ALMA MATER STUDIORUM · UNIVERSITÁ DI BOLOGNA

DIPARTIMENTO DI MATEMATICA
Corso di Laurea in Matematica

From Boltzmann's problem to Kac's chaos

Tesi di Laurea in Probabilitá e Statistica Matematica

Relatore:	**Presentata da:**
Chiar.mo Prof.	**EDOARDO PANDOLFO**
ANDREA PASCUCCI	

Anno Accademico 2024-2025

A Giacomo Bologna,
un ragazzo instancabile . . .
per ricordarmi di provare sempre
a vivere al massimo . . .

Introduction

The central problem that motivates this thesis originates from Ludwig Boltzmann's attempt to reconcile the reversible character of microscopic dynamics with the irreversible evolution described by thermodynamics. In his celebrated H-theorem (1872), Boltzmann proposed a statistical explanation of the second law, showing that entropy tends to increase in isolated systems. To support this approach, he also introduced the so-called ergodic hypothesis, suggesting that the long-time behavior of a dynamical system justifies the replacement of time averages with ensemble averages, thus justifying the foundations of statistical mechanics.

This view was later challenged by Zermelo, who raised his famous objection to Boltzmann, arguing that Poincaré's recurrence theorem contradicts the monotonic increase of entropy postulated by the H-theorem. This became one of the central points in the debate on the foundations of statistical mechanics, highlighting the tension between microscopic reversibility and macroscopic irreversibility.

To answer those doubts, simplified probabilistic models were introduced. Among them, the Ehrenfest urn model (1907) provided a schematic but illuminating description of the stochastic evolution of macrostates. This model makes explicit the statistical drift towards equilibrium while allowing a precise analysis of fluctuations and recurrence times.

Later, Mark Kac developed a rigorous probabilistic framework that refined Boltzmann's ideas. In particular, he introduced the concept of chaos, providing a mathematical interpretation of the *Stosszahlansatz* (molecular chaos assumption). In this setting, the statistical independence of particles in the thermodynamic limit is no longer a heuristic principle, but a well-defined probabilistic property.

The aim of this thesis is to revisit these topics by adding a modern probabilistic perspective where possible. By combining Boltzmann's original problem, the framework of ergodic theory, the Ehrenfest urn model, and Kac's notion of chaos, I seek to clarify the statistical meaning of irreversibility and show how the tension between determinism and irreversibility can be understood through the interplay of probability, combinatorics, and dynamical systems.

Notations and Conventions

r.v.	Random variable or random vector.
i.d.	Identically distributed.
i.i.d.	Independent identically distributed.
$\mathcal{P}(E)$	Set of probability measures on the measurable space E.
$C_b(E)$	Space of continuous bounded real-valued functions on E.
δ_x	The Dirac delta probability measure that assigns probability 1 to the element x and 0 elsewhere.
$1_A(x)$	Indicator function: equals 1 if $x \in A$ and 0 if $x \notin A$
$\langle \mu, \varphi \rangle, \ \langle x, y \rangle$	Respectively, the integral of the measurable function φ with respect to the probability measure μ, and the scalar product of x and y.
$\mathcal{B}$	Sigma-algebra of Borel.
$m\mathcal{F}$	Set of random variables from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.
$L^1(\Omega, \mathcal{F}, \mathcal{P})$	Set of random variables with finite expectation: $\{X \in m\mathcal{F} \mid \mathbb{E}[\lvert X \rvert] < +\infty\}$

Contents

Chapter 1

Boltzmann's problem

In this chapter, I present an overview of the historical development of thermodynamics and statistical mechanics. The explanation may sound somewhat imprecise at times; this is because I give an intuitive idea of the concepts as they were known in the 19th century, without a mathematically rigorous perspective. I am also not going to elaborate deeply on every topic as there are many and this is not the main purpose of this thesis.

1.1 Second law of Thermodynamics

Thermodynamics is the branch of physics that studies macroscopic processes involving the exchange and conversion of heat. Macroscopic means visible and perceptible to the naked eye, thus the thermodynamic study describes physical systems (e.g. gases) by a finite number of macroscopic measurable parameters, between which there are experimentally determinable relationships, for instance temperature, pressure and volume. The basis of the theory is the equivalence between heat and work. A description of any thermodynamic system employs the laws of thermodynamics, which were historically empirical findings, and now form an axiomatic basis. I will present a brief explanation of how it developed and focus on some elements important for the purpose of this thesis.

It was not until the end of the 16th century that humans had a way to quantify the idea of hotness and coldness. To do so, temperature was introduced.

Temperature T is a physical quantity of a body on which many other properties depend.

For example, most of materials expand when heated. Temperature can be measured with a thermoscope, an instrument that captures variations in this property of a body and translates them into a value. The first thermoscope dates back to Galileo Galilei (1597).

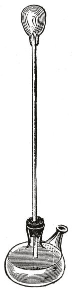


Figure 1.1:

The jar is heated with the hands and partially immersed, upside down, in a container filled with water. When the hands are removed, the air in the jar heats and the water level in the neck rises. The experiment demonstrated the changes in air density caused by temperature variations. If a proper scale is added to the neck, the thermoscope becomes an actual thermometer.

example: modern mercury thermometers work on the same principle.

The experiment is also an example of how the value shown by a thermoscope, when in contact with another body, changes, until it stops and the two bodies are said in thermal equilibrium.

Two bodies are said to be in *thermal equilibrium* if they have the same temperature.

It is empirically proven that two bodies in contact tend to reach a thermal equilibrium. This phenomenon is an example of thermodynamic process.

In general, a *thermodynamic process* is a process in which the thermodynamic state of a system changes. A change in a system is defined by a transition from an initial to a final state of thermodynamic equilibrium. In a system that is in its own state of internal thermodynamic equilibrium, not only is there an absence of macroscopic change, but there is an ”absence of any tendency toward change on a macroscopic scale.”

Def 1.1.1

Heat Q is the transfer of thermal energy between two systems, associated with a difference in temperature. For an object that remains in the same physical state (i.e. it does not melt or vaporize, or change state in general):

$$Q = mc\Delta T$$

where m is the mass of the object, c is its specific heat capacity and ΔT the difference in temperature of the object.

Historically, thermodynamics developed out of a desire to increase the efficiency of early steam engines, particularly through the work of French physicist Sadi Carnot. In 1824 he published a paper with the fundamental idea that a heat engine, in order to produce mechanical work using heat, must operate between two sources at different temperature. He believed that heat produces work because it spontaneously goes from the warmest source to the coldest.

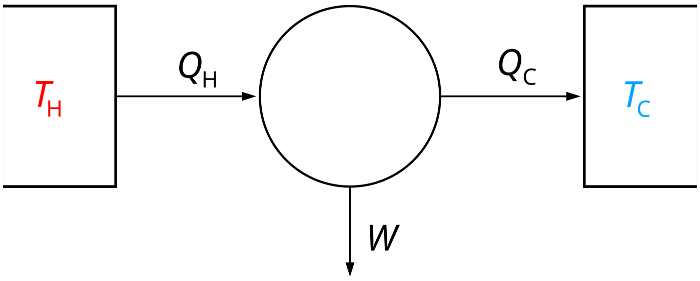


Figure 1.2: Carnot heat engine

In 1850 German physicist and mathematician Rudolf Clausius restated Carnot's principle known as the Carnot cycle and gave the theory of heat a sounder basis. Clausius distinguished the inner energy of a body (thermal energy, later associated with the kinetic and potential energies of the atoms composing the object) and the heat or energy transferred from or to the body. He defined the first two laws of thermodynamics:

First law of thermodynamics

$$\Delta E_{int} = Q - L$$

where L is the work done by the system, Q is the absorbed heat and $\Delta E_{int} = E_{int,f} - E_{int,i}$

Second law of thermodynamics

heat does not flow spontaneously from cold to hot bodies

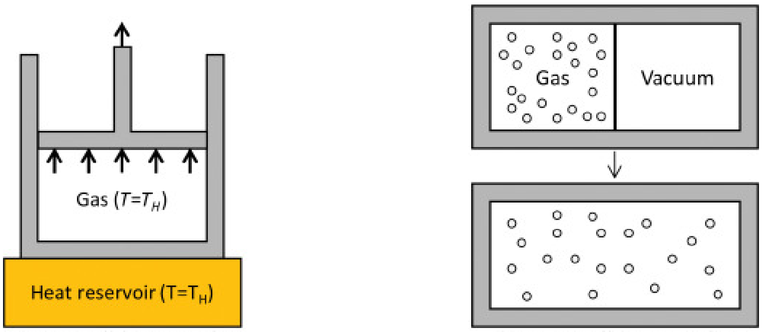
Entropy

An *isolated system* is a system that does not exchange mass, work, or heat with the environment. In this sense the universe, which is the system plus its environment (where the system is immersed and with whom it exchanges energy and heat), is an isolated system.

Def 1.1.2

A reversible process is a thermodynamic process that can be inverted without producing changes in the surroundings, returning both the environment and the system to their initial conditions.

An example of irreversible process is the free expansion of a gas. It is indeed hard to believe that it could spontaneously go back to its original state.



(a) Reversible expansion (b) Irreversible expansion

Figure 1.3: (a) isothermal expansion of a gas: the reservoir transfers heat to the gas while it expands, and maintains its temperature. The process produces work, with which it can be inverted by pushing the piston downwards. (b) free expansion into vacuum: It requires external energy to return to the original state

In 1864 Clausius introduced the state function ”entropy”. A state function is a function that depends only on the current thermodynamic equilibrium state of the system, not on the path the system has taken to reach that state.

Def 1.1.4

The variation of entropy S from an itial state A to B is

$$\Delta S = \int_A^B \frac{\mathrm{d}Q_{rev}}{T}$$

where $\mathrm{d}Q_{rev}$ is the amount of heat absorbed or released in a reversible and isotherm way by the system at temperature T in an infinitesimally short time interval.

