Mathematical Statistical Mechanics

Math Section

Thomas Richthammer

SoSe 2011, LMU München

July 28, 2011

Contents

1 Introduction **3**

2 Probability **3**
2.1 Probability measures 3
2.2 Random variables 6
2.3 Conditioning 8
2.4 Entropy 9
2.5 Limit laws 12
2.6 Convergence to equilibrium: Markov chains 17

3 Ensemble theory **21**
3.1 General setup 21
3.2 Microcanonical distribution 22
3.3 Canonical distribution 25
3.4 Grand canonical distribution 27
3.5 Stability 29
3.6 Boundary conditions 32
3.7 Infinite volume distributions 34

4 Ising model **37**
4.1 Definition of the Ising model 37
4.2 Symmetries 39
4.3 Ground states 40
4.4 Existence and uniqueness of Gibbs measures 41
4.5 Ising model in one dimension 46
4.6 Phase transition for the Ising model 48

1 Introduction

These lecture notes cover some of the topics discussed in the Math section of the lecture "Mathematical Statistical Mechanics" given at the LMU München in the summer semester 2011. I will try to update them regularly and if time permits I will try to produce an almost complete set of lecture notes. The Math part of the lecture will be self contained. Only some abstract measure theory will be assumed. However, this part of the lecture contains only very little in terms of motivation and justification for the models to be considered. Some of this will be discussed in the Physics part of the lecture.

2 Probability

2.1 Probability measures

Probability models can be used in situations where there is some degree of uncertainty about what happens. Such a situation will be called an experiment, and everything that possibly can happen will be called an outcome or a sample. Usually the uncertainty about the outcome of an experiment arises because the situation is too complicated to be analyzed in detail or because we choose to ignore some of the more complicated details to obtain a better understanding of what is going on.

Example: A very simple example of an experiment is to roll a die. In principle it should be possible to deduce the outcome of this experiment from the laws of physics provided that you have precise information about the "initial state" of the die (position, velocity,...) the moment you roll it and the shape of the terrain in which you roll it. But usually you either don't have this information or even if you have it, you may choose to ignore it, e.g. because the calculations involved are too complicated or you are not interested in a single experiment anyway, but in what can be said if the experiment is repeated many times with different initial states. In either case a probabilistic model is more appropriate. Such a model should consist of a list of possible outcomes (1,2,3,4,5,6) and their respective probabilities (1/6 for each outcome if the die is assumed to be fair).

The basic structure for all sorts of probability models is the same, so the following axiomatic definition has proven very useful:

Definition 2.1 *A probability model is given by* $(\Omega, \mathcal{F}, \mathbb{P})$. *Here* $\Omega \neq \emptyset$ *is called the sample space,* $\mathcal{F}$ *is a* σ*-algebra on* Ω *and its elements are called events and* $\mathbb{P}$ *is a probability measure (=distribution= normalized measure) on the measurable space* $(\Omega, \mathcal{F})$.

The elements of Ω are the possible outcomes of the experiment considered, and events are sets of outcomes that we would like to assign probabilities.

We now briefly review the basic conepts of measure theory. We will omit the construction of the integral w.r.t. an arbitrary measure and the statement of its basic properties, and we will omit proofs in this and the next section.

Definition 2.2 *$\mathcal{F} \subset \mathfrak{P}(\Omega)$ is called a σ-algebra on $\Omega \neq \emptyset$ if*

- $\Omega \in \mathcal{F}$
- $\forall A \in \mathcal{F} : A^c = \Omega - A \in \mathcal{F}$
- $\forall$ *countable* I, $A_i \in \mathcal{F} (i \in I)$: $\bigcup_i A_i \in \mathcal{F}$

Usually a σ-algebra is constucted from a class of "interesting sets":

Definition 2.3 *For $\mathcal{S} \subset \mathfrak{P}(\Omega)$ we let $\sigma(\mathcal{S})$ denote the smallest σ-algebra containing $\mathcal{S}$. $\mathcal{S}$ can be obtained as the intersection of all σ-algebras containing $\mathcal{S}$. It is called the σ-algebra generated by $\mathcal{S}$. If Ω is a topological space and $\mathcal{S}$ is the set of open sets, $\mathcal{B}(\mathcal{S}) := \sigma(\mathcal{S})$ is called the Borel-σ-algebra on Ω.*

In measure theory only those functions are of interest that are compatible with the measurability structure.

Definition 2.4 *A function $f : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ is called measurable if*

- $f^{-1}A' \in \mathcal{F}$ *for all* $A' \in \mathcal{F}'$.

Mostly we consider real valued functions. Basic examples of measurable functions $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ are indicator (or characteristic) functions: For $A \in \mathcal{F}$ let 1_A be defined by

$$1_A(\omega) = \begin{cases} 1 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A. \end{cases}$$

It will be convenient to consider general measures instead of just probability measures.

Definition 2.5 *$\mu : \mathcal{F} \to [0, \infty]$ is called a measure on a measurable space $(\Omega, \mathcal{F})$ if*

- $\forall$ *countable* I, *disjoint* $A_i \in \mathcal{F} (i \in I)$: $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$

μ is called σ-finite if there there are countably many $E_i \in \mathcal{F}$ with $\bigcup_i E_i = \Omega$ and $\mu(E_i) < \infty$. μ is called finite if $\mu(\Omega) < \infty$ and μ is called normalized if $\mu(\Omega) = 1$.

The measures to be considered here will always be assumed to be σ-finite, and mostly we will deal with probability measures, i.e normalized measures. A measure is determined by its values $\mu(A)$ for all $A \in \mathcal{F}$, but often it is sufficient to know the values for $A \in \mathcal{S}$, where $\mathcal{S} \subset \mathcal{F}$ is a suitable subset of $\mathcal{F}$.

Theorem 2.1 *(Uniqueness theorem.) Let μ be a probability measure on $(\Omega, \mathcal{F})$ and $\mathcal{S} \subset \mathcal{F}$ such that $\mathcal{S}$ is $\cap$-stable (i.e. $\forall A, B \in \mathcal{S} : A \cap B \in \mathcal{S}$) and a generator of $\mathcal{F}$ (i.e. $\sigma(\mathcal{S}) = \mathcal{F}$). Then μ is uniquely determined by its values on $\mathcal{S}$.*

Proof: Measure Theory. □

It is possible to define an integral w.r.t. a given measure μ on $(\Omega, \mathcal{F})$. There are several common notations for the integral

$$\int f d\mu = \int f(\omega) d\mu(\omega) = \int f(\omega) \mu(d\omega) \qquad \text{for } f : \Omega \to [0, \infty] \text{ measurable.}$$

The construction is the same as for the Lebesgue integral from the Lebesgue measure. The properties of the integral are also similar to those of the Lebesgue integral. If f_+, f_- denote the positive and negative part of a function f, i.e. $f_+, f_- \geq 0$ s.t. $f = f_+ - f_-$, we can set

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu$$

unless both integrals on the RHS are ∞. If both are finite we call f integrable. $\mathcal{L}^n(\mu)$ denotes the space of all measurable functions $f : \Omega \to [-\infty, \infty]$ such that f^n is integrable w.r.t. μ. Here $n \in \mathbb{N}$.

The most basic examples of (σ-finite) measures are

- The counting measure χ. Here Ω is countable, $\mathcal{F} = \mathfrak{P}(\Omega)$ and $\chi(A) := \#(A)$ is the number of elements of A. Integration w.r.t. χ is the same as summation.
- The Lebesgue measure λ^d ($d \geq 1$). Here $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}^d = \mathcal{B}(\mathbb{R}^d)$ and $\lambda^d(A)$ describes the d-dimensional volume of a set $A \in \mathcal{B}^d$. Integration w.r.t λ^d is the same as d-dimensional Lebesgue-integration.

The most basic examples of distributions are derived from these by means of densities.

Definition 2.6 *Let μ be a measure on $(\Omega, \mathcal{F})$ and let $f : \Omega \to [0, \infty]$ be measurable. Then*

$$\nu(A) := \int_A f d\mu := \int f 1_A d\mu$$

defines a new measure on $(\Omega, \mathcal{F})$ and f is called its density. If f is normalized, i.e. $\int f d\mu = 1$, then ν is a probability measure.

f plays the role of a penalty function. Large/small values of $f(x)$ increase/decrease the probability of the outcome $x \in \Omega$ as compared to the one w.r.t. μ. If ν has a density, this density is μ-a.s. uniquely determined. If a probability measure has a density w.r.t. $\mu = \chi$, it is called discrete, and if it has a density w.r.t. λ^d, it is called continuous. The most important examples of distributions of this type are the uniform distributions:

- If $\#(A) < \infty$ and $f = \frac{1}{\#(A)} 1_A$, the corresponding discrete distribution is called the (discrete) uniform distribution on A.
- If $\lambda^d(A) < \infty$ and $f = \frac{1}{\lambda^d(A)} 1_A$, the corresponding continuous distribution is called the (continuous) uniform distribution on A.

In both cases of the uniform distribution on A the outcomes $x \in A$ are equally likely, whereas outcomes $x \in A^c$ are impossible.

Remark: Every measure μ on $(\Omega, \mathcal{F})$ with countable Ω and $\mathcal{F} = \mathfrak{P}(\Omega)$ is discrete: To obtain the density we simply set $f(x) = \mu(\{x\})$ for all $x \in \Omega$.

Another possibility to generate measures is by taking products:

Definition 2.7 *Let I be an index set and for every $i \in I$ let $(\Omega_i, \mathcal{F}_i, \mu_i)$ be a measure space. We set $\Omega = \prod_i \Omega_i$ and $\mathcal{F} = \otimes_i \mathcal{F}_i := \sigma(\prod_i A_i : A_i \in \mathcal{F}_i)$ the product-σ-algebra. A measure μ on $(\Omega, \mathcal{F})$ is called product measure iff*

$$\mu(\prod_i A_i) = \prod_i \mu_i(A_i) \quad \forall A_i \in \mathcal{F}_i.$$

The product measure exists and is unique if either I is finite and every μ_i is σ-finite or I is arbitrary and every μ_i is a probability measure, and then it is denoted by $\mu = \otimes_i \mu_i$.

We note that the counting measure on a product set $\Omega = \prod_i \Omega_i$ is the product of the counting measures on each Ω_i, and similarly for the Lebesgue measure.

The other way round, we can also break down a given measure μ on a product space:

Definition 2.8 *Let μ be a measure on a product space $(\Omega, \mathcal{F})$ as above. The marginal distribution μ_i on $(\Omega_i, \mathcal{F}_i)$ is defined by $\mu_i(A_i) := \mu(\prod_j A_j)$, where $A_j := \Omega_j$ for $j \neq i$.*

There is a correspoding product property for densities:

Lemma 2.1 *Let ν be a measure with density f w.r.t. $\mu = \otimes_i \mu_i$.*

(a) *The marginal ν_i has a density f_i w.r.t. μ_i: $f_i(x_i) = \int f(., x_i) d\mu'$ with $\mu' = \otimes_{j \neq i} \mu_i$.*

(b) *ν is a product measure iff $f((x_i)_i) = \prod_i f_i(x_i)$ a.s..*

Proof: Measure theory. □

2.2 Random variables

Often we are not interested in the precise outcome $\omega \in \Omega$ of an experiment, but only in a few properties of ω. Such a property can be modelled by a function $X : \Omega \to \Omega'$.

Definition 2.9 *A measurable function $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ is called a (Ω'-valued) random variable (or observable) on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The probability measure $\mathbb{P}' = \mathbb{P} \circ X^{-1}$ on $(\Omega', \mathcal{F}')$ defined by*

$$\mathbb{P}'(A') = P(X^{-1}A') = P(X \in A')$$

is called the distribution of X.

If the distribution of X has a certain property, we say that X has this property, e.g. X is continuous/discrete if $\mathbb{P} \circ X^{-1}$ is. If we are only interested in the values of the random variables $X_1, X_2, ...$, it is sufficient to know their joint distribution, i.e. the distribution of $X = (X_1, X_2, ...)$. (The underlying probability measure $\mathbb{P}$ and the precise definition of the X_i as functions on Ω are then irrelevant.) The distribution of X_i is the i-th marginal of the joint distribution. An important special case for the joint distribution is that of independence:

Definition 2.10 *The random variables X_i, $i \in I$, are called independent if their joint distribution is the product measure of their distributions, i.e.*

$$\mathbb{P}(X_i \in A_i \forall i \in I) = \prod_i \mathbb{P}(X_i \in A_i) \quad \forall A_i \in \mathcal{F}_i.$$

The above Lemma implies that independence of random variables can be characterized in terms of their densities.

Example: In order to specify a model for rolling a die many times you have two possibilities. In each case we want X_i to denote the value of the i-th roll.

- Specify the probability space: Let $\Omega_1 = \{1,2,3,4,5,6\}$, $\mathcal{F}_1 = \mathfrak{P}(\Omega_1)$, $\mathbb{P}_1$ the uniform distribution on Ω_1. This defines a probability model for rolling a die once. Let $(\Omega, \mathcal{F}, \mathbb{P}) = \otimes_{i\in\mathbb{N}}(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ be the corresponding product space and $X_i : \Omega \to \Omega_1$ the projection on the i-th component.
- Specify the joint distribution: Let X_i be independent random variables, uniformly distributed on $\{1,2,3,4,5,6\}$.

For a real-valued random variable, a number of characteristics are of special interest.

Definition 2.11 *Let X be a real valued random variable. If the corresponding integrals exist we define its*

expectation	$\mathbb{E}(X) = \int X d\mathbb{P}$
variance	$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
moment generating function	$\varphi_X(t) := \mathbb{E}(e^{Xt}) = \sum_n \frac{1}{n!}\mathbb{E}(X^n)t^n, \quad t \in \mathbb{R}.$

$\mathbb{E}_{\mathbb{P}}(X) = \mathbb{E}(X)$ is the mean value of the random variable and $\mathbb{V}(X)$ can be interpreted as the average (quadratic) deviation of X from its mean. The standard deviation $\sigma_X := \sqrt{\mathbb{V}(X)}$ is a measure for the fluctuation of X. The expectation $\mathbb{E}(X^n)$ is called the n-th moment of X, and appears as a coefficient in the Taylor series of $\varphi_X(t)$. For the calculation of expectations the following is useful:

Theorem 2.2 *(Transformation Theorem.) Let μ be a probability measure on $(\Omega, \mathcal{F})$, $X : \Omega \to \Omega'$ be a random variable with distribution $\mu \circ X^{-1} =: \nu$ and $g : \Omega' \to \mathbb{R}$. We have*

$$\mathbb{E}_\mu(g(X)) = \mathbb{E}_\nu(g), \quad \text{i.e.} \int g(X) d\mu = \int g d\nu.$$

In particular we get $\mathbb{E}(g(X)) = \int g(x)f(x)\mu(dx)$ if X has density f w.r.t. μ.

Proof: Measure theory. □

2.3 Conditioning

In many applications we consider experiments such that the outcome of one experiment depends on the outcome of another. We thus need to introduce the notion of a probability distribution depending on an outcome, or equivalently the notion of conditional distribution of a random variable.

Definition 2.12 *Let $(\Omega_i, \mathcal{F}_i)$ be measurable spaces. A kernel from Ω_1 to Ω_2 is a function $p : (\Omega_1, \mathcal{F}_2) \to \mathbb{R}$ such that*

- $\forall x_1 \in \Omega_1 : p(x_1, .)$ *is a measure on* $(\Omega_2, \mathcal{F}_2)$
- $\forall A_2 \in \mathcal{F}_2 : p(., A_2)$ *is $\mathcal{F}_1$-measurable.*

p is called a probability kernel if in the first condition the measure is in fact a probability measure. If μ_1 is a measure on $(\Omega_1, \mathcal{F}_1)$ and p is a kernel from Ω_1 to Ω_2, then we define the measure $\mu = \mu_1 \otimes p$ on the product space by

$$\mu_1 \otimes p(A) := \int \mu_1(dx_1) \int p(x_1, dx_2) 1_A(x_1, x_2). \tag{2.1}$$

We then say that μ admits a disintegration w.r.t. μ_1 and that p is the conditional measure of μ w.r.t. μ_1.

Definition 2.13 *Let X_i be Ω_i-valued random variables and μ_1 a probability measure on Ω_1 and p a probability kernel from Ω_1 to Ω_2. We say that p is the conditional probability distribution of X_2, given X_1, if their joint distribution is of the form $\mu_1 \otimes p$.*

μ_1 is the first marginal of the joint distribution, and thus is uniquely determined, whereas the conditional distribution p may not exist and not be unique. However, the existence is guaranteed if $(\Omega_2, \mathcal{F}_2)$ is a standard Borel space (which is a relatively mild condition) and it is unique up to some suitable a.s.-considerations. If X_1, X_2 are independent the conditional distribution of X_2 given X_1 is just the marginal of X_2 (independent of the value of X_1).

Naively, for the conditional distribution you just take an appropriate cross-section of the distribution (for given x_1), and then renormalize. This procedure is made precise in the following important special case:

Lemma 2.2 *Let (X_1, X_2) have joint density f w.r.t. $\mu = \mu_1 \otimes \mu_2$ and let f_1 be the density of X_1. Then the conditional distribution of X_2 given X_1 has density*

$$f_2(x_2 | X_1 = x_1) := \frac{f(x_1, x_2)}{f_1(x_1)} 1_{\{f_1(x_1) \neq 0\}}.$$

Proof: Homework. □

In probability theory conditioning turns out to be extremely useful. (2.1) says that probabilities can be evaluated by first calculating the probability for a fixed value $X_1 = x_1$, and after that averaging over all possibilities for x_1. This disintegration property remains valid for expectations and many other probabilistic concepts:

Lemma 2.3 *(Tower property for conditional expectation.) Let X_1, X_2 be random variables such that X_1 has distribution μ_1 and X_2 has conditional distribution p for given X_1. Let $g : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be measurable. Then*

$$\mathbb{E}(g(X_1, X_2)) = \mathbb{E}_1(\mathbb{E}_2(g(X_1, X_2))) =: \mathbb{E}(\mathbb{E}(g(X_1, X_2)|X_1)).$$

Here $\mathbb{E}_2$ is the expectation w.r.t. to $p(x_1, .)$ for fixed value of $X_1 = x_1$ and $\mathbb{E}_1$ is the expectation w.r.t. μ_1.

Proof: Homework. □

2.4 Entropy

Entropy is supposed to quantify the amount of disorder within a complex system or equivalently the amount of uncertainty in a situation with uncertain outcome. Since a probability measure models such a situation, we would like to define its entropy. This probabilistic view of entropy is due to Boltzmann, and we will present some of the ideas that motivated his definition in the next chapter. Here we will give an axiomatic approach, showing that in the very simple setting of a distribution on a finite set a few very natural properties already determine how the entropy should be defined.

Theorem 2.3 *(Shannon 1949.) Let $\mathcal{P}(fin)$ be the set of probability distributions on finite state spaces. Suppose $S : \mathcal{P}(fin) \to \mathbb{R}$ satisfies the following properties:*

(a) *If μ, μ' are essentially the same distributions (i.e. ignoring the relabelling of states and states with probability 0), then $S(\mu) = S(\mu')$.*

(b) *S is a continuous function of the density.*

(c) *The uniform distribution has maximal entropy, i.e. if μ is a non-uniform distribution on a finite state space and ν is the uniform distribution on the same state space, then $S(\mu) < S(\nu)$.*

(d) *S satisfies the following additivity property under conditioning: For $\mu = \mu_1 \otimes p \in \mathcal{P}(fin)$*

$$S(\mu) = S(\mu_1) + \mathbb{E}_{\mu_1}(S_2(p)).$$

Here $S_2(p)$ is the entropy of $p(x_1, .)$ for a fixed value of x_1 and $\mathbb{E}_{\mu_1}$ is the expectation w.r.t. $\mu_1(dx_1)$.

Then $S(\mu) = -k \int f \log f d\chi$ for some $k > 0$, where f is the density of μ w.r.t. χ.

