Mathematical Statistical Physics

Imre Péter Tóth

July 22, 2012

Abstract

For the TMP course at LMU, SoSe2012. Constantly changing draft, with many typos, mistakes, errors and inconsistencies.

Contents

1 Probability warming-up **2**
- 1.1 Measure, measure space, probability, probability space 2
- 1.2 Measurable functions, random variables and their distributions 7
- 1.3 Integral and expectation 9
  - 1.3.1 Integral, integrability 9
  - 1.3.2 Exchanging limits with integrals 12
- 1.4 conditioning 15
- 1.5 Entropy 18

2 Equilibrium statistical ensembles in finite systems **19**
- 2.1 Identical particles and factorization of the phase space 21
- 2.2 Microcanonical ensemble 22
- 2.3 Canonical ensemble 24
- 2.4 Grand canonical ensemble 26

3 Thermodynamic limit **28**
- 3.1 Convergence of thermodynamic functions 28
- 3.2 Gibbs measures 30

4 Ising model **31**

5 Basics of ergodic theory **31**

Introduction

These notes are intended to contain as much of the material discussed in the lectures, as possible – with extensions at those points where there was no time to discuss thing I would be sorry to leave out. However, for a long time I will certainly be way behind with writing up the material. To still be of some help, I will publish the highly incomplete versions. So on the short term the main goal is to provide precise formulation of the definitions and theorems – in part to compensate for being vague on the lectures. Explanation about the background, motivation and logical structure will hopefully be added at some later time.

Nothing of the material presented is my own result. In fact, most of it is by now considered "classical" or "standard knowledge" in Mathematics, and I will not attempt to give references to the original papers of the true authors. Instead, I will (when I get to it) cite some textbooks and lecture notes where the material is easily accessible, and which I myself use to reduce the number of false statements written.

1 Probability warming-up

1.1 Measure, measure space, probability, probability space

Definition 1.1 (σ-algebra)**.** *For a nonempty set Ω, a family $\mathcal{F}$ of subsets of ω (i.e. $\mathcal{F} \subset 2^\Omega$, where $2^\Omega := \{A : A \subset \Omega\}$ is the power set of Ω) is called a σ-algebra over Ω if*

- $\emptyset \in \mathcal{F}$
- *if $A \in \mathcal{F}$, then $A^C := \Omega \setminus A \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under complement taking)*
- *if $A_1, A_2, \dots \in \mathcal{F}$, then $(\cup_{i=1}^\infty A_i) \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under countable union).*

Definition 1.2 (measurable space, measurable set)**.** *If Ω is a nonempty set and $\mathcal{F}$ is a σ-algebra over Ω, then the pair $(\Omega, \mathcal{F})$ is called a measurable space. The elements of $\mathcal{F}$ arer called measurable subsets of Ω.*

Lemma 1.3. *A σ-algerba is closed under finite intersection, countable union and finite union.*

Proof. Homework. □

Note that a σ-algebra is in general not closed under arbitrary intersection and union. For example, the Borel σ-algebra on the set $\mathbb{R}$ of real numbers (see later) contains every 1-element subset of $\mathbb{R}$, but it does not contain every subset (a fact we will not prove).

Two trivial examples of σ-algebra:

Definition 1.4 (indiscrete σ-algebra)**.** *For a nonempty set Ω, the family of subsets $\mathcal{F}_{ind} = \{\emptyset, \Omega\}$ is called the indiscrete or trivial σ-algebra over Ω.*

Definition 1.5 (discrete σ-algebra)**.** *For a nonempty set Ω, the family of subsets $\mathcal{F}_{discr} = 2^{\Omega}$ (the entire power set) is called the discrete σ-algebra over Ω.*

It is immediate from the definition that these are indeed σ-algebras over Ω.

Lemma 1.6. *The intersection of any (nonempty) family of σ-algebras over the same Ω is also a σ-algerba over Ω. That is, if Ω is a nonempty set and $\mathcal{F}_i$ is a σ-algerba over Ω for every $i \in I$ where I is* any *nonempty index set, then $\mathcal{F} := \cap_{i\in I}\mathcal{F}_i$ is also a σ-algebra over Ω.*

Proof. (trivial set algebra) By definition, $\emptyset \in \mathcal{F}_i$ for every $i \in I$, and I is nonempty, so $\emptyset \in \cap_{i\in I}\mathcal{F}_i = \mathcal{F}$. Similarly, if $A \in \mathcal{F} = \cap_{i\in I}\mathcal{F}_i$, then $A \in \mathcal{F}_i$ for every $i \in I$, so by definition $\Omega \setminus A \in \mathcal{F}_i$ for every $i \in I$, so $\Omega \setminus A \in \mathcal{F}$. Finally, if $A_1, A_2, \cdots \in \mathcal{F}$, then $A_1, A_2, \cdots \in \mathcal{F}_i$ for every $i \in I$, so by definition $(\cup_{k=1}^{\infty} A_k) \in \mathcal{F}_i$ for every $i \in I$, which means that $(\cup_{k=1}^{\infty} A_k) \in \mathcal{F}$. □

It is important to note that I being *any* nonempty set means in particular that it can well be a large set, having infinitely many, or even uncountably many, or possibly much more elements.

Corollary 1.7. *If Ω is a nonempty set and $H \subset 2^{\Omega}$ is any family of subsets, then there exists a unique σ-algebra $\sigma(H)$ over Ω, which is the smallest σ-algebra containig H in the following sense:*

- *$H \subset \sigma(H)$*
- *If $\mathcal{F}$ is any σ-algebra over Ω with $H \subset \mathcal{F}$, then $\mathcal{F} \subset \sigma(H)$.*

Proof. The family

$$\{\mathcal{F} : \ \mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \text{ and } H \subset \mathcal{F}\}$$

is nonempty, since it contains at least the discrete σ-algebra 2^{Ω}. Thus by the above lemma,

$$\sigma(H) := \cap\{\mathcal{F} : \ \mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \text{ and } H \subset \mathcal{F}\}$$

will do. Uniqueness also follows from the lemma: if there were two diferent such minimal σ-algebras, their intersection would also be a σ-algebra, but it would not contain them – a contradiction. □

Definition 1.8 (σ-algebra generated by a family of sets)**.** *The above $\sigma(H)$ is called the σ-algebra generated by H.*

Definition 1.9 (Borel σ-algebra)**.** *If $(\Omega, \mathcal{T})$ is a topological space (which means that it makes sense to talk about open subsets of Ω, and $\mathcal{T}$ is the set of these open subsets), then $\mathcal{B}(\Omega) := \sigma(\mathcal{T})$ is called the Borel σ-algebra on Ω.*

Remark 1.10. *The set $\mathcal{T}$ is called the* topology*, so the Borel σ-algebra is the σ-algebra generated by the topology. For those, who haven't heard, but are interested: A collection $\mathcal{T} \subset 2^\Omega$ of subsets of Ω is called (by definition) a* topology *over Ω if it contains $\emptyset$ and Ω, and it is closed under finite intersection and* arbitrary *union. Then the elements of $\mathcal{T}$ are called open sets, and the pair $(\Omega, \mathcal{T})$ is called a topological space. (So the definition only says that $\emptyset$ and Ω are open, the intersection of finitely many open set is open, and that the union of* any *family of open sets is open.)*

When we talk about the Borel sets on $\mathbb{R}$ or $\mathbb{R}^n$, we always think of the usual notion of open sets on these spaces.

Remark 1.11. *Not every subset of $[0, 1]$ is Borel. In fact, a non-Borel subset can be constructed (and not only the existence can be proven). We don't go into that.*

Notation 1.12. *We denote by $\mathbb{R}^+$ the set of nonnegative real numbers – that is, $\mathbb{R}^+ = [0, \infty)$. In particular, $\mathbb{R}^+$ includes zero.*

Definition 1.13 (measure space, measure)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space. The nonnegative extended real valued function μ on $\mathcal{F}$ (that is, $\mu : \mathcal{F} \to \mathbb{R}^+ \cup \{\infty\}$) is called a measure on Ω, if*

- *$\mu(\emptyset) = 0$,*
- *μ is σ-additive, meaning that if $\{A_i\}_{i \in I}$ is a countable family of pairwise disjoint measureable sets (with formulas: $A_i \in \mathcal{F}$ for every $i \in I$ where I is a* countable *index set, and $A_i \cap A_j = \emptyset$ for every $i \neq j$, $i, j \in I$), then*

$$\mu\left(\cup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i).$$

Then the triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space and $\mu(A)$ is called the measure of the set A.

Remark 1.14. *It is absolutely important that in the definition of σ-additivity I is countably infinite: σ-additivity is more than finite additivity and less than arbitrary additivity. For example, if μ is the Lebesgue measure on $\mathbb{R}$ (see later), every 1-element set $\{x\}$ has $\mu(\{x\}) = 0$, which implies that every countable set has zero measure, but of course*

$$1 = \mu([0,1]) = \mu(\cap_{x\in[0,1]}\{x\}) \neq \sum_{x\in[0,1]} \mu(\{0\}) = \sum_{x\in[0,1]} 0 = 0.$$

(Whatever the sum of uncountably many real numbers could mean.)