Remark: It follows from the formula that in any infinitesimal reversible transformation the change in S is given by $\mathrm{d}S = \frac{\mathrm{d}Q}{T}$

Note that the definition of entropy given above refers to reversible processes. To calculate the variation of entropy after an irreversible process in an isolated system, it is possible to substitute the transformation with any other reversible one having the same initial and final state.

In 1865 Clausius gave the best known phrasing of the second law:

The entropy of the universe tends to a maximum

Second law of thermodynamics v.2

In terms of time variation, the mathematical statement of the second law for an isolated system undergoing an arbitrary transformation is:

$$\frac{\mathrm{d}S}{\mathrm{d}t} \geq 0$$

Proof. see [16] section 1.4 □

Remark: The second law is naturally connected to the concept of irreversibility and the so-called arrow of time.This last version refers to the fact that the entropy of an isolated system increases if an irreversible process occurs and remains constant if it is reversible. Thus entropy quantifies this fundamental concept.

1.2 Boltzmann's work on the second law

Ludwig Boltzmann was born in Vienna in 1844. He had a major role in the development of statistical mechanics and the atomistic hypothesis whose stories are strongly interconnected. About the atomism, it may sound strange, but still in the late 19th century prominent scientists doubted the existence of atoms - including Max Planck, who later became the father of quantum mechanics. The first modern attempt to take up the atomistic idea is due to Daniel Bernoulli, who in 1738 proposed the theory describing matter as a collection of many microscopic particles (the atoms) in motion. Bernoulli was able to calculate atmospheric pressure by assuming air consists of a collection of particles and the pressure determined by the collisions of these particles with walls. Hence statistical mechanics had been born, as a program for determining macroscopic properties from microscopic dynamics. It was not until the first half of the nineteenth century that the first relevant development occurred with Clausius (1857), who was able to give an initial explanation of the thermodynamic behavior of gases in terms of particles moving at great speed colliding with each other and with the walls of the vessel containing them. However, as previously said, the thermodynamic theory was progressing and it seemed at odds with the atomistic hypothesis. How is it possible to reconcile the irreversible behaviors of the thermodynamics with the reversible ones of the atomistic dynamics?

Before Boltzmann, the first intellectual to use probability in physics was Maxwell. He discovered a method to calculate the number of particles with a velocity (in module) between v and $v+\Delta v$ with Δv sufficiently small. This discovery can be represented by a probability density f on $\mathbb{R}^+$ that allows one to calculate the probability for a particle to have a velocity between v_1

and v_2:

$$f(v) = 4\pi \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 e^{-\frac{m}{2kT}v^2} \tag{1.1}$$

where T is the temperature (constant) and k the Boltzmann's constant.

I am going to give the proof that (1.1) is a density

Proof. It is clearly measurable and ≥ 0 a.s.. Let $a = \frac{m}{2KT}$

$$\int_0^{+\infty} 4\pi (\frac{a}{\pi})^{\frac{3}{2}} v^2 e^{-av^2} \mathrm{d}v = \frac{4}{\sqrt{\pi}} \int_0^{+\infty} a^{\frac{3}{2}} v^2 e^{-av^2} \mathrm{d}v =$$

let $\quad x = \sqrt{a}v$

$$= \frac{4}{\sqrt{\pi}} \int_0^{+\infty} x^2 e^{-x^2} \mathrm{d}x =^*$$

reminding : $\qquad X \sim N_{0,1} \quad \Longrightarrow Var(X) = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \mathrm{d}y = 1$

$$\Longrightarrow \ ^* = \int_{-\infty}^{+\infty} \frac{4}{2\sqrt{\pi}} x^2 e^{-x^2} \mathrm{d}x = 1$$

with the substitution $\quad y = \sqrt{2}x$

□

Boltzmann, inspired by Maxwell's work, decided to pursue the construction of a statistical theory of thermodynamics. His main objective was to give an explanation of the second law of thermodynamics through a microscopic view, using classical mechanics and dynamics. Thus, he supported the atomistic hypothesis and based his studies on the idea that gases are systems of particles. He published his most important work in 1872, *Further Studies on the Thermal Equilibrium of Gas Molecules* [2], where he stated Boltzmann's equation and the H-theorem. I am going to present his work, referring to [1].

Def 1.2.1

A *thermodynamic macrostate* is a state of the system, identified by its temperature, pressure and volume.
A *microstate* of a system is a particular configuration of its N particles, thus their position and velocity.

Remark: One macrostate is usually associated with multiple microstates.

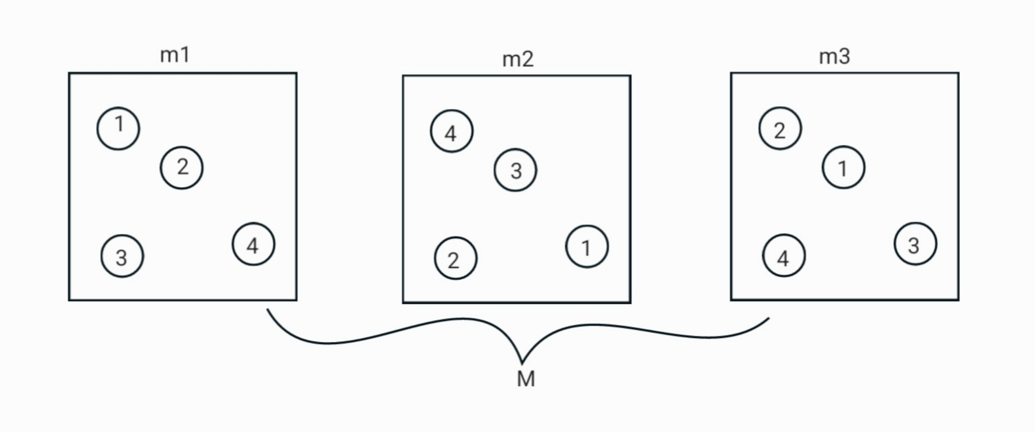


Figure 1.4: $m1, m2, m3$ are three different micorstates which correspond to the same macrostate

1.2.1 ergodic hypothesis

For a better understanding of the section refer to [3] chapter 1.2. and [4] chapter 1.3

Let us consider a system of N particles and denote $\left(q^{(i)}, p^{(i)}\right)$ the canonical Hamiltonian coordinates of the $i-th$ particle, where q is the position and p the momentum (note: $p = \frac{\partial \mathcal{L}}{\dot{q}}$ where $\mathcal{L}$ is the Lagrangian). Each particle has three degrees of freedom, thus the whole system can be represented by the vector $X(t) = (q^1(t), \ldots, q^N(t), p^1(t), \ldots, p^N(t))$ in the phase space Γ of dimension $6N$. Otherwise, we can think of N particles in the phase space γ of dimension 6. Measurable and observable properties of the system, such as its energy, are functions $\Phi(X)$. The particles dynamics follows classical mechanics, therefore it is described by the Hamilton's equations. Thus, if the Hamiltonian does not explicitly depend on time, the energy of the system stays constant, as in our case of study.

Note that, while a macroscopic property, such as the temperature, is being measured, the particles are still in motion (even if the system is in equilibrium), and they pass through many microstates. If the measured value refers to the function $\Phi(X)$, the macroscopic data is like a mean, and has to be confronted with an average, calculated through the evolution of the system:

$$\overline{\Phi}(t_0, \tau) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} \Phi(X(t)) \quad \mathrm{d}t \tag{1.2}$$

To compute this mean it would be necessary to know the exact microscopic configuration of a system in a precise time and the motion of every particle, which is clearly an unfulfillable request. The idea of Boltzmann was to substitute the temporal mean with one that uses a probability distribution.

Let Σ_E be the surface of constant energy, subset of the phase space in which all the configurations of microstates lie. Considering the errors in measuring, let us consider $\Gamma_{E\pm\Delta E}$, with ΔE much smaller than E, which is like a layer that contains Σ_E. Because of the conservation of energy, the dynamics of the system is confined within the layer.

If we think of N particles in space γ, we can divide the space into a sufficient number of cells $c_1, c_2 \dots$ of the same volume c and consider macroscopically equivalent all the microscopic states which differ in movements of particles within the cells and eventually differ in permutations of particles from one cell to another (the permutations within the cell are already accounted for in the movements). In this context the macroscopic states are characterized by the numbers of particles in each cell $N_1, N_2 \dots$ or the densities:

$$\delta_j = \frac{N_j}{N \cdot c}, \qquad \sum_j \delta_j \cdot c = 1$$

Boltzmann assumed that by sending the volume c to zero and increasing the number N it is possible to obtain an actual continuous probability density $\delta : \gamma \to \mathbb{R}$. Boltzmann did not worry about this limit, which in reality results extremely difficult to prove. Although, he studied the numbers N_j: for every choice of the numbers N_j of the cells, a set $C(N_1, N_2 \dots)$ in $\Gamma_{E \pm \Delta E}$ is related. Of this set, we can define and work with its volume $C_{N1,N2\dots} = \frac{N!}{N_1!N_2!\dots}c^N$ (proof of the well-defined nature of C can be found in [4] c. 1.3).

The **ergodic hypothesis** states that: the macroscopic dynamics are such that the point X in Γ, representing the microscopic state of the system, travels all the layer $\Gamma_{E \pm \Delta E}$ and spends in each volume C a time that is on average proportional to C itself. Therefore, if the system is observed at a randomly chosen instant over a very long period of time (to allow the trajectory time to explore the entire constant energy surface), the probability of finding the system in a generic set coincides with its volume C, up to normalization. There exists a maximum volume C^* (see [3] pag.4) of $C(N_1, N_2 \dots)$ with E and N fixed, and, if the ergodic hypothesis is true, one can conclude that:

1. However the system is set, even far away from thermodynamic equilibrium, (i.e. in a configuration represented by C_i) the microscopic dynamics will lead the system to C^*

2. Once the system is in C^* it will spend there most of the time, except for extremely rare, **but possible**, fluctuations.

The state C^* is called *Maxwell-Boltzmann state* and in this state the particles' velocities follows the distribution (1.1) (for the proof see [16] pag. 81-82). From a practical perspective, the equilibrium is indistinguishable from the Maxwell-Boltzmann state: in principle, however, equilibrium is a collection of macroscopic states, and although one of them is by far dominant, fluctuations are not a priori ruled out.

