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 and 2017.

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 for 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 basic 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 with some imagination the reader will have no trouble finding 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

As a preparation, we recall some facts about permutations and fix 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 you should think about.

Exercise I.1. Let σ be a uniformly distributed random permutation of the set $\{1, 2, \ldots, n\}$. 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.

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

$$\mathsf{P}\Big[\bigcup_{j=1}^{n} E_j\Big] = -\sum_{\substack{J\subset\{1,\ldots,n\}\\ J\neq\emptyset}} (-1)^{\#J}\, \mathsf{P}\Big[\bigcap_{j\in J} E_j\Big]$$

$$= \sum_{1\le j_1\le n} \mathsf{P}[E_{j_1}] - \sum_{1\le j_1<j_2\le n} \mathsf{P}[E_{j_1}\cap E_{j_2}] + \sum_{1\le j_1<j_2<j_3\le n} \mathsf{P}[E_{j_1}\cap E_{j_2}\cap E_{j_3}] - \cdots .$$

(b) What is the probability that a uniformly distributed random permutation of the set $\{1, 2, \ldots, n\}$ 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 $E_j = \{\text{the point } j \text{ is a fixed point}\}$.

1.1. Sorting algorithms

A common problem in programming and computer science is to sort a list of n elements, often for n very large. There are various sorting algorithms for this purpose, e.g., "Quicksort", "Merge sort", "Insertion sort", ... 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 performance of an algorithm is measured by its use of computational resources, mainly by the amount of processor time used before the algorithm outputs a sorted list (one could also consider other aspects such as memory requirement etc.). For sorting algorithms, the required processor time is (usually) well approximated by the number C of pairwise comparisons that were needed.

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, 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 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.

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

[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.

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 element $\mathbf{a} = (a_1, \ldots, a_n)$ 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 the 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,

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

This represents the average case performance of QuickSort. 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 $k \in \{0, 1, \ldots, n-1\}$ 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

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

One can similarly show that

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

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.

Exercise I.3. Check that m_n given by (I.4) satisfies the recursion (I.3)

Exercise I.4. Check that the asymptotical 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

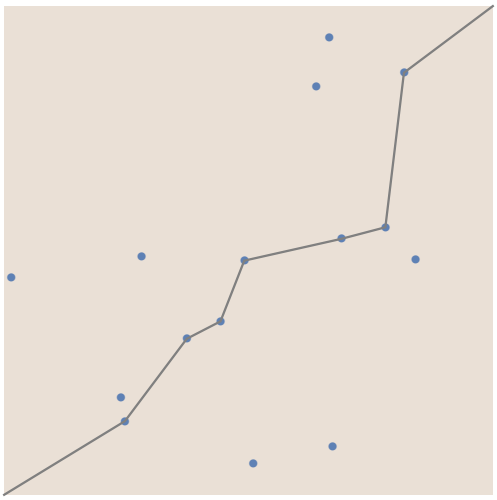
As our second example of random permutations in interesting applications, we will discuss a model of disordered material, with an interface in it.

The model we will consider is visualized in Figure I.1, 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.1.

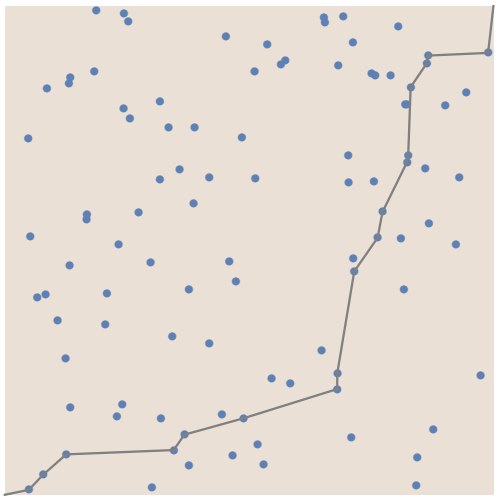
This is a simple model used in disordered materials physics. The easiest interpretation is that the square S represents a piece of some material, and the the random points $Z_1, \ldots, Z_n \in S$ represent impurities in this material. The material is considered to be weaker at the locations of the impurities, 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 explain that L_n arises in a simple way from random permutations.

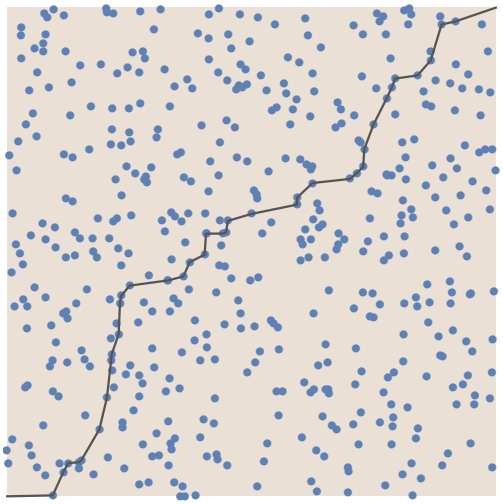
Fix the random points $Z_1 = (X_1, Y_1)$, $\ldots$, $Z_n = (X_n, Y_n)$. It is convenient to order these points by the values of their x-coordinates, so let us agree to relabel them so that $0 < X_1 < X_2 < \cdots < X_n < 1$ (note that the x-coordinate values are almost surely different, so no equalities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



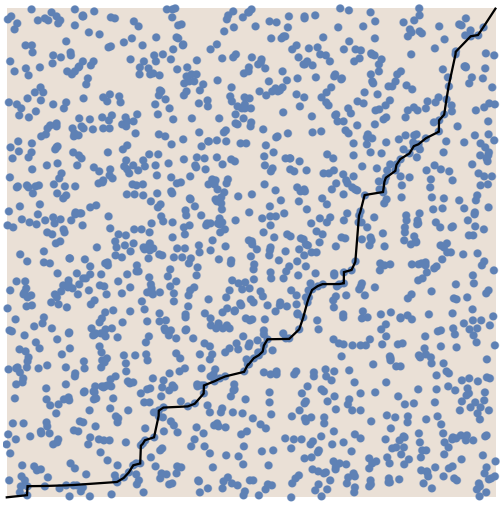
(a) A sample of $n = 15$ points in a square, and a NE path through $\ell = 7$ of them.



(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.1. 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 the uniform random permutations $\sigma \in \mathfrak{S}_n$,

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

Recall that the quantity L_n is of importance, for example as the energy advantage of the interface due to the impurities. 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. For $n \in \mathbb{N}$, consider a uniform random permutation $\sigma \in \mathfrak{S}_n$. 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 $k! \ge k^k e^{-k}$.
Hint: Compare $\log(k!) = \sum_{j=1}^{k} \log(j)$ and the integral $\int \log(x)\,\mathrm{d}x$.

(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 perhaps 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.

(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 $1, 2, \dots, n$, 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. The fully ordered deck is described by the delta-measure δ_{id} at the identity permutation $\mathrm{id} \in \mathfrak{S}_n$. This, or any other 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 common 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 of

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 a deck of cards.

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

Exercise I.8. 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:

How many shuffles are needed?

---

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.

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.

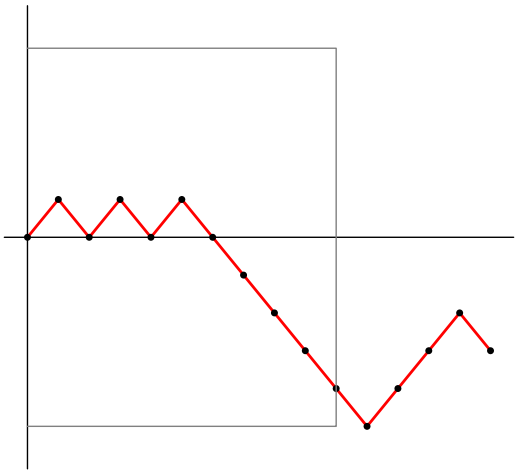
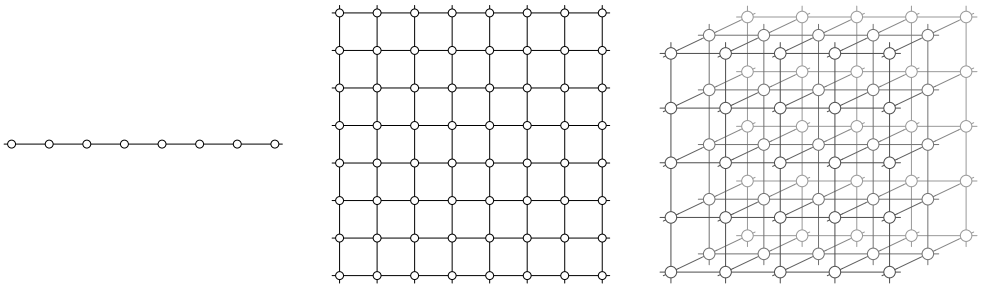


FIGURE I.2. A random walk on $\mathbb{Z}$.



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

FIGURE I.3. 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 good 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.3, 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 on $\mathbb{R}^d$ 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

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

The $\#\mathcal{N} = 2d$ 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 $\mathcal{N}$ 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.2. This should be a very familiar object, but even simple questions about X may require some thinking. The reader will surely have no trouble verifying that the law of the position $X(t)$ at a fixed time t is the following.

Exercise I.9. 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

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

A little bit more thinking is needed for example for the following.

Exercise I.10. 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 $p = \mathsf{P}[\tau < \infty]$. Provided that $\tau < \infty$, 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. Thus realizing L as a sum of the independent trial returns with a probability of success p, we see that $\mathsf{P}[L \geq k] = p^{k-1}$ for all $k \in \mathbb{Z}_{>0}$. The expected value of L is now 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,

$$L_\lambda(\vec{x}) \le \sum_{t=0}^{\infty} \lambda^t = \frac{1}{1-\lambda} < +\infty \qquad \Longrightarrow \qquad G_\lambda(\vec{x}) = \mathsf{E}\big[L_\lambda(\vec{x})\big] \le \frac{1}{1-\lambda} < +\infty.$$

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

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

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

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

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

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

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

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

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 $d \leq 2$ and converges if $d > 2$.

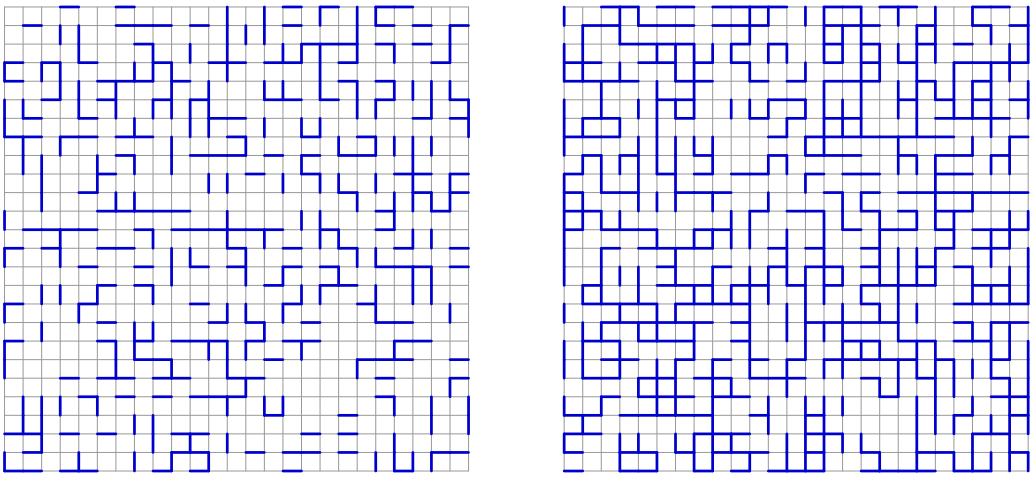
We conclude that $\mathsf{E}[L] = +\infty$ if $d \leq 2$ $\mathsf{E}[L] < +\infty$ if $d > 2$. By the characterization (I.10) this says that X is recurrent if $d \leq 2$ and transient if $d > 2$. □

[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 $C_1 \, g(x) \leq f(x) \leq C_2 \, g(x)$ for all x. For example here one specifically has $\frac{1}{2\sqrt{2}}\alpha^2 \leq 1 - \cos(\alpha) \leq \frac{1}{2}\alpha^2$ 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.

Alternatively, we could be modeling for example a forest fire (or a disease, or something else) in such a way that burning locations (or infected locations) are able to transmit fire (or disease) to nearby locations provided the passage between them is open. Again, fire (or disease) then spreads 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. 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. 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.

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

Exercise II.2. For the bond percolation on $\mathbb{Z}^d$ with any parameter $p < 1$, show that

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

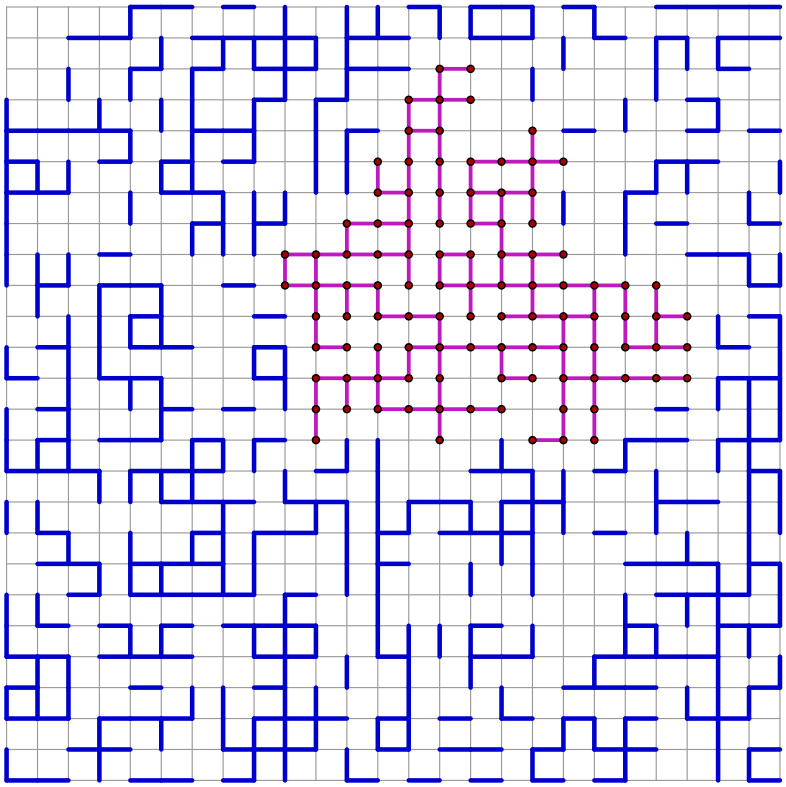


FIGURE II.2. A connected component in a bond percolation configuration on $\mathbb{Z}^2$.

1.1. 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, because of independence of the bonds. We make this intuition precise here, by our first (simple) monotonicity result.

The set of configurations $\Omega = \{0,1\}^{\mathrm{E}(\mathbb{Z}^d)}$ carries a natural partial order $\preceq$. Namely, 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 $\omega_e \le \omega'_e$.

