that from the positive association finishes the proof. □

The above directly implies the convergence of certain probabilities in the infinite volume limit.

Lemma X.6 (Plus spin probabilities with plus boundary conditions converge)**.**
Let $A\subset\mathbb{Z}^d$ be a finite subset. Then the following limit

$$p_A:=\lim_{n\to\infty}\;\mathsf{P}_n^+\big[\sigma_{|A}\equiv+1\big]$$

exists, and does not depend on the chosen sequence of subgraphs $(\mathcal{G}_n)_{n\in\mathbb{N}}$.

Proof. Observe first that $A\subset\mathrm{V}_n$ for large enough n, so it suffices to consider such n.

From Proposition X.5 we get that the sequence of probabilities is decreasing,

$$\mathsf{P}_n^+\big[\sigma_{|A}\equiv+1\big]\;\geq\;\mathsf{P}_{n+1}^+\big[\sigma_{|A}\equiv+1\big]\;\geq\;\mathsf{P}_{n+2}^+\big[\sigma_{|A}\equiv+1\big]\;\geq\;\cdots,$$

so the limit

$$p_A:=\lim_{n\to\infty}\;\mathsf{P}_n^+\big[\sigma_{|A}\equiv+1\big]$$

of this bounded decreasing sequence exists.

It remains to show that the limit p_A does not depend of the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of graphs. Consider another sequence $(\widetilde{\mathcal{G}}_n)_{n\in\mathbb{N}}$ of graphs $\widetilde{\mathcal{G}}_n = (\widetilde{\mathrm{V}}_n, \widetilde{\mathrm{E}}_n)$. We also assume these to be induced subgraphs of $\mathbb{Z}^d$ which exhaust the infinite lattice, $\widetilde{\mathrm{V}}_n \uparrow \mathbb{Z}^d$. Let us denote by $\widetilde{\mathsf{P}}_n^+$ the corresponding Ising probability measures with plus boundary conditions, and

$$\widetilde{p}_A := \lim_{n\to\infty} \widetilde{\mathsf{P}}_n^+\big[\sigma_{|A} \equiv +1\big]$$

the corresponding limit of plus spin probabilities. We claim that $p_A = \widetilde{p}_A$. For $\varepsilon > 0$, choose n_0 large enough so that $\mathsf{P}_{n_0}^+\big[\sigma_{|A} \equiv +1\big] < p_A + \varepsilon$. Then observe that the finite set $\mathrm{V}_{n_0} \subset \mathbb{Z}^d$ is contained in $\widetilde{\mathrm{V}}_n$ for large enough n, since $\widetilde{\mathrm{V}}_n \uparrow \mathbb{Z}^d$. For these large n we can argue by domain Markov property and FKG inequality as in Proposition X.5 that

$$\widetilde{\mathsf{P}}_n^+\big[\sigma_{|A} \equiv +1\big] \leq \mathsf{P}_{n_0}^+\big[\sigma_{|A} \equiv +1\big] < p_A + \varepsilon,$$

and in the limit $n \to \infty$ we obtain $\widetilde{p}_A \leq p_A + \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we get $\widetilde{p}_A \leq p_A$. Similarly one shows that that $p_A \leq \widetilde{p}_A$. It follows that $p_A = \widetilde{p}_A$, i.e., the limit p_A did not depend on the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$. □

The above lemma addresses the probabilities of events of observing only plus spins in a subset $A \subset \mathbb{Z}^d$. By inclusion-exclusion we can, however, express the probability of an arbitrary cylinder event in terms of these.

Exercise X.2 (Cylinder probabilities from plus spin probabilities)**.**
Consider a cylinder event of the form

$$C_{x_1,\dots,x_m}^{\epsilon_1,\dots,\epsilon_m} := \Big\{\sigma \in \{-1,1\}^{\mathbb{Z}^d} \;\Big|\; \sigma_{x_1} = \epsilon_1, \dots, \sigma_{x_m} = \epsilon_m\Big\},$$

where $x_1, \dots, x_m \in \mathbb{Z}^d$ and $\epsilon_1, \dots, \epsilon_m \in \{-1, +1\}$. Partition the marked points $x_1, \dots, x_m$ to the two subsets

$$A_+ := \big\{x_j \;\big|\; \epsilon_j = +1\big\} \qquad \text{and} \qquad A_- := \big\{x_j \;\big|\; \epsilon_j = -1\big\}$$

according to the required values ϵ_j of the spins σ_{x_j}. Prove that

$$\mathsf{P}_n^+\big[C_{x_1,\dots,x_m}^{\epsilon_1,\dots,\epsilon_m}\big] = \sum_{J\subset A_-} (-1)^{\#J}\, \mathsf{P}_n^+\Big[\sigma_{|A_+\cup J} \equiv +1\Big].$$

We are now ready to address the infinite volume limit of the Ising model with plus boundary conditions.

Proof of Theorem X.3: We will verify the existence of the weak limit of P_n^+ using Theorem VII.5 again. For this, we must establish the convergence of probabilities of cylinder events of the form

$$C_{x_1,\dots,x_m}^{\epsilon_1,\dots,\epsilon_m} := \Big\{\sigma \in \{-1,1\}^{\mathbb{Z}^d} \;\Big|\; \sigma_{x_1} = \epsilon_1, \dots, \sigma_{x_m} = \epsilon_m\Big\},$$

where $x_1, \dots, x_m \in \mathbb{Z}^d$ and $\epsilon_1, \dots, \epsilon_m \in \{-1, +1\}$. For a given event of this type, let us partition the marked points $x_1, \dots, x_m$ to the two subsets

$$A_+ := \big\{x_j \;\big|\; \epsilon_j = +1\big\} \qquad \text{and} \qquad A_- := \big\{x_j \;\big|\; \epsilon_j = -1\big\}$$

according to the required values ϵ_j of the spins σ_{x_j}. By Exercise X.2 we can write

$$\mathsf{P}_n^+\big[C_{x_1,\dots,x_m}^{\epsilon_1,\dots,\epsilon_m}\big] = \sum_{J\subset A_-} (-1)^{\#J}\, \mathsf{P}_n^+\Big[\sigma_{|A_+\cup J} \equiv +1\Big].$$

By Lemma X.6, the terms on the right hand side have limits, so we get

$$\lim_{n\to\infty} \mathsf{P}_n^+\big[C_{x_1,\dots,x_m}^{\epsilon_1,\dots,\epsilon_m}\big] = \sum_{J\subset A_-} (-1)^{\#J}\, p_{A_+\cup J}.$$

This establishes the convergence of the probabilities of cylinder events, and it follows from Theorem VII.5 that the sequence $(\mathsf{P}_n^+)_{n\in\mathbb{N}}$ of probability measures has a weak limit. Moreover, since $p_{A_+\cup J}$ above does not depend on the chosen sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of subgraphs

by Lemma X.6, it follows that the limits of probabilities of cylinder events do not depend on $(\mathcal{G}_n)_{n\in\mathbb{N}}$, either. Therefore by the uniqueness part in Theorem VII.5 the weak limit does not depend on $(\mathcal{G}_n)_{n\in\mathbb{N}}$. □

4. Phase transition in the Ising model

We finish our discussion of the Ising model by stating the mathematical result about its phase transition, which occurs as long as the spatial dimension is at least two, $d \geq 2$. The qualitative interpretation is the same as for Curie–Weiss model: in low temperatures the model has ferromagnetic behavior, whereas in high temperatures the behavior is paramagnetic.

The result below says that the ferromagnetic phase is distinguished from the paramagnetic phase by the property that the plus boundary conditions retain an influence even in the infinite volume limit: the limit exhibits a strictly positive spontaneous magnetization.

Theorem X.7 (Phase transition in the Ising model)**.**

Consider the infinite volume limit $\mathsf{P}_{\beta,0}^{(\mathbb{Z}^d;+)}$ of the Ising model with plus boundary conditions, without external magnetic field ($B = 0$), in dimension $d \geq 2$. Then there exists a critical value $\beta_c > 0$ of the inverse temperature parameter such that the magnetization $\mathsf{E}_{\beta,0}^{(\mathbb{Z}^d;+)}\big[\sigma_x\big]$ at any site $x \in \mathbb{Z}^d$ of the infinite lattice depends on inverse temperature as follows:

- *If $\beta < \beta_c$, then $\mathsf{E}_{\beta,0}^{(\mathbb{Z}^d;+)}\big[\sigma_x\big] = 0$.*
- *If $\beta > \beta_c$, then there exists $\bar{m} = \bar{m}_d(\beta) > 0$ such that $\mathsf{E}_{\beta,0}^{(\mathbb{Z}^d;+)}\big[\sigma_x\big] = \bar{m}$.*

By contrast, for the infinite volume limit with free boundary conditions one obviously has $\mathsf{E}_{\beta,0}^{(\mathbb{Z}^d;\mathbf{free})}\big[\sigma_x\big] = 0$ by symmetry, for any $\beta > 0$. Thus the above result in particular implies that in the low temperature ferromagnetic phase $\beta > \beta_c$ we have $\mathsf{P}_{\beta,0}^{(\mathbb{Z}^d;+)} \neq \mathsf{P}_{\beta,0}^{(\mathbb{Z}^d;\mathbf{free})}$. It can be shown that in the high temperature paramagnetic phase $\beta < \beta_c$, we have the coincidence $\mathsf{P}_{\beta,0}^{(\mathbb{Z}^d;+)} = \mathsf{P}_{\beta,0}^{(\mathbb{Z}^d;\mathbf{free})}$. In other words, the boundary conditions affect the infinite volume limits in the ferromagnetic phase, but not in the paramagnetic phase.

Lecture XI

Interacting particle systems

In this lecture we study a certain class of continuous time Markov processes known as "interacting particle systems". Among the typical features characteristic of these processes are

- state space $S^{\mathbb{Z}^d}$, where S is a finite set, and $\mathbb{Z}^d$ is the d-dimensional integer lattice
- flip rates of components are given by a translation invariant rule and they only depend on finitely many other components.

We will give one definition based on these properties below, which is general enough to include a number of important examples, but specific enough that we have a unified notation and that we can give a unified proof of the existence of these processes. Such a definition is, however, almost bound to be too narrow in some respect, and there are models of essentially the same flavor but which fail to fit exactly within the scope of the definition. It is of course fair to also call these essentially similar models interacting particle systems. The reader interested in studying the topic in more depth can start, e.g., from the lecture series [**Dur95**] or the textbook [**Lig05**].

The set S represents the possible states of a single component of the process, and the components are indexed by the sites of the lattice $\mathbb{Z}^d$ which represent a discretization of d-dimensional space. We will denote the process by $(\xi_t)_{t\in[0,\infty)}$, where for each fixed $t \geq 0$ the state of the process is a configuration $\xi_t = (\xi_t(x))_{x\in\mathbb{Z}^d} \in S^{\mathbb{Z}^d}$ of states of its components. Each component keeps flipping in time according to a jump process on S, which by itself is not Markovian. Rather, its jump rates depend on the states of some finite number of neighboring sites, which creates an "interaction" between the components of the process.

An essential difficulty in the study of interacting particle systems arises from the fact that the state space $S^{\mathbb{Z}^d}$ of the process is uncountably infinite.[1]

To constrast with something more familiar, recall that a stochastic process $(X_t)_{t\in[0,\infty)}$ on a finite state space $\mathcal{S}$ is a continuous time Markov process with *jump rates* $\lambda(a,b)$, $a, b \in \mathcal{S}$, $a \neq b$, if we have

$$\mathsf{P}\Big[X_{t+\varepsilon} = b \;\Big|\; X_t = a,\ \mathfrak{H}\Big] = \varepsilon\,\lambda(a,b) + o(\varepsilon) \tag{XI.1}$$

for any $t \geq 0$, any states $a, b \in \mathcal{S}$, $a \neq b$, and any event[2] $\mathfrak{H} \in \sigma(X_s : 0 \leq s \leq t)$ that only depends on the history of the process up to time t. We call $(X_t)_{t\in[0\infty)}$ a jump process, because $t \mapsto X_t$ stays constant over the time intervals between a locally finite set of jump times.

[1] Nevertheless, the space $S^{\mathbb{Z}^d}$ is topologically very well behaved, since it is a complete, separable metric space (even compact, since S is assumed finite) — see Appendix H.4.

[2] To be precise, it is meaningful to require (XI.1) to hold when the event $\mathfrak{H}$ is furthermore such that $\mathfrak{H} \cap \{X_t = a\}$ has positive probability, so that the conditional probability in the equation is meaningful.

For an interacting particle system $(\xi_t)_{t\in[0,\infty)}$, each of the infinitely many components of the process keeps flipping in time, so in any positive amount of time there are typically infinitely many flips overall, and the process $t \mapsto \xi_t$ is not constant over any nontrivial time interval. It is therefore not meaningful to talk about the jump rates of the process from one state to another, but rather only about the flip rates of its components. We would thus like to write an analogous defining property for the process as

$$\mathsf{P}\Big[\xi_{t+\varepsilon}(x) = a \;\Big|\; \xi_t = \eta,\ \mathfrak{H}\Big] = \varepsilon\, c_a^{(x)}(\eta) + o(\varepsilon), \tag{XI.2}$$

where $c_a^{(x)}(\eta)$ is the flip rate to state $a \in S$ at the site $x \in \mathbb{Z}^d$ given that the current configuration is $\eta \in S^{\mathbb{Z}^d}$ and $\eta(x) \neq a$. Even this requires some care, because the state space $S^{\mathbb{Z}^d}$ is uncountably infinite, and the event $\{\xi_t = \eta\}$ in the conditioning typically has zero probability. It is possible to interpret (XI.2) as a regular conditional probability (since $S^{\mathbb{Z}^d}$ is a complete separable metric space), but even without going into the details of the mathematical meaning of the equation, (XI.2) serves to give the key intuitive interpretation of the flip rates $c_a^{(x)}(\eta)$.

Just like continuous time Markov jump processes are usually defined by specifying the jump rates $\lambda(a, b)$, for $a, b \in \mathcal{S}$, $a \neq b$, interacting particle systems are usually defined by specifying the flip rates $c_a^{(x)}(\eta)$, for $x \in \mathbb{Z}^d$, $a \in S$, and $\eta \in S^{\mathbb{Z}^d}$ such that $\eta(x) \neq a$. The first obvious question that arises is:

> **Question 0:** Do the given flip rates $c_a^{(x)}(\eta)$ define a (unique) Markov process $(\xi_t)_{t\in[0,\infty)}$ on $S^{\mathbb{Z}^d}$.

Later in this lecture we will give an affirmative answer with some reasonable conditions on the flip rates, by concretely constructing the process. As for finite state space Markov processes, once the process has been constructed, the most natural questions concern its long time behavior and stationary distribution(s):

> **Question 1:** Describe the stationary distributions of the process $(\xi_t)_{t\in[0,\infty)}$.
> 1 a): Are there "nontrivial" stationary distributions?[3]
> 1 b): Does the process converge to a stationary distribution as time increases?

Let us now give some examples of interacting particle systems, and comment on their stationary measures.

Example XI.1 (Dynamical Ising model)**.**

Recall from Exercise IX.2 that the Glauber dynamics for the Ising model on a finite graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$ is a continuous time Markov process on the set of spin configurations $\{-1, +1\}^{\mathrm{V}}$, which models the effect of thermal motion in the magnetic material.

The dynamical Ising model is a version of the Glauber dynamics in the infinite volume Ising model. Set $S = \{-1, +1\}$, so that the states of the dynamical Ising model are spin configurations $\sigma \in \{-1, +1\}^{\mathbb{Z}^d}$ on the d-dimensional lattice $\mathbb{Z}^d$. As the flip rate of the spin

[3]Many interacting particle systems have stationary distributions which are in some sense obvious or "trivial". The question concerns the existence of other stationary distributions, and thus in particular the non-uniqueness of stationary distributions. The precise meaning of triviality of a stationary distribution depends on the model, and is illustrated by the examples below.

at x to value $\epsilon = \pm 1$, we use

$$c_\epsilon^{(x)}(\sigma) := \left(1 + \exp\Big(-2\epsilon\beta\,(B + \sum_{\substack{y\in\mathbb{Z}^d\\ \|y-x\|=1}} \sigma_y)\Big)\right)^{-1},$$

in analogy of the formula in Exercise IX.2, whenever $x \in \mathbb{Z}^d$ and $\sigma \in \{-1,+1\}^{\mathbb{Z}^d}$ is such that $\sigma(x) \neq \epsilon$.

The infinite volume Ising model probability distributions (with all possible boundary conditions) constructed in Lecture X can be shown to be stationary distributions for the process $(\xi_t)_{t\in[0,\infty)}$ with these flip rates. They are thus thermal equilibrium measures, when thermal motion is modeled by the process $(\xi_t)_{t\in[0,\infty)}$. The question of uniqueness of stationary distributions is now related to the uniqueness of infinite volume Ising probability measures, and therefore to the phase transition in the Ising model.

Example XI.2 (Voter model)**.**
In the voter model we take S to be some finite set of "opinions", e.g.,

$$S = \{\text{"republican"},\ \text{"democrat"}\}.$$

The lattice $\mathbb{Z}^d$ is thought of as an array of houses, each inhabited by a voter holding one of these opinions at any given instant of time. The voter model is a process $(\xi_t)_{t\in[0,\infty)}$ on $S^{\mathbb{Z}^d}$, in which $\xi_t(x) \in S$ represents the opinion of the resident of house x at time t. The opinion of the voter in house $x \in \mathbb{Z}^d$ changes in time, in such a way that the flip rate to opinion $a \in S$ is

$$c_a^{(x)}(\eta) := \#\left\{y \in \mathbb{Z}^d \;\middle|\; \|y-x\| = 1,\ \eta(y) = a\right\},$$

whenever $\eta(x) \neq a$. In other words, the rate at which the resident of house x changes her opinion to a is proportional to the number of her neighbors who currently hold the opinion a.

It is rather obvious that the consensus, where everybody holds the same opinion $a \in S$, will be a stationary for the voter model, because the old opinions only ever flip with positive rates to new opinions that are the opinions of some neighbors. Formally, if $a \in S$ is an opinion and $\bar{a} \in S^{\mathbb{Z}^d}$ denotes the consensus configuration $\bar{a}(x) = a$ for all $x \in \mathbb{Z}$, then the delta measure $\delta_{\bar{a}}$ on $S^{\mathbb{Z}^d}$ (all probability mass on the consensus configuration) is a "trivial" stationary distribution for the process $(\xi_t)_{t\in[0,\infty)}$. We will show in the next lecture that if $d \leq 2$, then the only stationary distributions for the voter model are the (trivial) consensus measures $\delta_{\bar{a}}$, $a \in S$, and their convex combinations, whereas if $d > 2$, there are also nontrivial stationary distributions in which different opinions can coexist.

Example XI.3 (Contact process)**.**
The contact process is a model for the spread of a plant population (or infectuous disease). We fix two parameters: a "birth rate" $\lambda > 0$ and a "death rate" $\delta > 0$, and we set $S = \{0,1\}$. The contact process $(\xi_t)_{t\in[0,\infty)}$ is a process with state space $\{0,1\}^{\mathbb{Z}^d}$, in which we interpret that there is a living plant at site x at time t if $\xi_t(x) = 1$, otherwise not. Living plants die with rate $\delta > 0$ in the sense that the flip rate to state 0 at x is

$$c_0^{(x)}(\eta) := \delta,$$

whenever $\eta(x) \neq 0$. A new living plant at x is born according to the flip rate to state 1 at x given by

$$\begin{aligned} c_1^{(x)}(\eta) &:= \lambda \sum_{\substack{y\in\mathbb{Z}^d\\ \|y-x\|=1}} \eta(y) \\ &= \lambda \times \#\left\{y \in \mathbb{Z}^d \;\middle|\; \|y-x\| = 1,\ \eta(y) = 1\right\}, \end{aligned}$$

whenever $\eta(x) \neq 1$, i.e., the rate is proportional to the number of neighboring sites that currently host a living plant.

FIGURE XI.1. Illustration of the one-dimensional voter model: the horizontal axis represents "space" $\mathbb{Z}$ and the vertical axis "time" $[0,\infty)$ — colors in the figure represent the different opinions held by the voters at the different spatial and temporal locations.

Again, a situation with no living plants anywhere constitutes a "trivial" stationary distribution. Formally, this is represented by the delta measure $\delta_{\bar{0}}$ on the constant configuration $\bar{0}$ of zeroes. It can be shown that if the ratio λ/δ of births to deaths is small, then there are no other stationary distributions except this trivial one $\delta_{\bar{0}}$, whereas if λ/δ is large enough, then there is a nontrivial stationary distribution (which, moreover, has translation invariance in space $\mathbb{Z}^d$).

1. Construction of interacting particle systems

In this section we present a construction of interacting particle systems in a generality that includes Examples XI.1, XI.2, and XI.3. The construction is analogous to a construction of finite state space Markov processes from Poisson processes, which we review first as a more pedestrian example.