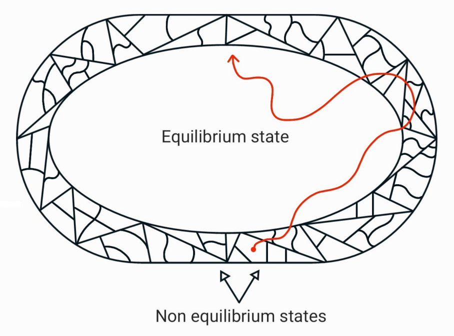


Figure 1.5: The equilibrium state occupies most of the volume, thus the system is spending there most of the time

This interpretation of the volume in phase space as the probability *a priori* of a set of microscopic states is commonly called *principle of equiprobability of microscopic states*. When the thermodynamic equilibrium is reached, the principle tells that its microscopic state is equally likely to be any state satisfying the macroscopic condition of the state. The entirety of statistical mechanics, in all its formulations, relies on this principle. Boltzmann's ergodic hypothesis represents a possible dynamic justification for it.

Hence, if τ is big enough to allow the trajectory time to explore the entire constant energy surface, the mean (1.2) depends only on the energy of the system and has the same value for all the evolutions with same energy. This value is calculable by doing an average on $\Phi(X)$ where the microstates with the same energy contribute in the same way:

The density function of uniform probability in the region $\Gamma_{E\pm\Delta E}$ is the so called *microcanonical ensemble*. Denoting this density as $\rho(E)$ and the element of volume of the phase space of the system $\mathrm{d}X = \mathrm{d}q^1, \dots, \mathrm{d}q^N, \mathrm{d}p^1, \dots, \mathrm{d}p^N$:

$$\rho(E) = \left(\int_{E-\Delta E \leq H(X) \leq E+\Delta E} \mathrm{d}X \right)^{-1} 1_{\Gamma_{E\pm\Delta E}}$$

and the ergodic hypothesis allow us to write:

$$\overline{\Phi} \equiv \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} \Phi(X(t)) \quad \mathrm{d}t = \int_{R^{6N}} \Phi(X)\rho(X)\mathrm{d}X \equiv \mathbb{E}[\Phi(X)] \equiv \langle \Phi \rangle \tag{1.3}$$

Note that, if this relation holds, we do not have neither to determine the initial state of the system, nor calculate the integral with respect to the time. (A modern rigorous view of the ergodicity and the proof of (1.3) can be found in [11] c.2)

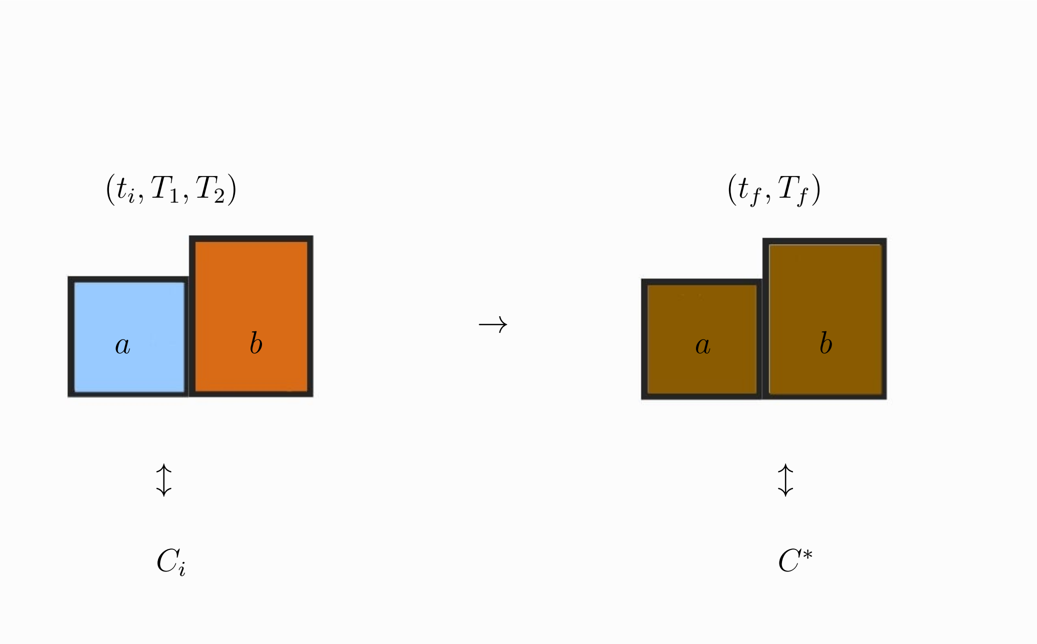
Boltzmann's Entropy

The principle of equiprobability allows us to define the microcanonical ensemble, previously introduced, that describes the equilibrium of an isolated system, whose macrostate is determined by its energy E, volume V, and number of particles N. To establish a connection between thermodynamics and the microscopical world, in 1877 Boltzmann redefined entropy as:

$$S = k_B \ln C(E, V, N) \tag{1.4}$$

where k_B is the Boltzmann's constant, and the volume C, previously defined, denotes the number (or phase-space volume) of microstates compatible with the macrostate. This definition is perfectly aligned with the second law of thermodynamics because the system tends to the state C^*, which is the one that occupies most of the volume in the phase space. The logarithm is monotonic, thus $\Delta S \geq 0$

example: Let a, b be two objects at different initial temperatures $T1, T2$. Suppose the bodies are brought into contact. The system of the two bodies together, has a state corresponding to C_i at time t_i. After sufficient time, at time t_f, the system will reach the equilibrium, the two bodies will have the same temperature T_f, and the state of the system will be the one of Maxwell-Boltzmann.



Intuitively, at t_0 the particles exhibit a more ordered arrangement than at t_f, hence there are less suitable microstates correspondent to that macrostate than at t_f, and thus the volume in the phase space is lower.

Considerations on the ergodic hypothesis

The validity of (1.3), i.e., the possibility of replacing the average of any observable along a time evolution with an average of the observable in phase space constitutes the *ergodic problem*. The problem is far from being demonstrable in practice, in physically realistic systems: but it has the great value of showing that in principle it makes sense to discuss probability in a deterministic system, and that macroscopic behavior is not incompatible with microscopic dynamics. Modern statistical mechanics uses instead of the ergodic hypothesis, a less strong concept, called *mixing*, to justify the principle of equiprobability.

1.2.2 The Boltzmann equation and the H theorem

The interesting observable property of thermodynamics, have a value (macroscopically) almost constant on the wholeness of the assigned surface of energy, but not all of it. Thus, they allow the existence of non-equilibrium states. If a system with fixed energy is initially in a non-equilibrium state, which occupies a "small" region on the surface of constant energy, it is expected that its time evolution will lead it towards the equilibrium region, which has a dominant volume. It is necessary to demonstrate that, at least in some simplified cases, this expected behavior is indeed a consequence of the laws of dynamics. A first step in this direction was taken by Boltzmann with his transport equation (1872), commonly known as Boltzmann equation. In particular, Boltzmann aimed to address the following problems:

- Deriving, within the framework of Newton mechanics, the Maxwell-Boltzmann distribution (1.1) of the molecules of gas in thermodynamic equilibrium.
- Showing that, starting from an arbitrary probability distribution, the system asymptotically approaches the Maxwell-Boltzmann distribution, still within the realm of classical mechanics.

The key idea is to study the time evolution of a certain "collective" variable that provides reduced information about the microscopic state of the system: namely, $f_X(q,p) : \mathbb{R}^6 \to \mathbb{R}$, the physical distribution function describing how the N molecules of a gas are distributed in the phase space γ. By definition, $f_X(q,p)\mathrm{d}q\mathrm{d}p$ counts the number of molecules that, in a given microscopic state X of the system, are found in the phase space volume element $\mathrm{d}q\mathrm{d}p$ around the point (q,p). It is important to stress that this function, as indicated by the subscript X, is defined in terms of the microscopic state $X = (q_1, \ldots, q_n, p_1, \ldots, p_n)$ of the entire system of N molecules. The quantity $\frac{f_X(q,p)}{N}$ can be interpreted as a sort of probability density, called "empirical" probability density (because it is not continuous), in the sense that $\int_{\mathbb{R}^6} \frac{f_X(q,p)}{N}\mathrm{d}q\mathrm{d}p = 1$. The function f_X evolves in time because it is determined by the microscopic state X, which evolves according to Hamilton's equations.

For a dilute gas, that is, one with sufficiently low density and high temperature, Boltzmann started from an exact equation describing the evolution of a generic $f_X(t)$ and, by introducing the hypothesis of **molecular chaos** derived his transport equation. From the Boltzmann equation it then follows:

- (a) there exist a stationary solution $\hat{f}$, given by the Maxwell-Boltzmann distribution
- (b) the time evolution drives any initial distribution f towards $\hat{f}$.

The Boltzmann equation thus describes the "correct" behavior for $f_X(t)$, but it is obtained at the cost of an additional non-dynamical assumption, namely the hypothesis of molecular chaos. (This hypothesis has been removed in the 70's by the Theorem of Lanford).

From the equation a specific quantity H can be extracted, and results that $\frac{\mathrm{d}H}{\mathrm{d}t} \leq 0$. By setting the entropy S proportional to $-H$ the H-theorem seemed to demonstrate the second law of thermodynamics, in terms of kinetic theory.

1.3 Ehrenfest model

Criticism on Boltzmann's Work

After the release of his H-theorem, many intellectuals were still skeptical: there was still no proof of the existence of atoms, furthermore, there was criticism about the apparent paradox of irreversibility. In contrast to the second law, Boltzmann's work seemed to have left space for the possibility that an irreversible process could reverse itself. For instance, one famous paradox was presented by E. Zermelo: he noticed that the H-theorem disagreed with the Poincaré Recurrence Theorem, which states that in every mechanical system, whose motion occurs within a bounded region of the phase space, at a certain finite time t_r, the system will go back as closed as desired to its original state. (For the proof see [16] pag. 91).

