Radboud University
Master Thesis
Mathematics

Irreversibility and Algorithmic Randomness

Nino Corak
August 11, 2025

Klaas Landsman (supervisor)
Eric Cator (second reader)

Abstract

Using a toy model as an illustrating case, Hiura and Sasa have shown that the trajectory generated by a Martin-Löf random microstate of a deterministic system, relative to a suitable probability measure, exhibits irreversibility when viewed macroscopically. We extend this result, which could prove useful in mathematically and conceptually justifying Boltzmann's ideas on randomness and the emergence of irreversibility, to systems with stochastic dynamics by illustrating it for the Ehrenfest urn model. The main difference with deterministic dynamics is that stochastic dynamics require the use probability measures on trajectories of microstates, rather than microstates themselves. In addition, we sketch a possible solution to the problem of non-random microstates satisfying macroscopic laws and discuss the Stosszahlansatz from the viewpoint of Martin-Löf randomness.

Contents

1 Introduction

The equations of motion governing microscopic particles, both in classical and quantum mechanics, are reversible in time, i.e., any evolution is just as physically plausible and possible when considered as occurring backward in time. In contrast, many processes on the macroscopic scale are irreversible, meaning their reversals are never observed. For example, friction causes a pendulum in motion to dissipate energy in the form of heat, thereby coming to a halt, but the reverse process of the pendulum spontaneously being set in motion by the energy contained in its surroundings never happens. If one believes in the unity of physics, in the sense that macroscopic laws should be derivable from microscopic ones, then this discrepancy is a problem, known as the reversibility paradox.

This paradox was first formulated in the 19th century within the context of kinetic theory, which is an attempt at deriving macroscopic properties of a gas from the motion of its constituent microscopic particles. Through the works of Maxwell and Boltzmann, it became clear that the paradox may be resolved by supplementing the microscopic mechanics of the gas by probability considerations. Essentially, while mechanics does allow for motions that do not satisfy the macroscopic laws, these are very unlikely when the number of particles N is large, hence are unobserved. Another great insight stemming from the use of probability theory is that macroscopic laws often take the form of a time-dependent law of large numbers when the number of particles becomes infinite. For deterministic systems, these laws look like

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_n(t) = F(t) \text{ for } \mathbb{P}\text{-almost all initial conditions } x \tag{1.1}$$

for some appropriate probability measure $\mathbb{P}$ and with $F(t)$ satisfying a differential equation if t is continuous and a difference equation if t is discrete. Here $X_n(t)$ is the state of the n'th particle at time t, which is a random variable, though only through its dependence on a random initial condition x. The quantity on the left-hand side of (1.1) is the macroscopic state of the system. For systems with stochastic dynamics, the entire trajectory becomes random and one uses a probability measure $\mathbb{P}$ on the space of trajectories rather than initial conditions.

The statement (1.1) does not tell us for which x the macroscopic law holds, only that it almost surely holds. Many theorems in mathematics in which measure theory is involved have the same problem: one knows that some property holds for almost all elements of a measure space, though not for which elements. The simplest example is the usual strong law of large numbers. To get a better understanding of the x for which the equality in (1.1) holds, [Hiura and Sasa, 2019] suggest using the theory of algorithmic randomness, which allows one to formulate a notion of $\mathbb{P}$-randomness and turn (1.1) into

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_n(t) = F(t) \text{ for all } \mathbb{P}\text{-random } x. \tag{1.2}$$

Intuitively, an element x is $\mathbb{P}$-random if it is generic with respect to all computable properties (hence the term 'algorithmic'), with genericity being measured by $\mathbb{P}$. Apart from the fact that one is now arguably closer to characterizing which x satisfy the equality in (1.1), the mathematical advantage of (1.2) over (1.1) is that the entire field of algorithmic randomness is now available as a tool for studying the emergence of macroscopic behaviour from microscopic laws, which is the main goal of statistical mechanics. Conceptually, it also does justice to Boltzmann's intuition that microscopic chaos

leads to irreversible macroscopic behaviour, by making the notion of chaos precise through algorithmic randomness.

The application of algorithmic randomness to statistical mechanics in the manner initiated by [Hiura and Sasa, 2019] is currently in its infancy, but [Landsman, 2023] believes that such a research program could lead to interesting developments and possibly solve and clarify long-standing problems in statistical mechanics, both mathematical and conceptual. Thus far, the idea has only been applied to the Kac ring model, a toy model with deterministic dynamics. In this thesis, we aim to make a small contribution to the program by applying it to a system with stochastic dynamics, specifically the Ehrenfest model. In addition, we will sketch a solution to the problem of non-random microstates satisfying macroscopic laws, which was mentioned in [Hiura and Sasa, 2019], and discuss the Stosszahlansatz from the viewpoint of algorithmic randomness, following a suggestion by [Landsman, 2023].

We begin by giving more details on kinetic theory in Chapter 2, so as to provide the reader with more context on the problem of deriving macroscopic laws from microscopic laws. In Chapter 3, the necessary basics of algorithmic randomness are presented. In Chapters 4 and 5 we apply algorithmic randomness to the two toy models, one with deterministic and one with stochastic dynamics, noting the differences between the two types of dynamics.

2 Statistical mechanics

The objective of statistical mechanics is to derive macroscopic behaviour from microscopic laws using probability theory. The earliest type of statistical mechanics, as well as its paradigmatic case, is kinetic theory, which seeks to derive macroscopic properties of gasses from the assumption that a gas consists of many particles behaving according to classical mechanics, usually only interacting by collisions. In this section, we will give a short description of kinetic theory, since the toy models of Chapters 4 and 5 are based on it. We also give a short overview of the conceptual explanation of irreversibility, known as the typicality account.

2.1 Kinetic theory

The philosophical roots of kinetic theory consist of ancient atomism and its 17th century mechanistic revival, which stated that all natural phenomena can be explained by small moving particles undergoing certain types of interactions such as collisions. This line of thinking remained mainly qualitative, the most notable exception being Daniel Bernoulli's microscopic derivation of Boyle's law. Kinetic theory could only flourish after it was acknowledged that heat is a form of motion, which was brought on by Joule and Mayer in the middle of the 19th century. Important contributions were made by Boltzmann, some of which we briefly present in this section. For more on the history of kinetic theory, see Chapters 3 and 4 in [Uffink, 2006] or [Brush, 1976] for even more.

2.1.1 Microscopic dynamics

The fundamental assumption of kinetic theory is that a gas forms a mechanical system, consisting of a large number N of particles, which we take to be identical for simplicity. If one takes these to be point particles, then the gas can be described using Hamiltonian mechanics. Its state is given by canonical coordinates $x = (\mathbf{q}_1, \ldots, \mathbf{q}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N)$, with $\mathbf{q}_i$ representing the position of the i'th point particle and $\mathbf{p}_i = m\mathbf{v}_i$ its momentum. Here m is the common mass of the particles. The set of all such states is called the phase space and denoted by Γ. The microscopic evolution of the gas is given by Hamilton's equations

$$\frac{d\mathbf{q}_i}{dt} = \frac{\partial H}{\partial \mathbf{p}_i} \quad \text{and} \quad \frac{d\mathbf{p}_i}{dt} = -\frac{\partial H}{\partial \mathbf{q}_i}. \tag{2.1}$$

The interactions between particles are encoded by the potential energy U appearing in the Hamiltonian $H = T + U$. Often, this potential is such that particles strongly repel at short distances and mildly attract at larger distances, which models particle collisions and intermolecular forces.

Instead of point particles, a gas can also be modelled as consisting of hard spheres, which only interact through elastic collisions. In that case, the Hamiltonian formalism becomes less practical to use, but remains the correct microscopic description in spirit. One speaks of an ideal gas if elastic collisions are the only kind of interaction between particles.

An important property of the equations (2.1) is that they are time-reversible in the following sense. Let $T : \Gamma \to \Gamma$ be the operation $(\mathbf{q}_1, \ldots, \mathbf{q}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N) \mapsto (\mathbf{q}_1, \ldots, \mathbf{q}_N, -\mathbf{p}_1, \ldots, -\mathbf{p}_N)$, which we call velocity reversal. Then, if the trajectory $x(t)$ through Γ satisfies (2.1), then so does $Tx(s-t)$ for any s. Consequently, if the initial state x evolves under (2.1) into x' during some time interval, then Tx' evolves into Tx during the same time interval.

2.1.2 Boltzmann equation

Without caring too much about rigour, let us now consider deriving macroscopic laws from the microscopic dynamics of a gas with N particles. The macroscopic quantity of interest is the distribution function

$$f(\mathbf{q},\mathbf{p}) = \frac{1}{N}\sum_{i=1}^{N} \delta^3(\mathbf{q}-\mathbf{q}_i)\delta^3(\mathbf{p}-\mathbf{p}_i), \tag{2.2}$$

which is a function of the state $x = (\mathbf{q}_1,\ldots,\mathbf{q}_N,\mathbf{p}_1,\ldots,\mathbf{p}_N)$ of the gas. It tells us how the positions and momenta of the particles are distributed in the state x, but not to which specific particles these belong. In this sense, $f(\mathbf{q},\mathbf{p})$ is a reduced description of the state. Actually, it would be more correct to call this a 'mesoscopic' quantity, since it is not really macroscopically observable. However, we will stick to the term 'macroscopic', essentially taking it to mean non-microscopic. If the number of particles N is large, $f(\mathbf{q},\mathbf{p})$ should become a continuous function such that

$$Nf(\mathbf{v})d^3\mathbf{q}d^3\mathbf{p} \approx \begin{array}{l}\text{number of particles with position in a volume } d^3\mathbf{q} \text{ around } \mathbf{q}\\ \text{and momentum in a volume } d^3\mathbf{p} \text{ around } \mathbf{p},\end{array}$$

though, like Boltzmann, we will not make this precise. To make things simpler, let us now switch to an ideal gas, for which the Hamiltonian description is less suitable, and assume that it is spatially homogeneous, so that we have the velocity distribution

$$f(\mathbf{v}) = \frac{1}{N}\sum_{i=1}^{N} \delta^3(\mathbf{v}-\mathbf{p}_i/m)$$

instead of the full distribution (2.2). One of the most important results in kinetic theory is that the particle velocities of an ideal gas in equilibrium are distributed according to the Maxwell-Boltzmann distribution

$$f(\mathbf{v}) = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} \exp\left(-\frac{m|\mathbf{v}|^2}{2k_B T}\right) \tag{2.3}$$

if $N\to\infty$, with T the temperature of the gas and k_B the Boltzmann constant. If the particle velocities of an isolated ideal gas are initially not distributed according to (2.3), the distribution function $f_t(\mathbf{v})$ becomes time-dependent and we should expect it to evolve in such a way that it attains the equilibrium (2.3). This evolution is supposed to be described by the Boltzmann equation, which is the prototypical macroscopic law in kinetic theory.

To derive his equation, Boltzmann considers collisions between two particles, modelled as spheres of diameter d, with incoming velocities $\mathbf{v}_1,\mathbf{v}_2$ and outgoing velocities $\mathbf{v}_1',\mathbf{v}_2'$. Following Section 3.3 in [Uffink, 2006], it is useful to set up cylindrical coordinates (b,φ,z) such that the first particle is resting at the origin and $\mathbf{v}_2-\mathbf{v}_1$ is aligned along the negative z-axis. The centre of the second particle then has coordinates given by $b(t)=b_0$, $\varphi(t)=\varphi_0$ and $z(t)=z_0-|\mathbf{v}_2-\mathbf{v}_1|t$ before the collision. The quantity b_0 is known as the impact parameter. A collision happens if and only if $b_0<d$. The new velocities $\mathbf{v}_1',\mathbf{v}_2'$ are uniquely determined as functions of the old ones $\mathbf{v}_1,\mathbf{v}_2$ and the two collision parameters φ_0,b_0.

If the number of particles is large, it is reasonable to assume that the number $N(\mathbf{v}_1,\mathbf{v}_2)$ of collisions of the type $\mathbf{v}_1,\mathbf{v}_2\to\mathbf{v}_1',\mathbf{v}_2'$ with collision parameters b,φ is proportional to relevant numbers of particles

$Nf_t(\mathbf{v}_1)d^3\mathbf{v}_1$, $Nf_t(\mathbf{v}_2)d^3\mathbf{v}_2$ and the volume $bdbd\varphi dz = |\mathbf{v}_2 - \mathbf{v}_1|bdbd\varphi dt$, in the space of collision parameters, so that

$$N(\mathbf{v}_1, \mathbf{v}_2) = N^2 f_t(\mathbf{v}_1) f_t(\mathbf{v}_2)|\mathbf{v}_2 - \mathbf{v}_1| d^3\mathbf{v}_1 d^3\mathbf{v}_2 bdbd\varphi dt. \tag{2.4}$$

Because the microscopic laws of motion governing gas particles are time-reversible, a collision with incoming velocities $\mathbf{v}_1', \mathbf{v}_2'$ should lead to outgoing velocities $\mathbf{v}_1, \mathbf{v}_2$, assuming the collision parameters are equal. This means that equation (2.4) should also hold with $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_1', \mathbf{v}_2'$ exchanged. In addition, conservation laws imply $|\mathbf{v}_2' - \mathbf{v}_1'| = |\mathbf{v}_2 - \mathbf{v}_1|$ and $d^3\mathbf{v}_1 d^3\mathbf{v}_2 = d^3\mathbf{v}_1' d^3\mathbf{v}_2'$, hence

$$N(\mathbf{v}_1', \mathbf{v}_2') = N^2 f_t(\mathbf{v}_1') f_t(\mathbf{v}_2')|\mathbf{v}_2 - \mathbf{v}_1| d^3\mathbf{v}_1 d^3\mathbf{v}_2 bdbd\varphi dt. \tag{2.5}$$

The change in the number of particles $Nf_t(\mathbf{v}_1)d^3\mathbf{v}_1$ due to only collisions of types $\mathbf{v}_1, \mathbf{v}_2 \to \mathbf{v}_1', \mathbf{v}_2'$ and $\mathbf{v}_1', \mathbf{v}_2' \to \mathbf{v}_1, \mathbf{v}_2$ with parameters b, φ is then given by the $N(\mathbf{v}_1', \mathbf{v}_2') - N(\mathbf{v}_1, \mathbf{v}_2)$. Assuming collisions between three or more particles are rare and can be neglected, we can integrate over $\mathbf{v}_2$ and b, φ to obtain the total change

$$Nf_{t+dt}(\mathbf{v}_1)d^3\mathbf{v}_1 - Nf_t(\mathbf{v}_1)d^3\mathbf{v}_1 \tag{2.6}$$

$$= N^2 \int_0^d bdb \int_0^{2\pi} d\varphi \int_{\mathbb{R}^3} |\mathbf{v}_2 - \mathbf{v}_1| \left(f_t(\mathbf{v}_1')f_t(\mathbf{v}_2') - f_t(\mathbf{v}_1)f_t(\mathbf{v}_2)\right) d^3\mathbf{v}_1 d^3\mathbf{v}_2 dt, \tag{2.7}$$

from which the Boltzmann equation

$$\frac{\partial f_t(\mathbf{v}_1)}{\partial t} = N \int_0^d bdb \int_0^{2\pi} d\varphi \int_{\mathbb{R}^3} |\mathbf{v}_2 - \mathbf{v}_1| \left(f_t(\mathbf{v}_1')f_t(\mathbf{v}_2') - f_t(\mathbf{v}_1)f_t(\mathbf{v}_2)\right) d^3\mathbf{v}_2 \tag{2.8}$$

follows upon dividing by N, $d^3\mathbf{v}_1$ and dt (recall that we make no pretences that this is a rigorous derivation). Recall that $\mathbf{v}_1', \mathbf{v}_2'$ are functions of $\mathbf{v}_1, \mathbf{v}_2$ and b, φ. Equation (2.8) is often called the homogeneous Boltzmann equation, since it is also possible to derive an equation for the full distribution function $f(\mathbf{q}, \mathbf{p})$. An important consequence of (2.8) is Boltzmann's H-theorem, which states that the quantity

$$H(f_t) = \int_{\mathbb{R}^3} f_t(\mathbf{v}) \log f_t(\mathbf{v}) d^3\mathbf{v} \tag{2.9}$$

always decreases except if f is of the form (2.3). Furthermore, $H(f) = 0$ if and only if f is of the form (2.3), which means the Maxwell-Boltzmann distribution is the global minimizer of H. Consequently, a gas whose particle velocities are not initially distributed according to (2.3) will evolve in such a way that they take on the distribution (2.3).

The derivation of the Boltzmann equation from microscopic motions is the prototypical example of a macroscopic law being derived from microscopic behaviour. It is specifically a law expressing the relaxation to equilibrium of an isolated system, which is also the kind of macroscopic law we will be interested in when looking at toy models in Chapters 4 and 5.

In the years after the Boltzmann equation was first published, controversies arose about its validity, in particular whether it really is a pure consequence of the microscopic dynamics of the gas. The two main objections brought forth were Loschmidt's Umkehreinwand and Zermelo's Wiederkehreinwand (see Sections 4.3 and 4.5 in [Uffink, 2006] for more details).

- The Umkehreinwand is based on the time-reversibility of microscopic dynamics. It follows from the considerations in Section 2.1.1 that there exist initial conditions such that the resulting evolution of the gas does not satisfy the Boltzmann equation. Simply take a non-equilibrium state that evolves into equilibrium (which presumably exist in the first place) and use the velocity reversed final state as new initial state. The resulting trajectory will be an evolution from equilibrium into non-equilibrium.
- The Wiederkehreinwand is based on a theorem proven by Poincaré (see Theorem 4.6 for a generalization of Poincaré's result), which states that a mechanical system that is spatially bounded and satisfies conservation of energy has the property that any trajectory will eventually return arbitrarily close to its initial state. This is known as recurrence and shows that the Boltzmann equation cannot hold for all times.

Both of these objections show that the Boltzmann equation cannot be derived solely from microscopic dynamics. Any true derivation necessarily needs to make use of non-mechanical assumptions. In the above derivation, the non-mechanical assumption is equation (2.4), which is known as the Stosszahlansatz, a term first introduced in [Ehrenfest and Ehrenfest, 1912]. This can be read as a probabilistic assumption: the velocity distributions of two particles are independent, hence one can take their product to calculate the probability of collision. Of course, making this precise, both conceptually and mathematically, is a more difficult matter. Boltzmann did make use of probability, but whether he believed it was only a tool for deriving his H-theorem or a necessary non-mechanical ingredient of the derivation remained unclear. In any case, the objections of Loschmidt and Zermelo forced Boltzmann to be more clear about his ideas on probability and move from a strict H-theorem to a statistical H-theorem: while $H(f_t)$ does not necessarily decrease when starting in non-equilibrium, there is a very high probability that it does, and upon reaching its minimum, it stays there for a very long but finite time.

2.2 Typicality

Despite Boltzmann's efforts at clarifying his ideas, his work did not lead to a completely satisfactory explanation of irreversibility. Consequently, physicists, mathematicians and philosophers explored alternative approaches to the problem of irreversibility in the 20th century (see Chapters 5 to 7 in [Uffink, 2006]). While many of these approaches were interesting and important, they did not lead to significantly more understanding. At the turn of the 21st century, articles written by a number of authors (see for example [Lebowitz, 1993], [Bricmont, 1995] and [Goldstein, 2001]) initiated a reappraisal of Boltzmann's ideas, especially those formulated in response to the objections of Loschmidt and Zermelo, known as the typicality account. We give a short overview of this modern formulation of Boltzmann's ideas. More details can be found in the original articles and in the more recent works [Lazarovici and Reichert, 2015] and [Bricmont, 2022].

The typicality account is a general conceptual explanation of irreversibility, specifically of relaxation to equilibrium. We present its ingredient in a list, illustrating each item the example of kinetic theory.

1. **Microstates.** Any system has a fundamental level of description, which we call its microstate x. The set of microstates is denoted by Ω. For a gas, the microstate is given by the canonical coordinates $(\mathbf{q}_1, \ldots, \mathbf{q}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N)$ and the set of microstates is the phase space Γ.
2. **Macrostates.** As already realised by Maxwell, irreversibility only occurs on macroscopic scales,

so that we need to switch to a reduced description of the system. This is accomplished by means of a coarse-graining map which assigns a macrostate $M(x)$ to each microstate x, in such a way that many microstates lie in the same macrostate. We denote the region in Ω corresponding to a macrostate M by Ω_M. For a gas, one maps a microstate $(\mathbf{q}_1, \ldots, \mathbf{q}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N)$ to its distribution function $f(\mathbf{q}, \mathbf{p})$.

3. **Probability.** By Loschmidt's Umkehreinwand, not all initial states lead to correct macroscopic behaviour. To account for this, one introduces a probability measure μ on Ω which should be invariant under the microscopic evolution. Then, relaxation to equilibrium is not necessary, but it is typical in the sense that it has high probability according to μ, from which the typicality account gets its name. For a gas, the invariant probability measure is given by the Liouville measure, which is invariant under the Hamiltonian dynamics (2.1) by Liouville's theorem (see [Goldstein et al., 2001]). Actually, this measure should be restricted to the hypersurface $\Gamma_E \subseteq \Gamma$ of microstates having energy E, to account for conservation of energy, which results in the microcanonical measure.

4. **Hierarchy.** The macrostates should be divided in a hierarchy of size as measured by μ. At the top of this hierarchy sits the equilibrium macrostate M_{eq}, which is characterized by being enormously larger than all other macrostates combined. This is known as the dominance of equilibrium. For a non-equilibrium macrostate M, the distance to equilibrium is measured by the size $\mu(\Omega_M)$. In kinetic theory, Boltzmann's famous combinatorial argument (see Section 4.4 in [Uffink, 2006]) essentially expresses the dominance of the Maxwell-Boltzmann distribution (2.3) for the ideal gas.

5. **Relaxation.** The explanation for relaxation to equilibrium is now as follows. If the system starts out in a non-equilibrium macrostate, then it is overwhelmingly likely, as measured by μ, that it will evolve into larger macrostates and eventually end up in the equilibrium macrostate. This depends crucially on the hierarchy of macrostates and the invariance of μ. To illustrate, let M be the initial macrostate and M' one which is further from equilibrium, i.e., $\mu(\Omega_{M'}) \ll \mu(\Omega_M)$. If we denote the evolution of x after time t by $\varphi_t(x)$ and define

$$\Omega_{M,t} = \{x \in \Omega_M \mid \varphi_t(x) \in \Omega_{M'}\},$$

then $\varphi_t(\Omega_{M,t}) \subseteq \Omega_{M'}$ and the invariance of μ together with the monotonicity of probability measures then imply

$$\mu(\Omega_{M,t}) = \mu(\varphi_t(\Omega_{M,t})) \leq \mu(\Omega_{M'}).$$

Dividing by $\mu(\Omega_M)$ turns this into

$$\frac{\mu(\Omega_{M,t})}{\mu(\Omega_M)} \leq \frac{\mu(\Omega_{M'})}{\mu(\Omega_M)} \ll 1,$$

which states that the probability of M evolving into M' at time t is very low. Intuitively, Ω_M does not fit into $\Omega_{M'}$. However, this does not exclude Ω_M from fitting into multiple smaller macrostates, in which case wrong macroscopic behaviour cannot be ruled out a priori. Actually proving that the microstates in Ω_M typically evolve towards equilibrium requires more detail, which is not available at this level of generality.

6. **Entropy.** If we define the Boltzmann entropy of a macrostate M by

$$S_B(M) = k_B \log \mu(\Omega_M) \tag{2.10}$$

and denote the evolution of the system's macrostate by $M(t)$, then the reasoning in the previous point implies that $S_B(M(t))$ very likely increases to the global maximum $S_B(M_{\text{eq}})$ if the system starts out in non-equilibrium. The quantity (2.10) can be related to the notion of entropy in thermodynamics, which was one of Boltzmann's great insights. However, we will not make any further use of this quantity.

Let us make some comments about this scheme.

- The reasoning of the typicality account, in particular the hierarchy of macrostates, is under the assumption that the number of particles N is large. In the thermodynamic limit $N \to \infty$, statements of high probability become statements of the form "property P holds almost surely".
- If we sample microstates using the invariant probability measure μ, then we will obtain equilibrium states with very high probability, due to dominance of equilibrium. In this sense, μ represents equilibrium.
- For many systems, the invariant probability measure μ is uniform in a certain sense, which is called the principle of equal a priori probabilities.
- Even though we illustrated the typicality account with deterministic dynamics, its reasoning is general enough to also work for stochastic dynamics, as we will illustrate in Chapter 5.

See [Lazarovici and Reichert, 2015] and Section 3.6 in [Frigg and Werndl, 2024] for critical discussions of the typicality account. The main critique brought against the typicality account is that it neglects the details of the microscopic dynamics: a microstate does not evolve into a region simply because that region is large. However, the typicality account does not pretend to give a mathematical proof that relaxation occurs. Rather, it is a conceptual explanation, and the details of a mathematical proof may differ case by case. Indeed, the typicality account can be seen as a reaction against the many futile attempts to find universal conditions on dynamics which lead to irreversible behaviour. The most famous of these is ergodicity, which already dates back to Boltzmann's early work (see Section 4.1 in [Uffink, 2006]). Intuitively, microscopic dynamics are ergodic if most initial states have trajectories which explore the entire state space, from which it follows that most trajectories end up staying the equilibrium macrostate for a long time. However, apart from simple systems, ergodicity is hard to prove and it is unclear whether physically relevant systems are ergodic at all (see Section 3.3 in [Frigg and Werndl, 2024]).

Another problem is that the coarse-graining map $x \mapsto M(x)$ and the invariant measure μ are in some sense arbitrary. However, even when there are multiple invariant measures, it is often the case that a natural one presents itself, such as the Liouville measure in the case of Hamiltonian mechanics. In many systems, the natural coarse-graining is given by a sum function, which is a sum over the microscopic constituents of the system. The distribution function (2.2) is an example. Sum functions do not exhaust all relevant macroscopic quantities, but form an important class since they often have dynamics given by a law of large numbers. The toy models we will look at in Chapters 4 and 5 are examples of systems for which this is the case.

Let us now return to the objections of Loschmidt and Zermelo and see how the typicality account deals with them. The Umkehreinwand is countered by the statement that relaxation to equilibrium, while not necessary, is overwhelmingly likely when starting out in non-equilibrium. The main point of the Wiederkehreinwand cannot be refuted: almost all initial states have recurrent trajectories. However, it can be proven that the expected recurrence time of a macrostate is inversely proportional

to its size according to the invariant measure μ (see Theorem 4.6), which means the recurrence of a non-equilibrium state takes an enormously large amount of time. Admittedly, this result depends on the microscopic dynamics being ergodic, but we still expect the Wiederkehreinwand to be practically irrelevant in many cases due to large recurrence times.

Mathematically, the Wiederkehreinwand is more significant, since it shows that rigorously derived irreversible macroscopic laws can only be valid on finite time intervals. This inconvenience can be circumvented by assuming the number of particles is infinite, i.e., taking a thermodynamic limit $N \to \infty$, since in that case the recurrence time becomes infinite. We will do this for the toy models in Chapters 4 and 5.

Let us make an important final note. The typicality account explains why a system which is initially in a non-equilibrium macrostate relaxes to equilibrium. Even though this can be safely described as irreversible behaviour, it does not provide an arrow of time, i.e., a way of distinguishing the two possible directions of time. The reason is that microscopic dynamics are still time-reversible, so that any conclusions which hold for one direction of time, also hold for the other. In particular, relaxation to equilibrium must also hold backward in time: if one observes a non-equilibrium state, then the theory predicts that it evolved from the equilibrium state in the past (see [Albert, 2000]). This is a major philosophical problem, which some try to solve by the postulate that the universe had an initial non-equilibrium state, known as the past hypothesis. However, this does not give a completely satisfying solution and the problem still remains a difficult point for the foundations of statistical mechanics (see [Earman, 2006]).