Poisson process

Let

$$\tau_1, \tau_2, \ldots \sim \mathrm{Exp}(\lambda) \qquad \text{i.i.d.}$$

be a sequence of independent random variables following exponential distributions with parameter $\lambda > 0$, i.e.,

$$\mathsf{P}[\tau_n > t] = e^{-\lambda t} \qquad \text{for } t \geq 0.$$

The random variables $\tau_1, \tau_2, \ldots$ are interpreted as waiting times between arrivals. For $n \in \mathbb{N}$, set

$$T_n = \tau_1 + \cdots + \tau_n,$$

interpreted as the time of n:th arrival. This gives rise to the arrival times $(T_n)_{n\in\mathbb{N}}$ of a Poisson process with intensity λ. The corresponding counting process $(N_t)_{t\in[0,\infty)}$ is defined by

$$N_t := \sup\left\{ n \in \mathbb{N} \;\middle|\; T_n \leq t \right\}.$$

A key property of the Poisson process is the following memoryless renewal property: for any $s \geq 0$ the increment process $(N_{s+t} - N_s)_{t\in[0,\infty)}$ is also a Poisson process with the same intensity λ, which is independent of the process $(N_u)_{u\in[0,t)}$ of earlier arrivals.[4] For the construction below, it is also useful to note that the Poisson process has the following thinning property: if we keep each arrival time with probability $p \in [0, 1]$ independently, and disregard the rest, then the arrival times that are kept form a Poisson process with intensity $p\lambda$.

Prelude: Construction of Markov jump processes on finite state spaces

Let $\mathcal{S}$ be a given finite set, and $\lambda(a, b) \geq 0$ given jump rates for $a, b \in \mathcal{S}$, $a \neq b$. We will construct a continuous time Markov process $X = (X_t)_{t\in[0,\infty)}$ on $\mathcal{S}$ with these jump rates. Note that we have

$$\max_{a,b\in\mathcal{S}} \lambda(a, b) < +\infty \qquad \text{(because of finiteness of } \mathcal{S}).$$

By performing a linear time change to decrease the jump rates if necessary, we can without loss of generality assume that

$$\lambda(a, b) \leq 1 \qquad \text{for all } a, b \in \mathcal{S}.$$

The construction uses two inputs of randomness. First, for each state $b \in \mathcal{S}$ we take an independent Poisson process with unit intensity $\lambda = 1$, and we denote its arrival times by $(T_n^b)_{n\in\mathbb{N}}$. For each state $b \in \mathcal{S}$ we also take a sequence $(U_n^b)_{n\in\mathbb{N}}$ of independent uniformly distributed random numbers on the unit interval $[0, 1]$, independently of each other and independently of all of the Poisson processes.

Given an initial state $x_0 \in \mathcal{S}$, the process is constructed as follows. Note first that on any bounded time interval there are almost surely only finitely many arrival times T_n^b of any of the Poisson processes, and all arrival times are distinct. A right-continuous process $X = (X_t)_{t\in[0,\infty)}$ which is constant in between the arrival times therefore becomes well defined if we specify its values at all arrival times and at time $t = 0$. We achieve this inductively as follows:

- Set $X_0 = x_0$ according to the initial condition.
- Once X_t has been defined for all times $t < T_n^b$, if the left limit of the process[5] at time T_n^b is

$$\lim_{t\uparrow T_n^b} X_t = a,$$

then we set

$$X_{T_n^b} = \begin{cases} b & \text{if } U_n^b \leq \lambda(a, b) \\ a & \text{if } U_n^b > \lambda(a, b). \end{cases}$$

Exercise XI.1 (The construction yields the desired Markov jump process)**.**
Show that the process $X = (X_t)_{t\in[0,\infty)}$ constructed as above satisfies the property (XI.1).

[4] In the terminology of Lecture VIII, the Poisson process $(N_t)_{t\in[0,\infty)}$ has stationary and independent increments.

[5] The left limit is simply the value of the process at the previous arrival time.

Construction of interacting particle system

We now mimic the construction of finite state space Markov jump processes to construct interacting particle systems with given flip rates.

Assumptions

We assume the flip rates to be translation invariant, and to depend on only on the states of finitely many other nearby components (as in Examples XI.1–XI.3 above), in the following precise sense. Let us denote by

$$\mathscr{N} = \{z_1, \dots, z_k\} \subset \mathbb{Z}^d$$

the set of sites which are allowed to affect the flip rates of the origin $0 \in \mathbb{Z}^d$. The flip rate of $x \in \mathbb{Z}^d$ to state $a \in S$ in configuration $\eta \in S^{\mathbb{Z}^d}$ such that $\eta(x) \neq a$ is now assumed to take the form

$$c_a^{(x)}(\eta) = g_a\big(\eta(x+z_1), \dots, \eta(x+z_k)\big),$$

where $g_a \colon S^n \to [0,\infty)$, for $a \in S$, are given functions. Note again that we have

$$\max_{a,s_1,\dots,s_k \in S} g_a(s_1, \dots, s_k) < \infty,$$

because there are only finitely many possibilities for $a, s_1, \dots, s_k$. By performing linear time change we may assume that the flip rates satisfy

$$c_a^{(x)}(\eta) \leq 1 \qquad \text{for all } x \in \mathbb{Z}^d,\ a \in S, \text{ and } \eta.$$

Attempt to construct the process

As inputs to our construction, we assume that for each pair $(x,a) \in \mathbb{Z}^d \times S$, we are given a unit intensity Poisson process with arrival times $(T_n^{x;a})_{n\in\mathbb{N}}$, and for each triple $(x,a,n) \in \mathbb{Z}^d \times S \times \mathbb{N}$, we are given a random variable $U_n^{x;a}$ uniformly distributed on the unit interval $[0,1]$, in such a way that the Poisson processes and the uniform random variables are all independent.

Let an initial configuration $\eta \in S^{\mathbb{Z}^d}$ be given. We seek to construct a process $\xi = (\xi_t)_{t\in[0,\infty)}$ such that for every $x \in \mathbb{Z}^d$, the function $t \mapsto \xi_t(x)$ describing the state of the single site x is right-continuous and has only finitely many changes over any bounded time interval. In particular, also the left limits $\xi_{t-}(x) := \lim_{s\uparrow t} \xi_s(x)$ will exist, since the state of the site will be constant on a interval of the form $[t-\varepsilon, t)$, for some $\varepsilon > 0$. We claim that such a process $\xi = (\xi_t)_{t\in[0,\infty)}$ can be constructed by the following rules:

- At time $t = 0$, we set $\xi_0 = \eta$ according to the given initial configuration.
- At time $t = T_n^{x;a}$ the state of site $x \in \mathbb{Z}$ is either changed to a or kept unchanged according to whether or not $U_n^{x;a} \leq c_a^{(x)}(\xi_{t-})$, i.e.,

$$\xi_t(x) = \begin{cases} a & \text{if } U_n^{x;a} \leq g_a\big(\xi_{t-}(x+z_1)), \dots, \xi_{t-}(x+z_k))\big) \\ \xi_{t-}(x) & \text{if } U_n^{x;a} > g_a\big(\xi_{t-}(x+z_1)), \dots, \xi_{t-}(x+z_k))\big). \end{cases}$$

Remark XI.4 (Why is well-definedness of the construction not obvious?)**.**
Note that in the present situation we can not simply define the changes inductively over the set of arrival times like we did for finite state space Markov jump processes. There are now infinitely many unit intensity Poisson processes $(T_n^{x;a})_{n\in\mathbb{N}}$, so there is no first arrival

time: on an arbitrarily small time interval there are already almost surely infinitely many arrivals. Consequenty, there is generally no first site x whose state should be updated. The most serious difficulty caused by this is an inherent chicken and egg problem: we need to know the flip rates to know how the states of the sites should be updated, but the states themselves affect the flip rates.

Fortunately, it turns out that the finite range of dependence and boundedness of the flip rates are just enough to ensure that we can break free from the vicious cycle.

Percolation of causal relations

Fix a small time duration $t_0 > 0$ to consider. In order to understand which sites can have a causal effect on others during the small time interval $[0, t_0]$, consider the following percolation model on $\mathbb{Z}^d$. Recall that we denote by $\mathscr{N} = \{z_1, \dots, z_k\} \subset \mathbb{Z}^d$ the finite set of sites which affect the flip rates of the state at the origin $0 \in \mathbb{Z}^d$. Let us denote by

$$\mathscr{N}^* := \{z_1, \dots, z_k, -z_1, \dots, -z_k\} \subset \mathbb{Z}^d$$

the symmetrized set, consisting of both those sites that affect the flip rates for the origin, as well as those sites whose flip rates are affected by the origin. This serves to capture all possible causal effects. Now with $t_0 > 0$ fixed and the Poisson processes $(T_n^{x;a})_{n\in\mathbb{N}}$ for all $(x, a) \in \mathbb{Z}^d \times S$ given, let us declare a "bond" $\{x, y\}$ between sites $x, y \in \mathbb{Z}^d$ such that $y - x \in \mathscr{N}^*$ to be open ($\omega_{\{x,y\}} = 1$) if one of the Poisson processes associated to either x or y has an arrival on the time interval $[0, t_0]$ and closed ($\omega_{\{x,y\}} = 0$) otherwise, i.e.,

$$\omega_{\{x,y\}} = \begin{cases} 1 & \text{if } T_1^{x;a} \le t_0 \text{ or } T_1^{y;a} \le t_0 \text{ for some } a \in S \\ 0 & \text{if } T_1^{x;a} > t_0 \text{ and } T_1^{y;a} > t_0 \text{ for all } a \in S. \end{cases}$$

This is a percolation model on $\mathbb{Z}^d$, but note that the different bonds are not independent of each other! As usual for percolation, we define two sites $x, y \in \mathbb{Z}^d$ to be connected if there exists a finite sequence $x = x_0, x_1, \dots, x_{\ell-1}, x_\ell = y \in \mathbb{Z}^d$ of distinct sites such that the bonds between consecutive sites of the path are open, i.e., $\omega_{\{x_{j-1},x_j\}} = 1$ for all $j = 1, \dots, \ell$. We denote this by $x \leftrightsquigarrow y$, and we thus obtain a (random) equivalence relation $\leftrightsquigarrow$ on $\mathbb{Z}^d$, whose equivalence classes are called the connected components.

Proposition XI.5 (Non-percolation of causal relations for small times)**.** *There exists a $t_0 > 0$ such that almost surely all connected components of the above percolation model are finite.*

Proof. Let us start with two observations regarding openness of bonds. First, for any bond $\{x_1, x_2\}$ such that $x_2 - x_1 \in \mathscr{N}^*$, the probability of the bond being open is

$$\begin{aligned} \mathsf{P}\big[\{x_1, x_2\} \text{ is open}\big] &= 1 - \mathsf{P}\big[\{x_1, x_2\} \text{ is not open}\big] \\ &= 1 - \mathsf{P}\Big[T_1^{x_1;a} > t_0 \text{ and } T_1^{x_2;a} > t_0 \text{ for all } a \in S\Big] \\ &= 1 - (e^{-t_0})^{2\#S} \\ &= 1 - e^{-2t_0\#S}. \end{aligned}$$

Moreover, whenever two bonds $\{x_1, x_2\}$ and $\{x_3, x_4\}$ do not share common endpoint sites (i.e., when $\{x_1, x_2\} \cap \{x_3, x_4\} = \emptyset$), then the openness states of the bonds are independent, because they are determined by independent Poisson processes.

The main task is to show that the connected component $C_x \subset \mathbb{Z}^d$ of any given site $x \in \mathbb{Z}^d$ is almost surely finite, $\mathsf{P}[\#C_x < \infty] = 1$. Indeed, it will then follow by the union bound that

$$\begin{aligned}
&\mathsf{P}\big[\text{all connected components are finite}\big] \\
&= 1 - \mathsf{P}\big[\text{for some } x \in \mathbb{Z}^d \text{ the connected component } C_x \text{ is infinite}\big] \\
&\geq 1 - \sum_{x \in \mathbb{Z}^d} \underbrace{\mathsf{P}[\text{the component } C_x \text{ is infinite}]}_{= 1 - \mathsf{P}[\#C_x < \infty]} \\
&= 1 - \sum_{x \in \mathbb{Z}^d} 0 \;=\; 1,
\end{aligned}$$

i.e., almost surely all components are finite.

So fix $x \in \mathbb{Z}^d$. If the component C_x is infinite, then for any $\ell \in \mathbb{N}$ we can find an open path of length ℓ starting from x, i.e., a sequence of distinct sites $x = x_0, x_1, \ldots, x_\ell$ such that $\{x_{j-1}, x_j\}$ is open for all $j = 1, \ldots, \ell$. The openness states of the bonds $\{x_0, x_1\}$, $\{x_1, x_2\}$, $\{x_2, x_3\}$, …, are not all independent, but if we only consider every other bond, $\{x_0, x_1\}$, $\{x_2, x_3\}$, $\{x_4, x_5\}$, …, then they use distinct sites and the openness states are independent. Consequently, we find the upper bound

$$\begin{aligned}
\mathsf{P}\big[\text{the path } x_0, x_1, \ldots, x_\ell \text{ is open}\big] &\leq \mathsf{P}\big[\{x_0, x_1\}, \{x_2, x_3\}, \ldots \text{ are open}\big] \\
&= \big(1 - e^{-2t_0 \# S}\big)^{\lceil \ell/2 \rceil}.
\end{aligned}$$

Recall also that we have assumed $\#\mathscr{N} = k$, which implies $\#\mathscr{N}^* \leq 2k$, so the number of different paths of length ℓ starting from x is at most $(2k)^\ell$. Monotonicity and the union bound now give

$$\begin{aligned}
\mathsf{P}\big[\#C_x = \infty\big] &\leq \mathsf{P}\big[\text{there exists an open path of length } \ell\big] \\
&\leq (2k)^\ell \left(1 - e^{-2t_0 \# S}\right)^{\lceil \ell/2 \rceil}.
\end{aligned}$$

Since $e^{-2t_0 \# S} \to 1$ as $t_0 \to 0$, we can choose a small $t_0 > 0$ such that $(1 - e^{-2t_0 \# S})^{1/2} < \frac{1}{2k}$. With such a choice, the right hand side bound above tends to zero as $\ell \to \infty$. This implies that $\mathsf{P}[\#C_x = \infty] = 0$, i.e., that the connected component of x is almost surely finite, and the proof is complete. □

Well-definedness of the process

It follows from the non-percolation of causal relations that the process can be defined, first at least for small times.

Theorem XI.6 (Well-definedness of the interacting particle system process)**.** *Almost surely the prescription above gives rise to a well defined process* $\xi = (\xi_t)_{t \in [0,\infty)}$ *on* $S^{\mathbb{Z}^d}$.

Proof. Fix $t_0 > 0$ small enough so that the non-percolation conclusion of Proposition XI.5 holds. Then (almost surely) all components of the percolation of causal relations are finite, and based on that, an unambiguous definition of the process ξ at least on the time interval $[0, t_0]$ is possible. Indeed, in order to determine $\xi_t(x)$ for any $t \in [0, t_0]$ and $x \in \mathbb{Z}^d$, it is sufficient to consider the Poisson processes $(T_n^{y;a})_{n \in \mathbb{N}}$ corresponding to the finitely many sites $y \in C_x$ in the connected component $C_x \subset \mathbb{Z}^d$ of x, and the finitely many states $a \in S$. These finitely many processes have (almost surely) only finitely many arrival times on the time interval, and these arrival times are (almost surely) distinct. Therefore we may determine all $\xi_t(y)$, for $y \in C_x$ and $t \in [0, t_0]$ inductively over the ordered set of arrival times (exactly like in the case of finite state space Markov process construction). Since this can be done for any $x \in \mathbb{Z}^d$ and all $t \in [0, t_0]$, the entire (random) configurations $\xi_t \in S^{\mathbb{Z}^d}$ become well-defined for all $t \in [0, t_0]$.

Once the process has been defined on this time interval $[0, t_0]$ of fixed positive duration, we may use the memoryless renewal property of the Poisson processes and observe that

the (random) configuration ξ_{t_0} can be used as a new initial condition, and we can argue by an exactly similar non-percolation of causal relations that the prescription is (almost surely) well-defined also on the time interval $[t_0, 2t_0]$. Continuing inductively, we define the process on time intervals of the form $[mt_0, (m+1)t_0]$, $m \in \mathbb{N}$, and thus the entire process $\xi = (\xi_t)_{t\in[0,\infty)}$ becomes defined. □

2. Markov semigroups of interacting particle systems

Let us denote by $\mathcal{C}(S^{\mathbb{Z}^d})$ the space of continuous functions

$$f\colon S^{\mathbb{Z}^d} \to \mathbb{R}.$$

Note that since $S^{\mathbb{Z}^d}$ is compact, a function $f \in \mathcal{C}(S^{\mathbb{Z}^d})$ is automatically bounded,

$$\|f\|_\infty := \sup_{\eta\in S^{\mathbb{Z}^d}} |f(\eta)| < +\infty,$$

and the space $\mathcal{C}(S^{\mathbb{Z}^d})$ inherits a metric from this norm via $\varrho(f,g) = \|f-g\|_\infty$.

For a given $t \geq 0$ and for any $f \in \mathcal{C}(S^{\mathbb{Z}^d})$, define a new function

$$T_t f\colon S^{\mathbb{Z}^d} \to \mathbb{R}$$

by the formula

$$(T_t f)(\eta) := \mathsf{E}_{\xi_0=\eta}\big[f(\xi_t)\big], \tag{XI.3}$$

where the expected value is over the process constructed above, using $\xi_0 = \eta$ as the given initial state.

Theorem XI.7 (The semigroup of an interacting particle system)**.**
For any continuous $f\colon S^{\mathbb{Z}^d} \to \mathbb{R}$ and any $t \geq 0$ we have the following:

$$T_t f\colon S^{\mathbb{Z}^d} \to \mathbb{R} \text{ is continuous,}$$
$$\text{and} \qquad \|T_t f\|_\infty \leq \|f\|_\infty.$$

In other words, for any fixed $t \geq 0$,

$$T_t\colon \mathcal{C}(S^{\mathbb{Z}^d}) \to \mathcal{C}(S^{\mathbb{Z}^d})$$

is a bounded linear operator on the Banach space $\mathcal{C}(S^{\mathbb{Z}^d})$ of continuous functions on $S^{\mathbb{Z}^d}$. Moreover, the operators $(T_t)_{t\geq 0}$ form a semigroup: we have

$$T_t \circ T_s = T_{t+s} \qquad \text{for all } s,t \geq 0.$$

Proof. Let $f \in \mathcal{C}(S^{\mathbb{Z}^d})$, and consider the process $(\xi_t)_{t\geq 0}$ started from the initial condition $\xi_0 = \eta$. The random variable $f(\xi_t)$ obviously satisfies the bound $|f(\xi_t)| \leq \|f\|_\infty$, so it follows from the definition (XI.3) and triangle inequality for expected values that

$$\big|(T_t f)(\eta)\big| = \Big|\mathsf{E}\big[f(\xi_t)\big]\Big| \leq \mathsf{E}\Big[\big|f(\xi_t)\big|\Big] \leq \|f\|_\infty.$$

Since this holds for any $\eta \in S^{\mathbb{Z}^d}$, we get $\|T_t f\|_\infty \leq \|f\|_\infty$.

To prove continuity of $T_t f\colon S^{\mathbb{Z}^d} \to \mathbb{R}$, consider first $t \in [0, t_0]$, where $t_0 > 0$ is small enough so that the non-percolation conclusion of Proposition XI.5 holds. Let $\eta^{(1)}, \eta^{(2)}, \ldots \in S^{\mathbb{Z}^d}$ be a sequence of configurations such that $\eta^{(n)} \to \eta$ in the topology of $S^{\mathbb{Z}^d}$. For continuity, we must then show that $(T_t f)(\eta^{(n)}) \to (T_t f)(\eta)$. Let us fix the Poisson processes $(T_n^{x;a})_{n\in\mathbb{N}}$ and random variables $U_n^{x;a}$ used in the construction of the processes. Then let $\xi^{(n)} = (\xi_t^{(n)})_{t\geq 0}$,

for $n \in \mathbb{N}$, and $\xi = (\xi_t)_{t\geq 0}$ be the processes constructed using the same Poisson processes and random variables, but using the different initial configurations $\xi_0^{(n)} = \eta^{(n)}$, and $\xi_0 = \eta$, respectively. Recall that for a fixed $x \in \mathbb{Z}^d$, the connected component $C_x \subset \mathbb{Z}^d$ of x is finite (almost surely). In view of the topology of $S^{\mathbb{Z}^d}$ (see see Appendix H.4) and assumed convergence $\eta^{(n)} \to \eta$, we can therefore find an N_0 such that for all $n \geq N_0$ we have $\eta^{(n)}(x') = \eta(x')$ for all $x' \in C_x$. By the construction of the processes, the state $\xi_t(x)$ (respectively $\xi_t^{(n)}(x)$) is only affected by the states of the sites $x' \in C_x$ in the initial condition η (respectively $\eta^{(n)}$). For $n \geq N_0$, the coincidence $\eta^{(n)}(x') = \eta(x')$ for all $x' \in C_x$ then implies that $\xi_t^{(n)}(x) = \xi_t(x)$. We have thus shown that (almost surely) for any $x \in \mathbb{Z}^d$ there exists an N_0 such that $\xi_t^{(n)}(x) = \xi_t(x)$ for $n \geq N_0$, and in view of the topology on $S^{\mathbb{Z}^d}$, this implies that

$$\xi_t^{(n)} \to \xi_t \qquad \text{as } n \to \infty \text{ (almost surely).}$$

By the assumed continuity of f, this immediately gives also

$$f(\xi_t^{(n)}) \to f(\xi_t) \qquad \text{as } n \to \infty \text{ (almost surely).}$$

Since f is bounded, the random variables above are bounded, so by the dominated convergence theorem (or bounded convergence theorem) we conclude that

$$(T_t f)(\eta^{(n)}) := \mathsf{E}\big[f(\xi_t^{(n)})\big] \underset{n\to\infty}{\longrightarrow} \mathsf{E}\big[f(\xi_t)\big] = (T_t f)(\eta),$$

which proves the continuity of $T_t f$ when $t \in [0, t_0]$.