Boltzmann answered by underlining that they were confusing the concept of "impossibility" with that of "probabilistically extremely rare", and that it was necessary to interpret the second law as a statistical theorem. According to him, it was indeed possible for the system to go back to its original state, but only in a time longer than the age of the universe, and thus not observable. The crucial point was that "at a certain time" of the Recurrence Theorem, for example, in $1\,\mathrm{cm}^3$ of gas, to return to the initial condition, with a precision of $10^{-9}\,\mathrm{m}$ for the positions and $1\,\mathrm{m/s}$ for the velocities, we would have to wait for a time proportional to $10^{10^{19}}$ years. Boltzmann's words were: *"The second law [of thermodynamics] can never be proved mathematically by means of the equations of dynamics alone. [...] What I have proved is the following: It is overwhelmingly probable that $H(t)$ will be very close to its minimum value; if it is larger, it may increase or decrease, but the probability that it decreases is far greater."*

Ehrenfest model

A probabilistic model that shows the Boltzmann's viewpoint was proposed by Paul and Tatiana Ehrenfest in 1907:

Suppose we have N numbered balls, such that they are distinguishable. Let A, B be two urns, and suppose the balls are distributed in the two urns. For example A contains q balls and thus B contains $N - q$ balls. Suppose now that we have a bag with N tickets numbered from 1 to N. Every second a ticket is extracted and its number is read, then it is reinserted in the bag. Every time a number is read, the corresponding ball is moved from its urn to the other. It is clearly more probable that the drawn number is in the urn with more balls, for instance if $N = 100$ and $q = 80$ the probability of extracting a ball in urn A is 80%. The urns A, B represent two bodies in contact, and we can think of the balls as representing the temperature of the body, where more balls mean warmer.

Again suppose $N = 100$ and $q = 80$. As long as there are more balls in A, we will witness the emptying of the urn. However, there will be fluctuations. The second law of thermodynamics stated that heat flows in only one direction and that at no time can the flow be inverted. Thus, that "no" now becomes "extremely rare".

For exact equipartition to be possible we assume N is even. Now let us consider $[N_A, N_B] \equiv [N_A, N - N_A]$, the number of balls in A and B. We can define $[N_A, N_B]$ as the macrostate, to which many microstates are associated, specifically W

$$W \equiv \frac{N!}{N_A! \cdot N_B!} = \binom{N}{N_A}$$

which is the number of ways we can choose N_A elements in a set of N. We can thus use the formula introduced by Boltzmann to calculate the entropy $S = k\ln(W)$. I am going to show that the maximum of S is reached for the macrostate $[\frac{N}{2}, \frac{N}{2}]$. The maximum for S is the maximum for $\ln(W)$ with $W = \binom{N}{N_A}$. For $N \gg 1$ the Stirling approximation ([10]) applies:

$$\lim_{N\to\infty} \frac{\ln N!}{N\ln N - N + \frac{1}{2}\ln(2\pi\mathrm{N})} = 1$$

Thus, using the properties of logarithm:

$$\begin{aligned} S(N_A) \sim & \Big[N\ln N - N + \frac{1}{2}\ln(2\pi N)\Big] \quad - \Big[N_A \ln N_A - N_A + \frac{1}{2}\ln(2\pi N_A) + \Big] \\ & - \Big[(N - N_A)\ln(N - N_A) - (N - N_A) + \frac{1}{2}\ln(2\pi(N - N_A)) + \Big] \\ & = N\ln N - N_A \ln N_A - (N - N_A)\ln(N - N_A) + \frac{1}{2}\ln\left(\frac{N}{2\pi N_A(N - N_A)}\right) \end{aligned}$$

And for $N \gg 1$ the last term can be ignored since it is $O(\frac{1}{N})$. Hence we can approximate

$$S \sim N\ln N - N_A \ln N_A - (N - N_A)\ln(N - N_A)$$

Now we derive S with respect to N_A:

$$\frac{dS}{dN_A} = -\ln N_A + \ln(N - N_A)$$

And setting $\frac{dS}{dN_A} = 0$:

$$\ln\left(\frac{N - N_A}{N_A}\right) = 0 \implies \frac{N - N_A}{N_A} = 1 \implies N_A = \frac{N}{2}$$

Now calculating the second derivative we establish that it is a maximum:

$$\frac{d^2S}{dN_A^2} = -\frac{1}{N_A} - \frac{1}{N - N_A} = \frac{-N}{N_A(N - N_A)} < 0 \quad \forall N_A,\ 0 < N_A < N$$

Thus the macrostate $[\frac{N}{2}, \frac{N}{2}]$ represents the thermodynamic equilibrium, as discussed in the previous chapters.

Now, I am going to show that the equilibrium is actually reached. Let us consider the so called "average excess" over $\frac{N}{2}$ of the number of balls in the first urn after s drawings. Let $X_s = N_A - \frac{N}{2}$, and suppose X_0 is given and determined. For example, $X_0 = 30$ means that in the initial situation there are $\frac{N}{2} + 30$ balls in the first urn, and X_1 has values either 29 or 31. In fact we can say that

$$X_{S+1} = X_s + Z_{s+1} \text{ where } Z_{S+1} = \begin{cases} -1 \text{ with probability } \frac{X_s + \frac{N}{2}}{N} \\ 1 \text{ with probability } \frac{-X_s + \frac{N}{2}}{N} \end{cases}$$

We want to calculate the expected value $\mathbb{E}[X_{s+1}]$ (see appendix A for probability reminders):

$$\mathbb{E}[X_{s+1}] = \mathbb{E}[\mathbb{E}[X_{s+1}|X_s]]$$

$$\mathbb{E}[X_{s+1}|X_s] = \mathbb{E}[X_s + Z_{s+1}|Xs] = \mathbb{E}[X_s] + \mathbb{E}[Z_{s+1}|X_s] = *$$

considering only the second element:

$$\begin{aligned}\mathbb{E}[Z_{s+1}|X_s] &= -1 \cdot \frac{X_s + \frac{N}{2}}{N} + 1 \cdot \frac{-X_s + \frac{N}{2}}{N} \\ &= -\frac{2X_s}{N}\end{aligned}$$

Thus:

$$* = \mathbb{E}[X_S](1 - \frac{2}{N})$$

and by iterating:

$$\mathbb{E}[X_{s+1}] = \mathbb{E}[X_0](1 - \frac{2}{N})^{s+1} = X_0(1 - \frac{2}{N})^{s+1}$$

This shows that for $s \gg 1$ the expected value tends to 0 and thus the system approaches equilibrium.

Let us also take into account the relative fluctuations of X_s:

$$Var(X_{s+1}) = \mathbb{E}[X_{s+1}^2] - \mathbb{E}[X_{s+1}]^2 = \mathbb{E}[X_{s+1}^2] - X_0^2(1 - \frac{2}{N})^{2(s+1)} = *$$

$$\begin{aligned}
&\mathbb{E}[X_{s+1}^2] = \mathbb{E}[X_s^2 + Z_{s+1}^2 + 2X_sZ_s + 1] = \mathbb{E}[\mathbb{E}[X_s^2 + Z_{s+1}^2 + 2X_sZ_{s+1}|X_s]] \\
&= \mathbb{E}[X_s^2] + \mathbb{E}[\mathbb{E}[Z_{s+1}^2|X_s]] + \mathbb{E}[\mathbb{E}[2X_sZ_{s+1}]] \\
&= \mathbb{E}[X_s^2] + \mathbb{E}[1] + \mathbb{E}[2X_s\mathbb{E}[Z_{s+1}|X_s]] \\
&= \mathbb{E}[X_s^2] + 1 - \frac{4}{N}\mathbb{E}[X_s^2] \\
\implies * &= \left(1 - \frac{4}{N}\right)\mathbb{E}[X_s^2] + 1 - X_0^2(1 - \frac{2}{N})^{2(s+1)}
\end{aligned}$$

Now if we consider the limit $s \to +\infty$, because the last term tends to 0:

$$\begin{aligned}
Var(X_\infty) &= \left(1 - \frac{4}{N}\right)Var(X_\infty) + 1 \\
\implies Var(X_\infty) &= \frac{N}{4}
\end{aligned}$$

Hence, the standard deviation is $\sigma = \sqrt{\frac{N}{4}}$ and if we consider the relative (i.e. with respect to the number of particles) standard deviation:

$$\frac{\sigma}{N} = \frac{\sqrt{N}}{2N} = \frac{1}{\sqrt{N}2}$$

which tends to 0 for $N \to +\infty$. So the relative fluctuations $\sigma/N \sim 1/\sqrt{N}$ become small for large N, which is consistent with thermodynamic behavior.

1.3.1 Recurrence Time

I showed that Ehrenfest's model is particularly suited for modeling Boltzmann's problem, and gave a logical statistical proof of why Boltzmann's theory made sense. However, Ehrenfest did not show an explicit way of calculating the Poincaré cycles (time intervals after which states "almost recur" for the first time) mentioned in Zermelo's paradox. I now present the one shown later by Mark Kac in [17]. Let first suppose that the time between two drawings is τ, and thus that the time necessary for doing c drawings is $c\tau$. Let us now set some notation: $P(X_{s+c} = m|X_s = n)$ is the probability that, given $X_s = n$, after c drawings, the excess of balls in urn A is m. Since $(X_n)_{n\in\mathbb{N}}$ is a homogeneous Markov chain, following Kac's notation, we set $P(X_{s+c} = m|X_s = n) \equiv P(n|m;c)$. Similarly, $P'(X_{s+c} = m|X_s = n) \equiv P'(n|m;c)$ is the probability that, after c drawings, m will be the excess observed on the first urn *for the first time*, if the excess was n at the beginning. In particular $P'(n|n;c)$ is the probability that the recurrence time of the state "n" is $c\tau$. The main contribution given by Kac is the following:

Theorem

$$P(m|n;c) = \frac{(-1)^{\frac{N}{2}+n}}{2^N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \left(\frac{2j}{N}\right)^c \cdot \mathrm{C}^{(-n)}_{\frac{N}{2}+j} \cdot \mathrm{C}^{(j)}_{\frac{N}{2}+m} \tag{1.5}$$

where the C's are defined by the identity:

$$(1-z)^{\frac{N}{2}-j}(1+z)^{\frac{N}{2}+j} = \sum_{k=0}^{N} \mathrm{C}^{(j)}_k z^k$$

Proof. See [17] Chapter 4. □

I now show two results:

1.
$$\sum_{c=1}^{\infty} P'(n|n;c) = 1 \tag{1.6}$$

This means that each state is bound to recur with probability 1. It is the equivalent of Poincaré's Theorem.