Once it came up, we mention the following, *absolutely non-important* definition:

Definition 1.15 (sum of many nonnegative extended real numbers)**.** *If $a_i \in \mathbb{R}^+ \cup \{\infty\}$ for every $i \in I$ where I is* any *index set, then we define the sum of all a_i as*

$$\sum_{i\in I} a_i := \sup\left\{ \sum_{i\in J} a_i \ : \ \ J \subset I \text{ and } J \text{ is finite} \right\}.$$

Note that it is important that the a_i are nonnegative. This definition obviously coincides with the usual sum of (an arbitrarily reorderable) series if I is countable. This new notion of an infinite sum is no serious extention of the well known notion of a countable series: it is easy to see that if the sum is finite, then at most countably many terms can be nonzero.

Remark 1.16. *In the definition of the measure, the first requirement $\mu(\emptyset) = 0$ is almost automatic from σ-additivity: it's only there to rule out the trivial nonsense $\mu(\emptyset) = \infty$. In fact it would be enough to require that at least one measurable set A has finite measure: σ-additivity implies*

$$\mu(A) = \mu(A\cup\emptyset\cup\emptyset\cup\emptyset\cup\dots) = \mu(A)+\mu(\emptyset)+\mu(\emptyset)+\mu(\emptyset)+\dots = \mu(A)+\sum_{i=1}^{\infty}\mu(\emptyset).$$

If $\mu(A) < \infty$, then this implies $0 = \sum_{i=1}^{\infty}\mu(\emptyset)$, so $\mu(\emptyset) = 0$.

Definition 1.17. *The measure χ on a nonempty set Ω equipped with the discrete σ-algebra 2^Ω defined as*

$$\chi(A) := \sharp A := \begin{cases} \text{number of elements in } A, & \text{if } A \text{ is finite} \\ \infty, & \text{if } A \text{ is infinite} \end{cases}$$

is called the counting measure on Ω. The restriction of χ to any σ-algebra $\mathcal{F} \subset 2^\Omega$ is still called a counting measure (on $\mathcal{F}$).

One of the most important examples of a measure is the Lebesgue measure on $\mathbb{R}$ or on $\mathbb{R}^d$

Definition 1.18 (Lebesgue measure vaguely)**.** *Consider the set $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$. The measure* Leb $: \mathcal{B} \to \mathbb{R}^+ \cup \{\infty\}$ *is called the Lebesgue measure on $\mathbb{R}$, if it assigns to every interval its length – that is, for every $a, b \in \mathbb{R}$, $a \leq b$ we have*

$$\text{Leb}((a,b)) = \text{Leb}((a,b]) = \text{Leb}([a,b)) = \text{Leb}([a,b]) = b - a.$$

The restriction of Leb *to a Borel subset of $\mathbb{R}$ (e.g. an interval $[c,d]$ or (c,∞)) is still called Lebesgue measure and is still denoted by* Leb*. (More precisely, if $(\mathbb{R}, \mathcal{B}, \text{Leb})$ is the Lebesgue measure space on $\mathbb{R}$, and $I \in \mathcal{B}$, than one can define the "restriction of* Leb *to I" as the measure space $(I, \mathcal{B}_I, \text{Leb}_I)$ where $\mathcal{B}_I := \{A \cap I \,:\, A \in \mathcal{B}\} = \{B \,:\, B \in \mathcal{B}, B \subset I\} \subset \mathcal{B}$ and $\text{Leb}_I := \text{Leb}|_{\mathcal{B}_I}$ is the restriction of* Leb *to $\mathcal{B}_I$.)*

Similarly, the "Lebesgue measure on $\mathbb{R}^d$" is the measure on Borel subsets of $\mathbb{R}^d$ which assigns to every box its d-dimensional volume, i.e. for every $a_1 \leq b_1, a_2 \leq b_2, \ldots a_d \leq b_d \in \mathbb{R}$ we have

$$\text{Leb}_d([a_1,b_1] \times [a_2,b_2] \times \cdots \times [a_d,b_d]) = \prod_{i=1}^{d} (b_i - a_i).$$

Restrictions to Borel subsets of $\mathbb{R}^d$ are still called Lebesgue measure, and denoted by Leb_d *or just* Leb*.*

Remark 1.19. *The above "definition" of the Lebesgue measure is far from being complete, and is not the usual definition – it's actually a characterization of the Lebesgue measure which shows its essence. It can be (and needs to be) shown that such a measure indeed exists, since in the "definition" we only gave the value of* Leb *for a few very special sets, and not every Borel set. Also uniqueness can and needs to be shown. These questions lead to the construction of measures based on their pre-known values on certain pre-chosen "to-be-measurable" sets, which can sometimes be of crucial importance, but we don't go into that here.*

Remark 1.20. *In the measure theory literature, Lebesgue measure is defined on a σ-algerba $\mathcal{F}$ which is* larger *than the Borel σ-algebra (i.e. $\mathcal{B} \subsetneq \mathcal{F}$), called the "set of Lebesgue measurable sets". In particular, $\mathcal{F}$ has the property that if $B \in \mathcal{F}$,* Leb$(B) = 0$ *and $A \subset B$, then $A \in \mathcal{F}$, which is not true for Borel sets. However, in probability theory it is usual to consider Lebesgue measure restricted to Borel sets only (as in the above definition).*

The following definition shows that the basic object of probability theory, called "the probability" is in fact a measure.

Definition 1.21 (Kolmogorov probability space)**.** *The triple* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called a Kolmogorov probability space (or just probability space) if it is a measure space and* $\mathbb{P}(\Omega) = 1$*. Then* $\mathbb{P}$ *is called the* probability *or a "probability measure", elements of* $\mathcal{F}$ *are called* events*, and elements of* Ω *are the* elementary events*. For* $A \in \mathcal{F}$*,* $\mathbb{P}(A) \in [0, 1]$ *is called the probability of the event* A*.*

Picture $\omega \in \Omega$ as possible outcomes of an experiment, so an event $A \in \mathcal{F}$ often consists of many possible outcomes of that experiment, which have some common property that we are interested in. By definition, an "event" is something which has a probability.

1.2 Measurable functions, random variables and their distributions

Notation 1.22. *For a function* $f : \Omega \to \Omega'$ *and a set* $A' \subset \Omega'$*, let* $f^{-1}(A')$ *denote, as usual, the complete inverse image of* A' *defined as* $f^{-1}(A') := \{\omega \in \Omega : f(\omega) \in A'\}$*. Note that this makes sense for any function and any* A' *– in particular,* f *need not be invertable.*

Definition 1.23 (measureable function, observable, random variable)**.** *Let* $(\Omega, \mathcal{F})$ *and* $(\Omega', \mathcal{F}')$ *be measurable spaces. The function* $X : \Omega \to \Omega'$ *is called* measurable *or* $(\mathcal{F}, \mathcal{F}')$*-measurable, if for every* $A' \in \mathcal{F}'$ *we have* $X^{-1}(A') \in \mathcal{F}$*. (That is, if the inverse image of any measurable set is also measurable.) In physical applications, when* Ω *is the (possibly complicated) phase space of a system and* Ω' *is a (usually simple) set of possible measurement results (e.g.* $\Omega' = \mathbb{R}$*), the same* X *is called an* observable*. In the context of probability theory, when* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space,* X *is called a (*Ω'*-valued)* random variable*.*

Note that the notion of measurability of a function depends on the choice of the σ-algebras $\mathcal{F}$ and $\mathcal{F}'$. However, in many cases when this choice is clear from the context, we simply say "measurable" instead of "$(\mathcal{F}, \mathcal{F}')$-measurable". When we talk about a random variable, and do not specify the range, usually $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B})$ is understood.

Remark 1.24. *If* $X : \Omega \to \Omega'$ *is a random variable and* $x \in \Omega'$*, then we denote by* $\{X = x\}$ *the set of elementary events where* X *takes the value* x *– that is,*

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} = X^{-1}(\{x\}).$$

Similarly, if $A' \subset \Omega'$, then $\{X \in A'\}$ denotes the set of elementary events where X takes values in A':

$$\{X \in A'\} := \{\omega \in \Omega \,:\, X(\omega) \in A'\} = X^{-1}(A).$$

With this in mind, the definition of a random variable as a measurable *function is very natural. The definition says exactly that is A' is a measurable subset of the range Ω', then the set $\{X \in A\}$ is indeed an* event *– that is, it makes sense to talk about its probability.*

Example 1.25. *coordinate, number rolled, sum of these*

Random variables are the central objects of study in probability theory. In a typical situation they extract fairly little information (e.g. a single number) from a big probability space containing many complicated possible outcomes of an experiment. So to "understand" a random variable $X : \Omega \to \Omega'$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ well, we need less information than the what $\mathbb{P}$ and X contain. From another point of view: when we consider a random variable, Ω is often not needed, or not even known. All we need to know is the possible values (in Ω') X can take, and the probability of these being taken. This information is contained exactly in a measure on Ω', as the following definition shows.