2.3 Ensembles

Other approaches to the problem of irreversibility focus on probability measures on the space of microstates, known as ensembles to physicists. These approaches stem from the work of Gibbs and one uses the term Gibbsian statistical mechanics to describe them collectively (see Chapter 5 in [Uffink, 2006]). In the case of Hamiltonian mechanics, a probability measure on the phase space Γ is given by a density $\rho(x)$ relative to the Liouville measure μ, i.e.,

$$R \mapsto \int_R \rho(x) d\mu(x)$$

for measurable regions $R \subseteq \Gamma$. It follows from Hamilton's equations (2.1) that the time evolution ρ_t of such a density function is given by the Liouville equation

$$\frac{\partial \rho_t}{\partial t} = \{H, \rho\} \tag{2.11}$$

(see [Goldstein et al., 2001]). The Gibbsian approach to statistical mechanics is much more systematic and useful in practice, especially when describing equilibrium states, but conceptually much less clear than the approach of Boltzmann. The main problem is the interpretation of the density function ρ_t and its behaviour given by (2.11). For example, even if one can prove that some sort of relaxation to equilibrium follows from (2.11), it is not immediately clear how that translates into actual systems relaxing to equilibrium. This in contrast to the approach of Boltzmann, which is interested in whether individual realisations of the system typically relax, instead of whether abstract probability distribution do so. A particularly radical form of Gibbsian statistical mechanics is that of Prigogine, which denies the existence of individual microscopic trajectories (see [Bricmont, 1995]).

Another conceptually problematic aspect of Gibbsian statistical mechanics is its notion of entropy, which is given by the Gibbs entropy

$$S_G(\rho) = -\int_\Gamma \rho(x) \log \rho(x) d\mu(x), \tag{2.12}$$

as a function of the phase space density ρ. It follows from Liouville's theorem that $S(\rho_t)$ is constant in time, unlike the Boltzmann entropy (2.10) which increases with high probability. This is problematic, since we expect entropy to increase during irreversible processes by the second law of thermodynamics. One can try to solve this by coarse-graining the phase space density ρ_t, but this introduces new problems (see Section 4.6 in [Frigg and Werndl, 2024]).

3 Algorithmic randomness

The notion of randomness is often intuitively understood as the lack of predictability and patterns. It is naturally related to the concept of probability, with the result of a probabilistic experiment, such as flipping a coin, being described as random. However, randomness does not exactly correspond to probability. For example, consider the binary sequences

$$10011001001101010101011111110111000101010000101010011011001100 11 \tag{3.1}$$

$$10101010101010101010101010101010101010101010101010101010101010 10 \tag{3.2}$$

which are possible outcomes of repeatedly flipping a coin. Their probabilities are both equal to $1/2^N$, with N the length of the sequences, but clearly the first would be described as more random than the second. The theory of algorithmic randomness is the attempt at formulating a precise definition of randomness that applies to intuitively random sequences like (3.1) but not to ordered sequences such as (3.2). It has been interwoven with probability theory since its inception in the work of Von Mises, who tried to base the foundations of probability theory on a precise notion of randomness. We do not intend to give a detailed exposition of algorithmic randomness and only a small part of the theory will be relevant to us. See [Dasgupta, 2011] for a general introduction and [Downey and Hirschfeldt, 2010], [Nies, 2008] and [Calude, 2002] for detailed expositions.

There are three approaches to formalizing the notion of randomness, corresponding to different intuitions:

- **Unpredictability.** We already mentioned this intuition in the first sentence of this section. One way of making this intuition more precise is through gambling. If one were to reveal the bits of a binary sequence one by one and bet on whether the next bit is a 0 or 1, then it would be very easy to make money if sequence were given by (3.2) rather than (3.1). This idea leads to formally defining the notion of a betting strategy and defining a binary sequence to be random if no betting strategy succeeds in making a large amount of money from that sequence, which is known as computable randomness.
- **Incompressibility.** Another sense in which (3.1) is more random than (3.2) is that the latter can be abbreviated by a simple rule, such as "print 10 thirty-two times", while the former seemingly has no shorter description than writing out the sequence itself. This intuition leads to the notion of Kolmogorov randomness, which essentially deems a sequence random if its shortest description, known as its Kolmogorov complexity, is not shorter than the sequence itself.
- **Typicality.** Even though the sequences (3.1) and (3.2) have equal probabilities, probability can still be used to describe why one is more random than the other. If a binary sequence is generated by flipping a coin, we expect 1 to be followed by 0 about half of the time. This is the case for (3.1), but not at all for (3.2). In this sense, the latter sequence is exceptional, while the former is typical. If a binary sequence is typical with respect to any statistical property, then it is deemed to be random. This is the essence of Martin-Löf randomness.

All three approaches require introducing the notion of computability, whence the term 'algorithmic'. This is most clear in the second approach: one needs to exactly define what a description of a binary sequence is. We will see in Section 3.1 why computability is needed for the typicality approach to randomness.

With randomness and probability being closely related to each other and the latter being fundamental to statistical mechanics, it is to be expected that randomness might have useful applications in statistical mechanics. In view of the typicality account, Martin-Löf randomness seems to be the most appropriate sense of randomness to use and we will describe possible ways of doing so in Section 3.3. This does not mean that the other two approaches might have no applications. On the contrary, Kolmogorov complexity has already seen more applications to physics than Martin-Löf randomness. See for example Chapter 8 in [Li and Vitányi, 2019] and Chapter 9 in [Calude, 2002]. Randomness through betting strategies has not yet been applied to problems in physics, but [Hiura and Sasa, 2019] expect it to also be useful.

3.1 Martin-Löf randomness

Even though we illustrated randomness for finite binary sequences, the rigorous definition of randomness in the sense of typicality is most clean when applied to infinite objects. The same holds for betting strategies, while Kolmogorov complexity only works for finite objects. Hence, we begin by considering Martin-Löf randomness of infinite binary sequences.

We first need to introduce some terminology and notation. The set of finite binary sequences of length N is denoted by 2^N and the set of infinite binary sequences by $2^{\mathbb{N}}$. The latter equipped with the product topology, using the discrete topology on each factor $\{0,1\}$, is known as the Cantor space. For $\sigma \in 2^N$, we define the cylinder set

$$\sigma 2^{\mathbb{N}} = \{x \in 2^{\mathbb{N}} \mid x(n) = \sigma(n) \text{ for } 0 \leq n \leq N-1\}, \tag{3.3}$$

which is the set of infinite binary sequences having initial segment σ (see also Appendix A). It is easily seen that each cylinder set is both open and closed. A general fact from topology regarding bases of product spaces implies that the collection of cylinder sets forms a basis for the topology of $2^{\mathbb{N}}$.

Intuitively, a sequence $x \in 2^{\mathbb{N}}$ is Martin-Löf random if it has no exceptional properties, as measured by some probability measure μ defined on the Borel σ-algebra of $2^{\mathbb{N}}$. This can be formalised in the following way. A subset $N \subseteq 2^{\mathbb{N}}$ is called a null set if there is a sequence $(U_n)_{n\in\mathbb{N}}$ of open sets such that $N \subseteq \bigcap_{n=0}^{\infty} U_n$ and $\mu(U_n) \leq 1/2^n$, which we interpret as N representing some exceptional property. Then, we might say that x is random if it is not contained in any null set. However, this proposed definition does not work, since it can imply that $\{x\}$ is a null set for any $x \in 2^{\mathbb{N}}$: each sequence is exceptional in virtue of being itself. For example, this happens for the Bernoulli measure $f^{\mathbb{N}}$ constructed from the fair probability measure f on $\{0,1\}$ satisfying $f(0) = 1/2$. Indeed, setting $U_n = x(0)\dots x(n)2^{\mathbb{N}}$, we have $x \in \bigcap_{n=0}^{\infty} U_n$ and $f^{\mathbb{N}}(U_n) = 1/2^n$.

The way out of this problem is to restrict to a smaller class of null sets. Since sequences that can be described by simple algorithms should be deemed non-random, such as the infinite version of (3.2), we restrict to effective null sets, i.e., those which can be described in an algorithmic manner. Making this precise requires the field of computability theory, which is too involved to present in detail here (see for example [Enderton, 2010]). The central objects of computability theory are computable functions, which are partial functions $f :\subseteq \mathbb{N} \to \mathbb{N}$, possible defined only on a subset $D_f \subseteq \mathbb{N}$, whose values $f(n)$ for $n \in D_f$ can be computed by some algorithm using n as input. The notion of an algorithm is formalised using a model of computation, such as the Turing machine. Given the notion of a computable function, computability can be defined for all kinds of other objects, such as subsets of $\mathbb{N}$ and real numbers. Instead of a precise definition, however, we will use the principle of "I know it when I see it" with regards to computability.

With the restriction to effective null sets, one obtains a satisfactory definition of randomness of infinite binary sequences, which was first given in [Martin-Löf, 1966]. However, we will give the definition of Martin-Löf randomness immediately for spaces more general than the Cantor space, following [Hertling and Weihrauch, 2003].

Definition 3.1. *An effective topological space is a pair (X, B) consisting of a topological space X, whose topology we denote by $\mathcal{O}(X)$, and an enumeration $B : \mathbb{N} \to \mathcal{O}(X)$ of a topological basis for X. An effective probability space (X, B, μ) consists of an effective topological space (X, B) together with a probability measure μ defined on the Borel σ-algebra of X.*

As an example, the probability space $(2^{\mathbb{N}}, f^{\mathbb{N}})$ representing an infinite number of coin tosses can be made effective by choosing some enumeration of the collection of cylinder sets (3.3). This is the effective probability space for which the original definition in [Martin-Löf, 1966] was given.

Definition 3.2. *An open set $U \subseteq X$ is computable if there exists a computable function $f : \mathbb{N} \to \mathbb{N}$ such that $U = \bigcup_{n\in\mathbb{N}} B(f(n))$. In words, it is possible to write U as a union of basis elements in an effective manner.*

Definition 3.3. *A sequence of open sets $(U_n)_{n\in\mathbb{N}}$ is uniformly computable if there exists some computable function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that $U_n = \bigcup_{m\in\mathbb{N}} B(g(n, m))$ for all $n \in \mathbb{N}$.*

Definition 3.4. *A Martin-Löf test is a uniformly computable sequence of open sets $(U_n)_{n\in\mathbb{N}}$ such that $\mu(U_n) \leq 1/2^n$ for each $n \in \mathbb{N}$. A subset $N \subseteq X$ is an effective null set if there exists a Martin-Löf test $(U_n)_{n\in\mathbb{N}}$ such that $A \subseteq \bigcap_{n=0}^{\infty} U_n$. An element $x \in X$ is Martin-Löf random if it is not contained in any effective null set.*

If we want to emphasize the dependence on the probability measure, we speak of 'μ-Martin-Löf randomness', or simply 'μ-randomness'. Note that this notion of randomness also depends on the choice of countable basis $B : \mathbb{N} \to \mathcal{O}(X)$. However, one can formulate a notion of equivalence between countable bases and show that equivalent countable bases give rise to equivalent notions of randomness (see [Hertling and Weihrauch, 2003]).

Proposition 3.5. *The set of Martin-Löf random elements of an effective probability space has measure one.*

Proof. Models of computation, such as Turing machines, are countable in nature, from which it follows that the number of computable functions is countably infinite. Consequently, there exists only a countably infinite number of Martin-Löf tests. For each test $(U_n)_{n\in\mathbb{N}}$, the intersection $\bigcap_{n=0}^{\infty} U_n$ is an effective null set. Each of these sets have measure zero, hence their union does too. Because all non-Martin-Löf random elements of X are contained in this union, the result follows. □

The following alternative definition of randomness can be seen as an effective version of the Borel-Cantelli lemma from probability theory (see Theorem 10.5 in [Jacod and Protter, 2004]). It is equivalent to Martin-Löf randomness and we will use it for applications of randomness to the toy models in Chapters 4 and 5.

Definition 3.6. *A computable sequence of open sets $(V_n)_{n\in\mathbb{N}}$ is a Solovay test if $\sum_{n=0}^{\infty} \mu(V_n) < \infty$. An element $x \in X$ is called Solovay random if for each Solovay test $(V_n)_{n\in\mathbb{N}}$ it is contained in only finitely many sets V_n.*

Proposition 3.7. *An element $x \in X$ of an effective probability space is Martin-Löf random if and only if it is Solovay random.*

Proof. See Theorem 6.2.8 in [Downey and Hirschfeldt, 2010]. □

Let us make some general comments on the notion of Martin-Löf randomness.

- Variants of Martin-Löf randomness can be obtained by modifying the class of null sets (see Sections 3.5 and 3.6 in [Nies, 2008]). For example, one obtains Schnorr randomness by demanding that the sequence of real numbers $(\mu(U_n))_{n\in\mathbb{N}}$ in Definition 3.4 is uniformly computable. One reason for preferring Martin-Löf randomness over its variants is because it is closely related to Kolmogorov randomness, in the sense that an infinite binary sequence is Martin-Löf random if and only if all its initial segments are Kolmogorov random (see Theorem 6.2.3 in [Downey and Hirschfeldt, 2010]). It is also equivalent to computable randomness (see Chapter 7 in [Nies, 2008]).
- If $x \in 2^{\mathbb{N}}$ is a computable sequence, then the sets $U_n = x(0)\dots x(n)2^{\mathbb{N}}$ from before are uniformly computable and hence form a Martin-Löf test, so that x is non-random. This fact implies that we got more out of our randomness definition than we may have wanted: not only are sequences given by simple rules non-random, for example the infinite version of (3.2), but sequences given by any computable rule whatsoever are also non-random. For example, the sequence

  $$10110101000001001111001100110011111110... \tag{3.4}$$

  may seem random, but it actually lists the digits in the binary expansion of $1/\sqrt{2}$, which means it is computable and hence non-random. Sequences like (3.4), which seem random but are given by a non-obvious rule, are called pseudorandom. This concept is relevant for statistical mechanics, since deterministic systems can exhibit chaotic behaviour which seems random but is actually pseudorandom.
- The fact that computable sequences are non-random means that Martin-Löf random sequences are inaccessible to us, in the sense that we cannot list the terms of a sequence which is Martin-Löf random (see Chapter 8 in [Calude, 2002]). This might seem worrying for applications of Martin-Löf randomness to physics, but one should always remember that mathematical models are idealizations anyway, especially if infinity is involved. However, there exist similar theorems for Kolmogorov randomness, which means that even finite random objects are inaccessible and mysterious in a certain sense.

3.2 Effective theorems

Whenever measure theory is involved in mathematics, one often comes across theorems of the form

$$\text{property } P \text{ holds for } \mu\text{-almost all } x \in X, \tag{3.5}$$

with (X, μ) some measure space, for us always a probability space. Such theorems state that typical elements of X satisfy P, but do not characterize these elements, which might be seen as a drawback. Since Martin-Löf randomness is a refined notion of typicality, especially in view of Proposition

3.5, we might hope to get a characterization of such elements by turning (3.5) into the randomness statement

$$\text{property } P \text{ holds for all } \mu\text{-Martin-Löf random } x \in X. \tag{3.6}$$

Whether this can be done and how depends on the specifics of the property P. If (3.5) can be proven using the Borel-Cantelli lemma and there are no problems regarding computability, then Proposition 3.7 can be used to prove (3.6). For example, this technique works for the strong law of large numbers, which in the simplest case of infinitely many coin tosses states that 1 almost surely appears about half of the time in the resulting infinite binary sequence, i.e.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} x(n) = \frac{1}{2} \text{ for } f^{\mathbb{N}}\text{-almost all } x \in 2^{\mathbb{N}}. \tag{3.7}$$

To prove this using the Borel-Cantelli lemma, define the exception sets

$$E_{N,m} = \left\{ x \in 2^{\mathbb{N}} \mid \left| \frac{1}{N} \sum_{n=0}^{N-1} x(n) - \frac{1}{2} \right| > \frac{1}{m} \right\} \tag{3.8}$$

and use the fourth-moment method (see Appendix B) to derive the bound $f^{\mathbb{N}}(E_{N,m}) \leq Cm^4/N^2$ for some constant $C \geq 0$, which is possible since the probability measure $f^{\mathbb{N}}$ makes the random variables $x(n)$ independent by construction. The Borel-Cantelli lemma then implies that for $f^{\mathbb{N}}$-almost all $x \in 2^{\mathbb{N}}$ there exists an N_0, which may depend on x, such that $x \notin E_{N,m}$ for all $N \geq N_0$, i.e.,

$$f^{\mathbb{N}}\left(\bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} E_{N,m}^c \right) = 1$$

for all $m \geq 1$. It then follows from elementary probability theory that

$$f^{\mathbb{N}}\left(\bigcap_{m=1}^{\infty} \bigcup_{N_0=1}^{\infty} \bigcap_{N=N_0}^{\infty} E_{N,m}^c \right) = 1,$$

which is equivalent to (3.7). To turn this into a randomness statement, note that the sequence $(E_{N,m})_{N\geq 1}$ of open sets is uniformly computable, since there obviously exists an algorithm to express (3.8) as a union of cylinder sets. Because (3.8) only depends on the first N terms of x, only finitely many cylinder sets need to be checked. Together with $f^{\mathbb{N}}(E_{N,m}) \leq Cm^4/N^2$, it follows that $(E_{N,m})_{N\geq 1}$ forms a Solovay test and Proposition 3.7 then implies that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} x(n) = \frac{1}{2} \text{ for all } f^{\mathbb{N}}\text{-Martin-Löf random } x \in 2^{\mathbb{N}}. \tag{3.9}$$

All random infinite binary sequences satisfy the law of large numbers, which is an improvement over (3.7), but this would only be a full characterization if the converse were also true. This is not the case, since the infinite version of (3.2) satisfies the law of large numbers, but is not random because it is given by an algorithm. The problem is that random sequences satisfy more statistical properties than just the law of large numbers. For example, the finite sequence 10 should occur a fourth of the time, which is not the case for (3.2). More generally, the frequency of any $\sigma \in 2^N$ should be

$1/2^N$. A sequence in which all frequencies are as expected is called Borel normal. However, even all these statistical properties together are not equivalent to Martin-Löf randomness, since the binary Champernowne sequence

$$0100011011000001010011100101110111...$$

is Borel normal. This sequence simply lists all binary strings in lexicographic order, which means it is given by an algorithm and hence not random. A further generalization of Borel normality is given by (a special case of) the ergodic theorem.

Theorem 3.8. *If $\varphi : S \to S$ is ergodic with respect to an invariant probability measure μ (see Section 4.1 for the definitions of these terms) and $A \subseteq S$ is measurable, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\varphi^k(x)}(A) = \mu(A) \tag{3.10}$$

for μ-almost all $x \in S$.

Proof. If $A \subseteq S$ is measurable, then $1_A \in L^1(S, \mu)$ and (3.10) follow from Theorem 4.5. □

Here δ_x is the point measure on S concentrated at x. Note that the sum on the left-hand side of (3.10) counts the number of times the trajectory $(\varphi^k(x))_{k\geq 0}$ lies in A. Theorem 3.9 states that in the long run, this trajectory spends a fraction $\mu(A)$ of its time in A. Restricting to the probability space $(2^{\mathbb{N}}, f^{\mathbb{N}})$ and the shift map $(Lx)(n) = x(n + 1)$, which leaves $f^{\mathbb{N}}$ invariant and is ergodic, we recover Borel normality as a special case by applying (3.10) to cylinder sets.

Theorem 3.8 is of the form (3.5) and we may expect that an effective version of the form (3.6) exists. There indeed does, if we make all objects computable. This effective ergodic theorem gives a characterization of μ-Martin-Löf randomness.

Theorem 3.9. *If $\varphi : S \to S$ is ergodic with respect to an invariant probability measure μ, and φ and μ are computable (see [Galatolo et al., 2010] for an explanation of what this means), then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\varphi^k(x)}(U) = \mu(U) \tag{3.11}$$

for all computable open $U \subseteq S$ if $x \in S$ is μ-Martin-Löf random. Conversely, if (3.11) *holds for all computable open $U \subseteq S$, then x is μ-Martin-Löf random.*

Proof. See Theorem 3.2.2 in [Galatolo et al., 2010] and Theorem 1.3 in [Pathak et al., 2014]. □

Theorem 3.9 states that the collection of computable properties (3.11) which x could satisfy, ranging over computable open subsets $U \subseteq S$, is large enough to yield a characterization of μ-Martin-Löf randomness. We now recognize the effective law of large numbers (3.9) as only one of these properties for the effective probability space $(2^{\mathbb{N}}, f^{\mathbb{N}})$.

3.3 Statistical mechanics

Now we describe the applications of algorithmic randomness to statistical mechanics envisioned by [Hiura and Sasa, 2019] and [Landsman, 2023]. When worked out rigorously in the thermodynamic limit $N \to \infty$, the typicality account amounts to a statement of the form (3.5), stating that the microstates of a given non-equilibrium macrostate typically lead to irreversible macroscopic behaviour, according to some probability measure. But which microstates actually have this property and what characterizes them? Following Section 3.2, we could attempt to answer these questions using Martin-Löf randomness.

First, we would like to prove that Martin-Löf random microstates lead to irreversible macroscopic behaviour, giving a precise class of microstates with correct macroscopic behaviour. Then one can study the properties of these microstates. In particular, do they satisfy the analogue of Boltzmann's Stosszahlansatz (2.4)? Recall that this is the infamous and crucial assumption in deriving the Boltzmann equation and a positive answer to the previous question would reveal that it actually follows from a more fundamental assumption of randomness. Finally, the converse question can be asked: are microstates with correct macroscopic behaviour Martin-Löf random? If not, then Martin-Löf randomness may be too strong to fully describe macroscopic irreversibility, in which case one might want to look at different notions of randomness.

To qualitatively illustrate how randomness of a microstate might be tied to irreversible macroscopic behaviour, consider the following two possible microstates of an ideal gas in a cylindrical container.

1. The particles are distributed homogeneously through the cylinder with velocities pointing in random directions.
2. The particles lie in a common circular cross section and all velocities are equal and aligned with the axis of the cylindrical container.

It is clear that the second microstate results in an evolution of the gas in which the particles periodically bounce off the two ends of the cylinder, which is not the expected irreversible behaviour. The first microstate intuitively relaxes to equilibrium, if it is not already there. The second microstate would clearly be described as non-random, being analogous to the sequence (3.2), while the first would be described as random and is more like (3.1). Hence we expect random microstates to give rise to irreversible macroscopic behaviour.

Once the relation between Martin-Löf randomness and irreversible macroscopic behaviour has been explored, one would like to do the same for systems with a finite number of particles, using a different notion of randomness appropriate to finite objects. We expect this to be more difficult, since randomness is most clean to work with in the case of infinite objects.

4 Deterministic dynamics

Even though it is quite important in physics, the Boltzmann equation has to this day not been made completely mathematically rigorous in a satisfactory manner. It is the hope of [Landsman, 2023] that algorithmic randomness may help in accomplishing this. Instead of immediately tackling the Boltzmann equation, it is better to begin with toy models of statistical mechanics. Since the equations of motion of classical mechanics are deterministic, it is natural to begin with deterministic dynamics. Specifically, we will look at the Kac ring model, which was introduced in [Kac, 1959] to illustrate Boltzmann's kinetic theory in a simplified setting. Other useful references for the model are [Gottwald and Oliver, 2009] and [Maes et al., 2009]. First we will need to introduce discrete-time dynamical systems in general.

4.1 Dynamical systems

The microscopic dynamics of a physical system are usually given in terms of differential equations, such as Hamilton's equations (2.1). However, an alternative description is more useful in statistical mechanics. Assuming some regularity of the Hamiltonian H, the existence and uniqueness theorem for ordinary differential equations implies that each microstate $x \in \Gamma$ determines a unique solution of (2.1) starting at x, which we denote by $t \mapsto \varphi_t(x)$. In this way, the differential equations (2.1) determine a collection of maps $\varphi_t : \Gamma \to \Gamma$, which obviously have the property $\varphi_s(\varphi_t(x)) = \varphi_{s+t}(x)$. Conversely, if we neglects some details, any collection of sufficiently regular maps $\varphi_t : \Gamma \to \Gamma$ with this property determines a system of differential equations, though not necessarily of Hamiltonian form. This leads to the following definition.

Definition 4.1. *A continuous-time dynamical system with state space S is a collection of maps $\varphi_t : S \to S$ parametrized by $t \geq 0$ such that $\varphi_{t+s} = \varphi_t \circ \varphi_s$.*

Since time is often assumed to be continuous in physics, it is most natural to use dynamical systems in continuous-time for studying statistical mechanics. However, the Kac ring has a discrete time variable, and discrete-time dynamical systems in general are mathematically much simpler, especially when restricting to finite state spaces.

Definition 4.2. *A discrete-time dynamical system with state space S is a collection of maps $\varphi_k : S \to S$ parametrized by $k \in \mathbb{N}$ such that $\varphi_{k+m} = \varphi_k \circ \varphi_m$.*

A discrete-time dynamical system satisfies the recursion relation $\varphi_{k+1} = \varphi \circ \varphi_k$, which means it is completely determined by the single function φ_1 through $\varphi_k = \varphi_1^k$. From now on, we restrict to discrete-time dynamical systems, which we interpret as maps $\varphi : S \to S$.

In the definitions one comes across in the literature, a dynamical system is usually supposed to include a probability measure π on S which is invariant under the dynamics. To us, it makes more sense to separate dynamics from invariant measures, which may or may not exist and be unique.

Definition 4.3. *A probability measure π on S is invariant under the dynamics $\varphi : S \to S$ if $\pi(\varphi^{-1}(A)) = \pi(A)$ for all measurable subsets $A \subseteq S$.*

As was mentioned in Chapter 2, the notion ergodicity was used in the 20th century as an attempt at giving a mathematical underpinning of irreversibility in physics. We can now give a precise definition of the concept.

Definition 4.4. *A dynamical system $\varphi : S \to S$ is ergodic with respect to a given probability measure μ if $\varphi^{-1}(A) = A$ implies $\mu(A) \in \{0, 1\}$ for any measurable $A \subseteq S$, i.e., S cannot be decomposed into two φ-invariant subsets with positive measure.*

Most definitions in the literature would include the condition that μ is invariant under φ. Ergodicity is actually only one of multiple properties forming the 'ergodic hierarchy', all of which try to mathematically capture the notion of irreversibility. To give an example of another property, a dynamical system $\varphi : S \to S$ is said to be metrically transitive with respect to μ if for any measurable $A, B \subseteq S$ such that $\mu(A) > 0$ and $\mu(B) > 0$, there exists a $k \geq 0$ such that $\mu(\varphi^{-k}(A) \cap B) > 0$. The study of such ergodic properties and their consequences is the field of ergodic theory (see [Walters, 1982]).

The central result of ergodic theory is the ergodic theorem, which states that for almost all trajectories of an ergodic dynamical system, the time average of an integrable function over the trajectory is equal to the spatial average of the function over the entire space. This expresses the intuition that most trajectories of an ergodic dynamical system fill up the entire state space.

Theorem 4.5. *If $\varphi : S \to S$ is ergodic with respect to an invariant probability measure μ and $f \in L^1(S, \mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} f(\varphi^k(x)) = \int_S f d\mu$$

for μ-almost all $x \in S$.

Proof. See Section 1.6 in [Walters, 1982]. □

The Poincaré recurrence theorem, which we mentioned in Chapter 2 within the context of kinetic theory, holds for more general dynamical systems. We also stated that the recurrence time of a subset can be related to its measure, assuming the dynamics are ergodic. This can now be made precise, though we will not go into the details of the proof.

Theorem 4.6. *Suppose $\varphi : S \to S$ is a dynamical system with invariant probability measure π. If a measurable $A \subseteq S$ satisfies $\pi(A) > 0$, then almost all $x \in A$ have the property that $\varphi^k(x) \in A$ for infinitely many $k \geq 0$. In addition, if φ is ergodic with respect to π, then the mean recurrence time*

$$\sum_{k=1}^{\infty} k \frac{\pi(A_k)}{\pi(A)}$$

is equal to $1/\pi(A)$. Here $A_k \subseteq A$ is the set of $x \in A$ which first return to A at time k.