2. The mean recurrence time, namely, the quantity

$$\theta_n = \sum_{c=1}^{\infty} c\tau P'(n|n;c)$$

is equal to

$$\tau \frac{(\frac{N}{2}+n)!(\frac{N}{2}-n)!}{N!} 2^N \tag{1.7}$$

This is the statistical analogue of a "Poincaré cycles", and it tells us, how long, on the average, one will have to wait for the state "n" to recur.

Remark: If $\frac{N}{2}+n$ and $\frac{N}{2}-n$ differ considerably θ_n is extremely long, while if they are nearly equal θ_n is short. This explain the concept of irreversibility of thermodynamic processes, in the sense that the statistical time for a system to go back to its original non-equilibrium state would be larger than the age of the universe.

Example: [extremely arranged state $[N,0]$**]** We set $N = 100,\ n = 50,\ \tau = 10^{-10} seconds$

$$\Longrightarrow\ \theta_n = 10^{-10} \cdot 1 \cdot 2^{100} \approx 4 \cdot 10^{12}\ years$$

approximately 4 times the age of the universe. Clearly θ_n explodes when N is very large.

Example: [equilibrium state $[\frac{N}{2}, \frac{N}{2}]$**]** We set $N = 100,\ n = 0,\ \tau = 10^{-10} seconds$

$$\Longrightarrow\ \theta_n = 10^{-10} \cdot \frac{50!^2}{100!} \cdot 2^{100} \approx 1.25 \cdot 10^{-9} s$$

using Stirling's approximation.

I now give the proofs of (1.6) and (1.7).

Lemma:

$$P(n|m;c) = P'(n|m;c) + \sum_{k=1}^{c-1} P'(n|m;k)P(m|m;c-k) \tag{1.8}$$

Proof. We consider all possible ways of reaching "m" from "n":

1. "m" is reached for the first time after n steps
2. "n" is reached for the first time in 1 step and then, starting from "m" it is again reached in $c-1$ steps.
3. "m" is reached for the first time in 2 steps and then, starting from "m", it is reached again in $c-2$ steps, and so forth. We note furthermore that the probability that "m" will be reached for the first time in k steps and then, starting from "m" it will be reached again in $c-k$ steps, is

   $$P'(n|m;k)P(m|m,c-k)$$

   where we are using the product probability because in our process the past is independent of the future since X_s is a Markov chain. When we know that $X_s = n$ the subsequent behavior of the process is independent of the way in which "m" was reached.

□

Now I introduce the generating functions:

$$h(n|m;z) = \sum_{c=1}^{\infty} P(n|m;c)z^c$$

$$g(n|m;z) = \sum_{c=1}^{\infty} P'(n|m;c)z^c$$

A generating function is a representation of an infinite sequence of numbers as the coefficient of a power series. We note that

$$h(n|m;z) = \sum_{c=1}^{\infty} P'(n|m;c)z^c + \sum_{c=1}^{\infty}\sum_{k=1}^{c-1} P'(n|m;k)P(m|m;c-k)z^c$$

introducing $j = c - k, \qquad k \geq 1,\ k \leq c-1,\ c \geq 1$

$$h(n|m;z) = g(n|m;z) + g(n|m;z)h(m|m;z)$$

equivalently

$$g(n|m;z) = \frac{h(n|m;z)}{1 + h(m|m;z)}$$

in particular:

$$g(n|n;z) = \frac{h(n|n;z)}{1 + h(n|n;z)} = 1 - \frac{1}{1 + h(n|n;z)} \tag{1.9}$$

And we note that:

$$\frac{\mathrm{d}g(n|n;z)}{dz} = \frac{\frac{\mathrm{d}h(n|n;z)}{dz}}{(1 + h(n|n;z))^2} \tag{1.10}$$

From the definition of $g(n|n;z)$ we obtain:

$$\lim_{z\to 1} g(n|n;z) = \sum_{c=1}^{\infty} P'(n|n;c) \tag{1.11}$$

$$\tau \lim_{z\to 1} \frac{\mathrm{d}g(n|n;z)}{dz} = \sum_{c=1}^{\infty} \tau c P'(n|n;c) \tag{1.12}$$

It is from this formulas that we derive (1.6) and (1.7).

From (1.5):

$$h(n|n;z) = \frac{(-1)^{\frac{N}{2}+n}}{2^N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \sum_{c=1}^{\infty} \left(\frac{2jz}{N}\right)^c \cdot \mathrm{C}^{(-n)}_{\frac{N}{2}+j} \cdot \mathrm{C}^{(j)}_{\frac{N}{2}+n}$$

and:

$$P(n|n;0) = 1 = \frac{(-1)^{\frac{N}{2}+n}}{2^N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \mathrm{C}^{(-n)}_{\frac{N}{2}+j} \cdot \mathrm{C}^{(j)}_{\frac{N}{2}+m}$$

we obtain:

$$1 + h(n|n;z) = \frac{(-1)^{\frac{N}{2}+n}}{2^N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \sum_{c=0}^{\infty} \left(\frac{2jz}{N}\right)^c \cdot \mathrm{C}^{(-n)}_{\frac{N}{2}+j} \cdot \mathrm{C}^{(j)}_{\frac{N}{2}+n} \tag{1.13}$$

$$= \frac{(-1)^{\frac{N}{2}+n}}{2^N} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}} \frac{1}{1 - \frac{2j}{N}z} \mathrm{C}^{(-n)}_{\frac{N}{2}+j} \cdot \mathrm{C}^{(j)}_{\frac{N}{2}+n} \tag{1.14}$$

Where I used $\sum_{c=0}^{\infty}\left(\frac{2jz}{N}\right)^c = \frac{1}{1-\frac{2j}{N}z}$ for $|z|<1$. All term on the right side of (1.14) are holomorphic in $z=1$ except for $j=\frac{N}{2}$, which has a simple pole at that point. Thus, we can write:

$$1+h(n|n;z) = p(z) + \frac{(-1)^{\frac{N}{2}+n}}{2^N}\mathrm{C}_N^{(-n)}\cdot \mathrm{C}_{\frac{N}{2}+n}^{(\frac{N}{2})}\frac{1}{1-z} \qquad |z|<1$$

where p(z) is holomorphic in $z=1$

Now we see that:

$$\lim_{z\to 1} 1+h(n|n;z) = \infty$$

And hence, using (1.9) and (1.11):

$$\sum_{c=1}^{\infty} P'(n|n;c) = \lim_{z\to 1} g(n|n;z) = \lim_{z\to 1} 1 - \frac{1}{1+h(n|n;z)} = 1$$

Now we can compute $\mathrm{C}_N^{(-n)}\cdot \mathrm{C}_{\frac{N}{2}+n}^{(\frac{N}{2})}$ thanks to the identities:

$$(1) \quad (1-z)^{\frac{N}{2}+n}(1+z)^{\frac{N}{2}-n} = \sum_{k=0}^{N}\mathrm{C}_k^{(-n)}z^k$$

$$(2) \quad (1+z)^N = \sum_{k=0}^{N}\mathrm{C}_k^{(\frac{N}{2})}z^k$$

$\mathrm{C}_N^{(-n)}$ is the maximum degree coefficient in equation (1) which corresponds to $(-1)^{\frac{N}{2}+n}$.
$\mathrm{C}_{\frac{N}{2}+n}^{(\frac{N}{2})}$ is the coefficient of $z^{\frac{N}{2}+n}$ in (2), that is, using the binomial theorem: $\binom{N}{\frac{N}{2}+n}$.

Hence,

$$\frac{(-1)^{\frac{N}{2}+n}}{2^N}\mathrm{C}_N^{(-n)}\cdot \mathrm{C}_{\frac{N}{2}+n}^{(\frac{N}{2})} = \binom{N}{\frac{N}{2}+n}\cdot\frac{1}{2^N} = \frac{1}{2^N}\frac{N!}{(\frac{N}{2}+n)!(\frac{N}{2}-n)!}$$

And we denote this last expression as w. To conclude,using (1.10) we have:

$$\begin{aligned}\sum_{c=1}^{\infty} \tau P'(n|n;c) &= \tau \lim_{z\to 1} \frac{\mathrm{d}g(n|n;z)}{dz} = \lim_{z\to 1} \tau \frac{\frac{\mathrm{d}h(n|n;z)}{dz}}{(1+h(n|n;z))^2} \\ &= \lim_{z\to 1} \tau \frac{p'(z) - w\frac{1}{(1-z)^2}}{(p(z)+w\frac{1}{1-z})^2} = \lim_{z\to 1} \tau \frac{(1-z)^2 p'(z) - w}{[(1-z)p(z)+w]^2} = \tau \frac{1}{w} \\ &= \tau \frac{(\frac{N}{2}+n)!(\frac{N}{2}-n)!}{N!} 2^N\end{aligned}$$

This concludes the proofs.

Chapter 2

Kac's definition of Chaos

As mentioned in section 1.2.2, the Boltzmann equation is derived from first principles under a crucial assumption, called *molecular chaos*. This assumption was already known from Maxwell and is often called the *Stosszahlansatz* since Ehrenfest. Informally, it translates the idea that, despite the multitude of interactions, two particles taken at random should be statistically independent when the total number of particles grows to infinity. It is not so clear how the appearance of probability theory should be interpreted in this context. The rigorous justification of the work of Boltzmann and the status of molecular chaos became true mathematical questions when Hilbert addressed them in his Sixth Problem at the Paris International Congress of Mathematicians in 1900. Quoting Hilbert: *"Boltzmann's work on the principles of mechanics suggests the problem of developing mathematically the limiting processes, there merely indicated, which lead from the atomistic view to the laws of motion of continua"*.