Definition 1.26 (distribution of a random variable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\Omega', \mathcal{F}')$ a measurable space and $X : \Omega \to \Omega'$ a random variable. Then the distribution of X is the measure μ on $(\Omega', \mathcal{F}')$ which is defined as*

$$\mu(A) := \mathbb{P}(\{X \in A\}) = \mathbb{P}(X^{-1}(A)) \quad \textit{for every } A \in \mathcal{F}'.$$

This can be written in short as

$$\mu := \mathbb{P} \circ X^{-1}.$$

μ is nothing else than the "push-forward" of the probability $\mathbb{P}$ by X to Ω'.

In the special case when $\Omega' = \mathbb{R}$ and X is discrete (meaning that it can take finitely many, or at most countably many values), there is a convenient alternative way to "desribe the distribution of X", by simply listing the possible values x_k and their probabilities $p_k := \mathbb{P}(X = x_k) := \mathbb{P}(\{X = x_k\})$. Then the information contained in the sequence of pairs $\{(x_k, p_k)\}_{k=1}^{N}$ (with possibly $N = \infty$) is called the *discrete probability distribution*. Having this information, one can calculate probabilities of events by summation:

$$\mu(A) = \mathbb{P}(\{X \in A\}) = \sum_{k : x_k \in A} p_k.$$

Similarly, in the special case when $\Omega' = \mathbb{R}$ and X is absolutely continuous (see later), there is convenient alternative "description of the distribution" by a density function $f : \mathbb{R} \to \mathbb{R}^+$ from which one can calculate probabilities of events by integration:

$$\mu(A) = \mathbb{P}(\{X \in A\}) = \int_A f(x)\,\mathrm{d}x.$$

The above notion of a probability distribution is a far-reaching generalization of both notions.

Example 1.27. *sum of two rolled numbers real number generated by a sequence of fair coin tosses real number generated by a sequence of biased coin tosses*

1.3 Integral and expectation

1.3.1 Integral, integrability

For an (extended) *real-valued* measurable functions $X : \Omega \to \mathbb{R}$ on a measure space $(\Omega, \mathcal{F}, \mu)$ it makes sense to talk about the integral $\int_\Omega X\,\mathrm{d}\mu$. This is an essential tool, and also an important object of study both in measure theory and in probability theory. We don't go deep into the definition and properties of the integral here – we don't want to, and we can't substitute a measure theory course now. I just give very briefly one of the shortest possible definitions, and point out a few main feaures.

Since we don't want to exclude the case when either a function or a measure takes the value ∞, we work with extended real numbers, and use the convention

$$0 \cdot \infty := \infty \cdot 0 := 0.$$

We start by defining the integral of nonnegative functions.

Definition 1.28 (integral of non-negative extended real valued functions)**.** *If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $X : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ is measurable, we introduce a sequence $X_n : \Omega \to \mathbb{R}^+$ of* simple *functions (i.e. taking only finitely many values) which approximate X (from below) as*

$$X_n(\omega) := \max\{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\} \,:\, x \le X(\omega)\}.$$

Than we define the n-th integral-approximating sum as

$$I_n := \sum_{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\}} x\mu(X_n^{-1}(\{x\})),$$

and the integral of X as

$$\int_\Omega X \,\mathrm{d}\mu := \lim_{n\to\infty} I_n.$$

The sets $X_n^{-1}(\{x\}) \subset \Omega$ in the definition of I_n are ensured to be $\mathcal{F}$-measurable by the assumption that X is measurable (and the fact that $\{x\}$ is Borel-measurable in $\mathbb{R}$), so $\mu(X_n^{-1}(\{x\}))$ makes sense. The sequence X_n of functions is cleverly defined to be increasing, and so is the sequence I_n, so the limit in the above definition exists, but is possibly infinite.

We can now go on to the general definition of the integral for extended real valued functions:

Definition 1.29 (integral)**.** *If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $X : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is measurable, we introduce the positive part X_+ and the negative part X_- of X as*

$$X_+(\omega) := \begin{cases} X(\omega), & \text{if } X(\omega) > 0, \\ 0, & \text{if not} \end{cases}, \quad X_-(\omega) := \begin{cases} -X(\omega), & \text{if } X(\omega) < 0, \\ 0, & \text{if not} \end{cases}.$$

Note that both X_+ and X_- are nonnegative and $X = X_+ - X_-$.

Now

- *If either $\int_\Omega X_+ \,\mathrm{d}\mu < \infty$ or $\int_\Omega X_- \,\mathrm{d}\mu < \infty$, then we define the integral of X as*
$$\int_\Omega X \,\mathrm{d}\mu := \int_\Omega X_+ \,\mathrm{d}\mu - \int_\Omega X_- \,\mathrm{d}\mu,$$
which can possibly be ∞ or $-\infty$.

- *If both $\int_\Omega X_+ \,\mathrm{d}\mu = \infty$ and $\int_\Omega X_- \,\mathrm{d}\mu = \infty$, then we say that the integral of X doesn't exist (or that it is undefined).*

In the usual mathematical language, there is an important distinction between the existence of an integral and the integrability of a function. We emphasize this in the following definition:

Definition 1.30 (integrability)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $X : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ measurable. If the integral $\int_\Omega X \,\mathrm{d}\mu$ exists and it is* finite*, then we say that X is integrable (with respect to μ).*

So integrability of X means $-\infty < \int_\Omega X \,\mathrm{d}\mu < \infty$, which is equivalent to both $\int_\Omega X_+ \,\mathrm{d}\mu$ and $\int_\Omega X_- \,\mathrm{d}\mu$ being finite.

Remark 1.31 (comparison to the Riemann integral)**.** *The above definition of the integral is similar to the construction of the good old Riemann integral: in both cases the domain of integration, Ω, is chopped up into pieces, on each of which the function X takes nearly constant values. Than the "size" of each piece is multiplied by the approximate value of the function there, and these products are added up to obtain an integral approximating sum. The crucial difference is that in the case of Riemann integral, these small pieces of Ω had to be* intervals*, while here they can be any measurable subset of Ω. In particular, the points in $X_n^{-1}(\{x\})$ don't need to be "close" to each other in any sense, thus Ω absolutely doesn't need to be the real line or anything similar. It doesn't need to have any additional structure that would give sense to the words "distance" or "being close". Really, any measure space will do.*

A trivial but important example of integrable functions:

Lemma 1.32 (bounded functions on finite measure spaces are integrable)**.** *If $(\Omega, \mathcal{F}, \mu)$ is a finite measure space (meaning that $\mu(\Omega) < \infty$) and $X : \Omega \to \mathbb{R}$ is measurable and bounded – meaning that there exists an $M \in \mathbb{R}$ such that $-M \leq X(\omega) \leq M$ for every $\omega \in \Omega$, then X is integrable w.r.t. μ.*

Proof. X being bounded by M implies that $0 \leq X_+(\omega), X_-(\omega) \leq M$. when calculating the integral of, say, X_+, we have $X_n(\omega) \leq X_+(\omega) \leq M$, and thus

$$\begin{aligned} I_n &:= \sum_{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\}} x\mu(X_n^{-1}(\{x\})) \\ &\leq M \sum_{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\}} \mu(X_n^{-1}(\{x\})) \\ &= M\mu(\cup_{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \ldots, 2^n - \frac{1}{2^n}, 2^n\}} X_n^{-1}(\{x\})) \\ &= M\mu(\Omega). \end{aligned}$$

So we get $\int_\Omega X_+ \, \mathrm{d}\mu < M\mu(\Omega)$. Similarly, $\int_\Omega X_- \, \mathrm{d}\mu < M\mu(\Omega)$, so

$$-\infty < -M\mu(\Omega) \leq \int_\Omega X \, \mathrm{d}\mu \leq M\mu(\Omega) < \infty.$$

□

A basic property of the integral with a huge importance is linearity in the function integrated (the integrand):

Theorem 1.33 (linearity of integrability and the integral)**.** *Let X_1 and X_2 be real valued measurable functions on the same probability space $(\Omega, \mathcal{F}, \mu)$,*

and let $\alpha_1, \alpha_2 \in \mathbb{R}$. *If both* $I_1 := \int_\Omega X_1 \,\mathrm{d}\mu$ *and* $I_2 := \int_\Omega X_2 \,\mathrm{d}\mu$ *exist and* $\alpha_1 I_1 + \alpha_2 I_2$ *is not of the form* $\infty - \infty$, *then* $\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2) \,\mathrm{d}\mu$ *exists and*

$$\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2) \,\mathrm{d}\mu = \alpha_1 \int_\Omega X_1 \,\mathrm{d}\mu + \alpha_2 \int_\Omega X_2 \,\mathrm{d}\mu.$$

As a consequence, if X_1 *and* X_2 *are both integrable, then so is* $\alpha_1 X_1 + \alpha_2 X_2$.