A 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 $f(\omega) \le f(\omega')$ whenever $\omega \preceq \omega'$. 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.5 below.

Proposition II.1. *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. All of the following are increasing functions of p:

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

etc.

The tool we use is a coupling of the probability measures P_p for different values of $p \in [0,1]$.

Definition II.1. A *coupling* of two probability space $(\Omega_1, \mathcal{F}_1, \mathsf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ is a new probability space $(\widetilde{\Omega}, \widetilde{\mathsf{P}}, \widetilde{\mathcal{F}})$ equipped with two random variables $X_1 \colon \widetilde{\Omega} \to \Omega_1$ and $X_2 \colon \widetilde{\Omega} \to \Omega_1$, such that the law of X_1 is P_1 and the law of X_2 is P_2.

Remark II.3. Another common definition of a coupling of $(\Omega_1, \mathcal{F}_1, \mathsf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ is a probability space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathsf{P}_{\text{cpl}})$ (the sample space is a Carthesian product, and the σ-algebra is the product σ-algebra) such that the two marginals are P_1 and P_2, i.e.,

$$\forall E_1 \in \mathcal{F}_1 : \qquad \mathsf{P}_{\text{cpl}}\big[E_1 \times \Omega_2\big] = \mathsf{P}_1\big[E_1\big]$$
$$\forall E_2 \in \mathcal{F}_2 : \qquad \mathsf{P}_{\text{cpl}}\big[\Omega_1 \times E_2\big] = \mathsf{P}_2\big[E_2\big].$$

These definitions are essentially equivalent. This second definition is indeed a special case of the first (take $\widetilde{\Omega} = \Omega_1 \times \Omega_2$ and X_1, X_2 the projections to the two components). Conversely, in the first definition, the law of the pair (X_1, X_2) is a probability measure on $\Omega_1 \times \Omega_2$ which satisfies the requirements of P_{cpl}.

Remark II.4. The product measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathsf{P}_1 \otimes \mathsf{P}_2)$ is a coupling, but not a very useful one as the two components are independent. Couplings of two probability measures are usually best if some relation holds between the two components, which allows to make comparisons between the original measures.

For monotonicity we construct a coupling of the percolation measures P_{p_1} and P_{p_2}, with two different parameters $p_1 \leq p_2$. The key is that the two components of the coupling can be made respect the partial order $\preceq$ on Ω.

Proposition II.5. *Let $0 \leq p_1 \leq p_2 \leq 1$. There exists a probability space $(\widetilde{\Omega}, \widetilde{\mathsf{P}}, \widetilde{\mathcal{F}})$ 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 $\omega^{(1)} \preceq \omega^{(2)}$.*

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 $\widetilde{\mathcal{F}}$ 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. Clearly $\widetilde{\mathsf{P}}[\omega_e^{(1)} = 1] = p_1$ by uniformity of U_e, 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 we have $\omega^{(2)} \sim \mathsf{P}_{p_2}$. □

Proof of Proposition II.1. Use the above monotone coupling to write $\mathsf{E}_p[f] = \widetilde{\mathsf{E}}\big[f(\omega^{(p)})\big]$, where $\widetilde{\mathsf{E}}$ is expected value w.r.t. $\widetilde{\mathsf{P}}$. If $p < p'$, we have $\omega^{(p)} \preceq \omega^{(p')}$, and thus $f(\omega^{(p)}) \le f(\omega^{(p')})$, since f is increasing. We get the conclusion

$$\mathsf{E}_p[f] = \widetilde{\mathsf{E}}\big[f(\omega^{(p)})\big] \le \widetilde{\mathsf{E}}\big[f(\omega^{(p')})\big] = \mathsf{E}_{p'}[f].$$

□

1.2. Phase transition by a zero-one law

We will show that there is the following phase transition in the parameter p, at a critical value[1] $p = p_c \in [0, 1]$.

Theorem II.6. *Let $d \in \mathbb{Z}_{>0}$. Then there exists a $p_c \in [0, 1]$ 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 $\mathcal{T}$, 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 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.7. The theorem does not tell whether an infinite cluster exists at the critical point $p = p_c$ or not. Note also the possibility that the transition happens at a trivial place $p_c = 0$ or $p_c = 1$. Fortunately, it turns out that such a triviality does not happen except in one dimension, $d = 1$.

Example II.8. As the simplest example, let $d = 1$, i.e., consider 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 are only countable number of sites in $\mathbb{Z}$, almost surely no component 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 $p_c(\mathbb{Z}^d) \in (0, 1)$.

Theorem II.9. *Let $d > 1$. Then the critical point of percolation on $\mathbb{Z}^d$ is nontrivial, $p_c \in (0, 1)$.*

[1]Note, that 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 $\mathbb{Z}^d$.

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.10. *We have*

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

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 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.9 will be done in two parts: we first show that $p_c > 0$ and then that $p_c < 1$.

1.3. 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.2. *For $p < \frac{1}{2d}$ we have $\theta(p) = 0$.*

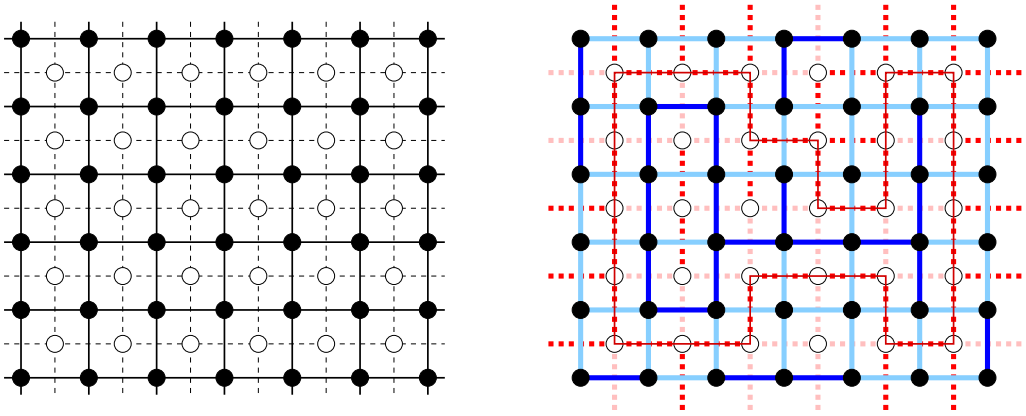
Proof. The number of self-avoiding paths of length n from the origin can obviously be bounded above by $\#\Gamma_n \leq (2d)^n$. We then note that for any n

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

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

1.4. Percolation for large p in two dimensions

We will first consider 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 bonds $\{q, r\}$ crosses exactly one bonds $\{z, w\}$ 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) 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 closed path (starting point and end point coincide) of the dual lattice $(\mathbb{Z}+\frac{1}{2})^2$, modulo cyclic reparametrization (any point on the closed path could be chosen as the starting point, and we identify such paths). Indeed, 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, we use $\#\Gamma^\circ_n \leq \frac{n}{2} 4^n$. 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.3. *For $p > \frac{3}{4} + \frac{1}{4\sqrt{2}}$ we have $\theta(p) > 0$.*

Proof. We estimate

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

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

1.5. Proof of non-triviality of the critical point

We now combine the above results to prove Theorem II.9.

Proof of Theorem II.9. By Lemma II.2 we have $\theta(p) = 0$ for $p < \frac{1}{2d}$. By Lemma II.10 this implies $p_c \ge \frac{1}{2d}$.

It remains to show that $p_c < 1$. For $d \ge 2$, 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.3 we have $\theta(p) > 0$ for $p > \frac{3}{4} + \frac{1}{4\sqrt{2}}$. By Lemma II.10 this implies $p_c \le \frac{3}{4} + \frac{1}{4\sqrt{2}} < 1$. □

2. Percolation on a regular infinite tree

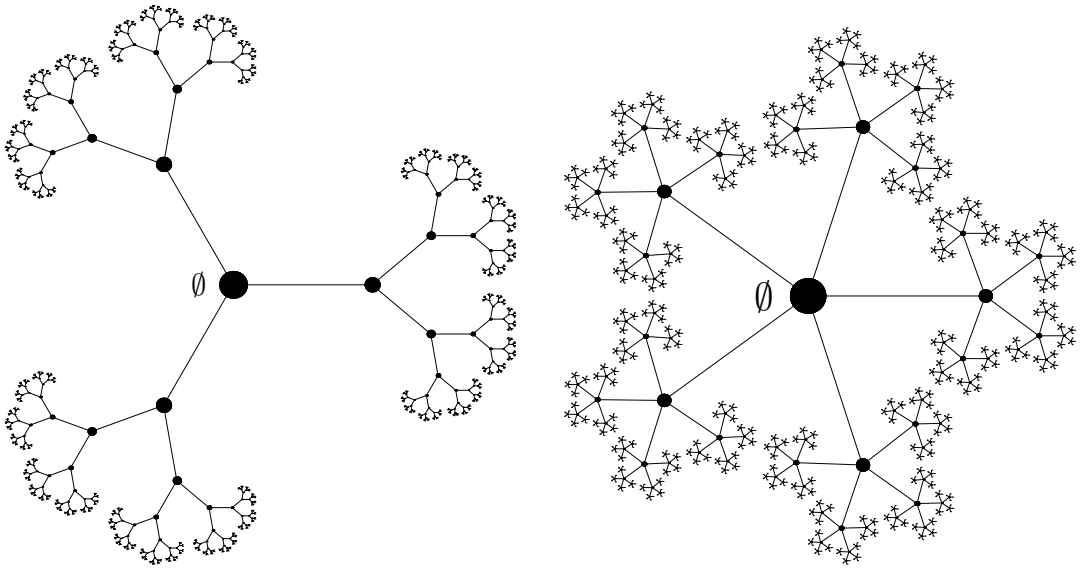
The model of percolation can be defined on any graph. On a regular infinite tree $\mathcal{T}$ of degree B, depicted in Figure II.4, percolation becomes particularly simple. In fact the analysis is essentially the the same as that of a Galton-Watson branching process.

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, $B \ge 3$. Percolation on a regular infinite tree $\mathcal{T}$ of any degree $B \ge 3$ behaves in some sense like percolation on $\mathbb{Z}^d$ for a very high dimension d.

Formally, the tree $\mathcal{T}$ can be defined as follows. The set of its sites is the set of finite words $(b_1, b_2, \dots, b_\ell)$ of any possible length $\ell \in \mathbb{Z}_{\ge 0}$, with first letter $b_1 \in \{1, 2, \dots, B\}$ and other letters $b_j \in \{1, 2, \dots, B-1\}$ for $j = 2, 3, \dots, \ell$. 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.3. Consider percolation with parameter $p \in [0, 1]$ on the regular infinite tree $\mathcal{T}$ of degree $B \ge 3$. 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.4. Consider percolation with parameter p on the regular infinite tree $\mathcal{T}$ of degree $B \ge 3$. Denote by $\mathcal{C}_\emptyset(r)$ the set of sites at distance r from the root which are connected to



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

FIGURE II.4. Regular infinite trees.

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.1 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. *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,*

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

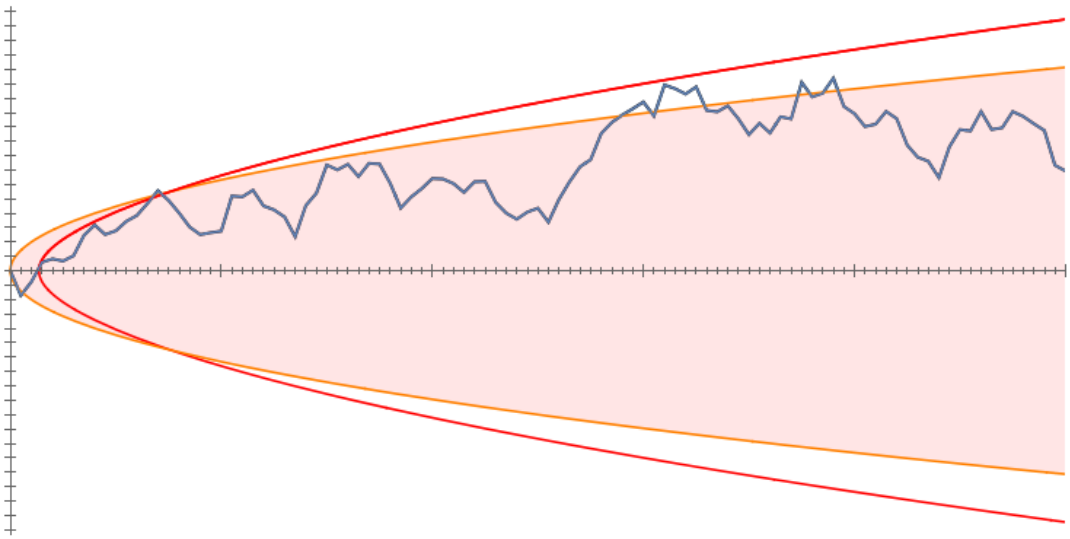


FIGURE III.1. A random walk with Gaussian steps. The shaded region bounded by the orange curves is $|x| \leq \sqrt{2n}$, and the red curves are the law of iterated logarithm tresholds $\pm\sqrt{2}\lambda(n) = \pm\sqrt{2n\log\log(n)}$.

1. Auxiliary estimates

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

Lemma III.1. *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.

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. 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.2. *For any $x > 0$, we have*

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

For any $C < \frac{1}{\sqrt{2\pi}}$ there exists $x_0 > 0$ such that for all $x > x_0$

$$\mathsf{P}\big[X_n > x\big] \ge C\frac{\sqrt{n}}{x}e^{-\frac{1}{2n}x^2}. \tag{III.5}$$

Also, we have

$$\mathsf{P}\big[|X_n| > x\big] \leq 2e^{-\frac{1}{2n}x^2}. \tag{III.6}$$

Exercise III.2. Let now $(X_n)_{n\in\mathbb{Z}_{\geq 0}}$ 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

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

Exercise III.3. Let λ be the function

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

Let $\alpha > 1$ and for $k \in \mathbb{N}$ let $n_k = \lfloor \alpha^k \rfloor$. 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,

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

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.2.

Lemma III.2. *For any $\alpha > 1$ denote by $n_k = \lfloor \alpha^k \rfloor$ the integer part of α^k. Then we have, for any $\beta > \sqrt{2}$,*

$$\mathsf{P}\Big[\limsup_{k\to\infty} \frac{X_{n_k}}{\lambda(n_k)} \leq \beta\Big] = 1.$$

Proof. By the inequality (III.4) we have

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

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

$$\sum_{k=0}^{\infty} \mathsf{P}\Big[\frac{X_{n_k}}{\lambda(n_k)} > \beta\Big] \leq \sum_{k=0}^{\infty} \log(n_k)^{-\frac{\beta^2}{2}} \leq \sum_{k=0}^{\infty} (ak+c)^{-\frac{\beta^2}{2}} < \infty.$$

Borel-Cantelli Lemma (Lemma E.1) then guarantees that $\frac{X_{n_k}}{\lambda(n_k)} > \beta$ only for finitely many k, or in other words,

$$\limsup_{k\to\infty} \frac{X_{n_k}}{\lambda(n_k)} \le \beta.$$

□

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

Lemma III.3 (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

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

□

We are then ready to prove the upper bound along the entire sequence by filling in the gaps.