Proof: We first determine $S(\nu_n)$, where ν_n is the uniform distribution on $\{1, ..., n\}$. Since ν_{nm} is just a relabelling of $\nu_n \otimes \nu_m$ we have $S(\nu_{nm}) = S(\nu_n) + S(\nu_m)$ by (a) and (d) and thus inductively $S(\nu_{n^m}) = mS(\nu_n)$. By (a) and (c) $S(\nu_n)$ is strictly increasing in n. Thus

$$2^l < n^m < 2^{l+1} \quad \Rightarrow \quad lS(\nu_2) < mS(\nu_n) < (l+1)S(\nu_2).$$

For $m \to \infty$ and $l = \lfloor m \log_2 n \rfloor$ we obtain $S(\nu_n) = \log_2 n \cdot S(\nu_2)$, i.e.

$$S(\nu_n) = k \log n \text{ for some } k > 0.$$

We next consider an arbitrary distribution μ such that the density has rational values only, say $f(x_i) = \frac{a_i}{n}$ for $1 \leq i \leq m$, where $\sum_i a_i = n$. We try to represent μ in terms of uniform distributions by conditioning: Let $\mu_1 = \mu$ and $p(i,.) = \nu_{a_i}$. Since $\mu_1 \otimes p(i,j) = \frac{a_i}{n}\frac{1}{a_i} = \frac{1}{n}$ for all $1 \leq i \leq m, 1 \leq j \leq a_i$, $\mu_1 \otimes p$ is the uniform distribution on a set with n elements. Using (a) and (d) we get $S(\nu_n) = S(\mu) + \sum_i f(x_i) S(\nu_{a_i})$, i.e.

$$S(\mu) = k \sum_i f(x_i)(\log n - \log a_i) = -k \sum_i f(x_i) \log f(x_i) = -k \int f \log f d\chi.$$

Now the result follows from (b) by rational approximation. □

Remark:

- Assumptions (a)-(d) are very natural. The interpretation of (d) is that the uncertainty in picking (x_1, x_2) is the sum of the uncertainty in picking x_1 and the one in picking x_2 for given x_1.
- From a mathematical point of view, the value of k is of little importance. Mathematics convention thus is to set $k = 1$, whereas Physics convention is to set k equal to Boltzmann's constant k_B (whose value depends on the choice of units of temperature and energy). We will ruthlessly switch between these conventions. Also in Mathematics a different sign convention is used sometimes, i.e. $k = -1$ so that in (b) maximal is replaced by minimal.
- In Mathematics it is more common to use the letter H instead of S.
- $S(\mu|\mu_1) := \mathbb{E}_{\mu_1}(S_2(p))$ in (d) is called conditional entropy of μ w.r.t. μ_1.
- In (c) the uniform distribution plays a special role. In many applications it is more appropriate to fix an arbitrary reference distribution ν, and consider an entropy that measures the disorder of a distribution μ w.r.t. ν. This entropy then should be maximal for the reference measure.

Definition 2.14 *Let ν be a measure and μ a distribution on the same space.*

$$S(\mu;\nu) := \begin{cases} -\int f \log f d\nu & \text{if } \mu \text{ has a density } f \text{ w.r.t. } \nu \\ -\infty & \text{otherwise} \end{cases}$$

is called the relative entropy of μ w.r.t. ν or just the entropy of μ if the choice of ν is obvious. If μ is discrete, we set $S(\mu) := S(\mu;\chi)$, and if μ is continuous we set $S(\mu) := S(\mu;\lambda^d)$.

If ν is not finite the integral in the above definition may not exist. The existence of the integral in $[-\infty, \infty)$ for finite ν follows as in part (a) of the next lemma, where we show that the entropy indeed has properties similar to (a)-(d) from the above theorem.

Lemma 2.4 *Let ν be a finite reference measure.*

(a) *$S(\mu;\nu)$ is maximal iff $\mu = \frac{1}{\nu(\Omega)}\nu$, and the maximal value is $\log \nu(\Omega)$. If ν is a probability measure, this implies that $S(\mu;\nu) \leq 0$ with equality iff $\mu = \nu$.*

(b) *$S(.;\nu)$ is concave (and strictly concave where it is finite).*

(c) *If $\nu = \nu_1 \otimes p_\nu$ and $\mu = \mu_1 \otimes p_\mu$, then $S(\mu;\nu) = S(\mu_1;\nu_1) + \mathbb{E}_{\mu_1}(S_2(p_\mu;p_\nu))$.*

Proof:
(a) We have

$$S(\mu;\nu) - \log\nu(\Omega) = -\int f\log f d\nu - \int \log\nu(\Omega)d\mu = -\int f(\log f + \log\nu(\Omega))d\nu$$
$$= \int f\log\frac{1}{\nu(\Omega)f}d\nu \leq \int f(\frac{1}{\nu(\Omega)f} - 1)d\nu = \int \frac{1}{\nu(\Omega)}d\nu - \int f d\nu = 1 - 1 = 0.$$

We have used $\log x \leq x - 1$, where equality holds iff $x = 1$. Thus we have equality above iff $f = 0$ or $\nu(\Omega)f = 1$ ν-a.s., i.e. iff $\mu = \frac{1}{\nu(\Omega)}\nu$.
(b),(c) Homework □

In statistical mechanics we frequently encounter the task of finding a probability distribution μ describing a complicated system in equilibrium, where the only information we have is that certain observables have certain given (average) values. This information does not determine μ uniquely. So additionally one assumes the "Principle of maximum entropy". This is analogous to symmetry assumptions that are made in many probabilistic models. For instance, suppose we have an experiment, where the only given thing is the set of possible outcomes. If we don't have further information about the experiment, the usual symmetry assumption is that all possible outcomes are equally likely. Thus we choose the distribution with maximal entropy so that we have the least possible information about the system, i.e. the maximal disorder of the system. In statistical mechanics this principle additionally is motivated by the second law of thermodynamics, which states that entropy never decreases in a closed system. So after a long time (long enough to reach equilibrium) it makes sense to assume that the system has reached a state of maximum entropy. The following theorem describes what kind of distributions with the given constraints have maximum entropy.

Theorem 2.4 *(Maximum entropy principle.) Let $X_1, ...X_n$ be measurable functions on $(\Omega, \mathcal{F}, \nu)$ and $m_i \in \mathbb{R}$. Let*

$$\mathcal{P}(m) := \{\mu \text{ probability measure on } (\Omega,\mathcal{F}) : \mathbb{E}_\mu(X_i) = m_i \forall i\}.$$

Suppose there are $t_i \in \mathbb{R}$ such that $\mu_t \in \mathcal{P}(m)$, where μ_t is the distribution with density

$$\rho_t = \frac{1}{Z_t}e^{-\sum_i t_i X_i} \quad \text{w.r.t. } \nu, \quad \text{with } Z_t := \int e^{-\sum_i t_i X_i}d\nu < \infty.$$

Then μ_t is the (unique) distribution with maximal entropy (w.r.t. ν) in $\mathcal{P}(m)$, and

$$S(\mu_t;\nu) = \sum_i t_i m_i + \log Z_t.$$

Proof: We have

$$S(\mu_t;\nu) = \int \rho_t\Big(\sum_i t_i X_i + \log Z_t\Big)d\nu = \sum_i t_i m_i + \log Z_t.$$

The rest of the proof relies on the the concavity of $\psi(x) = -x \log x$, which implies that

$$\psi(x) - \psi(y) \geq (x - y)\psi'(x) \quad \forall x, y > 0 \quad \text{with equality iff } x = y.$$

This can either be seen from a picture or by setting $f_x(y) := \text{LHS} - \text{RHS}$, and observing that $f_x'(y) = \psi'(x) - \psi'(y)$, which implies that f_x takes its minimal value at $y = x$ and thus $f_x \geq 0$. Now let $\mu \in \mathcal{P}(m)$ and assume w.l.o.g. that μ has a density ρ w.r.t. ν.

$$S(\mu_t; \nu) - S(\mu; \nu) = \int \psi(\rho_t) - \psi(\rho) d\nu \geq \int (\rho_t - \rho)\psi'(\rho_t) d\nu = \int (\rho - \rho_t)(1 + \log \rho_t) d\nu$$
$$= \int (\rho - \rho_t)(1 - \sum_i t_i X_i - \log Z_t) d\nu = \sum_i t_i m_i - \sum_i t_i m_i = 0$$

with equality iff $\rho_t = \rho$ ν-a.s.. □

2.5 Limit laws

We now consider a sequence of distributions μ_n. Think of μ_n describing the same type of phenomenon, indexed by a parameter n. We would like to investigate the behavior of μ_n in the limit $n \to \infty$. Several issues might be interesting:

- What is the approximate size a_n of typical outcomes w.r.t. μ_n for $n \to \infty$?
- What are the fluctuations around a_n?
- What is the probability of outcomes far away from a_n on an exponential scale?

The answers to these questions can be given in the form of the following limit laws:

Definition 2.15 *We say that distributions μ_n on a metric space (E, d) satisfy*

(a) *a (weak) Law of Large Number (LLN) with limit $c \in E$ iff $\mu_n \to c$ in probability, i.e.*

$$\mu_n(B_\epsilon(c)) \to 1 \qquad \forall \epsilon > 0.$$

(b) *a Central Limit Theorem (CLT) with limit distribution μ iff $\mu_n \to \mu$ in distribution, i.e.*

$$\int f d\mu_n \to \int f d\mu \qquad \forall \text{ bounded, continuous } f : E \to \mathbb{R}.$$

(c) *a Large Deviations Principle (LDP) with rates γ_n and rate function I iff*

$$\limsup_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(C) \leq -\inf I(C) \qquad \forall \text{ closed } C \subset E$$
$$\liminf_{n\to\infty} \frac{1}{\gamma_n} \log \mu_n(U) \geq -\inf I(U) \qquad \forall \text{ open } U \subset E.$$

Here a rate function is any function $I : E \to [0, \infty]$ with $I \not\equiv \infty$ and compact level sets $I^{-1}[0, c]$ $(c \geq 0)$.

Remark:

- It can be shown that the limits in (a) and (b) are uniquely determined, and so is the rate in (c) up to a multiplicative constant.
- (a) says that for large n outcomes near c have probability close to 1.
- For $E = \mathbb{R}$, the condition in (b) is equivalent to $\mu_n(A) \to \mu(A)$ for all $A \in \mathcal{B}(E)$ such that $\mu(\partial A) = 0$. It means that for large n probabilities of "nice" sets and expectations of "nice" random variables are close.
- (c) gives that for large n certain events have exponentially small probability: If $A \in \mathcal{B}(E)$ such that $r = \inf I(\overline{A}) = \inf I(A^o) \neq 0$, we get $\mu_n(A) \approx e^{-r\gamma_n}$.
- A sequence X_n of E-valued random variables is said to satisfy an LLN, CLT, LDP if their distributions do.
- The above limit laws obey a socalled contraction principle: If $h : E \to E'$ is a continuous function and distributions μ_n on E satisfy a LLN with limit c, a CLT with limit μ or a LDP with rates γ_n, I, then $\mu'_n := \mu_n \circ h^{-1}$ satisfy a LLN with limit $c' = h(c)$, a CLT with limit $\mu' = \mu \circ h^{-1}$ or a LDP with rates γ_n, $I'(x) := \inf I(h^{-1}(x))$.

Some classical limit laws (presented in lectures on probability theory) are the ones for sums of iid (independent, identically distributed) random variables:

Theorem 2.5 *Let X_i be independent, identically distributed (iid) real random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $S_n = X_1 + ... + X_n$.*

(a) *If $X_i \in \mathcal{L}^1$ with $m := E(X_i)$, then $\frac{S_n}{n}$ satisfy a LLN with limit m.*

(b) *If $X_i \in \mathcal{L}^2$ with $m := \mathbb{E}(X_i), v := \mathbb{V}(X_i) > 0$, then $S_n^* := \frac{S_n - nm}{\sqrt{n}}$ satisfy a CLT with limit $\mathcal{N}_{0,v}$ (the normal distribution with variance v).*

(c) *If the moment generating function $\varphi(t) = \varphi_{X_i}(t) = \mathbb{E}(e^{X_i t})$ exists for all $t \in \mathbb{R}$, then $\frac{S_n}{n}$ satisfy an LDP with rate $\gamma_n = n$ and rate function $I(x)$ that is the Legendre Transform of $\Lambda(t) = \log \varphi(t)$:*

$$I(x) = \Lambda^*(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda(t)), \quad x \in \mathbb{R}.$$

Proof: For X_n that take only finitely many values, the proof is homework. For a proof for general X_n we refer to a lecture in Probability theory. □

If the corresponding integrability assumptions are satisfied, this gives a lot of information about the limit behavior of the empirical mean $\frac{S_n}{n}$: The approximate size of $\frac{S_n}{n}$ is m. Fluctuations around this value have the shape of a Gaussian function and are of size $\sqrt{v/n}$. Finally for $a > m$, $\mathbb{P}(\frac{S_n}{n} \geq a) \approx e^{-n\Lambda^*(a)}$.

In order to present some ideas that may be used to derive limit laws, we now turn to an instructive example. Let $S \subset \mathbb{R}$ be a finite set and X_i be iid S-valued random

variables with distribution ν such that $\nu(s) > 0 \forall s \in S$. The empirical measure ξ_n is the measure on S with

$$\xi_n(s) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i = s\}}.$$

The empirical measure ξ_n records how often we obtain each possible result when an experiment with distribution ν is repeated independently n times. ξ_n takes values in

$$E_n := \Big\{ \Big(\frac{k_s}{n} \Big)_s \in \mathbb{R}^S : k_s \in \{0, ..., n\}, \sum_s k_s = n \Big\},$$

which we can consider a subset of the set E of all probability measures on $(S, \mathfrak{P}(S))$. E is equipped with the metric d defined by

$$d(\mu, \mu') = \sum_s |\mu(s) - \mu'(s)|.$$

For $\mu \in E_n$ we consider the renormalized version

$$\mu^* = \sqrt{n}(\mu - \nu).$$

In the following we will investigate the limit behavior of the sequence ξ_n.

Lemma 2.5 *Consider $\mu \in E_n$ in the above situation.*

(a) $\mathbb{P}_\nu(\xi_n = \mu) = c_n \prod_s \nu(s)^{n\mu(s)} = c_n e^{n(S(\mu;\nu) - S(\mu))}$, *where* $c_n = \frac{n!}{\prod_s (n\mu(s))!}$.

(b) *If μ maximizes the probability in (a), then μ has to be a $\frac{1}{n}$-discretization of ν:* $|\mu(s) - \nu(s)| < \frac{1}{n} \forall s \in S$.

(c) $(n+1)^{-\#S} e^{nS(\mu;\nu)} \leq \mathbb{P}_\nu(\xi_n = \mu) \leq e^{nS(\mu;\nu)}$.

(d) *If $\mu_n \in E_n$ such that $\mu_n \to \mu$ we have* $\frac{1}{n} \log \mathbb{P}_\nu(\xi_n = \mu_n) \to S(\mu; \nu)$.

(e) $\mathbb{P}(\xi_n = \mu_n) \sim \sqrt{\frac{2\pi n}{\prod_s 2\pi n \nu(s)}} e^{-\frac{1}{2} \sum_s (\mu_n^*(s))^2 / \nu(s)}$ *for $n \to \infty$ uniformly for all sequences $\mu_n \in E_n$ such that $\mu_n^*(s) \in [-c, c]$, where $c > 0$ is fixed.*

Proof: Let $k(s) := n\mu(s)$.
(a) Let A_n be the set of all sequences $(s_i)_i \in S^n$ containing the value s exactly $k(s)$ times. We note that $\#A_n = \frac{n!}{\prod_s k(s)!}$ is the multinomial coefficient and

$$\mathbb{P}_\nu(\xi_n = \mu) = \mathbb{P}_\nu\Big(\bigcup_{(s_i)_i \in A_n} \{X_1 = s_1, ..., X_n = s_n\} \Big) = \sum_{(s_i)_i \in A_n} \prod_{i=1}^{n} \nu(s_i) = \#A_n \prod_{s \in S} \nu(s)^{k(s)}.$$

The product can be rewritten using

$$S(\mu;\nu) - S(\mu) = \sum_s \mu(s) \log \mu(s) - \sum_s \nu(s) \frac{\mu(s)}{\nu(s)} \log \frac{\mu(s)}{\nu(s)} = \sum_s \mu(s) \log \nu(s).$$

(b) If μ does not have this property, there are s,t such that $k(s) \leq n\nu(s) - 1$ and $k(t) \geq n\nu(t) + 1$ and we modify μ to get $\mu' \in E_n$ such that $k'(s) = k(s) + 1$ and $k'(t) = k(t) - 1$. By (a)

$$\frac{\mathbb{P}_\nu(\xi_n = \mu')}{\mathbb{P}_\nu(\xi_n = \mu)} = \frac{k(s)!k(t)!}{(k(s)+1)!(k(t)-1)!}\frac{\nu(s)}{\nu(t)} = \frac{k(t)}{k(s)+1}\frac{\nu(s)}{\nu(t)} \geq \frac{n\nu(t)+1}{n\nu(s)}\frac{\nu(s)}{\nu(t)} > 1,$$

so for μ we don't get a maximal value.
(c) We use (a), but instead of estimating c_n directly, we note that for $\nu = \mu$ (a) gives $\mathbb{P}_\mu(\xi_n = \mu) = c_n e^{n(S(\mu;\mu)-S(\mu))} = c_n e^{-nS(\mu)}$, which implies

$$\mathbb{P}_\nu(\xi_n = \mu) = \mathbb{P}_\mu(\xi_n = \mu)e^{nS(\mu;\nu)}.$$

Now the result follows from $\mathbb{P}_\mu(\xi_n = \mu) \leq 1$ and from the fact that $\mathbb{P}_\mu(\xi_n = \mu')$ is maximal for $\mu' = \mu$ by (b), which implies $\mathbb{P}_\mu(\xi_n = \mu) \geq \frac{1}{\#E_n} \geq (n+1)^{-\#S}$.
(d) This follows from (c) since we have $S(\mu_n;\nu) \to S(\mu,\nu)$ and $\frac{1}{n}\log(n+1)^{-\#S} \to 0$.
(e) We first note that $\mu_n(s) = \nu(s) + \frac{1}{\sqrt{n}}\mu_n^*(s) \to \nu(s)$ for $n \to \infty$. Using Stirling's formula

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

we get the asymptotic size of c_n:

$$c_n \sim \frac{\sqrt{2\pi n}\big(\frac{n}{e}\big)^n}{\prod_s \sqrt{2\pi n\mu_n(s)}\big(\frac{n\mu_n(s)}{e}\big)^{n\mu_n(s)}} \sim \sqrt{\frac{2\pi n}{\prod_s 2\pi n\nu(s)}}e^{nS(\mu_n)}$$

since $\mu_n(s) \to \nu(s)$, the $\frac{n}{e}$ terms cancel and we can write $\mu_n(s)^{\mu_n(s)} = e^{\mu_n(s)\log\mu_n(s)}$. Combining this with (a) we obtain

$$\mathbb{P}_\nu(\xi_n = \mu_n) \sim \sqrt{\frac{2\pi n}{\prod_s 2\pi n\nu(s)}}e^{nS(\mu_n;\nu)}.$$

Using $\frac{\mu_n(s)}{\nu(s)} = 1 + \frac{\mu_n^*(s)}{\sqrt{n}\nu(s)}$ and $\psi(x) = -(1+x)\log(1+x)$ we can write

$$S(\mu_n;\nu) = -\sum_s \nu(s)\frac{\mu_n(s)}{\nu(s)}\log\frac{\mu_n(s)}{\nu(s)} = \sum_s \nu(s)\psi\Big(\frac{\mu_n^*(s)}{\sqrt{n}\nu(s)}\Big)$$

By Taylor expansion $\psi(x) = -x - \frac{1}{2}x^2 + O(x^3)$ and thus we get up to $O(\frac{1}{\sqrt{n^3}})$

$$S(\mu_n;\nu) \approx -\sum_s \nu(s)\frac{\mu_n^*(s)}{\sqrt{n}\nu(s)} - \frac{1}{2}\sum_s \nu(s)\Big(\frac{\mu_n^*(s)}{\sqrt{n}\nu(s)}\Big)^2 = -\frac{1}{2n}\sum_s \frac{\mu_n^*(s)^2}{\nu(s)}.$$

This gives the desired asymptotics. □

Remark:

- Part (d) in the case that ν is the uniform distributon on S was one of the motivations for Boltzmann to define the entropy by the formula given in the last section.

- Part (e) is a local central limit theorem, providing precise uniform asymptotics for probabilities. It is a generalization of the well known local central limit theorem of deMoivre and Laplace (1738).

The above Lemma already implies an LDP, a LLN and a CLT for the empirical measure.