Proof. See Sections III.4 and III.5 in [Kac, 1959]. □

For a dynamical system $\varphi : S \to S$ with finite state space, the notions of ergodicity and metrical transitivity coincide and are equivalent to φ having only one orbit: for any $x, y \in S$, there exists a $k \geq 0$ such that $\varphi^k(x) = y$. This orbit is periodic, i.e., there is some smallest $N \geq 1$ such that $\varphi^N(x) = x$ for all $x \in S$, so that the dynamics is recurrent. More generally, any dynamical system with finite state space can be partitioned into periodic orbits. The dynamics is still recurrent, but the recurrence time is now given by the least common multiple of all periods.

Finally, let us discuss the reversibility of dynamical systems. First, note that non-invertible dynamical systems cannot be said to be reversible, since those cannot be run backward in time. However, an invertible dynamical system is not necessarily reversible. Only if the behaviour forward and backward in time is essentially the same can we speak of reversibility.

Definition 4.7. *A dynamical system $\varphi : S \to S$ is reversible if it is invertible and there exists a non-zero involution $T : S \to S$ such that $\varphi^{-1} = T \circ \varphi \circ T$.*

The involution $T : S \to S$ is analogous to the operation of velocity reversal in kinetic theory (see Section 2.1.1) and we us the same name for it. If $\varphi : S \to S$ is invertible, then it is automatically reversible in case S is finite. This is because a finite dynamical system can be partitioned into periodic orbits, on each of which a velocity reversal can be defined. For example, if φ has only a single orbit $x, \varphi(x), \ldots, \varphi^{N-1}(x)$, then the velocity reversal is given by the involution

$$\varphi^k(x) \mapsto \varphi^{2N-k}(x).$$

A dynamical system can only be truly non-reversible according to Definition 4.7 if the state space is infinite, because then the orbits are not forced to be periodic. For example, the dynamical system $\varphi(x) = 2x$ on $\mathbb{R}$ is non-reversible. A corresponding velocity reversal T would have to satisfy the functional equation $2T(2x) = T(x)$, from which it follows that $T = 0$ under the assumption that T is continuous, which is a reasonable assumption to make.

4.2 Kac ring

Rather than directly writing down the dynamics of the Kac ring in a formal manner, it is better to first give an intuitive description of the model. To begin, there is a ring with N sites and at each site a ball that is either black or white. At each time step, every ball moves one site over in the counterclockwise direction. Some of the sites are marked, and if a ball departs from a marked site, it changes colour.

If we let the colours black and white correspond to 0 and 1, then the configuration of balls is equivalent to a finite binary sequence $x \in 2^N$. Similarly, if 0 corresponds to the absence of a mark and 1 to the presence, then the configuration of the sites is equivalent to $y \in 2^N$. Hence, we may use $S = 2^N \times 2^N$ as the state space of the Kac ring. Using the notation $(x, y)(i) = (x(i), y(i))$, the dynamics $\varphi : S \to S$ can be expressed as

$$\varphi(x, y)(i) = (x(i-1) + y(i-1)(1 - 2x(i-1)), y(i)). \tag{4.1}$$

The subtractions occurring in the arguments of x, y in (4.1) are modulo N. If we change the first factor 2^N into $\{-1, 1\}^N$, then we get a more physical interpretation of the Kac ring in terms of spins moving along a ring and being flipped by scatterers which may be present at some sites, but we will stick to coloured balls. The dynamics (4.1) are invertible with inverse

$$\varphi^{-1}(x, y)(i) = (x(i+1) + y(i)(1 - 2x(i+1)), y(i)), \tag{4.2}$$

and reversible when using the velocity reversal $T : S \to S$ defined by

$$T(x, y)(i) = (x(-i), y(-i-1)), \tag{4.3}$$

as is easily checked (Umkehreinwand). Again, the arguments of x, y are modulo N. The Kac ring is also recurrent in the sense that each state evolves back into itself after at most $2N$ time steps

(Wiederkehreinwand). In spite of this, the Kac ring exhibits irreversibility when viewed macroscopically. As pointed out by [Bricmont, 1995], this is noteworthy, since the Kac ring lacks properties which are often taken to be explanations of irreversibility, such as ergodicity. Indeed, the Kac ring cannot be ergodic since it obviously has multiple orbits.

Let us now walk through the steps of the typicality account to explain how irreversibility occurs in the Kac ring. We have already seen that the microstates form the set $2^N \times 2^N$, analogous to the phase space of a gas. The natural coarse-graining of this space is given by the two macroscopic quantities

$$f(x,y) = \frac{1}{N}\sum_{i=1}^{N} x(i) \quad \text{and} \quad g(x,y) = \frac{1}{N}\sum_{i=1}^{N} y(i), \tag{4.4}$$

which are the proportions of white balls and marked sites, analogous to Boltzmann's distribution function (2.2). Note that the second is non-dynamical, but obviously required to describe the dynamics of the first. This means that one should use only f in the typicality account, but conditioned on a fixed value of g. The macrostates corresponding to f should have differing sizes as measured by a probability measure π on S which is invariant under φ. Because S is finite, a probability measure on S is equivalent to its probability mass function and we use the same symbol for both. In addition, φ is invertible, which means π is invariant under φ if and only if

$$\pi(\varphi(x,y)) = \pi(x,y)$$

for all $(x,y) \in S$. It follows that π is constant on orbits of φ. Conversely, any probability measure on the set of orbits gives rise to an invariant probability measure on S. Because the Kac ring has multiple orbits, there is no unique invariant probability measure. However, in view of the principle of equal a priori probabilities (see Section 2.2), it seems natural to use the uniform probability measure

$$\pi(x,y) = \frac{1}{2^{2N}}. \tag{4.5}$$

Using π to measure sizes of subsets of S, the size of the region corresponding to the macrostate n/N is given by

$$\pi\left(\{f = n/N\}\right) = \frac{1}{2^N}\binom{N}{n}.$$

This shows that the equilibrium macrostate is given by $f \approx 1/2$, since it has the largest size. Also, the further n/N is from $1/2$, the smaller the corresponding region of the microstate space is, as should be the case for the typicality account to hold. The dominance of equilibrium is actually an asymptotic statement, which can be made precise as follows. The probability measure π makes the projections $x(i) : S \to \{0,1\}$ independent random variables with expectation $1/2$, hence we may use Hoeffding's inequality (see Appendix B) to conclude

$$\pi(\{|f - 1/2| < \varepsilon\}) \geq 1 - 2\exp(-2N\varepsilon^2),$$

which is approximately equal to 1 for large N. The buffer ε is necessary because the equilibrium macrostate would otherwise actually be smaller than all other macrostates combined (see Section 6.2 in [Bricmont, 2022]). Of course, this buffer is no problem when N is very large, as is the case for physical systems. Recall that in all of the previous steps, one should actually condition on a fixed value of g. Since f and g can be independently varied, it is clear that this does not affect the conclusions in an essential way.

By the typicality account, we are now justified in expecting that the Kac ring has a high probability of relaxing to equilibrium when starting in a non-equilibrium macrostate. Of course, as we made clear in Section 2.2, this is only an expectation. We can only be sure that relaxation occurs by giving a mathematical proof, which we will do in Section 4.3.

4.3 Macroscopic behaviour

4.3.1 Kac ring

First we non-rigorously derive the macroscopic law governing (4.4), after which we can prove that it is satisfied with high probability. We denote the time-evolution of the macroscopic quantity f by $f_k = f \circ \varphi^k$. We make the simplifying assumption that

$$\text{proportion of white balls at marked sites} = gf_k \tag{4.6}$$

and similarly

$$\text{proportion of black balls at marked sites} = g(1 - f_k) \tag{4.7}$$

hold for all relevant times k, which is the analogue of the Stosszahlansatz (2.4) for the Kac ring, stated in terms of proportions rather than numbers. One can easily come up with examples of microstates for which (4.6) and (4.7) do not hold, but we nevertheless expect them to typically hold with respect to a proper probability measure. The change in the proportion of white balls during the time step from k to $k+1$ is equal to the difference between (4.7) and (4.6), which results in the Kac ring's analogue of the Boltzmann equation:

$$f_{k+1} - f_k = g(1 - 2f_k). \tag{4.8}$$

Using standard techniques for solving difference equations, the solution of (4.8) with initial conditions $f = \alpha$ and $g = \beta$ is given by

$$f_k = \frac{1}{2} + \left(\alpha - \frac{1}{2}\right)(1 - 2\beta)^k \tag{4.9}$$

and exhibits relaxation to the equilibrium value of $1/2$ if $\beta \in (0, 1)$, monotonically if $\beta < 1/2$. The question is now whether the evolution equation (4.8) can also be obtained rigorously. As shown in Section III.14 of [Kac, 1959], one way of doing this is as follows. With respect to the probability measure π conditioned on fixed initial values $f = \alpha$ and $g = \beta$, the expectation $\mathbb{E}(f_k)$ becomes equal to (4.9) for fixed k upon taking a limit $N \to \infty$.

There is another method of rigorously obtaining (4.8) which is much easier and more elegant than the previous. Instead of fixing initial values $f = \alpha$ and $g = \beta$, one can use a probability measure μ in which they are allowed to vary, but still satisfy

$$\mathbb{E}(f) = \alpha \text{ and } \mathbb{E}(g) = \beta. \tag{4.10}$$

The best choice is the product measure $\mu = \mu_\alpha \times \mu_\beta$, with μ_p the Bernoulli measure on 2^N of parameter $p \in [0, 1]$. This makes all the coordinates $x(i), y(j)$ independent Bernoulli random variables such that

$$\mathbb{E}(x(i)) = \alpha \text{ and } \mathbb{E}(y(j)) = \beta.$$

In physics, this trick is known as switching from the microcanonical ensemble to the canonical ensemble. For the Kac ring, the former corresponds to π conditioned on $f = \alpha$ and $g = \beta$ while the latter

corresponds to μ. It is not the case that μ is wholly unrelated to π. One can show that μ minimizes the relative entropy $H(\mu|\pi)$ (see Definition 5.12) subject to the constraints (4.10), which is an instance of the maximum entropy principle (see Chapter 2 in [Mackey, 1992]). The quantity $H(\mu|\pi)$ is actually a negative entropy, hence is minimized instead of maximized. The following Proposition now makes (4.8) rigorous at the level of expectation using μ rather than π.

Proposition 4.8. *With respect to the probability measure μ, the expectation $F_k = \mathbb{E}(f_k)$ satisfies the difference equation $F_{k+1} - F_k = \beta(1-2F_k)$ with initial value $F_0 = \alpha$ for all $k < N$, and is consequently given by*

$$F_k = \frac{1}{2} + \left(\alpha - \frac{1}{2}\right)(1-2\beta)^k \tag{4.11}$$

for $k \leq N$. The finite time interval of validity is necessary in view of the Kac ring being recurrent. In fact, one can show that F_k evolves towards non-equilibrium for $N \leq k \leq 2N$ (see [Gottwald and Oliver, 2009]).

Note that both the canonical measure μ and the macroscopic law F_k actually depend on α and β, but we suppress this dependence to ease the notation.

Proof. The proof is most clean using the alternative variables $\eta, \varepsilon \in \{-1, 1\}^N$ defined by

$$\begin{cases} \eta(i) = 2x(i) - 1 \\ \varepsilon(i) = -2y(i) + 1. \end{cases} \tag{4.12}$$

With these, the dynamics (4.1) takes on the simple form $\varphi(\eta, \varepsilon)(i) = (\eta(i-1)\varepsilon(i-1), \varepsilon(i))$, from which the useful formula $\varphi^k(\eta, \varepsilon) = (\eta(i-k)\varepsilon(i-1)\cdots\varepsilon(i-k), \varepsilon(i))$ follows. Using this formula and

$$2f - 1 = \frac{1}{N}\sum_{i=1}^{N} \eta(i), \tag{4.13}$$

we have

$$2\mathbb{E}(f_{k+1}) - 1 = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} \eta(i-(k+1))\varepsilon(i-1)\cdots\varepsilon(i-(k+1))\right).$$

If $k < N$, the factors appearing in the summand are all independent. Using standard properties of the expectation, we can then bring a factor $\mathbb{E}(\varepsilon(i-1))$ out in front and use $\mathbb{E}(\varepsilon(j)) = 1 - 2\beta$ to get

$$2\mathbb{E}(f_{k+1}) - 1 = (1-2\beta)\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} \eta(i-(k+1))\varepsilon(i-2)\cdots\varepsilon(i-(k+1))\right).$$

The sum clearly does not change under the shift $i \mapsto i + v$, hence

$$\begin{aligned} 2\mathbb{E}(f_{k+1}) - 1 &= (1-2\beta)\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} \eta(i-k)\varepsilon(i-1)\cdots\varepsilon(i-k)\right) \\ &= (1-2\beta)(2\mathbb{E}(f_k) - 1) \end{aligned}$$

and a simple rearrangement of this equality finally yields $F_{k+1} - F_k = \beta(1-2F_k)$. The initial condition $F_0 = \alpha$ holds by (4.10). □

The evolution equation (4.8) has now been obtained rigorously, but only at the level of expectations. To actually prove the Kac ring exhibits irreversibility, it needs to be shown that f_k remains near to F_k on some finite time interval with high probability. This can be accomplished using π (see Section 10 in [Maes et al., 2009]), but it is once again mathematically more convenient to use μ.

Theorem 4.9. *Defining F_k by (4.11), for any $T \geq 0$ and $\varepsilon > 0$ we have*

$$\mu\left(\bigcap_{k=0}^{T} |f_k - F_k| < \varepsilon\right) \leq 1 - \frac{T^2}{4N\varepsilon^2}$$

if $T \leq N/2$. This means that for sufficiently large numbers of balls N, the Kac ring satisfies its macroscopic law on the time interval $0 \leq k \leq T$ with high probability.

Proof. It follows from Proposition 4.8 that $\mathbb{E}(f_k) = F_k$ for all $k \leq T$, since $T \leq N/2$. Chebyshev's inequality (see Appendix B) then implies

$$\mu(\{|f_k - F_k| \geq \varepsilon\}) \leq \frac{\mathbb{V}(f_k)}{\varepsilon^2}.$$

We will now derive an upper bound on the variance $\mathbb{V}(f_k)$. As in the proof of Proposition 4.8, this is easier if we use the variables (4.12), which means we are actually going to bound $\mathbb{V}(2f_k - 1) = 4\mathbb{V}(f_k)$. First, a standard formula from probability theory gives

$$\mathbb{V}(2f_k - 1) = \mathbb{E}((2f_k - 1)^2) - \mathbb{E}(2f_k - 1)^2.$$

It follows from (4.11) that $\mathbb{E}(2f_k - 1) = (2\alpha - 1)(1 - 2\beta)^k$. Together with (4.13), this means we can rewrite the previous formula as

$$\mathbb{V}(2f_k - 1) = \mathbb{E}\left(\left(\frac{1}{N}\sum_{i=1}^{N} \eta(i-k)\varepsilon(i-1)\cdots\varepsilon(i-k)\right)^2\right) - (2\alpha-1)^2(1-2\beta)^{2k}$$

$$= \frac{1}{N^2}\sum_{i,j=1}^{N} \left(\mathbb{E}\left(\eta(i-k)\eta(j-k)\varepsilon(i-1)\cdots\varepsilon(i-k)\varepsilon(j-1)\cdots\varepsilon(j-k)\right) - (2\alpha-1)^2(1-2\beta)^{2k}\right).$$

If the distance between sites i and j is greater than or equal to k, then all arguments of ε appearing in the second line are distinct and hence all factors are independent. Because the distance between sites cannot be greater than $N/2$, this can only happen if $k \leq N/2$, which is guaranteed by $T \leq N/2$ and $k \leq T$. Using $\mathbb{E}(\eta(i)) = 2\alpha - 1$ and $\mathbb{E}(\varepsilon(j)) = 1 - 2\beta$, the summand then vanishes. The summand does not necessarily vanish for pairs i, j with distance less than k, of which there are $N(2k-1)$. The summand is obviously bounded by 1 from above, hence $\mathbb{V}(2f_k - 1) \leq (2k-1)/N$ and from this

$\mathbb{V}(f_k) \leq (2k-1)/4N$ follows. By elementary properties of probability measures, we finally have

$$\begin{aligned}\mu\left(\bigcap_{k=0}^{T}|f_k - F_k| < \varepsilon\right) &= 1 - \mu\left(\bigcup_{k=0}^{T}\{|f_k - F_k| \geq \varepsilon\}\right)\\ &\leq 1 - \sum_{k=0}^{T}\mu(\{|f_k - F_k| \geq \varepsilon\})\\ &\leq 1 - \sum_{k=0}^{T}\frac{2k-1}{4N\varepsilon^2}\\ &= 1 - \frac{T^2}{4N\varepsilon^2}.\end{aligned}$$

□

4.3.2 Kac chain

Theorem 4.9 can be seen as a time-dependent weak law of large numbers. A strong version does not exist for the Kac ring, since the strong law of large numbers can only be formulated for a system with infinitely many components. For the same reason, Martin-Löf randomness cannot be applied to the Kac ring. To fix this, we switch to the Kac chain, which can be seen as the thermodynamic limit $N \to \infty$ of the Kac ring. It seems to have been introduced by [Hiura and Sasa, 2019].

The Kac chain has state space given by $S = 2^{\mathbb{Z}} \times 2^{\mathbb{Z}}$ with dynamics of the same form (4.1) as before. The interpretation is now an infinite line of coloured balls which change colour upon leaving a marked site. The dynamics of the Kac chain is invertible with the same inverse (4.2) and reversible with the same velocity reversal (4.3). The natural invariant probability measure is the infinite version of (4.5), which is the Bernoulli measure on $2^{\mathbb{Z}} \times 2^{\mathbb{Z}}$ built from the fair probability measure on each factor $\{0,1\}$, for which we use the same symbol π. With respect to this measure, the Kac chain is recurrent by Theorem 4.6. However, the mean recurrence time, which for the Kac ring is of order N, is now infinite. The relevant macroscopic quantities are now the limits of the truncations

$$f^N(x,y) = \frac{1}{2N+1}\sum_{i=-N}^{N} x(i) \quad \text{and} \quad g^N(x,y) = \frac{1}{2N+1}\sum_{i=-N}^{N} y(i) \tag{4.14}$$

as $N \to \infty$, which do not exist for all microstates. However, with respect to the infinite version of the probability measure μ from before, which we denote by the same symbol, these macroscopic quantities do exist for μ-almost all microstates and the first satisfies the macroscopic law (4.11). We denote the time evolution of the first quantity in (4.14) by $f_k^N = f^N \circ \varphi^k$.

Theorem 4.10. *Defining F_k by (4.11) and fixing $k \geq 0$, we have*

$$\lim_{N\to\infty} f_k^N(x,y) = F_k \quad \textit{and} \quad \lim_{N\to\infty} g^N(x,y) = \beta \tag{4.15}$$

for μ-almost all microstates $(x,y) \in S$.

Proof. Because the random variables $y(j)$ are independent and identically distributed, the second equality in (4.15) can be proven in the same way as the strong law of large numbers in Section 3.2.

For the other equality, we begin with the observation that the macroscopic law holds at the level of expectations in the sense that $\mathbb{E}(f_k^N) = F_k$ for all $N \geq 0$ and $k \geq 0$, which is proven in the same way as Proposition 4.8. Note that there is no restriction on the time interval of validity in this case, because the mean recurrence time of the Kac chain is infinite. The exception sets

$$E_{N,m} = \{|f_k^N - F_k| > 1/m\}$$

satisfy

$$\mu(E_{N,m}) \leq \frac{\mathbb{V}(f_k^N)}{1/m^2}$$

by Chebyshev's inequality. It can be shown that $\mathbb{V}(f_k^N) \leq (2k-1)/(4(2N+1))$ for $2N \geq k$ in the same way as the variance bound in the proof of Theorem 4.9. We then have

$$\mu(E_{N,m}) \leq \frac{(2k-1)m^2}{4(2N+1)},$$

which means that $\sum_{N=0}^{\infty} \mu(E_{N^2,m}) < \infty$. The Borel-Cantelli lemma now implies that for μ-almost all $(x,y) \in S$, there exists an $N_0 \geq 0$ such that $(x,y) \in E_{N^2,m}^c$ for all $N \geq N_0$, i.e.,

$$\mu\left(\bigcup_{N_0=0}^{\infty} \bigcap_{N=N_0}^{\infty} E_{N^2,m}^c\right) = 1$$

for all $m \geq 1$. It then follows from elementary probability theory that

$$\mu\left(\bigcup_{m=1}^{\infty} \bigcup_{N_0=0}^{\infty} \bigcap_{N=N_0}^{\infty} E_{N^2,\varepsilon}^c\right) = 1,$$

which is equivalent to $\lim_{N\to\infty} f_k^{N^2}(x,y) = F_k$ for μ-almost all $(x,y) \in S$. The convergence of this subsequence then implies the convergence of the entire sequence to F_k (see the proof of Theorem 3.3 in [Hiura and Sasa, 2019]). □

4.3.3 Randomness

We make the state space of the Kac chain effective by choosing some enumeration of the cylinder sets of $2^{\mathbb{Z}} \times 2^{\mathbb{Z}} \cong 2^{\mathbb{Z}\times\mathbb{Z}}$. Now we can implement the applications of Martin-Löf randomness to statistical mechanics described in Section 3.3 for the Kac chain. First, we turn the strong law of large numbers given by Theorem 4.10 into an effective law of large numbers, giving more information about which microstates have expected macroscopic behaviour.

Theorem 4.11. *Define F_k by (4.11) and assume that α, β are computable real numbers. Fixing $k \geq 0$, we have*

$$\lim_{N\to\infty} f_k^N(x,y) = F_k \quad \textit{and} \quad \lim_{N\to\infty} g_k^N(x,y) = \beta \tag{4.16}$$

for all μ-random microstates $(x,y) \in S$.

Proof. The second equality in (4.16) can be proven in the same way as the effective law of large numbers in Section 3.2. For the other equality, note that the exception sets $E_{N,m}$ defined in the proof

of Theorem 4.10 form a uniformly computable sequence of open sets, since there obviously exists an algorithm to express $E_{N,m}$ as union of cylinder sets. Since $E_{N,m}$ only depends on a finite segment of x, only finitely many cylinder sets need to be checked. Because $\sum_{N=0}^{\infty} \mu(E_{N^2,m}) < \infty$, the sequence $(E_{N^2,m})_{N\geq 0}$ is a Solovay test. If $(x, y) \in S$ is μ-random, then Proposition 3.7 implies that there is an $N_0 \geq 0$ such that $(x, y) \in E^c_{N^2,m}$ for all $N \geq N_0$. It now follows that $f^N_k(x, y)$ converges to F_k as before. □

While Theorem 4.11 does give a class of microstates for which expected macroscopic behaviour occurs, it should be kept in mind that this is with respect to the probability measure μ. Relative to other probability measures, the macroscopic behaviour may be different (see for example Section 5 in [Hiura and Sasa, 2019]). One way to argue for using μ is to say that the invariant probability measure π is natural and that μ can be derived from π in some natural sense, such as the maximum entropy principle.

An interesting consequence of Theorem 4.11 is the following randomness perspective on the Wiederkehreinwand, making more precise in what sense the velocity reversed state considered by Loschmidt is exceptional. For any $k \geq 0$, let μ_k be the image under φ^k, i.e., the probability measure defined by $\mu_k(A) = \mu(\varphi^{-k}(A))$ for all measurable $A \subseteq S$.

Theorem 4.12. *If (x, y) is μ-random, then $\varphi^k(x, y)$ is μ_k-random, but the velocity reversed microstate $(T \circ \varphi^k)(x, y)$ is non-random with respect to μ_k.*

Proof. See Theorem 5.1 in [Hiura and Sasa, 2019]. □

All μ-random microstates have correct macroscopic behaviour, but the converse is not true. That is to say, there exist microstates that are non-random, but for which (4.16) nevertheless hold. The easiest examples of such microstates are in equilibrium, such as (x, y) given by $y(i) = 1$ for all i and

$$x(i) = \begin{cases} 0 & i \text{ is even} \\ 1 & i \text{ is odd,} \end{cases}$$

though we expect that examples out of equilibrium can also be invented. One of the questions [Hiura and Sasa, 2019] have is what to make of such microstates. As with the law of large numbers in Section 3.2, the answer is that μ-random microstates satisfy more macroscopic laws than just (4.16). For example, the proportion of white balls followed by another white ball, which is the limit of

$$\frac{1}{2N+1} \sum_{i=-N}^{N} x(i)x(i+1) \tag{4.17}$$

as $N \to \infty$, is expected to behave as

$$\frac{1}{4}\left(1 + 2(2\alpha - 1)(1 - 2\beta)^k + (2\alpha - 1)^2(1 - 2\beta)^2\right)$$

by a similar reasoning as in the proof of Proposition 4.8. Note that this macroscopic evolution does not relax to its equilibrium value of $1/4$, which can be interpreted as persistence of correlations (see also Section III.15 in [Kac, 1959]). Like with the law of large numbers, we can try to get a characterization through the ergodic theorem. The relevant dynamical system is the two-sided shift map $L(x, y)(i) =$

$(x(i+1), y(i+1))$ on S with invariant probability measure μ. To show that L leaves μ invariant, it suffices to check this on cylinder sets. Unlike the one-sided shift in Section 3.2, the two-sided shift is invertible, so that a two-sided version of the ergodic theorem is more appropriate.

Theorem 4.13. *If $\varphi : S \to S$ is invertible and ergodic with respect to an invariant probability measure μ, then for any measurable $A \subseteq X$ we have*

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} \delta_{\varphi^k(x)}(A) = \mu(A) \tag{4.18}$$

for μ-almost all $x \in X$.

Proof. Applying Theorem 3.8 to both φ and φ^{-1} and adding the results yields (4.18) after some elementary manipulations. □

Similar to Theorem 3.8, we expect an effective version of Theorem 4.13 to exist, which should look something like the following theorem. We conjecture that its proof is similar to that of Theorem 3.9.

Theorem 4.14. *If $\varphi : S \to S$ is invertible and ergodic with respect to an invariant probability measure μ, and φ and μ are computable, then*

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{N} \delta_{\varphi^k(x)}(U) = \mu(U) \tag{4.19}$$

for all computable open $U \subseteq S$ if $x \in S$ is μ-random. Conversely, if (4.19) holds for all computable open $U \subseteq S$, then x is μ-random.

Now we apply Theorem 4.14 to the Kac chain. In the following theorem, φ denotes the dynamics of the Kac chain and L the two-sided shift.

Theorem 4.15. *For any computable open $U \subseteq S$ and any $k \geq 0$,*

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \delta_{\varphi^k(x,y)}(L^{-n}(U)) = \mu(\varphi^{-k}(U)) \tag{4.20}$$

for all μ-random microstates $(x, y) \in S$. Conversely, if (4.20) holds for each computable open $U \subseteq S$, then (x, y) is μ-random.

Proof. If $U \subseteq S$ is a computable open set, then so is $\varphi^{-k}(U)$, since φ is obviously computable and continuous. Applying Theorem 4.14 to the two-sided shift L and the computable open set $\varphi^{-k}(U)$ yields

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \delta_{L^n(x,y)}(\varphi^{-k}(U)) = \mu(\varphi^{-k}(U)) \tag{4.21}$$

for all μ-random microstates $(x, y) \in S$. One easily checks that φ and L commute. It follows that $\delta_{L^n(x,y)}(\varphi^{-k}(U)) = \delta_{\varphi^k(x,y)}(L^{-n}(U))$ and hence the left-hand side of (4.21) is equal to that of (4.20). The converse follows from this equality and the converse in Theorem 4.14. □

Theorem 4.15, if it is correct, answers the question of [Hiura and Sasa, 2019] on non-random microstates satisfying the macroscopic law (4.15). For each computable open $U \subseteq S$, the left-hand side of (4.20) is a macroscopic quantity of the Kac chain and the equality states that it has macroscopic law $\mu(\varphi^{-k}(U))$. For example, if $U = \{(x, y) \in S \mid x(0) = 1\}$, then

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \delta_{\varphi^k(x,y)}(L^{-n}(U)) = f_k^N$$

and (4.20) reduces to the first macroscopic law in (4.15) if one can show that $\mu(\varphi^{-k}(U)) = F_k$, with F_k defined by (4.11). If all these macroscopic laws are satisfied by a microstate (x, y), then it is μ-random. However, one can ask whether all of these laws are physically relevant. If not, then it might be better to use a definition of randomness which has less properties that need to be satisfied.