Proposition III.4. *We have*

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

Proof. Let $\beta > \sqrt{2}$ and $\varepsilon > 0$. We will use Lemma III.2 for a subsequence $n_k = \lfloor \alpha^k \rfloor$ with a suitably chosen $\alpha > 1$, and we will fill in the gaps up to tolerance ε with Lévy's inequality.

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

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

For this, set $S_m = X_{n_k+m} - X_{n_k}$ and $n = n_{k+1} - n_k$, and denote

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

Lemma III.3 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.6)

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

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 $k \to \infty$. Therefore, if we choose $\alpha < 1 + \frac{\varepsilon^2}{2}$ (for definiteness, e.g., $\alpha = 1 + \frac{\varepsilon^2}{4}$), we again have summable probabilities

$$\sum_{k=0}^{\infty} \mathsf{P}\Big[\max_{n_k < i \le n_{k+1}} |X_i - X_{n_k}| > 2\varepsilon\lambda(n_k)\Big] < \infty.$$

Recall from Lemma III.2 that we have $|X_{n_k}|/\lambda(n_k) \le \beta$ for all but finitely many k, and note in addition that by the summability of the above probabilities and Borel-Cantelli lemma we have that $\max_{n_k<i\le n_{k+1}} |X_i - X_{n_k}| \le 2\varepsilon\lambda(n_k)$ for all but finitely many k. In fact, both conditions hold simultaneously for all but finitely many k, and then for $n_k \le i \le n_{k+1}$, we have

$$\frac{|X_i|}{\lambda(i)} \le \frac{|X_{n_k}| + |X_i - X_{n_k}|}{\lambda(n_k)} \le \beta + 2\varepsilon.$$

This says that

$$\limsup_{i\to\infty} \frac{|X_i|}{\lambda(i)} \le \beta + 2\varepsilon.$$

Since $\beta > \sqrt{2}$ and $\varepsilon > 0$ were arbitrary, we obtain that in fact

$$\limsup_{i\to\infty} \frac{|X_i|}{\lambda(i)} \le \sqrt{2}.$$

□

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 smallest 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.5), 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.5. *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.5) 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.2 (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

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

Here, note that

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

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.4 and III.5 yields Theorem III.1.

Exercise III.4. Prove Theorem III.1 for the simple symmetric random walk with ± 1 steps.

Lecture IV

Weak convergence on the real line

Weak convergence (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 also the underlying idea is practically motivated: random variables are said to converge if all sufficiently well behaved observables of them converge. In the definition of weak convergence, sufficiently well behaved observables are taken to mean bounded continuous functions on the appropriate space.

We 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 practise, 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 **??**.

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 measure about the systems converges. We idealize the situation by declaring that the 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 be expressed in terms of such expected values. For weak convergence specifically, as the idealization of sufficiently well behaved observables, we take the bounded continuous functions (continuity guarantees in particular Borel-measurability and boundedness guarantees the existence of expected value).

Definition IV.1. A sequence $(\nu_n)_{n=1}^{\infty}$ 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.1. The condition (IV.1) uniquely characterizes the limit, a Borel probability measure ν_∞ on $\mathbb{R}$ (this follows from Theorem IV.3 below, see also **??**).

Remark IV.2. 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 $(\Omega_n, \mathcal{F}_n, \mathsf{P}_n)$, 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.3. *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*

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

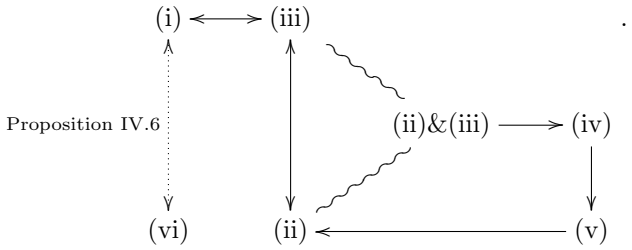
(ii) *For all open subsets $U \subset \mathbb{R}$ we have $\nu(U) \le \liminf_{n\to\infty} \nu_n(U)$.*
(iii) *For all closed subsets $F \subset \mathbb{R}$ we have $\nu(F) \ge \limsup_{n\to\infty} \nu_n(F)$.*
(iv) *For all Borel subsets $B \subset \mathbb{R}$ such that $\nu[\partial B] = 0$, we have $\nu_n(B) \to \nu(B)$.*
(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.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 **??**, since the proof remains identical in the more general context of probability measures on a general metric space $(\mathfrak{X}, \varrho)$. Below we therefore prove the implications (iv) $\Rightarrow$ (v) and (v) $\Rightarrow$ (ii), and separately in Proposition IV.6 a slightly strengthened version of the equaivalence (i) $\Leftrightarrow$ (vi).

Proof of implication (iv) $\Rightarrow$ (v) in Theorem IV.3. Assume condition (iv), i.e. $\nu_n[B] \to \nu[B]$ whenever $B \subset \mathbb{R}$ is a Borel subset such that $\nu[\partial B] = 0$. 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) $\Rightarrow$ (ii) in Theorem IV.3. 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

$$\nu\big[(a, b)\big] - \varepsilon \leq \nu\big[(a', b')\big] = \lim_{b'' \uparrow b'} \big(F(b'') - F(a')\big) = F(b') - F(a')$$

$$\overset{\text{(v)}}{=} \lim_{n\to\infty} \big(F_n(b'') - F_n(a')\big) = \lim_{n\to\infty} \nu_n\big[(a', b']\big] \leq \liminf_{n\to\infty} \nu_n\big[(a, b)\big].$$

We then consider a general open subset $V \subset \mathbb{R}$. It is an easy fact that the open set V is a disjoint union of at most countably many open intervals V_j (Exercise H.2). The case of a finite union is very easy, so assume that

$$V = \bigcup_{j=1}^{\infty} V_j,$$

where each V_j is an open interval, and the intervals V_j, $j \in \mathbb{Z}_{>0}$, are disjoint. Fix $\varepsilon > 0$. The case of a single open interval treated above implies that for any j we have

$$\nu\big[V_j\big] - 2^{-j}\,\varepsilon \leq \liminf_{n\to\infty} \nu_n\big[V_j\big].$$

Summing over j and using Fatou's lemma (Lemma A.8 in Appendix A) we get

$$\nu\big[V\big] - \varepsilon = \sum_{j=1}^{\infty} \Big(\nu\big[V_j\big] - 2^{-j}\varepsilon\Big) \le \sum_{j=1}^{\infty} \liminf_{n\to\infty} \nu_n\big[V_j\big]$$
$$\le \liminf_{n\to\infty} \sum_{j=1}^{\infty} \nu_n\big[V_j\big] = \liminf_{n\to\infty} \nu_n\big[V\big].$$

Since $\varepsilon > 0$ was arbitrary, (ii) follows. □

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

Exercise IV.1.

(a) Let $P \sim \text{Poisson}(\lambda)$, i.e., $\mathsf{P}\big[P = k\big] = \frac{\lambda^k}{k!} e^{-\lambda}$ for $k \in \mathbb{Z}_{\ge 0}$. 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 $k = 0, 1, \dots, n$. 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. 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 $M_n = \max_{1\le j\le n} X_j$, and consider the shifted maxima $R_n = M_n - \frac{1}{\lambda}\log(n)$. Calculate the cumulative distribution functions $F_n(x) = \mathsf{P}[R_n \le x]$, $n \in \mathbb{Z}_{>0}$, 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).

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.2. 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.3. *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.5. *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.3. We later return to this in a more general case. □

Exercise IV.3 (Convergent subsequence for a tight cdf sequence)**.** Let $F_n\colon \mathbb{R} \to [0,1]$ be cumulative distribution functions, for $n \in \mathbb{N}$.

(a) Show that the sequence $(F_n)_{n\in\mathbb{N}}$ has a subsequence $(F_{n_k})_{k\in\mathbb{N}}$, which converges pointwise to an increasing function $F\colon \mathbb{R} \to [0,1]$ at all continuity points of F.

(b) Suppose furthermore that the probability measures associated to the cumulative distribution functions F_n constitute a tight family. Show that in part (a), one can choose the subsequential limit F to be a cumulative distribution function.[1]

Hint: First form a subsequence which converges in a countable dense subset of $\mathbb{R}$, using the compactness of the unit interval and a diagonal extraction. Construct from this a right-continuous function F, and show, that F it is the desired subsequential limit at its points of continuity.

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.3. 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 **??** (random variables with values in general metric spaces).

Proposition IV.6. *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.7. From this proposition, we obtain in particular the equivalence of conditions (i) and (vi) in Theorem IV.3: the "only if" part shows that (i) ⇒ (vi), and the "if" part shows that (vi) ⇒ (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. The condition is thus more convenient to verify in practice.

Remark IV.8. 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. The strategy is common, but very important — we 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.*

[1] Which functions can serve as cumulative distribution functions is addressed in Proposition A.4 in Appendix A.

(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.9).

Proof. "Only if": Assume the weak convergence $\nu_n \xrightarrow{\mathrm{w}} \nu$. Since $x \mapsto e^{\mathrm{i}\theta x}$ 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 to the characteristic function φ of the measure ν, which is continuous at the origin.

"If": 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 model 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 are in the same state. The interaction is "ferromagnetic", in the sense that the probabilities of configurations are higher when there is more agreement 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. We also show that near the critical temperature of transition between the above paramagnetic and ferromagnetic phases, the magnetization has a power law type dependence on the parameters of the model — thus establishing what is known as critical exponents in physics.

The Curie–Weiss model disregards all possible spacial structure of the magnetic material — none of the spins are thought to be near or far from each other, but all pairs interact alike with each other. This obviously is a simplification in the model, which makes the model easier to study mathematically. Fortunately, even with such a simplification, there is a correct phase transition between the ferromagnetic and paramagnetic behaviors. Unfortunately, however, the values of the critical exponents turn out not to be the correct ones for magnetic materials in three, two, or one dimensions. The famous Ising model of ferromagnetism, which we will study in a later lecture, is essentially obtained by adding spacial structure to the Curie–Weiss model. It has also the correct critical behavior to account for materials known as uniaxial ferromagnets, but it becomes significantly harder to analyze mathematically. Because of the lack of spacial structure, the Curie–Weiss model is in physics called 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 interpretations 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 models has N *spins* (representing some sort of elementary magnetic units), each with two possible values $+1$ and -1. The sample space will therefore 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 $\mu_\beta^{(N)}$ on the finite sample space Ω_N, such that the individual outcomes have probabilities

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

and the partition function $Z_N(\beta)$ normalizes the total probability to one,

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

The parameter $\beta > 0$ in Boltzmann distributions is (proportional to) $\frac{1}{T}$, where T is the temperature of the material. We will call β the *inverse temperature*. With a suitable choice of units (of energy), we may assume $J = 1$.

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 the 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, but note that unlike common situations for for example laws of large numbers or central limit theorems, the terms in the sum are <u>not</u> independent — this is exactly what the ferromagnetic interactions are 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 phase transition at the critical value

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

of the inverse temperature parameter 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, the first theorem states that in the thermodynamical limit without external magnetic field, in the paramagnetic phase $\beta < \beta_c$ the empirical magnetization concenrates to the value zero, whereas in the ferromagnetic phase $\beta > \beta_c$ it concentrates at non-zero values $\pm\bar{m}(\beta)$ (*spontaneous magnetization*), with a random sign. The second theorem says 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, $B \downarrow 0$, is zero in the paramagnetic phase and a non-zero constant $\bar{m}(\beta) > 0$ in the ferromagnetic phase. The first two results thus express a qualitative phase transition. The third theorem concerns 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. *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 a spontaneous magnetization $\bar{m} = \bar{m}(\beta) > 0$ such that $\mathsf{P}[M_\infty = +\bar{m}] = \frac{1}{2}$ and $\mathsf{P}[M_\infty = -\bar{m}] = \frac{1}{2}$.*

Theorem V.2. *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.

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

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

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

1.1. 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)

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

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

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

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

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 $F_N = -\frac{1}{\beta}\log(Z_N)$, or rather the contribution $\frac{1}{N}F_N$ per spin, 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 Gibbs free energy $f = \lim_{N\to\infty} \frac{1}{N}F_N$ is a Legendre transform of the Helmholtz free energy g.

Proposition V.4. *Let $\beta > 0$ and $B \in \mathbb{R}$. As $N \to \infty$, we have*

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

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

Let us start from a lower bound for the Gibbs free energy, which is 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

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

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 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 $\inf_{m \in [-1,1]} \left(g(\beta, m) - Bm\right)$. □

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

$$\mu_{\beta,B}^{(N)}\big[M_N \approx m\big] \sim \exp\left[-N\beta\big(g(\beta, m) - Bm - f(\beta, B)\big)\right].$$

A precise formulation of this is given below.

Proposition V.5. *For any open set $A \subset [-1, 1]$ we have*

$$\lim_{N \to \infty} \frac{-1}{N} \log\left(\mu_{\beta,B}^{(N)}\big[M_N \in A\big]\right) = \beta \left[\inf_{m \in A} \left(g(\beta, m) - Bm\right) - f(\beta, h)\right].$$

Proof. The probability that $M_N \in A$ is

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

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

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

The asserted formula follows by taking logarithms and using also the previous result 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 startegy. 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. *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 $\mathcal{N} = \{x \in \mathbb{R} \,|\, \phi(x) = 0\}$ of zeroes of the rate function, i.e., we have $\mathsf{P}[X \in \mathcal{N}] = 1$.