Theorem 2.6 *(Sanov, 1957.) In the above setting, the ξ_n satisfy a LLN with limit ν and an LDP with rate $\gamma_n := n$ and rate function $I := -S(.;\nu)$.*

Proof:
We first note that $I = -S(.;\nu)$ is ≥ 0, $\not\equiv \infty$ and I is a continuous function on the compact space E and thus has compact level sets. To prove the LDP we note that for every $C \subset E$ by the preceding lemma

$$\mathbb{P}_\nu(\xi_n \in C) \leq \sum_{\mu \in C \cap E_n} e^{nS(\mu;\nu)} \leq \#(E_n \cap C) e^{-n \inf I(E_n \cap C)} \leq (n+1)^{\#S} e^{-n \inf I(C)}.$$

From this we get the upper bound of the LDP for arbitrary C. For the lower bound the preceding lemma gives

$$\mathbb{P}_\nu(\xi_n \in U) \geq \sum_{\mu \in U \cap E_n} (n+1)^{-\#S} e^{nS(\mu;\nu)} \geq (n+1)^{-\#S} e^{-n \inf I(U \cap E_n)}.$$

For $\mu \in U$ there are $\mu_n \in E_n \cap U, n \geq n_0$ such that $\mu_n \to \mu$. Thus $\inf I(U \cap E_n) \to \inf I(U)$ since I is continuous. From this the lower bound of the LDP follows.
The LLN can be obtained from the LDP: Let $\epsilon > 0$ such that $\partial B_\epsilon(\nu)$ does not contain any rational point (and note that only countably many ϵ do not have this property). Then $\mathbb{P}_\nu(\xi_n \in \partial B_\epsilon(\nu)) = 0$ and thus the LDP applied to the closure and the interior of $B_\epsilon(\nu))^c$ gives

$$\lim_n \frac{1}{n} \log \mathbb{P}_\nu(\xi_n \notin B_\epsilon(\nu)) = -\inf I(B_\epsilon(\nu)^c) = \sup S(B_\epsilon(\nu)^c;\nu),$$

where we also have used the continuity of I. To obtain the LLN, it suffices to show that the supremum is < 0. This follows since $S(\mu;\nu)$ takes its maximal value 0 iff $\mu = \nu \notin B_\epsilon(\nu)^c$ and $B_\epsilon(\nu)^c$ is compact. □

Theorem 2.7 *In the above setting, the $\xi_n^* := \sqrt{n}(\xi_n - \nu)$ satisfy a CLT with limit distribution γ_ν. γ_ν is a distribution on $E^* = \{\alpha \in \mathbb{R}^S : \sum_s \alpha(s) = 0\}$ with density*

$$\rho(\alpha) \sim \sqrt{\frac{2\pi}{\prod_s 2\pi\nu(s)}} e^{-\frac{1}{2}\sum_s \alpha(s)^2/\nu(s)} \ \ w.r.t.\ \lambda_{E^*}.$$

Let $Z_s \sim \mathcal{N}_{0,\nu(s)}$ be independent normal random variables with variance $\nu(s)$, $s \in S$. Then γ_ν is the distribution of $(X_s)_s$ conditioned on $\sum_s X_s = 0$.

Proof: Let $Q \subset \mathbb{R}^S$ be a cube. By the uniform convergence from (e) of the Lemma,

$$\mathbb{P}(\xi_n^* \in Q) = \sum_{\alpha \in E_n : \alpha^* \in Q} \mathbb{P}(\xi_n = \alpha) \sim \frac{1}{\sqrt{n}^{\#S-1}} \sum_{\alpha^* \in E_n^* \cap Q} \rho(\alpha^*) \to \int_{E^* \cap Q} \rho(\alpha) \lambda_{E^*}(d\alpha).$$

To justify the last step we observe that the sum considered is a Riemann sum; the points of E_n^* form a regular lattice and the density of points of E_n^* in E^* is $\frac{1}{\sqrt{n}^{\#S-1}}$ (the density of points of the form $\sqrt{n}(\frac{k}{n}-\nu)$ $(k \in \mathbb{Z}^S)$ in $\mathbb{R}^S$ is $\frac{1}{\sqrt{n}^{\#S}}$ and the density of sums of components of these vectors is $\frac{1}{\sqrt{n}}$). Thus we have shown

$$\mathbb{E}(g(\xi_n^*)) \to \int g(\alpha)\rho(\alpha)\lambda_{E^*}(d\alpha)$$

for every g of the form 1_Q. For arbitrary continuous bounded g we first choose a cube Q such that $\int_{E^*-Q} \rho(\alpha)\lambda_{E^*}(d\alpha) < \epsilon$. Then $\mathbb{P}(\xi_n^* \notin Q) \to \epsilon$ by the above, and on Q we can approximate g uniformly by functions that are constant on small cubes. This gives the CLT.
For the second characterization of γ_ν we observe that $\sum_s X_s$ is again a normal random variable with mean 0 and variance $\sum_s \nu(s) = 1$. Thus each X_s has density $\rho_s(t) = \frac{1}{\sqrt{2\pi\nu(s)}} e^{-\frac{1}{2}t^2/\nu(s)}$ and $\sum_s X_s$ has density $\rho'(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$. So the joint density of X conditioned on $\sum_s X_s = 0$ is $\frac{\prod_s \rho_s(\alpha_s)}{\rho'(0)} = \rho(\alpha)$. □

2.6 Convergence to equilibrium: Markov chains

In this section we consider a very simple model for convergence to equilibrium in discrete time $(n = 0, 1, 2, ...)$, given by a sequence X_n describing the value of some observable at time n. For example we might look at a container filled with gas and denote X_n the number of atoms in the left half of the container at time n. Since X_n does not give a complete description of the system at time n, the knowledge of the value $X_n = x_n$ does not determine the value of X_{n+1}. We will assume that X_{n+1} depends only on the value of X_n and some random input (which is independent of n).

Definition 2.16 *Let S be a countable set, μ a distribution on S and p a probability kernel from S to S . The sequence of random variables X_n, $n \geq 0$, is called a Markov Chain with state space S, initial distribution μ and transition matrix p if μ is the distribution of X_0 and $p(x_n, .)$ is the distribution of X_{n+1} given that $X_0 = x_0, ..., X_n = x_n$ for all $x_i \in S$.*

By the subsection on conditioning the definition is equivalent to stating that the distribution of $(X_0, ..., X_n)$ is given by $\mu \otimes p \otimes ... \otimes p$, i.e.

$$\mathbb{P}((X_0, ..., X_n) \in A) = \sum_{(x_0,...,x_n)\in A} \mu(x_0)p(x_0, x_1)p(x_1, x_2)...p(x_{n-1}, x_n) \quad \forall A \subset S^{n+1}.$$

If distributions are considered as row vectors $\mu \in \mathbb{R}^S$ and transition probabilities are considered as matrices $p \in \mathbb{R}^{S\times S}$, Markov chain calculations get particularly simple, e.g. $\mathbb{P}(X_n = x_n) = \sum_{x_0,...,x_{n-1}} \mu(x_0)p(x_0, x_1)p(x_1, x_2)...p(x_{n-1}, x_n)$, i.e. the distribution of X_n is given by the row vector μp^n.

An equilibrium state is a state that doesn't change over time, i.e. that is stationary:

Definition 2.17 *A distribution ρ on S is called stationary if $\rho = \rho p$.*

Under some natural assumptions, our model exhibits convergence to equilibrium; starting with any distribution, after a long time the system will be close to equilibrium:

Theorem 2.8 *(Ergodic Theorem for Markov Chains.) Let X_n be a Markov chain with state space S, initial distribution μ and transition matrix p. Suppose that there is a stationary distribution ρ and the chain has the irreducibility and aperiodicity property*

$$\forall x, y \in S \quad \exists N \geq 1 : \quad p^N(x,x) > 0, p^N(y,x) > 0. \tag{2.2}$$

Then the stationary distribution is unique and satisfies $\rho(x) > 0 \; \forall x \in S$, the distribution of X_n converges to ρ, i.e. $\mu p^n \to \rho$, and $S(\mu p^n; \rho) \uparrow S(\rho; \rho) = 0$ for $n \to \infty$.

Proof:
(a) $\rho(x) > 0$ for all $x \in S$: Homework.
(b) Monotonicity of the entropy. Let ν be an arbitrary initial distribution and $N \geq 1$. The joint distribution of (X_0, X_N) is $\nu \otimes p^N$ and has marginals ν and νp^N. The conditional distributions are

$$P(X_N = y | X_0 = x) = \frac{\nu(x) p^N(x,y)}{\nu(x)} = p^N(x,y) \quad \text{and}$$
$$P(X_0 = y | X_N = x) = \frac{\nu(y) p^N(y,x)}{\nu p^N(x)} =: p'_{\nu,N}(x,y).$$

By disintegration we get $\nu \otimes p^N = \nu p^N \otimes p'_{\nu,N}$. By the properties of the entropy

$$S(\nu \otimes p^N; \rho \otimes p^N) = S(\nu; \rho) + \mathbb{E}_\nu(S_2(p^N; p^N))$$
$$S(\nu p^N \otimes p'_{\nu,N}; \rho p^N \otimes p'_{\rho,N}) = S(\nu p^N; \rho p^N) + \mathbb{E}_{\nu p^N}(S_2(p'_{\nu,N}; p'_{\rho,N})).$$

The left hand sides are equal, $\rho p^N = \rho$ by stationarity and $S_2(p^N, p^N) = 0$, thus

$$S(\nu; \rho) = S(\nu p^N; \rho) + \mathbb{E}_{\nu p^N}(S_2(p'_{\nu,N}; p'_{\rho,N})) \leq S(\nu p^N; \rho). \tag{2.3}$$

This implies that $S(\mu p^n; \rho) \leq S(\mu p^m; \rho)$ for $n \leq m$.

(c) Strict monotonicity of the entropy. Suppose we have equality for all $N \geq 1$ in (2.3). This means that the conditional entropy is 0 and thus $p'_{\nu,N}(i,.) = p'_{\rho,N}(i,.)$ for all i such that $\nu p^N(i) > 0$. Now let $x \in S$ such that $\nu(x) > 0$ and $y \in S$ arbitrary. Choosing N according to (2.2), we get

$$1 = \frac{p'_{\nu,N}(x,y)}{p'_{\rho,N}(x,y)} = \frac{\frac{\nu(y)p^N(y,x)}{\nu p^N(x)}}{\frac{\rho(y)p^N(y,x)}{\rho p^N(x)}} = \frac{\nu(y)}{\rho(y)} \frac{\rho p^N(x)}{\nu p^N(x)} \quad \text{and similarly} \quad 1 = \frac{\nu(x)}{\rho(x)} \frac{\rho p^N(x)}{\nu p^N(x)}.$$

In particular this implies that $\nu(y) = \nu(x)\frac{\rho(y)}{\rho(x)} = c\rho(y)$ and thus $\nu = \rho$.

(d) Uniqueness of the stationary distribution. Suppose that ν is stationary, then $\nu = \nu p^N$ for all N and thus we have equality in (2.3). By (c) this implies $\nu = \rho$.

(e) The sequence μp^n is tight: $\forall \epsilon > 0 \; \exists$ finite $K_\epsilon \subset S \; \forall n : \mu p^n(K_\epsilon) \geq 1 - \epsilon$. Homework.

(f) Suppose that for some subsequence $\mu p^{n_k}(x) \to \nu(x)$ converges, then $\nu(S) = 1$. For finite $K \subset S$ we have $\nu(K) = \lim_{k\to\infty} \mu p^{n_k}(K) \leq 1$, since limits and finite sums interchange and μp^{n_k} is a probability measure for all k. Since $\nu(K) \to \nu(S)$ for $K \uparrow S$, we have $\nu(S) \leq 1$. To show that $\nu(S) \geq 1$, let $\epsilon > 0$ and choose K_ϵ according to (e). Then $\nu(S) \geq \nu(K_\epsilon) = \lim_{k\to\infty} \mu p^{n_k}(K_\epsilon) \geq 1 - \epsilon$, and thus $\nu(S) \geq 1$.

(g) Convergence. Suppose $\mu p^n \not\to \rho$, then by compactness of $[0,1]^S$ there is a subsequence n_k such that $\mu p^{n_k}(x) \to \nu(x) \in [0,1]$ for all x, where $\nu \neq \rho$. By (f) ν is a probability measure. By (c) there is an $N > 0$ such that

$$S(\nu;\rho) < S(\nu p^N;\rho) = \lim_k S(\mu p^{n_k} p^N;\rho) \leq \lim_k S(\mu p^{n_{k+N}};\rho) = S(\nu;\rho),$$

where we have also used $\nu = \lim_k \mu p^{n_k}$, the continuity of $S(.;\rho)$ and the monotonicity from (b) applied to $n_k + N \leq n_{k+N}$. The above is a contradiction. □

Remark:

- By irreducibility it is possible to get from any state to any other state in a finite number of steps. Otherwise there may be several components of S (each with its own stationary distribution) that can't be reached from each other.
- By aperiodicity a state x can be reached from other states in the same number of steps. Otherwise it might happen that x only shows up in the even (or odd) steps depending on the initial state. This might give a periodic behavior instead of convergence.
- Not every Markov chain has a stationary distribution, but if S is finite there is at least one.
- To find a stationary distribution, you have to solve the system of linear equations $\rho = \rho p$. Often it is possible, to guess what ρ might be.
- For a Markov chain started at x let T_x denote the time until the chain returns to x for the first time, and N_y the number of visits to y until that time. In the context of the ergodic theorem, it can be shown that $\rho(y) = \frac{\mathbb{E}(N_y)}{\mathbb{E}(T_x)}$ and $\rho(x) = \frac{1}{\mathbb{E}(T_x)}$. In particular $\mathbb{E}(T_x) < \infty$, i.e. x is "positive recurrent".

Example: Ehrenfest's urn model. We consider a container divided symmetrically by a wall. There is a gas in the left part and a vacuum in the right part. For the number of particles in the gas, $N \gg 10^{10}$ is realistic. We are only interested in the number of particles remaining in the left part after the wall is removed. As a toy model for the resulting diffusive behavior we consider a Markov chain, where at every discrete time step we choose a particle completely at random and with probability $\frac{1}{2}$ we put it into the other part of the container. (Otherwise we do nothing.) The chain has state space $S = \{0, ..., N\}$, initial distribution $\mu = \delta_N$ and a transition matrix given by

$$\begin{aligned} p(k,k) &= \frac{1}{2} && \text{for } k \in \{0,...,N\} \\ p(k,k+1) &= \frac{N-n}{N}\frac{1}{2} && \text{for } k \in \{0,...,N-1\} \\ p(k,k-1) &= \frac{n}{N}\frac{1}{2} && \text{for } k \in \{1,...,N\} \end{aligned}$$

and $p(k,l) = 0$ otherwise. In equilibrium every particle should be in the left part with probability $\frac{1}{2}$, independently from the other particles. From the discussion in the last section we thus guess that $\rho(k) = \frac{N!}{k!(N-k)!}\frac{1}{2^N}$ is stationary, and it is easy to verify $\rho p = \rho$. The irreducibility and aperiodicity condition (2.2) holds for the given N. Thus the ergodic theorem applies. By the results of this and the last section we conclude for the number X_n of particles in the left part:

- The distributino of X_n converges to ρ with increasing entropy.
- For large n, $\frac{X_n}{n} \approx \frac{1}{2}$ with normal fluctuations of size $\frac{1}{\sqrt{N}}$ and the probability to be bounded away from $\frac{1}{2}$ exponentially small.
- It is possible that at some time all particles will again be in the left part, but this is very unlikly with probability $\rho(N) \approx \frac{1}{2^N}$. The expected waiting time for this to happen is $\mathbb{E}(T_N) = 2^N$.

3 Ensemble theory

3.1 General setup

The aim in this section is to describe the equilibrium states of a complex system by distributions. The input from Physics is:

- The set Ω of all possible states of the system. $\omega \in \Omega$ is called microstate or configuration. A microstate should give complete information about the system.
- The Hamiltonian $H : \Omega \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. $H(\omega)$ describes the energy of the system when it is in state ω.
- A set of parameters (e.g. volume,...) so that for every choice of values for these parameters we expect to get an equilibrium state.
- A set of observables (e.g. pressure,...) that determine the behavior of the system on a macroscopic scale. A given set of values of these observables is called macrostate of the system. A definition of an observable as a function on Ω is called microscopic description of the observable. A definition of an observable as a function of the parameters of the system, is called macroscopic description.

To specify a probabilistic model (= ensemble) for this situation we need to define

- a set of equilibrium distributions indexed by the given parameters
- macroscopic descriptions of all observables
- what should be the thermodynamic limit, i.e. scaling rules for the parameters

The model should have at least the following properties:

- The equilibrium distributions should be invariant under the dynamics.
- The microscopic description of the observables should be equivalent to the macroscopic description, at least in the thermodynamic limit.

If we have a model for thermodynamics in the usual sense, we also should check

- the second law of thermodynamics holds, at least in the thermodynamic limit (orthodicity problem).

If we have more than one model for a given situation, we also should check that

- the models give the same predictions, at least in the the thermodynamic limit (equivalence of ensembles). Obviously these predictions should also match the results of experiments.

In the following we introduce a couple of classical ensembles: the microcanonical, the canonical and the grand canonical ensemble (as introduced by Boltzmann and Gibbs). We formulate each of these in the context of interacting particles, where each particle

has a mass m_i, a position q_i and a momentum p_i and the Hamiltonian for the interaction of N particles is of the form

$$H(p,q) = \sum_i \frac{p_i^2}{2m_i} + \sum_{i<j} \Phi(|q_i - q_j|),$$

where $\Phi : \mathbb{R}_+ := [0,\infty) \to \mathbb{R}$ is a given pair interaction. However, these ensembles can be formulated in many contexts, e.g.

- more general interactions
- ignoring the momenta
- further internal properties of particles, e.g. magnetic spin, charge, ...
- fixing particle positions at vertices of a lattice

3.2 Microcanonical distribution

Here we fix the number of (indistinguishable) particles N, the energy E and the domain $\Lambda \subset \mathbb{R}^d$ with volume $V = \lambda^d(\Lambda)$. (Often the shape of Λ is fixed, e.g. a cubic shape, so that Λ depends on V only.) An appropriate state space is

$$\Omega = (\Lambda \times \mathbb{R}^d)^N / \sim .$$

Here by $/\sim$ we identify configurations that differ only by the labelling of the particles. This is necessary since the particles are indistinguishable. One way to do this, is to define some order on $\Lambda \times \mathbb{R}^d$ and to admit only N-tuples into Ω that are in increasing order. We note that this identification procedure reduces the Lebesgue-volume of Ω by a factor of $\frac{1}{N!}$. Without this identification, we run into the so-called "Gibbs paradox" and have to introduce "correct Boltzmann" counting to resolve this paradox.

By Liouville's Theorem λ_Ω, the Lebesgue-measure on Ω, is invariant under the Hamiltonian dynamics, but we would like to restrict this measure to $\Omega_E = \{\omega \in \Omega : H(\omega) = E\}$. One way to do this is by conditioning:

Definition 3.1 *Suppose λ_Ω admits a disintegration w.r.t. the energy $\lambda_\Omega = \lambda \otimes p$ such that $p(E,.)$ is concentrated on Ω_E. If $\mu'_E := p(E,.)$ is a finite measure on Ω_E we define the microcanonical partition function by*

$$Z_\Lambda^{E,N} = Z(\Lambda, E, N) := \mu'_E(\Omega_E)$$

and the microcanonical distribution by the normalization of μ'_E: $\mu_\Lambda^{E,N} := \frac{1}{Z_\Lambda^{E,N}} \mu'_E$.

We can rewrite the above definition in the form

$$\lambda_\Omega(d\omega) = \lambda(dE) p(E, d\omega) = \lambda(dE) Z_\Lambda^{E,N} \mu_\Lambda^{E,N}(d\omega).$$

Thus $Z_\Lambda^{E,N}$ is the density of the energy w.r.t. λ and $\mu_\Lambda^{E,N}$ is the uniform distribution on the state space conditioned on the total energy to be E.

In order to motivate the above defintion by the maximum entropy principle we consider

$$\frac{1}{\delta}\lambda_{\Omega_{[E,E+\delta]}}(A) = \frac{1}{\delta}\lambda_{\Omega}(A \cap \{E \leq H \leq E+\delta\}) = \frac{1}{\delta}\int_E^{E+\delta} p(e,A)de \to p(E,A) = \mu'_E(A)$$

for $\delta \to 0$, which holds under suitable continuity assumptions. Thus we can think of μ'_E as the limit of $\frac{1}{\delta}\lambda_{\Omega_{[E,E+\delta]}}$ and of μ_E as the limit of the uniform distributions on $\Omega_{[E,E+\delta]}$. In particular, since Lebesgue-measure is the "most uniform" measure in the continuous setting, μ_E is the "most uniform" measure on Ω_E and thus a natural choice for the reference measure. By the properties of entropy, μ_E maximizes the entropy w.r.t. μ'_E, and we have

$$S(\mu_E;\mu'_E) = -k\int d\mu'_E \frac{1}{\mu'_E(\Omega_E)}\log\frac{1}{\mu'_E(\Omega_E)} = k\log\mu'_E(\Omega_E) = k\log Z(\Lambda,E,N).$$

Here by Physics convention we have inserted Boltzmann's constant k into the definition of relative entropy.

Definition 3.2 *The entropy of the micronanical ensemble is defined to be*

$$S(\Lambda,E,N) := k\log Z(\Lambda,E,N).$$

Our definition of the microcanonical ensemble is rather abstract. If the energy levels Ω_E are surfaces in Ω everything can also expressed in terms of the surface measure σ_E.