Let us then prove the semigroup property. For given $s, t \geq 0$, using the memoryless renewal property of the Poisson processes, it is clear that the process $\widetilde{\xi} = (\widetilde{\xi}_t)_{t\geq 0}$ defined by

$$\widetilde{\xi}_t := \xi_{s+t}$$

is a process constructed by the same procedure starting from the initial condition $\widetilde{\xi}_0 = \xi_s$ and using the renewed Poisson processes and remaining uniform random variables, which are independent of those used to construct $(\xi_u)_{u\in[0,s]}$. Using Fubini's theorem to integrate over the corresponding product measure in two stages, we find

$$\begin{aligned}(T_{s+t}f)(\eta) &= \mathsf{E}_{\xi_0=\eta}\big[f(\xi_{s+t})\big] \\ &= \mathsf{E}_{\xi_0=\eta}\Big[\mathsf{E}_{\widetilde{\xi}_0=\xi_s}\big[f(\widetilde{\xi}_t)\big]\Big] \\ &= \mathsf{E}_{\xi_0=\eta}\Big[(T_t f)(\xi_s)\Big] \;=\; \big(T_s(T_t f)\big)(\eta).\end{aligned}$$

This shows the semigroup property $T_{s+t} = T_s \circ T_t$. From this, it also follows that $T_t f$ is continuous for an arbitrary $t \geq 0$, since we can write T_t as a finite composition

$$T_t = \underbrace{T_{t_0} \circ \cdots \circ T_{t_0}}_{n \text{ times}} \circ T_{t'}, \qquad \text{where } n = \lfloor t/t_0 \rfloor \text{ and } t' = t - nt_0 < t_0,$$

and it was already proven that T_t preserves continuity for $t \leq t_0$. □

A Markov process for which the operators T_t, $t \geq 0$, given by (XI.3) map continuous functions to continuous functions is called Feller processes. The theorem above thus in particular says that interacting particle systems are Feller processes.

The dual action of the semigroup

Note already that T_t contains all information about the distribution of the random state ξ_t of the process at time t (in fact for any possible initial configuration).

Proposition XI.8 (The laws are determined by the semigroup operators)**.**
For any $t \geq 0$, the law of ξ_t is uniquely determined by the operator T_t.

Proof. The law of ξ_t is a Borel probability measure on the metric space $S^{\mathbb{Z}^d}$. According to Exercise H.8 the knowledge of expected values of all continuous bounded functions f is sufficient to characterize a Borel probability measure. By definition (XI.3), these expected values are exactly $(T_t f)(\eta)$. □

Exercise XI.2 (The semigroup determines the law of the process)**.**
Show that the semigroup $(T_t)_{t\geq 0}$ determines all finite dimensional distributions of the process, i.e., the joint laws of $(\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_m})$ for any $0 \leq t_1 < t_2 < \cdots < t_m$.

We can view the dual action of the semigroup as describing the evolution of the laws of the random configurations process.[6] Specifically, for $t \geq 0$ and a Borel probability measure ν on $S^{\mathbb{Z}^d}$, we can define another Borel probability measure $T_t^*\nu$ corresponding to the law of ξ_t started from a random initial condition ξ_0 sampled according to ν. By construction, it is then clear that $T_t^*\nu$ is characterized by the condition

$$\int_{S^{\mathbb{Z}^d}} f(\eta)\, \mathrm{d}(T_t^*\nu)(\eta) = \int_{S^{\mathbb{Z}^d}} (T_t f)(\eta)\, \mathrm{d}\nu(\eta). \tag{XI.4}$$

From the semigroup property $T_{s+t} = T_s \circ T_t$ it follows that

$$\begin{aligned}
\int_{S^{\mathbb{Z}^d}} f(\eta)\, \mathrm{d}(T_{s+t}^*\nu)(\eta) &= \int_{S^{\mathbb{Z}^d}} (T_{s+t} f)(\eta)\, \mathrm{d}\nu(\eta) \\
&= \int_{S^{\mathbb{Z}^d}} \big(T_s(T_t f)\big)(\eta)\, \mathrm{d}\nu(\eta) \\
&= \int_{S^{\mathbb{Z}^d}} (T_t f)(\eta)\, \mathrm{d}(T_s^*\nu) \\
&= \int_{S^{\mathbb{Z}^d}} f(\eta)\, \mathrm{d}\big(T_t^*(T_s^*\nu)\big)(\eta),
\end{aligned}$$

which shows the "dual semigroup property" $T_{s+t}^*\nu = T_t^*(T_s^*\nu)$.

Infinitesimal generators

The semigroup $(T_t)_{t\geq 0}$ can often be reconstructed from its infinitesimal generator G defined by

$$(Gf)(\eta) := \lim_{t\downarrow 0} \frac{(T_t f)(\eta) - f(\eta)}{t} = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathsf{E}_{\xi_0=\eta}\Big[f(\xi_t)\Big]$$

for functions f in a suitable dense subspace of $\mathcal{C}(S^{\mathbb{Z}^d})$, for which an expression can be written down in terms of the flip rates $c_a^{(x)}(\eta)$. For the reader interested in details of this approach, we suggest, e.g., [**Lig05**].

To illustrate the idea of infinitesimal generators, consider the following exercise about the semigroup of the standard Brownian motion.

Exercise XI.3 (Infinitesimal generator of Brownian motion)**.**
Consider the standard Brownian motion $B = (B_t)_{t\in[0,\infty)}$. The Brownian motion started

[6] One can formalize this a little bit more by noticing that Riesz-Markov-Kakutani representation theorem: the dual of the Banach space $\mathcal{C}(S^{\mathbb{Z}^d})$ is the space of regular signed Borel measures, and the adjoint of the bounded linear operator $T_t\colon \mathcal{C}(S^{\mathbb{Z}^d}) \to \mathcal{C}(S^{\mathbb{Z}^d})$ is an operator T_t^* on such measures.

from $x \in \mathbb{R}$ is the process defined by $B_t^{(x)} = x + B_t$. The generator G of the Brownian motion is the following operator. For a smooth and compactly supported function $f\colon \mathbb{R} \to \mathbb{R}$, set

$$Gf(x) = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} \mathsf{E}\Big[f(B_t^{(x)})\Big].$$

Show that

$$Gf(x) = \frac{1}{2} f''(x).$$

Hint: *Recall the distribution of B_t for a given $t > 0$. Perform a Taylor expansion of f at x to the second order. Control the error terms when this Taylor approximation is used in the defining formula of $Gf(x)$.*

3. Another example: totally asymmetric simple exclusion process

The *totally asymmetric simple exclusion process* (TASEP) is a paradigmatically important model for non-equilibrium phenomena statistical physics. Philosophically it clearly belongs to the category of interacting particle systems, although it fails to literally satisfy the definition given earlier in this chapter. Below we introduce the TASEP and cover a few of its basic properties in the form of exercises. The exercises also feature techniques entirely parallel to the constructions given earlier in this chapter for general interacting particle systems.

Informally, TASEP can be described as follows. The underlying space is discrete and one-dimensional, modeled by a (finite or infinite) row of sites. Each site can be either vacant or occupied by a particle at any given time. After exponentially distributed independent waiting times, each particle tries to jump to the next site in the row.[7] This jump takes place if the next site is currently vacant, but the jump is prevented if the next site is occupied.[8]

In the following exercise we study TASEP in the case of N particles on a finite periodic row of L sites, see Figure XI.2 for illustration.

Exercise XI.4 (TASEP on finite directed cycle)**.**

Fix parameters $L, N \in \mathbb{N}$ with $N \leq L$. Let $\mathcal{G}_L = (\mathrm{V}_L, \vec{\mathrm{E}}_L)$ be the directed cycle graph with the set of sites

$$\mathrm{V}_L = \{1, \dots, L\}$$

and the set of directed links

$$\vec{\mathrm{E}}_L = \big\{(x,y) \,\big|\, x, y \in \mathrm{V}_L,\ y - x \equiv 1 \pmod{L}\big\}.$$

We denote $x \curvearrowright y$, if $(x,y) \in \vec{\mathrm{E}}_L$. The totally asymmetric simple exclusion process (TASEP) on $\mathcal{G}_L$ with N particles and activation rate $v > 0$ is a continuous time Markov process $X = (X_t)_{t \in [0,\infty)}$ with

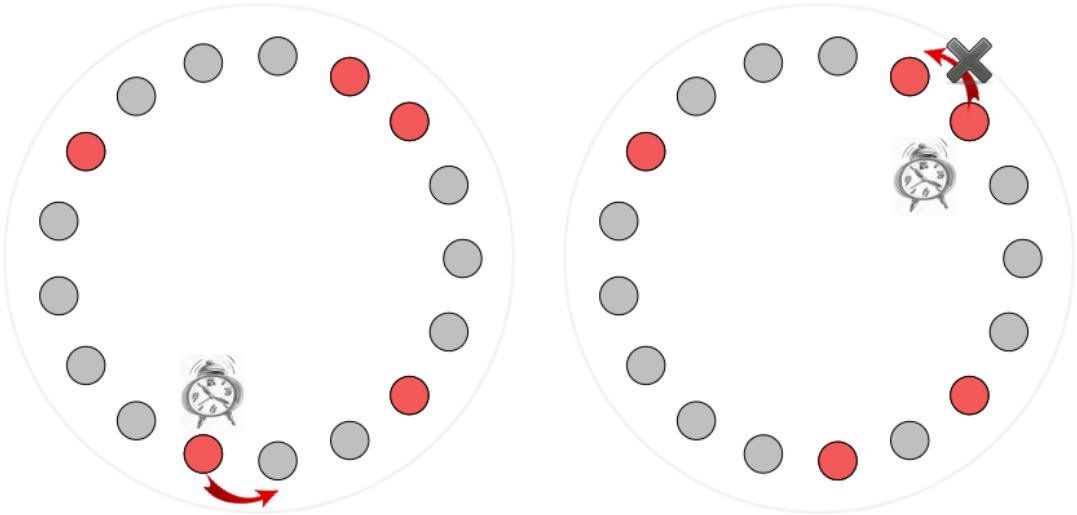
$$\text{state space} \qquad \mathcal{S}_L^{(N)} = \Big\{ Y \subset \mathrm{V}_L \;\Big|\; \#Y = N \Big\}$$

$$\text{and jump rates} \qquad \lambda(Y, Y') = \begin{cases} v & \text{if } Y' = (Y \setminus \{x\}) \cup \{y\} \text{ for some } x \curvearrowright y \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that the uniform distribution $\mu_{\text{unif.}}$ on $\mathcal{S}_L^{(N)}$ is the unique stationary distribution for the process $(X_t)_{t \geq 0}$.

[7] *Total asymmetry* refers to the chosen direction: only jumps forward to the next site in the row are attempted, but there are no backwards jumpts to the previous site.

[8] *Exclusion* refers to the prevention of jumps to already occupied sites.



(a) A particle can jump ahead if the site in front of it is vacant.

(b) If the site in front of a particle is occupied, no jump is allowed.

FIGURE XI.2. Illustration of the totally asymmetric simple exclusion process with $N = 5$ particles on a periodic directed graph with $L = 17$ sites.

(b) Define the average speed s in the stationary distribution as

$$s = \lim_{\varepsilon \downarrow 0} \frac{\mu_{\text{unif.}}\big[X_\varepsilon \neq X_0\big]}{\varepsilon\, N}.$$

Calculate s.

(c) Calculate s in the limit as $L \to \infty$, $\frac{N}{L} \to \rho \in (0,1)$. What is the optimal value of the density ρ for maximum speed s? What is the optimal value of ρ for maximum traffic flow $s \times N$ (optimality asymptotically as $L \to \infty$)?

The next two exercises concern the totally asymmetric simple exclusion process (TASEP) on $\mathbb{Z}$ is a process $\xi = (\xi_t)_{t \geq 0}$ which is constructed as follows. The state space $\mathcal{S}$ consists of all subsets $Y \subset \mathbb{Z}$, which we identify with $\mathcal{S} = \{0,1\}^{\mathbb{Z}}$ in such a way that Y corresponds to

$$\xi = \big(\xi(x)\big)_{x \in \mathbb{Z}} \qquad \text{with } \xi(x) = \begin{cases} 1 & \text{if } x \in Y \\ 0 & \text{if } x \notin Y. \end{cases}$$

Choose an initial configuration $\xi_0 = \big(\xi_0(x)\big)_{x \in \mathbb{Z}} \in \mathcal{S}$. For each $x \in \mathbb{Z}$, take an independent Poisson process with intensity $v > 0$, and denote its arrival times $(T_n^x)_{n \in \mathbb{N}}$. The rules to define $\xi_t \in \mathcal{S}$ for $t \geq 0$ are the following: for any $x \in \mathbb{Z}$

- $t \mapsto \xi_t(x)$ is continuous from the right, and constant on any time interval that does not contain any T_n^x or T_n^{x-1}
- at times $t = T_n^x$, $\xi_t(x)$ and $\xi_t(x+1)$ are determined in terms of the left limits:
  - if $\xi_{t-}(x) = 1$ and $\xi_{t-}(x+1) = 0$, then $\xi_t(x) = 0$ and $\xi_t(x+1) = 1$
  - otherwise, $\xi_t(x) = \xi_{t-}(x)$ and $\xi_t(x+1) = \xi_{t-}(x+1)$.

Exercise XI.5 (Well-definedness of TASEP on $\mathbb{Z}$)**.**
Show that the process $(\xi_t)_{t \geq 0}$ (the TASEP on $\mathbb{Z}$) becomes (almost surely) well defined by the rules given above.

Hint: *Show that (almost surely) for any $x \in \mathbb{Z}$ and $t \geq 0$ there are only finitely many Poisson process arrivals that could affect $\xi_t(x)$ according to the rules.*

Exercise XI.6 (Stationary distributions for TASEP on $\mathbb{Z}$)**.**
Let $\rho \in (0,1)$. Suppose that the initial state ξ_0 of the TASEP on $\mathbb{Z}$ is taken random and independent of the Poisson processes, so that its coordinates $\xi_0(x)$, $x \in \mathbb{Z}$, are independent and $\mathsf{P}\big[\xi_0(x) = 1\big] = \rho$ for each x. Show that for any $t \geq 0$ also the coordinates $\xi_t(x)$, $x \in \mathbb{Z}$, are independent and $\mathsf{P}\big[\xi_t(x) = 1\big] = \rho$.
Interpretation: *In other words, the product of Bernoulli measures $\mu = \bigotimes_{x\in\mathbb{Z}} \mathrm{Bernoulli}(\rho)$ is a stationary measure for the TASEP on $\mathbb{Z}$ — for any $\rho \in (0,1)$.*

Lecture XII

The voter model

This lecture is devoted to the voter model, introduced in Example XI.2 in the previous lecture. Recall that the voter model is an interacting particle system $\xi = (\xi_t)_{t\in[0,\infty)}$ with state space $S^{\mathbb{Z}^d}$, where S is some finite set of opinions, and with flip rates depending on the number of neighbors of a site $x \in \mathbb{Z}^d$ holding opinion $a \in S$,

$$n_a^{(x)}(\eta) = \#\left\{y \in \mathbb{Z}^d \;\middle|\; \|y-x\| = 1,\ \eta(y) = a\right\},$$

via

$$\lim_{\varepsilon\downarrow 0} \frac{1}{\varepsilon}\, \mathsf{P}\Big[\xi_{t+\varepsilon}(x) = a \;\Big|\; \xi_t = \eta,\ \mathfrak{H}\Big] = n_a^{(x)}(\eta)$$

when $\eta \in S^{\mathbb{Z}^d}$ is such that $\eta(x) \neq a$, and $\mathfrak{H}$ is an event that only depends on the history of the process up to time t.

We will prove the result announced in the previous lecture, which characterizes in which dimensions d there exists nontrivial stationary distributions:

- $d \leq 2$**:** The only stationary distributions of the voter model are convex combinations of trivial consensus measures.
- $d > 2$**:** The voter model has translation invariant stationary distributions, which are nontrivial, i.e., can not be written as convex combinations of consensus measures.

Besides the result itself, the technique used in its proof is worth knowing: we analyze the model itself via a dual process. In the case of the voter model, the dual process is coalescent random walks.

1. Voter model: alternative construction and dual process

Let the initial state $\xi_0 \in S^{\mathbb{Z}^d}$ be given. For each pair (x, y) with $x, y \in \mathbb{Z}^d$ such that $\|x - y\| = 1$, take an independent unit rate Poisson process with arrival times $(T_n^{x\to y})_{n\in\mathbb{N}}$. These are interpreted as the times at which the voter x asks and adopts the opinion of the voter y.[1] It is possible to state this more formally, and verify that the process becomes well defined, using a method analogous to the Harris' argument in the previous lecture. In the present case there is, however, a more instructive and even easier approach described below.

[1]Pay attention to the notation: we choose the arrow $x \to y$ to point to the direction of the voter y whose opinion is "sent" to the voter x, which is the opposite of the direction in which the opinion "travels". This choice of direction of arrows is natural for the dual process.

Fix a time $t > 0$ and a site $x \in \mathbb{Z}^d$. In order to determine what is the opinion $\xi_t(x)$ of x at time t, proceed as follows backwards in time from t to 0. Let

$$S_1 = \inf\left\{ s > 0 \;\middle|\; t - s \in \bigcup_{y,n} \{T_n^{x\to y}\} \right\}$$

be the amount of time needed to come backwards from t to get to the last time when x asked the opinion of one of its neighbors. Denote this neighbor by Y_1, so that $t - S_1 = T_n^{x\to Y_1}$ (for some n). The opinion $\xi_t(x)$ of x at time t must therefore be the same as the opinion $\xi_{t-S_1}(Y_1)$ of Y_1 at time $t - S_1$. Continue inductively, and define

$$S_{k+1} = \inf\left\{ s > S_k \;\middle|\; t - s \in \bigcup_{y,n} \left\{T_n^{Y_k\to y}\right\} \right\},$$

and define the site Y_{k+1} so that $t - S_{k+1} = T_n^{Y_k\to Y_{k+1}}$ (for some n). For some finite k' we will of course no longer find any such arrival times of the Poisson processes, and we naturally interpret $S_{k'} = \infty$ but do not attempt to define the corresponding $Y_{k'}$. For $k = 0$, on the other hand, it is natural to interpret $S_0 = 0$ and $Y_0 = x$. We can then collect these $S_0, S_1, \ldots, S_{k'-1}$ and $Y_0, Y_1, \ldots, Y_{k'-1}$ to fully describe the route along which the opinion has travelled: define $(D_s^{x;t})_{s\in[0,t]}$ by

$$D_s^{x;t} = Y_k \qquad \text{when } s \in [S_k, S_{k+1}).$$

The opinion $\xi_t(x)$ of x at time t is recovered via

$$\xi_t(x) = \xi_{t-s}(D_s^{x;t})$$

for any $s \in [0, t]$, and in particular we can trace back the source of the opinion all the way to the initial configuration ξ_0

$$\xi_t(x) = \xi_0(D_t^{x;t}).$$

We can do the same starting from any $x \in \mathbb{Z}^d$ and any $t > 0$, and our underlying collection of Poisson processes gives rise to processes $(D_s^{x;t})_{s\in[0,t]}$, $x \in \mathbb{Z}^d$, $t > 0$, from which the entire voter model configurations at any times can be recovered given the initial configuration ξ_0,

$$\xi_t = \big(\xi_t(x)\big)_{x\in\mathbb{Z}^d} = \big(\xi_0(D_t^{x;t})\big)_{x\in\mathbb{Z}^d}. \tag{XII.1}$$

From the above it is rather easy to verify that the construction gives rise to the correct flip rates. Let $\varepsilon > 0$ and consider the processes from time $t + \varepsilon$ backwards. Suppose that $\xi_t(x) \neq a$. Then we use $\xi_{t+\varepsilon}(x) = \xi_t(D_\varepsilon^{x;t+\varepsilon})$ together with

$$\mathsf{P}\big[D_\varepsilon^{x;t+\varepsilon} = y\big] = \varepsilon + o(\varepsilon) \qquad \text{for any } y \in \mathbb{Z}^d \text{ such that } \|y - x\| = 1,$$

where the leading term comes from having exactly one arrival of the Poisson process $(T_n^{x\to y})_{n\in\mathbb{N}}$ on the time interval $[t, t+\varepsilon]$. We leave the remaining details to the reader.