Regarding this request, I am going to present the first ideas of Mark Kac, referring to [5]. More than half a century after Hilbert, Kac gave the first rigorous mathematical definition of chaos and introduced the idea that for time-evolving systems, chaos should be propagated in time, a property that is therefore called the propagation of chaos.

2.1 Representation of symmetric particle systems

particle systems

Def (Polish metric Space)

A metric space (E, τ) is said *polish* if it is separable and complete.

example: $\mathbb{R}^d$ with the Euclidean distance is a Polish metric space.

From now on I will consider E as at least a Polish metric space.

Def (Stochastic process)

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. A *stochastic process* indexed by $I \neq \varnothing$, is a family $X = (X_i)_{i \in I}$ where X_i is a random variable for every i, with values in $(E, \mathcal{G})$, a measurable space.

example: If $I = \mathbb{N}$ and $E = \mathbb{R}$ the stochastic process X is a sequence of random variables $(X_n)_{n \in \mathbb{N}}$

A system of N particles that evolves in time can be represented as

$$\mathcal{X}_I^N \equiv (\mathcal{X}_t^N)_{t \in I} \equiv \left(X_t^{1,N}, ..., X_t^{N,N} \right)_{t \in I}$$

where each particle $(X_t^{i,N})_{t \in I}$ is a stochastic process with values in E (we consider it with the $Borel \quad \sigma - algebra$) which is defined on a time interval $I = [0, T]$ with $T \in [0, +\infty)$.
For simplicity, we can think of E as $\mathbb{R}^d$ (in particular $\mathbb{R}^6$ as it is the usual phase space for a particle in a gas)

In this case study, I will not consider the time evolution of the system, and thus analyze the system at $t = 0$ fixed. I can thus set the representation as a $N-$ *dimensional* vector of random variables with values in E:

$$(X_0^{1,N},\dots,X_0^{N,N}) \equiv (X^1,\dots,X^N) \equiv \mathcal{X}^N$$

The N-particles are not independent, they are said to interact, due to collisions and interaction forces.

Def (Exchangeability)

A family $(X^i)_{i\in I}$ of random variables is said to be *exchangeable* when the law of $(X^i)_{i\in I}$ is invariant under every permutation of a finite number of indices.

In our study, this means:

$$Law(X^1,\dots,X^N) = Law(X^{\sigma(1)},\dots,X^{\sigma(N)}) \quad \sigma\in S_N$$

where S_N is the group of permutations of N elements.

- Exchangeability implies that $(X^i)_{i=1,\dots,N}$ are identically distributed (i.d.).

Proof.

$$\begin{aligned}\mu_{X_j}(H) &= \mu_X(E\times\dots\underset{j-1}{E}\times\underset{j}{H}\times\dots\times E) \overset{exchangeability}{=} \\ &= \mu_X(E\times\dots\underset{i-1}{E}\times\underset{i}{H}\times\dots\times E) = \mu_{X_i}(H)\end{aligned}$$

and this is true for every $H\in\mathcal{G}$ and every $i,j=1,\dots,N$

□

-However, i.d. does not imply exchangeability.

example: Let $Y=(Y_1,Y_2,Y_3,Y_4)$ such that $P(Y=(1,0,1,0))=\frac{1}{2}$ and $P(Y=(0,1,0,1)=\frac{1}{2}$. Hence, $Y_1\sim Y_2\sim Y_3\sim Y_4\sim Be_{\frac{1}{2}}$, however, if we permute Y_2 and $Y_3 \Rightarrow P(Y=(0,0,1,1))=\frac{1}{2}\neq 0$ thus Y is not exchangeable.

-In addition, i.i.d implies exchangeability, but the opposite is not true:

Let $(Y_i)_{i=0,\dots,N}$ be a set of random variables i.i.d., hence:

$$\mu(Y_1, Y_2, \dots, Y_N) = \mu_{Y_1} \otimes \mu_{Y_2} \otimes \cdots \otimes \mu_{Y_N}$$
$$= \mu_{Y_2} \otimes \mu_{Y_N} \otimes \cdots \otimes \mu_{Y_i} = \mu(Y_2, Y_N, \dots, Y_i)$$

And this remains true for every permutation.

example: Let $Y_1 \sim Be_{\frac{1}{2}}$ and $Y_2 = 1 - Y_1$, thus $Y_2 \sim Y1$. If we consider (Y_1, Y_2) it has values $(1, 0)$ and $(0, 1)$ with probability $\frac{1}{2}$. If we now consider (Y_2, Y_1), the distribution of the vector stays the same, because $Y_1 = 1 - Y_2$ too, therefore this system is exchangeable. However Y_1 and Y_2 are clearly not independent.

The exchangeability is the only but crucial assumption that is made on the system of N-particles

2.2 Statistical description

Finite particle system

In statistical physics, the previous description in terms of stochastic process is sometimes called the *microscopic scale* because the trajectory of each individual particle is recorded. When N is large, the microscopic scale contains too much information and a statistical description is sought. I am going to present two statistical points of view on the particle system, which is considered finite:

1. The first one is the $N-$**particle distribution**, as the joint distribution function of the vector $(X^1, \dots, X^N)$. We call it $f^N \in \mathcal{P}_{sym}(E^N)$, the space of symmetric probability measures on the space E^N, where symmetric refers to the property of exchangeability, and not the usual symmetry of distributions $(\mu_X = \mu_{-X})$.

The symmetry assumption implies that for any $k \leq N$, we can define the $k-$th marginal distribution $f_{k,N}$ on E^k such that:

$$\forall \varphi_k \in C_b(E^k), \qquad \langle f^{k,N}, \varphi_k \rangle = \langle f^N, \varphi_k \otimes 1^{\otimes N-k} \rangle$$

where $C_b(E^k)$ is the set of real-valued bounded continuous functions on E^k. The φ_k are called test functions and can be interpreted as physical observables on k particles (e.g., if $k = 2$ and $E = \mathbb{R}^6$ it could be a coulomb interaction, $\varphi_2(x_1, p_1, x_2, p_2) = \frac{1}{|x_1 - x_2|}$).The symbol $\langle f, \varphi \rangle$ indicates the integral of a measurable function φ on a measure μ. The definition can be re-written as:

$$\int_{E^k} \varphi_k \ \ \mathrm{d}f^{k,N} = \int_{E^N} \varphi_k \otimes 1^{\otimes N-k} \ \ \mathrm{d}f^N$$

The symbol $\varphi \otimes \psi$, in this case, indicates the product function on $A \times B$ defined by $(\varphi \otimes \psi)(x, y) = \varphi(x)\psi(y)$ for two real valued functions φ, ψ respectively on A and B.

The exchangeability ensures that the term on the right-hand side does not depend on the indices of the k variables; for instance, it would be equivalent to take $1^{\otimes N-k} \otimes \varphi_k$as a test function instead of $\varphi_k \otimes 1^{\otimes N-k}$.

$f^{k,N} \in \mathcal{P}_{sym}(E^k)$, in fact, if we take the indicator function $\varphi_k = 1_{E^k}$:

$$P(E^K) = \int_{E^N} 1_{E^k} \otimes 1^{\otimes N-k} = \int_{E^N} 1^{\otimes N} \mathrm{d}f^N = 1$$

And it is still symmetric.

example: Let $X = (X^1, X^2, X^3), \quad X \sim N_{0,C} \sim f^3$ where $C = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{bmatrix}$ Therefore (thanks to Prop 2.2 Appendix A), the system is exchangeable (not i.i.d), and $f^2 \sim N_{0,C_2}$ with $C_2 = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}$

Remark: Exchangeability $\implies Cov(X^i, X^j) = Cov(X^k, X^t)$ $\forall\ i \neq j,\ k \neq t.$

In fact, if we suppose that $X \sim N_{0,C}$ with $C = \begin{bmatrix} 1 & 0.1 & 0.5 \\ 0.1 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{bmatrix}$ When we permute X_1 with X_2 we have $X^2 \sim N_{0,C^2}$ with $C_2 = \begin{bmatrix} 1 & 0.1 & 0.2 \\ 0.1 & 1 & 0.5 \\ 0.2 & 0.5 & 1 \end{bmatrix}$, thus the system is not exchangeable.

2. Alternatively, an exchangeable system is equivalently described by its (random) **empirical measure**: for a fixed realization of $N-$particles $x = (x^1, \ldots, x^N) \in E^N$, the empirical measure is defined as

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i} \ \in \ \mathcal{P}(E) \tag{2.1}$$

We can define the map:

$$\mu_N : E^N \to \mathcal{P}(E), \quad x \mapsto \ \mu_N(x) \equiv \mu_x \tag{2.2}$$

and thus study the random empirical measure $\mu_{\mathcal{X}^N} \equiv \mu_N(\mathcal{X}^N)$ where $\mathcal{X}^N$ is the random vector representing the particles. To further clarify, $\mu_{\mathcal{X}^N}$ is a random variable with values in $\mathcal{P}(E)$, $\quad \mu_{\mathcal{X}^N} : \Omega \to \mathcal{P}(E)$. The advantage of such representation is that it is possible to work with only one element which belongs to the fixed space $\mathcal{P}(E)$ (it does not grow along with N) in contrast to $\mathcal{P}(E^N)$ or to the N marginal distribution.