Proof. Define the opens sets $A_\varepsilon = \{x \in \mathbb{R} \,|\, \phi(x) > \varepsilon\}$ for $\varepsilon > 0$. Since $\inf_{x\in A_\varepsilon} \phi(x) \geq \varepsilon$ (by continuity), we infer from the large deviations upper bound that $\mathsf{P}[X_n \in A_\varepsilon] \to 0$ as $n \to \infty$ (these probabilities tend to zero exponentially fast). Thus by the characterization (ii) of Theorem IV.3 we have $\mathsf{P}[X \in A_\varepsilon] \leq \liminf \mathsf{P}[X_n \in A_\varepsilon] = 0$. Since the complement of the set $\mathcal{N}$ is a countable union of sets of the form A_ε, we get the desired conclusion

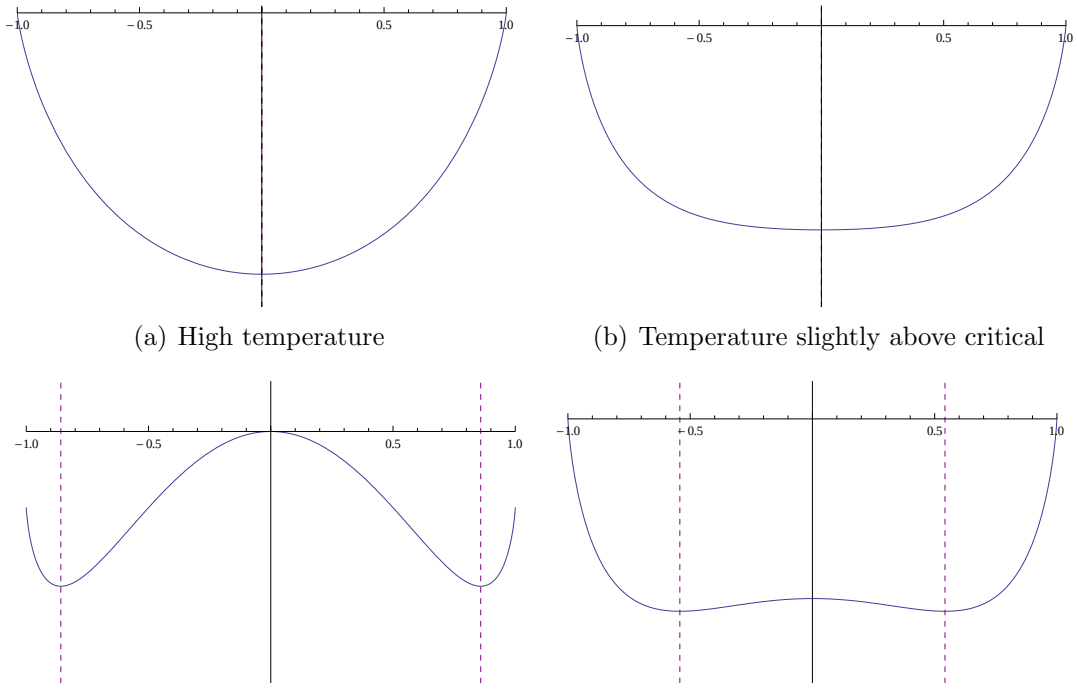
$$\mathsf{P}[X \notin N] = \mathsf{P}\Big[X \in \bigcup_{k=1}^{\infty} A_{1/k}\Big] \leq \sum_{k=1}^{\infty} \underbrace{\mathsf{P}[X \in A_{1/k}]}_{=0} = 0.$$

□

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. *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 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 $m = -\bar{m}(\beta)$.*



(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 minimun 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): $\beta = 0.75$.

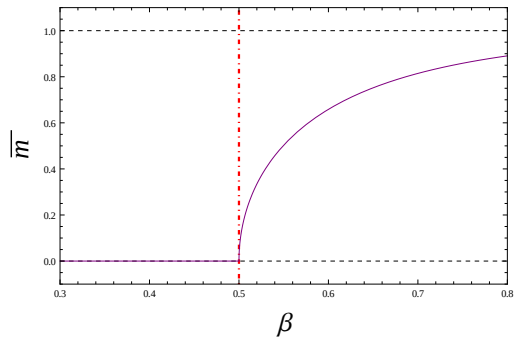
Proof. The continuity of the function ϕ is clear from the continuity of the function I on the interval $[-1, 1]$. Proposition V.4 implies that $f(\beta, 0)$ is the minimum of the function $m \mapsto g(\beta, m)$, so the minimum of ϕ is 0, and part (a) follows.

Moreover, ϕ is $\mathcal{C}^\infty$ on the open interval $(-1, 1)$, and it is an even function, $\phi(-m) = \phi(m)$. Thus clearly $\phi'(0) = 0$. An easy calculation gives $I''(m) = \frac{1}{1-m^2}$, and we get $\phi''(m) = \frac{1}{1-m^2} - 2\beta$. We see that for $\beta \leq \frac{1}{2}$ we have $\phi''(m) \geq 0$, and the only minimum is at $m = 0$. Correspondingly for $\beta > \frac{1}{2}$, the second derivative is negative on an interval around zero and non-negative elsewhere, so the function ϕ has a local maximum at zero and minima at the other two zeroes of the derivative. □

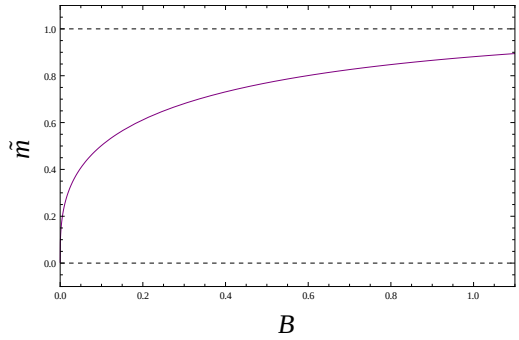
Proof of Theorem V.1. We wish to show that the empirical magnetizations M_N converge weakly towards the asserted law. Note first that the sequence $(M_N)_{N\in\mathbb{N}}$ of random variables is tight (recall Lecture IV, Section 3), because the values of each M_N are on the interval $[-1, 1]$. Theorem IV.5 implies that any subsequence $(M_{N_k})_{k\in\mathbb{N}}$ has some further subsequence which is convergent. It suffices to show that the limit of any convergent subsequence is of the asserted form.

In the case $\beta \leq \beta_c = \frac{1}{2}$, is $M_{N_k} \xrightarrow{\text{w}} M_\infty$, 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 ase $\beta > \beta_c = \frac{1}{2}$, if $M_{N_k} \xrightarrow{\text{w}} M_\infty$, then by Lemma V.6, Proposition V.5 and Lemma V.7(c) we have that $\mathsf{P}[M_\infty \in \{-\bar{m}, +\bar{m}\}] = 1$. Moreover, when $B = 0$ we obviously have the symmetry $\mathsf{P}[M_N < 0] = \mathsf{P}[M_N > 0]$, which implies the corresponding symmetry



(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): $B \mapsto \tilde{m}(\beta_c, B)$.

for the limit random variable M_∞, and therefore $\mathsf{P}[M_\infty = +\bar{m}] = \frac{1}{2} = \mathsf{P}[M_\infty = -\bar{m}]$. 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 Helmholz free energy is at the unique positive solution $\bar{m}(\beta, B) > 0$ of the equation $\frac{\partial}{\partial m}\big(g(\beta, m)\big) - Bm) = 0$. □

Proof of Theorem V.3. We leave the rest of the proof as Exercise V.1. □

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

(b) For fixed $B > 0$ and $\beta > 0$, let $\tilde{m} = \tilde{m}(\beta, B)$ be the unique positive solution of $\frac{\partial}{\partial m}\big(g(\beta, m) - Bm\big) = 0$. Calculate $\lim_{B \searrow 0} \frac{\tilde{m}(\frac{1}{2}, B)}{B^{1/3}}$.

Lecture VI

Weak convergence on metric spaces

We will next turn to probability theory on more general spaces. A natural generality for theory and applications will be that of Polish spaces (topological spaces which are separable and admit a complete metric). Relevant background in the topology of metric spaces is recalled in Appedix H.

We will in particular study weak convergence of probability measures. The definition of weak convergence would make sense on any topological spaces, but on Polish space we maintain the fundamental relation between tightness and precompactness (in the topology of weak convergence). Important applications of weak convergence on Polish spaces are for example the following two, which we impressionistically illustrate in Figure VI.1 and treat in later lectures:

- *Donsker's theorem* (Lecture VII):
  - Donsker's theorem states that appropriately rescaled random walks converge to Brownian motion.
  - This statement will be interpreted in terms of weak convergence of probability measures on the space $\mathcal{C}([0,T])$ of continuous functions.
- *Ising model thermodynamical limit* (Lecture VIII):
  - The idea is to define the Ising model on the infinite lattice $\mathbb{Z}^d$ by a limit of Ising models on increasing finite subgraphs of the lattice.
  - This requires weak convergence on the countable product of finite spaces $\{-1,+1\}^{\mathbb{Z}^d}$.
  - Remarkably, the phase transition in the Ising model (and other models) is related to the (non-)uniqueness of such an infinite volume limit.

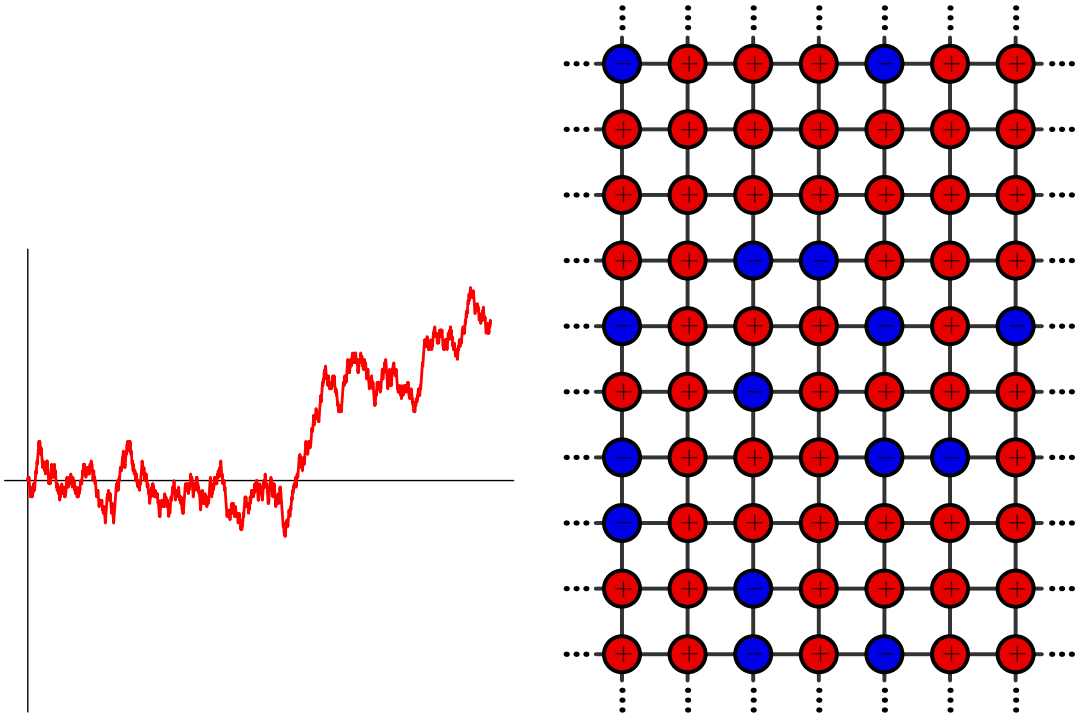
The excellent textbook [**Bil99**] is entirely devoted to the topic of weak convergence of probability measures in a general setup.

1. Weak convergence of probability measures

Recall the idea: what is considered reliably measurable by observations about a random system (in our mathematical idealization) are the expected values of bounded continuous functions. Weak convergence means exactly the convergence of all such observable quantities. For this definition, we only need $\mathfrak{X}$ to be a topological space, so that we may talk about continuous functions on it.

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

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



(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 are random objects on some spaces, which have the structure of a complete separable metric space.

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.

Remark VI.1. The weak limit ν is unique, if it exists. This follows from part (ii) of Exercise H.5 in Appendix H.

There are several equivalent characterizations of weak convergence.

Theorem VI.2 (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.3. The special case $\mathfrak{X} = \mathbb{R}$ completes the proof of Theorem IV.3 in Lecture IV, by establishing the implications whose proofs were omitted.

Proof of Theorem VI.2.

(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.

(ii)&(iii)⇒(iv): We assume the two equivalent conditions (ii) and (iii), and prove (iv). Suppose $E \in \mathcal{B}$, 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

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

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

(iv)⇒(iii): Suppose $F \subset \mathfrak{X}$ is closed. For $\delta > 0$, denote $F_\delta = \big\{x \in \mathfrak{X} \;\big|\; \varrho(x,F) \leq \delta\big\}$. Since $\mathfrak{X} \setminus F = \bigcup_{\delta>0} \big\{x \in \mathfrak{X} \;\big|\; \varrho(x,F) = \delta\big\}$ is a disjoint union, and $\nu[\mathfrak{X} \setminus F] \leq 1$, 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 $\partial F_\delta \subset \big\{x \in \mathfrak{X} \;\big|\; \varrho(x,F) = \delta\big\}$, 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 $\delta_k \searrow 0$ 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,

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

We have $F_{\delta_k} \downarrow F$ as $k \to \infty$, so the right hand side tends to $\nu[F]$, and we deduce (iii).

(i)⇒(iii): Suppose that $F \subset \mathfrak{X}$ is closed. For $\delta > 0$, denote $F_\delta = \big\{x \in \mathfrak{X} \;\big|\; \varrho(x,F) \leq \delta\big\}$. By Lemma H.3 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 $\varrho(x,F) \geq \delta$, that is,

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

We thus have

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

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

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

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

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

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

Then applying the same to the function $-f$, we can deduce that $\limsup \int_{\mathfrak{X}} f \,\mathrm{d}\nu_n \leq \int_{\mathfrak{X}} f \,\mathrm{d}\nu$. 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 $\nu[G] \leq \liminf \nu_n[G]$. Now on the probability space $(\mathfrak{X}, \mathcal{B}(\mathfrak{X}), \nu)$ consider the function $x \mapsto f(x)$ as a non-negative real-valued random variable. Apply Equation (B.2) to calculate its expected value

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

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

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

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

2. Metrizability of weak convergence

Weak convergence, Definition VI.1, says which sequences of probability measures on a metric space $\mathfrak{X}$ converge. This defines a topology on the set $\mathcal{M}_1(\mathfrak{X})$ 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 will follow from the following exercises.

Let $(\mathfrak{X}, \varrho)$ be a metric space, and let $\mathcal{M}_1(\mathfrak{X})$ be the set of Borel probability measures on $\mathfrak{X}$. For a Borel set $E \in \mathcal{B}$ 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 $\mu, \nu \in \mathcal{M}_1(\mathfrak{X})$, define

$$d(\mu, \nu) = \inf \Big\{ \delta > 0 \;\Big|\; \forall E \in \mathcal{B} :\ \mu[E] \le \nu[E^\delta] + \delta \text{ and } \nu[E] \le \mu[E^\delta] + \delta \Big\}.$$

This d is called the Lévy-Prohorov metric.

Exercise VI.1. Show that d is a metric on $\mathcal{M}_1(\mathfrak{X})$.

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

Exercise VI.2. Assume that $\nu_n \in \mathcal{M}_1(\mathfrak{X})$, $n \in \mathbb{N}$, are such that for some $\nu \in \mathcal{M}_1(\mathfrak{X})$ we have $d(\nu_n, \nu) \to 0$ 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 $n \to \infty$, $\nu[E^{\delta_n}] \searrow \nu[\bar{E}]$.
(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 d. In this setup, then, d metrizes the topology of weak convergence on the space $\mathcal{M}_1(\mathfrak{X})$ of probability measures.