Let us now consider the laws of the processes $(D_s^{x;t})_{s\in[0,t]}$ in more detail. Fix $t > 0$, and consider first just one $x \in \mathbb{Z}^d$. By construction, $s \mapsto D_s^{x;t}$ is a right continuous process on $\mathbb{Z}^d$, started from x when $s = 0$, and jumps to nearest neighbors on the lattice $\mathbb{Z}^d$ at unit rates — the jump times are up to reversal of time direction the arrival times of the Poisson processes $(T_n^{z\to y})$ at the appropriate location z. Thus $(D_s^{x;t})_{s\in[0,t]}$ is the continuous time random walk on $\mathbb{Z}^d$ started from x. Consider then $x, x' \in \mathbb{Z}^d$ with $x \neq x'$. Both $(D_s^{x;t})_{s\in[0,t]}$ and $(D_s^{x';t})_{s\in[0,t]}$ are continuous time random walks, started from x and x', respectively. They are not independent,

however: once they meet, they follow the same jumps, determined by the Poisson processes associated to their their current, coinciding locations. Letting

$$S^{x,x'} = \inf\left\{ s > 0 \;\middle|\; D_s^{x;t} = D_s^{x';t} \right\},$$

we have that the beginning parts $(D_s^{x;t})_{s\in[0,S^{x,x'}]}$ and $(D_s^{x';t})_{s\in[0,S^{x,x'}]}$ are independent (their jumps are determined by different Poisson processes, or occasionally by the arrivals of the same Poisson process but on disjoint time intervals), whereas for $s \geq S^{x,x'}$ the two random walkers have coalesced, $D_s^{x;t} = D_s^{x';t}$. The joint distributions of $(D_s^{x';t})_{s\in[0,t]}$ for larger collections of initial points x are described similary — any two walkers coalesce upon meeting. Such a collection of random walkers is called coalescent random walks. They are easiest constructed in terms of independent Poisson processes for all ordered pairs of neighboring sites of the lattice, exactly as we did backwards in time. Clearly we could continue the process of coalescent random walks beyond the time interval $s \in [0,t]$. What we prefer to do is to take just one process of coalescent random walks $(D_s^x)_{s\in[0,\infty),\, x\in\mathbb{Z}^d}$, and give up the almost sure equality in Equation (XII.1) in favor of just the equality in distribution

$$\xi_t = \big(\xi_t(x)\big)_{x\in\mathbb{Z}^d} \stackrel{\mathrm{d}}{=} \big(\xi_0(D_t^x)\big)_{x\in\mathbb{Z}^d}. \tag{XII.2}$$

In this form the process of coalescent random walks $(D_s^x)_{s\in[0,\infty),\, x\in\mathbb{Z}^d}$ is called the dual process of the voter model.

Dual processes exist for a number of different stochastic processes, and they may take various forms. In favorable cases, they provide easier ways to calculate properties of the original process. This is the case for the voter model, in particular — our analysis of the stationary measures is based on the study of the coalescent random walks.

2. Only trivial stationary measures in low dimension

Let us begin with the analysis of the case of dimension $d \leq 2$.

The crucial step is the following clustering property.

Proposition XII.1 (Clustering for voter model in low dimensions)**.**
Let $d \leq 2$. Let $x, y \in \mathbb{Z}^d$ be two distinct sites. Then for the voter model $\xi = (\xi_t)_{t\in[0,\infty)}$, we have

$$\mathsf{P}\Big[\xi_t(x) \neq \xi_t(y)\Big] \underset{t\to\infty}{\longrightarrow} 0.$$

Proof. Using the duality (XII.2), we have

$$\mathsf{P}_{\xi_0}\Big[\xi_t(x) \neq \xi_t(y)\Big] \leq \mathsf{P}\Big[D_t^x \neq D_t^y\Big],$$

since if the opinions of x and y at time t originate from the same place in the initial configuration ξ_0, they are necessarily equal. Given that the random walks D^x and D^y coalesce if they meet, we get that

$$\mathsf{P}_{\xi_0}\Big[\xi_t(x) \neq \xi_t(y)\Big] \leq \mathsf{P}\Big[D_s^x \neq D_s^y \text{ for all } s \leq t\Big].$$

Now consider the difference $D_t = D_t^x - D_t^y$. Each jump of D^x produces a jump of D and each jump of D^y produces a jump of D to the opposite direction. As long as $D_t^x \neq D_t^y$, these jumps are caused by independent Poisson processes, so $t \mapsto D_t = D_t^x - D_t^y$ follows a

random walk with rate 2 of jumps to nearest neighbors. But the random walk in dimension $d \leq 2$ is recurrent, so we have

$$\mathsf{P}\Big[D_s^x \neq D_s^y \text{ for all } s \leq t\Big] \underset{t\to\infty}{\longrightarrow} 0.$$

This proves the claim. □

After having proved this, the classification of stationary distributions is a straightforward corollary.

Theorem XII.2 (No nontrivial stationary measures for voter model in $d \leq 2$)**.** *Let $d \leq 2$. Then the only stationary measures of the voter model are the convex combinations of delta measures on constant configurations,*

$$\sum_{a\in S} p_a\, \delta_{\bar{a}} \qquad \textit{where } p_a \geq 0 \textit{ for all } a \in S \textit{ and } \sum_{a\in S} p_a = 1.$$

Proof. Let ν be a stationary distribution and use an initial state ξ_0 distributed according to ν. The previous proposition says that for any $x \neq y$ we have

$$\mathsf{P}\Big[\xi_t(x) \neq \xi_t(y)\Big] \underset{t\to\infty}{\longrightarrow} 0.$$

By stationarity, however, ξ_t is distributed according to ν for any $t \geq 0$, so the left hand side is constant in t, and we must therefore have

$$\mathsf{P}\Big[\xi_t(x) \neq \xi_t(y)\Big] = 0.$$

Now write the probability that ξ_t is not equal to any constant configuration as

$$\begin{aligned}\mathsf{P}\Big[\xi_t \notin \{\bar{a} \,|\, a \in S\}\Big] &\leq \mathsf{P}\Big[\exists x \neq y : \xi_t(x) \neq \xi_t(y)\Big] \\ &\leq \sum_{x\neq y} \mathsf{P}\Big[\xi_t(x) \neq \xi_t(y)\Big] = \sum_{x\neq y} 0 = 0.\end{aligned}$$

Therefore the complementary probability is

$$1 = \mathsf{P}\Big[\xi_t \in \{\bar{a} \,|\, a \in S\}\Big] = \sum_{a\in S} \mathsf{P}\Big[\xi_t = \bar{a}\Big].$$

This shows that the stationary distribution ν is a convex combination of delta measures $\delta_{\bar{a}}$ on the constant configurations $\bar{a}$, with coefficients $p_a = \mathsf{P}\big[\xi_t = \bar{a}\big]$. □

3. Nontrivial stationary measures in high dimension

Let us then analyze the case of dimension $d > 2$. To simplify the notation, we assume $S = \{0, 1\}$ below — the existence of nontrivial stationary measures for general S with $\#S \geq 2$ follows from this basic case easily.

Proposition XII.3 (Non-trivial limit distribution for voter model in $d > 2$)**.** *Let $d > 2$ and $\theta \in (0, 1)$. Consider the voter model with $S = \{0, 1\}$ started from the initial configuration ξ_0 whose components $(\xi_0(x))_{x\in\mathbb{Z}^d}$ are independent and $\mathsf{P}\big[\xi_0(x) = 1\big] = \theta$ for each $x \in \mathbb{Z}^d$, in other words*

$$\xi_0 \sim \text{Bernoulli}(\theta)^{\otimes\mathbb{Z}^d}.$$

Then the states ξ_t of the process converge in distribution

$$\xi_t \underset{t\to\infty}{\overset{\text{d}}{\longrightarrow}} \xi_\infty^{(\theta)},$$

to a limit $\xi_\infty^{(\theta)}$, whose distribution satisfies for any $x, y \in \mathbb{Z}^d$, $x \neq y$,

$$\mathsf{P}\Big[\xi_\infty^{(\theta)}(x) = 1\Big] = \theta \qquad \textit{and} \qquad \mathsf{P}\Big[\xi_\infty^{(\theta)}(x) = 1,\ \xi_\infty^{(\theta)}(y) = 1\Big] \neq \theta.$$

In particular, the distribution of $\xi_\infty^{(\theta)}$ is not a convex combination of delta measures of constant configurations.

Proof. Recall from Theorem VII.5 that to prove weak convergence of ξ_t on the space $\{0,1\}^{\mathbb{Z}^d}$, it is sufficient to show the convergence of probabilities of cylinder events. Moreover, as in the case of Ising model, by inclusion-exclusion it is in fact sufficient to show that the limits

$$\lim_{t\to\infty} \mathsf{P}\Big[\xi_t(x) = 1 \text{ for all } x \in A\Big]$$

exist for all finite subsets $A \subset \mathbb{Z}^d$.

Let therefore $A \subset \mathbb{Z}^d$ be finite, and consider the coalsecent random walks D_t^x for $x \in A$. Denote by

$$N_t^A := \#\left\{D_t^x \;\middle|\; x \in A\right\}$$

the number of distinct positions occupied by these random walks at time t. Clearly $t \mapsto N_t^A$ is non-increasing by virtue of coalescence, and thus we have a limit

$$N_t^A \downarrow N_\infty^A \qquad \text{as } t \to \infty.$$

On the other hand, by duality (XII.2), we have

$$\mathsf{P}\Big[\xi_t(x) = 1 \text{ for all } x \in A\Big] = \mathsf{P}\Big[\xi_0(D_t^x) = 1 \text{ for all } x \in A\Big].$$

Conditioning on $N_t^A = n$, the right hand side above becomes easy to calculate, since the components of the initial configuration ξ_0 are independent and follow Bernoulli distribution with parameter θ. Indeed, we have

$$\mathsf{P}\Big[\xi_0(D_t^x) = 1 \text{ for all } x \in A \;\Big|\; N_t^A = n\Big] = \theta^n,$$

$$\text{and thus} \qquad \mathsf{P}\Big[\xi_0(D_t^x) = 1 \text{ for all } x \in A\Big] = \mathsf{E}\Big[\theta^{N_t^A}\Big].$$

Now recalling that $N_t^A \downarrow N_\infty^A$, we get $\theta^{N_t^A} \uparrow \theta^{N_\infty^A}$. Using Monotone Convergence Theorem to this, and combining the above calculations, we get

$$\mathsf{P}\Big[\xi_t(x) = 1 \text{ for all } x \in A\Big] = \mathsf{E}\Big[\theta^{N_t^A}\Big] \underset{t\to\infty}{\longrightarrow} \mathsf{E}\Big[\theta^{N_\infty^A}\Big].$$

We conclude that the states ξ_t of the process converge weakly to a random variable $\xi_\infty^{(\theta)}$ characterized by

$$\mathsf{P}\Big[\xi_\infty^{(\theta)}(x) = 1\ \forall x \in A\Big] = \mathsf{E}\Big[\theta^{N_\infty^A}\Big].$$

We will show that the laws of $\xi_\infty^{(\theta)}$ are different for different θ (they will soon also be shown to be stationary distributions), and that for $0 < \theta < 1$ they are not convex combinations of delta-measures on constant configurations.

First take the set A to consist of a single point, $A = \{x\}$. We then necessarily have $N_t^{\{x\}} = 1$ for all t (a single random walk occupies exactly one position at any time t), so in particular $N_\infty^{\{x\}} = 1$ and

$$\mathsf{P}\Big[\xi_\infty^{(\theta)}(x) = 1\ \forall x \in A\Big] = \mathsf{E}\Big[\theta^1\Big] = \theta.$$

This formula has two simple but important consequences. First, it shows that for different values of θ, the laws of $\xi_\infty^{(\theta)}$ are indeed different. Second, the only convex combination of consensus measures $\delta_{\bar{a}}$ for which the above formula holds is

$$\theta\,\delta_{\bar{1}} + (1-\theta)\,\delta_{\bar{0}}.$$

Therefore, in order to show nontriviality of the law of $\xi_\infty^{(\theta)}$, it remains to show that the law is different from $\theta\delta_{\bar{1}} + (1-\theta)\delta_{\bar{0}}$.

Next consider $A = \{x, y\}$ for any two distinct points $x \neq y$. We have $N_t^{\{x,y\}} = 2$ until the random walks D_t^x and D_t^y started from x and y coalesce, and $N_t^{\{x,y\}} = 1$ after coalescence. The difference $D_t^x - D_t^y$ follows a continuous time random walk until hitting zero at the time of coalescence. In dimension $d > 2$ the random walk is transient, so there is a positive probability $q_{x,y} > 0$ that the difference $D_t^x - D_t^y$ never gets to zero, i.e., that coalescence never occurs. The limit random variable indicates the occurrence of coalescence,

$$N_\infty^{\{x,y\}} = \begin{cases} 1 & \text{if coalescence of } D_t^x \text{ and } D_t^y \text{ occurs at some finite } t \\ 2 & \text{otherwise.} \end{cases}$$

The probabilities of the two alternatives above are $1 - q_{x,y}$ and $q_{x,y}$, so we can calculate

$$\begin{aligned} \mathsf{P}\Big[\xi_\infty^{(\theta)}(x) = 1 \text{ and } \xi_\infty^{(\theta)}(x) = 1\Big] &= \mathsf{E}\Big[\theta^{N_\infty^{\{x,y\}}}\Big] \\ &= q_{x,y}\,\theta + (1 - q_{x,y})\,\theta^2. \end{aligned}$$

From this formula we conclude nontriviality of the law of $\xi_\infty^{(\theta)}$ for $0 < \theta < 1$, since for the convex combination $\theta\delta_{\bar{1}} + (1-\theta)\delta_{\bar{0}}$ the value above would be equal to θ. This finishes the proof. □

It is a general fact that limit distributions such as the ones in Proposition XII.3 are necessarily stationary distributions.

Lemma XII.4 (Limit distributions of Feller processes are stationary)**.** *Consider a Markov process $\xi = (\xi_t)_{t\in[0,\infty)}$ on a metric space $\mathfrak{X}$, such that the Markov semigroup $(T_t)_{t\geq 0}$ is Feller. Suppose that the states ξ_t of the process converge in distribution*

$$\xi_t \xrightarrow[t\to\infty]{\mathrm{d}} \xi_\infty.$$

Then the law μ_∞ of ξ_∞ is a stationary measure for the process.

Proof. Recall that a probability measure ν on $\mathfrak{X}$ is determined by the integrals $\int_{\mathfrak{X}} f(\eta)\,\mathrm{d}\nu(\eta)$ of bounded continuous functions $f\colon \mathfrak{X} \to \mathbb{R}$, see Exercise H.8.

For a given probability measure ν on $\mathfrak{X}$ and $s \geq 0$, define the probability measure $T_s^*\nu$ by

$$\begin{aligned} \int_{\mathfrak{X}} f(\eta)\,\mathrm{d}(T_s^*\nu)(\eta) &= \int_{\mathfrak{X}} T_s f(\eta)\,\mathrm{d}\nu(\eta) \\ &= \int_{\mathfrak{X}} \mathsf{E}_\eta\big[f(\xi_s)\big]\,\mathrm{d}\nu(\eta). \end{aligned}$$

The last expression shows that $T_s^*\nu$ is the distribution of the state ξ_s of the process at time s, when the process is started from the initial state ξ_0 distributed according to ν.

Suppose now that

$$\xi_t \xrightarrow[t\to\infty]{\mathrm{d}} \xi_\infty,$$

where the distribution of ξ_∞ is μ_∞. For $t \geq 0$, denote the law of ξ_t by μ_t, so that $\mu_t \xrightarrow{\mathrm{w}} \mu_\infty$ as $t \to \infty$. We first claim that also $T_s^*\mu_t \xrightarrow{\mathrm{w}} T_s^*\mu_\infty$ as $t \to \infty$. This is shown by the following calculation: for all bounded continuous functions $f\colon \mathfrak{X} \to \mathbb{R}$ also $T_s f\colon \mathfrak{X} \to \mathbb{R}$ is bounded and continuous because the semigroup is Feller, so we have

$$\begin{aligned} \lim_{t\to\infty} \int_{\mathfrak{X}} f(\eta)\,\mathrm{d}(T_s^*\mu_t)(\eta) &= \lim_{t\to\infty} \int_{\mathfrak{X}} T_s f(\eta)\,\mathrm{d}\mu_t(\eta) \\ &= \int_{\mathfrak{X}} T_s f(\eta)\,\mathrm{d}\mu_\infty(\eta) \\ &= \int_{\mathfrak{X}} f(\eta)\,\mathrm{d}(T_s^*\mu_\infty)(\eta). \end{aligned}$$

On the other hand, by Markov property we have

$$T_s^*\mu_t = \mu_{t+s}.$$

Taking the limit $t \to \infty$ and using also $T_s^* \mu_t \xrightarrow{\text{w}} T_s^* \mu_\infty$ yields

$$T_s^* \mu_\infty = \mu_\infty.$$

This proves the asserted stationarity of μ_∞. □

The combination of the above results finishes the proof of existence of non-trivial stationary measures for the voter model in dimensions $d > 2$.

Theorem XII.5 (Non-trivial stationary distributions for voter model in $d > 2$)**.** *Let $d > 2$. Then there exist stationary measures of the voter model which are not convex combinations of delta measures on constant configurations.*

Proof. For $\theta \in (0, 1)$, let $\mu_\infty^{(\theta)}$ denote the law of $\xi_\infty^{(\theta)}$ in Proposition XII.3. By that proposition, $\mu_\infty^{(\theta)}$ is not a convex combination of delta measures on constant configurations. Also, since $\xi_\infty^{(\theta)}$ is the limit in distribution of voter model states ξ_t, its law $\mu_\infty^{(\theta)}$ is stationary for the voter model by Lemma XII.4. □

Appendix A

Probability theory fundamentals

1. Measure spaces and probability spaces

Theorem A.1. *Dynkin's identification theorem.*

Theorem A.2. *Carathéodory's extension theorem.*

2. Random variables and their laws

Let $(\Omega, \mathscr{F}, \mathsf{P})$ be a probability space and $(\mathfrak{X}, \mathscr{X})$ a measurable space. Suppose that $X\colon \Omega \to \mathfrak{X}$ is a random variable, i.e., an $\mathscr{F}/\mathscr{X}$-measurable function. Then the *law* or *distribution* of X is the probability measure ν_X on $\mathfrak{X}$ given by

$$\nu_X[A] = \mathsf{P}\big[X \in A\big] \tag{A.1}$$

for all $A \in \mathscr{X}$.

Exercise A.1. Check that ν_X defined by (A.1) is indeed a probability measure on $\mathfrak{X}$.

Laws of real valued random variables

There are many convenient ways to describe the law of a real valued random variable. From here on, let $X\colon \Omega \to \mathbb{R}$ be a real valued random variable, and let ν_X be its law, a (Borel) probability measure on $\mathbb{R}$.

Probability density function

The most intuitive way of describing the distribution of a real valued random variable is the probability density function: the density is large at likely values and small at unlikely values. More precisely, we say that (the law of) a random variable X has *probability density function* $p_X\colon \mathbb{R} \to [0, +\infty]$ if for all Borel sets $B \subset \mathbb{R}$ we have

$$\nu_X[B] = \mathsf{P}\big[X \in B\big] = \int_B p_X(x)\,\mathrm{d}x. \tag{A.2}$$

Clearly the function p_X determines the law ν_X of X.

The description of a law by its density is very convenient when possible, but probability density can not be used as a general approach, since not all random variables have probability density functions. Below we therefore turn to two other approaches: cumulative distribution function and characteristic function. Both of them determine the law of a random variable and they exist for all real valued random variables.

Cumulative distribution function

The *cumulative distribution function* of X is the function $F_X\colon \mathbb{R} \to [0,1]$ given by

$$F_X(x) = \nu_X\big[(-\infty, x]\big] = \mathsf{P}\big[X \leq x\big]. \tag{A.3}$$

The term cumulative distribution function is often abbreviated *c.d.f.*

Lemma A.3. *The law ν_X of X is uniquely determined by the cumulative distribution function F_X of X via*

$$\nu_X\big[(a,b]\big] = F_X(b) - F_X(a).$$

for all $a, b \in \mathbb{R}$, $a < b$.