The proof is easy from the definition, but we don't discuss it here. See any measure theory book. It is useful to note that the integral is linear not only in the integral, but also in the measure:

Theorem 1.34 (linearity of the integral II.)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space,* μ_1 *and* μ_2 *measures on it,* X *a real valued measurable function and* $0 \leq \alpha_1, \alpha_2 \in \mathbb{R}$. *If both* $I_1 := \int_\Omega X \,\mathrm{d}\mu_1$ *and* $I_2 := \int_\Omega X \,\mathrm{d}\mu_2$ *exist and* $\alpha_1 I_1 + \alpha_2 I_2$ *is not of the form* $\infty - \infty$, *then* $\int_\Omega X \,\mathrm{d}(\alpha_1\mu_1 + \alpha_2\mu_2)$ *exists and*

$$\int_\Omega X \,\mathrm{d}(\alpha_1\mu_1 + \alpha_2\mu_2) = \alpha_1 \int_\Omega X \,\mathrm{d}\mu_1 + \alpha_2 \int_\Omega X \,\mathrm{d}\mu_2.$$

As a consequence, if X *is integrable w.r.t. both* μ_1 *and* μ_2, *then so it is w.r.t.* $\alpha_1\mu_1 + \alpha_2\mu_2$.

Remark 1.35 (bilinearity of the integral)**.** *In the last theorem, we required that* α_1 *and* α_2 *be nonnegative – otherwise* $\alpha_1\mu_1 + \alpha_2\mu_2$ *may not be a measure, since in our definition a measure has to be nonnegative. For the same reason, the measures on a measurable space do not form a linear space. In functional analysis, to overcome that limitation, it is common to introduce the notion of "signed measures" (say, as a difference of two measures), which already form a linear space (with the usual notion of addition and multiplication). Then the two-variable real-valued mapping*

$$\langle \mu, X \rangle := \int X \,\mathrm{d}\mu$$

can be defined on suitably chosen linear spaces of measures and functions (e.g. $\mu \in \{$*signed finite measures*$\}$, $X \in \{$*bounded measurable functions*$\}$*.) The above two theorems show that this mapping is bilinear, which is the property where functional analysis starts.*

1.3.2 Exchanging limits with integrals

It is utmost important that Theorem 1.33 is about the linear combination of *two* integrable functions. Of course, it immediately implies linearity of

the integral for *finite* linear combinations, but *does not say anything about infinite sums*. Indeed, linearity of the integral for infinite sums is not at all true in general. In fact, it is an important issue, in which cases exchanging an integral with a limit is possible – one has to be at least always careful. In the following we state (without proof) three theorems, which are the most frequently (and almost exclusively) used tools in checking exchangeability. In a situation where none of them works, exchangeability is usually hard to prove, and may very well not be true.

The first and most used tool is the Lebesgue dominated convergence theorem:

Theorem 1.36 (dominated convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \ldots$ measurable real valued functions on Ω which converge to the limit function pointwise, μ-almost everywhere. (That is, $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \Omega$, except possibly for a set of x-es with μ-measure zero.) Assume furthermore that the f_n admit a common* integrable dominating function*: there exists a $g : \Omega \to \mathbb{R}$ such that $|f_n(x)| \leq g(x)$ for every $x \in \Omega$ and $n \in \mathbb{N}$, and $\int_\Omega g \, \mathrm{d}\mu < \infty$. Then (all the f_n and also f are integrable and)*

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu.$$

As you see, it is enough to require "almost everywhere" convergence, which is no surprise, since changing a function on a set of measure zero doesn't change the integral. In fact, it would be enough to require that g dominates the f_n almost everywhere – moreover, it would be enough to require that the f_n and g be extended real valued and defined almost everywhere. This is not a serious generalization, so I decided to rather keep the formulation simple. In the literature usually the most general form is given.

The second and easiest tool is Beppo Levi's monotone convergence theorem:

Theorem 1.37 (monotone convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \ldots$ a sequence of measurable* nonnegative *real valued functions on Ω which is pointwise increasing. (That is, $0 \leq f_n(x) \leq f_{n+1}(x)$ for every $n \in \mathbb{N}$ and $x \in \Omega$.) Then the pointwise limit function f defined by $f(x) := \lim_{n\to\infty} f_n(x)$ is also measurable and*

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu.$$

The third and trickiest tool is the Fatou lemma:

Theorem 1.38 (Fatou lemma)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f_1, f_2, ... a sequence of measureabale functions $f_n : \Omega \to \mathbb{R}$, which are nonneagtive, e.g. $f_n(x) \geq 0$ for every $n = 1, 2, \ldots$ and every $x \in \Omega$. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n(x)\,\mathrm{d}\mu(x) \leq \liminf_{n\to\infty} \int_\Omega f_n(x)\,\mathrm{d}\mu(x)$$

(and both sides make sense).

Definition 1.39 (Absolute continuity of measures, singularity of measures)**.** *Let μ and ν be two measures on the same measurable space $\Omega, \mathcal{F}$. We say that ν is absolute continuous with respect to μ (notation: $\nu \ll \mu$) if for every $A \in \mathcal{F}$ for which $\mu(A) = 0$, also $\nu(A) = 0$. We say that ν and μ are singular (with respect to each other) (notation: $\nu \perp \mu$) if there exists and $A \in \mathcal{F}$ for which $\nu(A) = 0$ and $\mu(\Omega \setminus A) = 0$.*

The best known probability distributions are all examples of either one or the other of these: the "discrete" probability distributions are singular w.r.t. Lebesgue measurer on $\mathbb{R}$, while the ones that are loosely called "continuous" are actually absolutely continuous w.r.t. Lebesgue measurer on $\mathbb{R}$.

Theorem 1.40 (Radon-Nykodim)**.** *If μ and ν are two measures on the same measurable space $\Omega, \mathcal{F}$ and $\nu \ll \mu$, then there exists a measurable $f : \Omega \to \mathbb{R}$, called the **density** of ν w.r.t. μ, which satisfies $\nu(A) = \int_A f\,\mathrm{d}\mu$.*

absolute continuity w.r.t. counting measure
absolute continuity w.r.t. Lebesgue measure
integration w.r.t counting measure
integration w.r.t. Lebesgue measure
integration w.r.t. absolutely continouos measures
density function, distribution function
continuity of measures
expectation if it exists - real-valued - complex or R^d -valued - remark about more complicated spaces
integration by substitution

Theorem 1.41. *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, μ a measure on Ω, $X : \Omega \to \Omega'$ measurable and $\nu := \mu \circ X^{-1}$ the push-forward of μ by X (a measure on Ω') (the distribution of X, if μ is a probability). Then for any measurable $g : \Omega' \to \mathbb{R}$*

$$\int_\Omega g(X)\,\mathrm{d}\mu = \int_{\Omega'} g\,\mathrm{d}\nu.$$

expectation of a distribution
linearity of expectation
variance, standard deviation
moments, centered moments, moment-generating function, characteristic function
product measure space, product measure - be careful with infinite products
pairwise independence, marginal distributions
mutual independence
variance of linear combination of independent random variables
Markov inequality
weak law of large numbers
central limit theorem
weak convergence of measures
weak convergence of random variables
continuity of the characteristic function

1.4 conditioning

If A and B are events on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}(A) \neq 0$, then the conditional probability of B under the condition A is defined as

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(A)}.$$

What we do now is a generalization of this notion for the case when $\mathbb{P}(A) = 0$. We do not restrict to the case of probability measures.

The generalization happens through the "integral form" of the above equation, which is the "law of total probability": If $A_1, A_2, \ldots$ are pairwise disjoint and $\cup_i A_i = \Omega$, then

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\mathbb{P}(B|A_i).$$

As a result, we will not be able to generalize the notion of $\mathbb{P}(B|A)$ for a *single* event A with $\mathbb{P}(A) = 0$. Instead, we need a *family* of (possibly zero probability) conditions.

Definition 1.42. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A function $p : \Omega_1 \times \mathcal{F}_2 \to \mathbb{R}$ is called a* kernel *from Ω_1 to Ω_2, if*

- *$\forall x \in \Omega_1$, $p(x, .)$ is a measure on $(\Omega_2, \mathcal{F}_2)$*

- $\forall A \in \mathcal{F}_2$, $p(.,A)$ *is an* $\mathcal{F}_1$*-measurable function (from* Ω_1 *to* $\mathbb{R}$*).*

p is called a probability kernel *if* $\forall x \in \Omega_1$, $p(x,\Omega_2) = 1$.

In certain cases, the numbers $p(x,A)$ can play the role of "conditional probability of A under the condition x". In other cases $p(x,.)$ can be called a "conditional measure" on $(\Omega_2,\mathcal{F}_2)$.

Definition 1.43. *Let $(\Omega_1,\mathcal{F}_1)$ and $(\Omega_2,\mathcal{F}_2)$ be measurable spaces, μ_1 a measure on $(\Omega_1,\mathcal{F}_1)$ and $p : \Omega_1 \times \mathcal{F}_2$ a kernel. The "composition of μ_1 and p", denoted by $\int p\,\mathrm{d}\mu_1$ is a measure on $(\Omega_2,\mathcal{F}_2)$ defined by*

$$\left(\int p\,\mathrm{d}\mu_1\right)(A) := \int_{\Omega_1} p(x,A)\,\mathrm{d}\mu_1(x)$$

for every $A \in \mathcal{F}_2$.