Exercise VI.3. Assume that the metric space $(\mathfrak{X}, \varrho)$ is separable. Assume that $\nu_n \in \mathcal{M}_1(\mathfrak{X})$, $n \in \mathbb{N}$, are such that ν_n converges weakly to some $\nu \in \mathcal{M}_1(\mathfrak{X})$ as $n \to \infty$. Let $\varepsilon > 0$.

(a) Show that there exists a countable dense set $\big\{x_i \,\big|\, i \in \mathbb{Z}_{>0}\big\} \subset \mathfrak{X}$ and a radius $r \in (\frac{\varepsilon}{4}, \frac{\varepsilon}{2})$ such that $\nu[\partial B_r(x_i)] = 0$ for all $i \in \mathbb{Z}_{>0}$.
(b) Show that there exists some $k \in \mathbb{Z}_{>0}$ such that $\nu\big[\bigcup_{i=1}^k B_r(x_i)\big] \ge 1 - \varepsilon$.
(c) Show that there exist finitely many disjoint 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 $\mathcal{A} = \big\{\bigcup_{i \in I} A_i \,\big|\, i \subset \{1, \dots, k\}\big\}$ of subsets of $\mathfrak{X}$. Show that for any $A \in \mathcal{A}$ 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 $A \in \mathcal{A}$ and $n \ge N$.
(e) Let $E \subset \mathfrak{X}$ be a Borel set. Choose $A \in \mathcal{A}$ 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 $n \geq N$ we have $\nu_n[E] \leq \nu[E^\varepsilon] + 3\varepsilon$ and $\nu[E] \leq \nu_n[E^\varepsilon] + 2\varepsilon$.
(g) Conclude that $d(\nu_n, \nu) \to 0$ as $n \to \infty$.

Recall that a coupling of probability measures $\mu, \nu \in \mathcal{M}_1(\mathfrak{X})$ is a probability measure λ on $\mathfrak{X} \times \mathfrak{X}$ such that for all $E \in \mathcal{B}$ 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 d on the space $\mathcal{M}_1(\mathfrak{X})$ 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. Assume that there exists a coupling λ of μ and ν in which

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

Show that then $d(\mu, \nu) \leq \varepsilon$.

3. Tightness and Prohorov's theorem

We recall that an often practical strategy for proving convergence is that of Exercise H.9: 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. Keep in mind our results in the case of weak convergence on the real axis, in particular Definition IV.2, Theorem IV.5, and Exercise IV.3. This section generalizes these to the context of probability measures on complete separable metric spaces.

Definition VI.2. 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

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

The collection $(\nu_i)_{i \in I}$ is said to be *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.

Theorem VI.4 (Prohorov's theorem)**.** *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.

First, by tightness, choose compact sets $K_1 \subset K_2 \subset K_3 \subset \cdots \subset \mathfrak{X}$ such that $\nu_n[K_m] > 1 - \frac{1}{m}$ for all n, m. Each compact K_m is separable (Exercise H.8) 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 $\mathcal{A}$ consist of all open balls $B_r(s)$ with $r \in \mathbb{Q} \cap (0, \infty)$ and $s \in S$. Note that $\mathcal{A}$ is indeed countable. It has the following important property: if $G \subset \mathfrak{X}$ is an open set and $x \in G \cap \bigcup_m K_m$, then there exists some $A \in \mathcal{A}$ 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 $s \in S$ such that $\varrho(x,s) < \frac{\varepsilon}{2}$ and $r \in \mathbb{Q} \cap (\varrho(x,s), \frac{\varepsilon}{2})$. Then $A = B_r(s)$ works: clearly $x \in A$ and $\overline{B}_r(s) \subset \overline{B}_{r+\varepsilon/2}(x) \subset B_\varepsilon(x) \subset G$. This collection $\mathcal{A}$ is a fair starting point, but even better is the collection $\mathcal{H}$ which consists of all finite unions $\overline{A}_1 \cup \cdots \cup \overline{A}_n$ of closures of sets from the collection $\mathcal{A}$ (the empty set $\emptyset$ is allowed, as a union of closures of zero sets). The collection $\mathcal{H}$ is also countable, and also stable under finite unions.

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 $\mathcal{H}$ converge. For any $H \in \mathcal{H}$, 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 for all sets H in the countable collection $\mathcal{H}$, the limits

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

exist. The remaining task is to show that there exists a Borel probability measure ν on $\mathfrak{X}$ such that for all open sets G we have

$$\nu[G] = \sup_{H\subset G} \alpha[H].$$

Once we have such a probability measure ν, the proof will in fact be complete. Namely, note that if G is open and $H \subset G$ with $H \in \mathcal{H}$, then $\alpha[H] = \lim_{k\to\infty} \nu_{n_k}[H] \leq \liminf_{k\to\infty} \nu_{n_k}[G]$, so we also have $\nu[G] \leq \liminf_{k\to\infty} \nu_{n_k}[G]$. This establishes criterion (ii) of Theorem VI.2, so we may conclude that

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

The measure ν as above can be constructed as follows (following the idea in Proposition H.5 that a set can be well approximated from the inside with closed sets and from the outside with open sets). For open sets G we define

$$\beta[G] = \sup_{H\subset G} \alpha[H],$$

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

$$\gamma[E] = \sup_{G\supset E} \beta[G].$$

It can be shown that this γ is an outer measure, and when restricted to Borel sets it defines our desired probability measure ν. The details of this check can be found in [**Bil99**]. □

Lecture VII

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 $t \in \mathbb{R}$. Fixing a scale parameter $a > 0$ (imagine a very small), we then consider the walk $X^{(a)} = \big(X_t^{(a)}\big)_{t\in[0,\infty)}$, with steps of magnitude $\sqrt{a}$ occurring with time interval a, i.e., we define

$$X_t^{(a)} = \sqrt{a}\, S_{t/a}. \tag{VII.1}$$

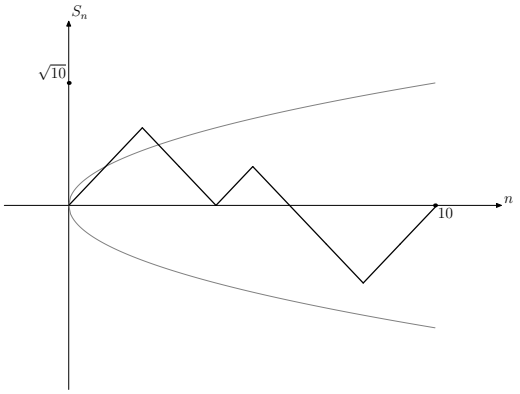
This linear interpolation and scaling of random walks is illustrated in Figure VII.1. Our goal is to prove the following result (the precise formulation and proof will be given after some preparations).

> **Donsker's theorem.** *The scaled random walks $X^{(a)}$ defined by* (VII.1) *converge weakly as $a \downarrow 0$ to Brownian motion, on a space of continuous functions.*

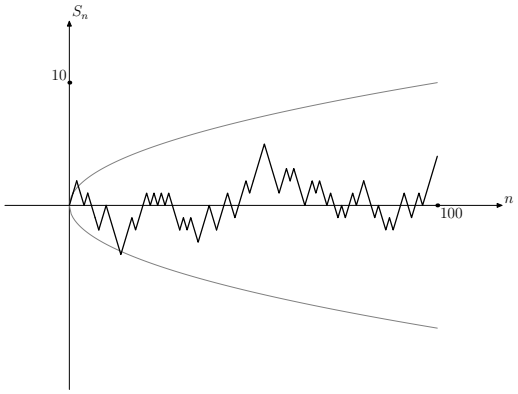
The proof strategy is the usual one, "precompactness plus identification of subsequential limits" (recall Exercise H.9 in Appendix H):

- 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 VI.4) that any sequence $X^{(a_m)}$ with $a_m \downarrow 0$ has convergent subsequences.
- Characterize explicitly any subsequential limit of the processes $X^{(a)}$.

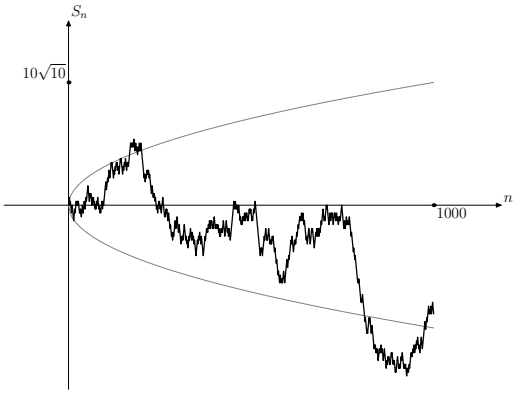
In Section 1 we give the defining properties of Brownian motion. In Section 2 we consider generalities about determining a Borel probability measure on the space of continuous functions, study compact subsets of a space of continuous functions, and establish a criterion for tightness. In Section 3 we 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) A random walk up to time $n = 10$.



(b) A continuation of the same walk up to time $n = 100$.



(c) A continuation of the same walk up to time $n = 1000$.

FIGURE VII.1. A random walk in different temporal and spatial scales. The plots include the curves $(t, \pm\sqrt{t})$, which remain invariant under the scaling given in formula (VII.1).

1. Brownian motion

In general, a (real-valued) stochastic process is a collection $(X_t)_{t\in\mathfrak{T}}$ 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 $\mathfrak{T} \subset \mathbb{R}$ 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 *trajectory* or *path* of the process. Often $0 \in \mathfrak{T}$ and $\mathfrak{T} \subset [0, +\infty)$, and we think that the process is started at time $t = 0$ from position $X_0 \in \mathbb{R}$, which may in general be random, although it is common to consider processes started deterministically from the origin, for example.

Brownian motion $(B_t)_{t\geq 0}$ is a stochastic process, which is crucial 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. We will be able to conclude the existence by showing that the rescaled random walks VII.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 VII.1. A stochastic process $X = (X_t)_{t \in \mathfrak{T}}$ is a *Gaussian process*, if for all n and $t_1, t_2, \ldots, t_n \in \mathfrak{T}$, the vector $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ is a Gaussian vector.

The laws of $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ for various choices $t_1, t_2, \ldots, t_n \in \mathfrak{T}$ are called the finite dimensional distributions of the process.

Example VII.1. The random walk with Gaussian steps considered in Lecture III is a Gaussian process with $\mathfrak{T} = \mathbb{Z}_{\geq 0}$.

Stationary and independent increments

Suppose now that for the process $(X_t)_{t \in \mathfrak{T}}$ the set of allowed time parameter values $\mathfrak{T} \subset \mathbb{R}$ is an additive semigroup, i.e., $0 \in \mathfrak{T}$ and whenever $t, s \in \mathfrak{T}$ then also $s + t \in \mathfrak{T}$ (for example $\mathfrak{T} = \mathbb{Z}_{\geq 0}$ or $\mathfrak{T} = \mathbb{R}$ or $\mathfrak{T} = [0, \infty)$). Then for a given $t, s \in \mathfrak{T}$ we can consider the increments $X_{s+t} - X_s$ 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 VII.2. A stochastic process $X = (X_t)_{t \in \mathfrak{T}}$ has *stationary increments*, if for all n and $t_1, t_2, \ldots, t_n \in \mathfrak{T}$ and $s \in \mathfrak{T}$, the vector $(X_{s+t_1} - X_s, \ldots, X_{s+t_n} - X_s)$ has the same law as the vector $(X_{t_1} - X_0, \ldots, X_{t_n} - X_0)$.

Definition VII.3. A stochastic process $X = (X_t)_{t \in \mathfrak{T}}$ has *independent increments*, if for all n and $t_0, t_1, \ldots, t_n \in \mathfrak{T}$ with $t_0 < t_1 < \cdots < t_n$ the collection $\big(X_{t_j} - X_{t_{j-1}}\big)_{j=1}^{n}$ of random variables is independent.

Continuity of a process

Suppose that $\mathfrak{T}$ is an interval, for example $\mathfrak{T} = [0,T]$, or $\mathfrak{T} = [0,\infty)$, or $\mathfrak{T} = \mathbb{R}$. In such a setup of continuous time, it is meaningful to consider te continuity of the paths of a process $(X_t)_{t\in\mathfrak{T}}$.

Definition VII.4. A stochastic process $X = (X_t)_{t\in\mathfrak{T}}$ has *continuous paths*, if

$$\mathsf{P}\Big[\,\Big\{\omega\in\Omega\;\Big|\; t\mapsto X_t(\omega) \text{ is continuous}\Big\}\,\Big] = 1.$$

Remark VII.2. 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 $\mathfrak{T}$ is an interval, this event depends on values X_t of the process at uncountably many times t, so the probability space $(\Omega,\mathcal{F},\mathsf{P})$ can not be chosen carelessly!

Defining properties of Brownian motion

Proposition VII.3. *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 $\mathsf{E}[X_sX_t] = \min(s,t)$, for all $s,t\geq 0$.*

Proof. (i) $\Rightarrow$ (ii): Let $0 = t_0 < t_1 < \cdots < t_m$ and $a_1,\dots,a_m \in \mathbb{R}$. Then $\sum_{k=1}^m a_k X_{t_k} = \sum_{k=1}^m b_k\big(X_{t_k} - X_{t_{k-1}}\big)$ for certain coefficients $b_1,\dots,b_m\in\mathbb{R}$. By assumption (i), the increments $X_{t_k} - X_{t_{k-1}}$, $k = 1,\dots,m$, 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 $s\leq t$

$$\begin{aligned}\mathsf{E}\big[X_s\,X_t\big] &= \mathsf{E}\Big[X_s\big(X_s + X_t - X_s\big)\Big]\\ &= \mathsf{E}\big[X_s^2\big] + \mathsf{E}\Big[X_s\big(X_t - X_s\big)\Big]\\ &= \mathsf{E}\big[X_s^2\big] + \mathsf{E}\big[X_s\big]\,\mathsf{E}\big[X_t - X_s\big] = s + 0\cdot 0 = s = \min(s,t).\end{aligned}$$

Thus the property (ii) follows.

(ii) $\Rightarrow$ (i): Suppose that $(X_t)_{t\in[0,\infty)}$ is a Gaussian process with $\mathsf{E}[X_t] = 0$ and $\mathsf{E}[X_sX_t] = \min(s,t)$. Then the increment $X_t - X_s$ for $s\leq t$ is Gaussian with mean

$$\mathsf{E}\big[X_t - X_s\big] = \mathsf{E}\big[X_t\big] - \mathsf{E}\big[X_s\big] = 0 - 0 = 0$$

and variance

$$\begin{aligned}\mathsf{E}\big[(X_t - X_s)^2\big] &= \mathsf{E}\big[X_t^2\big] - 2\,\mathsf{E}\big[X_t\,X_s\big] + \mathsf{E}\big[X_s^2\big]\\ &= t - 2\,\min(s,t) + s = t - 2s + s = t - s.\end{aligned}$$

This shows that indeed $X_t \sim \mathrm{N}(0,t)$ and $X_t - X_s \sim X_{t-s}$ for $s\leq t$. Finally consider, for $0 = t_0 < t_1 < \cdots < t_m$, the vector

$$(X_{t_1} - X_{t_0},\dots,X_{t_m} - X_{t_{m-1}}),$$

which 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 byt Proposition G.2. This calculation also implies the stationarity of the increments. □

Definition VII.5. 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 VII.3.