Proof. This is a consequence of the Dynkin's identification theorem, Theorem A.1, with the π-system that consists of all intervals of the form $(a, b]$. □

Proposition A.4. *A function $F\colon \mathbb{R} \to [0,1]$ is a cumulative distribution function of some random variable if and only if the following conditions hold*

(a) *F is non-decreasing, i.e., $x \leq y \;\Rightarrow\; F(x) \leq F(y)$*
(b) *F is right continuous, i.e., if $x_n \downarrow x$ then $F(x_n) \to F(x)$*
(c) *$F(x) \to 0$ as $x \to -\infty$ and $F(x) \to 1$ as $x \to +\infty$.*

Characteristic function

The *characteristic function* of X is the function $\varphi_X\colon \mathbb{R} \to \mathbb{C}$ given by

$$\varphi_X(\theta) = \mathsf{E}\Big[e^{\mathrm{i}\,\theta X}\Big]. \tag{A.4}$$

Theorem A.5. *The law ν_X of X is uniquely determined by the characteristic function φ_X of X.*

Proof. See, e.g., the course **MS-E1600 Probability Theory**, or [**Wil91**]. □

Proposition A.6. *A function $\varphi\colon \mathbb{R} \to \mathbb{C}$ is a characteristic function of some random variable if and only if the following conditions hold*

(a) *$\varphi(0) = 1$*
(b) *$\varphi\colon \mathbb{R} \to \mathbb{C}$ is continuous*
(c) *for any $z_1, \ldots, z_n \in \mathbb{C}$ and $\theta_1, \ldots, \theta_n \in \mathbb{R}$, we have*

$$\sum_{i,j=1}^{n} z_i\, \overline{z_j}\; \varphi(\theta_i - \theta_j) \geq 0.$$

Example A.7 (Gaussian distribution)**.** Let $\mu \in \mathbb{R}$ and $\sigma \geq 0$. Then the function

$$\varphi(\theta) = \exp\Big(\mathrm{i}\,\theta\mu - \frac{1}{2}\theta^2\sigma^2\Big) \tag{A.5}$$

satisfies the properties (a), (b), (c) of Proposition A.6, so it is a characteristic function of some random variable X. The law of such an X is the *Gaussian distribution* (or *normal*

distribution) with mean μ and variance σ^2, denoted by $\mathrm{N}(\mu, \sigma^2)$. If $\sigma > 0$, then this law has the probability density function

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{1}{2\sigma^2}(x-\mu)^2 \Big) \tag{A.6}$$

in the sense of (A.2). If $\sigma = 0$, then the law is a delta measure δ_μ at the point $\mu \in \mathbb{R}$, i.e., the random variable X almost surely takes the value μ.

Exercise A.2. Suppose that $X_1, \dots, X_n$ are independent random variables, each with a Gaussian law, $X_j \sim \mathrm{N}(\mu_j, \sigma_j^2)$ for all $j = 1, \dots, n$. Show that the sum $S = X_1 + \cdots + X_n$ is also Gaussian, $S \sim \mathrm{N}(\sum_{j=1}^n \mu_j, \sum_{j=1}^n \sigma_j^2)$.

3. Integrals, expected values and convergence theorems

Lemma A.8. *Fatou's lemma*

Theorem A.9. *Monotone convergence theorem.*

Theorem A.10. *Lebesgue's dominated convergence theorem.*

Appendix B

Product measures

This appendix is about product measures, but there are in fact two distinct results that we need about them. A standard topic in measure theory (and probability theory) is the construction of the product of finitely many σ-finite measures. Fubini's theorem, in particular, states that under suitable assumptions (non-negativity of the integrand or integrability of the integrand), the orders of integrations can be interchanged. In the specific case of probability measures, the product measure has the interpretation of independence. In that case we also care about products of infinitely many probability spaces. The construction of the infinite products of probability measures is slightly more subtle than the case of products of only finitely many (σ-finite) measures.

1. Product sigma-algebra

Let I be an index set, and let $\big((\mathfrak{X}_i, \mathscr{X}_i)\big)_{i\in I}$ be a collection of measurable spaces. Consider the Cartesian product

$$\mathfrak{X} = \prod_{i\in I} \mathfrak{X}_i = \Big\{ (x_i)_{i\in I} \;\Big|\; \forall i \in I :\ x_i \in \mathfrak{X}_i \Big\}$$

of the sets $\mathfrak{X}_i$, $i \in I$. The product σ-algebra, also called the cylinder σ-algebra, is denoted by

$$\bigotimes_{i\in I} \mathscr{X}_i,$$

and is defined as the smallest σ-algebra $\mathscr{X}$ on the Cartesian product $\mathfrak{X} = \prod_{i\in I} \mathfrak{X}_i$ such that for each $i \in I$, the projection to the i:th component

$$\mathrm{pr}_i\Big((x_i)_{i\in I}\Big) := x_i$$

is a $\mathscr{X}\,/\,\mathscr{X}_i$-measurable function $\mathrm{pr}_i \colon \mathfrak{X} \to \mathfrak{X}_i$.

If the index set I is finite, then the product σ-algebra $\bigotimes_{i\in I} \mathscr{X}_i$ can equivalently be defined as the smallest σ-algebra $\mathscr{X}$ on the finite Cartesian product $\mathfrak{X} = \prod_{i\in I} \mathfrak{X}_i$ which contains all sets of the form $\prod_{i\in I} A_i$, where $A_i \in \mathscr{X}_i$ for each $i \in I$. Sets of this form are sometimes called measurable rectangles, since they are Cartesian products of measurable sets.

If the index set I is infinite, then the product σ-algebra $\bigotimes_{i\in I} \mathscr{X}_i$ still has a characterization of broadly similar flavor. It can equivalently be defined as the smallest σ-algebra $\mathscr{X}$ on the infinite Cartesian product $\mathfrak{X} = \prod_{i\in I} \mathfrak{X}_i$ which contains all sets of the form $\prod_{i\in I} A_i$, where $A_i \in \mathscr{X}_i$ for each $i \in I$, and $A_i = \mathfrak{X}_i$ for all but finitely many indices $i \in I$. Sets of this form are often called cylinder sets.

2. Products of finitely many σ-finite measures

Assume now that the index set I is finite, and that we are given a collection $\big((\mathfrak{X}_i, \mathscr{X}_i, \mu_i)\big)_{i\in I}$ of measure spaces. Assume moreover, that each of these measure spaces is σ-finite.[1] Then there is a product measure denoted by $\bigotimes_{i\in I}\mu_i$ on the Cartesian product space $(\prod_{i\in I}\mathfrak{X}_i, \bigotimes_{i\in I}\mathscr{X}_i)$, characterized by the property that

$$\big(\bigotimes_{i\in I}\mu_i\big)\big[\prod_{i\in i}A_i\big] = \prod_{i\in I}\mu_i[A_i]$$

whenever $A_i \in \mathscr{X}_i$ for all $i \in I$.

In this situation, Fubini's theorem gives sufficient conditions under which the order of integrations can be interchanged. In order to keep the notation simple, we only state Fubini's theorem for the product of two σ-finite measure spaces $(\mathfrak{X}_1, \mathscr{X}_1, \mu_1)$ and $(\mathfrak{X}_2, \mathscr{X}_2, \mu_2)$ and their product measure $\mu_1\otimes\mu_2$ on the Cartesian product $\mathfrak{X}_1\times\mathfrak{X}_2$.

Theorem B.1. *Let $(\mathfrak{X}_1, \mathscr{X}_1, \mu_1)$ and $(\mathfrak{X}_2, \mathscr{X}_2, \mu_2)$ be two σ-finite measure spaces, and let $(\mathfrak{X}_1 \times \mathfrak{X}_2, \mathscr{X}_1 \otimes \mathscr{X}_2, \mu_1 \otimes \mu_2)$ be their product measure space.*

(a) *If $f\colon \mathfrak{X}_1\times\mathfrak{X}_2 \to [0,+\infty]$ is $\mathscr{X}_1\otimes\mathscr{X}_2$-measurable, then we have the equalities*

$$\begin{aligned}&\int_{\mathfrak{X}_1\times\mathfrak{X}_2} f \,\mathrm{d}(\mu_1\otimes\mu_2) \qquad\qquad (B.1)\\ &= \int_{\mathfrak{X}_2}\left(\int_{\mathfrak{X}_1} f(x_1,x_2)\,\mathrm{d}\mu_1(x_1)\right)\mathrm{d}\mu_2(x_2)\\ &= \int_{\mathfrak{X}_1}\left(\int_{\mathfrak{X}_2} f(x_1,x_2)\,\mathrm{d}\mu_2(x_2)\right)\mathrm{d}\mu_1(x_1)\end{aligned}$$

of numbers in $[0,+\infty]$.

(b) *If $f\colon \mathfrak{X}_1\times\mathfrak{X}_2 \to \mathbb{R}$ is $\mathscr{X}_1\otimes\mathscr{X}_2$-measurable, and if at least one of the integrals*

$$\begin{aligned}&\int_{\mathfrak{X}_1\times\mathfrak{X}_2} |f| \,\mathrm{d}(\mu_1\otimes\mu_2)\\ &\int_{\mathfrak{X}_2}\left(\int_{\mathfrak{X}_1} \big|f(x_1,x_2)\big|\,\mathrm{d}\mu_1(x_1)\right)\mathrm{d}\mu_2(x_2)\\ &\int_{\mathfrak{X}_1}\left(\int_{\mathfrak{X}_2} \big|f(x_1,x_2)\big|\,\mathrm{d}\mu_2(x_2)\right)\mathrm{d}\mu_1(x_1)\end{aligned}$$

is finite, then we have the equalities (B.1) *of real numbers.*

Proof. See, e.g., the course **MS-E1600 Probability Theory**, or [**Wil91**]. □

Some practical consequences for the calculation of expected values are given below.

Corollary B.2. *Let $X\colon \Omega \to [0,+\infty]$ be a non-negative random variable. Then we have*

$$\mathsf{E}\big[X\big] = \int_0^\infty \mathsf{P}\big[X > s\big]\,\mathrm{d}s. \qquad (B.2)$$

[1] Recall that $(\mathfrak{X}_i, \mathscr{X}_i, \mu_i)$ is said to be σ-finite if the the whole space can be expressed as a countable union $\mathfrak{X}_i = \bigcup_{n\in\mathbb{N}} A_n^{(i)}$ of measurable sets $A_1^{(i)}, A_2^{(i)}, \ldots \in \mathscr{X}_i$ of finite measure, $\mu_i[A_n^{(i)}] < +\infty$ for all $n \in \mathbb{N}$.

If $X\colon \Omega \to \mathbb{Z}_{\geq 0}$ a random variable with non-negative integer values, then

$$\mathsf{E}\big[X\big] = \sum_{k=1}^{\infty} \mathsf{P}\big[X \geq k\big]. \tag{B.3}$$

Proof. For clarity we emphasize here also the dependence on the outcome $\omega \in \Omega$. Write, applying Fubini's theorem for non-negative integrands,

$$\begin{aligned}\mathsf{E}\big[X(\omega)\big] = \mathsf{E}\Big[\int_0^{X(\omega)} 1\,\mathrm{d}s\Big] &= \mathsf{E}\Big[\int_0^{\infty} \mathbb{I}_{\{s<X(\omega)\}}\,\mathrm{d}s\Big] \\ &= \int_0^{\infty} \mathsf{E}\big[\mathbb{I}_{\{s<X(\omega)\}}\big]\,\mathrm{d}s = \int_0^{\infty} \mathsf{P}\big[X(\omega) > s\big]\,\mathrm{d}s.\end{aligned}$$

This shows the first claim. The second can be proven similarly, or can be seen as an easy consequence of the first one by noting that if X takes integer values then $s \mapsto \mathsf{P}[X > s]$ is constant $\mathsf{P}[X \geq k]$ on the interval $s \in [k-1, k)$. □

3. Products of infinitely many probability measures

Appendix C

Finite state space Markov processes

Stochastic processes model random phenomena with dependence on a time parameter. At any possible time, the *state* of the process is a random variable. The set $\mathcal{S}$ of possible values of this random variable is called the *state space* of the process.

Example C.1. In this course we encounter for example stochastic processes with the following state spaces:

- For finite state space Markov chains (and Markov processes), the set $\mathcal{S}$ is some finite set. Some specific examples are:
  - Shuffling of a deck of n cards defines a Markov chain on the set $\mathcal{S} = \mathfrak{S}_n$ of all possible orderings of the deck.
  - The thermal motion of the Ising model on a finite graph $\mathcal{G}$ is modeled by Glauber dynamics, which is a Markov process on the set $\mathcal{S} = \{+1, -1\}^{\mathcal{G}}$ of possible spin configurations on the graph.
- For the simple random walk, the state space is $\mathcal{S} = \mathbb{Z}$, and more generally for the d-dimensional simple random walk the state space is $\mathcal{S} = \mathbb{Z}^d$.
- For the Brownian motion the state space is $\mathcal{S} = \mathbb{R}$.
- For interacting particle systems, the state space is typically a countable product of discrete sets: a space of the form $\mathcal{S} = S^I$, where I is a countable index set and S is a finite or countable set.

Exercise C.1. Since the state of a stochastic process at a given time is a random variable with values in $\mathcal{S}$, the set $\mathcal{S}$ must in fact be equipped with a sigma algebra Σ. There is a standard choice of Σ for each of the example cases above. What are these standard choices?

Hint: *For the last case, note that S^I has a natural topology described in Appendix H.4.*

Let the underlying probability space be denoted, as usual, by $(\Omega, \mathscr{F}, \mathsf{P})$, and let $(\mathcal{S}, \Sigma)$ be a measurable space. A *stochastic process* on the state space $\mathcal{S}$ is a collection $(X_t)_{t\in\mathbb{T}}$ of random variables $X_t \colon \Omega \to \mathcal{S}$ indexed by a "time parameter" t. The set $\mathbb{T} \subset \mathbb{R}$ of allowed time parameter values may be either a discrete set or an interval, and the process is correspondingly said to have discrete or continuous time. The random variable X_t is called the *state* of the process at time t. We often refer to the stochastic process $(X_t)_{t\in\mathbb{T}}$ simply as X.

A stochastic process is called a Markov process, if, roughly speaking, its future does not depend on the past, but only depends on the current state of the process. This appendix is a summary of relevant results about Markov processes on finite state space $\mathcal{S}$. Section 1 is about discrete time Markov chains, and Section 2 about continuous time Markov jump processes.

For the rest of this appendix, we assume that the state space $\mathcal{S}$ is a finite set (and Σ is the sigma-algebra of all subsets of $\mathcal{S}$).

1. Markov chains

The time of a Markov chain X is indexed by non-negative integers, $\mathbb{T} = \mathbb{Z}_{\geq 0}$. The process is started at time $t = 0$ from some (possibly random) state $X_0 \in \mathcal{S}$, and time is counted in "steps" so that the (random) state after t steps is $X_t \in \mathcal{S}$.

The most important information about a Markov chain is the probabilities of steps from any state $x \in \mathcal{S}$ to any other state $y \in \mathcal{S}$. These probabilities are encoded in the transition matrix $\mathfrak{P}$. The rows and columns of the transition matrix are indexed by states, so $\mathfrak{P} \in \mathbb{R}^{\mathcal{S}\times\mathcal{S}}$.[1]

2. Continuous time Markov processes

[1]If the finite state space $\mathcal{S}$ has n elements, and if we choose an enumeration of the states so as to identify $\mathcal{S}$ with $\{1, \ldots, n\}$, then the transition matrix can be interpreted as a real $n \times n$ matrix $\mathfrak{P} \in \mathbb{R}^{n\times n}$. This remark is only important for realizing that usual matrix operations make sense and have familiar properties. It would usually be rather awkward to choose an enumeration of states, and we prefer to think of $\mathfrak{P} = (\mathfrak{P}_{x,y})_{x,y\in\mathcal{S}}$ as a matrix with rows and columns indexed by states x and y.

Appendix D

Couplings

1. Coupling of probability measures

Definition D.1 (Coupling of probability measures)**.**
Let $(\mathfrak{X}_1, \mathscr{X}_1, \nu_1)$ and $(\mathfrak{X}_2, \mathscr{X}_2, \nu_2)$ be two probability spaces. A *coupling* of the probability measures ν_1 and ν_2 is a probability measure ν on the Cartesian product space $\mathfrak{X}_1 \times \mathfrak{X}_2$ (equipped with the product sigma algebra $\Sigma_1 \otimes \Sigma_2$) such that

$$\begin{aligned} &\forall E_1 \in \Sigma_1 : \quad \nu\big[E_1 \times \mathfrak{X}_2\big] = \nu_1[E_1] \qquad \text{(D.1)} \\ \text{and} \qquad &\forall E_2 \in \Sigma_2 : \quad \nu\big[\mathfrak{X}_1 \times E_2\big] = \nu_2[E_2]. \end{aligned}$$

Remark D.2 (Independent coupling)**.**
The product measure space $(\mathfrak{X}_1 \times \mathfrak{X}_2, \mathscr{X}_1 \otimes \mathscr{X}_2, \nu_1 \otimes \nu_2)$ is a coupling, but not a very useful one as the two components are independent. Couplings of two probability measures are usually most useful if some relation holds between the two components, which allows to make comparisons between the two measures.

An essentially equivalent, often convenient, definition of couplings can be given in terms of random variables defined on a common probability space.

Definition D.3 (Coupling in terms of random variables)**.**
Let $(\Omega_1, \mathscr{F}_1, \mathsf{P}_1)$ and $(\Omega_2, \mathscr{F}_2, \mathsf{P}_2)$ be two probability spaces and $X_1 \colon \Omega_1 \to \mathfrak{X}_1$ and $X_2 \colon \Omega_2 \to \mathfrak{X}_2$ random variables on each. Let

$$\pi_1 \colon\ \mathfrak{X}_1 \times \mathfrak{X}_2 \to \mathfrak{X}_1 \qquad \text{and} \qquad \pi_2 \colon\ \mathfrak{X}_1 \times \mathfrak{X}_2 \to \mathfrak{X}_2$$

be the projections to the first and second component, respectively. Then a *coupling* of X_1 and X_2 is a random variable $X \colon \Omega \to \mathfrak{X}_1 \times \mathfrak{X}_2$ on some probability space $(\Omega, \mathscr{F}, \mathsf{P})$ such that the law of the first projection $\pi_1(X)$ coincides with the law of X_1 and the law of $\pi_2(X)$ coincides with the law of X_2.

2. Coupling and order

Throughout this section, let $(\mathfrak{X}, \Sigma)$ be a measurable space and let $\preceq$ be a *partial order* on $\mathfrak{X}$, i.e., a binary relation such that

$$\begin{aligned} x \preceq y \ \text{ and } \ y \preceq x \quad &\iff \quad x = y \qquad \text{(D.2)} \\ x \preceq y \ \text{ and } \ y \preceq z \quad &\implies \quad x \preceq z. \qquad \text{(D.3)} \end{aligned}$$

Example D.4 (Coordinatewise partial order)**.**
Let $S \subset \mathbb{R}$ be a subset, and I be some index set. Then the product space

$$S^I = \big\{(s_i)_{i\in I} \;\big|\; s_i \in S \text{ for all } i \in I\big\}$$

has a natural partial order by coordinatewise comparison of values:

$$(s_i)_{i\in I} \preceq (s'_i)_{i\in I} \quad \Longleftrightarrow \quad s_i \le s'_i \text{ for all } i \in I.$$

Example D.5 (Coordinatewise partial orders in models studied in this course)**.**
The above example in fact contains a number of special cases of importance.

- For bond percolation on the hypercubic lattice $\mathbb{Z}^d$, we used $S = \{0, 1\}$ with the interpretation 1 = "open bond", 0 = "closed bond" and $I = \mathrm{E}(\mathbb{Z}^d)$ the set of bonds of the lattice $\mathbb{Z}^d$, so $\{0,1\}^{\mathrm{E}(\mathbb{Z}^d)}$ is the set of percolation configurations such as the ones illustrated in Figures II.1 in Lecture II.
- For the Ising model on a graph $\mathcal{G}$, we used $S = \{+1, -1\}$, the set of values of a single spin, and $I = \mathcal{G}$ the set of vertices (sites) of the graph, so that $\{+1,-1\}^{\mathcal{G}}$ is the set of spin configurations on the graph.

In the setting of a set with a partial order $\preceq$, we can talk about increasing (real valued) functions. A (measurable) function $f\colon \mathfrak{X} \to \mathbb{R}$ is said to be *increasing* if

$$x \preceq y \quad \Longrightarrow \quad f(x) \le f(y). \tag{D.4}$$

Similarly, f is said to be *decreasing* if $-f$ is increasing.

The fundamental use of couplings for monotonicity is based on the following observation: when two random variables taking values on a partially ordered set can be coupled in such a way that the first is always smaller than the second, then the expected values of all increasing functions of the first are always smaller than those of the corresponding expected values of the second.

Lemma D.6 (Monotone couplings)**.**
Suppose that ν_1 and ν_2 are two probability measures on the partially ordered set $\mathfrak{X}$, and suppose that there exists a coupling ν of ν_1 and ν_2 such that $\nu\big[\,\{(x,y) \in \mathfrak{X}\times\mathfrak{X} \mid x \preceq y\}\,\big] = 1$. Then for all increasing functions $f\colon \mathfrak{X} \to \mathbb{R}$ (non-negative or integrable, so that the integrals below exist) we have

$$\int_{\mathfrak{X}} f \,\mathrm{d}\nu_1 \;\le\; \int_{\mathfrak{X}} f \,\mathrm{d}\nu_2.$$

Proof. Let ν be a coupling of ν of ν_1 as above, and let $f\colon \mathfrak{X} \to \mathbb{R}$ be an increasing function. Note that at almost every point $(x, y) \in \mathfrak{X} \times \mathfrak{X}$ with respect to the measure ν, we have $x \preceq y$ and therefore $f(x) \le f(y)$. We can then calculate

$$\int_{\mathfrak{X}} f(x)\,\mathrm{d}\nu_1(x) = \int_{\mathfrak{X}\times\mathfrak{X}} f(x)\,\mathrm{d}\nu(x,y) \le \int_{\mathfrak{X}\times\mathfrak{X}} f(y)\,\mathrm{d}\nu(x,y) = \int_{\mathfrak{X}} f(y)\,\mathrm{d}\nu_2(y).$$

□

Corollary D.7 (Monotone couplings of random variables)**.**
Let $X\colon \Omega \to \mathfrak{X}\times\mathfrak{X}$ be a random variable, and denote by $X_1 = \pi_1(X)$ and $X_2 = \pi_2(X)$ its two components. Suppose that we have $X_1 \preceq X_2$ almost surely, where $\preceq$ is a partial order on $\mathfrak{X}$. Then for all increasing functions $f\colon \mathfrak{X} \to \mathbb{R}$ we have

$$\mathsf{E}\Big[f(X_1)\Big] \le \mathsf{E}\Big[f(X_2)\Big].$$

Proof. The joint law of (X_1, X_2) is a coupling of the laws of the components X_1 and X_2. The assumption $X_1 \preceq X_2$ almost surely says that this coupling satisfies the conditions in Lemma D.6, and the assertion readily follows from it. □

The property in the above lemma is important enough to merit a definition.