Note that, since $\mu_{\mathcal{X}^N}$ is a random variable, there exists its law $F^N \in \mathcal{P}(\mathcal{P}(E))$. This fact makes sense because if E is Polish, $\mathcal{P}(E)$ is polish itself ([13], theorem 6.15). Distances that make it Polish are the Wasserstein distances:

Let (E, d_E) be a Polish space. For $p \geq 1$, a measure μ in $\mathcal{P}(E)$ admits a finite p-th moment when there exists $x_0 \in E$ such that:

$$\mathbb{E}_{X\sim\mu}[d_E(X,x_0)^p] = \int_E d_E(x,x_0)^p \mu(\mathrm{d}x) < +\infty$$

. The space of probability measures with finite p-th moment is denoted by $\mathcal{P}_P(E)$. In this space the Wasserstein distances are:

Def

For $p \geq 1$ the Wasserstein-p distance between the probability measures μ and ν is defined by:

$$\begin{aligned} W_{d_E,p}(\mu,\nu) &= \inf_{\pi\in\prod(\mu,\nu)} \left(\int_{E\times E} d_E(x,y)^p \pi(\mathrm{d}x,\mathrm{d}y) \right)^{\frac{1}{p}} \\ &= \mathop{\mathbb{E}}_{\substack{X\sim\mu \\ Y\sim\nu}} [d_E(X,Y)^p]^{\frac{1}{p}} \end{aligned}$$

where $\prod(\mu,\nu)$ is the set of all couplings of μ and ν, that is to say, the set of probability measures on $E \times E$ with first and second marginals respectively equal to μ and ν. We can generalize to the case of $P(E)$ by replacing the distance on E with a bounded one inducing the same topology (e.g. $\overline{d} = \frac{d}{1+d}$) in order to have finite p-th moment for every $\mu \in P(E)$ and every p

To clarify again what F^N represents, I introduce the definition of push-forward measure, which can be found in [7]:

Def

Whenever $g : X \to Y$ is a map from a measure space X, equipped with a measure ν, to an arbitrary space Y, we denote by $g_{\#}\nu$ the push-forward measure of ν by g. Explicitly

$$g_{\#}\nu(B) = \nu(g^{-1}(B))$$

where $g^{-1}(B) = \{x \in X \mid g(x) \in B\}$

example Let Y be a r.v. on $(\Omega, \mathcal{F}, \mathcal{P})$ with real values. $Y_{\#}\mathcal{P}(B) = P(X^{-1}(B)) \quad B \in \mathcal{B}$ and it coincides with the distribution μ_Y on $\mathbb{R}$.

Now we can say that, from the point of view of measure theory, the representation (2.1) means that the law f^n is replaced by its push-forward by the map (2.2) by:

$$F^N := (\mu_N)_{\#} f^N \in \mathcal{P}(\mathcal{P}(E))$$

From a probabilistic point of view F^N is the law of the random empirical measure $\mu_{\mathcal{X}^N}$ where $\mathcal{X}^N \sim f^N$

Infinite particle systems

Gases have a number of molecules proportional to the Avogadro number $\sim 6 \times 10^{23}$, therefore the next step is the description of the limit dynamics when $N \to \infty$. In the previous section, finite exchangeable particle systems are described either by the marginal distributions or by the empirical measure. In this section the framework to take the limit $N \to \infty$ is presented. An infinite set of exchangeable random variables $(X^1, X^2, \dots)$ is described by one of the two following objects:

1. The infinite hierarchy of marginal distributions $f^k \in \mathcal{P}_{sym}(E^k), \quad k \in \mathbb{N}$ such that:
$$f^k = Law(X^1, X^2, \dots).$$
They satisfy the relation: for every $1 \leq j \leq k$
$$\forall \varphi_j \in C_b(E^j), \qquad \langle f^k, \varphi_j \otimes 1^{\otimes k-j} \rangle = \langle f^j, \varphi_j \rangle$$
that is to say, the $j-$particle marginal of f^k is f^j.

2. The infinite sequence of random empirical measure of size N, $N \in \mathbb{N}$:
$$\mu_{\mathcal{X}^N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i}$$

The choice of these two probability measures has been made because when N grows, they remain in the same space, respectively $\mathcal{P}(E^k)$ and $\mathcal{P}(E)$, in contrast to the $N-$ particle distribution. In fact, it becomes more difficult to work with f^N when $N \to \infty$. Moreover, physically speaking, f^N is almost useless, because the observable functions typically depend on interactions between few particles or global means.

example: [**Temperature**] Suppose we have a gas in thermal equilibrium. From the Maxwell-Boltzmann distribution (1.1) it is possible to derive the temperature T of a gas as follows:
We set $a = \frac{m}{2\pi kT}$ rewrite the distribution as a distribution on $\mathbb{R}^3$ of the vector velocities :

$$a^{\frac{3}{2}} \int_{\mathbb{R}} 4\pi v^2 e^{-a\pi v^2} \mathrm{d}v = a^{\frac{3}{2}} \int_0^{\pi} \mathrm{d}\phi \int_0^{2\pi} sin(\theta) \mathrm{d}\theta \int_{\mathbb{R}^+} v^2 e^{-a\pi v^2} \mathrm{d}v$$

$$= a^{\frac{3}{2}} \int_{\mathbb{R}^3} e^{-a\pi(v_x^2+v_y^2+v_z^2)} \mathrm{d}v_x \mathrm{d}v_y \mathrm{d}v_z$$

Where I used the change of coordinates from spherical to cartesian. Hence, we obtained the density distribution of the velocities as vectors:

$$G_v = (\frac{m}{2\pi kT})^{\frac{3}{2}} e^{-\frac{m(v_x^2+v_y^2+v_z^2)}{2kT}}$$

We notice that it represent the density of a multi-normal distribution with mean $\nu = 0$ and covariance matrix $C = \frac{kT}{m} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Thus, (thanks to Prop 2.2 Appendix A) in every direction the probability distribution of the velocities is: $\mu_{v_x} = N_{0,\frac{kT}{m}}$, which implies

$$\mathbb{E}[v_x^2] = \frac{kT}{m}$$

This means that we only need the limit $N \to \infty$ of $\mu_{\overline{X}^N} \in \mathcal{P}(E)$ (supposed $E = \mathbb{R}^6$) to work with the observable T.

example: [**Pressure**] For an ideal gas (that is to say without interactions between particles) in thermodynamical equilibrium, the formula $P = \frac{Nm\mathbb{E}[v_x^2]}{V} = \frac{NkT}{V}$ applies.

However, in the general case, where there are interactions between particles it is necessary to add the contribution of the latter:

$$P = \frac{NkT}{V} + \frac{1}{3V}\mathbb{E}\left[\sum_{i\leq j}\langle \vec{r_{ij}}, \vec{F_{ij}}\rangle\right]$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ and $\vec{F}_{ij}(||\vec{r}_{ij}||)$ is the force between two particles, function only of the distance. Although, even in this case, we only need the $2-$particle distribution f^2 to compute the expected value:

$$\int_{\mathbb{R}^6\times\mathbb{R}^6} \langle \vec{r}_{12}, \vec{F}_{12}(||\vec{r}_{ij}||)\rangle f^2(\vec{r}_1, \vec{v}_1, \vec{r}_2, \vec{v}_2)\mathrm{d}\vec{r}_1\mathrm{d}\vec{v}_1\mathrm{d}\vec{r}_2\mathrm{d}\vec{v}_2$$

2.3 Kac's Chaos

Kac's definition of chaos

Let $f \in \mathcal{P}(E)$. A sequence $(f^N)_{N\geq 1}$ of symmetric probability measures (i.e. with the property of exchangeability) on E^N is said to be $f-$*chaotic* when for any $k \in \mathbb{N}$ and any function $\varphi_k \in C_b(E^k)$,

$$\lim_{N\to\infty}\langle f^N, \varphi_k \otimes 1^{\otimes N-k}\rangle = \lim_{N\to\infty}\langle f^{k,N}, \varphi_k\rangle = \langle f^{\otimes k}, \varphi_k\rangle \tag{2.3}$$

This means that $f^{k,N}$ converges weakly towards $f^{\otimes k}$, that is the joint probability measure that characterizes independence for a vector of k random variables i.i.d. In fact, Interpreting f^N as the law of an exchangeable system of N particles, the property (2.3) means that for any group of k particles, the particles become statistically independent as N tends to $+\infty$, hence the terminology of chaos.

Lemma [characterization of Kac's choas]

each of the following assertions is equivalent to Kac's chaos:

$i)$ There exist $k \geq 2$ such that $f^{k,N}$ converges weakly towards $f^{\otimes k}$

$ii)$ The random measure $\mu_{\mathcal{X}^N}$ with $X^N \sim f^N$ converges in Law toward the deterministic measure f

Remark (chaos as a law of large numbers)

In the theorem, the expression "$\mu_{\mathcal{X}^N}$ with $\mathcal{X}^N \sim f^N$ converges in Law toward the deterministic measure f" means that $F^N := Law(\mu_{\mathcal{X}^N}) \to \delta_f$ and

$$\forall\ \psi \in C_b(\mathcal{P}(E)) \qquad \lim_{N\to\infty} \mathbb{E}[\psi(\mu_{\mathcal{X}^N})] = \mathbb{E}[\psi(f)] = \psi(f)$$

For $\varphi \in C_b(E)$ the function $\psi : \mathcal{P}(E) \to \mathbb{R},\ \mu \mapsto \mathbb{E}_{X\sim\mu}[\varphi(X)] = \langle \mu, \varphi\rangle$ is bounded because of φ and continuous with respect to the weak-* topology on $\mathcal{P}(E)$. $\mathcal{P}(E)$ is endowed with the weak-*topology because it can be seen as a subset of the dual space of $C_b(E)$ throughout the identification $\nu \mapsto \int_E \phi \mathrm{d}\nu$ (ψ is continuous by the definition of the weak-* topology). We can also notice that $\mathbb{E}_{X\sim\mu_{\mathcal{X}^N}}[\varphi(X)] = \frac{\varphi(X^{1,N})+\cdots+\varphi(X^{N,N})}{N}$, and that for the property i of the lemma $\langle f^{1,N}, \varphi\rangle \overset{N\to\infty}{\to} \langle f, \varphi\rangle$. Thus, the convergence in law of $\mu_{\mathcal{X}^N}$ to f implies:

$$\frac{\varphi(X^{1,N}) + \cdots + \varphi(X^{N,N})}{N} - \mathbb{E}[\varphi(X^{1,N})] = \langle \mu_{\mathcal{X}^N}, \varphi\rangle - \langle f^{1,N}, \varphi\rangle \overset{N\to\infty}{\to} 0 \quad (2.4)$$

Where the convergence is the convergence in law. Let us clarify this last statement: the last limit is not a limit of a determined sequence in $\mathbb{R}$ because $\psi(\mu_{\mathcal{X}^N})$ is a random variable.