Theorem 3.1 *Suppose that H is differentiable such that $\nabla H \neq 0$, then μ'_E has density $\frac{1}{\|\nabla H\|}$ w.r.t. surface measure σ_E, and in particular $Z(\Lambda,E,N) = \int \frac{1}{\|\nabla H(\omega)\|}\sigma_E(d\omega)$. Here $\|.\|$ is the Euclidean norm. We note that here σ_E denotes surface measure w.r.t. the identification $/\sim$, i.e. surface measure with an additional factor $\frac{1}{N!}$.*

Proof: Since $\nabla H \neq 0$, we can decompose Ω into open sets Ω_i such that $\partial_{\omega_i}F(\omega,E) \neq 0$ for all $\omega \in \Omega_i$. On Ω_i we can locally solve $H(\omega) = E$ for ω_i, i.e. there is a differentiable function f_i such that for $\omega' = (\omega_j)_{j\neq i}$ we have $H(\omega_i,\omega') = E$ iff $\omega_i = f_i(\omega',E)$. For the surface measure σ_E we have the representation

$$\sigma_E(d\omega) = \sqrt{1+\|\nabla_{\omega'}f_i(\omega',E)\|^2}d\omega' = \frac{\|\nabla H(f_i(\omega',E),\omega')\|}{|\partial_{\omega_i}H(f_i(\omega',E),\omega'|}d\omega'.$$

The first equality is by definition of the surface masure and the second can be seen be a picture, or by differentiating $H(f_i(\omega',E),\omega') = E$ w.r.t. ω', which gives

$$\partial_{\omega_i}H(f_i(\omega',E),\omega')\nabla_{\omega'}f_i(\omega',E) + \nabla_{\omega'}H(f_i(\omega',E),\omega') = 0,$$

and using $\nabla H = (\partial_{w_i}H, \partial_{\omega'}H)$, which implies $\|\nabla H\|^2 = |\partial_{w_i}H|^2 + \|\partial_{\omega'}H\|^2$. Combining the above transformation with the transformation theorem for Lebesgue measure applied to $\omega_i \mapsto E = H(\omega_i,\omega')$, we get for arbitrary measurable $A \subset \Omega_i$:

$$\begin{aligned}\int d\omega 1_A(\omega) &= \int d\omega' \int d\omega_i 1_A(\omega_i,\omega') = \int d\omega' \int dE \frac{1}{|\partial_{\omega_i}H(f_i(\omega',E),\omega')|}1_A(f_i(\omega',E),\omega') \\ &= \int dE \int \sigma_E(d\omega)\frac{1}{\|\nabla H(\omega)\|}1_A(\omega)\end{aligned}$$

By decomposition of A, the above is also true for arbitrary measurable $A \subset \Omega$, which gives the desired disintegration. $\square$

After defining distributions corresponding to equilibrium states we can now give macroscopic descriptions of the most common observables and define the thermodynamic limit. The second law of thermodynamics in the form $dS = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}dN$ motivates the definition of T, P, μ as partial derivatives of $S(E, V, N)$:

Definition 3.3 *In the microcanonical ensemble the energy E, the number of particles N and the volume V are given as parameter values, and we define*

$$\begin{aligned} &T(E,V,N) \text{ (temperature) by} && \frac{1}{T} = \partial_E S(E,V,N) = \frac{\partial S}{\partial E}\Big|_{V,N} \\ &P(E,V,N) \text{ (pressure) by} && \frac{P}{T} = \partial_V S(E,V,N) = \frac{\partial S}{\partial V}\Big|_{E,N} \\ &\mu(E,V,N) \text{ (chemical potential) by} && -\frac{\mu}{T} = \partial_N S(E,V,N) = \frac{\partial S}{\partial N}\Big|_{E,V} \end{aligned}$$

For the thermodynamic limit we fix an energy density u and a particle density ρ. We then let $N, E, V \to \infty$ such that $\frac{E}{V} \to u$ and $\frac{N}{V} \to \rho$.

We note that in the definition of μ we are a bit sloppy since we take a derivative w.r.t. a discrete variable. Here ∂_N can be interpreted to be a discrete derivative. For the thermodynamic limit, we stress that we have chosen a specific domain (e.g. cubic) for every value of V. Of course, we now can also introduce the various free energies/thermodynamic potentials A, Φ, G, H in terms of the above observables, e.g. $A = E - TS$.

For any given Hamiltonian, we can now investigate properties of the microcanonical ensemble, e.g. the behavior of interesting random variables (such as the kinetic energy per particle, the microcanonical pressure, the number of particles in a certain region,...) or try to derive the equation of state.

Example: Let us consider the free gas in three dimensions with indistinguishable particles. Here $\Phi = 0$ so that

$$H = \sum_i \frac{p_i^2}{2m}.$$

$\Omega_E = (\Lambda^N \times S_{3N,R})/\sim$, where $S_{3N,R}$ is the surface of a $3N$-dimensional sphere with radius $R = \sqrt{2mE}$. On Ω_E

$$\|\nabla H\| = \|(\frac{p_i}{m})_i\| = \frac{R}{m}.$$

Thus μ_E' has density $\sqrt{\frac{m}{2E}}$ w.r.t. $\sigma_E = \lambda_\Lambda^N \otimes \sigma_{S_{3N,R}}$, In particular μ_E is uniform on $(V^N \times S_{3N,R})/\sim$ and since the surface of an n-dimensional sphere with radius r is given by

$$\sigma(n, r) := r^{n-1} \frac{2\pi^{n/2}}{\Gamma(n/2)},$$

the partition function is

$$Z(V,E,N) = \frac{m}{R}\sigma_E(\Omega_E) = \frac{m}{R}\lambda(\Lambda^N/\sim)\sigma(3N,R) = \frac{m}{R}\frac{V^N}{N!}R^{3N-1}\frac{2\pi^{3N/2}}{\Gamma(3N/2)}$$
$$= \frac{1}{E}\frac{V^N}{N!\Gamma(3N/2)}(2\pi mE)^{3N/2} \sim \frac{1}{\sqrt{6}\pi NE}\Big(\frac{Ve}{N}\Big)^N\Big(\frac{4\pi emE}{3N}\Big)^{3N/2}$$

for $N \to \infty$ using Stirling. From the partition sum we immediately get the entropy

$$S(V,E,N) = k\Big(-\log(\sqrt{6}\pi NE) + N\log\frac{Ve}{N} + \frac{3N}{2}\log\frac{4\pi emE}{3N}\Big) + o(1).$$

Taking partial derivatives we obtain

$$\frac{1}{kT} = \frac{3N-2}{2E}, \quad \frac{P}{kT} = \frac{N}{V}, \quad -\frac{\mu}{kT} = \log\frac{Z(V,E,N)}{Z(V,E,N-1)} = \log\frac{V(2\pi mE)^{3/2}\Gamma(\frac{3N-3}{2})}{N\Gamma(\frac{3N}{2})}.$$

Thus the equation of state $PV = NkT$ holds for all parameter values, and in the thermodynamic limit we obtain

$$\frac{S}{N} \sim k\Big(\log\frac{Ve}{N} + \frac{3}{2}\log\frac{4\pi emE}{3N}\Big), \quad T \sim \frac{2}{3}\frac{E}{kN}, \quad \mu \sim -kT\Big(\log\frac{V}{N} + \frac{3}{2}\log\frac{4\pi mE}{3N}\Big).$$

In particular we see that N, V, E, S are extensive quantities (scaling with N in the thermodynamic limit), whereas T, P, μ are intensive quantities (scaling as constants).

We also might be interested in the distribution of the position X and momentum Y of a single particle. By definition of μ_E, X is uniformly distributed on V and independent of Y, and the distribution of Y can be calculated as the marginal of the uniform distribution on S_R. We get the density

$$f_Y(p) = \frac{1}{\sigma(3N,R)}\sigma(3N-3,\sqrt{R^2-p^2}) = \frac{(1-\frac{p^2}{R^2})^{(3N-4)/2}}{R^3\pi^{3/2}}\frac{\Gamma(3N/2)}{\Gamma((3N-3)/2)}$$

In the thermodynamic limit $(1-\frac{p^2}{R^2})^{(3N-4)/2} \sim e^{-\frac{p^2}{2}\frac{3N}{2mE}}$ (uniformly on compact sets) and $\frac{\Gamma(3N/2)}{R^3\pi^{3/2}\Gamma((3N-3)/2)} \sim \frac{(3N/2)^{3/2}}{(2\pi mE)^{3/2}} = (\frac{3N}{4\pi mE})^{3/2}$. Thus

$$f_Y(p) \sim (\frac{3N}{4\pi mE})^{3/2}e^{-\frac{p^2}{2}\frac{3N}{2mE}} = \prod_{i=1}^{3}\frac{1}{\sqrt{2\pi v}}e^{-\frac{p_i^2}{2v}}, \quad \text{where } v = \frac{2mE}{3N},$$

uniformly on compact sets. This implies that Y satisfies a CLT in the thermodynamic limit. The limit distribution is called the Maxwell distribution: the three momentum components are independent and normally distributed.

3.3 Canonical distribution

The microcanonical distribution is very natural, but since it is defined by conditioning, calculations get complicated. The canonical distribution is computationally much easier. Here we fix the number of (indistinguishable) particles N, and the domain

$\Lambda \subset \mathbb{R}^d$ with volume $V = \lambda^d(\Lambda)$, but instead of the energy we fix the temperature. By convention the parameter used is the inverse temperature

$$\beta := \frac{1}{kT}.$$

As in the microcanonical case, the state space is

$$\Omega = (\Lambda \times \mathbb{R}^d)^N / \sim .$$

Again we stress that because of the identification $/\sim$, λ_Ω has an additional factor of $\frac{1}{N!}$ as compared to the usual Lebesgue-measure. W.r.t. a canonical distribution the energy is not fixed, but it is a random variable H; however we would like the expectation of H to have a specific appropriate value E. By the principle of maximum entropy this motivates to define the canonical distribution in terms of an exponential density of the form $e^{-\beta H}$, where β has to be chosen to produce the given value of E.

Definition 3.4 *We define the canonical partition function by*

$$Z_\Lambda^{\beta,N} = Z(\Lambda, \beta, N) := \int_\Omega e^{-\beta H} d\lambda_\Omega$$

and the canonical distribution $\mu_\Lambda^{\beta,n}$ to be the probability measure with density

$$\rho_\Lambda^{\beta,N} := \frac{1}{Z_\Lambda^{\beta,N}} e^{-\beta H} \ \textit{w.r.t.}\ \lambda_\Omega.$$

For this definition to be meaningful we have to assume that the partition function has a finite value. By the above motivation the energy E should be defined to be the expectation of H w.r.t. the corresponding distribution, and by the maximum entropy theorem

$$S(\mu_\Lambda^{\beta,n}; \lambda_\Omega) = k(\beta E + \log Z(\Lambda, \beta, N)).$$

By the thermodynamics definition of the (Helmholtz) free energy $A = E - TS$ we see that we can identify A with $-\frac{1}{\beta} \log Z$. This quantity plays a fundamental role in the definition of the other thermodynamic variables: We note that

$$\partial_\beta \log Z_\Lambda^{\beta,N} = \frac{\partial_\beta Z_\Lambda^{\beta,N}}{Z_\Lambda^{\beta,N}} = \frac{1}{Z_\Lambda^{\beta,N}} \int \partial_\beta e^{-\beta H} d\lambda = -\int H d\mu_\Lambda^{\beta,n} = -E$$

under suitable integrability conditions, and $A = E - TS = -PV + \mu N$ implies that P and μ can be obtained from A as corresponding partial derivatives.

Definition 3.5 *In the canonical ensemble the temperature $T = \frac{1}{k\beta}$, the number of particles N and the volume V are given as parameter values, and we define*

$E(\beta, V, N)$ *(energy) by*	$E = -\partial_\beta \log Z(\beta, V, N)$
$A(\beta, V, N)$ *(free energy) by*	$A = -\frac{1}{\beta} \log Z(\beta, V, N)$
$P(\beta, V, N)$ *(pressure) by*	$-P = \partial_V A(\beta, V, N) = \frac{\partial A}{\partial V}\big\vert_{\beta,N}$
$\mu(\beta, V, N)$ *(chemical potential) by*	$\mu = \partial_N A(\beta, V, N) = \frac{\partial A}{\partial N}\big\vert_{\beta,V}$

For the thermodynamic limit we fix a a particle density ρ. We then let $N, V \to \infty$ such that $\frac{N}{V} \to \rho$, while keeping β fixed.

The calculations in the free gas example are muchg easier in the canonical ensemble.

Example: For doing computations for the canonical ensemble of the free gas, again we have to start by computing the partition function

$$Z = \int e^{-\beta H} d\lambda = \frac{1}{N!} V^N \prod_{i=1}^{3N} e^{-\beta \frac{\omega_i^2}{2m}} d\omega_i = \frac{1}{N!} V^N (\frac{2\pi m}{\beta})^{3N/2}.$$

So

$$E = \frac{3N}{2\beta} \quad \text{and} \quad A = \frac{1}{\beta}\Big(\log N! - N \log V + \frac{3N}{2} \log \frac{\beta}{2\pi m}\Big),$$

and

$$P = \frac{N}{\beta V} \quad \text{and} \quad \mu = \frac{1}{\beta}\Big(- \log \frac{V}{N} + \frac{3}{2} \log \frac{\beta}{2\pi m}\Big).$$

Again A, E, S can be seen to be extensive and P, μ are intensive. The corresponding limits are the same as for the microcanonical ensemble (equivalence of ensembles!) and we get the same equation of state $N = \beta V P$. Directly from the definition of the canonical ensemble we get that the positions and momenta of the particles are independent of each other and the positions are uniformly distributed over Λ and the momenta have Maxwell-distribution (even for finite N). In the canonical case H is a random variable with expectation $\mathbb{E}(H) = E$, and we might also be interested in the fluctuations. Since H is a sum of iid random variables, the iid CLT implies that the fluctuations are of the order $\sqrt{N}$. Another way to see this, is to observe that

$$\partial_\beta^2 \log Z = \partial_\beta (\frac{1}{Z} \partial_\beta Z) = \frac{1}{Z} \partial_\beta^2 Z - \frac{1}{Z^2} (\partial_\beta Z)^2 = \mathbb{E}(H^2) - \mathbb{E}(H)^2 = \mathbb{V}(H)$$

and thus $\mathbb{V}(H) = -\partial_\beta E = \frac{3N}{2\beta^2}$. So the fluctuation of H is $\sqrt{\frac{3N}{2\beta^2}}$. In particular the energy density $\frac{H}{N}$ has vanishing fluctuations in the thermodynamic limit, i.e. $\frac{H}{N} \to \frac{E}{N}$.

3.4 Grand canonical distribution

In the canonical distribution we already have abandoned the idea of fixed energy, and we now want to do the same for the number of particles: For the grand canonical distribution we fix the domain $\Lambda \subset \mathbb{R}^d$ with volume $V = \lambda^d(\Lambda)$, the inverse temperature β and the chemical potential μ, and we allow the energy and particle number to vary. The state space now is

$$\Omega = \bigcup_{n \geq 0} \Omega_n \quad \text{with } \Omega_n = (\Lambda \times \mathbb{R}^d)^n / \sim .$$

The elements of Ω are all finite unordered sets of tuples (x, p) with $x \in \Lambda$ and $p \in \mathbb{R}^d$. The most natural choice of a reference measure λ_Ω is a measure such that its restriction to Ω_n is the reference measure from the previous section, i.e. $\lambda = \sum_n \lambda_{\Omega_n}$. Let $H(\omega)$ and $N(\omega)$ denote the energy and the number of particles of a configuration $\omega \in \Omega$. The principle of maximum entropy motivates to define the grand canonical distribution in terms of an exponential density of the form $e^{-\beta H - \beta' N}$, where the parameters β, β' determine the expectation of the energy and number of particles.

Definition 3.6 *We define the grand canonical partition function by*

$$Z_\Lambda^{\beta,\mu} = Z(\Lambda,\beta,\mu) := \int_\Omega e^{-\beta H+\beta\mu N} d\lambda_\Omega$$

and the canonical distribution $\mu_\Lambda^{\beta,\mu}$ to be the probability measure with density

$$\rho_\Lambda^{\beta,\mu} := \frac{1}{Z_\Lambda^{\beta,\mu}} e^{-\beta H+\beta\mu N} \ \ w.r.t.\ \lambda_\Omega.$$

For this definition to be meaningful we have to assume that the partition function has a finite value. Instead of μ sometimes the activity parameter z is used, which is defined as

$$z = e^{\beta\mu}.$$

Energy E and particle number $\bar{N}$ should be defined to be the expectation $\mathbb{E}(H)$ and $\mathbb{E}(N)$ w.r.t. the corresponding distribution, and by the maximal entropy theorem

$$S(\mu_\Lambda^{\beta,\mu};\lambda_\Omega) = k(\beta E - \beta\mu\bar{N} + \log Z(\Lambda,\beta,\mu)).$$

By the thermodynamics definition of the grand free energy $G = E - TS - \mu\bar{N}$ we see that we can identify G with $-\frac{1}{\beta}\log Z$. Again $\log Z$ plays a fundamental role: E and $\bar{N}$ can be expressed in terms of $\partial_\beta \log Z$ and $\partial_\mu \log Z$ and the pressure P can be obtained from $G = E - TS - \mu\bar{N} = -PV$.

Definition 3.7 *In the canonical ensemble the temperature T, the chemical potential μ and the volume V are given as parameter values, and we define*

$\bar{N}(\beta,V,\mu)$ *(number of particles) by*	$\beta\bar{N} = \partial_\mu \log Z(\beta,V,\mu)$
$E(\beta,V,\mu)$ *(energy) by*	$-E + \mu\bar{N} = \partial_\beta \log Z(\beta,V,N)$
$G(\beta,V,\mu)$ *(grand free energy) by*	$G = -\frac{1}{\beta}\log Z(\beta,V,N)$
$P(\beta,V,\mu)$ *(pressure) by*	$P = -\frac{G}{V}.$

For the thermodynamic limit we let $V \to \infty$ such that β,μ are kept fixed.

At least in the definition of the thermodynamic limit it can be seen that the grand canonical ensemble is even easier than the canonical for computations.

Example: For the grand canonical ensemble of the free gas, we compute

$$Z = \int e^{-\beta H+\beta\mu N} d\lambda = \sum_n \frac{1}{n!} e^{\beta\mu n} V^n \prod_{i=1}^{3n} e^{-\beta\frac{\omega_i^2}{2m}} d\omega_i = \sum_n \frac{1}{n!} e^{\beta\mu n} V^n (\frac{2\pi m}{\beta})^{3n/2},$$

which gives

$$\log Z = V e^{\beta\mu} (\frac{2\pi m}{\beta})^{3/2}.$$

From this we obtain

$$\bar{N} = V e^{\beta\mu} (\frac{2\pi m}{\beta})^{3/2}, \quad E = -\frac{3}{2\beta} V e^{\beta\mu} (\frac{2\pi m}{\beta})^{3/2}, \quad G = -\frac{1}{\beta} V e^{\beta\mu} (\frac{2\pi m}{\beta})^{3/2}$$

and thus $P = \frac{1}{\beta V} V e^{\beta\mu} (\frac{2\pi m}{\beta})^{3/2}$, i.e. $P = \frac{\bar{N}}{\beta V}$. Again in the thermodynamic limit all quantities have the appropriate scaling behavior. As in the canonical ensemble it can be seen that both energy and particle number have fluctuations that are small compared to their values in the thermodynamic limit. The distribution of N is called a Poisson distribution with parameter $\alpha = \log Z$.

3.5 Stability

For the free gas we have seen that the entropy, energy and pressure are extensive quantities, scaling with the system size, and the distribution of the position of particles is rather homogeneous. Although this is the way we would like our system to behave, this need not be the case in general. Some bad things that might happen are

- Coalescence catastrophe: This happens if the interaction is too attractive (either the long range or the short range part) so that all particles lump together.
- Evaporation catastrophe: This happens if the interaction is too repulsive (mostly the long range part), so that the particles try to maximize their mutual distance, i.e. most particles will stick to the boundary of Λ.

There are various sorts of conditions on the interaction that are sufficient for preventing the above catastrophes, usually called temperedness and stability conditions. Here we consider the following ones:

Definition 3.8 *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be an interaction.*

(a) *Φ is called tempered if there is an $R > 0$ such that $\Phi(r) \leq 0$ for $r > R$.*

(b) *Φ is called stable if there is a $B > 0$ such that*

$$\Phi(q) := \sum_{1 \leq i < j \leq N} \Phi(|q_i - q_j|) \geq -NB \quad \forall N \in \mathbb{N}, q_i \in \mathbb{R}^d$$

As an important application of these conditions, we show the existence of the thermodynamic limit for the free energy in the canonical ensemble. We have already seen that (for every kind of interaction) the momenta are independent of the positions with a Maxwell distribution, and else there is nothing interesting to say about the momenta, so we will ignore them from now on and consider a Hamiltonian of the form

$$H(q) = \sum_{i<j} \Phi(|q_i - q_j|).$$

In the case of the free gas we now get

$$Z = \int e^{-\beta H} d\lambda_\Omega = \frac{1}{N!} V^N \quad \text{and} \quad A = \frac{1}{\beta}\Big(\log N! - N \log V \Big) \sim \frac{N}{\beta} \log \frac{N}{eV}.$$

In the thermodynamic limit $\frac{N}{V} \to \rho$ and thus $\frac{A}{V} \to a_{free}(\rho, \beta) := \frac{1}{\beta}\rho(\log \rho - 1)$. In the presence of interaction it usually is not possible to calculate Z and thus A explicitly. However, the free energy density exists and has some nice properties.