Lastly, we discuss the Kac chain's analogue of the Stosszahlansatz from the perspective of Martin-Löf randomness. The proof of Theorem 4.11 show that the Stosszahlansatz can be bypassed in proving the macroscopic laws (4.16). In other words, the assumption of randomness is more fundamental than the Stosszahlansatz. This becomes especially clear by looking at macroscopic laws other than (4.16). For example, (4.17) cannot be proven from (4.6) and (4.7), but requires different Stosszahlansatz-like assumptions, all of which are implied by the more fundamental assumption of randomness.

If one wants, the first of the macroscopic laws (4.16) can be proven through the Stosszahlansatz, but this requires more work than the direct proof given above. Denoting the components of the first factor of $\varphi^k(x, y)$ by $x_k(i)$, the Stosszahlansatz for the Kac chain is given by

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} x_k(i)y(i) = \left(\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} y(i) \right) \left(\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} x_k(i) \right)$$

and

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} (1 - x_k(i))y(i) = \left(\lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} y(i) \right) \left(1 - \lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} x_k(i) \right).$$

Proving this requires first showing that the two limits on the right-hand sides, which we denote by $g(x, y)$ and $f_k(x, y)$ respectively, exist for each $k \geq 0$. This is most easily done through the law of large numbers, which also yields the macroscopic laws (4.16) which we are trying to prove, making the Stosszahlansatz redundant. However, let us proceed anyway. Using the formula $x_{k+1}(i) = x_k(i-1) + y(i-1)(1 - 2x_k(i-1))$, the change

$$\sum_{i=-N}^{N} x_{k+1}(i) - \sum_{i=-N}^{N} x_k(i)$$

in the number of white balls on sites $-N$ through N is equal to

$$\begin{aligned}
&\sum_{i=-N}^{N} (x_k(i-1) + y(i-1)(1-2x_k(i-1))) - \sum_{i=-N}^{N} x_k(i) \\
&= \sum_{i=-N}^{N} y(i-1)(1-2x_k(i-1)) + x_k(-N-1) - x_k(N) \\
&= \left(\sum_{i=-N}^{N} y(i)(1-x_k(i)) - \sum_{i=-N}^{N} y(i)x_k(i) \right) + x_k(-N-1) - x_k(N) \\
&\quad + y(-N-1)(1-2x_k(-N-1)) - y(N)(1-2x_k(N)),
\end{aligned}$$

which is the difference between the number of black balls at marked sites and white balls at marked sites in the same interval of sites, up to boundary terms. This implies

$$\begin{aligned}
&\left| \left(\frac{1}{2N+1} \sum_{i=-N}^{N} x_{k+1}(i) - \frac{1}{2N+1} \sum_{i=-N}^{N} x_k(i) \right) \right. \\
&\quad \left. - \left(\frac{1}{2N+1} \sum_{i=-N}^{N} y(i)(1-x_k(i)) - \frac{1}{2N+1} \sum_{i=-N}^{N} y(i)x_k(i) \right) \right| \leq \frac{4}{2N+1},
\end{aligned}$$

from which in turn

$$f_{k+1}(x,y) - f_k(x,y) = g(x,y)(1-2f_k(x,y))$$

follows, when combined with the Stosszahlansatz above. This concludes the derivation of the Kac chain's analogue of the Boltzmann equation through the Stosszahlansatz.

It may be the case that the Stosszahlansatz is not necessary here because the Kac chain is a simple toy model. Perhaps more complicated models have macroscopic laws which are more easily proven through a Stosszahlansatz, which in turn is first proven from an assumption of randomness on the microstate. This could be explored in the Ehrenfests' wind-tree model, which is more complicated than the Kac chain but still reasonably simple (see [Brown et al., 2009]).

5 Stochastic dynamics

When studying irreversibility in a deterministic model, the use of probability is restricted to random initial conditions. Once the initial state is given, the dynamics determine its orbit completely. A system in which transitions of state are themselves random is said to have stochastic dynamics. Models with stochastic dynamics have long been used in statistical mechanics, going back to the work of Einstein, Smoluchowski and Langevin on Brownian motion in the early 20th century [Ebeling et al., 2008]. Even if the fundamental microscopic laws of physics are presumed to be deterministic, stochastic dynamics are still relevant: microscopic conditions cannot be fully known, which means one necessarily needs to work with reduced descriptions, in which information of the microstate is lost. This lack of information through coarse-graining turns a deterministic process into a stochastic one.

Following the mathematical development of the theory of stochastic processes in the second half of the 20th century [Meyer, 2022], stochastic approaches to non-equilibrium systems have flourished. In the mathematical literature, one has the field of interacting particle systems, which studies stochastic processes modelling systems both in and outside physics [Liggett, 1985]. In the physics literature, there is the field of stochastic thermodynamics [Seifert, 2025]. See Chapter 7 of [Uffink, 2006] for more on stochastic approaches to the foundations of non-equilibrium statistical mechanics.

In this section, we look at one of the oldest and simplest models with stochastic dynamics, the Ehrenfest model[1], which was introduced in [Ehrenfest and Ehrenfest, 1907] to discuss the objections of Loschmidt and Zermelo against Boltzmann's H-theorem. The canonical reference for the model's mathematical details is [Kac, 1959]. We use the Ehrenfest model to illustrate the typicality account for stochastic dynamics. Furthermore, we use it as a test case for extending the ideas of [Hiura and Sasa, 2019] on algorithmic randomness to stochastic dynamics, noting the differences with the deterministic case.

5.1 Markov chains

Before describing the Ehrenfest model, it is necessary to explain part of the mathematical theory of stochastic processes, specifically that part concerning Markov chains. We will find that the theory of Markov chains is a miniature version of kinetic theory, having analogues of the Boltzmann equation, relaxation, recurrence and the H-theorem. We are not interested in giving a full and completely rigorous exposition of either Markov chains or stochastic processes in general. The reader may consult [Norris, 1997] for the former and [Karlin and Taylor, 1975] and [Karlin and Taylor, 1981] for the latter. We do provide some background material in Appendix A, but only that which is required for our purposes.

5.1.1 Construction

A stochastic process with state space S is essentially a collection of random variables $(X(t))_{t\in T}$ taking values in S, one for each value of time $t \in T$, and which may be correlated across time (see Appendix A for more details). In general, a stochastic process may have arbitrary correlations across time. However, in most models of interest, including those used in physics, one often makes simplifying assumptions which limit the possible correlations. The most common is the Markov property, which states that the future development of the process only depends on its current state. In other words,

[1] It is also often called the 'Ehrenfest urn model', to prevent confusion with other models named after Ehrenfest, such as the wind-tree model.

the process is memoryless. In the discrete-time case, such processes are called Markov chains. These can be defined for general state spaces (see Section 4.1 in [Kurtz and Ethier, 1986]), but for countable state spaces one has the following definition.

Definition 5.1. *A discrete-time stochastic process $(X(k))_{k\geq 0}$ with countable state space S is said to be a Markov chain if*

$$\mathbb{P}(X(k+1) = x_{k+1} \mid X(0) = x_0, \dots X(k) = x_k) = \mathbb{P}(X(k+1) = x_{k+1} \mid X(k) = x_k) \tag{5.1}$$

for all $k \geq 0$ and $x_0, \dots, x_{k+1} \in S$ such that the conditional probabilities exist (which is the case if and only if $\mathbb{P}(X(0) = x_0, \dots, X(k) = x_k) \neq 0$).

From now on, we restrict to Markov chains with finite state space. Most of what will be presented also works, after suitable modifications, for countably infinite state spaces (see [Norris, 1997]). Let us introduce some more terminology. The distribution[2] of $X(0)$ is called the initial distribution, which we often denote by μ. The probabilities $\mathbb{P}(X(m) = y \mid X(k) = x)$ are called transition probabilities. We restrict to time-homogeneous Markov chains, for which these probabilities depend only on the difference $m - k$. In that case, all transition probabilities are determined by the transition function $P_k(x, y) = \mathbb{P}(X(k) = y \mid X(0) = x)$, which we view as a time-dependent $|S| \times |S|$ matrix.

Definition 5.2. *A matrix P is called stochastic if its entries are all non-negative and all rows have sum 1, i.e., $P(x, \cdot)$ forms a probability distribution for each $x \in S$. If the transpose P^T is also stochastic, one calls P doubly stochastic. If each $P(x, \cdot)$ is deterministic, the matrix P is called deterministic. Equivalently, P is deterministic if there is a function $\varphi : S \to S$ such that $P(x, \cdot) = \delta_{\varphi(x)}$.*

It follows from elementary probability theory that each P_k is a stochastic matrix. A simple application of the law of total probability combined with the Markov property and time-homogeneity shows that the transition function satisfies the Chapman-Kolmogorov equations

$$P_{k+m}(x, z) = \sum_{y \in S} P_k(x, y) P_m(y, z). \tag{5.2}$$

Rewriting these in matrix form $P_{k+m} = P_k P_m$, it becomes clear that $P_k = P^k$, with $P = P_1$ the matrix of one-step transition probabilities. In the language of semigroups, $(P_k)_{k\geq 1}$ form a discrete-time Markov matrix-semigroup with generator P.

Proposition 5.3. *A discrete-time stochastic process $(X(k))_{k\geq 0}$ is a Markov chain if and only if there is a distribution μ and stochastic matrix P such that the finite-dimensional distributions are given by*

$$\mathbb{P}(X(0) = x_0, \dots, X(k) = x_k) = \mu(x_0) P(x_0, x_1) \cdots P(x_{k-1}, x_k). \tag{5.3}$$

Proof. If $(X(k))_{k\geq 0}$ is a Markov chain, one sets μ equal to the initial distribution and P equal to the one-step transition probabilities. Then (5.3) follows from repeated use of the Markov property and time-homogeneity. Conversely, suppose that $\mathbb{P}(X(0) = x_0, \dots, X(k) = x_k) \neq 0$. Then (5.3) implies that (5.1) holds using the definition of conditional probability, from which it follows that X is a Markov chain. □

[2] While the term 'distribution' often refers to the law of a random variable, in this thesis it will refer to the probability mass function of a random variable (see Appendix A).

It follows that a Markov chain is uniquely determined by its initial distribution μ and one-step transition probabilities P. Indeed, given a distribution μ and stochastic matrix P, Kolmogorov's extension theorem (see Appendix A) allows us to construct a unique discrete-time stochastic process satisfying (5.3), which consequently is a Markov chain by Proposition 5.3. This reduction to μ and P is useful, since it allows for the application of linear algebra to the theory of Markov chains. We will see that many probabilistic properties can be expressed in the language of vectors and matrices.

Although we have defined a Markov chain as a certain type of stochastic process, analogy with deterministic dynamics suggests to regard a Markov chain as being defined solely by a matrix of one-step transition probabilities P, just as one may consider equations of motions without choosing initial conditions. Appending an initial distribution μ then yields a stochastic process, which is a Markov chain in the sense of Definition 5.1. We will use both perspectives throughout this thesis, as sometimes one is more convenient than the other.

5.1.2 Master equation

Now that we have seen how to construct Markov chains, let us look at their properties. An obvious quantity of interest is the distribution of $X(k)$, which we denote by μ_k. An application of the law of total probability yields the relation

$$\mu_{k+1}(y) = \sum_{x \in S} \mu_k(x) P(x, y), \tag{5.4}$$

through which all distributions can be determined inductively beginning from the initial distribution. If we represent probability distributions by row vectors, equation (5.4) can be written as the matrix multiplication $\mu_{k+1} = \mu_k P$. Iteration yields $\mu_k = \mu P^k$, which could also have been derived from the Chapman-Kolmogorov equations (5.2). Assuming P is diagonalizable[3], it follows that one may determine μ_k by diagonalizing P. However, at the moment we are less interested in the solutions of (5.4) and more in the equation itself, which we see as the evolution equation of a time-dependent probability distribution. When rewritten in the form

$$\mu_{k+1}(y) - \mu_k(y) = \sum_{x \in S} \left(\mu_k(x) P(x, y) - \mu_k(y) P(y, x) \right) \tag{5.5}$$

through use of $\sum_{x \in S} P(y, x) = 1$, equation (5.4) gets called the master equation. Though trivial, this rewriting has conceptual advantages. Following [Uffink, 2006], we interpret the master equation as follows. The probability of state y increases as result of transitions from other states x to y, and decreases as a result of transitions from y to other states x. Hence, the probability of a state remains constant in time if these two effects cancel each other. With respect to this balancing idea, the master equation is similar to the Boltzmann equation (2.8). However, there are some aspects in which the similarity fails:

1. The master equation presented here is in discrete-time, while the Boltzmann equation is in continuous-time.
2. The master equation is linear, while the Boltzmann equation is non-linear.
3. The master equation determines the evolution of a probability distribution, while the Boltzmann equation concerns a distribution of single-particle states.

[3] An important case in which P is diagonalizable is when it is reversible (see Section 5.3).

The first is not a significant problem, since the theory of Markov jump processes, which are in continuous-time, also has a master equation analogous to the homogeneous Boltzmann equation (see Section 5.5, in particular equation (5.53)). The other two non-similarities are more significant. They imply that the true analogue of the master equation is actually the Liouville equation (2.11), which is similarly linear and an equation for a time-dependent probability distribution. In this sense, the general theory of Markov chains is more Gibbsian than Boltzmannian. However, it is still possible to construct a closer analogy between Boltzmann's kinetic theory and Markov chains by restricting to certain types of Markov chains, as we will detail in Sections 5.2 and 5.4.

5.1.3 Relaxation, recurrence and H-theorem

Continuing the analogy with kinetic theory, the theory of Markov chains even has notions of relaxation and recurrence and an analogue of the H-theorem. To understand this, we first need to introduce some more concepts.

Definition 5.4. *A Markov chain is said to be stationary if the distribution of $X(k)$ does not change with time. Equivalently, its initial distribution μ is invariant under the master equation* (5.5)*:*

$$\sum_{x \in S} (\mu(x)P(x,y) - \mu(y)P(y,x)) = 0. \tag{5.6}$$

Changing perspective, we call a distribution μ stationary with respect to the Markov chain given by a stochastic matrix P if (5.6) *holds, i.e., if the process built from μ and P is stationary.*

Stationary distributions of Markov chains play the same role as the Liouville measure in Hamiltonian mechanics and the invariant measure of the Kac ring: they are the natural measures corresponding to the dynamics and characterize the equilibrium of the system. A finite-state Markov chain always has at least one stationary distribution. There are many ways to prove this. The simplest is to notice that since P is stochastic, it has 1 as eigenvalue and hence must have a corresponding left-eigenvector, which is a stationary distribution. Uniqueness of the stationary distribution does not hold for all Markov chains, but can only fail if the Markov chain is decomposable in a certain sense.

Definition 5.5. *A Markov chain is said to be irreducible if for any $x, y \in S$, there exists some $k \in \mathbb{N}$ such that $P^k(x,y) \neq 0$, i.e., from any state x it is always possible to reach any other state y after a sufficiently large number of time steps.*

Proposition 5.6. *An irreducible Markov chain has a unique stationary distribution π such that $\pi(x) > 0$ for all $x \in S$.*

Proof. For any distribution π, there is at least one $x \in S$ such that $\pi(x) > 0$. Given $z \in S$, choose $k \geq 0$ such that $P^k(x,z) > 0$. If π is stationary, then $\pi P^k = \pi$ and hence

$$\pi(z) = \sum_{y \in S} \pi(y)P^k(y,z) \geq \pi(x)P^k(x,z) > 0.$$

To show that the stationary distribution is unique, we follow the hint given for Exercise 1.17 in [Levin and Peres, 2017]. If π_1 and π_2 are both stationary distributions, let $x \in S$ be a minimizer of

$\pi_1(x)/\pi_2(x)$. Then for any $k \geq 0$, we have

$$\begin{aligned}\pi_1(x) &= \sum_{y\in S} \pi_1(y)P^k(y,x) \\ &= \sum_{y\in S} \frac{\pi_1(y)}{\pi_2(y)}\pi_2(y)P^k(y,x) \\ &\geq \frac{\pi_1(x)}{\pi_2(x)} \sum_{y\in S} \pi_2(y)P^k(y,x) = \pi_1(x),\end{aligned}$$

from which it follows that

$$\sum_{y\in S} \left(\frac{\pi_1(y)}{\pi_2(y)} - \frac{\pi_1(x)}{\pi_2(x)}\right) P^k(y,x) = 0.$$

The term between the brackets is non-negative by definition of x, so that the summand vanishes for each $y \in S$. For each $y \in S$, there is a $k \geq 0$ such that $P^k(y,x) > 0$, from which it follows that the term between the brackets vanishes. This implies $\pi_1 = \pi_2$. □

Definition 5.7. *Given a state $x \in S$, let $T(x) = \{k \geq 1 \mid P^k(x,x) \neq 0\}$ be the set of times for which a return to x is possible when starting the process in x. The period of x is defined to be* $\gcd T(x)$. *A Markov chain is aperiodic if all its states have period 1.*

If a Markov chain is irreducible, then its states have a common period, call it d (see Lemma 1.6 in [Levin and Peres, 2017]). If $d > 1$, then the state space can be partitioned in subsets $C_1 \ldots, C_d$ such that $P(x,y) \neq 0$ if and only if $x \in C_i$ and $y \in C_j$ with $j \equiv i+1 \pmod d$. The matrix P^d then defines an irreducible and aperiodic Markov chain on each C_i (see Section 6.3 in [Kemeny et al., 1976]).

Definition 5.8. *A Markov chain is ergodic if it is both irreducible and aperiodic.*

This definition agrees with ergodicity defined for deterministic dynamics in the following sense. To any Markov chain with given initial distribution, one can associate a dynamical system $(S^{\mathbb{N}}, \mathbb{P}, T)$, with $\mathbb{P}$ the Kolmogorov representation of the process (see Appendix A) and $T : S^{\mathbb{N}} \to S^{\mathbb{N}}$ the shift map, mapping $(x_k)_{k\geq 0}$ to $(x_{k+1})_{k\geq 0}$. This dynamical system is metrically transitive, hence ergodic, if the Markov chain is ergodic in the sense of Definition 5.8.

Proposition 5.9. *A Markov chain is ergodic if and only if there is some $k \geq 1$ such that $P^k(x,y) \neq 0$ for all $x, y \in S$.*

Proof. See Proposition 1.7 in [Levin and Peres, 2017] for one direction and Theorems 8.5.2 and 8.5.3 in [Horn and Johnson, 1985] for the converse. □

Ergodic Markov chains exhibit relaxation in the sense that any initial distribution evolves into the stationary distribution. The rate of this convergence is even exponential, which is particularly useful for algorithms based on Markov chains, such as Markov chain Monte Carlo methods (see Chapter 3 in [Levin and Peres, 2017]). To measure this rate of convergence, we need to introduce a distance on the space $\operatorname{Prob} S$ of probability distributions on S. A natural metric (see Section 4.1 in [Levin and Peres, 2017]) is the total variation distance

$$\|\mu - \nu\| = \frac{1}{2} \sum_{x\in S} |\mu(x) - \nu(x)|. \tag{5.7}$$

Theorem 5.10 (Fundamental theorem of Markov chains)**.** *If the Markov chain given by the stochastic matrix P is ergodic, then it has a unique stationary distribution π and any initial distribution μ evolves into π with exponential rate relative to the total variation distance, i.e., $\mu P^k \to \pi$ in* $\operatorname{Prob} S$ *as $k \to \infty$ and there are $C \geq 0$ and $a \in (0,1)$ such that*

$$\|\mu P^k - \pi\| \leq C a^k.$$

Like other results in mathematics which are fundamental in their respective fields, this theorem has many proofs. One common proof is based on the Perron-Frobenius theorem from linear algebra (see Chapter 8 in [Horn and Johnson, 1985]). Another proof uses Banach's fixed point theorem. More probabilistic proofs can also be given. A probabilistic approach to the existence and uniqueness of π, which also works for Markov chains with countably infinite state space, involves the notion of recurrence.

Given a state $x \in S$, we use the notation $\mathbb{P}_x$ for probabilities when the Markov chain has deterministic initial distribution δ_x. The state x is said to be recurrent if

$$\mathbb{P}_x\left(X(k) = x \text{ for infinitely many } k\right) = 1,$$

i.e., the process returns infinitely often to x when starting from x. We define the random variable

$$T_x = \min\{k \geq 1 \mid X(k) = x\},$$

which is the first time the process arrives at state x (after possibly starting at x). The state x is said to be positive recurrent if it is recurrent and $\mathbb{E}_x(T_x) < \infty$, i.e., it is expected to recur within a finite number of time steps. Here $\mathbb{E}_x$ denotes expectation relative to $\mathbb{P}_x$. See Section 1.5 in [Norris, 1997] for more details on recurrence. All states of an irreducible Markov chain with finite state space are positive recurrent, which is intuitively obvious (see Theorem 1.7.7 in [Norris, 1997]). This is analogous to the Poincaré recurrence theorem, as first pointed out by [Kac, 1947a], in which the following result was proven.

Lemma 5.11. *If π is the stationary distribution of an irreducible Markov chain P with finite state space, then $\pi(x) = 1/\mathbb{E}(T_x)$ for any $x \in S$*

Proof. See Lemma 21.12 in [Levin and Peres, 2017]. □

The probabilistic proof of the existence and uniqueness of the stationary distribution π essentially hinges on Lemma 5.11: for an irreducible Markov chain in which each state is positive recurrent, one can construct π from the expected recurrence times $\mathbb{E}_x(T_x)$ (see Section 1.7 in [Norris, 1997]).

The convergence and its exponential rate in Theorem 5.10 also have a probabilistic proof (see Theorem 4.9 in [Levin and Peres, 2017]). We will give another kind of proof for the convergence based on entropy. In order to be able to do this, we first need to look at the analogue of Boltzmann's H-theorem for Markov chains.

Definition 5.12. *The relative entropy of two probability distributions μ, ν on S is defined by*

$$H(\mu|\nu) = \sum_{x \in S} \mu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right) \tag{5.8}$$

if $\mu \ll \nu$, with the convention that the summand in (5.8) is zero if both $\mu(x)$ and $\nu(x)$ vanish. Otherwise, it is set equal to ∞.

Proposition 5.13. *For any distributions μ, ν, $H(\mu|\nu) \geq 0$ and equality holds if and only if $\mu = \nu$.*

Proof. We assume that $\nu(x) > 0$ for all $x \in S$, since that is the only case we will be interested in. If we set $\rho(x) = \mu(x)/\nu(x)$, then (5.8) can be rewritten as

$$H(\mu|\nu) = \sum_{x \in S} \nu(x)\eta(\rho(x)),$$

with $\eta : [0, \infty) \to \mathbb{R}$ the function defined by $\eta(t) = t \log t$ and $\eta(0) = 0$. It is easy to check that η is convex, so that we may use Jensen's inequality to conclude

$$H(\mu|\nu) \geq \eta\left(\sum_{x \in S} \nu(x)\rho(x)\right) = \eta\left(\sum_{x \in S} \mu(x)\right) = \eta(1) = 0.$$

Because η is even strictly convex and $\nu(x) > 0$ for all $x \in S$, equality holds if and only if ρ is constant, which is equivalent to $\mu(x)\nu(y) = \mu(y)\nu(x)$ for all $x, y \in S$. Summing over y then yields $\mu(x) = \nu(x)$ for all $x \in S$. □

Proposition 5.14. *Given any two distributions μ, ν and a stochastic matrix P, we have*

$$H(\mu P|\nu P) \leq H(\mu|\nu).$$

Proof. We follow the proof of Theorem 3.1 in [Mackey, 1992]. Again, we restrict to the case that $\nu(x) > 0$ for all $x \in S$. To begin,

$$H(\mu P|\nu P) = \sum_{y \in S} \nu P(y)\eta\left(\frac{\mu P(y)}{\nu P(y)}\right),$$

using the same function η as in the proof of Proposition 5.13. Using $\rho(x) = \mu(x)/\nu(x)$ and Jensen's inequality again, we have

$$\eta\left(\frac{\mu P(y)}{\nu P(y)}\right) = \eta\left(\frac{\sum_{x \in S} \rho(x)\nu(x)P(x,y)}{\sum_{x \in S} \nu(x)P(x,y)}\right) \leq \frac{\sum_{x \in S} \eta(\rho(x))\nu(x)P(x,y)}{\sum_{x \in S} \nu(x)P(x,y)} \tag{5.9}$$

and hence

$$H(\mu P|\nu P) \leq \sum_{x,y \in S} \eta(\rho(x))\nu(x)P(x,y) = \sum_{x \in S} \eta(\rho(x))\nu(x) = H(\mu|\nu).$$

□

By Proposition 5.15, the stationary distribution π of an irreducible Markov chain satisfies $\pi(x) > 0$ for all $x \in S$ and $H(\mu|\pi)$ is consequently always defined by (5.8). This is our analogue of Boltzmann's

H-quantity, though it is more correct to say that it is analogous to the negative of the Gibbs entropy (2.12), which becomes clear upon rewriting (5.8) as

$$H(\mu|\pi) = \sum_{x\in S} \rho(x)\log(\rho(x))\pi(x) \tag{5.10}$$

and comparing with (2.12), where $\rho(x) = \mu(x)/\pi(x)$ the analogue of the phase space density function. If μ_k is the evolution of an initial distribution μ under a Markov chain P, i.e., $\mu_k = \mu P^k$, then it follows from Proposition 5.14 that the quantity $H(\mu_k|\pi)$ is non-increasing, which is the analogue of Boltzmann's H-theorem. Looking at (2.9) one may think that the correct analogue of Boltzmann's H-quantity should be

$$H(\mu) = \sum_{x\in S} \mu(x)\log\mu(x), \tag{5.11}$$

which is the negative Shannon entropy of μ. However, this quantity is not necessarily non-increasing. In fact, by Proposition 5.15, $H(\mu_k)$ is non-increasing if and only if the stationary distribution π is uniform, at least for ergodic Markov chains. This means that $H(\mu_k)$ is non-increasing if and only if the system satisfies the principle of equal a priori probabilities.

Proposition 5.15. *For any ergodic Markov chain P, the following are equivalent.*

1. *The stationary distribution π is uniform.*
2. *The matrix P is doubly stochastic.*
3. *For any initial distribution μ, $H(\mu_k)$ is non-increasing.*

Proof. The first two conditions are equivalent, since both are equivalent to

$$\sum_{x\in S} \frac{1}{|S|} P(x,y) = \frac{1}{|S|}.$$

If π is uniform, then $H(\mu) = H(\mu|\pi) - \log|S|$ and hence $H(\mu_k)$ is non-increasing by Proposition 5.14. Conversely, suppose π is non-uniform. Let μ be the uniform distribution. By Theorem 5.10 (which does not depend on Proposition 5.15) and the continuity of relative entropy, $H(\mu_k) \to H(\pi)$ as $k \to \infty$. But, one can show that (5.11) is maximal for the uniform distribution, so that $H(\mu_k)$ must decrease at some point. □

The following lemma says that in an ergodic Markov chain $H(\mu_k|\pi)$ actually decreases, which will be used in our proof of Theorem 5.10. Because $\mu \mapsto H(\mu \mid \pi)$ is clearly continuous, it follows from Theorem 5.10 and Proposition 5.13 that $H(\mu_k|\pi)$ decreases to zero. Note that our reasoning is inverse to that of Boltzmann. He proves his H-theorem from relaxation in the form of the Boltzmann equation, while we use the analogue of the H-theorem to prove relaxation.