Definition D.8 (Stochastic domination)**.**
If ν_1 and ν_2 are two probability measures on the partially ordered set $\mathfrak{X}$ such that for all increasing functions $f\colon \mathfrak{X} \to \mathbb{R}$ we have

$$\int_{\mathfrak{X}} f \,\mathrm{d}\nu_1 \;\leq\; \int_{\mathfrak{X}} f \,\mathrm{d}\nu_2,$$

then we say that the measure ν_2 *stochastically dominates* the measure ν_1, and denote $\nu_1 \preceq_{\mathrm{st}} \nu_2$.

3. Holley's criterion

In this section, we include a criterion for stochastic domination due to Holley [**Hol74**], which we use in Lecture IX to verify the FKG inequality for the Ising model. Although more straightforward proofs of the FKG inequality for the Ising model are possible, Holley's criterion seems appropriate, because it is quite efficient and general, and its proof is instructive.

Suppose now that I is a finite index set and $\Omega = \{-1, +1\}^I$. Then Ω is finite, $\#\Omega = 2^{\#I} < \infty$, and it has the natural partial order defined by

$$\omega \preceq \omega' \quad\Longleftrightarrow\quad \omega_i \leq \omega'_i \quad \forall i \in I.$$

We define the following operations that modify a configuration ω at a particular position. For $j \in I$ and $\omega \in \{-1, +1\}^I$ we define the configurations $\omega^{\uparrow j}, \omega^{\downarrow j} \in \{-1, 1\}^I$ by

$$\omega_i^{\uparrow j} = \begin{cases} \omega_i & \text{if } i \neq j \\ +1 & \text{if } i = j \end{cases} \qquad \text{and} \qquad \omega_i^{\downarrow j} = \begin{cases} \omega_i & \text{if } i \neq j \\ -1 & \text{if } i = j. \end{cases} \tag{D.5}$$

Throughout this section we assume that P and Q are two probability measures on Ω such that for each $\omega \in \Omega$ we have

$$P[\{\omega\}] = p(\omega) > 0 \qquad \text{and} \qquad Q[\{\omega\}] = q(\omega) > 0.$$

Our goal is to prove that a certain criterion that only compares the probabilities of configurations modified at a particular position is a sufficient condition for stochastic domination of P by Q. This is known as Holley's criterion.

Proposition D.9 (Holley's criterion)**.**
Let P and Q be probability measures on $\Omega = \{-1, +1\}^I$ as above. Suppose that for all $\sigma, \tau \in \{-1, +1\}^I$ with $\sigma \preceq \tau$ and for all $j \in I$ we have

$$\frac{p(\sigma^{\uparrow j})}{p(\sigma^{\downarrow j})} \;\leq\; \frac{q(\tau^{\uparrow j})}{q(\tau^{\downarrow j})}. \tag{D.6}$$

Then we have $P \preceq_{\mathrm{st}} Q$.

We first consider a particular Monte-Carlo Markov chain method of sampling from the probability measure P, which conveys a key idea in the proof of Holley's criterion. The sampling is based on setting up a Markov jump process X on the state space $\{-1,+1\}^I$ where the jumps can only change one coordinate of the state at once. Each coordinate gets flipped down at unit rate, and the rate at which the coordinate gets flipped up is determined by a ratio of the probabilities under P of the configuration before and after the flip.

Lemma D.10 (Monte-Carlo Markov chain sampling for the P-component)**.** *Let $X = (X_t)_{t\in[0,\infty)}$ be a continuous time Markov jump process on the state space $\{-1,+1\}^I$ with the jump rates*

$$\lambda(\sigma,\tau) = \begin{cases} \frac{p(\tau)}{p(\sigma)} & \text{if } \tau = \sigma^{\uparrow j} \neq \sigma \text{ for some } j \in I \\ 1 & \text{if } \tau = \sigma^{\downarrow j} \neq \sigma \text{ for some } j \in I \\ 0 & \text{if } \#\left\{i \in I \;\middle|\; \sigma_i \neq \tau_i\right\} \neq 1. \end{cases} \tag{D.7}$$

Then we have, for all $\sigma \in \{-1,+1\}^I$

$$\lim_{t\to\infty} \mathsf{P}\big[X_t = \sigma\big] = p(\sigma).$$

Proof. We claim that the Markov process is irreducible and P is a stationary measure for it. The assertion then follows from general theory of (continuous time) Markov processes on finite state spaces.

To show irreducibility, we must show that any state τ is accessible from any other state σ via a finite sequence of jumps that all have a positive rate in (D.7). This is clear, because we can first perform jumps that one by one lower the coordinates j of σ for which $\sigma_j = +1$, to access the configuration of all -1's. Each of these lowering jumps have rate 1. Then from the configuration of all -1's we can perform jumps that one by one raise those coordinates j for which $\tau_j = +1$. Each of these jumps has rate given by a ratio of values of p, which are assumed positive.

It thus only remains to show that P is stationary for the process X. We claim that P in fact satisfies the detailed balance condition

$$p(\sigma)\,\lambda(\sigma,\tau) = p(\tau)\,\lambda(\tau,\sigma) \qquad \forall \sigma,\tau \in \{-1,+1\}^I .$$

Indeed, according to (D.7), both sides are zero unless σ and τ differ at exactly one coordinate $j \in I$. If σ and τ only differ at one coordinate j, then one of the rates $\lambda(\sigma,\tau)$ and $\lambda(\tau,\sigma)$ equals 1 and the other is a ratio of $p(\sigma)$ and $p(\tau)$, so the detailed balance condition is verified. The detailed balance implies stationarity of P. □

The proof of Holley's criterion relies on building a similar process with two components, one with stationary distribution P and the other with stationary distribution Q, and such that the first component is deterministically lower than the second. The ideas are no more difficult than in the previous lemma — only the number of cases to be considered increases.

Proof of Proposition D.9. We will define a continuous time Markov jump process $Z = (Z_t)_{t\in[0,\infty)}$ on a state space R, which is an appropriate subset of $\{-1,+1\}^I \times \{-1,+1\}^I$, namely

$$R := \left\{(\sigma,\tau) \;\middle|\; \sigma,\tau \in \{-1,+1\}^I,\ \sigma \preceq \tau\right\}.$$

We denote $Z_t = (X_t, Y_t)$, so that the component processes $X = (X_t)_{t\in[0,\infty)}$ and $Y = (Y_t)_{t\in[0,\infty)}$ both have state space $\{-1,+1\}^I$. The jump rates $\lambda\big((\sigma,\tau),(\sigma',\tau')\big)$ of Z are

defined to be the following: for $\sigma \preceq \tau$ and $j \in I$,

$$\lambda\big((\sigma, \tau^{\uparrow j}), (\sigma^{\downarrow j}, \tau^{\downarrow j})\big) = 1 \tag{D.8}$$

$$\lambda\big((\sigma^{\downarrow j}, \tau), (\sigma^{\uparrow j}, \tau^{\uparrow j})\big) = \frac{p(\sigma^{\uparrow j})}{p(\sigma^{\downarrow j})} \tag{D.9}$$

$$\lambda\big((\sigma^{\downarrow j}, \tau^{\downarrow j}), (\sigma^{\downarrow j}, \tau^{\uparrow j})\big) = \frac{q(\tau^{\uparrow j})}{q(\tau^{\downarrow j})} - \frac{p(\sigma^{\uparrow j})}{p(\sigma^{\downarrow j})} \tag{D.10}$$

and

$$\lambda\big((\sigma, \tau), (\sigma', \tau')\big) = 0 \qquad \text{for all other } (\sigma, \tau), (\sigma', \tau') \in R.$$

By (D.8), the j:th coordinate of the second component process Y is flipped down at unit rate, and whenever this happens, also the corresponding coordinate of the first component process X is flipped down if it is not down already. By (D.9), the j:th coordinate of the first component process X is flipped up at a certain rate depending on P, and whenever this happens, also the corresponding coordinate of the second component process Y is flipped up if it is not up already. Finally, by (D.10) it is possible that the j:th coordinate of the second component process Y is flipped up, while the corresponding coordinate of the first component process X remains down. The rate (D.10) of this last possibility is a difference, which is non-negative by the assumption (D.6) in Holley's criterion. Also these jumps from a state $(\sigma, \tau) \in R$ can only arrive at other states in the subset R. Thus the rates indeed define a continuous time Markov process $Z = (Z_t)_{t\in[0,\infty)}$ on R.

Observe that necessarily there exists a probability measure ν on R such that for all $(\sigma, \tau) \in R$ we have

$$\lim_{t\to\infty} \mathsf{P}\Big[Z_t = (\sigma, \tau)\Big] = \nu\big[\,\{\sigma, \tau\}\,\big].$$

Since we have not shown that the process Z is irreducible, the measure ν might depend on the initial state Z_0 of the process, but for our purposes it is sufficient to have any such limit measure. Our goal is to show that the measure ν is a coupling between P and Q.

Now consider only the first component process $X = (X_t)_{t\in[0,\infty)}$ of $Z = (Z_t)_{t\in[0,\infty)}$. We claim that X is itself a continuous time Markov process on $\{-1, +1\}^I$, and its jump rates are (D.7). It is first of all clear that the jumps of Z only affect one coordinate of X at a time. If $j \in I$, and $\sigma \in \{-1, +1\}^I$ is such that $\sigma_j = +1$, then for any $\tau \succeq \sigma$ the jump rate (D.8) from (σ, τ) to $(\sigma^{\downarrow j}, \tau^{\downarrow j})$ is 1. Moreover, this is the only jump of Z that affects the j:th coordinate of X, so X has jumps from σ to $\sigma^{\downarrow j}$ at rate 1. Similarly, if $\sigma \in \{-1, +1\}^I$ is such that $\sigma_j = -1$, then for any $\tau \succeq \sigma$ the jump rate (D.9) from (σ, τ) to $(\sigma^{\uparrow j}, \tau^{\uparrow j})$ is $p(\sigma^{\uparrow j})/p(\sigma)$, and this is the only jump of Z that affects the j:th coordinate of X, so X has jumps from σ to $\sigma^{\uparrow j}$ at rate $p(\sigma^{\uparrow j})/p(\sigma)$. This shows that X is the Markov process as in Lemma D.10. In particular, from the conclusion of the lemma, we get that

$$\lim_{t\to\infty} \mathsf{P}\big[X_t = \sigma\big] = p(\sigma),$$

which implies that

$$\nu\Big[\,\{\sigma\} \times \{-1, +1\}^I\,\Big] = \sum_\tau \lim_{t\to\infty} \mathsf{P}\Big[Z_t = (\sigma, \tau)\Big] = \lim_{t\to\infty} \mathsf{P}\big[X_t = \sigma\big] = p(\sigma),$$

i.e., the marginal of ν on the first component is P.

Similarly, consider only the second component process $Y = (Y_t)_{t\in[0,\infty)}$. We claim that Y is itself a continuous time Markov process on $\{-1, +1\}^I$, and its jump rates are as in (D.7), except with p replaced by q. The argument is almost identical to the case of X, except that there are two jumps of Z that can flip a coordinate of Y up. Indeed, if $\tau \in \{-1, +1\}^I$ is such that $\tau_j = -1$, then for $\sigma \preceq \tau$ the rate (D.9) equals $p(\sigma^{\uparrow j})/p(\sigma)$, and the rate (D.10) equals $q(\tau^{\uparrow j})/q(\tau) - p(\sigma^{\uparrow j})/p(\sigma)$. For any σ, these rates add up to $q(\tau^{\uparrow j})/q(\tau)$, which is then the rate of jumps from τ to $\tau^{\uparrow j}$ for the process Y. From Lemma D.10 we thus get that

$$\lim_{t\to\infty} \mathsf{P}\big[Y_t = \tau\big] = q(\tau),$$

which implies that the marginal of ν on the second component is Q.

We have proven that ν is a coupling of P and Q. Since ν is supported on $R = \{(\sigma, \tau) \,|\, \sigma \preceq \tau\}$, Lemma D.6 implies that $P \preceq_{\text{st}} Q$. This finishes the proof of Holley's criterion. □

Appendix E

Zero-one laws

Often in infinite random systems, some apparently random properties that are not too sensitive to the individual components of the system become certain (they hold with probability one). Zero-one laws are probabilistic results that guarantee such behavior. They are particularly often used to formulate precisely phenomena of phase transitions, where there is an abrupt change in some qualitative property of a system as some parameter is varied.

Below we recall two such zero-one law results, Kolmogorov's 0-1 law and Borel-Cantelli lemmas. As applications in the lectures, we had a phase transition result about percolation (Lecture II) and law of iterated logarithm for random walk (Lecture III).

1. Tail sigma-algebra and Kolmogorov 0-1-law

Kolmogorov's zero-one-law states roughly that in a system with infinite number of independent components, any event that remains unchanged by changes in any finite number of the components must be trivial in the sense that its probability is either 0 or 1. It is sometimes easier to prove that the probability is either 0 or 1 than to decide which of these two extremes actually happens.

The general formulation is the following. Let $(\Omega, \mathscr{F}, \mathsf{P})$ be a probability space. Suppose that $\mathscr{A}_1, \mathscr{A}_2, \ldots$ are sub-sigma algebras of $\mathscr{F}$ which are independent (meaning that if $A_1 \in \mathscr{A}_1, \ldots, A_n \in \mathscr{A}_n$, then $\mathsf{P}[A_1 \cap \cdots \cap A_n] = \mathsf{P}[A_1] \cdots \mathsf{P}[A_n]$). Define

$$\mathscr{T}_n = \sigma\left(\bigcup_{m \geq n} \mathscr{A}_m\right)$$

to be the sigma algebra generated by $\mathscr{A}_m$, $m \geq n$, and then define the tail sigma algebra

$$\mathscr{T}_\infty = \bigcap_{n>0} \mathscr{T}_n. \tag{E.1}$$

Then we have:

Theorem E.1 (Kolmogorov's $0-1$ law)**.** *If $E \in \mathscr{T}_\infty$, then $\mathsf{P}[E] \in \{0, 1\}$.*

Proof. See, e.g., the course **MS-E1600 Probability Theory**, or [**Wil91**]. □

The above result has an essentially equivalent more concrete reformulation, also called Kolmogorov's 0-1 law.

Corollary E.2. *Suppose that $X_1, X_2, X_3, \ldots$ are independent random variables (taking values in an arbitrary measurable space). Suppose that E is an event that is measurable with respect to $\sigma(X_n, X_{n+1}, X_{n+2}, \ldots)$ for any n. Then $\mathsf{P}[E] \in \{0, 1\}$.*

Proof. Set $\mathscr{A}_n = \sigma(X_n)$. We have $\mathscr{T}_n = \sigma(X_n, X_{n+1}, X_{n+2}, \ldots)$. If $E \in \mathscr{T}_n$ for any n, then $E \in \mathscr{T}_\infty = \bigcap_n \mathscr{T}_n$. Thus the assertion follows from the previous formulation. □

We finally make a remark about indexing of the independent components. Suppose that I is any countable index set, and $(\mathscr{A}_i)_{i \in I}$ is a collection of independent sub-sigma algebras of $\mathscr{F}$. We can give some enumeration $i_1, i_2, \ldots$ of all elements of I, and define the tail sigma algebra using this as

$$\mathscr{T}_\infty = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{m \geq n} \mathscr{A}_{i_m} \right).$$

This definition of $\mathscr{T}_\infty$ does not depend on the chosen enumeration (exercise: prove this!). In fact Kolmogorov's 0-1 law is often naturally applied in such a setup.

2. Borel-Cantelli lemmas

The first Borel-Cantelli lemma states that in a sequence of events, almost surely only finitely many occur, if the sum of probabilities is finite. The second Borel-Cantelli lemma states that in a sequence of independent events, almost surely infinitely many occur, if the sum of probabilities is infinite. In this sense, the Borel-Cantelli lemmas are zero-one laws.

Recall the following terminology and notation about a sequence $(A_n)_{n \in \mathbb{N}}$ of events. We denote

$$\limsup(A_n) = \bigcap_{m \in \mathbb{N}} \bigcup_{k \geq m} A_k \tag{E.2}$$

Exercise E.1. To make sure you have understood the meaning of (E.2), verify the following:

(i) Show that the definition (E.2) is equivalent to

$$\limsup(A_n) = \left\{ \omega \in \Omega \;\middle|\; \omega \in A_n \text{ for infinitely many different } n \in \mathbb{N} \right\}. \tag{E.3}$$

(ii) Show that the indicators satisfy

$$\mathbb{I}_{\limsup(A_n)}(\omega) = \limsup_{n \to \infty} \mathbb{I}_{A_n}(\omega) \qquad \forall \omega \in \Omega, \tag{E.4}$$

where the left-hand-side is the lim sup of the real-valued sequence of the indicator values.

In plain English, the event $\limsup(A_n)$ reads "A_n (occurs) infinitely often". We abbreviate this as: A_n i.o.

We now recall the Borel-Cantelli lemmas.

Lemma E.3 (Borel-Cantelli lemmas)**.** *Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of events.*

(i) *If $\sum_n \mathsf{P}[A_n] < \infty$, then we have $\mathsf{P}\big[A_n \text{ i.o.}\big] = 0$.*
(ii) *If (A_n) are independent and $\sum_n \mathsf{P}[A_n] = \infty$, then we have $\mathsf{P}\big[A_n \text{ i.o.}\big] = 1$.*

Proof. See, e.g., the course **MS-E1600 Probability Theory**, or [**Wil91**]. □

Appendix F

Large deviations

1. The idea of large deviations

2. Simple large deviations examples

Exercise F.1. Let $(X_j)_{j\in\mathbb{N}}$ be independent, $X_j \sim \mathrm{N}(\mu,\sigma^2)$, and $S_n = \sum_{j=1}^n X_j$.

(a) Show that for all $x > 0$ we have

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\exp\left(-\frac{x^2}{2}\right) \le \int_x^\infty \exp\left(-\frac{y^2}{2}\right)\mathrm{d}y \le \frac{1}{x}\exp\left(-\frac{x^2}{2}\right).$$

(b) For $a > \mu$, calculate the following rate of large deviations

$$-\lim_{n\to\infty}\frac{1}{n}\log \mathsf{P}[S_n \ge na].$$

Exercise F.2. Recall that $P \sim \mathrm{Poisson}(\lambda)$ if $\mathsf{P}\big[P = k\big] = \frac{1}{k!}e^{-\lambda}\lambda^k$ for all $k \in \mathbb{Z}_{\ge 0}$.

(a) Let P_1, P_2 be two independent Poisson distributed random variables, $P_1 \sim \mathrm{Poisson}(\lambda_1)$ and $P_2 \sim \mathrm{Poisson}(\lambda_2)$. Show that $P_1 + P_2 \sim \mathrm{Poisson}(\lambda_1 + \lambda_2)$.

(b) Let $(X_j)_{j\in\mathbb{N}}$ be independent, $X_j \sim \mathrm{Poisson}(1)$, and $S_n = \sum_{j=1}^n X_j$. Show that, when $a > 1$, the following rate of large deviations holds

$$-\lim_{n\to\infty}\frac{1}{n}\log \mathsf{P}\Big[\frac{S_n}{n} \ge a\Big] = -a + 1 + a\log a.$$

Exercise F.3. This exercise concerns the Cramér entropy, and a related rate of large deviations for the simple random walk.

(a) Let $x \in (-1,1)$, and suppose a sequence $(k_n)_{n\in\mathbb{N}}$ of integers k_n satisfies $k_n = \frac{n}{2}(1+x) + \mathcal{O}(1)$. Using the Stirling approximation to show that

$$\log\binom{n}{k_n} = n\Big(\log(2) - I(x)\Big) + \mathcal{O}(\log(n)),$$

where $I(x)$ is the Cramér entropy

$$I(x) = \frac{1+x}{2}\log(1+x) + \frac{1-x}{2}\log(1-x).$$

(b) Consider the simple random walk, $X_n = \sum_{s=1}^n \xi_s$, where $(\xi_s)_{s\in\mathbb{N}}$ are i.i.d. steps with $\mathsf{P}\big[\xi_s = \pm 1\big] = \frac{1}{2}$. Show that for $0 < a < 1$ we have the following rate of large deviations

$$-\lim_{m\to\infty}\frac{1}{2m}\log \mathsf{P}\Big[\frac{X_{2m}}{2m} \ge a\Big] = I(a).$$

Appendix G

Calculus facts

1. Stirling approximation

Consider the factorial

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1. \tag{G.1}$$

We frequently use the following well known result:

Theorem G.1 (Stirling approximation)**.** *As $n \to \infty$, we have*

$$n! = n^n e^{-n} \sqrt{2\pi n} \big(1 + \mathcal{O}(n^{-1})\big).$$

2. Multi-dimensional Gaussian distribution

Recall from Example A.7, that a real-valued random variable X is said to have a Gaussian distribution with mean $\mu \in \mathbb{R}$ and variance σ^2, denoted $X \sim \mathrm{N}(\mu, \sigma^2)$, if its characteristic function is $\varphi(\theta) = \mathsf{E}\big[e^{\mathbb{i}\,\theta X}\big] = e^{\mathbb{i}\,\theta\mu - \frac{1}{2}\theta^2\sigma^2}$. Recall also from Exercise A.2 that sums of independent Gaussian random variables are themselves Gaussian. In view of these, the following definition of multi-dimensional Gaussian distribution is the following.