Theorem 3.2 *(Fisher, Ruelle, 1963) We consider the canonical ensemble for interacting particles in $\mathbb{R}^d$, ignoring the momenta. If the interaction Φ is tempered and stable, then the free energy density $a(\rho,\beta)$ exists in the thermodynamic limit. As a function of ρ, $a(.,\beta)$ is convex and continuous and $a(\rho,\beta) \sim \frac{\rho\log\rho}{\beta}$ for $\rho \to 0$.*

Proof: (For cubic domains.) Let $\rho,\beta > 0$. For simplicity we consider $B := \log Z = -\beta A$ instead of A. We first consider a particular sequence of cubes with side lengths $l_k = 2^k$ containing $N_k = 2^{dk}\rho$ particles. (For non-integer particle numbers we interpolate corresponding functions linearly.) We define Λ_k to be the cube with side length $l_k - R$ and let $V_k := \lambda^d(\Lambda_k)$ and Z_k the corresponding partition sum. We note

$$\frac{N_k}{V_k} = \frac{N_k}{(l_k-R)^d} = \frac{2^{dk}\rho}{(2^k-R)^d} \to \rho \quad \text{for } k \to \infty.$$

Λ_{k+1} can be decomposed into 2^d copies $\Lambda_k(i)$ of Λ_k that keep distance R from each other. In Y_{k+1} we restrict our attention to the set of positional configurations $\Omega_{k+1} \subset \Lambda_{k+1}^{N_{k+1}}$ with exactly N_k particles in each $\Lambda_k(i)$. There are $\frac{N_{k+1}!}{N_k!^{2^d}}$ ways to choose particles for these regions. Suppose now that particles $q(i) := (q_j)_{(i-1)N_k<j\le iN_k}$ are in $\Lambda_k(i)$. By temperedness $\Phi(q) \le \sum_i \Phi(q(i))$ and we get

$$\begin{aligned} Z_{k+1} &= \frac{1}{N_{k+1}!}\int_{\Lambda_{k+1}^{N_{k+1}}} e^{-\beta\Phi(q)}dq \ge \frac{1}{N_{k+1}!}\int_{\Omega_{k+1}} e^{-\beta\Phi(q)}dq \\ &\ge \frac{1}{N_{k+1}!}\frac{N_{k+1}!}{N_k!^{2^d}}\prod_i \int_{\Lambda_k(i)^{N_k}} e^{-\beta\Phi(q(i))}dq(i) = \Big(\frac{1}{N_k!}\int_{\Lambda_k^{N_k}} e^{-\beta\Phi(q)}dq\Big)^{2^d} = Z_k^{2^d} \end{aligned}$$

and thus $b_k(\rho) := \frac{1}{2^{dk}}\log Z_k$ is increasing in k. By stability

$$Z_k = \frac{1}{N_k!}\int_{\Lambda_k^{N_k}} e^{-\beta\Phi(q)}dq \le \frac{1}{N_k!}\int_{\Lambda_k^{N_k}} e^{\beta BN_k}dq \le \frac{1}{N_k!}V_k^{N_k}e^{\beta BN_k}$$

and thus

$$\frac{b_k(\rho)}{\rho} \le \beta B + \log V_k - \frac{1}{N_k}\log(N_k!) \le \beta B - \log\rho + 1,$$

where we have used $V_k \le \frac{N_k}{\rho}$ and Stirling's formula in the form $\log n - \frac{1}{n}\log n! \le 1$. So b_k is increasing and bounded from above and thus convergent:

$$b_k(\rho) \uparrow b(\rho).$$

We next establish some properties of b. The upper bound from above immediately carries over:

$$b(\rho) \le -\rho\log\rho + \rho(\beta B + 1).$$

We also have a lower bound for $\rho \to 0$. For small particle density we decompose Λ_k into m_k^d cubes of side length l_k with $(l_k+R)m_k = 2^k$ so that they keep distance $\ge R$. Choosing $m_k = \lceil \rho^{1/d}2^k \rceil$ we have $m_k^d \ge N_k$, so we can put one particle in the first N_k of these little cubes and no particles in the remaining cubes, which gives $Z_k \ge (l_k^d)^{N_k}$ as above, and thus

$$b(\rho) \ge b_k(\rho) \ge \frac{N_k d}{2^{dk}}\log l_k = \rho d\log(\frac{2^k}{m_k} - R) \to \rho d\log(\rho^{-1/d} - R) \quad \text{for } k \to \infty.$$

Combining the two bounds we get in particular $b(\rho) \sim -\rho \log \rho$ for $\rho \to \infty$. By a homework problem b is midpoint concave. Midpoint concavity and boundedness on an interval $(0, \epsilon]$ immediately give that b is continuous and convex. Because b_k and b are continuous and $b_k \uparrow b$, it finally follows that the convergence $b_k \to b$ is uniform on compact intervals. See the lemmas below for details.

Now let $\Lambda_n, N_n, V_n, l_n, b_n$ denote the quantities for a given sequence of arbitrary cubes and $\Lambda'_k, N'_k, V'_k, l'_k, b'_k$ those for the special sequence considered above. For fixed $n \ll k$ let $m_{n,k}$ be the maximal number of copies of Λ_n that fit into Λ'_k in the above manner, i.e. $m_{n,k} := \left\lfloor \frac{l'_k}{l_n+R} \right\rfloor^d$, and fill each one with N_n particles. The particle density in Λ'_k is

$$\rho_{n,k} = \frac{m_{n,k} N_n}{V'_k} \to \frac{N_n}{(l_n + R)^d} \quad \text{for } k \to \infty.$$

As above we have $Y'_k(\rho_{n,k}) \geq Y_n(N_n, \Lambda_n)^{m_{n,k}}$ and thus

$$b'_k(\rho_{n,k}) \geq \frac{m_{n,k}}{V'_k} \log Y_n(N_n, \Lambda_n) = \frac{m_{n,k} V_n}{V'_k} b_n$$

Letting $k \to \infty$ and using the uniform convergence from above we obtain

$$b'\Big(\frac{N_n}{(l_n + R)^d}\Big) \geq \frac{V_n}{(l_n + R)^d} b_n \quad \text{and thus} \quad b'(\rho) \geq \limsup_n b_n,$$

where we have used $\frac{N_n}{V_n} \to \rho$ and the continuity of b'. For the converse inequality we fix $n \gg k$ and let $M_{n,k}$ be the maximal number of copies of Λ'_k that fit into Λ_n in the above manner, i.e. $M_{n,k} := \lfloor \frac{l_n}{l'_k+R} \rfloor^d$, and fill each one with $N'_{n,k} := \frac{N_n}{M_{n,k}}$ particles. The particle density in each Λ'_k is then

$$\rho_{n,k} = \frac{N_n}{M_{n,k} V'_k} \to \frac{(l'_k + R)^d}{V'_k} \rho =: \rho'_k \quad \text{for } n \to \infty.$$

As above we have $Y_n(N_n, \Lambda_n) \geq Y'_k(\rho_{n,k})^{M_{n,k}}$ and thus

$$b_n \geq \frac{M_{n,k}}{V_n} \log Y'_k(\rho_{n,k}) = \frac{M_{n,k} V'_k}{V_n} b'_k(\rho_{n,k}),$$

so

$$\liminf_n b_n \geq \frac{V'_k}{(l'_k + R)^d} b'_k(\rho'_k) \quad \text{and thus} \quad \liminf_n b_n \geq b'(\rho)$$

by continuity and the uniform convergence of the b'_k. □

Lemma 3.1 *Let $f : (0, \infty) \to \mathbb{R}$ be midpoint concave, i.e. $f(\frac{x_1+x_2}{2}) \geq \frac{1}{2}(f(x_1)+f(x_2))$ for all $x_1, x_2 > 0$, and bounded on $(0, \epsilon]$, where $\epsilon > 0$. Then f is bounded on any bounded interval, continuous and concave.*

Proof: Let $c > 0$ such that $|f(x)| \leq c$ for all $0 < x \leq \epsilon$. Let $\epsilon < x \leq \frac{3}{2}\epsilon$. We have

$$3c \geq 2f(\epsilon) - f(2\epsilon - x) \geq f(x) \geq \frac{1}{2}(f(2\epsilon) + f(2x - 2\epsilon)) \geq \frac{1}{2}(f(2\epsilon) - c)$$

by midpoint concavity. Thus f is bounded on $(0, \frac{3}{2}\epsilon)$. By induction f is bounded on $(0, (\frac{3}{2})^k\epsilon)$ for every k, i.e. on every bounded interval. By midpoint concavity and induction on n

$$f\Big(\frac{k_1}{2^n}x_1 + \frac{k_2}{2^n}x_2\Big) \geq \frac{k_1}{2^n}f(x_1) + \frac{k_2}{2^n}f(x_2) \quad \forall n \geq 0\, \forall k_1, k_2 \geq 0 : k_1 + k_2 = 2^n.$$

Let $x, x', x'' > 0$ such that $f(x'') \geq f(x')$. f is bounded near x by some c. For fixed n choose x_1, x_2 such that $x' = \frac{1}{2^n}x_1 + (1 - \frac{1}{2^n})x_2$ and $x'' = x_2$. If x' and x'' are sufficiently close to x then x_1 is also close to x. The above implies that

$$f(x') \geq \frac{1}{2^n}f(x_1) + (1 - \frac{1}{2^n})f(x''), \qquad \text{i.e. } f(x'') - f(x') \leq \frac{1}{2^n}2c.$$

Thus f is continuous in x. Approximating real numbers by numbers of the form $\frac{k}{2^n}$, the above inequality implies that f is concave. □

Lemma 3.2 *Let $K \subset \mathbb{R}$ be compact and $f_n, f : K \to \mathbb{R}$ be continuous such that $f_n \uparrow f$. Show that $f_n \to f$ uniformly.*

Proof: Suppose otherwise, then there are $\epsilon > 0$, $x_n \in K$ such that $f_n(x_n) < f(x_n) - \epsilon$. By compactness we have a convergent subsequence, and by ignoring the other terms we may assume that $x_n \to x$. By monotonicity $f_k(x_n) \leq f_n(x_n) < f(x_n) - \epsilon$ for all $k \leq n$. By continuity of f_k, f this implies $f_k(x) \leq f(x) - \epsilon$ and we get a contradiction letting $k \to \infty$. □

3.6 Boundary conditions

For many systems it can be observed that although most combinations of parameter values describe a unique state of the system, there may be some combinations of parameter values that do not determine the state of the system uniquely. A typical example for this behavior is water, which (at a certain temperature) can both be in a liquid state and in a solid (frozen) state. This phenomenon is called phase transition and the possible states of the system are called phases. So far we have ignored this phenomenon: The ensemble distribution corresponding to given parameter values describes only one phase. How do we get the other phases? Often it is sufficient to enlarge the ensembles by considering boundary configurations. The idea is that there are particles outside of the volume in some configuration typical for the phase. If these are interacting with the particles inside the volume, this may force these particles also into configurations typical for this phase. The idea of boundary configurations can be made rigorous for all ensembles considered so far, but for the sake of definiteness we will just present it for the grand canonical ensemble and for particles in $\mathbb{R}^d$ ignoring the momentum. In the previous sections we have seen that corresponding configuration space is the collection of all finite sets of points in a domain Λ. To be able to give a more formal description of this space we need some notations and definitions. For $A \in \mathcal{B}^d$ let

$$\mathcal{B}_A^{<\infty} := \{E \in \mathcal{B}^d : E \subset A, \lambda^d(E) < \infty\}$$

denote the set of all subsets of A of finite volume. For $E \subset \mathbb{R}^d$ let

$$N_E : \mathcal{B}^d \to \{0, 1, 2, ...\} \cup \infty, \quad N_E(A) := \#(E \cap A),$$

the number of points a given set has in E. The N_E are called counting random variables. Let

$$\Omega_A = \{\omega \subset A : N_E(\omega) < \infty \; \forall E \in \mathcal{B}_A^{<\infty}\}.$$

$\omega \in \Omega_A$ is called locally finite: ω may consist of infinitely many points, but in any bounded set there are only finitely many points, i.e. ω does not have accumulation points. A natural σ-algebra on Ω_A is given by

$$\mathcal{F}_A = \sigma(\{N_E = k\} : k \in \mathbb{N}, E \in \mathcal{B}_A^{<\infty}).$$

Now we can set up grand canonical distributions with boundary configurations. Let $\Lambda \in \mathcal{B}_{\mathbb{R}^d}^{<\infty}$ be a domain. Ω_Λ is the collection of all finite (unordered) sets of points in Λ. This is equivalent to the description of the state space in the preceding sections: Since Ω_Λ consists of sets rather than ordered tuples of points, the identification $\sim$ is already built into the sate space. The set of boundary configurations is given by Ω_{Λ^c}. We note that a configuration $\omega \in \Omega_\Lambda$ is always finite, whereas a boundary configuration $\eta \in \Omega_{\Lambda^c}$ may be infinite. The way the boundary condition enters the definition of the distribution is via the Hamiltonian. Instead of considering H we now consider

$$H_\Lambda(\omega|\eta) := \frac{1}{2} \sum_{q \neq q' \in \omega} \Phi(|q - q'|) + \sum_{q \in \omega, q' \in \eta} \Phi(|q - q'|) \quad \text{for } \omega \in \Omega_\Lambda, \eta \in \Omega_{\Lambda^c}.$$

Thus we consider all interactions terms of particles in ω and particles in ω with the boundary configuration η. (The factor $\frac{1}{2}$ gets rid of double counting.) We note that the second sum may be infinite, so we have to impose additional conditions on ϕ (such as temperedness) to make sure $H_\Lambda(.|\eta)$ is well defined. Using this conditional Hamiltonian instead of the usual one, we define grand canonical distributions with boundary conditions:

Definition 3.9 *Let $\Lambda \in \mathcal{B}_{\mathbb{R}^d}^{<\infty}$, $\beta, z > 0$ and $\eta \in \Omega_{\Lambda^c}$. The grand canonical distribution $\gamma_\Lambda^{\beta,z}(.|\eta)$ is defined to be the probability measure on $(\Omega_\Lambda, \mathcal{F}_\Lambda)$ with density*

$$\rho_\Lambda^{\beta,z}(.|\eta) := \frac{1}{Z_\Lambda^{\beta,z}(\eta)} e^{-\beta H_\Lambda(.|\eta)} z^N \quad w.r.t. \quad \lambda_{\Omega_\Lambda} = \sum_{n \geq 0} \frac{1}{n!} \lambda_{\Lambda^n},$$

where the partition function $Z_\Lambda^{\beta,z}(\eta)$ is the appropriate normalizing constant. For $\eta \in \Omega$ we define $\gamma_\Lambda^{\beta,z}(.|\eta) := \gamma_\Lambda^{\beta,z}(.|\eta_{\Lambda^c}) \otimes \delta_{\eta_{\Lambda^c}}$ to get a probability measure on $(\Omega, \mathcal{F})$ that looks like η outside of Λ. Here $\eta_B := \eta \cap B$ is the corresponding restriction.

Here we have used the so called activity $z = e^{\beta\mu}$ as a parameter instead of the chemical potential μ. All other definitions carry over. In particular, in the thermodynamic limit the boundary condition is also kept fixed.

Example: For the free gas there is no interaction between particles. Thus the boundary condition does not have any influence on the grand canonical distribution.

The type of boundary condition described above is sometimes called configurational boundary condition. A special case is the one of a free boundary condition which corresponds to $\eta = \emptyset$. (This amounts to the same as not considering boundary conditions at all.) Often also periodic boundary conditions are considered in the case that Λ has a rectangular shape. Here Λ is considered a torus by identifying opposite faces of Λ or equivalently $\mathbb{R}^d$ is decomposed into translates of Λ and η is defined to be the union of all translates of ω_Λ.

3.7 Infinite volume distributions

All ensembles considered so far give distributions for fixed volume, and only in the thermodynamic limit this volume is taken to infinity. Since we are only interested in what happens in the thermodynamic limit anyway, it would be much more convenient to have a model, where the distributions are defined for infinite volume right away. There is a general procedure to do this for all ensembles considered above, but for the sake of definiteness we will only treat the case of the grand canonical ensemble as set up in the previous section. Thus our state space is

$$\Omega := \Omega_{\mathbb{R}^d} = \{\omega \subset \mathbb{R}^d : N_E(\omega) < \infty \ \forall E \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}\}.$$

We now could define a reference measure on Ω similar to the Poisson measure and try to use the maximum entropy principle. The problem is that in general $N(\omega) = \infty$ and $H(\omega)$ is not well defined since it is an infinite sum with infinitely many positive and negative terms. However, we already have a distribution $\gamma^{\beta,z}_\Lambda(.|\eta)$ on Ω_Λ, which can be used to define a distribution μ on Ω implicitly: We demand that the samples obtained from μ that look like η outside of Λ are distributed according to $\gamma^{\beta,z}_\Lambda(.|\eta)$ inside Λ. This is a disintegration property according to the decomposition $\Omega = \Omega_{\Lambda^c} \times \Omega_\Lambda$ of the configuration space that can be written in the form

$$\mu = \mu_{\Lambda^c} \otimes \gamma^{\beta,z}_\Lambda, \quad \text{i.e. } \mu(d\omega, d\eta) = \mu_{\Lambda^c}(d\eta)\gamma^{\beta,z}_\Lambda(d\omega|\eta) = \mu(d\eta)\gamma^{\beta,z}_\Lambda(d\omega|\eta_{\Lambda^c})$$

where μ_{Λ^c} is the corresponding marginal of μ and $\gamma^{\beta,z}_\Lambda$ is considered a probability kernel from Ω_{Λ^c} to Ω_Λ. Since $\gamma^{\beta,z}_\Lambda$ can also be considered a probability kernel from Ω to Ω (when defined with boundary condition $\eta \in \Omega$), we can also write this in the form

$$\mu = \mu \otimes \gamma^{\beta,z}_\Lambda, \quad \text{i.e. } \mu(d\omega) = \mu(d\eta)\gamma^{\beta,z}_\Lambda(d\omega|\eta).$$

Here we have used that the last measure only produces samples such that $\omega_{\Lambda^c} = \eta_{\Lambda^c}$.

Definition 3.10 *Let $\beta, z > 0$. A distribution μ on $(\Omega, \mathcal{F})$ is called a Gibbs measure with parameters β, z iff $\mu = \mu \otimes \gamma^{\beta,z}_\Lambda$ for every $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$, i.e. iff*

$$\mu(D) = \int \mu(d\eta)\gamma^{\beta,z}_\Lambda(D|\eta) \quad \forall \Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}, D \in \mathcal{F}. \tag{3.1}$$

Let $\mathcal{G}^{\beta,z}$ denote the set of all such Gibbs measures. If $\#(\mathcal{G}^{\beta,z}) > 1$, we say that there is a (first-order) phase transition for the parameter values β, z.

We note that (3.1) holds for $\gamma^{\beta,z}_{\Lambda'}(.|\eta')$ instead of μ, where $\Lambda' \supset \Lambda$:

Lemma 3.3 *(Consistency property.) For $\beta, z > 0$, we have $\gamma^{\beta,z}_{\Lambda'} = \gamma^{\beta,z}_{\Lambda'} \otimes \gamma^{\beta,z}_{\Lambda}$ for all $\Lambda \subset \Lambda' \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$, i.e.*

$$\gamma^{\beta,z}_{\Lambda'}(D|\eta') = \int \gamma^{\beta,z}_{\Lambda'}(d\eta|\eta')\gamma^{\beta,z}_{\Lambda}(D|\eta) \quad \forall \Lambda \subset \Lambda' \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}, D \in \mathcal{F}, \eta' \in \Omega.$$

Proof: Homework. □

We note that instead of testing (3.1) for all D, Λ it is sufficient to consider special D, Λ:

Lemma 3.4 *If (3.1) holds for all D from a $\cap$-stable generator of $\mathcal{F}$ and for a specific sequence $\Lambda_n \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$ such that $\Lambda_n \uparrow \mathbb{R}^d$, then it holds for all $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}, D \in \mathcal{F}$.*

Proof: Both sides of (3.1) define probability measures in D. If they agree on a $\cap$-stable generator of $\mathcal{F}$, then they agree on $\mathcal{F}$ by the uniqueness theorem. For the second part let $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$ and choose an n such that $\Lambda_n \supset \Lambda$. If $\mu = \mu \otimes \gamma^{\beta,z}_{\Lambda_n}$, then

$$\mu = \mu \otimes \gamma^{\beta,z}_{\Lambda_n} = \mu \otimes (\gamma^{\beta,z}_{\Lambda_n} \otimes \gamma^{\beta,z}_{\Lambda}) = (\mu \otimes \gamma^{\beta,z}_{\Lambda_n}) \otimes \gamma^{\beta,z}_{\Lambda} = \mu \otimes \gamma^{\beta,z}_{\Lambda}$$

by the above lemma. □

Remark:

- (3.1) is called DLR-condition in honor of Dobrushin, Lanford, Ruelle.
- The definition of $\mathcal{G}^{\beta,z}$ is very much implicit. Usually it is very difficult to get an explicit description of $\mathcal{G}^{\beta,z}$ or to determine whether there is a phase transition.
- A useful $\cap$-stable generator of $\mathcal{F}$ conists of all sets of the form $D \times \Omega_{\Lambda^c}$, where $D \in \mathcal{F}_\Lambda$ and $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$ such that $\Lambda \supset \Lambda_0 \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$.
- The set of Gibbs measures is convex:

  $$\forall \mu_i \in \mathcal{G}^{\beta,z}, \alpha_i \geq 0 \text{ with } \sum_i \alpha_i = 1 : \quad \sum_i \alpha_i \mu_i \in \mathcal{G}^{\beta,z}.$$

  Let $Ex(\mathcal{G}^{\beta,z})$ denote the set of extremal elements of the convex set $\mathcal{G}^{\beta,z}$. These extremal Gibbs measures are also called pure states or pure phases and correspond to what is called phases in physics. All other Gibbs measures are statistical mixtures of these: The Theorem of Krein-Milman guarantees that every Gibbs measure can be written as a convex combination of elements of $Ex(\mathcal{G}^{\beta,z})$.