X_N, Y **real-valued random variable**	μ_N, ν **random measures**
$X_N \overset{\mathcal{L}}{\to} Y$	$\mu_N \overset{\mathcal{L}}{\to} \nu$
$\underset{\underset{\mathbb{R}}{\in}}{\mathbb{E}[\varphi(X_N)]} \overset{N\to\infty}{\to} \underset{\underset{\mathbb{R}}{\in}}{\mathbb{E}[\varphi(Y)]}$	$\underset{\underset{\mathbb{R}}{\in}}{\mathbb{E}[\psi(\mu_N)]} \overset{N\to\infty}{\to} \underset{\underset{\mathbb{R}}{\in}}{\mathbb{E}[\psi(\nu)]}$
limit in $\mathbb{R}$ -Absolute convergence	$\underset{\underset{m\mathcal{F}}{\in}}{\psi(\mu_N)} \overset{N\to\infty}{\to} \underset{\underset{m\mathcal{F}}{\in}}{\psi(\nu)}$
	convergence for random variables, -weak convergence

The relation (2.4) is reminiscent of the weak law of large numbers (A). It is not the strong one because the $X^{i,N}$ are only exchangeable (not i.i.d.). Moreover, there is a slight difference since f is the law of a typical particle in the limit system, and not the law of $X^{1,N}$ as in the i.i.d. case, because $X^{1,N}$ still depends on N (i.e. on the other particles).

The physical interpretation is that in a system with many particles (e.g. a gas) it is possible to think of every particle as independent from the others, at least statistically. The quantity: $\frac{\varphi(X^{1,N})+\cdots+\varphi(X^{N,N})}{N}$ is the empirical average of an observable φ (e.g., velocity, energy, position). The relation (2.4) says that for $N \gg 1$ the average observed on the particles coincides with the theoretical mean, that is the expected value on a limit distribution f, which describes the behavior of a typical particle.

Conclusions

In this thesis I have revisited the fundamental problem posed by Boltzmann: how can microscopic, time-reversible laws give rise to macroscopic irreversibility? Boltzmann's statistical explanation of the second law, together with the crucial assumption of molecular chaos, proved to be both insightful and fertile. These ideas did not remain merely philosophical: they received confirmation from different directions, experimental and mathematical.

Experimentally, Einstein's analysis of Brownian motion together with Perrin's measurements provided a direct and quantitative validation of the atomic hypothesis: microscopic fluctuations produce observable macroscopic effects and allow the determination of physically meaningful constants (e.g. Avogadro's number). Mathematically, the derivation of macroscopic transport equations from particle dynamics in suitable asymptotic regimes most notably the Boltzmann-Grad limit, and the partial rigorous results such as Lanford's theorem has shown how the Boltzmann equation can emerge from microscopic interactions in an appropriate limit.

Moreover, the probabilistic formalization introduced by Kac clarified the meaning of chaos and the propagation of statistical independence, bridging Boltzmann's physical intuition and a precise mathematical framework. Altogether, these developments support Boltzmann's core message: irreversibility is a statistically overwhelmingly probable phenomenon, while exact recurrences are possible in principle but practically irrelevant for macroscopic systems.

This story exemplifies the productive interplay between physical intuition, experiment, and mathematical rigor: heuristic ideas motivate precise formulations, experiment tests them, and rigorous analysis refines our theoretical understanding.

Appendix A

Probability reminders

I am going to recall some classical probability results that can be found in [12]

Let $Y = (Y_1, \dots, Y_d)$ be a $d-$dimensional real valued random variable. μ_Y is the joint probability distribution of $(Y_1, \dots, Y_d)$ and μ_{Y_i} is the marginal distribution of $i-$th component

Prop 2.1

Let Y be a $d-$dimensional r.v.:

1. $\mu_{Y_1}(H) = \mu_Y(H \times \mathbb{R}^{d-1}) \quad \forall\ H \in \mathcal{B}$
2. If $Y \in AC$ with joint probability density γ_Y, then $Y_1 \in AC$ and $\gamma_{Y_1}(z) = \int_{\mathbb{R}^{d-1}} \gamma_Y(z, y_2, \dots, y_d) \mathrm{d}y_2 \dots \mathrm{d}y_d \ z \in \mathbb{R}$

Def

Let Z, Y be two r.v. on $(\Omega, \mathcal{F}, \mathcal{P})$. $Z, Y : \Omega \rightarrow (E, \mathcal{G})$ where $(E, \mathcal{G})$ is a measurable space, Z and Y are said *stochastically independent* on $\mathcal{P}$ if $P((Y \in H) \cap (Z \in K)) = P(Y \in H)P(Z \in K)$ for every $H, K \in \mathcal{G}$

Usually $\mathcal{G}$ is the Borel $\sigma-$Algebra $(\mathcal{B})$, if E has a topology.

Theorem

They are equivalent:

1. Y and Z are independent
2. $\mu_{(Z,Y)} = \mu_Z \otimes \mu_Y$ where $\otimes$ indicates the product measure

Also, The general formula of the density of a $d-$dimensional multi-normal is: $\Gamma(X) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} exp\left(-\frac{1}{2}\langle C^{-1}(X-\nu), X-\nu\rangle\right)$ where $C > 0$ and symmetric and $X \in \mathbb{R}^d$.
I also recall the following proposition:

Prop 2.2

Let $X \sim N_{\nu,C} \quad \nu \in \mathbb{R}^d, C \in \mathbb{R}^{d\times d} \Rightarrow \alpha X + \beta \sim N_{\alpha\nu+\beta,\alpha C\alpha^*}$ where $\alpha \in \mathbb{R}^{N\times d}, \quad \beta \in \mathbb{R}^N$

Weak law of large numbers

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of real valued r.v. in $L^1(\Omega, P)$, with expected value $\upsilon = \mathbb{E}[X_1]$. Then $M_n = \frac{X_1+\cdots+X_N}{N}$ converges weakly to the constant r.v. υ:

$$M_N \xrightarrow{\mathcal{L}} \upsilon$$

The following are results about some properties of the expected value:

Prop 2.3

Let $X, Y \in L^1(\Omega, \mathcal{F}, \mathcal{P})$ and $\mathcal{G}$ be a $\sigma - Algebra$, then the following properties hold true:

1. $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$
2. $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$
3. if $Y \in m\mathcal{G}$ is such that $YX,\ YE[X|G] \in L^1(\Omega, \mathcal{P})$, then $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$

Bibliography

[1] Massimo Falcioni and Angelo Vulpiani. Ludwig Boltzmann: a tribute on his 170th birthday. *Lett. Mat.*, 2(4):171−183, 2015.

[2] Boltzmann, L. (1872). *Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen.* Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Classe, 66, 275−370.

[3] Benettin G., *Introduzione alla teoria ergodica, notes*. https://www.math.unipd.it/ benettin/postscript-pdf/teorerg.pdf

[4] G. E. Uhlenbeck and G. W. Ford. *Lectures in Statistical Mechanics, volume Vol. I of Lectures in Applied Mathematics (Proceedings of the Summer Seminar, Boulder, Colorado, 1960).* American Mathematical Society, Providence, RI, 1963. With an appendix on quantum statistics of interacting particles by E. W. Montroll.

[5] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *Kinet. Relat. Models*, 15(6):895−1015, 2022.

[6] S. Mischler. *Kac's chaos and Kac's program* In Séminaire Laurent Schwartz Équations aux dérivées partielles et applications. Année 2012−2013, Sémin. Équ. Dériv. Partielles, pages Exp. No. XXII, 17. École Polytech., Palaiseau, 2014.

[7] Cédric Villani. *Topics in Optimal Transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003.

[8] (Web resource) *Equipartition Theorem*, wikipedia, 08/2025, `https://en.wikipedia.org/wiki/Equipartition_theorem`.

[9] Rapaport, Dennis, *The Art of Molecular Dynamics Simulation*,online PDF, `https://www.r-5.org/files/books/computers/algo-list/fuzzy-logic/modeling/Dennis_Rapaport-The_Art_of_Molecular_Dynamics_Simulation-EN.pdf`, pag. 19.

[10] (Web resource) *Striling Approximation*, wikipedia, 08/2025,`https://it.wikipedia.org/wiki/Approssimazione_di_Stirling`.

[11] Benettin G., *A smooth introduction to Ergodic Theory*, notes. https://www.math.unipd.it/ benettin/links-ergodic/ergodic.pdf

[12] Andrea Pascucci. *Probability theory. Vol. I. Random variables and distributions*, volume 165 of *Unitext*. Springer, Cham, [2024] © 2024. Translated from the 2020 Italian, La Matematica per il 3+2.

[13] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften* [Fundamental Principles of Mathematical Sciences]. SpringerVerlag, Berlin, 2009. Old and new.

[14] (Web resource) *Thermodynamics*, wikipedia, 07/2025, `https://en.wikipedia.org/wiki/Thermodynamics`.

[15] (Web resource), Treccani, 07/2025 *L'ottocento: fisica. La termodinamica (it)*, `https://www.treccani.it/enciclopedia/l-ottocento-fisica-la-termodinamica_(Storia-della-Scienza)`.

[16] Kerson Huang. *Statistical mechanics.* John Wiley & Sons, Inc., New York, second edition, 1987.

[17] Mark Kac. Random walk and the theory of Brownian motion. *Amer. Math. Monthly*, 54:369−391, 1947.

[18] Mark Kac. *Enigmas of chance.* Alfred P. Sloan Foundation. Harper & Row, Publishers, New York, 1985. An autobiography.

.