Lemma 5.16. *For an ergodic Markov chain P with stationary distribution π, there is some $m \geq 1$ such that $H(\mu P^m|\pi) < H(\mu|\pi)$ for any initial distribution $\mu \neq \pi$.*

Proof. By Proposition 5.9, there exists some $m \geq 1$ such that P^m has non-zero entries. If we go through the proof of Proposition 5.14 with P^m instead of P and $\nu = \pi$, then the weights $\pi(x)P^m(x,y)$ in Jensen's inequality (5.9) are non-zero, which means the inequality $H(\mu P^m|\pi P^m) \leq H(\mu|\pi)$ becomes strict, since we assume $\mu \neq \pi$. Because π is stationary, it follows that $H(\mu P^m|\pi) < H(\mu|\pi)$. □

As a side remark, note that Lemma 5.16 also provides a proof of the uniqueness of the stationary distribution π, at least for ergodic Markov chains. For, if μ is also stationary, then $H(\mu P^m|\pi) = H(\mu|\pi)$, which means we must have $\mu = \pi$. The exponential rate of convergence can also be proven using entropy methods (see [Caputo, 2022]). We wonder whether the same is true for the existence of the stationary distribution.

Now we can give our proof of Theorem 5.10 using relative entropy. This proof is similar to the one given in Section II.4 of [Liggett, 1985]. The same reference also contains some remarks on the history of the entropy method.

Proof of Theorem 5.10. We first prove an auxiliary result. If $A \subseteq \operatorname{Prob} S$ is compact and closed under right multiplication by P, then $\pi \in A$. For, $\mu \mapsto H(\mu|\pi)$ is continuous and hence attains a minimum on A, say at μ. If $\mu \neq \pi$, then Lemma 5.16 implies $H(\mu P^m|\pi) < H(\mu|\pi)$ for some $m \geq 1$, which contradicts μ being a minimum. Hence, $\mu = \pi$.

Given any $\mu \in \operatorname{Prob} S$, let $A = \{\mu P^k \mid k \geq 0\}$. Then $\overline{A}$ is compact and closed under right multiplication by P. Indeed, if $\mu P^{k_i} \to \nu$, then $\mu P^{k_i+1} \to \nu P$ since right multiplication by P is obviously continuous. It follows that $\pi \in \overline{A}$, which means that $\mu P^{m_i} \to \pi$ for some sequence $(m_i)_{i\geq 1}$ and hence $H(\mu P^{m_i}|\pi) \to 0$. Since $H(\mu P^k|\pi)$ is a non-increasing sequence by Proposition 5.14, we have $H(\mu P^k|\pi) \to 0$. If $\mu P^{k_i} \to \nu$, then

$$H(\nu|\pi) = \lim_{i\to\infty} H(\mu P^{k_i}|\pi) = 0,$$

which can only be the case if $\nu = \pi$. Consequently, all subsequences of $(\mu P^k)_{k\geq 0}$ converge to π, allowing us to conclude $\mu P^k \to \pi$. □

5.2 Ehrenfest model

We pointed out in the previous section that the general theory of Markov chains, when viewed as a miniature version of kinetic theory, is more Gibbsian than Boltzmannian, since it deals with probability distributions rather than empirical distributions of single-particle states. This is remedied by restricting to population Markov chains, which have state space of the form $S \subseteq A^N$, i.e., the state consisting of N individual states in A. In models of kinetic theory, elements of A are the single-particle states. Writing such a process as $X(k) = (X_1(k), \ldots, X_N(k))$, the analogue of Boltzmann's distribution function (2.2) is the time-dependent empirical measure

$$\frac{1}{N}\sum_{n=1}^{N} \delta_{X_n(k)}, \tag{5.12}$$

whose evolution equation would be the true analogue of the Boltzmann equation, rather than the master equation (5.5) governing the probability distribution μ_k on S. However, (5.12) is a stochastic process, which does not satisfy a deterministic evolution equation. We will see how to deal with this problem in Section 5.4. One of the earliest studied population Markov chains, and at the same time the earliest Markov chain model used in statistical mechanics, is the Ehrenfest model, which we will look at together with a similar variant which is more suited for our purposes.

5.2.1 Original model

The original Ehrenfest model, introduced in [Ehrenfest and Ehrenfest, 1907], has single-particle state space $A = \{0, 1\}$ and stochastic dynamics given by the following simple rule. At each time step, one

out of N particles is chosen at random and its state is changed. The transition probabilities are thus given by $P(x, y) = 1/N$ if x, y differ in only one single-particle state and $P(x, y) = 0$ in all other cases. There are multiple ways of physically interpreting the Ehrenfest model.

- The two states may represent two locations between which the particles randomly move, making it a model of diffusion (see [Scalas et al., 2007]). Less physically, one often hears talk of balls being moved between urns or fleas hopping between dogs.
- Redefining $A = \{-1, 1\}$, the states can be interpreted as spins which randomly flip, making it a model of thermal agitation of magnetic particles. A slight generalization leads to the Glauber dynamics, which includes magnetic interaction (see [Falk, 1980]).
- Using again $A = \{-1, 1\}$, the states can be interpreted as velocities, giving a model of a gas.

The last interpretation in this list is, of course, the most natural for kinetic theory. However, the state transitions are not due to collisions between particles, but due to random fluctuations. Hence the Ehrenfest model is trivial as a kinetic model. Nevertheless, it exhibits irreversibility in the sense that it relaxes to equilibrium, as we will see in Section 5.4, showing that collisions are not the only mechanism of irreversibility.

When one encounters the Ehrenfest model in the literature, it is more often than not the process

$$Y(k) = \sum_{n=1}^{N} X_n(k)$$

which is meant, rather than the more fundamental process $X(k)$. This process has state space $\{0, 1, 2, \ldots, N\}$ and it can be shown that it is also a Markov chain. As is clear from its definition, $Y(k)$ represents the number of particles in state 1. However, it can also be given interpretations independent of an underlying particle interpretation. For example, [Kac, 1947b] interprets it as a random walk with restoring force. In [Kac, 1959], the Ehrenfest model in the sense of $Y(k)$ is used as an analogy for Hamiltonian dynamics, to show that time-reversibility and recurrence can in principle be reconciled with irreversibility. In view of the critical distinction between microstates and macrostates in the typicality account, it seems to us more appropriate to base an analogy with kinetic theory on the underlying population process $X(k)$, which is what we do next. At the same time, this analogy illustrates the typicality account for a stochastic model.

To start our analogy, the state-space of the Ehrenfest model, which we denote by 2^N, consists of the possible microstates, corresponding to the microstates making up the phase space of a gas. The stochastic dynamics of the Ehrenfest model, given by its transition probabilities, is analogous to the Hamiltonian dynamics on the phase space of a gas.

Proposition 5.17. *The Ehrenfest model is an irreducible Markov chain with period 2.*

Proof. Two states $x, y \in 2^N$ differ in at most N single-particle states. Changing one single-particle state of x at a time, one constructs a path from x to y with non-zero probability, showing that P is irreducible. Because the parity of the number of ones changes at each time step, the Ehrenfest model can only return to its original state after an even number of time steps, which means the period is equal to 2. □

It follows from Proposition 5.6 that the Ehrenfest model has a unique stationary distribution π, which plays the same role as the invariant probability measure in the typicality account. It is easily seen that the Ehrenfest model has a doubly stochastic matrix of transition probabilities. Hence, by Proposition 5.15, the stationary distribution is uniform and given by $\pi(x) = 1/2^N$.

The Ehrenfest model is periodic and consequently cannot be ergodic. This means that the convergence statement in Theorem 5.10 does not apply. For example, the probability distribution $P^k(x, \cdot)$ does not converge to π as $k \to \infty$, since its support keeps switching back and forth. However, by the remarks following Definition 5.7, the state space 2^N can be partitioned into two subsets

$$C_{\text{even}} = \left\{ x \in 2^N \mid \sum_{n=1}^{N} x_n \text{ is even} \right\} \quad \text{and} \quad C_{\text{odd}} = \left\{ x \in 2^N \mid \sum_{n=1}^{N} x_n \text{ is odd} \right\}.$$

such that P^2 is ergodic on both. The Ehrenfest model thus still exhibits relaxation to equilibrium, in the sense that P^{2k} converges to π on C_{even} and C_{odd} separately. The periodicity of 2 is essentially a mathematical inconvenience, which is one reason for looking at a variant of the Ehrenfest model in Section 5.2.2.

Irreversibility is, as we have seen before, a phenomenon which only occurs when switching from a microscopic description to a macroscopic (or mesoscopic) description of the system by means of coarse-graining. For the Ehrenfest model, the relevant macroscopic quantity is the time-dependent empirical measure (5.12). Since there are only two single-particle states, this random measure is equivalent to the random variable

$$f(k) = \frac{1}{N} \sum_{n=1}^{N} X_n(k), \tag{5.13}$$

which is a scaled version of the the process $Y(k)$ define earlier and is a Markov chain. Its state space is given by $\{m/N \mid m = 0, 1, 2, \ldots, N\}$. It is not hard to see that the non-zero transition probabilities are given by

$$P\left(\frac{n}{N}, \frac{n+1}{N}\right) = 1 - \frac{n}{N} \quad \text{and} \quad P\left(\frac{n}{N}, \frac{n-1}{N}\right) = \frac{n}{N}. \tag{5.14}$$

This macroscopic process is similarly irreducible and has period 2. Since all microstates have equal probability $1/2^N$, it is intuitively obvious that the stationary distribution of the macroscopic process is given by

$$\frac{n}{N} \mapsto \frac{1}{2^N} \binom{N}{n}. \tag{5.15}$$

We can equivalently first define a coarse-graining map $f : 2^N \to [0, 1]$ by

$$f(x) = \frac{1}{N} \sum_{n=1}^{N} x_n$$

and then express $f(k)$ as f applied to $X(k)$. Corresponding to this choice of coarse-graining, there is an equilibrium macrostate consisting of microstates having $f \approx 1/2$. It should be characterized by having the largest size among the macrostates, according to the stationary distribution π. Indeed, the projections $X_n : 2^N \to \{0, 1\}$ are independent random variables with expectation $1/2$ under the

probability distribution π on 2^N, so that we may appeal to Hoeffding's inequality (Proposition B.4) to conclude

$$\pi(\{x \in 2^N \mid |f(x) - 1/2| < \varepsilon\}) \geq 1 - 2\exp(-2N\varepsilon^2), \tag{5.16}$$

for any $\varepsilon > 0$. The bound in (5.16) is approximately equal to 1 for large N. As with the Kac ring, the buffer ε is required to actually make the equilibrium macrostate dominant (see Section 4.2). The size of a macrostate according to π is equal to (5.15), which becomes smaller the further away it is from equilibrium, as is needed for the typicality account to work.

If the Ehrenfest model starts in a non-equilibrium macrostate, corresponding to a value of f far from $1/2$, the resulting stochastic evolution of $f(k)$ should very likely be an essentially monotonic convergence from the initial value to $1/2$, assuming the number of particles N is large. The typicality account explains this qualitatively through the differences in size of the macrostates. In the case of the Ehrenfest model, this qualitative explanation is more specific: the Ehrenfest model can be viewed as a random walk on the hypercube 2^N, so that it is no surprise that it wanders into larger portions of this space and stays there for long times. After switching to a variant of the Ehrenfest model, we will study the evolution of $f(k)$ quantitatively in Section 5.4 and prove that it indeed relaxes to equilibrium.

The objections of Loschmidt and Zermelo to Boltzmann's kinetic theory can be illustrated in the Ehrenfest model. Recall that the model was invented for this purpose. Beginning with the Umkehreinwand, the Ehrenfest model has no analogue of velocity reversal, but microscopic time-reversibility can still be formulated in the following form.

Proposition 5.18. *Given any two microstates $x, y \in 2^N$, the probability of transitioning from x to y in a given time k is equal to the probability of transitioning from y to x in the same time, i.e., $P^k(x, y) = P^k(y, x)$*

Proof. The way in which the Ehrenfest model's transition probabilities are defined makes it clear that $P(x, y) = P(y, x)$ holds for any $x, y \in 2^N$, i.e., P is a symmetric matrix. It follows that P^k is symmetric as well. □

Loschmidt's paradox is now the observation that trajectories going from non-equilibrium to equilibrium are as likely as trajectories which do the opposite, so why do we only observe the former? Of course, this not a true paradox, since the almost sure relaxation to equilibrium can be mathematically proven, as we will do in Section 5.4. Instead, we have here a conceptual paradox. The solution lies in coarse-graining. Suppose x belongs to some non-equilibrium macrostate and y belongs to the equilibrium macrostate. Even though $x \to y$ does have the same probability as $y \to x$, due to (5.16) there are many more microstates like y than there are like x, making an evolution from equilibrium to non-equilibrium much less likely than the reverse when viewed macroscopically. Stated differently, while the microscopic Ehrenfest model satisfies the symmetry $P(x, y) = P(y, x)$, the macroscopic version does not.

The Ehrenfest model exhibits Zermelo's Wiederkehreinwand in the sense that it is an irreducible Markov chain with finite state space, hence every microstate is positive recurrent. Likewise, the macroscopic process $f(k)$ is also positive recurrent. Does this not imply that the Ehrenfest model will have anti-thermodynamic behaviour? It does, since recurrence is a mathematical theorem. However, this does not impede the very high likelihood of relaxation when starting in non-equilibrium. In addition, the recurrence of an initial non-equilibrium state will never be observed in practice due to

the recurrence time being enormously large for large N, which follows from Lemma 5.11 and the macroscopic stationary distribution (5.15).

Not all fundamental aspects of kinetic theory have an analogue in the Ehrenfest model, or population Markov chains in general. An example is the non-invariance of the Gibbs entropy (5.10). We will show in Section 5.3.3 that the Gibbs entropy is invariant for each initial distribution μ if and only if P is deterministic (see Definition 5.2). It follows that in the Ehrenfest model $H(\mu_k|\pi)$ increases for any non-stationary μ, unlike its analogue (2.12) in kinetic theory. The explanation is that stochastic dynamics inherently dissipate information, which is measured by entropy: one cannot perfectly predict the future or retrodict the past from the present state, expect in the degenerate case of deterministic P.

5.2.2 Variant model

We introduce here a variant of the Ehrenfest model, which is studied and seemingly introduced for the first time in [Hauert et al., 2004], for the following three reasons.

1. The particles in the original Ehrenfest model are not truly independent, since only a single particle can change state at each time step. Of course, in kinetic theory one is interested in models in which particles interact, but the interaction in the original Ehrenfest model is of an artificial kind. Hence, we seek a Markov chain model in which the particles are truly independent.
2. The original Ehrenfest model is not ergodic, which is mathematically inconvenient.
3. Most significantly, it is impossible to formulate the original Ehrenfest model with an infinite number of particles, simply because the dynamics involves choosing a particle at random and there is no uniform probability measure on a countably infinite set. This is a problem for us, because Martin-Löf randomness can only be applied to infinite objects.

These problems can be solved using the following method. Let P_n be stochastic matrices for $n = 1, \dots, N$ defining Markov chains with state space A. We then define the stochastic matrix P with state space $S = A^N$ by

$$P(x, y) = \prod_{n=1}^{N} P_n(x_n, y_n).$$

Let us denote the corresponding process by $X(k) = (X_1(k), \dots, X_N(k))$. Upon appending an initial distribution of the form

$$\mu(x) = \mu_1(x_1) \cdots \mu_N(x_N), \tag{5.17}$$

each projection $(X_n(k))_{k\geq 1}$ is itself a Markov chain and at each k the random variables $(X_n(k))_{1\leq n\leq N}$ are independent, which follows from Proposition 5.3 combined with the form of μ and P. To obtain our variant of the Ehrenfest model, we set $A = \{0, 1\}$ and let each P_n be equal to the stochastic matrix

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \tag{5.18}$$

corresponding to a single particle with possible states 0 and 1, jumping from its current state to the other with probability p at each time step. The stochastic matrix P defines the variant of the Ehrenfest model considered by [Hauert et al., 2004]. The Markov chain given by (5.18) is the most general form of a two-state Markov chain which is reversible in the sense of Proposition 5.18, making it the simplest Markov chain relevant to statistical mechanics.

The variant of the Ehrenfest model thus constructed has essentially the same properties as the original model. It has the same stationary distribution $\pi(x) = 1/2^N$, is reversible in the sense that $P(x, y) = P(y, x)$ and exhibits recurrence, with the same recurrence times due to Lemma 5.11. Unlike the original, the variant is ergodic since there is always a non-zero probability that the microstate does not change. Furthermore, it is possible to construct a version with an infinite number of particles by simply taking an infinite number of copies of the single-particle Markov chain, which we detail in Section 5.4.5. This infinite version will be used to formulate a time-dependent strong law of large numbers, which will be turned into a statement about Martin-Löf randomness of microstates. A disadvantage of the variant is that its transition probabilities are more complicated than those of the original.

5.3 Reversibility

We saw in the previous section that the Ehrenfest model is reversible in the sense that $P(x, y) = P(y, x)$ for any microstates $x, y \in 2^N$, i.e., the matrix of transition probabilities is symmetric. However, other definitions of reversibility are also possible and it is not immediately clear which is correct, unlike the case of deterministic dynamics.

5.3.1 Possible definitions

Definition 5.19. *A Markov chain with matrix of transition probabilities is said to be state-symmetric if $P(x, y) = P(y, x)$ holds for any $x, y \in S$.*

Proposition 5.20. *For a Markov chain with matrix of transition probabilities P, the following properties are equivalent.*

1. $\mathbb{P}(X(1) = y \mid X(0) = x) = \mathbb{P}(X(1) = x \mid X(0) = y)$ *for any $x, y \in S$, i.e., P is symmetric.*
2. $\mathbb{P}(X(k) = y \mid X(0) = x) = \mathbb{P}(X(k) = x \mid X(0) = y)$ *for any $x, y \in S$ and any $k \geq 1$, i.e., P^k is symmetric.*
3. $\mathbb{P}(X(0) = x_0, \ldots, X(k) = x_k \mid X(0) = x_0) = \mathbb{P}(X(0) = x_k, \ldots, X(k) = x_0 \mid X(0) = x_k)$ *for any $k \geq 1$ and $x_0, \ldots, x_k \in S$.*

Proof. The first two are obviously equivalent. The third follows from the first using the definition of conditional probability and (5.3). Conversely, the first follows from the third by setting $k = 1$. □

While state-symmetry seems to be the natural analogue to reversibility of deterministic dynamics (see Proposition 5.26), and the failure of state-symmetry of a coarse-grained system seems like a reasonable explanation of irreversibility (see the discussion following Proposition 5.18), it is almost non-existent in the mathematical literature on stochastic processes. The more common definition of reversibility is the following.

Definition 5.21. *A Markov chain with matrix of transition probabilities P and initial distribution μ is said to satisfy detailed balance if the equality*

$$\mu(x)P(x, y) = \mu(y)P(y, x) \tag{5.19}$$

holds for any $x, y \in S$.

Proposition 5.22. *For a Markov chain with matrix of transition probabilities P and initial distribution μ, the following are equivalent.*

1. $\mu(x)P(x,y) = \mu(y)P(y,x)$ *for any* $x, y \in S$
2. $\mathbb{P}(X(k+1) = y \mid X(k) = x) = \mathbb{P}(X(k) = y \mid X(k+1) = x)$ *for any* $k \geq 0$ *and* $x, y \in S$.
3. $\mathbb{P}(X(0) = x_0, \ldots, X(k) = x_k) = \mathbb{P}(X(0) = x_k, \ldots, X(k) = x_0)$ *for any* $k \geq 1$ *and* $x_0, \ldots, x_k \in S$.

Proof. If 1. holds, then 3. follows from the decomposition (5.3) and repeated use of $\mu(x)P(x,y) = \mu(y)P(y,x)$. Conversely, if 3. holds, then $\mathbb{P}(X(0) = x, X(1) = y) = \mathbb{P}(X(0) = y, X(1) = x)$ in particular, which is equivalent to $\mu(x)P(x,y) = \mu(y)P(y,x)$. For the equivalence between 1. and 2., if 1. holds, then the Markov chain is stationary, as is clear from (5.5). It then follows that

$$\begin{aligned} \mathbb{P}(X(k) = y \mid X(k+1) = x) &= \mathbb{P}(X(k+1) = x \mid X(k) = y) \frac{\mathbb{P}(X(k) = y)}{\mathbb{P}(X(k+1) = x)} \\ &= P(y,x) \frac{\mu(y)}{\mu(x)} = P(x,y), \end{aligned} \tag{5.20}$$

which is equal to $\mathbb{P}(X(k+1) = y \mid X(k) = x)$ by time-homogeneity. This requires the assumption that $\mu(x) > 0$, which means we actually need to restrict 2. to x such that $\mu(x) > 0$, or restrict the theorem to irreducible Markov chains. Conversely, suppose that 2. holds. Then

$$\mu_{k+1}(x)P(x,y) = \mu_k(y)P(y,x)$$

by (5.20), so that it suffices to prove that the Markov chain is stationary. Condition 2. implies that $\mathbb{P}(X(k) = y \mid X(k+1) = x)$ is independent of k. It then follows from (5.20) that

$$\frac{\mathbb{P}(X(k) = y)}{\mathbb{P}(X(k+1) = x)}$$

is independent of k. If we turn

$$\frac{\mathbb{P}(X(k) = y)}{\mathbb{P}(X(k+1) = x)} = \frac{\mathbb{P}(X(0) = y)}{\mathbb{P}(X(1) = x)},$$

into

$$\mathbb{P}(X(1) = x)\mathbb{P}(X(k) = y) = \mathbb{P}(X(k+1) = x)\mathbb{P}(X(0) = y).$$

and sum over x, then this results in $\mathbb{P}(X(k) = y) = \mathbb{P}(X(0) = y)$, which means the Markov chain is stationary. Again, this reasoning depends on some probabilities being non-zero, which means the theorem is actually not precisely true as stated, but does hold for irreducible Markov chains. □

In view of condition 2 in the previous proposition, another natural name for detailed balance is time-symmetry, which we will use from now on. A time-symmetric Markov chain is necessarily stationary, though the converse does not hold. State-symmetry is a stronger property, it the sense that any stationary state-symmetric Markov chain is time-symmetric. We discuss two problematic aspects about Definition 5.21.

Problem 1. Should reversibility be a property of a stochastic process, or of a set of transition probabilities?

Reversibility according to Definition 5.21 suggests that it is a property of a process, which is necessarily stationary. However, relaxation to equilibrium, the irreversible phenomenon we are interested in, only occurs when the system is initially out of equilibrium, which entails a non-stationary initial distribution. Paraphrasing [Uffink, 2006], Definition 5.21 thus trivializes the problem of reconciling reversibility and irreversibility, since it implies that the two cannot even coexist.

On the other hand, analogy with deterministic dynamics suggests that reversibility should be a property of the transition probabilities, since they are analogous to the equations of motion. We will see in Section 5.3.3 that a direct analogy with Definition 4.7 does not work. Fortunately, there is a simple way of reconciling this intuition with Definition 5.21, which is the following definition.

Definition 5.23. *A matrix of transition probabilities P is said to be time-symmetric if there exists a distribution π such that detailed balance* (5.19) *holds.*

Such a distribution π is necessarily stationary, of course, but we need not set the initial distribution of a process with transition probabilities P equal to it. Most texts on Markov chains define reversibility as a property of processes, for example [Norris, 1997], [Kelly, 1979] and [Brémaud, 2020], though some do use Definition 5.23, such as [Liggett, 2010]. As with many other concepts in theory of Markov chains, time-symmetry in the sense of Definition 5.23 can be expressed using linear algebra. For a given distribution π such that $\pi(x) > 0$ for all $x \in S$, we define the inner product

$$\langle f, g \rangle_\pi = \sum_{x \in S} f(x) g(x) \pi(x)$$

on functions $f, g : S \to \mathbb{R}$. It is then easy to show that P is time-symmetric if and only if there exists a distribution π such that P is self-adjoint with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$, i.e., $\langle Pf, g \rangle_\pi = \langle f, Pg \rangle_\pi$ for all functions f, g. Here Pf is the function $x \mapsto \sum_{y \in S} P(x, y) f(y)$.

Having defined time-symmetry as a property of transition probabilities rather than of processes, there still remains the question whether it has consequences for processes starting from a non-stationary distribution. First, the stationary distribution itself, whether it satisfies detailed balance or not, already determines the recurrence properties of the Markov chain by Lemma 5.11. If detailed balance is satisfied, one additional consequence is that P is self-adjoint and hence has real eigenvalues. These determine the mixing time of the Markov chain, which quantifies how fast an arbitrary initial distribution converges to the stationary distribution. This is particularly useful for Markov chain Monte Carlo algorithms (see [Levin and Peres, 2017]).

Problem 2. The theory of Markov chains is seemingly inherently time-asymmetric.

There are two ways in which the theory may be said to be time-asymmetric. First, Markov chains are usually considered as processes with time taking values in $\mathbb{N}$, which means the process has distinguished initial time and time-direction. If one is interested in the reversibility of Markov chains, it seems more appropriate to consider them as processes with time taking values in $\mathbb{Z}$. In that case, we have the following discrete-time version of a special case of a theorem by Kolmogorov.

Theorem 5.24. *Given a stochastic matrix P, there exists a unique time-homogeneous Markov chain $(X(k))_{k \in \mathbb{Z}}$ with forward transition probabilities given by P if and only if $P^k(x, \cdot)$ converges to some distribution π as $k \to \infty$ for any $x \in S$. In that case, $X(k)$ has distribution π for any $k \in \mathbb{Z}$.*

Proof. See [Kolmogorov, 1936]. □

By Theorem 5.10, it follows that an ergodic Markov chain is necessarily stationary when using $\mathbb{Z}$ as the set of times. Note that this move not only gets rid of asymmetries, but also solves Problem 1, since there is no longer any choice of initial distribution. As for the definition of reversibility in this context, one can again either use state-symmetry or time-symmetry.

The second asymmetry is the fact that Markov chains are constructed using forward transition probabilities P, which means one knows how to evolve a given state into the future, though not immediately how to 'evolve' it back into the past. In the case of deterministic dynamics, the same problem appears, since the map $\varphi : S \to S$ represents evolution in one direction of time. This is solved by assuming that φ is invertible, upon which the question of reversibility becomes the question whether φ and φ^{-1} are equal up to velocity reversal (see Definition 4.7). The analogous solution in the stochastic case is to assume there is a second stochastic matrix Q such that

$$\mu PQ = \mu \quad \text{and} \quad \mu QP = \mu \tag{5.21}$$

for any distribution μ. This implies that Q is the matrix inverse of P, since any row vector can be normalized to a probability distribution. The following proposition states that this can only happen if P is a deterministic matrix (see Definition 5.2), which spoils the solution.

Proposition 5.25. *If P is an invertible stochastic matrix with inverse Q, then Q is stochastic if and only if P is deterministic, in which case Q is also deterministic.*

Proof. For each $x \in S$, there exists a $y \in S$ such that $P(x, y) > 0$. This means that we can choose a function $\varphi : S \to S$ such that $P(x, \varphi(x)) > 0$ for all $x \in S$. Because $PQ = I$, we have

$$\sum_{y \in S} P(x, y)Q(y, z) = 0$$

for all $x \neq z$. If Q is stochastic, the summand is non-negative and hence $P(x, y)Q(y, z) = 0$ for all $x, y, z \in S$ such that $x \neq z$. Setting $y = \varphi(x)$, it follows that $Q(\varphi(x), z) = 0$ for all $x, z \in S$ such that $x \neq z$. In other words, $Q(\varphi(x), \cdot) = \delta_x$. This implies that φ is a bijection (if $\varphi(x) = \varphi(x')$, then $\delta_x = \delta_{x'}$ and hence $x = x'$) and we get $Q(x, \cdot) = \delta_{\varphi^{-1}(x)}$, so that Q is deterministic. By symmetry, P is also deterministic. The converse follows from the same reasoning, since a deterministic matrix is stochastic. □

As with the non-invariance of the Gibbs entropy, the reason why this fails is because stochastic dynamics dissipate information: one cannot perfectly reconstruct the original distribution by applying a backward stochastic matrix, except if the dynamics is deterministic.