Example: Free gas. Let $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$, $\beta, z > 0$, $\eta \in \Omega_{\Lambda^c}$, $\Phi = 0$. Here $\gamma^{\beta,z}_{\Lambda}(.|\eta)$ in fact does not depend on β and η: γ^z_Λ is defined to have density $\rho^z_\Lambda = \frac{1}{Z^z_\Lambda} z^N$ w.r.t. Lebesgue-Poisson measure λ_{Ω_Λ}. We have

$$Z^z_\Lambda = \int z^N d\lambda_{\Omega_\Lambda} = \sum_n \frac{1}{n!} \int_{\Lambda^n} z^n dx_1 ... dx_n = \sum_n \frac{1}{n!} z^n \lambda(\Lambda)^n = e^{z\lambda(\Lambda)},$$

and thus the density is simply

$$\rho_\Lambda^z = e^{-z\lambda(\Lambda)} z^N.$$

To extend γ_Λ^z to infinite sets Λ let C_i be the decomposition of $\mathbb{R}^d$ into unit cubes. We set

$$\nu_{A,z} := \otimes_i \gamma^z_{A\cap C_i} \quad \text{for arbitrary } A \in \mathcal{B}^d.$$

Below we will show that $\mathcal{G}^{\beta,z} = \{\nu_{\mathbb{R}^d,z}\}$ after presenting some important properties of the distributions $\nu_{A,z}$.

Lemma 3.5 *If $A = \bigcup_i A_i$ is a decomposition of A, then $\nu_{A,z} = \otimes_i \nu_{A_i,z}$, and for $A \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$ we have $\nu_{A,z} = \gamma^z_A$.*

Proof: Homework. □

Theorem 3.3 *Let $\beta, z > 0$. For the free gas we have $\mathcal{G}^{\beta,z} = \{\nu_{\mathbb{R}^d,z}\}$.*

Proof: First we note that for any $\Lambda \in \mathcal{B}^{<\infty}_{\mathbb{R}^d}$ we have $\nu_{\mathbb{R}^d,z} = \nu_{\Lambda^c,z} \otimes \nu_{\Lambda,z} = (\nu_{\mathbb{R}^d,z})_{\Lambda^c} \otimes \gamma^z_\Lambda$ by the above lemma, i.e. $\nu_{\mathbb{R}^d,z} \in \mathcal{G}^{\beta,z}$. On the other hand let $\mu \in \mathcal{G}^{\beta,z}$. For $D \in \mathcal{F}_\Lambda$ with finite Λ we have

$$\mu(D\times\Omega_{\Lambda^c}) = \mu_{\Lambda^c}\otimes\gamma^z_\Lambda(D\times\Omega_{\Lambda^c}) = \mu_{\Lambda^c}(\Omega_{\Lambda^c})\gamma^z_\Lambda(D) = \gamma^z_\Lambda(D) = \nu_{\Lambda,z}(D) = \nu_{\mathbb{R}^d,z}(D\times\Omega_{\Lambda^c}).$$

By the uniqueness theorem this implies that $\mu = \nu_{\mathbb{R}^d,z}$. □

In Mathematics $\nu_{A,z}$ is known as the Poisson-measure. It plays in important role for many applications.

Lemma 3.6 *Let $z > 0$ and $A \in \mathcal{B}^d$. $\nu_{A,z}$ is the Poisson-measure (or distribution of the Poisson process) on A with intensity z, i.e. $\nu_{A,z}$ is the unique probability measure on $(\Omega_A, \mathcal{F}_A)$ such that*

- *For every $E \in \mathcal{B}^{<\infty}_A$ N_E is a Poisson random variable with parameter $z\lambda^d(E)$, i.e.*

$$\nu_{A,z}(N_E = k) = e^{-z\lambda^d(E)} \frac{(z\lambda^d(E))^k}{k!}.$$

- *For disjoint $E_i \in \mathcal{B}^{<\infty}_A$, the random variables N_{E_i} are independent.*

Proof: Homework. □

For calculations, using the two defining properties of the Poisson measure usually is easier than working with the explicit construction in terms of the grand canonical distributions.

4 Ising model

4.1 Definition of the Ising model

In this section we introduce and investigate the Ising model, a toy model for ferromagnetism that was introduced by Wilhelm Lenz in 1920 and investigated by his student Ernst Ising. It is a prototype model for statistical mechanics. A lot of things can be computed explicitly, and many of the techniques used there carry over to other models. In fact in dimension one and two the free energy can be computed using algebraic methods ("Onsager solution"). Here we will use probabilistic and geometric ideas instead.

For the Ising model we consider particles with fixed positions, where the positions form a regular lattice, say $\mathbb{Z}^d$. We will also assume that only particles at adjacent positions will interact, so we consider the set of edges (or bonds)

$$E = E(\mathbb{Z}^d) = \{\{i,j\} : i,j \in \mathbb{Z}^d, |i-j| = 1\}.$$

$(\mathbb{Z}^d, E(\mathbb{Z}^d))$ forms a graph, and we will denote edges by $ij := \{i,j\}$, and write $i \sim j$ if $ij \in E(\mathbb{Z}^d)$. For a subset $\Lambda \subset \mathbb{Z}^d$ we also consider the corresponding edge set $E_\Lambda = \{ij \in E : i,j \in \Lambda\}$. A domain will be any finite subset of $\mathbb{Z}^d$, and we will set $\mathcal{Z} = \mathcal{Z}^d := \{\Lambda \subset \mathbb{Z}^d : \Lambda \text{ finite}\}$. Sometimes for a given domain $\Lambda \subset \mathbb{Z}^d$ it is convenient to consider its boundary and closure

$$\partial\Lambda := \{i \in \Lambda^c : \exists j \in \Lambda : i \sim j\} \quad \text{and} \quad \bar{\Lambda} = \Lambda \cup \partial\Lambda.$$

Each particle has an internal property, a so called (magnetic) spin that takes the values $\{-1,1\}$ (or $\{-,+\}$ or $\{\downarrow,\uparrow\}$). A suitable configuration space is $\Omega = \{-1,1\}^{\mathbb{Z}^d}$ with product σ-algebra $\mathcal{F}$. In a configuration $\sigma = (\sigma_i)_{i\in\mathbb{Z}^d} \in \Omega$, σ_i denotes the spin of the particle with position $i \in \mathbb{Z}^d$. For $\Lambda \subset \mathbb{Z}^d$ let $\Omega_\Lambda := \{-1,1\}^\Lambda$ denote the spin configuration of particles in Λ. For $\sigma \in \Omega$ and $\Lambda \subset \mathbb{Z}^d$ let $\sigma_\Lambda = (\sigma_i)_{i\in\Lambda} \in \Omega_\Lambda$ denote the restriction to Λ, and for $\sigma \in \Omega_\Lambda$ and $\sigma' \in \Omega_{\Lambda'}$ where $\Lambda \cap \Lambda' = \emptyset$ let $\sigma\sigma' \in \Omega_{\Lambda\cup\Lambda'}$ denote the concatenation of σ and σ'. Let $S_i : \Omega \to \{-1,1\}$ denote the projection on the i-th component, i.e. S_i is supposed to describe the spin of the particle at position $i \in \mathbb{Z}^d$, and let $S_\Lambda = (S_i)_{i\in\Lambda}$ denote the projection on $\Lambda \subset \mathbb{Z}^d$. This notation is useful for describing events, e.g. the event that the particles in Λ have given spins $\sigma \in \Omega_\Lambda$ is

$$\{\omega \in \Omega : \omega_\Lambda = \sigma\} = \{S_\Lambda = \sigma\}.$$

A single particle can be described by $(i,\sigma_i) \in \mathbb{Z}^d \times \{-1,1\}$, so the two-particle interaction can be given in terms of a function $\Phi : (\mathbb{Z}^d \times \{-1,1\})^2 \to \bar{\mathbb{R}}$. For the Ising model we set

$$\Phi((i,\sigma_i),(j,\sigma_j)) = -\sigma_i\sigma_j 1_{\{i\sim j\}}.$$

Thus only adjacent particles interact, and the interaction only depends on whether their spins are aligned or not. Aligned spins are preferred in terms of lower energy. The Hamiltonian H is the sum of all interaction terms, i.e.

$$H(\sigma) = \frac{1}{2}\sum_{i,j\in\Lambda} \Phi((i,\sigma_i),(j,\sigma_j)) = \sum_{ij\in E_\Lambda} \Phi((i,\sigma_i),(j,\sigma_j)) \qquad \text{for } \sigma \in \Omega_\Lambda.$$

In order to define a (grand) canonical model we fix a finite domain $\Lambda \in \mathcal{Z}^d$, the inverse temperature $\beta > 0$ and a boundary configuration $\eta \in \Omega_{\Lambda^c}$. The grand canonical distribution γ_Λ^β is defined to be the probability measure on $(\Omega_\Lambda, \mathcal{F}_\Lambda)$ with density

$$\rho_\Lambda^\beta = \frac{1}{Z_\Lambda^\beta(\eta)} e^{-\beta H(.|\eta)} \quad \text{w.r.t.} \quad \chi_{\Omega_\Lambda}.$$

Here χ_{Ω_Λ} is the counting measure on Ω_Λ and for $\sigma \in \Omega_\Lambda$ we set

$$H(\sigma|\eta) = \sum_{ij \in E_\Lambda} \Phi((i,\sigma_i),(j,\sigma_j)) + \sum_{i\in\Lambda, j\in\Lambda^c} \Phi((i,\sigma_i),(j,\eta_j)) = \sum_{ij\in E_{\bar{\Lambda}}} \Phi((i,(\sigma\eta)_i),(j,(\sigma\eta)_j)).$$

Since this sum is finite, H is well defined. Also $Z_\Lambda^\beta(\eta) = \sum_{\sigma\in\Omega_\Lambda} e^{-\beta H(\sigma|\eta)}$ is finite and positive, which implies in particular that γ_Λ^β is well defined. For $\eta \in \Omega$ we consider $\gamma_\Lambda^\beta(.|\eta)$ as a measure on $(\Omega, \mathcal{F})$, as before. We are mainly interested in the set of infinite volume Gibbs measures $\mathcal{G}^\beta$ for our model. Here $\mu \in \mathcal{G}^\beta$ iff

$$\mu(D) = \int \mu(d\eta)\gamma_\Lambda^\beta(D|\eta) \quad \forall \Lambda \in \mathcal{Z}^d, D \in \mathcal{F}.$$

Our main goal for the next sections is to investigate whether this system exhibits a phase transition.

We note that the Ising model is a very special case of a more general class of models, so called lattice models. Some possible generalizations are the following:

- Consider a lattice different from $\mathbb{Z}^d$.
- Consider a more general pair interaction Φ.
- Consider more general spin values.
- Consider additional one-particle-interaction terms (e.g. modelling an external magnetic field h) so that $H = \sum_{ij} \Phi((i,\sigma_i),(j,\sigma_j)) + \sum_i h(i,\sigma_i)$.

Some models that have been investigated are the following:

- Generalized Ising model: $E = \{-1,1\}$, $\Phi((i,\sigma_i),(j,\sigma_j)) = -J_{ij}\sigma_i\sigma_j$ for some $J_{ij} \in \mathbb{R}$. Here any pair of particles may interact and the coupling constants J_{ij} give the value of the interaction. If $J_{ij} \geq 0$ the interaction is called ferromagnetic (equal spins are preferred), and if $J_{ij} \leq 0$ the interaction is called anti-ferromagnetic (different spins are preferred).
- n-vector or $O(n)$-model: $E = S^{n-1} \subset \mathbb{R}^n$, $\Phi((i,\sigma_i),(j,\sigma_j)) = -\sigma_i \cdot \sigma_j 1_{\{i\sim j\}}$. For special choices of n we get the Ising model ($n=1$), the XY-model ($n=2$) and the Heisenberg model ($n=3$).
- Potts model: $E = \{1,...,q\}$, $\Phi((i,\sigma_i),(j,\sigma_j)) = -1_{\{\sigma_i=\sigma_j, i\sim j\}}$.
- Gaussian model: $E = \mathbb{R}$, $\Phi((i,\sigma_i),(j,\sigma_j)) = (\sigma_i - \sigma_j)^2 1_{\{i\sim j\}}$.

4.2 Symmetries

For investigating whether a certain model has a phase transition, it is often useful to look at the symmetries of the model. We will define symmetries in the context of the Ising model. Basically a symmetry is supposed to be any transformation of particles that leaves the energy and the reference measure invariant.

Definition 4.1 *A transformation $\tau : \mathbb{Z}^d \times \{-1,1\} \to \mathbb{Z}^d \times \{-1,1\}$ is a symmetry if*

- *τ is bijective and the transformation of the position does not depend on the spin, i.e. there is a bijective map $\tau_* : \mathbb{Z}^d \to \mathbb{Z}^d$ such that $\tau(i, \pm 1) = (\tau_*(i), \pm 1)$.*
- *Φ is τ-invariant, i.e. $\Phi(\tau(i,\sigma_i), \tau(j,\sigma_j)) = \Phi((i,\sigma_i),(j,\sigma_j))\ \forall (i,\sigma_i),(j,\sigma_j)$.*

By abuse of notation we identify τ and τ_*. By transforming every single particle, we can apply τ to a configuration of particles, and thus get $\tau : \Omega_\Lambda \to \Omega_{\tau\Lambda}$ or $\tau : \Omega \to \Omega$. We note that because of the bijectivity, the reference measure is automatically invariant under a symmetry: $\chi_{\Omega_\Lambda} \circ \tau^{-1} = \chi_{\Omega_{\tau\Lambda}}$. Since τ doesn't change energy and reference measure, it is easy to see how τ affects the grand canonical distributions and the Gibbs measures.

Theorem 4.1 *If τ is a symmetry, then $\gamma^\beta_\Lambda(.|\eta) \circ \tau^{-1} = \gamma^\beta_{\tau\Lambda}(.|\tau\eta)$ for every $\Lambda \in \mathcal{Z}$, $\beta > 0$, $\eta \in \Omega_{\Lambda^c}$, and $\mu \circ \tau^{-1} \in \mathcal{G}^\beta$ for every $\mu \in \mathcal{G}^\beta$.*

Proof: Let $\Lambda \in \mathcal{Z}$, $\beta > 0$ and $\eta \in \Omega_{\Lambda^c}$. For all $\omega \in \Omega_\Lambda$ we have $H_{\tau\Lambda}(\tau\omega|\tau\eta) = H_\Lambda(\omega|\eta)$ since the Hamiltonian is a sum of interaction terms that are invariant under τ. Thus for every $D \in \mathcal{F}_\Lambda$

$$\begin{aligned} Z^\beta_\Lambda(\eta)\gamma^\beta_\Lambda(\tau^{-1}D|\eta) &= \sum_{\omega \in \tau^{-1}D} e^{-\beta H_\Lambda(\omega|\eta)} = \sum_{\omega' \in D} e^{-\beta H_\Lambda(\tau^{-1}\omega'|\eta)} \\ &= \sum_{\omega' \in D} e^{-\beta H_{\tau\Lambda}(\omega'|\tau\eta)} = Z^\beta_{\tau\Lambda}(\tau\eta)\gamma^\beta_{\tau\Lambda}(D|\tau\eta). \end{aligned}$$

For $D = \Omega_{\tau\Lambda}$ this implies $Z^\beta_\Lambda(\eta) = Z^\beta_{\tau\Lambda}(\tau\eta)$, and thus $\gamma^\beta_\Lambda(\tau^{-1}D|\eta) = \gamma^\beta_{\tau\Lambda}(D|\tau\eta)$. Now let $\mu \in \mathcal{G}^\beta$ and $D \in \mathcal{F}$, then the measure transformation theorem gives

$$\begin{aligned} \int \mu \circ \tau^{-1}(d\eta)\gamma^\beta_\Lambda(D|\eta) &= \int \mu(d\eta')\gamma^\beta_\Lambda(D|\tau\eta') \\ &= \int \mu(d\eta')\gamma^\beta_{\tau^{-1}\Lambda}(\tau^{-1}D|\eta') = \mu(\tau^{-1}D) = \mu \circ \tau^{-1}(D), \end{aligned}$$

since μ is a Gibbs measure. We thus have shown that $\mu \circ \tau^{-1}$ satisfies the DLR-condition and so $\mu \circ \tau^{-1} \in \mathcal{G}^{\beta,z}$. □

It may be somewhat surprising that a symmetry τ doesn't necessarily leave each Gibbs measure invariant. The reason for this is that τ changes the boundary configuration, which may have an influence on spins near the origin even if Λ is huge.

Definition 4.2 *A Gibbs measure $\mu \in \mathcal{G}^{\beta,z}$ is said to break a symmetry τ, if $\mu \neq \mu \circ \tau^{-1}$.*

The above theorem immediately implies that whenever a symmetry is broken, there is a phase transition, since we have $\mu \neq \mu \circ \tau^{-1} \in \mathcal{G}^{\beta,z}$ and thus $\#(\mathcal{G}^{\beta,z}) \geq 2$. Since we are interested in a possible phase transition of the ising model, let us look at its symmetries. Here a lattice automorphism is a bijective transformation $\tau : \mathbb{Z}^d \to \mathbb{Z}^d$ such that $i \sim j \Leftrightarrow \tau(i) \sim \tau(j)$. Examples of lattice automorphisms are translations by lattice vectors, reflections in the coordinate axes, and rotations by multiples of $\pi/2$. The spin-flip transformation is defined by

$$\tau : \{-1, 1\} \to \{-1, 1\}, \quad \tau(\sigma) := -\sigma.$$

Lemma 4.1 *For the Ising model every lattice automorphism and the spin-flip are a symmetries.*

Proof: A lattice automorphism τ defines a 1-particle transformation by $\tau(i, \sigma_i) := (\tau(i), \sigma_i)$, which is bijective and doesn't change the given Φ and thus is a symmetry. The spin-flip transformation τ defines a 1-particle transformation by $\tau(i, \sigma_i) := (i, \tau(\sigma_i))$, which is bijective and doesn't change the given Φ and thus is a symmetry. □

What kind of phase transition can be expected for a ferromagnet? For low temperature a ferromagnet shows spontaneous magnetization, i.e. most of the spins are aligned. For high temperature this phenomenon does not occur. If this behavior is reflected in our model, for low temperature there should be a Gibbs measure μ where the majority of spins are $+1$. In this case the spin-flip symmetry is broken, since then $\mu \circ \tau^{-1}$ has a majority of -1 spins and thus $\mu \circ \tau^{-1} \neq \mu$.

4.3 Ground states

Another concept related to phase transitions is that of ground states. Suppose that $\eta \in \Omega$ and $\omega, \omega' \in \Omega$ with $\omega_{\Lambda^c} = \omega'_{\Lambda^c} = \eta_{\Lambda^c}$ such that $H_\Lambda(\omega_\Lambda|\eta_{\Lambda^c}) < H_\Lambda(\omega'_\Lambda|\eta_{\Lambda^c})$, then

$$\frac{\gamma^\beta_\Lambda(\{\omega'\}|\eta)}{\gamma^\beta_\Lambda(\{\omega\}|\eta)} = e^{-\beta(H_\Lambda(\omega'_\Lambda|\eta_{\Lambda^c}) - H_\Lambda(\omega_\Lambda|\eta_{\Lambda^c}))} \to 0 \quad \text{for } \beta \to \infty.$$

So for "$\beta = \infty$" the grand canonical distribution $\gamma^\beta_\Lambda(.|\eta)$ can be thought of being concentrated on the configurations $\omega \in \Omega$ such that and $H_\Lambda(\omega_\Lambda|\eta_{\Lambda^c})$ is minimal. In the infinite volume limit we thus get measures concentrated on configurations $\omega \in \Omega$ such that for every Λ the Hamiltonian is minimal under changes of ω inside of Λ. These configurations are so called ground states.