Remark VII.4 (On the existence of standard Brownian motion)**.** It is not a priori clear that the two requirements in Definition VII.5 can be simultaneously satisfied, i.e., it is not clear if a standard Brownian motion exists. The equivalent conditions of Proposition VII.3 could in principle rule out the possibility of continuous paths (and they indeed rule out for example differentiable paths: a standard Brownian motion is everywhere non-differentiable almost surely). We will, however, later in this lecture be able to conclude that a standard Brownian motion exists.

Remark VII.5 (On the uniqueness of standard Brownian motion)**.** Besides existence of a standard Brownian motion, another issue is whether the two requirements in Definition VII.5 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 $(B_{t_1}, \ldots, B_{t_n})$ involving finitely many times $t_1, \ldots, t_n \in [0, \infty)$. 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.

Remark VII.6 (A counterexample)**.** Suppose that $B = (B_t)_{t\in[0,\infty)}$ is a standard Brownian motion, for which the two conditions of Definition VII.5 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 $\mathsf{P}[\tau \in \{t_1, \ldots, t_n\}] = 0$), and in particular $\widetilde{B}$ satisfies the equivalent conditions of Proposition VII.3. 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 measure 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 $[0, \infty)$. For convenience, we will however first restrict our attention to a compact time interval $[0, T]$, and consider the space

$$\mathcal{C}([0,T]) = \{f\colon [0,T] \to \mathbb{R} \text{ continuous}\}$$

of continuous functions of time. 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 $\mathcal{B}\big(\mathcal{C}([0,T])\big)$, and consider probability measures on it.

Finite dimensional distributions and Borel sigma algebra

We first address the concern raised in Remark VII.5 by showing that finite dimensional distributions uniquely specify a Borel probability measure on $\mathcal{C}([0,T])$.

Proposition VII.7. *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 $0\le t_1<t_2<\dots<t_n\le T$ 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 sigma 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.4. 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 sigma algebra generated by the above collection. But we can write, by 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

$$\nu[E]=\mathsf{P}\left[\Big\{\omega\in\Omega\;\Big|\;\big(t\mapsto B_t(\omega)\big)\in\mathcal{E}\Big\}\right]. \tag{VII.2}$$

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 VII.7, 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 (VII.2) a unique Borel probability measure 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 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 VII.6. 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

$$\mathfrak{w}_f(\delta) = \sup_{\substack{t,s\in[0,T]\\ |s-t|\le\delta}} |f(s) - f(t)|. \tag{VII.3}$$

Remark VII.8. 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.10. For the present purposes, a convenient equivalent way of stating that theorem is the following.

Theorem VII.9. *A subset $\Phi \subset \mathcal{C}([0,T])$ is precompact if and only if the following two conditions are satisfied:*

$$\sup_{f\in\Phi} |f(0)| < \infty \tag{VII.9-i}$$

$$\lim_{\delta\downarrow 0} \sup_{f\in\Phi} \mathfrak{w}_f(\delta) = 0. \tag{VII.9-ii}$$

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 VII.10. *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:*

$$\forall \varepsilon > 0 \;\; \exists M < \infty, n_0 \in \mathbb{Z}_{>0} : \quad n \geq n_0 \;\Rightarrow\; \nu_n\Big[\big\{f \;\big|\; |f(0)| \geq M\big\}\Big] \leq \varepsilon \quad \text{(VII.10-i)}$$

$$\forall \varepsilon, \eta > 0 \;\; \exists \delta > 0, n_0 \in \mathbb{Z}_{>0} : \quad n \geq n_0 \;\Rightarrow\; \nu_n\Big[\big\{f \;\big|\; \mathfrak{w}_f(\delta) \geq \eta\big\}\Big] \leq \varepsilon. \quad \text{(VII.10-ii)}$$

Proof. "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 $\nu_n[\Phi] \geq 1-\varepsilon$ for all $n \in \mathbb{Z}_{>0}$. By Theorem VII.9, for large enough $M > 0$ we have $\Phi \subset \big\{f \;\big|\; |f(0)| < M\big\}$. This shows that (VII.10-i) holds. Fix also $\eta > 0$. By Theorem VII.9, for small enough $\delta > 0$ we have $\Phi \subset \big\{f \;\big|\; \mathfrak{w}_f(\delta) < \eta\big\}$. This shows that (VII.10-ii) holds.

"if": Assume conditions (VII.10-i) and (VII.10-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 $\Psi = \big\{f \;\big|\; |f(0)| \geq M\big\}$ satisfies $\nu_n[\Psi] \leq \varepsilon$ for all n. For every $k \in \mathbb{Z}_{>0}$ choose $\delta_k > 0$ such that $\Psi_k = \big\{f \;\big|\; \mathfrak{w}_f(\delta_k) \geq \frac{1}{k}\big\}$ satisfies $\nu_n[\Psi_k] \leq 2^{-k}\varepsilon$ for all n. Then the intersection of the complements $\Phi = \Psi^c \cap \bigcap_k \Psi_k^c$ satisfies the conditions of Theorem VII.9, so its closure $\overline{\Phi}$ is compact, and we see that for all n

$$\nu_n\big[\,\overline{\Phi}\,\big] \geq \nu_n\big[\Phi\big] = 1 - \nu_n\Big[\Psi \cup \bigcup_k \Psi_k\Big] \geq 1 - \varepsilon - \sum_k 2^{-k}\varepsilon = 1 - 2\varepsilon.$$

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

$$X_t^{(a)} = \sqrt{a}\left(\sum_{\ell=1}^{\lfloor t/a\rfloor} \xi_\ell + \Big(\frac{t}{a} - \Big\lfloor\frac{t}{a}\Big\rfloor\Big)\xi_{\lfloor t/a\rfloor+1}\right). \tag{VII.4}$$

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 VII.10.

We will obtain bounds for the modulus of continuity ultimately by the following.

Lemma VII.7. *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*

$$\mathfrak{w}_f(\delta) \leq 3 \max_{1\leq j\leq k} \max_{s\in[t_{j-1},t_j]} |f(s) - f(t_{j-1})|.$$

Proof. Denote, for brevity, the maximum on the right hand side by

$$M = \max_{1\leq j\leq k} \max_{s\in[t_{j-1},t_j]} |f(s) - f(t_{j-1})|.$$

Suppose that $s, t \in [0,T]$ are such that $|s-t| \leq \delta$, so they need to be considered in the definition (VII.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

$$|f(s)-f(t)| \le |f(s)-f(t_{j-1})| + |f(t_{j-1})-f(t)| \le 2M.$$

If $s \in [t_{j-1},t_j]$ and $t \in [t_j,t_{j+1}]$, then by triangle inequalities

$$|f(s)-f(t)| \le |f(s)-f(t_{j-1})| + |f(t_{j-1})-f(t_j)| + |f(t_j)-f(t)| \le 3M.$$

In either case $|f(s)-f(t)| \le 3M$ and thus $\mathfrak{w}_f(\delta) \le 3M$. □

To verify the important condition (VII.10-ii) for the random walks (VII.4), we use two more lemmas, formulated below with slightly different assumptions about the sequence $(\xi_\ell)_{\ell\in\mathbb{N}}$ of steps.

Lemma VII.11. *Suppose that $(\xi_\ell)_{\ell\in\mathbb{N}}$ is a stationary sequence, i.e., for any $\ell' \in \mathbb{N}$ and $L \in \mathbb{N}$, $(\xi_{L+1},\dots,\xi_{L+\ell'})$ has the same law as $(\xi_1,\dots,\xi_{\ell'})$. Suppose also that $(a_n)_{n\in\mathbb{N}}$ is a sequence of scales $a_n > 0$ tending to zero, $a_n \downarrow 0$. Then the laws of $\big(X^{(a_n)}\big)_{n\in\mathbb{N}}$ defined by* (VII.4) *for $t \in [0,T]$ are tight if*

$$\lim_{\lambda\to\infty} \limsup_{h\to\infty} \lambda^2\, \mathsf{P}\Big[\max_{1\le\ell'\le h} \Big| \sum_{\ell=1}^{\ell'} \xi_\ell \Big| \ge \lambda\sqrt{h} \Big] = 0. \tag{VII.5}$$

Proof. Condition (VII.10-i) holds trivially for the laws of $\big(X^{(a_n)}\big)_{n\in\mathbb{N}}$ given by (VII.4) since $X_0^{(a_n)} = 0$ for all n. We only need to verify condition (VII.10-ii), i.e. that for all $\eta > 0$

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge \eta \Big] = 0.$$

Now fix $\delta > 0$. By Lemma VII.7, we may estimate

$$\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge 3\varepsilon \Big] \le \sum_{j=1}^{k} \mathsf{P}\Big[\max_{s\in[t_{j-1},t_j]} |X_s^{(a_n)} - X_{t_{j-1}}^{(a_n)}| \ge \varepsilon \Big]$$

when $0 = t_0 < t_1 < \cdots < t_k \ge T$ are such that $t_j - t_{j-1} \ge \delta$ for all j. 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 discretization points $t_j = jh_na_n$ for $j = 0,1,\dots,k$. Note that $h_na_n \to \delta$ 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 VII.7 then reads

$$\begin{aligned} \mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge 3\varepsilon \Big] &\le \sum_{j=1}^{k} \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ (j-1)h_n<\ell'\le jh_n}} \Big| \sqrt{a_n} \sum_{\ell=(j-1)h_n+1}^{\ell'} \xi_\ell \Big| \ge \varepsilon \Bigg] \\ &= k\, \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ 0<\ell'\le h_n}} \Big| \sum_{\ell=1}^{\ell'} \xi_\ell \Big| \ge \frac{\varepsilon}{\sqrt{a_n}} \Bigg] \end{aligned}$$

where the last equality uses stationarity. Recall that for large enough n we have $a_n \le \frac{2\delta}{h_n}$ and also $k \le \frac{2T}{\delta}$, since only $\delta \le T$ are meaningful. We can thus further estimate the last expression to get the upper bound

$$\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \ge 3\varepsilon \Big] \le \frac{2T}{\delta} \mathsf{P}\Bigg[\max_{\substack{\ell'\in\mathbb{Z} \\ 0<\ell'\le h_n}} \Big| \sum_{\ell=1}^{\ell'} \xi_\ell \Big| \ge \frac{\varepsilon\sqrt{h_n}}{\sqrt{2\delta}} \Bigg].$$

Now for a fixed $\eta > 0$, set $\varepsilon = \frac{\eta}{3}$ above, and denote $\lambda = \frac{\varepsilon}{\sqrt{2\delta}} = \frac{\eta}{\sqrt{18\,\delta}}$, so that $\delta \downarrow 0$ corresponds to $\lambda \uparrow \infty$. Then rewrite the estimate as

$$\mathsf{P}\Big[\mathfrak{w}_{X^{(a_n)}}(\delta) \geq \eta\Big] \leq \frac{36T}{\eta^2}\,\lambda^2\,\mathsf{P}\bigg[\max_{\substack{\ell' \in \mathbb{Z} \\ 0<\ell' \leq h_n}} \Big|\sum_{\ell=1}^{\ell'} \xi_\ell\Big| \geq \lambda\sqrt{h_n}\bigg].$$

By the assumption (VII.5), we can make the right hand side here arbitrarily small apart from finite number of exceptional n, by choosing a sufficiently small $\delta > 0$ (i.e., sufficiently large λ). This finishes the proof. □

We also use the following Etemadi's inequality, very similar to Lévy's inequality (Lemma III.3) in Lecture III.

Lemma VII.12. *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*

$$\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq 3\Lambda\Big] \leq 3 \max_{1\leq k\leq h} \mathsf{P}\Big[\big|S_k\big| \geq \Lambda\Big].$$

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 VII.13. *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. Tightness of the laws is equivalent to the conditions (VII.10-i) and (VII.10-ii), which we will verify using Lemma VII.11.

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

$$\begin{aligned}\mathsf{E}\big[S_k^4\big] &= \sum_{\ell=1}^k \mathsf{E}\big[\xi_\ell^4\big] + 6 \sum_{1\leq \ell<m\leq k} \mathsf{E}\big[\xi_\ell^2\big]\,\mathsf{E}\big[\xi_m^2\big] \\ &= k\,\mathsf{E}\big[\xi_1^4\big] + 3k(k-1)\,\mathsf{E}\big[\xi_1^2\big]^2\end{aligned}$$

Noting that $|\xi_1| = 1$, the expected values are trivial to compute, and we can in particular deduce the upper bound $\mathsf{E}\big[S_k^4\big] \leq 3k^2$.

Then set $\Lambda = \frac{1}{3}\lambda\sqrt{h}$ in Lemma VII.12, and use furthermore Markov's inequality to get

$$\begin{aligned}\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq \lambda\sqrt{h}\Big] &\leq 3 \max_{1\leq k\leq h} \mathsf{P}\Big[\big|S_k\big| \geq \frac{1}{3}\lambda\sqrt{h}\Big] \\ &\leq 3 \max_{1\leq k\leq h} \frac{3k^2}{\big(\frac{1}{3}\lambda\sqrt{h}\big)^4} \\ &= 3^6\,\lambda^{-4}.\end{aligned}$$

From here we deduce the estimate

$$\lim_{\lambda\to\infty} \limsup_{h\to\infty}\ \lambda^2\,\mathsf{P}\Big[\max_{1\leq k\leq h} \big|S_k\big| \geq \lambda\sqrt{h}\Big] = 0,$$

so Lemma VII.11 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 VII.14. *For all $a > 0$, let $X^{(a)}$ be the scaled random walk given by* (VII.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 VII.13 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.9). 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 VII.7 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 (VII.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

$$(X^{(a_n)}_{t_1} - X^{(a_n)}_{t_0}, \dots, X^{(a_n)}_{t_m} - X^{(a_n)}_{t_{m-1}})$$

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 VII.1. Suppose that $B = (B_t)_{t\geq 0}$ is a standard Brownian motion and define three other stochastic processes $W^{(k)} = (W_t^{(k)})_{t\geq 0}$, $k = 1, 2, 3$, by setting

$$W_t^{(1)} = B_{s+t} - B_s$$
$$W_t^{(2)} = \lambda^{-1/2} B_{\lambda t}$$
$$W_t^{(3)} = \begin{cases} tB_{1/t} & , \text{ when } t > 0 \\ 0 & , \text{ when } t = 0 \end{cases}$$

where $s \geq 0$ and $\lambda > 0$ are constants. Show that all these three stochastic processes $W^{(k)} = (W_t^{(k)})_{t\geq 0}$, $k = 1, 2, 3$, are also standard Brownian motions.

Hint: *You may use results about existence and uniqueness of a standard Brownian motion.*

Exercise VII.2. Suppose that $B = (B_t)_{t\geq 0}$ 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 $(s,t) \mapsto \mathsf{Cov}[X_s, X_t] = \mathsf{E}[X_s X_t] - \mathsf{E}[X_s]\,\mathsf{E}[X_t]$.