Remark 1.44. *On the lecture I used the notation $\mu \otimes p$ to denote two different notions (one of them was the above $\int p\,\mathrm{d}\mu$, se below for the other), and caused a lot of confusion. I will try to avoid this notation here.*

This way we have built up a measure from many component measures. The phrases *conditional measure* and *conditional probability* are typically used with slightly different meanings:

In the following definition, imagine Ω_1 as a small set of parameters and Ω_2 as a big set, decomposed into disjoint subsets Ω_x with $x \in \Omega_1$.

Definition 1.45 (Conditional measure)**.** *Let $(\Omega_1,\mathcal{F}_1)$ and $(\Omega_2,\mathcal{F}_2)$ be measurable spaces and $\pi : \Omega_2 :\to \Omega_1$ measurable. Use the notation*

$$\Omega_x := \pi^{-1}(x) := \{\omega \in \Omega_2 : \pi(\omega) = x\} \quad \text{for every } x \in \Omega_1.$$

For a measure μ on Ω_2 we say that it admits a decomposition w.r.t. *π if there exist μ_1 and p such that*

- *μ_1 is a measure on $(\Omega_1,\mathcal{F}_1)$,*
- *p is a kernel from Ω_1 to Ω_2,*
- *$p(x,.)$ is concentrated on Ω_x for every $x \in \Omega_1$ – that is, $p(x,\Omega_2 \setminus \Omega_x) = p(x,\{\omega \in \Omega_2 : \pi(\omega) \neq x\}) = 0$,*
- *$\mu = \int p\,\mathrm{d}\mu_1$.*

In this case we also say that $p(x,.)$ is the conditional measure of μ under the condition $\{\pi = x\}$.

Remark 1.46. *It is easy to see that μ_1 has to be the push-forward of μ w.r.t. π and that if μ admits two decompositions p_1 and p_2, then $p_1(x,.) = p_2(x,.)$ for μ_1-almost every x. Rokhlin's theorem ensures that a decomposition (and thus conditional measures) exist in the cases we will be interested in.*

Remark 1.47. *Note that in the above definition, Ω_2 doesn't need to have a true product structure such that Ω_1 is one of the components. Indeed, in our most important application, Ω_2 will be the phase space of a Hamiltonian system and π will be the Hamiltonian function H, so the Ω_x will be the "energy level surfaces", which are absolutely not flat – it's better to picture them like the surface of a potato.*

As opposed to the above notion of "conditional measure", the word "conditional probability" is used for a decomposition of the *joint distribution* of two random variables, which is a measure on the product space $\mathbb{R} \times \mathbb{R}$:

Definition 1.48 (Conditional probability)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \Omega_1$, $Y : \Omega \to \Omega_2$ random variables. (You can think of $\Omega_1 = \Omega_2 = \mathbb{R}$.) Let $\mu_1 := \mathbb{P} \circ X^{-1}$ denote the distribution of X, and let p be a probability kernel from*
Omega$_1$ to Ω_2. We say that p is the conditional (probability) distribution of Y under the condition X if the joint distribution can be decomposed in the form

$$\mathbb{P}((X,Y) \in D) = \left(\int \hat{p} \, \mathrm{d}\mu_1 \right)(D) \quad \text{for every } D \subset \Omega_1 \times \Omega_2 \text{ measurable,} \tag{1}$$

where $\hat{p}$ is a kernel from Ω_1 to $\Omega_1 \times \Omega_2$ constructed from p as

$$\hat{p}(x, D) := p(x, D_x) := p(x, \{y \in \Omega_2 : (x,y) \in D\}).$$

Remark 1.49. *This D_x is the projection to the Ω_2 (the "y axis") of the intersection of D with the $\{x\} \times \Omega_2$ (the "vertical line at x"), so $\hat{p}(x,.)$ is concentrated on $\{x\} \times \Omega_2 \subset \Omega_1 \times \Omega_2$. The definition (1) can be written without introducing $\hat{p}$ in the slightly more scaring form*

$$\mathbb{P}((X,Y) \in D) = \int_{\Omega_1} \left(\int_{\Omega_2} \mathbf{1}_D(x,y) p(x, \mathrm{d}y) \right) \mathrm{d}\mu_1(x).$$

Of course, it is enough to demand this for $D \subset \Omega_1 \times \Omega_2$ that are of product from $A \times B$, so (1) holds iff

$$\mathbb{P}(X \in A, Y \in B) = \int_A p(x, B) \, \mathrm{d}\mu_1(x) \quad \text{for } \forall A \subset \Omega_1, B \subset \Omega_2 \text{ measurable.}$$

1.5 Entropy

Definition 1.50. *Let $M = \{x_1, x_2, \ldots, x_n\}$ be a finite set. For a probability measure μ on M, the* entropy *of μ is*

$$S(\mu) := -\sum_{k=1}^{n} p_k \log p_k,$$

with the convention $0 \log 0 := 0$, where $\{p_k\}$ is the "discrete probability vector" defined as $p_k := \mu(\{x_k\})$. This is clearly independent of the order of points in the set $M = \{x_k\}$.

Many interesting properties of this notion were discussed. Most importantly,

1. $S \in [0, \log n]$,
2. $S(\mu)$ is zero iff μ is concentrated on a single point.
3. On a given set $M = \{x_1, x_2, \ldots x_n\}$, the entropy $S(\mu)$ is maximal and equal to $\log n$ iff μ is the uniform distribution.
4. If a measure is decomposed into conditional measures – that is, there is a (finite) index set $I = \{i\}$ with a measure μ_1 on it and a kernel P from I to M such that the measures $\nu_i := P(i, .)$ are concentrated on disjoint subsets of M and $\mu = \int P \, \mathrm{d}\mu_1$, then

   $$S(\mu) = S(\mu_1) + \sum_{i \in I} p_i S(\nu_i),$$

   where $p_i := \mu_1(\{i\})$.

   In particluar, if μ is a product measure $\mu = \mu_1 \otimes \mu_2$, then $S(\mu) = S(\mu_1) + S(\mu_2)$.

The next definition generalizes this notion to measures on possibly non-finite sets:

Definition 1.51. *Let μ and ν be two measures on the same measurable space $(\Omega, \mathcal{F})$. The* relative entropy *of μ with respect to ν is*

$$S(\mu|\nu) := \begin{cases} -\int_\Omega f \log f \, \mathrm{d}\nu, & \text{if } \mu \text{ has density } f \text{ w.r.t. } \nu \\ -\infty, & \text{if } \mu \text{ is not absolutely continuous w.r.t. } \nu \end{cases}.$$

Properties:

1. The old notion of entropy is exactly the relative entropy w.r.t. counting measure.

2. As opposed to the previous notion of entropy, this relative entropy can in general be negative.

3. If ν is a finite measure with $\nu(\Omega) = K < \infty$, then $S(\mu|\nu) \leq \log K$ and the maximum $\log K$ is obtained iff μ is uniform w.r.t. ν – that is, $F \equiv \frac{1}{K}$, of simply $\mu = \frac{1}{K}\nu$.

4. However, if $\nu(\Omega) = \infty$, then $S(\mu|\nu)$ is not bounded and there is no probability measure with maximal relative entropy.

5. However[2], the problem of measures with maximal relative entropy *under some constraints* (on expectations of observables) is of special interest in Physics. See Homework 4.6, 5.2 and 5.3.

2 Equilibrium statistical ensembles in finite systems

In this section we consider "finite systems", by which we mean Hamiltonian systems in some finite box $\Lambda \subset \mathbb{R}^d$. This, however, will not mean that the number of particles is also limited to some finite number. In fact, in the first two cases (the microcanonical and the canonical ensemble) the particle number will be fixed, but in the third case (the grand canonical ensemble) it can be unbounded (although finite with probability 1).

In all three cases we describe the system "in equilibrium", under different conditions. The definitions will not follow in any strict sense from "first principles" – that is, laws of motion for the underlyin Hamiltonian system – but we will argue that it is a good and natural idea to define equilibrium states this way.

The first key idea is contained the following definition:

Definition 2.1. *A* statistical ensemble *is a probability measure on the phase space of a physical (here: Hamiltonian) system. An* equilibrium *statistical ensemble is a probability measure on the phase space which is* invariant under the dynamics.

It is important to understand the distinction between a "state" or "microstate" of the Hamiltonian system, which is nothing else than a point in the phase space, and a "state" or "macrostate" of a "Statistical Physical system" or just "Statistical ensemble". This can be pictured as a huge set

of copies of the same Hamiltonian system, being in different phase points ("microstates"). As a mathematical abstraction, we start the have a system in "every possible microstate", with some weight associated to it. The mathematical notion is exactly the notion of a @emphmeasure on the phase space: it reflects the uncertaintiy in our knowledge of the microstate, so that we can not say in which microstate our system is in – instead, we can only talk about probabilities of the microstate having this or that property.

The notion of invariance of a measure will be discussed in detail in Section 5.

Consider a Hamiltonian system of N particles in a finite box $\Lambda \subset \mathbb{R}^d$ with phase space Ω and Hamiltonian function $H : \Omega \to \mathbb{R}$. Maybe surprisingly for the first look, we will not be very interested in describing the microscopic dynamics – we will only make use of a single feature of Hamiltonian systems: the existence of a natural invariant measure, called Liouville measure.