However, it is possible to construct a stochastic matrix Q which in some sense is inverse to P, at least in the irreducible case. Let π be the stationary distribution. If $(X(k))_{k \geq 1}$ is a Markov chain with forward transition probabilities P and initial distribution π, then its backward transition probabilities are also time-homogeneous, since

$$\begin{aligned} \mathbb{P}(X(k) = y \mid X(m) = x) &= \frac{\mathbb{P}(X(m) = x \mid X(k) = y)\mathbb{P}(X(k) = y)}{\mathbb{P}(X(m) = x)} \\ &= \frac{P^{m-k}(y, x)\pi(y)}{\pi(x)} \end{aligned}$$

if $k \leq m$. This means that the one-step backward transition probabilities are given by the stochastic matrix Q, defined by

$$Q(x, y) = P(y, x)\frac{\pi(y)}{\pi(x)}. \tag{5.22}$$

This stochastic matrix is a kind of inverse to P, though not in the sense of (5.21). If one wants to know how a stochastic dynamics behaves when going backward in time, one should arguably apply Q. The difference between going forward and backward in time disappears if $Q(x, y) = P(x, y)$, which is equivalent to time-symmetry in the form of detailed balance (5.19).

While the matrix (5.22) is certainly well-defined, it still needs to be mathematically clarified in which sense it represents a backward evolution. Furthermore, there remain philosophical issues with the asymmetry between forward and backward transition probabilities, which is sometimes even taken to be the origin of irreversibility in physics. See for example [Watanabe, 1965], [Arntzenius, 1995] and [Sober, 1993] for opinions of this kind and [Bacciagaluppi, 2010] and [Di Biagio and Rovelli, 2024] for criticisms thereof.

Barring the conceptual problem of inherent asymmetry, we now have two possible definitions of reversibility for stochastic dynamics: state-symmetry (Definition 5.19) and time-symmetry (Definition 5.23). The former is a stronger form of the latter, so that the question becomes whether we should restrict to it or not. The mathematical literature clearly prefers using the more general Definition 5.23, and there are also philosophical reasons for preferring it over Definition 5.19 (see [Holster, 2003]). The main argument for state-symmetry is that it corresponds to the principle of equal a priori probabilities, since a state-symmetric P is doubly stochastic and hence has a uniform stationary distribution by Proposition 5.15. This is a good argument insofar as one believes that physical systems will always satisfy the principle of equal a priori probabilities in equilibrium.

5.3.2 Velocity reversal

In the presence of velocities, the above definitions of reversibility should be modified so as to include the operation of velocity reversal (or more general reversals of state, such as reversal of magnetic fields[4]), which is an involution $T : S \to S$. All states appearing on the right-hand sides of the equations in Definitions 5.19, 5.21 and 5.23 and Propositions 5.20 and 5.22 should have velocity reversal applied to them. In addition, Definitions 5.21 and 5.23 pick up the condition that $\pi(Tx) = \pi(x)$ for all $x \in S$. For any distribution μ, the function $\mu \circ T : x \mapsto \mu(Tx)$ is another distribution, called the velocity reversal of μ.

Loschmidt's paradox, in the sense that reversing velocities will bring the system back to its original state, does not occur for stochastic systems with velocities. This has, once again, the same reason as the non-invariance of the Gibbs entropy: stochastic dynamics inherently dissipate information, which means the past state is forgotten. Consequently, there is no direct analogue of Theorem 4.12.

As an example, let us consider a stochastic model in which there are N particles which can be located on any of M sites on a discrete ring. Each particle has, apart from its position on the ring, a velocity (or rather, direction of motion) that is either -1 or 1, representing respectively clockwise and counter-clockwise motion. The single-particle state space is thus $\{1, \ldots, M\} \times \{-1, 1\}$. At each time step, each particle takes a step in its current direction of motion, after which it reverses its velocity with

[4]For such more general reversals, one uses the term 'time reversal' instead of 'velocity reversal', though we will only use the latter.

probability p. It is not hard to show that this system is state-symmetric, so that the stationary distribution is uniform.

5.3.3 Deterministic dynamics

In Section 5.3.1 we mentioned that direct analogy with deterministic dynamics does not yield a working definition of reversibility of stochastic dynamics. Let us see how. The analogy consists of trying to replicate Definition 4.7 in the stochastic case at the level of probability distributions. A given initial distribution μ evolves into μP over a single time step. Applying P again to the velocity reversal of μP, we might demand that this returns the velocity reversal of μ for any initial distribution, i.e.,

$$\begin{aligned} \mu(Tz) &= \sum_{y\in S} (\mu P)(Ty) P(y,z) \\ &= \sum_{x,y\in S} \mu(x) P(x,Ty) P(y,z) \end{aligned} \tag{5.23}$$

for any $z \in S$. Actually, this tentative condition is more than an analogy. The stochastic matrix P induces a dynamical system Φ on $\operatorname{Prob} S$ given by $\mu \mapsto \mu P$. Condition (5.23) then amounts to Φ being reversible in the sense of Definition 4.7 with velocity reversal given by $\mu \mapsto \mu \circ T$. Making μ equal to $\delta_{x'}$ and setting $z = Tz'$, (5.23) becomes

$$\delta(x', z') = \sum_{y\in S} P(x', Ty) P(y, Tz'),$$

so that the matrix $R(x,y) = P(x,Ty)$, which is obviously also stochastic, satisfies $R^2 = I$, i.e., it is invertible and equal to its own inverse. However, Proposition 5.25 implies that this can only be the case if R is deterministic, in which case P is too. Hence, this approach does not work in general.

A definition of reversibility directly analogous to the deterministic case is thus not possible. However, deterministic dynamics are a special case of stochastic dynamics in the following sense. Any map $\varphi : S \mapsto S$ defines a deterministic matrix P by

$$P(x,y) = \delta(y, \varphi(x)). \tag{5.24}$$

Conversely, if P is a deterministic matrix, there is a unique function $\varphi : S \to S$ such that (5.24) holds. We denote the deterministic matrix corresponding to φ by P_φ. We may now ask whether Definition 5.23 reduces to Definition 4.7 when applied to deterministic matrices. It turns out that it even reduces to the stronger Definition 5.19. This explains the statement made in Section 5.3.1 that state-symmetry is the natural analogue of reversibility in the deterministic setting.

Proposition 5.26. *The mapping $\varphi : S \to S$ is reversible in the sense of Definition 4.7 if and only if P_φ is state-symmetric.*

Proof. Suppose there is an involution $T : S \to S$ such that $\varphi^{-1} = T \circ \varphi \circ T$. We need to show that $P_\varphi(Ty, Tx) = P_\varphi(x,y)$. Using the reversibility of φ, we have

$$\begin{aligned} P_\varphi(Ty, Tx) &= \delta(Tx, (\varphi \circ T)(y)) \\ &= \delta(Tx, (T \circ \varphi^{-1})(y)) \end{aligned}$$

and because T and φ are invertible, this is equal to $\delta(x, \varphi^{-1}(y)) = \delta(\varphi(x), y)$. Conversely, suppose there is an involution T such that $P_\varphi(Ty, Tx) = P_\varphi(x, y)$ holds for all $x, y \in S$. Then

$$\begin{aligned}\delta(y, \varphi(x)) &= \delta(Tx, (\varphi \circ T)(y)) \\ &= \delta(x, (T \circ \varphi \circ T)(y))\end{aligned}$$

implies that φ is invertible with inverse $T \circ \varphi \circ T$, hence reversible. □

In fact, all properties of the Markov chain P_φ should and indeed do reduce to the corresponding properties of the deterministic dynamics φ. As a sample, a distribution π is stationary relative to φ in the sense of Definition 4.3 if and only if it is stationary relative P_φ in the sense of Definition 5.4, which follows from the equality

$$(\pi P_\varphi)(x) = \pi(\varphi^{-1}(\{x\})). \tag{5.25}$$

Finally, we stated at the end Section 5.2.1 that the Gibbs entropy is invariant for any initial distribution if and only if P is deterministic, which we now prove.

Proposition 5.27. *Suppose P is an irreducible Markov chain with stationary distribution π. Then, the Gibbs entropy $H(\mu_k|\pi)$ is time invariant for each initial distribution μ if and only if P is deterministic.*

Proof. Suppose P is deterministic and let $P(x, y) = \delta(y, \varphi(x))$. Because P is irreducible, φ has only a single orbit and hence is invertible. This follows from the equality $P^k(x.y) = \delta(y, \varphi^k(x))$. It is easily checked that $\mu P = \mu \circ \varphi^{-1}$. If μ, ν are probability measures such that $\nu(x) > 0$ for all $x \in S$, then

$$\begin{aligned}H(\mu P|\nu P) &= \sum_{x \in S} \mu(\varphi^{-1}(x)) \log\left(\frac{\mu(\varphi^{-1}(x))}{\nu(\varphi^{-1}(x))}\right) \\ &= \sum_{x \in S} \mu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right) = H(\mu|\nu)\end{aligned}$$

and this yields the invariance of the Gibbs entropy. Conversely, suppose $H(\mu_k|\pi)$ is constant for any given initial distribution μ, i.e., $H(\mu P|\pi) = H(\mu|\pi)$ for all μ. This means that Jensen's inequality (5.9) is an equality. If μ is chosen such that $\rho(x)$ is different for each $x \in S$, then the equality can only hold if there is only a single non-zero weight, which implies $P(x, \cdot)$ is deterministic. □

5.4 Macroscopic behaviour

We have seen that the empirical distribution $f(k)$, which is the analogue of Boltzmann's distribution function, exhibits relaxation to equilibrium when the Ehrenfest model starts out in a non-equilibrium macrostate. Now we would like to have an equation that describes this macroscopic evolution. As mentioned before, since $f(k)$ is a random variable, it does not satisfy a deterministic evolution equation. However, the expected value $\mathbb{E}(f(k))$ does, and we will see that $f(k)$ itself almost surely does too, upon taking a thermodynamic limit $N \to \infty$. Such a limit is only possible for the variant of the Ehrenfest model discussed in Section 5.2.2, to which we restrict unless stated otherwise. We will also always assume that the initial distribution is of the form (5.17), such that the processes $(X_n(k))_{k \geq 1}$ are independent.

5.4.1 Expectation

As with the Kac ring, the expected macroscopic behaviour can first be derived non-rigorously using an analogue of the Stosszahlansatz. For the Ehrenfest model, this analogue is given by the assumption that at each time step,

$$\text{proportion of particles in state 1 which change state} = pf(k) \tag{5.26}$$

and

$$\text{proportion of particles in state 0 which change state} = p(1-f(k)). \tag{5.27}$$

Again, it is easy to give examples of trajectories for which (5.27) and (5.26) do not hold, but we expect them to typically hold. The change in the proportion of particles in state 1 during the time step from k to $k+1$ is given by the difference between (5.27) and (5.26), which results in the Ehrenfest model's analogue of the Boltzmann equation:

$$f(k+1) - f(k) = p(1-2f(k)).$$

This can be made rigorous by working with the expectation $F(k) = \mathbb{E}(f(k))$. To determine the equation governing its evolution, we take advantage of the fact that each of the single-particle processes $(X_n(k))_{k\geq 1}$ is itself a Markov chain with transition probabilities (5.18). First, it follows from equation (5.13) and the linearity of expectation that

$$F(k) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}(X_n(k)) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{P}(X_n(k)=1), \tag{5.28}$$

the second equality holding because $X_n(k)$ takes values in $\{0,1\}$. The probabilities $\mathbb{P}(X_n(k)=1)$ satisfy the master equation of the two-state Markov chain (5.18), which is given by

$$\begin{aligned}\mathbb{P}(X_n(k+1)=1) - \mathbb{P}(X_n(k)=1) = \;& p\mathbb{P}(X_n(k)=0) + (1-p)\mathbb{P}(X_n(k)=1) \\ & - p\mathbb{P}(X_n(k)=1) - (1-p)\mathbb{P}(X_n(k)=1) \\ = \;& p(1-2\mathbb{P}(X_n(k)=1)),\end{aligned} \tag{5.29}$$

using $\mathbb{P}(X_n(k)=0) = 1-\mathbb{P}(X_n(k)=1)$ for the second equality. Summing over n and dividing by N, it follows from (5.28) combined with (5.29) that $F(k)$ satisfies the difference equation

$$F(k+1) - F(k) = p(1-2F(k)), \tag{5.30}$$

which is the rigorously derived analogue of the Boltzmann equation. It follows from standard techniques for solving difference equations that its solution is given by

$$F(k) = \frac{1}{2} + \left(F(0) - \frac{1}{2}\right)(1-2p)^k, \tag{5.31}$$

which exhibits relaxation to the equilibrium value $1/2$ as long as $p \in (0,1)$, monotonically if $p < 1/2$. Hence, on the average, the Ehrenfest model relaxes. However, as in the case of the Kac ring, one wants to prove something stronger, which is that actual realizations of the model typically relax if the number of particles N is large, i.e., $f(k)$ is close to its expected value $F(k)$. This is a law of large numbers with time-dependence, which we will formulate more precisely and prove in Sections 5.4.3 and 5.4.4.

This will require a bound on the variance of $f(k)$, which we now calculate. Because the processes $(X_n(k))_{1\leq n\leq N}$ are independent, we may use equation (B.1) together with a standard formula for the variance to write

$$\begin{aligned}\mathbb{V}(f(k)) &= \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{V}(X_n(k)) \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\left(\mathbb{E}(X_n(k)^2) - \mathbb{E}(X_n(k))^2\right) \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\left(\mathbb{P}(X_n(k)=1) - \mathbb{P}(X_n(k)=1)^2\right),\end{aligned}$$

the last equality holding because $X_n(k)\in\{0,1\}$. Solving the master equation (5.33) and substituting the solution into the previous expression leads to

$$\mathbb{V}(f(k)) = \frac{1}{4N} - \frac{1}{N^2}\sum_{n=1}^{N}\left(\mathbb{P}(X_n(0)=1) - \frac{1}{2}\right)^2 (1-2p)^{2k},$$

which gives the bound

$$\mathbb{V}(f(k)) \leq \frac{1}{4N}. \tag{5.32}$$

5.4.2 Interaction

Before moving on to the time-dependent law of large numbers, let us make some comments about the evolution equation (5.30) and how such equations can be derived for other systems. Note that (5.30) is identical to the master equation (5.29) of the single-particle Markov chain (5.18). This is true for any two-state[5] population Markov chain such that all processes $(X_n(k))_{k\geq 1}$ are Markov chains with identical transition probabilities, since (5.28) still holds.

If interactions are present, then the single-particle processes $(X_n(k))_{k\geq 1}$ usually are no longer Markov chains themselves, since the state of a single particle is not enough to predict its evolution. However, if we assume that these processes are identically distributed at each $k\geq 1$, say with common distribution ν_k, then from (5.28) follows $F(k)=\nu_k(1)$, so that $F(k)$ has the same evolution as the probability distribution of a single particle. A natural case in which this happens is when the transition probabilities P and the initial distribution μ are symmetric, i.e., invariant under permutation of the particles,[6] since the master equation (5.5) then implies that μ_k is symmetric at each $k\geq 1$. Physically, this corresponds to the particles being indistinguishable.

The question remains how to determine the evolution equation of the single-particle distribution ν_k. If the transition probabilities have the property that

$$Q(x,y_1) = \sum_{y_2,\ldots,y_N\in A} P(x_1,\ldots,x_N,y_1,\ldots,y_N)$$

[5] We only consider population Markov chains with single-particle state space $\{0,1\}$ here, but similar statements hold for more general single-particle state spaces.

[6] Defining the action of a permutation σ on $x\in 2^N$ by $\sigma x = (x_{\sigma(1)},\ldots,x_{\sigma(N)})$, invariance under permutations of the particles means that $\mu(\sigma x)=\mu(x)$ and $P(\sigma x,\sigma y)=P(x,y)$ for all $x,y\in 2^N$ and any permutation σ.

only depends on x_1 and y_1, in which case we denote it by $Q(x_1, y_1)$, then one obtains the evolution equation of ν_k by summing over $y_2, \ldots, y_N$ in the master equation (5.5), which yields the linear difference equation

$$\nu_{k+1}(y_1) - \nu_k(y_1) = \sum_{x_1 \in A} (\nu_k(x_1)Q(x_1, y_1) - \nu_k(y_1)Q(y_1, x_1)).$$

This assumption typically holds if only a single particle changes state at each time step. As an example, let us go through this procedure for original Ehrenfest model, which has non-independent particles and obviously symmetric transition probabilities. The master equation is given by

$$\begin{aligned} \mu_{k+1}(y) - \mu_k(y) &= \sum_{x \in 2^N} (\mu_k(x)P(x, y) - \mu_k(y)P(y, x)) \\ &= \frac{1}{N} \sum_{n=1}^{N} (\mu_k(y^n) - \mu_k(y)), \end{aligned} \tag{5.33}$$

with y^n the result of changing the state of particle n in microstate y. Let ν_k again denote the marginal distribution of μ_k with respect to a single particle, say the first. Summing over all particles expect the first in (5.33) yields

$$\begin{aligned} \nu_{k+1}(y_1) - \nu_k(y_1) &= \frac{1}{N} (\nu_k(1 - y_1) + (N - 1)\nu_k(y_1)) - \nu_k(y) \\ &= \frac{1}{N} (\nu_k(1 - y_1) - \nu_k(y_1)) \\ &= \frac{1}{N} (1 - 2\nu_k(y_1)), \end{aligned}$$

which is the evolution equation of all single-particle marginals as well as the empirical distribution $F(k)$ of the original Ehrenfest model. Note that if we set $y_1 = 1$, then this equation coincides with (5.29) if $p = 1/N$. In this sense, the variant of the Ehrenfest model reduces to the original upon setting $p = 1/N$ (see also [Hauert et al., 2004]).

For population Markov chain models in kinetic theory, it is more realistic that two particles should change state at each time step, representing binary collisions. In that case,

$$Q(x, y_1, y_2) = \sum_{y_3, \ldots, y_N \in A} P(x_1, \ldots, x_N, y_1, \ldots, y_N)$$

should depend only on x_1, x_2 and y_1, y_2, which we then write as $Q(x_1, x_2, y_1, y_2)$. Summing over $y_2, \ldots, y_N$ in the master equation (5.5) now only results in

$$\nu_{k+1}(y_1) - \nu_k(y_1) = \sum_{x_1, x_2, y_2 \in A} (\rho_k(x_1, x_2)Q(x_1, x_2, y_1, y_2) - \rho_k(y_1, y_2)Q(y_1, y_2, x_1, x_2)), \tag{5.34}$$

with ρ_k the common distribution of all pairs $(X_n(k), X_m(k))$ such that $n \neq m$, which are likewise identically distributed assuming symmetry. We would prefer to have ν_k instead of ρ_k on the right-hand side of this equation. If the stochastic dynamics are such that

$$\rho_0(x_1, x_2) = \nu_0(x_1)\nu_0(x_2) \quad \text{implies} \quad \rho_k(x_1, x_2) = \nu_k(x_1)\nu_k(x_2) \tag{5.35}$$

for all $k \geq 1$, then (5.34) turns into

$$\nu_{k+1}(y_1) - \nu_k(y_1) = \sum_{x_1,x_2,y_2 \in A} (\nu_k(x_1)\nu_k(x_2)Q(x_1,x_2,y_1,y_2) - \nu_k(y_1)\nu_k(y_2)Q(y_1,y_2,x_1,x_2)), \quad (5.36)$$

when restricting to initial distributions such that $\rho_0(x_1,x_2) = \nu_0(x_1)\nu_0(x_2)$ holds. Equation (5.36) is a non-linear evolution equation for ν_k, analogous to the Boltzmann equation. The property (5.35) is known as propagation of chaos and was first introduced in [Kac, 1956]. The Stosszahlansatz is often interpreted as this factorization condition $\rho_k(x_1,x_2) = \nu_k(x_1)\nu_k(x_2)$ (see [Uffink, 2006]).

5.4.3 Weak law of large numbers

The law of large numbers comes in a weak version and a strong version. We begin by proving the former. Note that it cannot be proven that $f(k)$ stays close to $F(k)$ at all times with high probability. In fact, this probability is zero because the Ehrenfest model is recurrent. Instead, we will at first fix a finite time interval, say T time steps. Let

$$A_k = \{|f(k) - F(k)| < \varepsilon\},$$

suppressing the dependence of $f(k)$ and A_k on N to ease the notation. We aim to show that the probability $\mathbb{P}\left(\cap_{k=0}^{T} A_k\right)$ is high for large N. Because of the fixed time interval, this simply follows from the usual weak law of large numbers for each random variable $f(k)$ together with the bound

$$\mathbb{P}\left(\bigcap_{k=0}^{T} A_k\right) \geq 1 - \sum_{k=0}^{T} \mathbb{P}\left(A_k^c\right), \quad (5.37)$$

since each $\mathbb{P}(A_k^c)$ individually converges to zero as $N \to \infty$. If the terminal time T_N grows with N and the error ε_N decreases, it is still possible to prove a law of large numbers through (5.37) by making use of explicit bounds on $\mathbb{P}(A_k^c)$.

Proposition 5.28. *For any sequences $(T_N)_{N\geq 1}$ in $\mathbb{N}$ and $(\varepsilon_N)_{N\geq 1}$ in $(0,\infty)$, we have the bound*

$$\mathbb{P}\left(\bigcap_{k=0}^{T_N} A_k\right) \geq 1 - 2T_N \exp(-2N\varepsilon_N^2). \quad (5.38)$$

Proof. The random variables $(X_n(k))_{1\leq n\leq N}$ are independent for fixed k and take values in $[0,1]$, so that we may use Hoeffding's inequality (Proposition B.4) to obtain (5.38). □

For the specific case of power laws $\varepsilon_N = 1/N^a$ and $T_N = \lfloor N^b \rfloor$, the second term on the right-hand side of (5.38) converges to zero if and only if $a < 1/2$. We can even make the terminal time grow exponentially in N. For example, letting $\varepsilon_N = \varepsilon$ be constant and setting $T_N = \exp(N\varepsilon^2)$ turns the right-hand side of (5.38) into $1 - \exp(-N\varepsilon^2)$. However, if ε is small, which we would like it to be, then both the time interval and bound grow very slowly, making them of little practical use.

Following [Baldovin et al., 2019], we can obtain better bounds by exploiting the fact that $(f(k))_{k\geq 1}$ is a Markov chain, which we did not do in the above method based on Hoeffding's inequality.

Theorem 5.29. *For any sequences $(T_N)_{N\geq 1}$ in $\mathbb{N}$ and $(\varepsilon_N)_{N\geq 1}$ in $(0,\infty)$, the bound*

$$\mathbb{P}\left(\bigcap_{k=0}^{T_N} A_k\right) \geq \left(1 - 2\exp(-2N\varepsilon_N^2)\right)\left(1 - \frac{1}{N\varepsilon_N^2}\right)^{T_N} \tag{5.39}$$

holds for N large enough that $\mathbb{P}(A_0) \geq 1/2$.

Proof. It follows from the Markov property that

$$\mathbb{P}\left(\bigcap_{k=0}^{T_N} A_k\right) = \mathbb{P}(A_0)\prod_{k=1}^{T_N}\mathbb{P}(A_k|A_{k-1}). \tag{5.40}$$

The probability $\mathbb{P}(A_0)$ can be bounded using Hoeffding's inequality, so it remains to bound the transition probabilities appearing in (5.40). Using elementary probability theory and the Markov property once again,

$$\begin{aligned}\mathbb{P}(A_k|A_0) &= \mathbb{P}(A_k|A_{k-1}\cap A_0)\mathbb{P}(A_{k-1}|A_0) + \mathbb{P}(A_k|A_{k-1}^c\cap A_0)\mathbb{P}(A_{k-1}^c|A_0)\\ &= \mathbb{P}(A_k|A_{k-1})\mathbb{P}(A_{k-1}|A_0) + \mathbb{P}(A_k|A_{k-1}^c)\mathbb{P}(A_{k-1}^c|A_0)\\ &\leq \mathbb{P}(A_k|A_{k-1}) + \mathbb{P}(A_{k-1}^c|A_0)\end{aligned}$$

and from this follows the bound

$$\mathbb{P}(A_k|A_{k-1}) \geq 1 - \big(\mathbb{P}(A_k^c|A_0) + \mathbb{P}(A_{k-1}^c|A_0)\big) \tag{5.41}$$

on transition probabilities. To bound the two terms between the parentheses, note that

$$\mathbb{P}(A_k^c|A_0) = \mathbb{P}(A_k^c\cap A_0)/\mathbb{P}(A_0) \leq \mathbb{P}(A_k^c)/\mathbb{P}(A_0)$$

by monotonicity of the probability measure. Assuming N is large enough that $\mathbb{P}(A_0)\geq 1/2$, Chebyshev's inequality (Proposition B.2) combined with the bound (5.32) on the variance of $f(k)$ results in

$$\mathbb{P}(A_k^c|A_0) \leq 1/2N\varepsilon_N^2.$$

Using this in (5.41) gives

$$\mathbb{P}(A_k|A_{k-1}) \geq 1 - \frac{1}{N\varepsilon_N^2},$$

which together with Hoeffding's inequality for $f(0)$ and (5.40) finally implies that (5.39) holds for N large enough that $\mathbb{P}(A_0)\geq 1/2$. □

Specializing again to the case of power laws $\varepsilon_N = 1/N^a$ and $T_N = \lfloor N^b\rfloor$, the right-hand side of (5.39) is bounded from below by

$$\left(1 - 2\exp(-2N^{1-2a})\right)\left(1 - \frac{1}{N^{1-2a}}\right)^{N^b} \tag{5.42}$$

assuming that $a < 1/2$. Using the Taylor expansion $\log(1+x) = x + o(x)$, the second factor in (5.42) can be expressed as

$$\exp\left(-N^{2a+b-1} + o(N^{2a+b-1})\right)$$

and hence converges to 1 if and only if $2a+b<1$. It follows that (5.42) converges to 1 as well. Even though the convergence is not exponential, it is still much better than what one gets from (5.38).

5.4.4 Strong law of large numbers

To formulate a strong law of large numbers for the Ehrenfest model it is necessary to assume that all random variables $X_n(k)$ are defined on the same probability space. Our construction of Markov chains given in Section 5.1.1 is only valid for finite state spaces, hence is not applicable here. However, one can still use the Kolmogorov representation to construct an appropriate probability space. Since the specific probability space on which the $X_n(k)$ exist is not at all important for the law of large numbers, we leave this construction to the next section, in which it is more important.

In the Ehrenfest model with an infinite number of particles, the relevant macroscopic quantity is now given by

$$f(k) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_n(k),$$

which need not necessarily converge. However, assuming a convergence condition on the initial distributions, we will show that it almost surely converges for all $k \geq 1$. It is convenient to introduce the truncations

$$f_N(k) = \frac{1}{N} \sum_{n=1}^{N} X_n(k) \quad \text{and} \quad F_N(k) = \mathbb{E}(f_N(k)).$$

Theorem 5.30. *Suppose that $F_N(0)$ converges to some number $F(0)$ as $N \to \infty$. If we define $F(k)$ by equation (5.31) and fix $k \geq 0$, then almost surely*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X_n(k) = F(k). \tag{5.43}$$

That is to say, the macroscopic quantity $f(k)$ becomes deterministic in the thermodynamic limit $N \to \infty$, satisfying our analogue of the Boltzmann equation. This illustrates how deterministic macroscopic laws may emerge from stochastic microscopic laws as a result of the law of large numbers.