(b) Show that X and B_1 are independent.

Hint: *The distribution of X is determined by the finite dimensional marginals $\mathsf{P}[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n]$, for $0 \leq t_1 < \cdots < t_n \leq 1$ and $A_1, \ldots, A_n \subset \mathbb{R}$ Borel. Consider also using the Gaussianity of the processes (recall: Gaussians are independent iff they are uncorrelated).*

Exercise VII.3. 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 $0 \leq s \leq t \leq 1$. Let $Y \sim N(0,1)$ be a random variable independent of X. Define a stochatic process $W = (W_t)_{t\in[0,1]}$ by setting $W_t = X_t + tY$. Show that W is a standard Brownian motion (on the time interval $[0,1]$). Define the *conditional distribution* of W given $W_1 = y$ by first conditioning W on the event $|W_1 - y| < \varepsilon$, which has positive probability, and then taking the limit $\varepsilon \searrow 0$. Find the conditional distribution of W given $W_1 = y$.

Lecture VIII

Ising model

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_v) located at the sites of the graph $(v \in \mathrm{V})$, taking two possible values: up or down ($\sigma_v = +1$ or $\sigma_v = -1$). 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 study the Ising model model first on finite subgraphs of $\mathbb{Z}^d$, and then consider how the model can be defined in "thermodynamical limit", on the infinite lattice $\mathbb{Z}^d$ via a weak limit.

The Ising model is defined very similarly to 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 Ising model can be shown to have a phase transition between paramagnetic and ferromagnetic phases similar to the Curie-Weiss model. The crucial refinement in the definition 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 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 a preference of local alignment, and it is also applicable to various other phenomena besides ferromagnetism.

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 set 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 $\sigma_v \in \{-1, +1\}$ (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 σ is taken to be

$$H_{\mathcal{G}}(\sigma) = -\sum_{\{x,y\}\in \mathrm{E}} \sigma_x\sigma_y - B\sum_{x\in \mathrm{V}} \sigma_x. \tag{VIII.1}$$

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

$$\mathsf{P}^{(\mathcal{G})}_{\beta,B}\left[\{\sigma\}\right] = \frac{1}{Z_{\mathcal{G}}(\beta,B)} e^{-\beta H_{\mathcal{G}}(\sigma)} \tag{VIII.2}$$

and the partition function $Z_{\mathcal{G}}(\beta,B)$ normalizes the total probability to one,

$$Z_{\mathcal{G}}(\beta,B) = \sum_{\sigma\in\Omega_{\mathcal{G}}} e^{-\beta H_{\mathcal{G}}(\sigma)}. \tag{VIII.3}$$

Exercise VIII.1. Consider the Ising model on a finite graph $\mathcal{G}$. 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

$$M := \mathsf{E}^{(\mathcal{G})}_{\beta,B}\Big[\frac{1}{\#\mathrm{V}}\sum_{x\in\mathrm{V}}\sigma_x\Big].$$

Show that

$$M = \frac{1}{\beta\,\#\mathrm{V}}\,\frac{\partial}{\partial B}\Big(\log Z_{\mathcal{G}}(\beta,B)\Big),$$

where the partition function $Z_{\mathcal{G}}$ is viewed as a function of β and B.

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 spaces $\Omega_{\mathcal{G}} = \{-1,+1\}^{\mathrm{V}}$ have natural partial orders $\preceq$, 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 we use the following monotonicity result known as the *FKG inequality*.

Theorem VIII.1 (FKG inequality for Ising model)**.** *Let $\mathcal{G} = (\mathrm{V},\mathrm{E})$ be a finite graph, and let $\mathsf{P}^{(\mathcal{G})}_{\beta,B}$ be the Ising model probability measure on $\Omega_{\mathcal{G}} = \{-1,+1\}^{\mathrm{V}}$. Denote, for brevity, $\mathsf{P} = \mathsf{P}^{(\mathcal{G})}_{\beta,B}$ and the expected value with respect to this measure by E. If f, g are increasing functions $\Omega_{\mathcal{G}} \to \mathbb{R}$, then the following FKG inequality holds*

$$\mathsf{E}\big[f\,g\big] \geq \mathsf{E}\big[f\big]\,\mathsf{E}\big[g\big]. \tag{VIII.4}$$

Monotonicity will be treated more generally in Section 3.2, and the proof of the above result in particular will be presented in Section 3.2.3. There exists also more straightforward proofs, and the reader is invited to try to prove the FKG inequality directly.

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 VIII.2. *Consider the Ising model on a finite graph, with notations as in Theorem VIII.1. If $A, B \subset \mathrm{V}$, then we have*

$$\mathsf{P}\Big[\sigma_{|A} \equiv +1 \,\Big|\, \sigma_{|B} \equiv +1\Big] \geq \mathsf{P}\Big[\sigma_{|A} \equiv +1\Big].$$

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

$$\begin{aligned}\mathsf{E}[f] &= \mathsf{P}\big[\sigma_{|A} \equiv +1\big] \\ \frac{\mathsf{E}[fg]}{\mathsf{E}[g]} &= \frac{\mathsf{P}\big[\sigma_{|A} \equiv +1 \text{ and } \sigma_{|B} \equiv +1\big]}{\mathsf{P}\big[\sigma_{|B} \equiv +1\big]} = \mathsf{P}\Big[\boldsymbol{\sigma}_{|A} \equiv +1 \,\Big|\, \boldsymbol{\sigma}_{|B} \equiv +1\Big].\end{aligned}$$

By Theorem VIII.1, the left hand side of the second line is greater than or equal to the left hand side of the first line. □

2. Weak convergence on countable products of finite sets

A criterion for weak convergence

Let $(\mathfrak{X}, \varrho)$ be a metric space, and $\mathcal{B} = \mathcal{B}(\mathfrak{X})$ its Borel sigma algebra.

Proposition VIII.3. *Suppose that $\mathcal{E} \subset \mathcal{B}$ is a collection of Borel subsets of $\mathfrak{X}$ such that the following conditions hold:*

$$\mathcal{E} \text{ is stable under finite intersections} \tag{VIII.3a}$$
$$\text{i.e., if } E_1, E_2 \in \mathcal{E} \text{ then also } E_1 \cap E_2 \in \mathcal{E}$$
$$\text{any open set } G \subset \mathfrak{X} \text{ is a countable union of sets from } \mathcal{E} \tag{VIII.3b}$$
$$\text{i.e., } G = \bigcup_{i=1}^{\infty} E_i \text{ with } E_i \in \mathcal{E}.$$

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 (VIII.3a) and (VIII.3b).

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 (VIII.3a) also the 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 (VIII.3b) 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

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

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 establishes weak convergence $\nu_n \xrightarrow{\text{w}} \nu$ by characterization (iii) of Portmanteau Theorem (Theorem VI.2). □

Weak convergence with 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. For the application to the Ising model, we will take $S = \{-1, +1\}$ and $I = \mathbb{Z}^d$, so that $\mathfrak{X} = S^I = \{-1, +1\}^{\mathbb{Z}^d}$ is the set of all spin configurations on the infinite hypercubic lattice $\mathbb{Z}^d$.

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

$$C = \pi_{i_1}^{-1}(S_1) \cap \cdots \cap \pi_{i_k}^{-1}(S_k) \subset S^I \tag{VIII.5}$$

are called *cylinder sets*. Recall also from Appendix H.4 that the collection $\mathcal{C}$ of all cylinder sets have the following properties:

(i) The collection $\mathcal{C}$ is countable.
(ii) The collection $\mathcal{C}$ is stable under finite intersections.
(iii) Any open set $G \subset S^I$ is the union of all cylinder sets C contained in it.
(iv) Any cylinder set C is both open and closed.

It follows from (i), (ii), (iii) that the collection $\mathcal{C}$ of all cylinder sets satisfies the two properties of Proposition VIII.3.

Theorem VIII.4. *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. *"only if"*: Suppose that $\nu_n \xrightarrow{\text{w}} \nu$, and let $C \subset S^I$ be a cylinder set. By property (iv) above, C is both open and closed, 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.2) we have $\nu_n[C] \to \nu[C]$.

"if": As remarked above, the collection $\mathcal{C}$ of all cylinder sets satisfies the two properties of Proposition VIII.3. 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 $C \in \mathcal{C}$, so it remains to show that α is a probability measure restricted to cylinder sets. Recall now that S^I is compact, and therefore $(\nu_n)_{n\in\mathbb{N}}$ is automatically tight. By Prohorov's theorem (Theorem VI.4) there exists some subsequence $(\nu_{n_k})_{k\in\mathbb{N}}$ which converges weakly $\nu_{n_k} \xrightarrow{\text{w}} \nu$. Again as in the "only if" part, since a cylinder set C is both open and closed, we get that $\nu[C] = \lim_{k\to\infty} \nu_{n_k}[C] = \alpha[C]$ for all $C \in \mathcal{C}$. This concludes the proof. □

3. Infinite volume limit by monotonicity

We will now treat the infinite volume limit of the Ising model. The idea is to define 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 $\mathbb{Z}^d$.

There are certain advantages of considering the Ising model with so called "plus boundary conditions", and we choose to do so in the rest of this lecture. The reader is encouraged to think of how to modify the approach below so as to work with the so called "free boundary conditions" which are a more literal interpretation of just defining the Ising model probability measures on finite subgraphs of $\mathbb{Z}^d$.

Limit via increasing sequence of finite subgraphs

We are ready to attempt to define the Ising model on the d-dimensional infinite lattice $\mathbb{Z}^d$. Fix parameters $\beta > 0$ (inverse temperature) and $B \in \mathbb{R}$ (external magnetic field), and for notational simplicity, explicit references to β and B will be omitted below.

Fix a sequence $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of finite induced subgraphs of the lattice $\mathbb{Z}^d$:

- $\mathcal{G}_n = (\mathrm{V}_n, \mathrm{E}_n)$, 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
- to account for boundary conditions, denote also
$$\partial\mathrm{E}_n = \big\{\{x,y\} \;\big|\; x \in \mathrm{V}_n,\ y \in \mathbb{Z}^d \setminus \mathrm{V}_n,\ \|x-y\| = 1\big\}\,,$$
the set of bonds connecting one lattice sites inside and one outside of V_n
- assume that $\mathrm{V}_n \uparrow \mathbb{Z}^d$, i.e. $\mathrm{V}_1 \subset \mathrm{V}_2 \subset \cdots$ and $\bigcup_{n\in\mathbb{N}} \mathrm{V}_n = \mathbb{Z}^d$.

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 $\Omega_n = \{-1,+1\}^{\mathrm{V}_n}$. 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 $\mathbb{Z}^d$. With the plus boundary conditions that we will consider, the natural way to do this is to extend a spin configuration $\sigma^{(n)} = (\sigma_x^{(n)})_{x\in\mathrm{V}_n}$ as constant $+1$ to the outside of the finite subset $\mathrm{V}_n \subset \mathbb{Z}^d$, i.e. to define the corresponding $\sigma = (\sigma_x)_{x\in\mathbb{Z}^d}$ by

$$\sigma_x = \begin{cases} \sigma_x^{(n)} & \text{if } x \in \mathrm{V}_n \\ +1 & \text{if } x \in \mathbb{Z}^d \setminus \mathrm{V}_n. \end{cases} \tag{VIII.6}$$

This way we interpret $\Omega_n \subset \Omega$ as a subset, and any probability measure P_n on Ω_n also as a probability measure of Ω (in particular the Ising model probability measure on $\mathcal{G}_n$ with given parameters β, B). Note that the space $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$ is of the form S^I considered in Section 2 and Appendix H.4: its topology comes from a complete separable metric, and we have a convenient criterion for weak convergence in terms of probabilities of cylinder events.

Before proceeding further to the precise analysis of the Ising model on the increasing sequance $(\mathcal{G}_n)_{n\in\mathbb{N}}$ of finite subgraphs of $\mathbb{Z}^d$, we make an important remark about what one really needs to check in order to establish weak convergence of the laws of the random spin configurations.

Lemma VIII.5. *Let $(\mathsf{P}_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$. Suppose that for any $k \in \mathbb{N}$ and all $x_1, \dots, x_k \in \mathbb{Z}^d$ the probabilities*

$$\mathsf{P}_n\Big[\sigma_{x_1} = +1, \dots, \sigma_{x_k} = +1\Big] \tag{VIII.7}$$

have limits as $n \to \infty$. Then the sequence $(\mathsf{P}_n)_{n\in\mathbb{N}}$ converges weakly. The weak limit P of this sequence is uniquely characterized by the condition that for any $x_1, \dots, x_k \in \mathbb{Z}^d$ the probability $\mathsf{P}\big[\sigma_{x_1} = +1, \dots, \sigma_{x_k} = +1\big]$ is the limit of (VIII.7).

Proof. Recall from Theorem VIII.4 that for weak convergence of P_n it is sufficient that the limits $\alpha[C] = \lim_{n\to\infty} \mathsf{P}_n[C]$ exist for all cylinder events C, and that the weak limit probability measure P is then uniquely specified by these $\alpha[C]$. A general a cylinder event C is of the form

$$C = \big\{\sigma \in \Omega \,\big|\, \sigma_{x_1} = \epsilon_1, \dots, \sigma_{x_k} = \epsilon_k\big\},$$

where $x_1, \dots, x_k \in \mathbb{Z}^d$ and $\epsilon_1, \dots, \epsilon_k \in \{-1,+1\}$. Considering $P = \big\{x_i \,\big|\, \epsilon_i = +1\big\}$ and $M = \big\{x_i \,\big|\, \epsilon_i = -1\big\}$, we see that the condition amounts to the convergence of the probabilities

$$\mathsf{P}_n\Big[\sigma_x = +1 \;\forall x \in P \text{ and } \sigma_y = -1 \;\forall y \in M\Big],$$

for all $P, M \subset \mathbb{Z}^d$ disjoint finite subsets of sites. However, by inclusion-exclusion formula we can write

$$\begin{aligned}
&\mathsf{P}_n\Big[\sigma_x = +1 \;\forall x \in P \text{ and } \sigma_y = -1 \;\forall y \in M\Big] \\
&= \mathsf{P}_n\Big[\sigma_x = +1 \;\forall x \in P\Big] - \mathsf{P}_n\Big[\sigma_x = +1 \;\forall x \in P \text{ and } \exists y \in M : \sigma_y = +1\Big] \\
&= \sum_{J\subset M} (-1)^{\#J}\, \mathsf{P}_n\Big[\sigma_z = +1 \;\forall z \in P \cup J\Big].
\end{aligned}$$

Thus we can deduce the convergence of the probabilities of all cylinder events from the existence of the limits of (VIII.7). □

In this context it is natural to extend the configuration $\boldsymbol{\sigma} \in \{-1,+1\}^{\Lambda_R}$ also to the complement of the box Λ_R as constant $+1$, so that each ν_R^+ can be interpreted as a probability measure on the space $\Omega = \{-1,+1\}^{\mathbb{Z}^d}$, which is of the form considered in Section **??**.