Definition 2.2. *Consider a Hamiltoninan system with n degrees of freedom, phase space Ω and canonical coordinates $q_1, \dots, q_n$ with associated momenta $p_1, \dots p_n$. The Liouville measure ν of this system is the measure on Ω which is locally the $2n$-dimensional Lebesgue measure: $\mathrm{d}\nu = \mathrm{d}^n q \, \mathrm{d}^n p$.*

Remark 2.3. *Physicists often call this measure "phase space volume". The reason for the tricky definition is that in general, the phase space of a Hamiltonian system can be some (Riemannian) manifold, so speaking about Lebesgue measure doesn't really make sense. In the simplest case of point particles where the q_i are actually Cartesian coordinates of particle positions and the p_i are velocities (times mass), Ω is just a piece of $\mathbb{R}^{2n}$ and the Liouville measure is exactly the Lebesgue measure. Since this is not a course on Hamiltonian mechanics, we will only consider this simplest case in our explicit calculations.*

Theorem 2.4. *The Liouville measure is invariant under the Hamiltonian dynamics.*

Physicists often furmulate this theorem as "Phase space volume is invariant". The proof is an easy calculation, but we don't go into it now – in particluar the precise meaning of "invariance of a measure", or the "time evolution of a measure" will be discussed in Section 5. We can do without those now, because we are only building *equilibrium* Statistical Physics at the moment: we are not interested in the "dynamics" of the statistical physical system, which would be the time-dependence of the macrostates=ensembles=measures. The only thing we care about is to make sure that those measures, which we will call *equilibrium ensembles*, don't evolve anywhere, but are time-invariant.

2.1 Identical particles and factorization of the phase space

To get physically correct results, we make the following

Ansatz 2.5 (Identification of identical particles)**.** *Identical particles (particles of the "same kind") are indistinguishable. For that reason, we consider the phase space of the statistical physical system to be the Hamiltonian phase space factorized by the eqivalence relation $\sim$, where $\omega_1 \sim \omega_2$ for $\omega_1, \omega_2 \in \Omega$ (the states ω_1 and ω_2 are equivalent) if ω_1 can be transformed into ω_2 by renumbering identical particles:*

$$\Omega := \Omega^{Ham} / \sim .$$

This ansatz, which does not follow in any way from our (classical mechanical) microscopic descripiton of the physical systems (i.e. Newton's law), should be considered as a **physical law** based on experience, which has to be included as an **axiom** in the mathematical discussion. It has an explanation in quantum mechanics, but we don't go into that.

The reader may be worried that an explicit description of the dynamics (writing the equations of motion) on this factorized phase space could be difficult, but we will never need to do that. In fact, the only things we use about Hamiltonian dynamics are that

- it leaves the Liouville measure invariant, and
- energy is preserved.

The only practical consequence of the above ansatz is the following

Definition 2.6. *Let π denote the canonical projection from Ω^{Ham} to $\Omega := \Omega^{Ham} / \sim$. If a system contains N identical particles, the "Liouville measure" of a set $A \in \Omega := \Omega^{Ham} / \sim$ is*

$$\nu(A) := \frac{1}{N!} \tilde{\nu}(\pi^{-1}(A))$$

where $\tilde{\nu}$ denotes the Liouville measure on Ω^{Ham} (before factorization).

This definition results in the utmost important appearence of the factor $\frac{1}{N!}$ in the formulas involving the microcanonical, canonical and grand canonical measures we will introduce. The consequences were discussed in detail in class.

2.2 Microcanonical ensemble

It is obvious that the Liouville measure is not the only invariant measure, and not even the only interesting invariant measure for a Hamiltonian system. In fact, since Hamiltonian dynamics preserves the energy, H is an invariant function, and consequently the level sets of H are invariant sets:

$$\{H = E\} := \Omega^\Lambda_{N,E} := \{\omega \in \Omega^\Lambda_N\} : H(\omega) = E$$

can never be left by any phase point which was once in it. Similarly, set of the form $E_{min} \leq H \leq E_{max}$ are invariant, and the restriction of the Liouville measure to such sets is also invariant.

In the definition of the microcanonical ensemble we would like to look at such a "restriction" to a single level set $\{H = E\}$, which however (usually) has zero Liouville measure, so the sensible notion is that of conditioning in Definition 1.45:

Definition 2.7 (Microcanonical ensemble)**.** *Consider a Hamiltonian system of N particles in a finite box $\Lambda \subset \mathbb{R}^d$ with phase space Ω^Λ_N (with indistinguishability of identical particles (Ansatz 2.5) in mind) and Hamiltonian function $H : \Omega^\Lambda_N \to \mathbb{R}$. For possible values E of H set $\{H = E\} := \Omega^\Lambda_{N,E} := \{\omega \in \Omega^\Lambda_N : H(\omega) = E\}$. If the conditional measure of the Liouville measure, under the condition $\{H = E\}$ exists (that is, the Liouville measure admits a decomposition w.r.t. H) and the conditional measure is finite, denote it by $\nu_{\Lambda,N,E}$ and call it the (microcanonical) reference measure. In this case we say that the microcanonical ensemble $\mu_{micr} = \mu_{\Lambda,N,E}$ exists, and it is the probability measure on $\Omega^\Lambda_{N,E}$ which is uniform w.r.t. $\nu_{\Lambda,N,E}$.*

Of course μ_{micr} being the "probability measure which is uniform w.r.t. $\nu_{\Lambda,N,E}$" is just a tricky way of saying that μ_{micr} is the normalized version of $\nu_{\Lambda,N,E}$:

$$\mu_{\Lambda,N,E} = \frac{1}{Z_{micr}(\Lambda, N, E)} \nu_{\Lambda,N,E},$$

where

Definition 2.8 (Microcanonical partition function)**.**

$$Z_{micr}(\Lambda, N, E) := \nu_{\Lambda,N,E}(\Omega^\Lambda_{N,E}) = \nu_{\Lambda,N,E}(\Omega^\Lambda_N)$$

is the microcanonical partition function.

Definition 2.9 (Microcanonical entropy)**.** *The microcanonical entropy is defined as the relative entropy*

$$S_{micr}(\Lambda, N, E) := H(\mu_{\Lambda,N,E}|\nu_{\Lambda,N,E}).$$

It is immediate from the definition that

$$S_{micr}(\Lambda, N, E) = \log Z_{micr}(\Lambda, N, E).$$

Acually, in Physics $S_{micr}(\Lambda, N, E) = k \log Z_{micr}(\Lambda, N, E)$ where k is Boltzamann's constant, but we will omit k (which is equivalent to choosing units so that $k = 1$).

Remark 2.10. *The definition of the microcanonical enseble is justified by the* belief *that in true physical systems the microcanonical measure cannot be further decomposed into invariant measures – in a way, it is the only physically interesting invariant measure on the energy level set $\{H = E\}$, at least with good approximation when the system is big (consisting of many particles). This is roughly the content of the Boltzmann ergodic hypothesis, which belongs to Section 5.*

Having introduced the microcanonical entropy, we can define the microcanonical temperature T, pressure P and chemical potential μ. For this, we assume that Z_{micr} depends on Λ only through the volume V of Λ with good approximation. In this case we set

$$\frac{1}{T(V, N, E)} := \beta(T, N, E) := \frac{\partial}{\partial E} S(V, N, E) \tag{2}$$

$$\frac{P(V, N, E)}{T(V, N, E)} := \frac{\partial}{\partial V} S(V, N, E) \tag{3}$$

$$-\frac{\mu(V, N, E)}{T(V, N, E)} := \frac{\partial}{\partial N} S(V, N, E). \tag{4}$$

(The differentiation w.r.t. N is usually performed formally, but we can also mean the discrete derivative, so $\frac{\partial}{\partial N} S(V, N, E) := S(V, N + 1, E) - S(V, N, E)$.)

However, in the present context of the microcanonical ensemble, these notions have no probabilistic interpretation, since we are differentiating w.r.t. parameters whose different choices result in completely different measures on different spaces, and the normalizing factor Z comes from the pretty arbitrary choice of the reference measure. The probabilistic interpretations (at least for the temperature and the chemical potential) will come from the canonical and the grand canonical ensembles.

To make the calculation of the of the microcanonical measure and partition functiion easier, we mention without proof the following (easy) theorem:

Theorem 2.11 (Calcualtion with the microcanonical measure)**.** *Suppose that H is differentiable and $\nabla H \neq 0$ on $H = E$. Then $\{H = E\}$ is a surface in*

the Hamiltoninan phase space and $\nu_{\Lambda,N,E}$ is absolutely continuous w.r.t. the surface measure σ_E, with density $\frac{1}{||\nabla H||}$, where $||.||$ denotes Euclidean norm. Consequently, $Z_{micr}(\Lambda, N, E) = \int \frac{1}{||\nabla H||} \, \mathrm{d}\sigma_E$.

Note that we consider the Hamiltonian phase space with the identification of indistinguishable particles, so the surface measure σ_E contains the factor $\frac{1}{N!}$ is we have N identical particles. For an application, see Homework 4.7.