Proof. The proof is similar to that of the strong law of large numbers in Section 3.2. Using the reasoning of Section 5.4.1, F_N satisfies (5.31) and hence $F_N(k) \to F(k)$. By the triangle inequality, $|f_N(k) - F(k)| \geq 2/m$ implies $|f_N(k) - F_N(k)| \geq 1/m$ for N large enough that $|F_N(k) - F(k)| < 1/m$. Using monotonicity and the fourth moment bound (Proposition B.3), it follows that

$$\begin{aligned} \mathbb{P}(|f_N(k) - F(k)| \geq 2/m) &\leq \mathbb{P}(|f_N(k) - F_N(k)| \geq 1/m) \\ &\leq m^4 \mathbb{F}(f_N(k)) \\ &\leq Cm^4/N^2 \end{aligned}$$

for sufficiently large N, with some constant $C \geq 0$, since the random variables $X_n(k)$ are independent and uniformly bounded (see equation (B.2)). It now follows from the Borel-Cantelli lemma that $|f_N(k) - F(k)| < 2/m$ almost surely eventually holds for any $m \geq 1$, from which (5.43) follows as in the proof of the strong law of large numbers in Section 3.2. □

Note that in Theorem 5.30, the initial distribution of the process is arbitrary, except that its truncations are of the form (5.17) and have convergent expectations at the initial time. Thus while we could use

the same physically relevant, entropy-maximizing distribution $\mathbb{P}(X_n(0) = 1) = \alpha$ which we used in Section 4.3, we could also, at the other entropy-extreme, concentrate the initial distribution on a single microstate $x \in 2^{\mathbb{Z}_+}$ such that $N^{-1}\sum_{n=1}^{N} x_n$ converges. Thus, in contrast to deterministic dynamics, every initial state with well defined macroscopic value leads to relaxation.

5.4.5 Randomness

Having proven the strong law of large numbers with explicit and summable bounds, it is now a simple application of Solovay randomness to prove the corresponding randomness statement. However, while in probability theory the specific underlying probability space is almost never important, it is in the theory of algorithmic randomness, if only because one needs to speak of the possible computability of its subsets. Hence we first take some time to construct a proper probability space for the Ehrenfest model with infinitely many particles.

The Ehrenfest model with an infinite number of particles is the process $X(k) = (X_n(k))_{n\geq 1}$ with state space $2^{\mathbb{Z}_+}$ such that $X_n(k)$ are independent copies of the Markov chain (5.18). The natural probability space is the Kolmogorov representation $(2^{\mathbb{Z}_+})^{\mathbb{N}}$, with probability measure $\mathbb{P}$ making the random variables $(X_n(k))_{n\geq 1}$ independent for each $k \in \mathbb{N}$ and which makes $(X_n(k))_{k\geq 0}$ a Markov chain for each $n \in \mathbb{Z}_+$ with transition probabilities (5.18). Our earlier construction is not applicable, since the state space is uncountably infinite. While it is possible to construct Markov chains like we did earlier for more general states spaces (see Section 4.1 in [Kurtz and Ethier, 1986]), we will take a different approach here.

For each $n \in \mathbb{Z}_+$, let $\mathbb{P}_n$ be the Kolmogorov representation (see Appendix A) of the Markov chain $X_n(k)$ on $2^{\mathbb{N}}$, with arbitrary initial distribution, the dependence on which we suppress. By Theorem A.5, there is a unique probability measure $\mathbb{P}$ on the product space $(2^{\mathbb{N}})^{\mathbb{Z}_+}$, equipped with the cylindrical σ-algebra it inherits from the same on $2^{\mathbb{N}}$, such that

$$\mathbb{P}\left(\bigcap_{n\in F} X_n^{-1}(A_n)\right) = \prod_{n\in F} \mathbb{P}_n(A_n),$$

with $X_n : (2^{\mathbb{N}})^{\mathbb{Z}_+} \to 2^{\mathbb{N}}$ the projection onto the n'th factor and $A_n \subseteq 2^{\mathbb{N}}$ measurable sets indexed by a finite subset $F \subseteq \mathbb{Z}_+$. Appealing to Proposition A.6, the obvious bijection $(2^{\mathbb{N}})^{\mathbb{Z}_+} \cong (2^{\mathbb{Z}_+})^{\mathbb{N}}$ is actually an isomorphism of measurable spaces, hence we may view $\mathbb{P}$ as a probability measure on $(2^{\mathbb{Z}_+})^{\mathbb{N}}$.

Proposition 5.31. *The probability measure $\mathbb{P}$ on $(2^{\mathbb{Z}_+})^{\mathbb{N}}$ makes the random variables $X(k)$ independent for each fixed $k \in \mathbb{N}$ and makes each $X_n(k)$ with $n \in \mathbb{Z}_+$ fixed a Markov chain with transition probabilities given by* (5.18).

Proof. The random variables $X(k)$ are independent by construction (see the remarks following Theorem A.5), hence so are $X_n(k)$, since they are measurable functions of them. To show that $X_n(k)$ is a copy of the Markov chain 5.18 for each $n \in \mathbb{Z}_+$, we use Proposition 5.3. For any $a_k \in \{0, 1\}$, the event $X_n(k) = a_k$ is equal to the event $X_n \in A_k$, where $A_k = \{x \in 2^{\mathbb{N}} \mid x(k) = a_k\}$. Hence, for any fixed

$n \in \mathbb{Z}_+$ we have

$$\begin{aligned}\mathbb{P}(X_n(T) = a_T, \dots, X_n(0) = a_0) &= \mathbb{P}(X_n \in A_T, \dots, X_n \in A_0) \\ &= \mathbb{P}(X_n^{-1}(A_0 \cap \cdots \cap A_T)) \\ &= \mathbb{P}_n(A_0 \cap \cdots \cap A_T) \\ &= p_n(a_0)P(a_0, a_1) \cdots P(a_{T-1}, a_T).\end{aligned}$$

□

We make the probability space $((2^{\mathbb{Z}_+})^{\mathbb{N}}, \mathbb{P})$ effective by choosing some enumeration of the cylinder sets of $2^{\mathbb{Z}_+ \times \mathbb{N}} \cong (2^{\mathbb{Z}_+})^{\mathbb{N}}$. This completes our construction of an appropriate effective probability space for the infinite Ehrenfest model. Its elements are sequences in the microstate space $2^{\mathbb{Z}_+}$ and the following theorem says that such a sequence follows our analogue (5.30) of the Boltzmann equation if it is $\mathbb{P}$-random, under the same condition on the initial distributions as in Theorem 5.30.

Theorem 5.32. *Suppose that the initial distributions are such that $F_N(0)$ converges to some computable number $F(0)$. Fixing $k \geq 1$ and defining $F(k)$ by equation (5.31), any $\mathbb{P}$-random $x \in (2^{\mathbb{Z}_+})^{\mathbb{N}}$ satisfies*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} x_n(k) = F(k). \tag{5.44}$$

Proof. The exception sets $E_{N,m} = \{|f_N(k) - F(k)| > 2/m\}$ are obviously computable and summable in N by the inequalities in the proof of Theorem 5.30, hence form a Solovay test. By Proposition 3.7, if $x \in (2^{\mathbb{Z}_+})^{\mathbb{N}}$ is $\mathbb{P}$-random, then $x \in E_{N,m}$ for finitely many N. Since this holds for all $m \geq 1$, this implies (5.44) as in the proof of the effective strong law of large numbers in Section 3.2. □

With this theorem, we have shown that Martin-Löf randomness may also be applied to a system with stochastic dynamics, in the sense that its macroscopic evolution holds for random trajectories of microstates. We could also look at the other two aspects of randomness we looked at for the Kac chain, i.e., characterization of randomness through the ergodic theorem and the role of the Stosszahlansatz, but this is similar to the case of the Kac chain, so we leave the details for another time. Let us now note the important differences with deterministic dynamics:

- With deterministic dynamics, one uses probability measures on initial microstates, while stochastic dynamics require probability measures on trajectories of microstates.
- With deterministic dynamics, not all initial microstates lead to correct macroscopic behaviour. In the case of stochastic dynamics, any initial microstate leads to correct macroscopic behaviour with high probability. The analogue of bad initial microstates is given by bad trajectories.
- For stochastic dynamics with velocities, velocity reversal at a single time does not reverse the macroscopic evolution.

5.5 Continuous-time

The Kac chain model used in [Hiura and Sasa, 2019] is a two-state system in discrete-time, making the appearance of the Cantor space $2^{\mathbb{N}}$ natural and the basic theory of Martin-Löf randomness almost

immediately applicable. We tried to follow them in this by using the discrete-time Ehrenfest model, which is also a two-state system. However, we get a closer analogy with the Boltzmann's kinetic theory by working in continuous-time, which we do in this section.

5.5.1 Continuum limit

The Boltzmann equation is formulated in terms of a continuous time variable, while our analogue (5.30) and the Ehrenfest model itself are in discrete-time. We can turn our analogue into a continuous-time equation by a continuum limit, in which we make both the time step and transition probability p scale with the number of particles N and let $N \to \infty$. Making the time step equal to $1/N$ and setting $p = \lambda/N$, equation (5.30) becomes

$$F_N\left(\frac{k+1}{N}\right) - F_N\left(\frac{k}{N}\right) = \frac{\lambda}{N}\left(1 - 2F_N\left(\frac{k}{N}\right)\right), \tag{5.45}$$

with F_N now a function on the rational numbers of the form k/N instead of integers. We extend F_N to all other times by linear interpolation, making it a continuous function. As N increases, the number of discrete time points in any given finite interval becomes denser and we expect (5.45) to turn into a differential equation. Rewriting (5.45) as

$$\frac{F_N(t+1/N) - F_N(t)}{1/N} = \lambda(1 - 2F_N(t)) \tag{5.46}$$

with $t = k/N$, we can conjecture that the functions F_N should converge to solutions of the differential equation $F'(t) = \lambda(1 - 2F(t))$, which is the continuous analogue of (5.30). The following proposition makes this precise.

Proposition 5.33. *If for each N the function F_N satisfies the difference equation* (5.45) *and the sequence $F_N(0)$ converges to some number $F(0)$, then $\lim_{N\to\infty} F_N(t) = F(t)$ exists for each $t \geq 0$ and is a differentiable function satisfying the continuous-time analogue of the Boltzmann equation:*

$$F'(t) = \lambda(1 - 2F(t)). \tag{5.47}$$

Proof. Since F_N is defined by linearly interpolating between its values at discrete time points, we can write

$$\begin{aligned} F_N(t) &= (1 - Nt + \lfloor Nt \rfloor)F_N\left(\frac{\lfloor Nt \rfloor}{N}\right) + (Nt - \lfloor Nt \rfloor)F_N\left(\frac{\lfloor Nt \rfloor + 1}{N}\right) \\ &= F_N\left(\frac{\lfloor Nt \rfloor}{N}\right) + (Nt - \lfloor Nt \rfloor)\left(F_N\left(\frac{\lfloor Nt \rfloor + 1}{N}\right) - F_N\left(\frac{\lfloor Nt \rfloor}{N}\right)\right) \end{aligned}$$

for any $t \geq 0$. Rewriting the last term using (5.46), we get

$$F_N(t) = F_N\left(\frac{\lfloor Nt \rfloor}{N}\right) + (Nt - \lfloor Nt \rfloor)\frac{\lambda}{N}\left(1 - 2F_N\left(\frac{\lfloor Nt \rfloor}{N}\right)\right).$$

It follows from the solution of (5.46) that F_N is bounded. Since $Nt - \lfloor Nt \rfloor$ is obviously also bounded and $\lambda/N \to 0$, it follows that $\lim_{N\to\infty} F_N(t) = \lim_{N\to\infty} F_N(\lfloor Nt \rfloor/N)$, assuming the latter limit exist.

It remains to show that this limit indeed exists. It follows from (5.31) that

$$F_N\left(\frac{\lfloor Nt\rfloor}{N}\right) = \frac{1}{2} + \left(F_N(0) - \frac{1}{2}\right)\left(\left(1 - \frac{2\lambda}{N}\right)^N\right)^{\frac{\lfloor Nt\rfloor}{N}},$$

which converges to

$$\frac{1}{2} + \left(F(0) - \frac{1}{2}\right)e^{-2\lambda t}$$

by the assumption $F_N(0) \to F(0)$ and the standard limits $(1 + x/N)^N \to e^x$ and $\lfloor Nt\rfloor/N \to t$. One easily checks that $F(t)$ satisfies the differential equation (5.47). □

5.5.2 Markov processes

The continuum limit $F(t) = \lim_{N\to\infty} F_N(t)$ disregards the fact that F_N and its equation (5.46) arise from a discrete-time stochastic process. It would be more fundamental to show that this stochastic process itself has a continuum limit in a certain sense. However, instead of taking a continuum limit of a discrete-time process, we could also work exclusively with continuous-time processes from the start. This is the more usual approach in the mathematical literature, for example in the theory of interacting particle systems. While continuous-time processes require more measure-theoretic machinery to set up, the theory is mathematically much more convenient than the discrete-time counterpart. As before, we restrict to processes satisfying the Markov property.

Definition 5.34. *A continuous-time stochastic process $(X(t))_{t\geq 0}$ on S is said to be a Markov process if for all $s, t \geq 0$ with $t \leq s$ the equality*

$$\mathbb{P}(X(s) \in \Gamma \mid \mathcal{F}_t) = \mathbb{P}(X(s) \in \Gamma \mid X(t))$$

holds for all measurable $\Gamma \subset S$. Here $\mathcal{F}_s = \sigma(X(t), 0 \leq t \leq s)$ is the canonical filtration generated by the process $(X(t))_{t\geq 0}$.

Markov processes with continuous sample paths are called diffusions.[7] Important examples are solutions of stochastic differential equations, which stem from the work of Einstein, Smoluchowski and Langevin on Brownian motion, and which were rigorously defined by Itô. We do not look at these kinds of processes and instead restrict to Markov processes with discrete state space. For these, the Markov property is easier to check according to the following proposition which we state without proof.

Proposition 5.35. *Suppose $(X(t))_{t\geq 0}$ is a continuous time stochastic process with countable state space S. If the equality*

$$\mathbb{P}(X(s) = x_s \mid X(t_0) = x_0, \ldots, X(t_n) = x_n) = \mathbb{P}(X(s) = x_s \mid X(t_n) = x_n) \tag{5.48}$$

holds for all $s, t_0, \ldots, t_n \geq 0$ and $x_s, x_0, \ldots, x_n \in S$ such that these conditional probabilities exist (which is the case if and only if $\mathbb{P}(X(t_0) = x_0, \ldots, X(t_n) = x_n) \neq 0$), then $(X(t))_{t\geq 0}$ is a Markov process.

Analogous to the discrete-time case, the probability mass function of $X(0)$ is called the initial distribution and the probabilities $\mathbb{P}(X(t) = y \mid X(s) = x)$ are called the transition probabilities. We restrict to time-homogeneous Markov processes, for which these depend on the times s, t only through their

[7] Actually, one often includes the strong Markov property in the definition of a diffusion.

difference $s - t$. In that case, the transition function is defined by $P_t(x, y) = \mathbb{P}(X(t) = y \mid X(0) = x)$, which we see as a time-dependent $|S| \times |S|$ matrix. The Chapman-Kolmogorov equations

$$P_{s+t}(x, z) = \sum_{y \in S} P_s(x, y) P_t(y, z)$$

are proven as in the discrete-time case and may be expressed in matrix notation as $P_{s+t} = P_s P_t$, which is called the semigroup property. Together with the obvious equality $P_0 = I$ and the fact that each matrix P_t is stochastic, this means that the operators $(P_t)_{t \geq 0}$ form a matrix-semigroup of stochastic matrices, also known as a Markov matrix-semigroup.

Unlike in the discrete-time case, a general Markov process on a discrete state space could be badly behaved, for example jumping infinitely often in each arbitrarily short time interval (see Section 2.4 in [Liggett, 2010]). This motivates the following definition.

Definition 5.36. *A Markov jump process*[8] *is Markov process with discrete state space and right-continuous sample paths.*

The discreteness of S together with the right-continuity of $(X(t))_{t \geq 0}$ imply that the process actually waits a finite time in each state before jumping to another state. Thus, as a function of the sample path, one has jumping times J_k at which the state changes and holding times $S_k = J_k - J_{k-1}$ giving the duration of each visited state. The Markov property together with the assumption of time-homogeneity imply that the holding time S_k is memoryless, hence exponentially distributed, conditional on $X(J_{k-1})$ having some fixed value. This means that for each $x \in S$, there is a parameter $\lambda(x) \in [0, \infty)$ such that

$$\mathbb{P}(S_k > t \mid X(J_{k-1}) = x) = e^{-\lambda(x) t}. \tag{5.49}$$

Thus, a Markov jump process stays in a state for an exponentially distributed amount of time and then jumps to another state. Conditional on $X(J_{k-1})$ having some fixed value, the holding time S_k and the new state $X(J_k)$ are independent, i.e.,

$$\begin{aligned} \mathbb{P}(X(J_k) = y, S_k \leq t \mid X(J_{k-1}) = x) &= \mathbb{P}(X(J_k) = y \mid X(J_{k-1}) = x) \mathbb{P}(S_k \leq t \mid X(J_{k-1}) = x) \\ &= a(x, y) \left(1 - e^{-\lambda(x) t}\right). \end{aligned} \tag{5.50}$$

Here we defined $a(x, y) = \mathbb{P}(X(J_k) = y \mid X(J_{k-1}) = x)$, which are the transition probabilities of the discrete-time stochastic process $Y(k) = X(J_k)$. It easily follows from the Markov property of $(X_t)_{t \geq 0}$ that $(Y(k))_{k \geq 0}$ is a Markov chain, called the jump chain.

In Section 5.1.1, we saw that all transition probabilities of a Markov chain can be reduced to the one-step transition probabilities. Disregarding technicalities, the analogous fact in the continuous-time case is that all transition probabilities can be recovered from the infinitesimal description of the semigroup $(P_t)_{t \geq 0}$ near 0, which is its derivative at 0, also known as the generator. The existence of this derivative depends on the continuity of the semigroup, which is guaranteed by the right-continuity of the process.

[8] Such processes are also often called 'continuous-time Markov chains'. Furthermore, some sources, such as [Liggett, 2010], include the condition that the number of jumps be finite in any finite interval of time, while others do not, such as [Norris, 1997]. For Markov jump processes with finite-state spaces, this condition is always satisfies by Theorem 2.7.1 in [Norris, 1997].

Proposition 5.37. *The transition probabilities P_t of a Markov jump process are continuous in the sense that*

$$\lim_{t\to 0^+} P_t(x,y) = \delta(x,y)$$

for any $x, y \in S$.

Proof. See Definition 2.2 and Theorem 2.12 in [Liggett, 2010]. □

In other words, the Markov matrix-semigroup $(P_t)_{t\geq 0}$ associated to a Markov jump process is strongly continuous. If the state space S is finite, then Theorem 2.8 in [Engel and Nagel, 2006] allows us to conclude abstractly, i.e., without regard to the underlying stochastic process,[9] that the generator

$$Q(x,y) = \lim_{t\to 0^+} \frac{P_t(x,y) - \delta(x,y)}{t}$$

exists and is a rate matrix,[10] which means that it satisfies $Q(x,y) \geq 0$ for all $x, y \in S$ such that $x \neq y$, and $\sum_{y\in S} Q(x,y) = 0$ for each $x \in S$. Furthermore, the matrices P_t satisfy the differential equations

$$\frac{dP_t}{dt} = QP_t \quad \text{and} \quad \frac{dP_t}{dt} = P_tQ \tag{5.51}$$

and either of these, together with the initial condition $P_0 = I$, imply that $P_t = \exp(tQ)$. Conversely, any rate matrix Q uniquely determines a corresponding Markov matrix-semigroup $P_t = \exp(tQ)$.

Returning to Markov jump processes, it follows that all transition probabilities P_t can be recovered from the generator Q through either of the equations (5.51), which in this context are called the backward and forward Kolmogorov equations respectively. How should we interpret the rate matrix Q in probabilistic terms? First, note that the probability

$$\mathbb{P}(X(J_{k-1} + t) = x \mid X(J_{k-1}) = x) = 1 + Q(x,x)t + o(t)$$

is equal to (5.49). Since $\exp(-\lambda(x)t) = 1 - \lambda(x)t + o(t)$, it follows that $\lambda(x) = -Q(x,x)$. Hence, $Q(x,x)$ determines the rate at which the state x jumps to some other state, in the sense of probability per unit time. For $Q(x,y)$ with $x \neq y$, note that the probability (5.50) is equal to

$$\mathbb{P}(X(J_{k-1} + t) = y \mid X(J_{k-1}) = x) = Q(x,y)t + o(t).$$

Since $a(x,y)(1 - \exp(-\lambda(x)t)) = a(x,y)\lambda(x)t + o(t)$, it follows that $Q(x,y) = a(x,y)\lambda(x)$. Hence, $Q(x,y)$ is the rate at which the state x changes to y, which is determined by the rate at which the process leaves x and the jump chain transition probabilities. Combining the two cases, we can write $Q(x,y) = \lambda(x)(a(x,y) - \delta(x,y))$.

We have now seen that the transition probabilities of any Markov jump process form a strongly continuous Markov matrix-semigroup, which is equivalent to its corresponding generator. Similar to what we did in Section 5.1.1, we can ask the converse question whether we can construct a Markov jump process with given initial distribution and transition probabilities.

[9] It is also possible to conclude this from the stochastic process itself, see Theorem 2.1 in [Brémaud, 2020]. Also, for non-finite S abstract reasoning is possible, see Theorem 2.14 in [Liggett, 2010].

[10] A more common name for such a matrices is 'Q-matrix', used for example in [Norris, 1997].

To begin, if $(X(t))_{t\geq 0}$ is a Markov process (not necessarily having right-continuous sample-paths) with initial distribution μ and transition probabilities $(P_t)_{t\geq 0}$, the equality

$$\mathbb{P}(X(t_1) = x_1, \ldots, X(t_n) = x_n) = \sum_{x\in S} \mu(x) P_{t_1}(x, x_1) P_{t_2-t_1}(x_1, x_2) \cdots P_{t_n-t_{n-1}}(x_{n-1}, x_n)$$

follows from and is equivalent to the Markov property in the form equation (5.48). Thus, the finite-dimensional distributions of a Markov process are fixed by the initial distribution and the transition probabilities. Appealing to Kolmogorov's extension theorem, it follows that they uniquely determine the Markov process as a random variable in the space $S^{[0,\infty)}$. However, if one wants to construct processes which are right-continuous or even continuous, then this method does not work directly.

Instead of constructing a Markov jump process directly from given transition probabilities, it is easier to construct it from the corresponding rate matrix, since that is a simpler object. To see how, note that by the considerations following Definition 5.36, it seems reasonable that a Markov jump process $(X(t))_{t\geq 0}$ can equivalently be characterized by the parameters $(\lambda(x))_{x\in S}$ of its holding times and its jump chain $(Y(k))_{k\geq 0}$. These, in turn, are related to the rate matrix by its probabilistic interpretation.

This leads to the following natural construction of a Markov jump process with a given rate matrix Q (see Section 2.6 in [Norris, 1997]). Define $\lambda(x) = -Q(x, x) \geq 0$, $a(x, y) = Q(x, y)/\lambda(x) - \delta(x, y)$ if $\lambda(x) > 0$ and $a(x, y) = 1-\delta(x, y)$ if $\lambda(x) = 0$. In the latter case, x is absorbing, which means the process stays forever in x if it reaches that state. Let $(T_k)_{k\geq 1}$ be independent exponential random variables with parameter 1 and set $S_k = T_k/\lambda(Y_k)$, making S_k exponentially distributed with parameter $\lambda(x)$, conditional on $Y_k = x$. Setting $J_k = S_1 + \cdots + S_k$, we define $(X_t)_{t\geq 0}$ by $X_t = Y_k$ if $J_k \leq t < J_{k-1}$. This requires that $J_k \to \infty$ as $k \to \infty$, which is guaranteed by a simple fact about exponential random variables (see Theorem 2.3.2 in [Norris, 1997]). Note that under this construction, the process obviously has right-continuous sample paths.

Now that we have seen how to construct Markov jump processes, we could look at properties such as stationarity, irreducibility and convergence, like we did for Markov chains in Section 5.1. Much of this is essentially the same as for Markov chains, so we skip most of it (see instead Chapter 3 in [Norris, 1997]). However, we promised in Section 5.1.2 that a continuous-time version of the master equation (5.5) exists, which is what we show now. Let $\mu_t(x) = \mathbb{P}(X(t) = x)$ be the probability distribution at time t. Then

$$\mu_t(y) = \sum_{x\in S} \mu_0(x) P_t(x, y), \tag{5.52}$$

which shows that μ_t is differentiable, since P_t is. Furthermore, it follows from the forward Kolmogorov equation (5.51) that it satisfies the differential equation

$$\frac{d\mu_t(y)}{dt} = \sum_{x\in S} \mu_t(x) Q(x, y),$$

which can be rewritten in the form

$$\frac{d\mu_t(y)}{dt} = \sum_{x\in S} (\mu_t(x) Q(x, y) - \mu_t(y) Q(y, x)) \tag{5.53}$$

using $\sum_{x\in S} Q(y, x) = 0$. Equation (5.53) is the continuous-time analogue of the discrete-time master equation (5.5). Again, the interpretation is that of gains and losses, but now of probability rates, rather than probabilities themselves.

Let us now look at some examples. First, the continuous-time analogue of the Markov chain (5.18) is the two-state Markov jump process with rate matrix

$$\begin{pmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{pmatrix}, \tag{5.54}$$

in which the process jumps between the two states with equal rates. Like its discrete-time counterpart, this is arguably the simplest possible reversible Markov jump process. The probability distribution μ_t is completely determined by $\mu_t(1)$, which according to (5.53) satisfies the differential equation

$$\begin{aligned} \frac{d\mu_t(1)}{dt} &= \lambda\mu_t(0) - \lambda\mu_t(1) \\ &= \lambda(1 - 2\mu_t(1)). \end{aligned} \tag{5.55}$$

This is the analogue of the single-particle master equation (5.29) of the Markov chain (5.18) and has solution given by

$$\mu_t(1) = \frac{1}{2} + \left(\mu_0(1) - \frac{1}{2}\right) e^{-2\lambda t}, \tag{5.56}$$

which shows that μ_k relaxes to the uniform distribution, which is stationary in the sense that it makes the right-hand side of (5.53) vanish.

For the next example, consider N independent copies $(X_n(t))_{t\geq 0}$ of the two-state Markov jump process with generator (5.54), i.e., we use Theorem A.5 to view them as independent processes on the same probability space.

Proposition 5.38. *The stochastic process given by $X(t) = (X_1(t), \dots, X_N(t))$ with state space 2^N is a jump Markov process.*

Proof. Due to independence, both sides of (5.48) can be factorised and the equality holds because each process $X_n(t)$ satisfies the Markov property. It follows that $X(t)$ is a Markov process. For each $t \geq 0$, there are $\varepsilon_n > 0$ such that $X_n(s) = X_n(t)$ if $t \leq s < t + \varepsilon_n$. Taking $\varepsilon = \min_{n=1,\dots,N} \varepsilon_n$, it follows that $X(s) = X(t)$ if $t \leq s < t + \varepsilon$. Hence, $(X(t))_{t\geq 0}$ is right-continuous. □

The process $(X(t))_{t\geq 0}$ is constructed similarly to the variant Ehrenfest model in Section 5.2.2. However, an important difference is that in the process $(X(t))_{t\geq 0}$, the particles almost surely change state one at a time, due to the continuity of time. This makes the process more like the original Ehrenfest model of Section 5.2.1, despite its construction being more like the variant. Indeed, the original Ehrenfest model is actually the jump chain of this process.

As in the discrete-time case, one may now use the natural coarse-graining map $f : 2^N \to [0, 1]$ to define a macroscopic version

$$f(t) = \frac{1}{N} \sum_{n=1}^{N} X_n(t)$$

of the process $X(t)$, which also satisfies the Markov property and is right-continuous using the same argument as in the proof of Proposition 5.38, hence is a Markov jump process as well.