Definition 4.3 *A configuration $\omega \in \Omega$ is called a ground state if $H_\Lambda(\omega_\Lambda|\omega_{\Lambda^c}) \leq H_\Lambda(\omega'|\omega_{\Lambda^c})$ for all $\Lambda \in \mathcal{Z}$ and $\omega' \in \Omega_\Lambda$. A ground state is called isolated if the above inequality is strict whenever $\omega' \neq \omega_\Lambda$.*

We note that the above energy minimality condition is an analogue for the DLR-condition in the case "$\beta = \infty$". Thus the ground states may be considered the equilibrium states for "$\beta = \infty$". In the Ising model we have at least two isolated ground states corresponding to the expected breaking of the spin-flip symmetry.

Lemma 4.2 *The configurations $\sigma^+, \sigma^- \in \Omega$ defined by $\sigma_i^+ := 1, \sigma_i^- := -1 \quad \forall i \in \mathbb{Z}^d$ are isolated ground states.*

Proof: Let $\Lambda \in \mathcal{Z}$ and let n be the number of bonds in $E_{\bar{\Lambda}}$. We have

$$H_\Lambda(\sigma_\Lambda^+ | \sigma_{\Lambda^c}^+) = \sum_{ij \in E_{\bar{\Lambda}}} \Phi((i,+1),(j,+1)) = -n,$$

whereas for any $\omega' \neq \sigma_\Lambda^+$ there is at least one bond where the spins are not aligned, thus at least one of the values of Φ changes from -1 to $+1$, i.e. $H_\Lambda(\omega' | \sigma_{\Lambda^c}^+) \geq -n+2$. So σ^+ is an isolated ground state, and by a similar argument so is σ^-. □

Obviously σ^+ and σ^- are related to each other by a spin flip, and it is not hard to see that in general for every ground state σ and every symmetry $\tau \quad \tau(\sigma)$ is again a ground states. Since $\sigma^+ \neq \sigma^- = \tau(\sigma^+)$ we say that the spin flip symmetry is broken for these ground states. It may be somewhat surprising that σ^+ and σ^- are not the only ground states. σ^+ and σ^- are certainly the only global minima of the energy, however ground states are defined in terms of local minima (where spins may only be changed in a bounded set Λ). An example for an additional (non-isolated) ground state in one dimension is $\sigma \in \Omega$ defined by $\sigma_i := 1$ if $i \geq 0$ and $\sigma_i := -1$ if $i > 0$.

Our motivation to consider ground states was that ground states correspond to Gibbs measures at "$\beta = \infty$", and it is tempting to assume that for large but finite β there should be Gibbs measures that look like small perturbations of corresponding ground states. Arguing along these lines the breaking of a symmetry at "$\beta = \infty$" should imply the breaking of this symmetry at large but finite β, and multiple ground states should imply a phase transition for large but finite β. However, we will see that this is not necessarily the case. Still, looking at the ground states at least gives some candidates for possible symmetries to be broken and it gives some ideas on how the corresponding Gibbs measures should look like. In the example of the Ising model, for sufficiently large β we thus might hope to find a Gibbs measure μ^+ that is a small perturbation of σ^+, i.e. that is concentrated on configurations that look like a huge sea of $+$ with some small islands of $-$.

4.4 Existence and uniqueness of Gibbs measures

In this section we will see that the Ising model always has at least one infinite volume Gibbs measure, and for large temperature this Gibbs measure is unique. The proof that there is a Gibbs measure uses several facts from functional analysis and topology.

Let $k_i \mathbb{N}, i \in \mathbb{Z}^d$, be an enumeration of $\mathbb{Z}^d$. On $\Omega := \{-1,1\}^{\mathbb{Z}^d}$ we consider the product topology defined by the metric $d(\omega,\omega') = \sum_i 2^{-k_i} 1_{\{\omega_i \neq \omega_i'\}}$. Ω is compact by Tychonoff's theorem. We consider the Banach space $C(\Omega)$ of continuous functions on Ω with the supremum norm $\|.\|$.

Definition 4.4 *$f : \Omega \to \mathbb{R}$ is called local if it only depends on the spins in Λ, i.e.*

$$f(\sigma\eta) = f(\sigma\eta') \quad \forall \sigma \in \Omega_\Lambda, \eta, \eta' \in \Omega_{\Lambda^c},$$

where Λ is finite. f is called quasilocal if there are local $f_k : \Omega \to \mathbb{R}$ with $\|f_k - f\| \to 0$.

Lemma 4.3 *For $f : \Omega \to \mathbb{R}$ we have the equivalence*

$$f \in C(\Omega) \;\Leftrightarrow\; \forall \epsilon > 0 \exists \text{ finite } \Lambda : \sup_{\sigma \in \Omega_\Lambda, \eta, \eta' \in \Omega_{\Lambda^c}} |f(\sigma\eta) - f(\sigma\eta')| < \epsilon \;\Leftrightarrow\; f \text{ quasilocal.}$$

Proof: Homework. □

The dual space $C(\Omega)^*$ is the collection of continuous linear maps form $C(\Omega)$ to $\mathbb{R}$. On $C(\Omega)^*$ we consider the operator norm $\|.\|$ and the weak-*-topology. By definition $T_n \to T$ in the weak-*-topology iff $T_n(f) \to T(f)$ for all $f \in C(\Omega)$. Let $M(\Omega)$ denote the set of all probability measures on Ω. For every $\mu \in M(\Omega)$ we define $T_\mu \in C(\Omega)^*$ by $T_\mu(f) = \mathbb{E}_\mu(f)$, and we will identify $T_\mu = \mu$. With this identification, $\mu_n \to \mu$ w.r.t. the weak-*-topology is the same as $\mu_n \to \mu$ in distribution.

Lemma 4.4 *$M(\Omega)$ is compact w.r.t. the weak-*-topology.*

Proof: The Banach-Alaoglu Theorem states that the closed unit ball B^* in $C(\Omega)^*$ w.r.t. $\|.\|$ is compact w.r.t. the weak-*-topology. With the above identification we have $M(\Omega) \subset B^*$, so it suffices to show that $M(\Omega)$ is a closed subset of B^*. Let $\mu_n \in M(\Omega)$ such that $\mu_n \to T \in B^*$ w.r.t. the weak-*-topology. For every $f \in C(\Omega)$ such that $f \geq 0$ we have $T(f) \geq 0$ (since $\mu_n(f) = \mathbb{E}_{\mu_n}(f) \geq 0$) and we have $T(1) = 1$ (since $\mu_n(1) = \mathbb{E}_{\mu_n}(1) = 1$). The Riesz-Markov Theorem implies that $T = T_\mu$ for some $\mu \in M(\Omega)$. □

Theorem 4.2 *(Existence Theorem: Dobrushin '68.) Let $\beta > 0$, $\Lambda_n \uparrow \mathbb{R}^d$ and $\eta_n \in \Omega$. Let $\gamma_n = \gamma^\beta_{\Lambda_n}(.|\eta_n)$. Then there is a Gibbs measure $\mu \in \mathcal{G}^\beta$ such that a subsequence γ_{n_k} converges to μ in distribution. In particular we have $\mathcal{G}^\beta \neq \emptyset$.*

Proof: By the preceding lemma we have a convergent subsequence with some limit $\mu \in M(\Omega)$, and it suffices to check that μ satisfies the DLR-condition for finite Λ and $D = \{S_{\Lambda'} = \sigma\}$, where $\sigma \in \Omega_{\Lambda'}$ and Λ' is finite. If n is sufficiently large we have

$$\int \gamma_n(d\omega) 1_D(\omega) = \int \gamma_n(d\omega) \gamma^\beta_\Lambda(D|\omega),$$

using the consistency condition. Both 1_D and $\gamma^\beta_\Lambda(D|.)$ are local functions ($f(\omega) = \gamma^\beta_\Lambda(D|\omega)$ only depends on $\omega_{(\Lambda' - \Lambda) \cup \partial\Lambda}$) and thus continuous. Convergence in distribution implies the convergence of the integrals on both sides, and so

$$\int \mu(d\eta) 1_D(\eta) = \int \mu(d\eta) \gamma^\beta_\Lambda(D|\eta).$$

Thus we have verified the DLR-condition for μ. □

Theorem 4.3 *(Uniqueness Theorem: Dobrushin '68.)* $\#(\mathcal{G}^\beta) \leq 1$ *for* $0 < \beta < \frac{1}{2d}$.

Proof: Here we give the outline of the proof. All details that require further explanation are relegated to the lemmas below. For $k \in \mathbb{Z}^d$ we set

$$\Phi_k(f)(\eta) = \int \gamma^\beta_{\{k\}}(d\omega|\eta) f(\omega) = \gamma^\beta_k(+|\eta) f(\eta^+_k) + \gamma^\beta_k(-|\eta) f(\eta^-_k),$$

where $\eta^\pm_k$ is η with the spin at k replaced by ± 1, and $\gamma^\beta_k(\pm|\eta) = \gamma^\beta_{\{k\}}(\{\pm 1\}|\eta)$. The map $\Phi_k : C(\Omega) \to C(\Omega)$ is linear with operator norm $\|\Phi_k\| = 1$. Next we define

$$T : C(\Omega) \to C(\Omega), \quad T := \lim_{n\to\infty} T_n, \qquad \text{where } T_n = \Phi_1 \circ ... \circ \Phi_n$$

for some given enumeration of $\mathbb{Z}^d$, where the limit is defined pointwise. Every $\mu \in \mathcal{G}^\beta$ satisfies $\mu(f) = \mu(T(f))$ for all $f \in C(\Omega)$, and we have

$$\inf f \leq \inf T(f) \leq \sup T(f) \leq \sup f \quad \forall f \in C(\Omega).$$

We set

$$\Delta(f) := \sum_k \Delta_k(f), \quad \text{where } \Delta_k(f) = \sup_\omega |f(\omega^+_k) - f(\omega^-_k)|.$$

$\Delta_k(f)$ describes the maximal oscillation of f when changing the spin at k. We have

$$\sup f - \inf f \leq \Delta(f) \quad \forall f \in C(\Omega) \qquad \text{and}$$

$$\Delta(T(f)) \leq (2d\beta)\Delta(f) \quad \forall f \in C_\Delta(\Omega) := \{f \in C(\Omega) : \Delta(f) < \infty\}.$$

Combining these estimates we get

$$\sup T^n(f) - \inf T^n(f) \leq \Delta(T^n(f)) \leq (2d\beta)^n \Delta(f) \to 0,$$

and thus $T^n(f)$ converges for $n \to \infty$ to some constant $c(f) \in \mathbb{R}$. Thus we get

$$\mu(f) = \mu(T^n(f)) \to c(f), \text{ i.e. } \mu(f) = c(f) \quad \forall \mu \in \mathcal{G}^\beta.$$

This gives the desired uniqueness of the Gibbs measure. □

Lemma 4.5 $\Phi_k(f) \in C(\Omega)$ *for all* $f \in C(\Omega)$, *and* Φ_k *is linear with* $\|\Phi_k\| = 1$.

Proof: Since $\eta \mapsto \eta^\pm_k$ is continuous and $\eta \mapsto \gamma^\beta_k(\pm|\eta)$ is local, we have $\Phi_k(f) \in C(\Omega)$. The linearity of Φ_k is obvious and we have $\|\Phi_k(f)\| \leq \sup_\eta(\gamma^\beta_k(+|\eta)\|f\| + \gamma^\beta_k(-|\eta)\|f\|) = \|f\|$ for all $f \in C(\Omega)$ with equality for constant functions. □

Lemma 4.6 *For every* $f \in C(\Omega)$ $T_n(f)$ *converges in supremum norm.*

Proof: First assume that f is local, i.e. f only depends on the spins from a finite set Λ. For all $k \notin \Lambda$ we have $f(\eta) = f(\eta_k^+) = f(\eta_k^-)$ and thus $\Phi_k(f) = f$. If n is sufficiently large the enumeration $\{1, ..., n\}$ contains all $k \in \Lambda$, so the above implies that $T_m(f) = T_n(f)$ for all $m \geq n$, so the limit exists. If f is continuous and thus quasilocal, we have local functions f_k such that

$$\|T_n(f - f_k)\| \leq \|\Phi_1\|...\|\Phi_n\|\|f - f_k\| = \|f - f_k\| \to 0.$$

This implies for all N

$$\sup_{m,n \geq N} \|T_m(f) - T_n(f)\| \leq \sup_{m,n \geq N} \|T_m(f_k) - T_n(f_k)\| + 2\|f - f_k\|$$

using the linearity of T_n. Letting $N \to \infty$ the first term on the RHS vanishes, and then letting $k \to \infty$ we see that $T_n(f)$ is a Cauchy sequence and thus converges. □

Lemma 4.7 *For every $f \in C(\Omega)$ we have*

$$\inf f \leq \inf T(f) \leq \sup T(f) \leq \sup f \quad \text{and} \quad \mu(f) = \mu(T(f)) \forall \mu \in \mathcal{G}^\beta.$$

Proof: We first show the assertions for Φ_k instead of T. We have $\inf \Phi_k(f) \geq \inf_\eta(\gamma_k^\beta(+|\eta) \inf f + \gamma_k^\beta(+|\eta) \inf f) = \inf f$ and similarly for the supremum. The DLR-condition gives $\mu(\Phi_k(f)) = \mu \otimes \gamma_k^\beta(f) = \mu(f)$. By induction we get the assertions for T_n. Letting $n \to \infty$ we get the assertions for T. □

Lemma 4.8 *For every $f \in C(\Omega)$ we have* $\sup f - \inf f \leq \Delta(f)$.

Proof: By compactness we have $\omega, \omega' \in \Omega$ such that

$$\sup f - \inf f = f(\omega) - f(\omega') \leq |f(\omega) - f(\omega_\Lambda \omega'_{\Lambda^c})| + |f(\omega_\Lambda \omega'_{\Lambda^c}) - f(\omega')|$$

for every finite Λ. Flipping the spins in Λ one spin at a time we see that the second term on the RHS is bounded by $\sum_{k \in \Lambda} \Delta_k(f) \leq \Delta(f)$. By continuity of f and Lemma 4.3 the first term vanishes for $\Lambda \uparrow \mathbb{Z}^d$. □

Lemma 4.9 *For $k \neq j \in \mathbb{Z}^d$ and $f \in C_\Delta(\Omega)$ we have*

$$\Delta_k \Phi_k(f) = 0 \quad \text{and} \quad \Delta_j \Phi_k(f) \leq \Delta_j(f) + \beta \Delta_k(f) 1_{\{k \sim j\}}.$$

Proof: The first assertion is obvious, since $\Phi_k(f)$ does not depend on the spin at k. For the second assertion, in case of $\eta_j = -1$, $\eta_k = -1$ we get

$$\begin{aligned}
&|\Phi_k(f)(\eta_j^+) - \Phi_k(f)(\eta_j^-)| \\
&= \left|\gamma_k^\beta(+|\eta_j^+) f(\eta_{kj}^{++}) + \gamma_k^\beta(-|\eta_j^+) f(\eta_{kj}^{-+}) - \gamma_k^\beta(+|\eta_j^-) f(\eta_{kj}^{+-}) - \gamma_k^\beta(-|\eta_j^-) f(\eta_{kj}^{--})\right| \\
&\leq \left|\gamma_k^\beta(+|\eta_j^+) f(\eta) + \gamma_k^\beta(-|\eta_j^+) f(\eta) - \gamma_k^\beta(+|\eta_j^-) f(\eta) - \gamma_k^\beta(-|\eta_j^-) f(\eta)\right| \\
&\quad + \left|\gamma_k^\beta(+|\eta_j^+)\big(f(\eta_{kj}^{++}) - f(\eta_k^+) + f(\eta_k^+) - f(\eta)\big) + \gamma_k^\beta(-|\eta_j^+)\big(f(\eta_j^+) - f(\eta)\big) - \gamma_k^\beta(+|\eta_j^-)\big(f(\eta_k^+) - f(\eta)\big)\right| \\
&\leq |f(\eta) - f(\eta)| + (\gamma_k^\beta(+|\eta_j^+) + \gamma_k^\beta(-|\eta_j^+))\Delta_j(f) + \left|\big(\gamma_k^\beta(+|\eta_j^+) - \gamma_k^\beta(+|\eta_j^-)\big)\big(f(\eta_k^+) - f(\eta)\big)\right| \\
&\leq \Delta_j(f) + \left|\gamma_k^\beta(+|\eta_j^+) - \gamma_k^\beta(+|\eta_j^-)\right| \Delta_k(f),
\end{aligned}$$

and similarly in the other cases. Since $\gamma_k^\beta(+|\eta_j^+) - \gamma_k^\beta(+|\eta_j^-) = \gamma_k^\beta(-|\eta_j^-) - \gamma_k^\beta(-|\eta_j^+)$, we get

$$\Delta_j \Phi_k(f) \le \sup_\eta \left|\gamma_k^\beta(+|\eta_j^+) - \gamma_k^\beta(+|\eta_j^-)\right| \Delta_k(f) + \Delta_j(f).$$

Now it suffices to estimate $g(\eta) := |\gamma_k^\beta(+|\eta_j^+) - \gamma_k^\beta(+|\eta_j^-)|$ from above. We note that $g(\eta) = 0$ unless $k \sim j$, and $\gamma_k^\beta(+|\eta)$ only depends on how many of the spins at the $2d$ sites adjacent to k are $+$. If m of these spins are $+$ we have

$$\gamma_k^\beta(+|\eta) = \frac{e^{\beta(m-(2d-m))}}{e^{\beta(m-(2d-m))} + e^{\beta((2d-m)-m)}} = \frac{1}{1+e^{4\beta(d-m)}} =: h(m),$$

and thus

$$\sup_\eta g(\eta) = \max_{0 \le m \le 2d-1} |h(m+1) - h(m)| \le \max_{x \in [0,2d]} |h'(x)|$$

using the mean value theorem. We have $h'(x) = \frac{4\beta e^{4\beta(d-x)}}{(1+e^{4\beta(d-x)})^2} = 4\beta \frac{y}{(1+y)^2}$, which gets maximal for $y = 1$. Thus $\sup_\eta g(\eta) \le \beta$. □

Lemma 4.10 *Let $2d\beta \le 1$. For all $n \ge 1$ and $f \in C_\Delta(\Omega)$ we have*

$$\Delta(T_n(f)) \le 2d\beta \sum_{j \le n} \Delta_j(f) + \sum_{j > n} \Delta_j(f).$$

Proof: Indcution on n. For $n = 0$ we have equality. For the inductive step we have

$$\begin{aligned}
\Delta(T_{n+1}(f)) &= \Delta(T_n(\Phi_{n+1}f)) \le 2d\beta \sum_{j \le n} \Delta_j(\Phi_{n+1}f) + \sum_{j > n} \Delta_j(\Phi_{n+1}f) \\
&\le 2d\beta \sum_{j \le n} (\Delta_j(f) + \beta \Delta_{n+1}(f) 1_{\{j \sim n+1\}}) + \sum_{j > n+1} (\Delta_j(f) + \beta \Delta_{n+1}(f) 1_{\{j \sim n+1\}}) \\
&\le 2d\beta \sum_{j \le n} \Delta_j(f) + \sum_{j > n+1} \Delta_j(f) + \sum_{j \sim n+1} \beta \Delta_{n+1}(f)
\end{aligned}$$

using the previous lemma. Since $n+1$ has $2d$ neighbors we are done. □

Lemma 4.11 *Let $2d\beta \le 1$. For all $f \in C_\Delta(\Omega)$ we have $\Delta(T(f)) \le 2d\beta \Delta(f)$.*

Proof: We first show that for $g_n \to g$ we have $\Delta(g) \le \lim_n \Delta(g_n)$. We have

$$|g(\omega_k^+) - g(\omega_k^-)| \le |g_n(\omega_k^+) - g_n(\omega_k^-)| + 2\|g_n - g\|$$

and taking the supremum over all ω and summing over k we get

$$\sum_{k=1}^m \Delta_k(g) \le \sum_{k=1}^m \Delta_k(g_n) + 2m\|g_n - g\| \le \Delta(g_n) + 2m\|g_n - g\|.$$

Letting $n \to \infty$ and then $m \to \infty$ we get $\Delta(g) \le \lim_n \Delta(g_n)$. Applying this to $g_n = T_n(f)$ the previous lemma gives $\Delta(T(f)) \le 2d\beta\Delta(f)$. □

4.5 Ising model in one dimension

In one dimension everything can be described explicitly by means of Markov chains.

Lemma 4.12 *Let $\beta > 0$, $\Lambda = \{0, ..., n\}$. Let γ_Λ^β denote the grand canonical distribution (without boundary configuration), then for every $\sigma \in \{-1, 1\}^\Lambda$*

$$\gamma_\Lambda^\beta(\{\sigma\}) = \frac{1}{Z_\Lambda^\beta} e^{-\beta H_\Lambda(\sigma)} = \frac{1}{2} \prod_{i=0}^{n-1} p(\sigma_i, \sigma_{i+1}),$$

where $p(1,1) = p(-1,-1) = c_\beta e^\beta$, $p(1,-1) = p(-1,1) = c_\beta e^{-\beta}$ and $c_\beta = \frac{1}{e^\beta + e^{-\beta}}$.