2.3 Canonical ensemble

Similarly to the microcanonical ensemble, the canonical ensemble we are about to define is a measure on the Hamiltonian phase space. However, as opposed to the microcanonical setting, we will not fix the energy, but allow it to be a random variable. This models a system in a heat bath, with which it can exchange energy (unlike the microcanonical ensemble, which models a closed system).

The density w.r.t. Liouville measure will only depend on the energy:

Definition 2.12 (Canonical ensemble)**.** *Consider a Hamiltonian system of N particles in a finite box $\Lambda \subset \mathbb{R}^d$ with phase space Ω_N^Λ (with indistinguishability of identical particles (Ansatz 2.5) in mind) and Hamiltonian function $H : \Omega_N^\Lambda \to \mathbb{R}$. Let $\beta \in \mathbb{R}^+$. The canonical ensemble with inverse temperature β is the probability measure $\mu_{can} = \mu_{\Lambda,N,\beta}$ on Ω_N^Λ which is absolutely continuous w.r.t. the Liouville measure $\nu_{\Lambda,N}$ and has the density*

$$\frac{\mathrm{d}\mu_{\Lambda,N,\beta}}{\mathrm{d}\nu_{\Lambda,N}}(\omega) := \frac{1}{Z_{can}(\Lambda, N, \beta)} e^{-\beta H(\omega)}$$

(if it exists).

Definition 2.13 (Canonical partition function)**.** *The normalizing factor*

$$Z_{can}(\Lambda, N, \beta) := \int_{\Omega_N^\Lambda} e^{-\beta H} \, \mathrm{d}\nu_{\Lambda,N}$$

is the microcanonical partition function.

Clearly, μ_{can} exists exactly if $Z_{can} < \infty$.

The fact that the density $\frac{\mathrm{d}\mu_{\Lambda,N,\beta}}{\mathrm{d}\nu_{\Lambda,N}}$ depends on H only, ensures that we again have an invariant measure under the Hamiltonian dynamics (meaning: not taking into account the effect of the environment). Actually, it's a composition of the microcanonical measures on the different energy surfaces. Precisely, if we condition the canonical measure on the $\{H = E\}$

surfaces (and normalize), we get exactly the microcanonical measures back. Of course, such a composition of the microcanonical measures colud be done in many ways – in other words, one could choose many different densities (w.r.t. Liouville measure) that depend on the energy only. The definition of the canonical measure specifies one of these many possible choices. The reason for the choice was discussed in detail in the lectures by considering some trivial model of the heat bath.

Definition 2.14 (Canonical energy)**.** *The canonical energy is defined as the expectation of the Hamiltonian w.r.t. the canonical measure:*

$$E = E_{can}(\Lambda, N, \beta) := \int H(\omega)\, \mathrm{d}\mu_{\Lambda,N,\beta}(\omega).$$

An easy calculation gives

Theorem 2.15.

$$E_{can} = -\frac{\partial}{\partial\beta} \log Z_{can}(\Lambda, N, \beta).$$

Definition 2.16 (Canonical entropy)**.** *The canonical entropy is defined as the relative entropy*

$$S_{can}(\Lambda, N, \beta) := H(\mu_{\Lambda,N,\beta}|\nu_{\Lambda,N}).$$

Theorem 2.17.

$$S_{can} = \beta E_{can} + \log Z_{can}$$

meaning

$$S_{can}(\Lambda, N, \beta) = \beta E_{can}(\Lambda, N, \beta) + \log Z_{can}(\Lambda, N, \beta).$$

Proof. Homework 5.2. □

Having introduced the canonical partition function, we can define the temperature T, free energy A, the canonical pressure P and chemical potential μ. As before, we assume that Z depends on Λ only through the volume V of Λ. In this case we set

$$\begin{aligned} T &:= \frac{1}{\beta} \\ -\beta A(V, N, \beta) &:= \log Z(V, N, \beta) \\ -P(V, N, \beta) &:= \frac{\partial}{\partial V} \log Z(V, N, \beta) \\ -\mu(V, N, \beta) &:= \frac{\partial}{\partial N} \log Z(V, N, \beta). \end{aligned}$$

Note that – as opposed to the microcanonical setting, the temperature is a parameter and the energy has a clear probabilistic interpretation.

2.4 Grand canonical ensemble

Assume now that our system can exchange not only energy, but also particle with its environment – so it is not only in a "heat bath", but also in a "particle bath". This models e.g. a volume in space which is only defined by the observing scientist (and not surrounded by walls). Now not only the energy, but also the number of particles will be a random variable. We reach this by manually building up a "grand canonical phase space" as a disjoint union of the Hamiltonian (canonical) phase spaces for the different N. The Hamiltonian dynamics will of course never change the particle number, so the construction may seem unnatural, but remember that the canonical phase space is also the disjoint union of the microcanonical phase spaces, and the dynamics doesn't change the energy either. If fact, in both cases our model "ignores" the effect of the environment – more preciely, it assumes that the effect of the environment can be seen as setting the parameters of the ensemble. In the canonical ensemble we only had one parameter, the temperature, which turned out to fix the expectation of the energy. Now we will have two parameters, the temperature and the chamical potantial, which will fix the expectation of both the energy and the particle number.

Definition 2.18 (Grand canonical ensemble)**.** *Consider, for every $n = 0, 1, 2, \dots$ a Hamiltonian system of n particles in a finite box $\Lambda \subset \mathbb{R}^d$ with phase space Ω_n^Λ (with indistinguishability of identical particles (Ansatz 2.5) in mind) and Hamiltonian function $H_n : \Omega_n^\Lambda \to \mathbb{R}$. ($\Omega_n^\Lambda$ consists of a single point $\emptyset$ and $H_0(\emptyset) := 0$.) The grand canonical phase space is the disjoint union*

$$\Omega^\Lambda := \bigcup_{n\geq 0} \Omega_n^\Lambda := \{(n, \omega_n) : n \in \mathbb{N}, \omega_n \in \Omega_n^\Lambda\},$$

and the grand canonical reference measure on Ω^Λ is $\nu_\Lambda := \sum_{n\geq 0} \nu_{\Lambda,n}$ meaning

$$\nu_\Lambda(A) := \sum_{n\geq 0} \nu_{\Lambda,n}(\{\omega_n \in \Omega_n^\Lambda : (n, \omega_n) \in A\}),$$

where $\nu_{\Lambda,n}$ is the Liouville measure on Ω_n^Λ.

On Ω^Λ define the observables N (particle number) and H (Hamiltonian) as

$$N((n, \omega_n)) := n \quad , \quad H((n, \omega_n)) := H_n(\omega_n).$$

Let $\beta \in \mathbb{R}^+$ and $\beta' \in \mathbb{R}$. The grand canonical ensemble with parameters β, β' is the probability measure $\mu_{gr} = \mu_{\Lambda,\beta,\beta'}$ on Ω^Λ which is absolutely continuous w.r.t. reference measure ν_Λ and has the density

$$\frac{\mathrm{d}\mu_{\Lambda,\beta,\beta'}}{\mathrm{d}\nu_\Lambda}(\omega) := \frac{1}{Z_{gr}(\Lambda, \beta, \beta')} e^{-\beta H(\omega) - \beta' N(\omega)}$$

(if it exists).

Definition 2.19 (Grand canonical partition function)**.** *The normalizing factor*

$$Z_{gr}(\Lambda, \beta, \beta') := \int_{\Omega^\Lambda} e^{-\beta H - \beta' N} \, \mathrm{d}\nu_\Lambda$$

is the grand canonical partition function.

Clearly, μ_{gr} exists exactly if $Z_{gr} < \infty$.
Using the definition of ν_Λ, N and H we get that

$$Z_{gr}(\Lambda, \beta, \beta') = \sum_{n \geq 0} e^{-\beta' n} \int_{\Omega_n^\Lambda} e^{-\beta H} \, \mathrm{d}\nu_{\Lambda,n} = \sum_{n \geq 0} e^{-\beta' n} Z_{can}(\Lambda, n, \beta).$$

Let E denote the expectation of H and $\bar{N}$ denote the expectation of N:

$$E = E_{gr}(\Lambda, \beta, \beta') := \int H(\omega) \, \mathrm{d}\mu_{\Lambda,\beta,\beta'}(\omega),$$
$$\bar{N} = \bar{N}_{gr}(\Lambda, \beta, \beta') := \int N(\omega) \, \mathrm{d}\mu_{\Lambda,\beta,\beta'}(\omega).$$

An easy calculation gives

Theorem 2.20.

$$E_{gr} = -\frac{\partial}{\partial \beta} \log Z_{gr}(\Lambda, \beta, \beta'),$$
$$\bar{N}_{gr} = -\frac{\partial}{\partial \beta'} \log Z_{gr}(\Lambda, \beta, \beta').$$

Definition 2.21 (Grand canonical entropy)**.** *The grand canonical entropy is defined as the relative entropy*

$$S_{gr}(\Lambda, \beta, \beta') := H(\mu_{\Lambda,\beta,\beta'} | \nu_\Lambda).$$

Theorem 2.22.

$$S_{gr} = \beta E_{gr} + \beta' \bar{N}_{gr} + \log Z_{gr}$$

meaning

$$S_{gr}(\Lambda, N, \beta) = \beta E_{gr}(\Lambda, \beta, \beta') + \beta' \bar{N}_{gr}(\Lambda, \beta, \beta') + \log Z_{gr}(\Lambda, \beta, \beta').$$

Proof. Homework 5.3. □

Having introduced the grand canonical partition function, we can define the temperature T, chemical potential μ, grand free energy G, the grand canonical pressure P. As before, we assume that Z depends on Λ only through the volume V of Λ. In this case we set

$$\begin{aligned} T &:= \frac{1}{\beta} \\ \beta' &:= -\beta\mu \\ G(V,\beta,\beta') &:= -\frac{1}{\beta}\log Z(V,\beta,\beta') \\ P(V,\beta,\beta') &:= -\frac{1}{V}G(V,\beta,\beta') = \frac{1}{\beta V}\log Z(V,\beta,\beta') \end{aligned}$$

Note that this time both the temperature and the chemical potential are parameters, and the energy and the average particle number have probabilistic interpretations.