Definition G.2. A random vector $V = (V_1, V_2, \ldots, V_n) \in \mathbb{R}^n$ is a *Gaussian vector*, if for all vectors $a = (a_1, a_2 \ldots, a_n) \in \mathbb{R}^n$, the characteristic function of the random variable

$$a \cdot V = \sum_{k=1}^{n} a_k V_k$$

is of the form $\mathsf{E}\big[e^{\mathbb{i}\,\theta\, a \cdot V}\big] = e^{\mathbb{i}\,\theta\mu - \frac{1}{2}\sigma^2\theta^2}$ for some $\mu \in \mathbb{R}$ and $\sigma \geq 0$.

The distribution of a Gaussian vector V is determined by the mean (vector), $m = (m_1, \ldots, m_n) \in \mathbb{R}^n$, and the covariance (matrix) $C \in \mathbb{R}^{n \times n}$

$$m_j = \mathsf{E}[V_j], \qquad C_{ij} = \mathsf{Cov}(V_i, V_j).$$

To see this, note that if $V = (V_1, \ldots, V_n)$ is a Gaussian vector, then considering the random variables $\theta \cdot V = \sum_k \theta_k V_k$, for $\theta \in \mathbb{R}^n$, and the defining property of the Gaussian vector, we get the n-dimensional characteristic function of V

$$\varphi(\theta) := \mathsf{E}\Big[e^{\mathbb{i}\theta \cdot V}\Big] = e^{\mathbb{i}\mu - \frac{1}{2}\sigma^2}.$$

But clearly μ is linear in θ, and expressible as $\mu = \theta \cdot M$. Similarly, σ^2 is quadratic in θ, and expressible as $\theta^\top C\, \theta$.

A consequence of this observation is the following.

Proposition G.3. *If $V = (V_1, V_2, \dots, V_n)$ is a Gaussian vector, then the following are equivalent*

(1) *$V_1, V_2, \dots, V_n$ are independent.*
(2) *$C = (C_{jk})$ is diagonal, i.e., $\mathsf{Cov}(V_j, V_k) = 0$ for $j \neq k$.*

We leave it as an exercise to connect the above definition of Gaussian vectors with the common definition in terms of densities.

Exercise G.1. Let $n \in \mathbb{N}$, and let $M \in \mathbb{R}^n$ be a vector and $C \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix, i.e., a matrix such that $C_{ij} = C_{ji}$ for all $i, j = 1, \dots, n$ and $\vec{v}^{\top} C\vec{v} > 0$ for all non-zero $v \in \mathbb{R}^n$. Define a function $p : \mathbb{R}^n \to \mathbb{R}$ by

$$p(x) = \frac{1}{Z} \exp\left(-\frac{1}{2}(x-m)^{\top}\, C^{-1}\, (x-m)\right),$$

where Z is a constant.

(a) Calculate $\int_{\mathbb{R}^n} p(x)\, \mathrm{d}^n x$, and show that p is a (correctly normalized) probability density on $\mathbb{R}^n$ if $Z = (2\pi)^{n/2}\sqrt{\det(C)}$.
Hint: *First do a change of variables (translation) to reduce to the case $M = 0$. Then do an orthogonal change of variables to a basis in which C is diagonal.*
(b) Choose Z as in part (a), and suppose that $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ is a random vector in $\mathbb{R}^n$, which has probability density[1] $p : \mathbb{R}^n \to \mathbb{R}$ as above. Calculate the characteristic function $\varphi(\theta) = \mathsf{E}\big[e^{\mathrm{i}\,\theta\cdot\xi}\big]$, for $\theta \in \mathbb{R}^n$ where $\theta \cdot \xi = \sum_{j=1}^n \theta_j \xi_j$ denotes the inner product.
(c) Let ξ be the random vector as in (b), and let $a_1, \dots, a_n \in \mathbb{R}$. Show that the linear combination $\sum_{j=1}^n a_j \xi_j$ is a random number with Gaussian distribution.

Remark G.4. Recall that in the one-dimensional Gaussian case, a Gaussian random variable with vanishing variance $\sigma^2 = 0$ is almost surely a constant and does not have a probability density. Similarly, a Gaussian random vector is degenerate if the covariance matrix C is not invertible — the random vector almost surely belongs to an affine subspace of $\mathbb{R}^n$ of lower dimension, and its law has no density with respect to the n-dimensional Lebesgue measure.

[1] This means that $\mathsf{P}[\xi \in E] = \int_E p(x)\, \mathrm{d}^n x$ for all Borel sets $E \subset \mathbb{R}^n$.

Appendix H

Background in topology

1. Topological properties of the real line

Exercise H.1. Show that any open set $V \subset \mathbb{R}$ is the union of at most countably many open intervals.

Hint: *Show that every point $x \in V$ is contained in some interval $(a, b) \subset V$ with rational endpoints $a, b \in \mathbb{Q}$.*

Exercise H.2. Show that any open set $V \subset \mathbb{R}$ is the union of at most countably many disjoint open intervals.

Hint: *Show that every point $x \in V$ is contained in a unique maximal interval (a, b) within the set V. Use the previous exercise to show that there are at most countably many different such maximal intervals.*

Exercise H.3. Show that a monotone function $f\colon \mathbb{R} \to \mathbb{R}$ has at most countably many points of discontinuity.

Hint: *Consider f restricted to an interval $[k, k+1]$. For a given $m \in \mathbb{Z}_{>0}$, how many jumps of size at least $\frac{1}{m}$ can f have on such an interval?*

2. Metric space topology

Basic concepts of metric space topology

Recall that a metric space is a set $\mathfrak{X}$ equipped with a metric, i.e., a function $\varrho\colon \mathfrak{X} \times \mathfrak{X} \to [0, \infty)$ such that

$$\varrho(x, y) = 0 \quad \Leftrightarrow \quad x = y \qquad (\varrho\text{-Sep})$$

$$\varrho(x, y) = \varrho(y, x) \qquad \forall x, y \in \mathfrak{X} \qquad (\varrho\text{-Sym})$$

$$\varrho(x, y) \leq \varrho(x, z) + \varrho(z, y) \qquad \forall x, y, z \in \mathfrak{X}. \qquad (\varrho\text{-Tri})$$

We will use the following definitions that the reader is assumed to be familiar with.

For $x \in \mathfrak{X}$ and $r > 0$, the (open) *ball* of radius r centered at x is the subset $B_r(x) = \{y \in \mathfrak{X} \mid \varrho(x, y) < r\}$. A set $A \subset \mathfrak{X}$ is *open*, if for all its points some ball centered at that point is contained in the set A, and *closed*, if its complement $\mathfrak{X} \setminus A$ is open. A sequence $(x_n)_{n\in\mathbb{N}}$ of points $x_n \in \mathfrak{X}$ *converges* to $x \in \mathfrak{X}$ if $\varrho(x_n, x) \to 0$ as $n \to \infty$ — we denote $x_n \to x$ or $\lim_{n\to\infty} x_n = x$. A sequence $(x_n)_{n\in\mathbb{N}}$ of points $x_n \in \mathfrak{X}$ is a *Cauchy sequence*, is for all $\varepsilon > 0$ there exists some $N \in \mathbb{N}$ such that $\varrho(x_n, x_m) < \varepsilon$ whenever $n, m \geq N$. Note that any convergent sequence is a Cauchy sequence. If $(\mathfrak{X}^{(1)}, \varrho^{(1)})$ and $(\mathfrak{X}^{(2)}, \varrho^{(2)})$ are two metric spaces, then a function $f\colon \mathfrak{X}^{(1)} \to \mathfrak{X}^{(2)}$ is *continuous* if for any convergent sequence $(x_n)_{n\in\mathbb{N}}$ of points $x_n \in \mathfrak{X}^{(1)}$ the sequence $(f(x_n))_{n\in\mathbb{N}}$ converges in $\mathfrak{X}^{(2)}$ and $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$.

The notions of convergence of sequences and continuity of functions can be formulated in purely topological terms, without reference to metric, only using the notion of open sets. Indeed, a sequence $(x_n)_{n\in\mathbb{N}}$ converges to $x \in \mathfrak{X}$ if and only if, for all open sets $V \subset \mathfrak{X}$ containing x, we have $x_n \in V$ for all but finitely many n. Also, a function $f\colon \mathfrak{X}^{(1)} \to \mathfrak{X}^{(2)}$ is continuous if and only if for any open set V in $\mathfrak{X}^{(2)}$ the preimage $f^{-1}(V) = \bigl\{x \in \mathfrak{X}^{(1)} \bigm| f(x) \in V\bigr\}$ is open in $\mathfrak{X}^{(1)}$.

Exercise H.4. Verify that the convergence of a sequence and continuity of a function can be equivalently defined in terms of open sets as stated above.

The *closure* $\overline{A}$ of $A \subset \mathfrak{X}$ is the intersection of all closed sets containing A, and the *interior* A° of $A \subset \mathfrak{X}$ is the union of all open sets contained in A. The *boundary* of $A \subset \mathfrak{X}$ is $\partial A = \overline{A} \setminus A^\circ$. A set A is *dense* in $\mathfrak{X}$, if its closure is the whole space, $\overline{A} = \mathfrak{X}$.

Definition H.1. A metric space $(\mathfrak{X}, \varrho)$ is

- *complete*, if all Cauchy sequences in $\mathfrak{X}$ converge.
- *separable*, is there exists a countable dense subset $A \subset \mathfrak{X}$.

Below are a few familiar examples.

Example H.2. The set of real numbers $\mathbb{R}$ equipped with the usual metric $\varrho(x, y) = |x - y|$ is a complete separable metric space.

Example H.3. The set of rational numbers $\mathbb{Q} \subset \mathbb{R}$ is not complete (but it is separable).

Example H.4. Denote by $\ell^\infty(\mathbb{N})$ the set of all bounded sequences $a = (a_n)_{n\in\mathbb{N}}$ of real numbers. The most commonly used metric on $\ell^\infty(\mathbb{N})$ is $\varrho\bigl(a, b\bigr) = \|a-b\|_\infty$, inherited from the uniform norm $\|a\|_\infty = \sup_{n\in\mathbb{N}} |a_n|$. The space $\ell^\infty(\mathbb{N})$ is not separable (but it is complete).

We devote separate sections to two examples that are important for some probabilistic applications in the lectures. The space $\mathcal{C}([0, T])$ of continuous functions $f\colon [0, T] \to \mathbb{R}$ equipped with the metric inherited from the uniform norm is treated in Section H.3 — this is the space used for the Brownian motion in Lecture VIII. Countable Cartesian products of finite sets are the topic of Section H.4 — they were used in Lecture **??** for the Ising model, and in Lectures XI for interacting particle systems. Both of these are complete separable metric spaces.

The notion of *distance to a set* is used frequently. The following exercise establishes its basic properties.

Exercise H.5. Let $A \subset \mathfrak{X}$ be a non-empty subset. For $x \in \mathfrak{X}$, define

$$\varrho(x, A) := \inf_{y\in A} (\varrho(x, y)). \tag{H.1}$$

(a) Show that for any $x, y \in \mathfrak{X}$ we have $|\varrho(x, A) - \varrho(y, a)| \le \varrho(x, y)$.
(b) Conclude from (a) that $x \mapsto \varrho(x, A)$ is a continuous function $\mathfrak{X} \to [0, \infty)$.
(c) Show that

$$\overline{A} = \Bigl\{x \in \mathfrak{X} \Bigm| \varrho(x, A) = 0\Bigr\},$$

and conclude in particular that for A closed we have $x \in A \Leftrightarrow \varrho(x, A) = 0$.

Exercise H.6 (Thickening of a set)**.**
Let $A \subset \mathfrak{X}$ be a non-empty subset, and define $\varrho(x, A)$ as in (H.2). For $\delta > 0$, consider the *δ-thickening*

$$A_\delta := \left\{ x \in \mathfrak{X} \;\middle|\; \varrho(x, A) < \delta \right\} \tag{H.2}$$

of the set A.

(a) Show that

$$\bigcap_{\delta > 0} A_\delta = \overline{A},$$

and conclude in particular that for A closed we have $A_{\delta_k} \downarrow A$ for any sequence $\delta_k \downarrow 0$.

(b) Show that

$$\partial A_\delta \subset \left\{ x \in \mathfrak{X} \;\middle|\; \varrho(x, A) = \delta \right\}.$$

Find an example in which the inclusion is strict.

Borel probability measures

Definition of the Borel sigma algebra on a topological space

Let $\mathfrak{X}$ be a topological space (e.g., a metric space). In order to consider probability measures on $\mathfrak{X}$, we need to equip it with a σ-algebra.

Definition H.5. The Borel σ-algebra $\mathscr{B} = \mathscr{B}(\mathfrak{X})$ on a topological space $\mathfrak{X}$ is the smallest sigma algebra which contains all open sets $U \subset \mathfrak{X}$.

Remark H.6. By the properties of σ-algebras, the Borel σ-algebra $\mathscr{B}$ contains all closed sets $F \subset \mathfrak{X}$ (as the complements of open sets), all countable intersections of open sets, all countable unions of closed sets, etc.

In what follows, we consider probability measures on topological spaces $\mathfrak{X}$ (usually metric spaces), equipped with their Borel σ-algebras. The probability triples are thus of the form $\big(\mathfrak{X}, \mathscr{B}, \nu\big)$. Often the Borel σ-algebra is used without explicit mention, but when we want to emphasize this choice, we call ν a Borel probability measure on $\mathfrak{X}$.

Exercise H.7. Suppose that $(\mathfrak{X}, \varrho)$ is a separable metric space, and that $G \subset \mathfrak{X}$ is an open set.

(a) Show that G can be written as a countable union of open balls. Conclude that the Borel sigma-algebra $\mathscr{B}(\mathfrak{X})$ is generated by open balls.
(b) Show that G can be written as a countable union of closed balls. Conclude that the Borel sigma-algebra $\mathscr{B}(\mathfrak{X})$ is generated by closed balls.

Regularity of Borel probability measures on metric spaces

In practise, when we do probability theory, the topology of $\mathfrak{X}$ will always be metrizable. Suppose now, therefore, that the space $\mathfrak{X}$ is equipped with a metric ϱ which gives its topology.

In this setup, any probability measure on $\mathfrak{X}$ is regular in the sense that we can approximate Borel sets from below by closed sets and from above by open sets as follows.

Proposition H.7. *Let ν be a Borel probability measure on a metric space $(\mathfrak{X}, \varrho)$. For any Borel set $E \subset \mathfrak{X}$ and any $\varepsilon > 0$, there exists a closed set $F \subset \mathfrak{X}$ and an open set $G \subset \mathfrak{X}$ such that $F \subset E \subset G$ and $\nu[G \setminus F] < \varepsilon$.*

Proof. Suppose first that E is closed. Then the open sets $G_\delta = \{x \in \mathfrak{X} \,|\, \varrho(x, E) < \delta\}$ approximate E from above: $G_\delta \downarrow E$ as $\delta \searrow 0$. Therefore we have $\lim_{\delta \searrow 0} \nu[G_\delta] = \nu[E]$, and we can choose $F = E$ and $G = G_\delta$ for some small enough $\delta > 0$. Thus all closed sets have the desired property. It is easy to see that the collection of sets which has the desired property forms a σ-algebra, and it must thus be the entire Borel σ-algebra $\mathscr{B}$. □

We also record the following trivial but very useful lemma.

Lemma H.8. *Let $(\mathfrak{X}, \varrho)$ be a metric space. For any closed $F \subset \mathfrak{X}$ and any $\varepsilon > 0$, there exists a continuous function $f \colon \mathfrak{X} \to [0, 1]$ such that $f(x) = 1$ if and only if $x \in F$, and $f(x) = 0$ if and only if $\varrho(x, F) \geq \varepsilon$.*

Proof. If $F = \emptyset$, then the zero function $f(x) \equiv 0$ works. If $F \neq \emptyset$, define $\varrho(x, F) = \inf_{y \in F}(\varrho(x, y))$ as in Exercise H.5 and use for example the function $f(x) = \max\{1 - \frac{\varrho(x,F)}{\varepsilon}, 0\}$. □

It is also frequently important to consider what information about a probability measure is needed to fully determine the measure. We leave the following as an exercise.

Exercise H.8. Let $(\mathfrak{X}, \varrho)$ be a metric space, and let ν_1, ν_2 be two (Borel) probability measures on $\mathfrak{X}$. Then either of the following is a sufficient condition for $\nu_1 = \nu_2$:

(i) for all closed sets $F \subset \mathfrak{X}$ we have $\nu_1[F] = \nu_2[F]$
Hint: *Recall Dynkin's identification theorem (Theorem A.1).*
(ii) for all bounded continuous functions $f \colon \mathfrak{X} \to \mathbb{R}$ we have $\int_{\mathfrak{X}} f \,\mathrm{d}\nu_1 = \int_{\mathfrak{X}} f \,\mathrm{d}\nu_2$.

The second part of the above exercise in particular shows that the condition appearing in the definition of weak limit, Definition VI.1, does indeed uniquely determine the limit as a Borel probability measure.

Compactness and sequential compactness

For a topological space $\mathfrak{X}$, we give the following definitions.

Definition H.9. We say that $\mathfrak{X}$ is *sequentially compact* if every sequence $(x_n)_{n \in \mathbb{N}}$ of points in $\mathfrak{X}$ has a subsequence $\left(x_{n_k}\right)_{k \in \mathbb{N}}$ which converges to some limit $x \in \mathfrak{X}$, i.e. $\lim_{k \to \infty} x_{n_k} = x$.

We say that $\mathfrak{X}$ is *compact* if every collection $(U_i)_{i \in I}$ of open sets $U_i \subset \mathfrak{X}$ that covers the space, $\bigcup_{i \in I} U_i = \mathfrak{X}$, has a finite subcollection $U_{i_1}, \ldots, U_{i_m}$ which also covers the space, $\bigcup_{k=1}^{m} U_{i_k} = \mathfrak{X}$.

For metric spaces the two definitions are equivalent.

Proposition H.10. *A metric space $(\mathfrak{X}, \varrho)$ is compact if and only if it is sequentially compact.*

Exercise H.9. Prove Proposition H.10 (or recall its proof).

Familiar examples of compact sets are e.g. closed intervals $[a, b] \subset \mathbb{R}$ of the real line, and more generally, Heine-Borel theorem characterizes compact subsets of Euclidean spaces as follows.

Example H.11. In d-dimensional Euclidean space, a subset $A \subset \mathbb{R}^d$ is compact if and only if it is bounded and closed.

Compactness has many important consequences. A few frequently used ones are given below.

Proposition H.12. *Suppose that $(\mathfrak{X}, \varrho)$ is compact metric space and $f\colon \mathfrak{X} \to \mathbb{R}$ is a continuous function. Then we have:*

- *f is bounded: there exists some $M < \infty$ such that*
$$|f(x)| \leq M \qquad \text{for all } x \in \mathfrak{X}.$$
- *f attains its maximum and minimum: there exists points $x_{\max}, x_{\min} \in \mathfrak{X}$ such that*
$$f(x_{\min}) = \inf_{x \in \mathfrak{X}} f(x) \qquad \text{and} \qquad f(x_{\max}) = \sup_{x \in \mathfrak{X}} f(x).$$
- *f is uniformly continuous: for any $\varepsilon > 0$ there exists a $\delta > 0$ such that*
$$x, y \in \mathfrak{X},\ \varrho(x, y) < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$

Exercise H.10. Prove Proposition H.12 (or recall its proof).

Exercise H.11. Show that a compact metric space is separable.

We also occasionally use precompactness.

Definition H.13. We say that a subset $A \subset \mathfrak{X}$ is *precompact* if every sequence $(x_n)_{n \in \mathbb{N}}$ of points in A has a subsequence $\left(x_{n_k}\right)_{k \in \mathbb{N}}$ which converges to some limit $x \in \mathfrak{X}$, i.e. $\lim_{k \to \infty} x_{n_k} = x$.

Equivalently, $A \subset \mathfrak{X}$ is precompact if and only if its closure $\overline{A} \subset \mathfrak{X}$ is sequentially compact. Precompactness is "relative" notion: it depends not only on the set A but also on the whole space $\mathfrak{X}$ in which A is viewed as a subset.