5.5.3 Macroscopic evolution

By the same reasoning as in Section 5.4.1, the expectation $F(t) = \mathbb{E}(f(t))$ satisfies the differential equation $F'(t) = \lambda(1 - 2F(t))$ which is identical to the single-particle master equation (5.55). Its solution

$$F(t) = \frac{1}{2} + \left(F(0) - \frac{1}{2}\right) e^{-2\lambda t} \tag{5.57}$$

is similarly identical to (5.56) and shows that $F(t)$ exhibits relaxation to the equilibrium value $1/2$ when starting in a non-equilibrium state. As in Section 5.4.3, one may prove that with high probability $f(t)$ stays near $F(t)$ on a finite time interval. However, let us go directly to the strong law of large numbers. Suppose we have an infinite number of independent copies $(X_n(t))_{t\geq 0}$ of the Markov jump process with generator (5.54) defined on the same probability space, for example using Theorem A.5. Define the truncations

$$f_N(t) = \frac{1}{N}\sum_{n=1}^{N} X_n(t) \quad \text{and} \quad F_N(t) = \mathbb{E}(f_N(t)).$$

The following theorem is proven in exactly the same way as Theorem 5.30. Indeed, both are essentially nothing more than the usual strong law of large numbers applied at each point of time.

Theorem 5.39. *Suppose that $F_N(0)$ converges to some number $F(0)$ as $N \to \infty$. If we define $F(t)$ by (5.57) and fix $t \geq 0$, then almost surely*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} X_n(t) = F(t).$$

An analogue of Theorem 5.32 would require us to first construct a proper probability space for the process given by $X(t) = (X_n(t))_{n\geq 1}$. However, this is significantly more involved than in the discrete-time case. We would like $(X(t))_{t\geq 0}$ to have sample paths that are right-continuous and have left limits. The state space $2^{\mathbb{Z}_+}$ is not discrete, so the construction in Section 5.5.2 is not applicable. However, it is still possible to construct the process using a generator. Because the state space is not discrete, this require some operator theory. Formally, the generator is given by its action

$$Lf(x) = \sum_{n=1}^{\infty} \lambda(f(x^n) - f(x)), \tag{5.58}$$

on continuous functions $f \in C(2^{\mathbb{Z}_+})$. Here x^n is the microstate obtained by changing the state of the n'th particle in the microstate x. To make the sum well defined, one first restricts to local functions f, which only depend on a finite number of particles. These are dense in $C(2^{\mathbb{Z}_+})$ and one uses the Hille-Yosida theorem to show that the closure of L generates a strongly continuous Markov semigroup. By a general theorem relating Markov processes and semigroups, one has then constructed the process in the form of a probability measure $\mathbb{P}$ on the Skorokhod space $D_{[0,\infty)}(2^{\mathbb{Z}_+})$ of right-continuous paths in $2^{\mathbb{Z}_+}$ with left limits. See Chapter 1 in [Liggett, 1985] for the many details needed to make this precise.

With the appropriate probability space $(D_{[0,\infty)}(2^{\mathbb{Z}_+}), \mathbb{P})$ constructed, one needs to choose a countable basis to make it effective. There is a general theorem stating that $D_{[0,\infty)}(S)$ is second-countable if S

is a second-countable metric space (see Chapter 3, Theorem 5.6 in [Kurtz and Ethier, 1986]), which implies that this is possible in principle, but the effective basic sets will likely be less simple than the cylinder sets we encountered in the discrete-time case. Nonetheless, assuming no problems regarding computability, it should not be hard to turn Theorem 5.39 into a statement about $\mathbb{P}$-randomness, though we leave the details for another time.

6 Conclusion

In this thesis, we have used two toy models, the Kac ring model and the Ehrenfest urn model, to illustrate the emergence of irreversible macroscopic behaviour from reversible microscopic laws for deterministic and stochastic dynamics. We also used these models as test cases for the research programme of applying algorithmic randomness to statistical mechanics, initiated by [Hiura and Sasa, 2019]. Our contributions to this research programme are as follows:

- It was shown in [Hiura and Sasa, 2019], using the Kac chain as an illustrating case, that random microstates of a deterministic system, give rise to trajectories that exhibit irreversibility when viewed macroscopically, relative to a suitable probability measure. We extended this result to stochastic dynamics, specifically for the Ehrenfest urn model, with our Theorem 5.32. The main difference with the deterministic case is that stochastic dynamics require the use of probability measures on trajectories of microstates, rather than microstates themselves.
- The problem of what to make of non-random microstates whose trajectories nevertheless satisfy macroscopic laws was brought up in [Hiura and Sasa, 2019]. We proposed that this problem can be resolved by the realization that the trajectories of such microstates satisfy some microscopic laws, but not all. This is made precise by the effective ergodic theorem, resulting in our Theorem 4.15 which states that a microstate is random if and only if its trajectory satisfies all relevant macroscopic laws.
- Following [Landsman, 2023], we discussed the Stosszahlansatz from the viewpoint of algorithmic randomness, using the Kac chain as a test case. We concluded that macroscopic laws can be proven directly from the assumption of randomness, bypassing the Stosszahlansatz. Since each macroscopic law requires its own Stosszahlansatz-like assumptions, which we expect to be consequences of randomness, we also conclude that randomness of the microstate is a more fundamental assumption than the Stosszahlansatz. However, macroscopic laws in more complicated laws may require first proving a Stosszahlansatz from the randomness assumption.

The use of algorithmic randomness in statistical mechanics as described by [Hiura and Sasa, 2019] and [Landsman, 2023] is in its infancy and much work is still to be done. The holy grail would be to rigorously derive the Boltzmann equation using algorithmic randomness. Recently, important work on the rigorous derivation of the Boltzmann equation has appeared (see [Bodineau et al., 2023] and [Deng et al., 2025]), and it would be interesting to see whether algorithmic randomness can be useful for such work. Some more immediate directions of possible future research are the following:

- While the Kac chain and Ehrenfest urn model illustrate the essence of how algorithmic randomness can be applied to statistical mechanics, it would be useful to use more complicated models as stepping stones towards the case of kinetic theory, in particular those in which the particles can interact.
- Both the Kac chain and the Ehrenfest urn model are in discrete-time. To get closer to the Boltzmann equation, it will be useful to apply algorithmic randomness to a model which is in continuous-time. We already indicated how this may be done in Section 5.5, but more work needs to be done.
- In this thesis, the Stosszahlansatz was discussed from the viewpoint of algorithmic randomness within the context of the Kac chain, but we feel that better conclusions can be drawn by using

a more complicated model, such as the Ehrenfest wind-tree model (see [Brown et al., 2009]).

- The most clean notion of algorithmic randomness is Martin-Löf randomness, which is based on measure theory. However, it can only be sensibly applied to infinite objects. This requires taking thermodynamic limits of physical models. It would be interesting to derive irreversible macroscopic laws for finite systems, using a notion of randomness appropriate to finite objects, i.e., Kolmogorov randomness.

7 References

[Albert, 2000] Albert, D. (2000). *Time and Chance*. Harvard University Press.

[Arntzenius, 1995] Arntzenius, F. (1995). Indeterminism and the direction of time. *Topoi*, 14:67–81.

[Bacciagaluppi, 2010] Bacciagaluppi, G. (2010). *Probability and time symmetry in classical Markov processes: Probabilities, Causes and Propensities in Physics*, volume 347 of *Synthese Library*, pages 41–60. Springer.

[Baldovin et al., 2019] Baldovin, M., Caprini, L., and Vulpiani, A. (2019). Irreversibility and typicality: A simple analytical result for the Ehrenfest model. *Physica A: Statistical Mechanics and its Applications*, 524:422–429.

[Bodineau et al., 2023] Bodineau, T., Gallagher, I., Saint-Raymond, L., and Simonella, S. (2023). On the dynamics of dilute gases. *Eur. Math. Soc. Mag.*, 128:13–22.

[Bricmont, 1995] Bricmont, J. (1995). Science of Chaos or Chaos in Science? *Physicalia Magazine*, 17:159–208.

[Bricmont, 2022] Bricmont, J. (2022). *Making Sense of Statistical Mechanics*. Springer.

[Brown et al., 2009] Brown, H. R., Myrvold, W., and Uffink, J. (2009). Boltzmann's H-theorem, its discontents, and the birth of statistical mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 40(2):174–191.

[Brush, 1976] Brush, S. (1976). *The Kind of Motion We Call Heat*. North-Holland Publishing Company.

[Brémaud, 2020] Brémaud, P. (2020). *Markov Chains*. Springer, 2nd edition.

[Calude, 2002] Calude, C. (2002). *Information and Randomness An Algorithmic Perspective*. Springer, 2nd edition.

[Caputo, 2022] Caputo, P. (2022). Lecture Notes on Entropy and Markov Chains. `https://www.mat.uniroma3.it/users/caputo/entropy.pdf`.

[Dasgupta, 2011] Dasgupta, A. (2011). Mathematical Foundations of Randomness. In Bandyopadhyay, P. S. and Forster, M. R., editors, *Philosophy of Statistics*, volume 7 of *Handbook of the Philosophy of Science*, pages 641–710. North-Holland, Amsterdam.

[Deng et al., 2025] Deng, Y., Hani, Z., and Ma, X. (2025). Long time derivation of the Boltzmann equation from hard sphere dynamics.

[Di Biagio and Rovelli, 2024] Di Biagio, A. and Rovelli, C. (2024). On the Time Orientation of Probabilistic Theories. *Phil. Phys.*, 2(1):16.

[Downey and Hirschfeldt, 2010] Downey, R. and Hirschfeldt, D. (2010). *Algorithmic Randomness and Complexity*. Springer.

[Earman, 2006] Earman, J. (2006). The "Past Hypothesis": Not even false. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 37(3):399–430. The arrows of time, 2006.

[Ebeling et al., 2008] Ebeling, W., Gudowska-Nowak, E., and Sokolov, I. (2008). On Stochastic Dynamics in Physics - Remarks on History and Terminology. *Acta Physica Polonica*, B39.

[Ehrenfest and Ehrenfest, 1907] Ehrenfest, P. and Ehrenfest, T. (1907). Über zwei bekannte einwände gegen das boltzmannsche h-theorem. *Physikalische Zeitschrift*, pages 311–314.

[Ehrenfest and Ehrenfest, 1912] Ehrenfest, P. and Ehrenfest, T. (1912). Begriffliche grundlagen der statistischen auffassung in der mechanik. In Klein, F. and Müller, C., editors, *Enzycl. d. Mathem. Wissenschaften*, pages 773–860. Vieweg+Teubner Verlag, Wiesbaden.

[Enderton, 2010] Enderton, H. (2010). *Computability Theory: An Introduction to Recursion Theory*.

[Engel and Nagel, 2006] Engel, K. and Nagel, R. (2006). *A Short Course on Semigroups*. Springer.

[Falk, 1980] Falk, H. (1980). Discrete-time Glauber model as a generalized Ehrenfest urn model. *Physica A: Statistical Mechanics and its Applications*, 104(3):459–474.

[Frigg and Werndl, 2024] Frigg, R. and Werndl, C. (2024). *Foundations of Statistical Mechanics*. Elements in the Philosophy of Physics. Cambridge University Press.

[Galatolo et al., 2010] Galatolo, S., Hoyrup, M., and Rojas, C. (2010). Effective symbolic dynamics, random points, statistical behavior, complexity and entropy. *Information and Computation*, 208(1):23–41.

[Goldstein et al., 2001] Goldstein, H., Poole, C., and Safko, J. L. (2001). *Classical Mechanics*. Addison-Wesley, 3rd edition.

[Goldstein, 2001] Goldstein, S. (2001). *Boltzmann's Approach to Statistical Mechanics*, pages 39–54. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Gottwald and Oliver, 2009] Gottwald, G. A. and Oliver, M. (2009). Boltzmann's Dilemma: An Introduction to Statistical Mechanics via the Kac Ring. *SIAM Rev.*, 51(3):613–635.

[Hauert et al., 2004] Hauert, C., Nagler, J., and Schuster, H. (2004). Of Dogs and Fleas: The dynamics of N uncoupled two-state systems. *J Stat Phys*, pages 1453–1469.

[Hertling and Weihrauch, 2003] Hertling, P. and Weihrauch, K. (2003). Random elements in effective topological spaces with measure. *Information and Computation*, 181(1):32–56.

[Hiura and Sasa, 2019] Hiura, K. and Sasa, S. (2019). Microscopic reversibility and macroscopic irreversibility: From the viewpoint of algorithmic randomness. *J Stat Phys*, pages 727–751.

[Hoeffding, 1963] Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30.

[Holster, 2003] Holster, A. T. (2003). The criterion for time symmetry of probabilistic theories and the reversibility of quantum mechanics. *New Journal of Physics*, 5:130.

[Horn and Johnson, 1985] Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press.

[Jacod and Protter, 2004] Jacod, J. and Protter, P. (2004). *Probability Essentials*. Springer, 2nd edition.

[Kac, 1947a] Kac, M. (1947a). On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. Soc.*, 53:1002–1010.

[Kac, 1947b] Kac, M. (1947b). Random Walk and the Theory of Brownian Motion. *The American Mathematical Monthly*, 54(7):369–391.

[Kac, 1956] Kac, M. (1956). Foundations of Kinetic Theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*.

[Kac, 1959] Kac, M. (1959). *Probability and Related Topics in Physical Sciences*. Interscience Publishers Inc.

[Karlin and Taylor, 1975] Karlin, S. and Taylor, H. (1975). *A First Course in Stochastic Processes*. Academic Press, 2nd edition.

[Karlin and Taylor, 1981] Karlin, S. and Taylor, H. (1981). *A Second Course in Stochastic Processes*. Academic Press.

[Kelly, 1979] Kelly, F. (1979). *Reversibility and Stochastic Networks*. Wiley.

[Kemeny et al., 1976] Kemeny, J., Snell, J., and A.W., K. (1976). *Denumerable Markov Chains*. Springer, 2nd edition.

[Klenke, 2020] Klenke, A. (2020). *Probability Theory A Comprehensive Course*. Springer, 3rd edition.

[Kolmogorov, 1936] Kolmogorov, A. (1936). Zur Theorie der Markoffschen Ketten. *Mathematische Annalen*, 112:155–160.

[Kurtz and Ethier, 1986] Kurtz, T. and Ethier, S. (1986). *Markov Processes: Characterization and Convergence*. Wiley.

[Landsman, 2023] Landsman, K. (2023). Typical = Random. *Axioms*, 12(8).

[Lazarovici and Reichert, 2015] Lazarovici, D. and Reichert, P. (2015). Typicality, Irreversibility and the Status of Macroscopic Laws. *Erkenntnis*, 80:689–716.

[Lebowitz, 1993] Lebowitz, J. L. (1993). Macroscopic laws, microscopic dynamics, time's arrow and Boltzmann's entropy. *Physica A: Statistical Mechanics and its Applications*, 194(1):1–27.

[Levin and Peres, 2017] Levin, D. and Peres, Y. (2017). *Markov Chains and Mixing Times*. American Mathematical Society, 2nd edition.

[Li and Vitányi, 2019] Li, M. and Vitányi, P. (2019). *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 4th edition.

[Liggett, 1985] Liggett, T. (1985). *Interacting Particle Systems*. Springer.

[Liggett, 2010] Liggett, T. (2010). *Continuous Time Markov Processes: An Introduction*. American Mathematical Society.

[Mackey, 1992] Mackey, M. (1992). *Time's Arrow: Origin of Thermodynamics Behavior*. Springer.

[Maes et al., 2009] Maes, C., Neto˘cn´y, K., and Shergelashvili, B. (2009). *A Selection of Nonequilibrium Issues*, pages 247–306. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Martin-Löf, 1966] Martin-Löf, P. (1966). The Definition of Random Sequences. *Information and Control*, 9(6):602–619.

[Meyer, 2022] Meyer, P. (2022). *Stochastic Processes in the Decades after 1950*, pages 169–202. Trends in the History of Science. Springer Science and Business Media Deutschland GmbH.

[Nies, 2008] Nies, A. (2008). *Computability and randomness*. Clarendon Press.

[Norris, 1997] Norris, J. R. (1997). *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[Pathak et al., 2014] Pathak, N., Rojas, C., and Simpson, S. G. (2014). Schnorr randomness and the Lebesgue differentiation theorem. *Proc. Am. Math. Soc.*, 142:335–349.

[Scalas et al., 2007] Scalas, E., Martin, E., and Germano, G. (2007). Ehrenfest urn revisited: Playing the game on a realistic fluid model. *Phys. Rev. E*, 76:011104.

[Seifert, 2025] Seifert, U. (2025). *Stochastic Thermodynamics*. Cambridge University Press.

[Sober, 1993] Sober, E. (1993). Temporally Oriented Laws. *Synthese*, 94(2):171–189.

[Uffink, 2006] Uffink, J. (2006). Compendium of the Foundations of Classical Statistical Physics. In Butterfield, J. and Earman, J., editors, *Handbook of the philosophy of physics*. Kluwer Academic Publishers.

[Walters, 1982] Walters, P. (1982). *An Introduction to Ergodic Theory*. Springer.

[Watanabe, 1965] Watanabe, S. (1965). Conditional Probability in Physics. *Progress of Theoretical Physics Supplement*, E65:135–160.

A Stochastic processes

We summarize in this appendix some parts of the theory of stochastic processes which are needed in this thesis. The reader is referred to any standard text for a more complete exposition, for example [Karlin and Taylor, 1975] and [Karlin and Taylor, 1981].

Recall from probability theory that a random variable is a measurable function $X : \Omega \to S$, with $(\Omega, \Sigma, \mathbb{P})$ a probability space and $(S, \mathcal{S})$ a measurable space. The specific probability space is often not important. Rather, the fundamental object is really the law[11] P_X, which is the probability measure on S defined by $P_X = \mathbb{P} \circ X^{-1}$. Thus, an S-valued random variable is essentially the same as a probability measure on S.

A stochastic process on S is simply a collection of random variables $X_t : \Omega \to S$, indexed by some set T, which we often interpret as time, but can also be interpreted in non-temporal ways (in which case we speak of a random field rather than a process). The set S is called the state space. Through currying, we may equivalently view the process as a function $X : \Omega \to S^T$. A natural choice of σ-algebra on S^T makes X a random variable.

Definition A.1. *Given a measurable space $(S, \mathcal{S})$, the cylindrical σ-algebra on S^T is the smallest σ-algebra making all the evaluation maps $\pi_t : S^T \mapsto S$, $x \mapsto x(t)$ measurable. Equivalently, it is the σ-algebra generated by all (one-dimensional rectangular) cylinder sets*

$$\pi_t^{-1}(A) = \{x \in S^T \mid x(t) \in A\},$$

with $A \in \mathcal{S}$ and $t \in T$.

Throughout this thesis, we use the term 'cylinder set' in the sense of a subset of S^T having the form

$$\bigcap_{t \in F} \pi_t^{-1}(\{a_t\}) = \{x \in S^T \mid x(t) = a_t \text{ for all } t \in F\},$$

with $F \subseteq T$ finite and $a_t \in S$ for each $t \in F$, i.e., subsets whose elements have fixed values at a finite collection of times.

Proposition A.2. *There is a one-to-one correspondence between stochastic process $(X_t)_{t \in T}$ on S and random variables X on S^T, given by $X_t = \pi_t \circ X$.*

Proof. Given a stochastic process $X_t : \Omega \to S$ indexed by T, there is a unique function $X : \Omega \to S^T$ satisfying $\pi_t \circ X = X_t$ for all $t \in T$, i.e., $X(\omega) : t \mapsto X_t(\Omega)$. This function is measurable since preimages $X^{-1}(\pi_t^{-1}(A)) = X_t^{-1}(A)$ of cylinder sets are measurable. Conversely, a given random variable $X : \Omega \to S^T$ gives rise to random variables $X_t = \pi_t \circ X$, since the evaluation maps π_t are measurable by construction. □

By our general comment on random variables made above, it follows that a stochastic process X is essentially the same as its law P_X on S^T. We call the probability space $(S^T, \mathcal{F}, P_X)$ the Kolmogorov representation of X. The probability measures $P_X \circ \pi_F^{-1}$ on S^F with $F \subseteq T$ finite are called finite-dimensional distributions. By a result from measure theory, these uniquely determine P_X, since

[11] This is sometimes called the distribution of X, a name which I prefer but do not use, since I already reserve that term for probability mass functions.

the collection of cylinder sets is closed under intersection. Conversely, any consistent collection of probability measures p_F on each finite product S^F determines a probability measure on S^T, assuming S is sufficiently regular. This is known as Kolmogorov's extension theorem. Consistent here means that

$$p_F = p_G \circ \pi_{F,G}^{-1}$$

for any finite subsets $F, G \subseteq T$ such that $F \subseteq G$, with $\pi_{F,G} : S^G \to S^F$ the natural projection of S^G onto S^F.

Theorem A.3 (Kolmogorov's extension theorem)**.** *If S is Polish, i.e., a separable complete metric space, then any consistent collection of probability measures p_F on S^F with $F \subseteq T$ finite determines a unique probability measure P on S^T such that $p_F = P \circ \pi_F^{-1}$ for each F.*

Proof. See Section 14.3 in [Klenke, 2020]. □

A discrete-time process is one with indexing set $T = \mathbb{N}$ or $T = \mathbb{Z}$, the former corresponding to processes with an initial time and the latter to those without an initial time. In the first case, one can restrict to finite sets of the form $F = \{0, 1, 2, \ldots, N\}$ in Theorem A.3, because any finite subset of $\mathbb{N}$ is contained such a set. Similarly, in the second case one may restrict to $F = \{-N, \ldots, -1, 0, 1, \ldots, N\}$.

Definition A.4. *Given a collection $(S_i, \mathcal{S}_i)_{i \in I}$ of measurable spaces, the product σ-algebra on $\prod_{i \in I} S_i$ is the smallest σ-algebra making all the projections $\pi_j : \prod_{i \in I} S_i \to S_j$ measurable.*

Theorem A.5. *Given a collection $(S_i, \mathcal{S}_i, P_i)_{i \in \mathbb{N}}$ of probability spaces, there is a unique probability measure P on the product σ-algebra of $\prod_{i \in \mathbb{N}} S_i$ such that*

$$P\left(\bigcap_{i \in F} \pi_i^{-1}(E_i)\right) = \prod_{i \in F} P_i(E_i)$$

for any finite $F \subseteq \mathbb{N}$ and $E_i \in \mathcal{S}_i$. With respect to the probability measure P, the random variables π_i are independent.

Proof. See Corollary 14.33 in [Klenke, 2020]. □

Consider the double product space $S^{I \times J}$. It is obviously bijective to the iterated product spaces $(S^I)^J$ and $(S^J)^I$. However, the generating class of cylinder sets appearing in Definition A.1 is different for each of these products. Denoting the elements of $S^{I \times J}$ by double sequences $x(i, j)$, these classes are given by

$$\begin{aligned} C_1 &: \text{sets } \{x \in S^{I \times J} \mid x(i,j) \in E\} \text{ with } i \in I,\ j \in J \text{ and } E \subseteq S \text{ measurable} \\ C_2 &: \text{sets } \{x \in S^{I \times J} \mid x(\cdot, j) \in F\} \text{ with } j \in J \text{ and } F \subseteq S^I \text{ measurable} \\ C_3 &: \text{sets } \{x \in S^{I \times J} \mid x(i, \cdot) \in G\} \text{ with } i \in I \text{ and } G \subseteq S^J \text{ measurable.} \end{aligned} \tag{A.1}$$

Proposition A.6. *The three classes of cylinder sets* (A.1) *generate the same σ-algebra on $S^{I \times J}$.*

Proof. We prove only $\sigma(C_1) = \sigma(C_2)$, the equality $\sigma(C_1) = \sigma(C_3)$ being proven in the same way by symmetry. Define

$$C_1(E, i, j) = \{x \in S^{I\times J} \mid x(i, j) \in E\}$$
$$C_2(F, j) = \{x \in S^{I\times J} \mid x(\cdot, j) \in F\}.$$

and denote the projections $S^I \to S$ by π_i. It is obvious that

$$C_1(E, i, j) = C_2(\pi_i^{-1}(E), j), \tag{A.2}$$

which implies $C_1 \subseteq C_2$ and hence $\sigma(C_1) \subseteq \sigma(C_2)$. For the converse, let Σ be the cylindrical σ-algebra on S^I and define

$$\Sigma'_j = \{F \in \Sigma \mid C_2(F, j) \in \sigma(C_1)\}.$$

It is easily checked that Σ'_j is itself a σ-algebra. The sets $\pi_i^{-1}(E)$ generate Σ and are contained in Σ'_j due to (A.2). It follows that $\Sigma'_j = \Sigma$ for each $j \in J$, so that $C_2 \subseteq \sigma(C_1)$ and hence $\sigma(C_2) \subseteq \sigma(C_1)$. □

B Concentration inequalities

As was stated in Section 2 and illustrated in Sections 4 and 5, coarse-graining together with the law of large numbers is the essence of emergent irreversible behaviour on the macroscopic scale. When referring to the law of large numbers, it usually the version for independent variables which is meant, though this can be generalized to correlated variables and indeed needs to be generalized for systems with interactions.

Underlying laws of large numbers for both independent and dependent random variables is the concentration phenomenon, which says that a function $f(X_1, \ldots, X_N)$ of random variables $X_1, \ldots, X_N$ will be close to its expected value if it depends only weakly on each X_i and if the random variables are not strongly correlated. This idea is made rigorous by concentration inequalities or, even better, large deviations principles. We collect a few concentration inequalities in this appendix which we use throughout the thesis.

Proposition B.1 (Markov's inequality)**.** *If X is a non-negative random variable, then for any $a > 0$ we have*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Proof. See Theorem 5.1 in [Jacod and Protter, 2004]. □

From Markov's inequality, further and more useful concentration inequalities can be derived as follows. Suppose that φ is a non-negative and non-decreasing function and a is such that $\varphi(a) > 0$. Then

$$\begin{aligned}\mathbb{P}(X \geq a) &= \mathbb{P}(\varphi(X) \geq \varphi(a)) \\ &\leq \frac{\mathbb{E}(\varphi(X))}{\varphi(a)}.\end{aligned}$$

Applying this to the random variable $|X - \mathbb{E}(X)|$ with the function $\varphi(x) = x^m$, we get so called moment bounds, assuming X has finite expectation. Two particular cases are the following. Note that a finite moment implies finite expectation by Hölder's inequality.

Proposition B.2 (Chebyshev's inequality)**.** *If the real-valued random variable X has finite variance, then*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Proposition B.3 (Fourth moment bound)**.** *If the real-valued random variable X has finite fourth moment, then*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{F}(X)}{\varepsilon^4}.$$

Here $\mathbb{V}(X)$ is our notation for the variance $\mathbb{E}((X - \mathbb{E}(X))^2)$ and $\mathbb{F}(X)$ is defined as the fourth moment $\mathbb{E}((X - \mathbb{E}(X))^4)$. As is well known, if the random variables $X_1, \ldots, X_N$ are independent, then the identity

$$\mathbb{V}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2}\sum_{n=1}^{N} \mathbb{V}(X_n) \tag{B.1}$$

holds, which makes Chebyshev's inequality useful. For example, if the random variables are such that their variances are uniformly bounded by a constant C, then the right-hand side of (B.1) can be bounded by C/N. The corresponding identity for the fourth moment is given by

$$\mathbb{F}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^4}\left(\sum_{n=1}^{N} \mathbb{F}(X_n) + 6 \sum_{\substack{n,m=1 \\ n<m}}^{N} \mathbb{V}(X_n)\mathbb{V}(X_m)\right). \tag{B.2}$$

In this case, if the fourth moments of the random variables are uniformly bounded by C, the right-hand side of (B.2) can be bounded by C/N^2.

While moment bounds are in many cases useful, they are quite crude. At the cost of restricting to random variables of the form $N^{-1}\sum_{n=1}^{N} X_n$ with $X_1, \ldots, X_N$ independent, we have the following much tighter bound.

Proposition B.4 (Hoeffding's inequality)**.** *Suppose the random variables $X_1, \ldots, X_N$ are independent, bounded, say $a \leq X_n \leq b$, and have finite expectations. Then*

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}(X_n)\right| \geq \varepsilon\right) \leq 2\exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right).$$

Proof. See [Hoeffding, 1963]. □