3 Thermodynamic limit

3.1 Convergence of thermodynamic functions

In the thermodynamic limit, meaning $V \to \infty$, we hope to see that the thermodynamic quantites scale with V as expected from their physical interpretations. In particular

- In the canonical setting, when fixing β and choosing the parameters V and N so that the density convegres (meaning $\frac{N(V)}{V} \to \rho$ as $V \to \infty$) we hope to see that the energy and also the free energy scale linearly with V which means exactly that $\frac{1}{V}\log Z$ convegres to a (finite nonzero) limit.
- In the grand canonical setting, when fixing β and β', we hope to see that the grand free energy scales linearly with V which means exactly that $\frac{1}{V}\log Z$ (and also the pressure) converges to a (finite nonzero) limit.

Remark 3.1. *The existence of these limits in turn imply the proper scaling of the expectations (energy in both cases and also the average particle number in the grand canonical case) unless there is some issue with exchangeability of the limit and the differentiation. Investigating such issues (i.e. non-differentiability of the limiting quantities) is the main interest in equilibrium statistical physics, being related to the presence of a phase transition in the model.*

The above thermodynamic limits don't always exist. To study this issue, we restrict to the simplest possible Hamiltonian system: identical point particles with mass m interacting only via a spherically symmetric pair interaction potential. In particular, we fix a domain $\Lambda \subset \mathbb{R}^d$, the Hamiltonian phase space to be

$$\Omega_n^\Lambda = \Lambda^n \times \left(\mathbb{R}^d\right)^n$$

and the Hamiltonian to be

$$H_n(q_1, \ldots, q_n, p_1, \ldots, p_n) := \sum_{i=1}^n \frac{p_i^2}{2m} + \frac{1}{2} \sum_{i \neq j} \Phi(|q_i - q_j|)$$

where $\Phi : \mathbb{R}^+ \to \mathbb{R}$ is the pair interaction potential.

If we consider the canonical ensemble for this Hamiltonian, it is clear that the positions and the velocities are independent (random variables), and that the velocities can be described explicitly – i.e. they are just the same as in the free gas with $\Phi \equiv 0$ discussed in Homework 5.4. This means that it's enough to understand the case when we ignore the velocities completely by rather setting the phase space to be

$$\Omega_n^\Lambda = \Lambda^n \times$$

and the Hamiltonian to be

$$H_n(q_1, \ldots, q_n, p_1, \ldots, p_n) := \frac{1}{2} \sum_{i \neq j} \Phi(|q_i - q_j|).$$

This model is called the *configuration gas*. Its relation to the original model on the level of the partition function is discussed in Homework 10.2.

The definitions and theorem below give a sufficient condition for the existence of the limiting free energy per particle in the canonical setting.

Definition 3.2 (tempered pair interaction)**.** *The pair interaction potential $\Phi : \mathbb{R}^+ \to \mathbb{R}$ is tempered if there is a real $R < \infty$ for which $\Phi(r) \leq 0$ for all $r > R$.*

Definition 3.3 (stable pair interaction)**.** *The pair interaction potential $\Phi : \mathbb{R}^+ \to \mathbb{R}$ is stable (in d dimensions) if there is a real $B < \infty$ for which*

$$\Phi(q) := \sum_{q \leq i < j \leq N} \Phi(|q_i - q_j|) \geq -BN$$

holds for every $N \in \mathbb{N}$ and every $q \in \left(\mathbb{R}^d\right)^N$.

Note that the definition is very implicit and that the same B has to work for every N.

Theorem 3.4 (Fisher and Ruelle, 1963)**.** *Consider the configuration gas in $\Lambda \subset \mathbb{R}^d$ with the pair interaction potential Φ, in the canonical ensemble. Suppose that Φ is tempered and stable. Then in the thermodynamic limit $\Lambda \nearrow \mathbb{R}^d$ and $\frac{N}{V} \to \rho$, the free energy density converges to a (non-constant) function $a(\rho, \beta)$ which is convex (and continuous) in ρ.*

Note that to be precise, one would need to set some conditions on how and in what sense $\Lambda \nearrow \mathbb{R}^d$, ensuring not only that every point is sooner or later included, but also that the shape of Λ is "nice" in the sense that points close to the boundray become less and less important as Λ grows. As an example, a sequance of bowes where the ratios of sidelengths are bounded away form 0 would do. On the lecture the proof was presented for a special sequence of squares.

3.2 Gibbs measures

Another approach to thermodynamic limits is not to consider limiting behaviour of the partition function, but the limiting behaviour of the measure itself. This is highly non-trivial, since

- First of all, simply setting $\Lambda = \mathbb{R}^d$ clearly doesn't make sense, because in the interesting cases the particle number would be infinite and the Hamiltoninan also infinite or minus infinity.

- If we consider the Hamiltonian system in a big box λ_2, but only look at what we see in a smaller box $\Lambda_1 \subset \Lambda_2$, we don't get the "system with Λ_1" back – e.g. neither the particle number, nor the energy is preserved by the dynamics. What we could hope is that the *grand canonical distribution* of the big system, when viewed only in Λ_1, gives the *grand canonical distribution* of the system with Λ_1 back. This is indeed the case, and this "compatibility" property will be used to define "a measure with $\Lambda = \mathbb{R}^d$" called a *Gibbs measure*. However, it is important to understand what we exactly mean by "the *grand canonical distribution* of the big system, when viewed only in Λ_1". This cannot simply be a *marginal* of the distribution, since then we would ignore the interaction between the susbystems in and outside Λ_1, which are or course there in a Hamiltonian systems, and are responsible for all the interesting statistical physical phenomena.

For that reason, we have to start by introducing a *conditional grand canonical distribution* on a finite Λ,taking also the environment into account. This will be defined just like the good old grand canonical distribution, but using a *conditional Hamiltonian*, which also takes into account the interaction with the environment.

We restrict to the case of pair intercations. The conditional Hamiltonian will contain the contribution to the total energy by the particles inside λ as well as pairs that consist of a particle inside and one outside. The positions of the particles outside are considered as parameters, and the state of those inside are the variables of or function.

Since the environment typically contains infinitely many particles, we have to define the phase space Ω for $\Lambda = \mathbb{R}^d$, which is the generalization of our earlier grand canonical phase space ω^Λ. This is not difficult: instead of a list of particles (meaning an element of $\left(\Omega_1^\Lambda\right)^n = \left(\Lambda \times \mathbb{R}^d\right)^n$) we consider a *set of particles* meaning a subset of the one-particle phase space $\Omega_1^\Lambda = \Lambda \times \mathbb{R}^d$. We only have to make sure that the number of particles in every finite box is finite:

Definition 3.5 (locally finite configuration)**.** *$\omega \subset \mathbb{R}^d \times \mathbb{R}^d$ is called a locally finite configuration if $\omega \cap (\Lambda \times \mathbb{R}^d)$ is finite for every finite box $\Lambda \subset \mathbb{R}^d$.*

(Note that this is not exactly the notion of a locally finite subset of $\mathbb{R}^{2d}$, because the box is given in terms of the positions only (and not the velocities) - e.g. infinitely many particles with the same position but different velocities, are, for us, not allowed.)

Definition 3.6 (infinite volume phase space)**.**

$$\Omega := \left\{\omega \,:\, \omega \subset \mathbb{R}^d \times \mathbb{R}^d \text{ is a locally finite configuration}\right\}.$$

For an $\omega \in \Omega$ and a finite box $\Lambda \subset \mathbb{R}^d$, the set of particles in Λ, which is $\omega \cap \Lambda$ can be naturally identified with an element of the grand canonical phase space Ω^Λ – with factorization due to non-distinguishability of particles automatically taken care of. Thus Ω can be equipped with a natural σ-algerba by simply requiring that sets of the form $\{\omega \in \Omega \,:\, \omega \cap \Lambda \in A\}$ be measurable whenever $A \subset \Omega^\Lambda$ is measurable. We set $\mathcal{F}$ to be the σ-algebra generated by such events.

4 Ising model

5 Basics of ergodic theory