At many places in the lectures we used an argument of the following type to verify the convergence of a sequence:

Exercise H.12. Let $(\mathfrak{X}, \varrho)$ be a metric space. Suppose that $(x_n)_{n \in \mathbb{N}}$ is a sequence of points $x_n \in \mathfrak{X}$, which satisfies two conditions:

1°) any subsequence $\left(x_{n_k}\right)_{k \in \mathbb{N}}$ has a further subsequence $\left(x_{n_{k_j}}\right)_{j \in \mathbb{N}}$ that converges
2°) the limits of any two convergent subsequences of $(x_n)_{n \in \mathbb{N}}$ are the same.

Prove that the sequence (x_n) converges.

This can be summarized as: "precompactness plus uniqueness of subsequential limits implies convergence".

3. Space of continuous functions

In Lecture VIII we consider Brownian motion as a random element of a space of continuous functions. We considered the Brownian motion defined on some time interval $[0, T]$, and we equipped the space of continuous real valued function on the interval with the metric inherited from the uniform norm

Here we recall some topological properties of this space of continuous functions.

Metric on the space of continuous functions

The space

$$\mathcal{C}\big([0,T]\big) = \{f\colon [0,T] \to \mathbb{R} \text{ continuous}\} \tag{H.3}$$

of continuous real-valued functions on the interval $[0, T]$ is a vector space with the addition and scalar multiplication defined pointwise. This vector space is equipped with the uniform norm

$$\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|. \tag{H.4}$$

By compactness of $[0, T]$, any $f \in \mathcal{C}([0, T])$ is bounded, so indeed $\|f\|_\infty$ is finite. It is easy to check that $\|\cdot\|_\infty$ is a norm, i.e.,

$$\begin{aligned}
&\|f\|_\infty = 0 \;\Leftrightarrow\; f \equiv 0 && \text{for } f \in \mathcal{C}([0,T])\\
&\|\lambda f\|_\infty = |\lambda|\,\|f\|_\infty && \forall f \in \mathcal{C}([0,T]),\ \lambda \in \mathbb{R}\\
&\|f_1 + f_2\|_\infty \le \|f_1\|_\infty + \|f_2\|_\infty && \forall f_1, f_2 \in \mathcal{C}([0,T]).
\end{aligned}$$

The norm is used to define a metric in the usual way: for $f_1, f_2 \in \mathcal{C}([0, T])$ we set

$$\varrho(f_1, f_2) = \|f_1 - f_2\|_\infty = \sup_{t\in[0,T]} |f_1(t) - f_2(t)|. \tag{H.5}$$

The following is not difficult to prove.

Theorem H.14. *The space $\big(\mathcal{C}([0, T]), \varrho\big)$ is complete and separable.*

Arzelà-Ascoli theorem

In Lecture VIII we used the following characterization of compact subsets in $\big(\mathcal{C}([0, T]), \varrho\big)$.

Theorem H.15 (Arzelà-Ascoli theorem)**.** *A subset $\Phi \subset \mathcal{C}([0, T])$ is precompact if and only if the following two conditions are satisfied:*

(i) *the collection Φ of functions is uniformly bounded: there exists $M < \infty$ such that for all $f \in \Phi$ and $t \in [0, T]$ we have $|f(t)| \le M$.*

(ii) *the collection Φ of functions is (uniformly) equicontinuous: for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $f \in \Phi$ and $t, s \in [0,T]$ we that that $|t-s| < \delta$ implies $|f(t) - f(s)| < \varepsilon$.*

Proof of necessity of the conditions: Suppose that $\Phi \subset \mathcal{C}([0,T])$ is precompact. Then $\overline{\Phi}$ is compact.

By compactness of $[0,T]$, any $f \in \mathcal{C}([0,T])$ is bounded, i.e., $\sup_{t\in[0,T]} |f(t)| < \infty$. Therefore as $M > 0$ varies, the sets

$$G_M := \Big\{ f \in \mathcal{C}([0,T]) \;\Big|\; \sup_{t\in[0,T]} |f(t))| < M \Big\}$$

cover $\mathcal{C}([0,T])$, i.e.,

$$\mathcal{C}([0,T]) \subset \bigcup_{M>0} G_M.$$

These sets G_M are open as preimages of open sets under the continuous function $f \mapsto \sup_{t\in[0,T]} |f(t)|$ from $\mathcal{C}([0,T])$ to $\mathbb{R}$. By compactness of $\overline{\Phi}$, some finite collection of these open sets suffices to cover $\overline{\Phi}$, i.e., we have

$$\overline{\Phi} \subset G_{M_1} \cup \cdots \cup G_{M_n}.$$

But taking $M = \max(M_1, \ldots, M_n)$ we then have $\overline{\Phi} \subset G_M$. This implies that for every $f \in \Phi$ and any $t \in [0,T]$ we have $|f(t)| < M$ since $f \in \Phi \subset \overline{\Phi} \subset G_M$. This establishes (i).

Fix $\varepsilon > 0$. By compactness of $[0,T]$, any $f \in \mathcal{C}([0,T])$ is uniformly continuous, so we have

$$\lim_{\delta \downarrow 0} \sup_{\substack{t,s\in[0,T] \\ |t-s|\le\delta}} |f(t) - f(s)| \;=\; 0.$$

This implies that as $\delta > 0$ varies, the sets

$$G'_\delta := \left\{ f \in \mathcal{C}([0,T]) \;\middle|\; \sup_{\substack{t,s\in[0,T] \\ |t-s|\le\delta}} |f(t) - f(s)| < \varepsilon \right\}$$

cover $\mathcal{C}([0,T])$, i.e.,

$$\mathcal{C}([0,T]) \subset \bigcup_{\delta>0} G'_\delta.$$

These sets G'_δ are open as preimages of open sets under the continuous function

$$f \mapsto \sup_{\substack{t,s\in[0,T] \\ |t-s|\le\delta}} |f(t) - f(s)|$$

from $\mathcal{C}([0,T])$ to $\mathbb{R}$. By compactness of $\overline{\Phi}$, some finite collection of these open sets suffices to cover $\overline{\Phi}$, i.e., we have

$$\overline{\Phi} \subset G'_{\delta_1} \cup \cdots \cup G'_{\delta_m}.$$

But taking $\delta = \min(\delta_1, \ldots, \delta_m)$ we then have $\overline{\Phi} \subset G'_\delta$. This implies that for every $f \in \Phi$ and any $t, s \in [0,T]$ with $|t-s| < \delta$ we have $|f(t) - f(s)| < \varepsilon$ since $f \in \Phi \subset \overline{\Phi} \subset G'_\delta$. This establishes condition (ii).

Proof of sufficiency of the conditions: Suppose that $\Phi \subset \mathcal{C}([0,T])$ satisfies conditions (i) and (ii). Let $f_n \in \Phi$ for $n \in \mathbb{N}$. To prove precompactness we must extract a convergent subsequence from $(f_n)_{n\in\mathbb{N}}$. We will do this by diagonal extraction on a countable dense set.

Let $Q \subset [0,T]$ be a countable dense set, for example $Q = [0,T] \cap \mathbb{Q}$. Fix an enumeration of this set, $Q = \{q_1, q_2, \ldots\}$.

By condition (i), we have that $|f_n(q_j)| \le M$. Thus from the bounded real sequence $\big(f_n(q_1)\big)_{n\in\mathbb{N}}$ we can pick some subsequence $\big(f_{n_k^{(1)}}(q_1)\big)_{k\in\mathbb{N}}$ such that $f_{n_k^{(1)}}(q_1)$ converges as $k \to \infty$ to a limit denoted by $\phi(q_1) := \lim_{k\to\infty} f_{n_k^{(1)}}(q_1)$. Similarly, we continue picking further subsequences determined $n_k^{(j)}$ such that $\lim_{k\to\infty} f_{n_k^{(j)}}(q_i) = \phi(q_i)$ for all $i \le j$. Finally,

the diagonal subsequence defined by $n_k = n_k^{(k)}$ is such that $\lim_{k\to\infty} f_{n_k}(q) = \phi(q)$ for all $q \in Q$.

We claim that the sequence $(f_{n_k})_{k\in\mathbb{N}}$ is Cauchy. First, for any $\varepsilon > 0$ by condition (ii) we can find a $\delta > 0$ such that for all n, whenever $|t - s| < \delta$, we have

$$|f_n(t) - f_n(s)| < \frac{\varepsilon}{4}.$$

We can now take some finite number m (approximately $m \approx T/\delta$) of intervals $I_r \subset [0, T]$, $r = 1, \ldots, m$, each with positive length less than δ, such that $[0, T] \subset I_1 \cup \cdots \cup I_m$. By density of Q there is some $q^{(r)} \in Q \cap I_r$ for each $r = 1, \ldots, m$, and we fix a choice of these $q^{(1)}, \ldots, q^{(m)}$. Since $\big(f_{n_k}(q^{(r)})\big)_{k\in\mathbb{N}}$ converges, it is Cauchy in $\mathbb{R}$, and thus there exists an N_r such that we have

$$\big|f_{n_k}(q^{(r)}) - f_{n_l}(q^{(r)})\big| < \frac{\varepsilon}{2} \qquad \text{for all } k, l \geq N_r.$$

Now any $t \in [0, T]$ belongs to some of these intervals, $t \in I_r$ for some r. When $k, l \geq N := \max(N_1, \ldots, N_m)$, we can estimate by triangle inequality

$$\begin{aligned}\big|f_{n_k}(t) - f_{n_l}(t)\big| &\leq \big|f_{n_k}(t) - f_{n_k}(q^{(r)})\big| + \big|f_{n_k}(q^{(r)}) - f_{n_l}(q^{(r)})\big| + \big|f_{n_l}(q^{(r)}) - f_{n_l}(t)\big| \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon.\end{aligned}$$

As this holds for all $t \in [0, T]$, we have

$$\|f_{n_k} - f_{n_l}\|_\infty \leq \varepsilon \qquad \text{for all } k, l \geq N := \max(N_1, \ldots, N_m).$$

This shows that sequence $(f_{n_k})_{k\in\mathbb{N}}$ is Cauchy in $\mathcal{C}([0, T])$, so by completeness of $\mathcal{C}([0, T])$ is converges. This $(f_{n_k})_{k\in\mathbb{N}}$ is the desired convergent subsequence of $(f_n)_{n\in\mathbb{N}}$. □

4. Countable products of discrete spaces

In Lecture **??** and Lectures XI we considered models that are defined on spaces that are countably infinite Cartesian products of finite sets. The finite sets are naturally equipped with the discrete topology, and the product space with the product topology. In the case of countable products, the product topology is metrizable. Here we recall some topological properties of spaces of this type.

Metric on countable product of discrete spaces

Let S be a finite or countably infinite set. We consider the Cartesian product of copies of S

$$S^I = \Big\{ (x_i)_{i\in I} \;\Big|\; x_i \in S \Big\} \tag{H.6}$$

where I is a countable index set. In common applications we might have for example $I = \mathbb{Z}^d$. We fix an enumeration $I = \{i_1, i_2, \ldots\}$ of the countable index set, and using it, define a metric on S^I by

$$\varrho(x, y) = \sum_{\substack{k\in\mathbb{Z}_{>0} \\ x_{i_k} \neq y_{i_k}}} 2^{-k} \tag{H.7}$$

$$\text{for } x, y \in S^I \text{ with components } x = (x_i)_{i\in I},\ y = (y_i)_{i\in I}.$$

The metric ϱ depends on the chosen enumeration of I, but the topology induced by the metric is in fact independent of the choice — it is the natural product topology. The space S^I becomes a complete, separable metric space, and moreover compact

if S is finite. We leave the verification of this as an exercise. Note that by choosing the enumeration we may assume $I = \mathbb{N}$.

Exercise H.13. Define $\varrho\colon S^{\mathbb{N}} \times S^{\mathbb{N}} \to [0,\infty)$ by the formula

$$\varrho(x,y) = \sum_{\substack{i\in\mathbb{N}\\ x_i\neq y_i}} 2^{-i} \qquad \text{for } x,y \in S^{\mathbb{N}} \text{ with components } x=(x_i)_{i\in\mathbb{N}},\ y=(y_i)_{i\in\mathbb{N}}.$$

(a) Show that ϱ is a metric on the set $S^{\mathbb{N}}$.
(b) Show that $(S^{\mathbb{N}}, \varrho)$ is complete.
(c) Show that $(S^{\mathbb{N}}, \varrho)$ is separable.
(d) Show that if S is finite, then $(S^{\mathbb{N}}, \varrho)$ is compact.

Cylinder sets

For $j \in I$, denote the projection to the corresponding component by

$$\pi_j\colon S^I \to S, \qquad \pi_j\big((x_i)_{i\in I}\big) = x_j.$$

Let $i_1,\dots,i_m \in I$ and $A_1,\dots,A_m \subset S$. Subsets of S^I of the form

$$\begin{aligned} C &= \pi_{i_1}^{-1}(A_1) \cap \cdots \cap \pi_{i_m}^{-1}(A_m) \\ &= \Big\{ (x_i)_{i\in I} \;\Big|\; x_{i_1} \in A_1, \dots, x_{i_m} \in A_m \Big\} \end{aligned} \tag{H.8}$$

are called *cylinder sets* or *cylinder events*, when we emphasize the probabilistic interpretation and $\Omega = S^I$ is our sample space.

Proposition H.16.

(i) *Every cylinder set $C \subset S^I$ is open.*
(ii) *Every cylinder set $C \subset S^I$ is closed.*
(iii) *Every open set $G \subset S^I$ is the union of cylinder sets contained in it.*
(iv) *The intersection of two cylinder sets is a cylinder set, i.e., the collection Π of all cylinder sets is a π-system.*

Proof of (i): Fix the enumeration of the countable index set $I = \{i_1, i_2, \dots\}$ used in the definition (H.7) of the metric ϱ on S^I.

Consider the cylinder set

$$C = \Big\{ (x_i)_{i\in I} \;\Big|\; x_{i_1} \in A_1, \dots, x_{i_m} \in A_m \Big\}.$$

Let $r = 2^{-m}$. Suppose that $x = (x_i)_{i\in I} \in C$. To prove that the cylinder C is open, we show that the ball of radius r around x is contained in the cylinder, $B_r(\omega) \subset C$. Suppose that $y = (y_i)_{i\in I}$ is a point in this ball, so that $\varrho(x,y) < r$. Then we must have

$$y_{i_1} = x'_{i_1} \in A_1,\ \dots,\ y_{i_m} = x'_{i_m} \in A_m,$$

because otherwise the sum

$$\varrho(x,y) = \sum_{\substack{k\in\mathbb{Z}_{>0}\\ x_{i_k}\neq y_{i_k}}} 2^{-k}$$

would contain a term at least $2^{-m} = r$, which would contradict $\varrho(x,y) < r = 2^{-m}$. This shows that $y \in C$. We thus conclude that $B_r(x) \subset C$, and that the cylinder set C is open.

Proof of (ii): Consider the cylinder set

$$C = \Big\{ (x_i)_{i\in I} \;\Big|\; x_{i_1} \in A_1, \dots, x_{i_m} \in A_m \Big\}.$$

To show that C is closed, we have to show that its complement $S^I \setminus C$ is open. The complement is

$$\begin{aligned} S^I \setminus C &= \left\{ (x_i)_{i\in I} \;\middle|\; \text{for some } k = 1, \dots, m \text{ we have } x_{i_k} \in S \setminus A_k \right\} \\ &= \bigcup_{k=1}^{m} \left\{ (x_i)_{i\in I} \;\middle|\; x_{i_k} \in S \setminus A_k \right\} = \bigcup_{k=1}^{m} \pi_{i_k}^{-1}(S \setminus A_k). \end{aligned}$$

This expression shows that the complement $S^I \setminus C$ of the cylinder C is a finite union of cylinders. By (i), cylinders are open, so $S^I \setminus C$ is open as a union of opens sets, and therefore C is closed.

Proof of (iii): Let $G \subset S^I$ be an open set. Suppose that $y \in G$. Then there exists some $r > 0$ such that $B_r(y) \subset G$. Choose an $m \in \mathbb{N}$ such that $2^{-m} < r$. Form the cylinder set

$$C_y = \left\{ (x_i)_{i\in I} \;\middle|\; x_{i_1} = y_{i_1}, \dots, x_{i_m} = y_{i_m} \right\}.$$

Then for any $x \in C_y$ we have

$$\varrho(x, y) = \sum_{\substack{k\in\mathbb{Z}_{>0} \\ x_{i_k} \neq y_{i_k}}} 2^{-k} \leq \sum_{k=m+1}^{\infty} 2^{-k} = 2^{-m} < r.$$

This implies that $C_y \subset B_r(y) \subset G$. Also clearly $y \in C_y$, and therefore

$$G = \bigcup_{y\in G} C_y,$$

which expresses the open set G as a union of cylinders.[1]

Proof of (iv): It is clear from the definition (H.8) that finite intersections of cylinder sets are cylinder sets. □

Lemma H.17. *If S is finite, then there are only countably many different cylinder sets $C \subset S^I$.*

Proof. Suppose that S is finite. If we fix $m \in \mathbb{N}$, then we only have finitely many choices for the subsets $A_1, \dots, A_m \subset S$ in the defining formula of a cylinder set

$$C = \pi_{i_1}^{-1}(A_1) \cap \cdots \cap \pi_{i_m}^{-1}(A_m).$$

There are countably many choices of m, and the countable union over m of the finitely many choices for each m gives in total a countable number of choices of cylinders. □

Corollary H.18. *If S is finite, then the Borel sigma algebra $\mathscr{B}$ on S^I coincides with the sigma algebra $\sigma(\Pi)$ generated by the collection Π of all cylinder events $C \subset S^I$.*

Proof. By part (i) of Proposition H.16, cylinder sets are open, and they are thus contained in the Borel sigma algebra, which implies that $\sigma(\Pi) \subset \mathscr{B}$.

By part (iii) of Proposition H.16, any open set $G \subset S^I$ is a union of cylinders. If S is finite, then by Lemma H.17 there are only countably many different cylinders, so in fact G is a countable union of cylinders. Therefore it belongs to the sigma algebra generated by cylinders, $G \in \sigma(\Pi)$. Since the Borel sigma algebra $\mathscr{B}$ is the smallest sigma algebra into which all open sets belong, we have $\mathscr{B} \subset \sigma(\Pi)$.

The two inclusions in opposite directions show that $\mathscr{B} = \sigma(\Pi)$. □

[1] In principle in $G = \bigcup_{y\in G} C_y$, one cylinder C_y is needed for each different point $y \in G$, but it is worth observing that cylinders associated to some different y may in fact coincide.

Corollary H.19. *Let S be finite, and let ν_1 and ν_2 be two Borel probability measures on S^I. Suppose that $\nu_1[C] = \nu_2[C]$ for all cylinder events $C \subset S^I$. Then we have $\nu_1 = \nu_2$.*

Proof. The collection Π of cylinder sets forms a π-system on S^I by Proposition H.16(iv). When S is finite, this π-system Π generates the Borel sigma algebra by Corollary H.18, so the assertion follows from Dynkin's identification theorem, Theorem A.1. □

Index

References

Bibliography

[AD99] David Aldous and Persi Diaconis. Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. *Bull. Amer. Math. Soc.*, 36(4):413–432, 1999.

[BD92] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2):294–313, 1992.

[BDJ99] J. Baik, P.A. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12:1119–1178, 1999.

[Bil99] Partrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 2nd edition, 1999.

[DC12] Hugo Duminil-Copin. Théorie mathématique de la percolation. lecture notes, 2012. http://www.unige.ch/ duminil/teaching/percolation/polycopie.pdf.

[Dur95] Rick Durrett. Ten lectures on interacting particle systems. In *Lectures on Probability Theory: Ecole d'Eté de Probabilités de Saint-Flour XXIII—1993*, volume 1608 of *Lecture Notes in Mathematics*. Springer, 1995.

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22(2):89–103, 1971.

[FV15] Sacha Friedli and Yvan Velenik. *Equilibrium Statistical Mechanics of Classical Lattice Systems: a Concrete Introduction*. 2015.

[Gri99] Geoffrey Grimmett. *Percolation*. Grundlehren der matematischen Wissenschaften. Springer, 1999.

[Hol74] R. Holley. Remarks on the FKG inequalities. *Comm. Math. Phys.*, 36(3):227–231, 1974.

[Knu97] Donald E. Knuth. *The art of computer programming, Vol. 3: Sorting and searching*. Addison Wesley Longman, 2nd edition, 1997.

[Kyt19] Kalle Kytölä. Probability theory, 2019. Lecture notes of course MS-E1600, Aalto University. `https://math.aalto.fi/~kkytola/files_KK/ProbaTh2019/ProbaTh-2019.pdf`.

[KZ87] M Kardar and Y-C. Zhang. Scaling of directed polymers in random media. *Phys. Rev. Lett.*, 58(20):2087–2090, 1987.

[Lig05] Thomas M. Liggett. *Interacting Particle Systems*. Classics in Mathematics. Springer, 2005.

[LL10] Gregory F. Lawler and Vlada Limic. *Random Walk: A Modern Introduction*. Cambridge Studies in Advanced Mathematics. Cambridge Univ. Press, 2010.

[Wer09] Wendelin Werner. *Percolation et modèle d'Ising*. Cours spécialisés. Societé Mathématique de France, 2009.

[Wil91] David Williams. *Probability with Martingales*. Cambridge Univ. Press, 1991.