3.1. Glauber dynamics for the Ising model

3.2. Stochastic domination and monotonicity in general

3.2.1. *Stochastic domination*

3.2.2. *Holley's criterion*

3.2.3. *Proof of FKG inequality for the Ising model*

3.3. Existence of phase transition

3.3.1. *Low temperature expansion*

3.3.2. *High temperature expansion*

3.4. Further results on the Ising model

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 $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and $(\mathfrak{X}, \Sigma)$ a measurable space. Suppose that $X\colon \Omega \to \mathfrak{X}$ is a random variable, i.e., an $\mathcal{F}/\Sigma$-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 $A \in \Sigma$.

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

1. Product sigma-algebra

2. Products of finitely many measures

3. Fubini's theorem

Theorem B.1. *Let $(\mathfrak{X}_1, \Sigma_1, \mu_1)$ and $(\mathfrak{X}_2, \Sigma_2, \mu_2)$ be two σ-finite measure spaces, and let $(\mathfrak{X}_1 \times \mathfrak{X}_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \otimes \mu_2)$ be their product measure space.*

(a) *If $f\colon \mathfrak{X}_1 \times \mathfrak{X}_2 \to [0, +\infty]$ is $\Sigma_1 \otimes \Sigma_2$-measurable, then we have the equalities*

$$\begin{aligned} &\int_{\mathfrak{X}_1 \times \mathfrak{X}_2} f \,\mathrm{d}(\mu_1 \otimes \mu_2) \\ &= \int_{\mathfrak{X}_2} \left(\int_{\mathfrak{X}_1} f(x_1, x_2) \,\mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \\ &= \int_{\mathfrak{X}_1} \left(\int_{\mathfrak{X}_2} f(x_1, x_2) \,\mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \end{aligned} \tag{B.1}$$

of numbers in $[0, +\infty]$.

(b) *If $f\colon \mathfrak{X}_1 \times \mathfrak{X}_2 \to \mathbb{R}$ is $\Sigma_1 \otimes \Sigma_2$-measurable, and if at least one of the integrals*

$$\begin{gathered} \int_{\mathfrak{X}_1 \times \mathfrak{X}_2} |f| \,\mathrm{d}(\mu_1 \otimes \mu_2) \\ \int_{\mathfrak{X}_2} \left(\int_{\mathfrak{X}_1} \big|f(x_1, x_2)\big| \,\mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) \\ \int_{\mathfrak{X}_1} \left(\int_{\mathfrak{X}_2} \big|f(x_1, x_2)\big| \,\mathrm{d}\mu_2(x_2) \right) \mathrm{d}\mu_1(x_1) \end{gathered}$$

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*

$$\mathsf{E}\big[X\big] = \int_0^\infty \mathsf{P}\big[X > s\big] \,\mathrm{d}s. \tag{B.2}$$

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 $s \in [k-1, k)$. □

4. Countable products of 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 $(\Omega, \mathcal{F}, \mathsf{P})$, and let $(\mathcal{S}, \Sigma)$ be a measurable space. A *stochastic process* on the state space $\mathcal{S}$ is a collection $(X_t)_{t\in\mathfrak{T}}$ of random variables $X_t \colon \Omega \to \mathcal{S}$ indexed by a "time parameter" t. The set $\mathfrak{T} \subset \mathbb{R}$ 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 $(X_t)_{t\in\mathfrak{T}}$ 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, $\mathfrak{T} = \mathbb{Z}_{\geq 0}$. 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. Let $(\mathfrak{X}_1, \Sigma_1, \nu_1)$ and $(\mathfrak{X}_2, \Sigma_2, \nu_2)$ 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

$$\begin{aligned} &\forall E_1 \in \Sigma_1 : \quad \nu\big[E_1 \times \mathfrak{X}_2\big] = \nu_1[E_1] \qquad (D.1)\\ \text{and} \qquad &\forall E_2 \in \Sigma_2 : \quad \nu\big[\mathfrak{X}_1 \times E_2\big] = \nu_2[E_2]. \end{aligned}$$

It is often convenient to phrase a coupling equivalently with random variables as follows.

Definition D.2. Let $(\Omega_1, \mathcal{F}_1, \mathsf{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathsf{P}_2)$ 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 $(\Omega, \mathcal{F}, \mathsf{P})$ 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

Suppose that $(\mathfrak{X}, \Sigma)$ is a measurable space and $\preceq$ is a *partial order* on $\mathfrak{X}$, i.e., a binary relation such that

$$x \preceq y \ \text{ and } \ y \preceq x \quad \Longleftrightarrow \quad x = y \qquad (D.2)$$

$$x \preceq y \ \text{ and } \ y \preceq z \quad \Longrightarrow \quad x \preceq z. \qquad (D.3)$$

Example D.1. 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 componentwise 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.2. 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.

TIn the setting of a set with a partial order $\preceq$, we can talk about increasing (real valued) functions. A function $f\colon \mathfrak{X} \to \mathbb{R}$ is said to be *increasing* if

$$x \preceq y \quad \Longrightarrow \quad f(x) \leq f(y). \tag{D.4}$$

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.3. *Let $(\mathfrak{X}, \Sigma)$ be a measurable space and $\preceq$ a partial order on $\mathfrak{X}$. Suppose that ν_1 and ν_2 are two probability measures on $\mathfrak{X}$, and suppose that there exists a coupling ν of ν_1 and ν_2 such that $\nu\big[\,\{(x, y) \in \mathfrak{X} \times \mathfrak{X} \;\big|\; x \preceq y\}\,\big] = 1$. Then for all increasing functions $f\colon \mathfrak{X} \to \mathbb{R}$ we have*

$$\int_{\mathfrak{X}} f \,\mathrm{d}\nu_1 \;\leq\; \int_{\mathfrak{X}} f \,\mathrm{d}\nu_2.$$

Corollary D.4. *Let $(\mathfrak{X}, \Sigma)$ be a measurable space and $\preceq$ a partial order on $\mathfrak{X}$. 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 $X_1 \preceq X_2$ almost surely. Then for all increasing functions $f\colon \mathfrak{X} \to \mathbb{R}$ we have*

$$\mathsf{E}\Big[f(X_1)\Big] \leq \mathsf{E}\Big[f(X_2)\Big].$$

3. 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 $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. Suppose that $\mathcal{A}_1, \mathcal{A}_2, \ldots$ are sub-sigma algebras of $\mathcal{F}$ which are independent (meaning that if $A_1 \in \mathcal{A}_1, \ldots, A_n \in \mathcal{A}_n$, then $\mathsf{P}[A_1 \cap \cdots \cap A_n] = \mathsf{P}[A_1] \cdots \mathsf{P}[A_n]$). Define

$$\mathcal{T}_n = \sigma\left(\bigcup_{m \geq n} \mathcal{A}_m\right)$$

to be the sigma algebra generated by $\mathcal{A}_m$, $m \geq n$, and then define the tail sigma algebra

$$\mathcal{T} = \bigcap_{n>0} \mathcal{T}_n. \tag{E.1}$$

Then we have:

Theorem E.1 (Kolmogorov's $0-1$ law)**.** *If $E \in \mathcal{T}$, 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 $\mathcal{A}_n = \sigma(X_n)$. We have $\mathcal{T}_n = \sigma(X_n, X_{n+1}, X_{n+2}, \ldots)$. If $E \in \mathcal{T}_n$ for any n, then $E \in \mathcal{T} = \bigcap_n \mathcal{T}_n$. 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 $(\mathcal{A}_i)_{i \in I}$ is a collection of independent sub-sigma algebras of $\mathcal{F}$. We can give some enumeration $i_1, i_2, \ldots$ of all elements of I, and define the tail sigma algebra using this as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{m \geq n} \mathcal{A}_{i_m} \right).$$

This definition of $\mathcal{T}$ 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.1 (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 $\varphi(\theta) = \mathsf{E}\big[e^{\mathrm{i}\,\theta X}\big] = e^{\mathrm{i}\,\theta\mu - \frac{1}{2}\theta^2\sigma^2}$. 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.1. 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 $\mathsf{E}\big[e^{\mathrm{i}\,\theta\, a\cdot V}\big] = e^{\mathrm{i}\,\theta\mu - \frac{1}{2}\sigma^2\theta^2}$ 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

$$\varphi(\theta) := \mathsf{E}\Big[e^{\mathrm{i}\theta\cdot V}\Big] = e^{\mathrm{i}\mu - \frac{1}{2}\sigma^2}.$$

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.2. *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.3. 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

2.1. 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 $\mathfrak{X}^{(1)}$.

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.1. 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.2. The set of rational numbers $\mathbb{Q} \subset \mathbb{R}$ is not complete (but it is separable).

Example H.3. Denote by $\ell^\infty(\mathbb{N})$ the set of all bounded sequences $a = (a_n)_{n\in\mathbb{N}}$ of real numbers. A natural metric 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 VII. Countable Cartesian products of finite sets are the topic of Section H.4 — they were used in Lecture VIII for the Ising model, and in Lectures **??** for interacting particle systems. Both of these are complete separable metric spaces.

2.2. 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.2. The Borel σ-algebra $\mathcal{B} = \mathcal{B}(\mathfrak{X})$ on a topological space $\mathfrak{X}$ is the smallest sigma algebra which contains all open sets $U \subset \mathfrak{X}$.

Remark H.4. By the properties of σ-algebras, the Borel σ-algebra $\mathcal{B}$ 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 $\big(\mathfrak{X}, \mathcal{B}, \nu\big)$. 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.4. Suppose that $(\mathfrak{X}, \varrho)$ is a separable metric space, and that $G \subset \mathfrak{X}$ is an open set.

- Show that G can be written as a countable union of open balls. Conclude that the Borel sigma-algebra $\mathcal{B}(\mathfrak{X})$ is generated by open balls.
- Show that G can be written as a countable union of closed balls. Conclude that the Borel sigma-algebra $\mathcal{B}(\mathfrak{X})$ 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.5. *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 $\mathcal{B}$. □

We also record the following trivial but very useful lemma.

Lemma H.3. *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. When $F \neq \emptyset$, set $\varrho(x, F) = \inf_{y \in F}(\varrho(x, y))$ and use for example $f(x) = \max\{1 - \frac{\varrho(x,F)}{\varepsilon}, 0\}$. □

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.5. 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 $\nu_1[F] = \nu_2[F]$
(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.

2.3. Compactness and sequential compactness

For a topological space $\mathfrak{X}$, we give the following definitions.

Definition H.4. 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 $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ 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 $U_{i_1}, \dots, U_{i_m}$ which also covers the space, $\bigcup_{k=1}^{m} U_{i_k} = \mathfrak{X}$.

For metric spaces the two definitions are equivalent.

Proposition H.6. *A metric space $(\mathfrak{X}, \varrho)$ is compact if and only if it is sequentially compact.*

Exercise H.6. Prove Proposition H.6 (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.7. 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.8. *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*
$$|f(x)| \le M \qquad \text{for all } x \in \mathfrak{X}.$$
- *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.7. Prove Proposition H.8 (or recall its proof).

Exercise H.8. Show that a compact metric space is separable.

We also occasionally use precompactness.

Definition H.5. 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 $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ 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.9. 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 $\big(x_{n_k}\big)_{k\in\mathbb{N}}$ has a further subsequence $\big(x_{n_{k_j}}\big)_{j\in\mathbb{N}}$ 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 VII 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

$$\mathcal{C}\big([0,T]\big) = \{f\colon [0,T] \to \mathbb{R} \text{ continuous}\} \tag{H.1}$$

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

$$\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|. \tag{H.2}$$

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.,

$$\|f\|_\infty = 0 \;\Leftrightarrow\; f \equiv 0, \quad \|\lambda f\|_\infty = |\lambda|\,\|f\|_\infty, \quad \|f_1 + f_2\|_\infty \le \|f_1\|_\infty + \|f_2\|_\infty.$$

The norm is used to define a metric in the usual way: for $f_1, f_2 \in \mathcal{C}([0,T])$ we set

$$\varrho(f_1, f_2) = \|f_1 - f_2\|_\infty = \sup_{t\in[0,T]} |f_1(t) - f_2(t)|. \tag{H.3}$$

Note that a sequence $(f_n)_{n\in\mathbb{N}}$ of functions $f_n \in \mathcal{C}([0,T])$ converges with respect to the metric ϱ induced by the norm $\|\cdot\|_\infty$ if and only if it converges uniformly. A basic result in calculus states that if a sequence of continuous functions converges uniformly, then the limit is continuous. It follows that the space $\big(\mathcal{C}([0,T]), \varrho\big)$ is complete. By uniform continuity of any $f \in \mathcal{C}([0,T])$ it is easy to see that the countable collection of piecewise linear interpolations of functions with rational values at dyadic points is dense in $\mathcal{C}([0,T])$ (alternatively, from Weierstrass approximation theorem it follows that the countable collection of polynomial functions with rational coefficients is dense). Therefore we have:

Theorem H.9. *The space $\big(\mathcal{C}([0,T]), \varrho\big)$ is complete and separable.*

In Lecture VII we used the following characterization of compact subsets in $\big(\mathcal{C}([0,T]), \varrho\big)$.

Theorem H.10 (Arzelà-Ascoli theorem)**.** *A subset $\Phi \subset \mathcal{C}([0,T])$ is precompact if and only if the following two conditions are satisfied:*

(1) *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 $|f(t)| \le M$.*
(2) *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 $|f(t) - f(s)| < \varepsilon$.*

4. Countable products of discrete spaces

In Lecture VIII and Lectures **??** 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

$$S^I = \Big\{ (s_i)_{i\in I} \;\Big|\; s_i \in S \Big\} \tag{H.4}$$

where I is a countable index set. In common applications we might have for example $I = \mathbb{Z}^d$, but it is convenient to assume a given enumeration of I, so that we may identify $I = \mathbb{N}$. We make S^I a complete separable metric space with a metric ϱ which depends on the chosen enumeration, but so that the topology induced by the metric is independent of the choice.

Exercise H.10. Define $\varrho\colon S^{\mathbb{N}} \times S^{\mathbb{N}} \to [0,\infty)$ by the formula

$$\varrho(\omega,\omega') = \sum_{\substack{i\in\mathbb{N}\\ \omega_i\neq\omega'_i}} 2^{-i} \qquad \text{for } \omega,\omega' \in S^{\mathbb{N}},\ \omega=(\omega_i)_{i\in\mathbb{N}},\ \omega'=(\omega'_i)_{i\in\mathbb{N}}.$$

(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.

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.

[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.

[Knu97] Donald E. Knuth. *The art of computer programming, Vol. 3: Sorting and searching*. Addison Wesley Longman, 2nd edition, 1997.

[KZ87] M Kardar and Y-C. Zhang. Scaling of directed polymers in random media. *Phys. Rev. Lett.*, 58(20):2087–2090, 1987.

[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.