Proof: By the definition of the interaction

$$e^{-\beta H_\Lambda(\sigma)} = \prod_{i=0}^{n-1} e^{\beta \sigma_i \sigma_{i+1}} = c(n, \beta) \prod_{i=0}^{n-1} p(\sigma_i, \sigma_{i+1}),$$

and by normalization the above equality follows. □

Definition 4.5 *Let $\beta > 0$ and $c_\beta = \frac{1}{e^\beta + e^{-\beta}}$. The Markov chain with state space $\{-1, 1\}$, initial distribution $(\frac{1}{2}, \frac{1}{2})$ and transition matrix $p = \begin{pmatrix} c_\beta e^\beta & c_\beta e^{-\beta} \\ c_\beta e^{-\beta} & c_\beta e^\beta \end{pmatrix}$ is called the Ising chain on $\mathbb{N}$.*

We note that $(\frac{1}{2}, \frac{1}{2})p = (\frac{1}{2}, \frac{1}{2})$, i.e. the initial distribution is stationary. Our aim is to extend the Ising chain $X = (X_n)_{n \in \mathbb{N}}$ to all times $n \in \mathbb{Z}$. In the following we consider special forms of domains $\Lambda_k := \{-k, ..., k\}$.

Definition 4.6 *Let $\beta > 0$. There is a sequence of $\{-1, 1\}$-valued random variables $Y = (Y_n)_{n \in \mathbb{Z}}$ such that for every $k > 0$ and every $\sigma \in \{-1, 1\}^{\Lambda_k}$ we have*

$$\mathbb{P}(Y_{\Lambda_k} = \sigma) = \frac{1}{2} \prod_{i=-k}^{k-1} p(\sigma_i, \sigma_{i+1}).$$

Y is called the stationary Ising chain on $\mathbb{Z}$. Its distribution is uniquely determined by the above condition.

Proof: (Uniqueness and existence.) The distribution of Y is a probability measure on $(\Omega, \mathcal{F})$, where $\Omega = \{-1, 1\}^{\mathbb{Z}}$ and $\mathcal{F}$ is the corresponding product-σ-algebra. The sets of the type $\{S_{\Lambda_k} = \sigma\}$ form a $\cap$-stable generator of $\mathcal{F}$. Since the given condition specifies the probabilities for these sets, the uniqueness theorem gives the uniqueness of such a distribution. For the existence let X and X' be two stationary Ising chains on $\mathbb{N}$ such that $X_0 = X'_0$ and the transitions of X are independent of the transitions of X'. We set $Y_n := X_n$ and $Y_{-n} = X'_n$ for $n \geq 0$.

$$\mathbb{P}(Y_{\Lambda_k} = \sigma) = \mathbb{P}(Y_0 = \sigma_0, X_n = \sigma_n \forall n \leq k, X'_m = \sigma_{-m} \forall m \leq k)$$

$$= \frac{1}{2} \mathbb{P}_{\sigma_0}(X_n = \sigma_n \forall n \leq k) \mathbb{P}_{\sigma_0}(X'_m = \sigma_{-m} \forall m \leq k) = \frac{1}{2} \prod_{i=1}^{k} p(\sigma_{i-1}, \sigma_i) \prod_{i=1}^{k} p(\sigma_{-i+1}, \sigma_{-i}).$$

Since $p(\sigma_{-i+1}, \sigma_{-i}) = p(\sigma_{-i}, \sigma_{-i+1})$ this gives the desired property. □

By the above lemma, the grand canonical distribution in finite volume (without b.c.) is equal to the distribution of the Ising chain restricted to this volume. In the following we will see that this carries over in a very natural way to finite volume equilibrium distributions (with b.c.) and infinite volume equilibrium distributions.

Lemma 4.13 *Let $k < m$, $\Lambda_k^m := \Lambda_m - \Lambda_k$, $\eta \in \Omega_{\Lambda_k^m}$ and let $\gamma^\beta_{\Lambda_k}(.|\eta)$ be the corresponding grand canonical distribution. Let Y be the Ising chain on $\mathbb{Z}$. Then $\gamma^\beta_{\Lambda_k}(.|\eta)$ is the distribution of $(Y_n)_{n\in\Lambda_k}$ given that $Y_n = \eta_n$ for all $n \in \Lambda_k^m$.*

Proof: By the definition of the conditional distribution we get for all $\sigma \in \{-1,1\}^{\Lambda_k}$

$$\begin{aligned}\mathbb{P}(Y_{\Lambda_k} = \sigma | Y_{\Lambda_k^m} = \eta) &= \frac{\mathbb{P}(Y_{\Lambda_m} = \sigma\eta)}{\mathbb{P}(Y_{\Lambda_k^m} = \eta)} = c\frac{1}{Z^\beta_{\Lambda_m}} e^{-\beta H_{\Lambda_m}(\sigma\eta)} \\ &= c' e^{-\beta(H_{\Lambda_k^m}(\eta) + H_{\Lambda_k}(\sigma|\eta))} = c'' e^{-\beta H_{\Lambda_k}(\sigma|\eta)} = \gamma^\beta_{\Lambda_k}(\sigma|\eta),\end{aligned}$$

where c, c', c'' denote constants that do not depend on σ (but only on η, β, k, m). By normalization c'' equals the partition function, which gives the last equality. □

Theorem 4.4 *For the 1-dimensional Ising model we have $\mathcal{G}^\beta = \{\mu^\beta\}$, where μ^β is the distribution of the Ising chain on $\mathbb{Z}$. In particular there is no phase transition and the spinflip-symmetry is not broken, i.e. there is no spontaneous magnetization.*

Proof: We first show that μ^β is a Gibbs measure. It suffices to show the DLR-condition (3.1) for $\Lambda = \Lambda_n$ and $D_\sigma = \{S_{\Lambda_m} = \sigma\}$, where $\sigma \in \{-1,1\}^{\Lambda_m}$ and $m > n$:

$$\begin{aligned}\mu^\beta \otimes \gamma^\beta_{\Lambda_n}(D_\sigma) &= \int \mu^\beta(d\eta)\gamma^\beta_{\Lambda_n}(D_\sigma|\eta) = \int \mu^\beta(d\eta) 1_{\{\eta_i = \sigma_i \forall i \in \Lambda_n^m\}} \gamma^\beta_{\Lambda_n}(\{\sigma_{\Lambda_n}\}|\eta_{\Lambda_n^c}) \\ &= \mathbb{P}(Y_{\Lambda_n^m} = \sigma_{\Lambda_n^m})\mathbb{P}(Y_{\Lambda_n} = \sigma | Y_{\Lambda_n^m} = \sigma_{\Lambda_n^m}) = \mathbb{P}(Y_{\Lambda_m} = \sigma) = \mu^\beta(D_\sigma).\end{aligned}$$

To show that μ^β is the only Gibbs measure, let $\mu \in \mathcal{G}^\beta$, let $k > 0$ and $\sigma \in \{-1,1\}^{\Lambda_k}$. Let $n \geq k$. The DLR-condition (3.1) implies that

$$\mu(D_\sigma) = \mu \otimes \gamma^\beta_{\Lambda_{n-1}}(D_\sigma) = \int \mu(d\eta) \int \gamma^\beta_{\Lambda_{n-1}}(D_\sigma|\eta_{\Lambda_{n-1}^c}) = \int \mu(d\eta)\mathbb{P}(Y_{\Lambda_k} = \sigma | Y_{\pm n} = \eta_{\pm n}),$$

where we have used the above lemma. We note that the last probability equals

$$\frac{\mathbb{P}(Y_{-n} = \eta_{-n}, Y_{\Lambda_k} = \sigma, Y_n = \eta_n)}{\mathbb{P}(Y_{-n} = \eta_{-n}, Y_n = \eta_n)} = \frac{\frac{1}{2}p^{n-k}(\eta_{-n}, \sigma_{-k}) \prod_{i=-k}^{k-1} p(\sigma_i, \sigma_{i+1}) p^{n-k}(\sigma_k, \eta_n)}{\frac{1}{2}p^{2n}(\eta_{-n}, \eta_n)}.$$

The ergodic theorem for Markov chains implies that $p^n(i,j) \to \frac{1}{2}$ for $n \to \infty$ and arbitrary $i, j \in \{-1,1\}$. Thus the above fraction converges to $\frac{\frac{1}{2}\frac{1}{2}\prod_{i=-k}^{k-1} p(\sigma_i,\sigma_{i+1})\frac{1}{2}}{\frac{1}{2}\frac{1}{2}}$. So

$$\mu(D_\sigma) = \int \mu(d\eta) \frac{1}{2} \prod_{i=-k}^{k-1} p(\sigma_i, \sigma_{i+1}) = \frac{1}{2} \prod_{i=-k}^{k-1} p(\sigma_i, \sigma_{i+1}) = \mathbb{P}(Y_\Lambda = \sigma) = \mu^\beta(D_\sigma),$$

which implies $\mu = \mu^\beta$. □

4.6 Phase transition for the Ising model

So far we don't know whether the Ising model has a phase transition in some dimension. We know that this is not the case in dimension $d = 1$ and for larger d when the temperature is sufficiently high. In this section we will see that for $d \geq 2$ the spinflip symmetry is broken for low temperature, and thus we get a phase transition.

To show this phase transition we consider interfaces between regions with spin + and regions with spin −. A large interface requires a lot of energy and thus should have small probability. However, the absence of long interfaces implies that almost all of the spins are the same, i.e. we have a sea of + with small islands of − (or the other way round). In the following we will make these ideas rigorous using an idea of Robert Peierls that goes back to 1936.

Definition 4.7 *We consider the graph $(\mathbb{Z}^d, E(\mathbb{Z}^d))$. Let $V_1, V_2 \subset \mathbb{Z}^d$ form a disjoint decomposition of $\mathbb{Z}^d$ into two connected components such that V_1 is finite. The finite set of bonds $B = \{x_1x_2 : x_1 \in V_1, x_2 \in V_2\}$ is called a surface, $V_1 =: int(B)$ is called its interior and $V_2 =: ext(B)$ its exterior.*

If every bond b of a surface B is replaced by a $(d-1)$-dimensional cube b^* that is perpendicular to b and has the same midpoint, then the collection of all such surface elements b^* indeed forms a $d-1$-dimensional surface in $\mathbb{R}^d$. (Try to visualize this for $d = 2$ and $d = 3$.)

Definition 4.8 *For a given surface B, let $I_-(B)$ denote the set of all configurations $\omega \in \Omega$ such that $\omega_i = -1$ for all $i \in \partial ext(B)$ and $\omega_i = +1$ for all $i \in \partial int(B)$. If $\omega \in I_-(B)$ we call B a $-$-interface for ω. Similarly we define $I_+(B)$ and a $+$-interface.*

We note that a −-interface is a surface so that spins adjacent to the surface in the interior are − and adjacent to the surface in the exterior are +. The following lemma shows that every −-spin inside a domain with +-boundary configuration is surrounded by a −-interface.

Lemma 4.14 *Let $\omega \in \Omega$ and $a \in \Lambda \subset \mathbb{Z}^d$ with finite Λ. If $\omega_a = -1$ and $\omega_i = 1$ for all $i \in \Lambda^c$, then there is a surface B such that $\omega \in I_-(B)$, $a \in int(B)$ and $\Lambda^c \subset ext(B)$.*

Proof: Let Λ_1' denote the set of all $x \in \mathbb{Z}^d$ such that there is a path $a = x_0, x_1, ...x_n = x$ with $\omega_{x_i} = -1$ for all i. $\mathbb{Z}^d - \Lambda_1'$ can have several connected components. Let Λ_2 be the one containing Λ^c. Let $\Lambda_1 := \mathbb{Z}^d - \Lambda_2$ and $B = \{xy : x \in \Lambda_1, y \in \Lambda_2\}$. B is a surface with interior Λ_1 and exterior Λ_2: By definition Λ_1 and Λ_2 form a disjoint decomposition of $\mathbb{Z}^d$ into two connected sets and $\Lambda_1 \subset \Lambda$ is finite. To see that $\omega \in I_-(B)$, we note that the points of Λ_1 adjacent to points of Λ_2 are all in Λ. Thus for each $xy \in B$ with $x \in \Lambda_1$ and $y \in \Lambda_2$ we have in fact $x \in \Lambda$ and $y \notin \Lambda$. Thus by definition $\omega_x = -1$ and $\omega_y = 1$. □

The following lemma contains the energy estimate, which is the heart of the Peierls argument.

Lemma 4.15 *Let $\beta > 0$, $\eta \in \Omega$ and $\Lambda \subset \mathbb{Z}^d$ be finite, and let B be a surface contained in Λ. We have*

$$\gamma_\Lambda^\beta(I_-(B)|\eta) \leq e^{-2\beta\#(B)}.$$

Proof: We define a bijective one-particle transformation τ that flips the spins in the interior of B:

$$\tau(i,\sigma_i) = (i,\sigma_i) \text{ if } i \in ext(B) \text{ and } = (i,-\sigma_i) \text{ if } i \in int(B).$$

For $\omega \in I_-(B)$ we compare the energies of ω and $\tau(\omega)$: For every bond $b = xy$ such that $xy \subset int(B)$ or $xy \subset ext(B)$ we have $\omega_x\omega_y = \tau\omega_x\tau\omega_y$, since either both spins are flipped or none. For every bond $b = xy \in B$ we have $\omega_x\omega_y = -1$ and $\tau\omega_x\tau\omega_y = 1$ since exactly one of the spins is flipped. This implies

$$H_\Lambda(\omega|\eta) = H_\Lambda(\tau\omega|\eta) + 2\#(B),$$

and thus

$$\gamma_\Lambda^\beta(\omega|\eta) = \frac{1}{Z_\Lambda^\beta(\eta)} e^{-\beta H_\Lambda(\omega|\eta)} = \frac{1}{Z_\Lambda^\beta(\eta)} e^{-\beta H_\Lambda(\tau\omega|\eta)} e^{-2\beta\#(B)} = e^{-2\beta\#(B)} \gamma_\Lambda^\beta(\tau\omega|\eta).$$

Summing over all $\omega \in I_-(B)$ and using the bijectivity of τ we get

$$\gamma_\Lambda^\beta(I_-(B)|\eta) \leq e^{-2\beta\#(B)} \gamma_\Lambda^\beta(\tau I_-(B)|\eta) \leq e^{-2\beta\#(B)}.$$

□

The last ingredient of the proof is an estimate of the number of surfaces.

Lemma 4.16 *The number of surfaces B of size k such that $0 \in int(B)$ is bounded from above by $\frac{k-2}{18} 3^k$.*

Proof: For a given surface B we move along a fixed coordinate axis until we leave $int(B)$ for the first time. The corresponding bond in B will be denoted by b_0. We also have a corresponding bond b_0' in the opposite direction. We note that $d(b_0, b_0') \leq \frac{k-2}{2}$ since the remaining $k-2$ surface elements have to connect b_0^* and $b_0'^*$, covering this distance at least twice. Thus we have at most $\frac{k-2}{2}$ possible choices for b_0. We now build a surface with the given properties by first choosing b_0, and then start building the surface from b_0^*. Let b^* be a surface element of the surface constructed so far that has a boundary element c^*, where an additional surface element can be added. There are at most 3 ways to attach this additional surface element to c^*. (c^* has dimension $d-2$, so there are $2 \cdot 2 = 4$ surface elements that share this boundary element, one of which is the original b^*.) The last surface element falls into place automatically. Thus the number of surfaces B is bounded from above by

$$\frac{k-2}{2} \cdot 3 \cdot \ldots \cdot 3 \cdot 1 = \frac{k-2}{2} 3^{k-2}.$$

□

Theorem 4.5 *(Phase transition for $d \geq 2$.) For $d \geq 2$ and for $\beta \geq \log 2$ there are at least two Gibbs measures $\mu_+^\beta, \mu_-^\beta \in \mathcal{G}^\beta$ such that $\mu_+^\beta(S_i = 1) > \frac{1}{2}$ for every $i \in \mathbb{Z}^d$. Furthermore for finite $\Lambda \in \mathbb{Z}^d$ we have $\mu_+^\beta(S_\Lambda = \sigma_\Lambda^+) \to 1$ for $\beta \to \infty$.*

Proof: Let $\Lambda_n \uparrow \mathbb{Z}^d$ be a sequence of finite domains. By the existence theorem the sequence $\gamma_n := \gamma_{\Lambda_n}^\beta(.|\sigma^+)$ has a subsequential limit μ_+^β. If n is sufficiently large we have $i \in \Lambda_{n-1}$ and if $S_i = -1$ and we have + spins outside of Λ_n there is a $-$-interface B separating i from Λ_n. Let $\mathcal{B}_k$ the set of all such surfaces of length k, then

$$\gamma_n(S_i = -1) \leq \gamma_n\Big(\bigcup_k \bigcup_{B \in \mathcal{B}_k} I_-(B)\Big) \leq \sum_k \sum_{B \in \mathcal{B}_k} \gamma_n(I_-(B)) \leq \sum_{k \geq 2} \frac{k-2}{18} 3^k e^{-2\beta k}.$$

Introducing $\alpha = 3e^{-2\beta} \leq \frac{3}{4}$ and substituting $l = k - 2$ we obtain

$$\gamma_n(S_i = -1) \leq \frac{\alpha^3}{18} \sum_{l \geq 0} l\alpha^{l-1} := c(\alpha) < \frac{1}{18} \frac{\alpha^2}{(1-\alpha)^2} \leq \frac{1}{2}.$$

Here we have used

$$\sum_{l \geq 0} l\alpha^{l-1} = \frac{d}{d\alpha} \sum_{l \geq 0} \alpha^l = \frac{d}{d\alpha} \frac{1}{1-\alpha} = \frac{1}{(1-\alpha)^2}.$$

and the fact that $\frac{\alpha^2}{(1-\alpha)^2} = \frac{1}{(\frac{1}{\alpha}-1)^2}$ is increasing in α and $= 9$ for $\alpha = \frac{3}{4}$. Since $1_{\{S_i=-1\}}$ is local and thus continuous we obtain

$$\mu_+^\beta(S_i = -1) = \lim_{k \to \infty} \gamma_{n_k}(S_i = -1) \leq c(\alpha) < \frac{1}{2}.$$

Similarly we get for finite Λ

$$\mu_+^\beta(S_\Lambda \neq \sigma_\Lambda^+) \leq \sum_{i \in \Lambda} \mu_+^\beta(S_i = -1) \leq \#(\Lambda) c(\alpha) \to 0,$$

since for $\beta \to \infty$ we have $\alpha \to 0$ and thus $c(\alpha) \to 0$. Finally we note that for μ_+^β the spinflip symmetry is broken, since for $\mu_-^\beta := \mu_+^\beta \circ \tau^{-1}$ we have $\mu_-^\beta(S_i = 1) = \mu_+^\beta(S_i = -1)$ $= 1 - \mu_+^\beta(S_i = 1) < \frac{1}{2}$. □

Actually much more is known for the Ising model. In the following we collect some of these facts:

- For $d = 1$ there is a unique Gibbs measure for every β, and we have an explicit description in terms of the stationary Ising chain on $\mathbb{Z}$.
- For $d \geq 2$ there is a critical $\beta_c \in (0, \infty)$ such that
  - For $\beta < \beta_c$ there is a unique Gibbs measure μ^β.
  - for $\beta > \beta_c$ there is more than one Gibbs measure, thus we have a phase transition.

We have $\gamma^{\beta}_{\Lambda_n}(.|\sigma^+) \to \mu^{\beta}_+$ and $\gamma^{\beta}_{\Lambda_n}(.|\sigma^-) \to \mu^{\beta}_-$ (without choosing subsequences) for any sequences $\Lambda_n \uparrow \mathbb{Z}^d$. The limit measures do not depend on the choice of the sequence Λ_n. $\mu^{\beta}_+, \mu^{\beta}_-$ are extremal Gibbs measures and $\mu^{\beta}_+ = \mu^{\beta}_-$ iff $\beta < \beta_c$.

- For $d = 2$ it is known that $\beta_c = \frac{1}{2}\log(1+\sqrt{2})$. For $\beta > \beta_c$
  - $\mu^{\beta}_+, \mu^{\beta}_-$ are the only extremal Gibbs measures, so the set of all Gibbs measures is $\mathcal{G}^{\beta} = \{\alpha_+\mu^{\beta}_+ + \alpha_-\mu^{\beta}_- : \alpha_+, \alpha_- \geq 0 \text{ such that } \alpha_+ + \alpha_- = 1\}$. In particular the spinflip symmetry is the only broken symmetry, so μ^{β}_+ is invariant under translations and rotations.
  - The average magnetization w.r.t. μ^{β}_+ is given by $\mathbb{E}^{\beta}_+(S_i) = (1-(\sinh 2\beta)^{-4})^{\frac{1}{8}}$

  Furthermore for $d = 2$ and $\beta = \beta_c$ the random shape of the interface between $+$ and $-$ regions can be described explicitly. If the mesh size of the lattice converges to 0, the random shape converges to a random process called SLE_3.

- For $d \geq 3$ the translational and rotational symmetry is also broken. For low temperature there are Gibbs measures corresponding to the ground state σ which is positive in one half space and negative in the other half space.