慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	Chapter 4 : Boltzmann's equilibrium statistical mechanics
Sub Title	
Author	石川, 史郎(Ishikawa, Shiro)
Publisher	Keio University Press Inc.
Publication year	2006
Jtitle	Mathematical Foundations of Measurement Theory (測定理論の数学的基礎). (2006.) ,p.75- 90
JaLC DOI	
Abstract	
Notes	
Genre	Book
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO52003001-00000000-0075

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

Chapter 4

Boltzmann's equilibrium statistical mechanics

As mentioned in Chapters 2 and 3, we see that (pure) measurement theory (= PMT) is formulated as follows:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} . \qquad \underset{(=(1.4))}{(4.1)}$$

The purpose of this chapter[1] is to understand Boltzmann's equilibrium statistical mechanics [2] (i.e., "the principle of equal a priori probability" and "the ergodic hypothesis") as one of applications of PMT. We believe that our approach completely justifies the the thermodynamical weight method (i.e., the Gibbs method, *cf.* [26])[3].

4.1 Introduction

In spite that equilibrium statistical mechanics is generally believed to be based on Newtonian mechanics, the term "probability" frequently appears in equilibrium statistical mechanics. Therefore, if we want to understand equilibrium statistical mechanics in the framework of Newtonian mechanics, a certain rule concerning "probability" should be added. That is, we hope to understand equilibrium statistical mechanics such as:

[1]It may be recommended that this chapter is skipped if readers want to study statistics in the framework of PMT firstly (*cf.* Chapters 5 and 6).

[2]In this chapter readers are not required to have much knowledge of statistical mechanics.

[3]In this book, we think that statistical mechanics should be understood as one of applications of measurement theory and not theoretical physics, (*cf.* Table (1.7)). Thus, it should be noted that no serious test has been conducted in statistical mechanics. What we know is nothing but the fact that statistical mechanics is quite useful (*cf.* Table (1.8)). Or, statistical mechanics is "almost empirically true" to such a degree that statistical mechanics is assured to be useful in usual situations. Cf. the (I_9) in §1.2.

$$\underset{}{\text{“equilibrium statistical mechanics”}} = \underset{\text{[Axiom 2 (3.26)]}}{\text{“Newton equation”}} + \underset{\text{[Axiom 1 (2.37)]}}{\text{“probabilistic rule”}} \tag{4.2}$$

in PMT.

First we must answer the following question:

(Q_1) What is the "probabilistic rule" in (4.2)?

Recall Example 2.16 (the urn problem), which is the most fundamental in the classical measurement. Thus in order to understand "probabilistic rule (=Axiom 1) in (4.2)", it suffices to note the following simplest example:

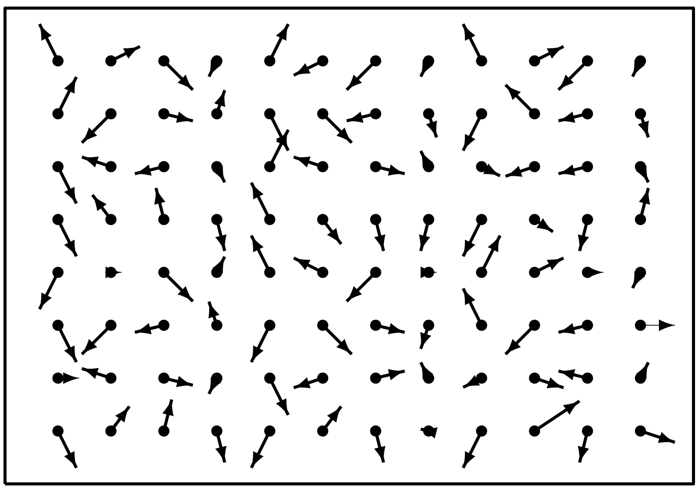
(A_1) *"Consider a box containing* 7×10^{23} *white balls and* 3×10^{23} *black balls, and choose a ball at random from the box. Then the probability that the ball is white is given as 0.7."*

Even without the knowledge of measurement theory (in Chapters 2 and 3), every reader surely agrees that the probability appearing in urn (i.e., box) problems is most typical in statistics.

Next we must refer to "Newtonian mechanics" in (4.2). Namely we must solve the following question.

(Q_2) What kinds of conditions are imposed on the Newton equation in (4.2)?

In equilibrium statistical mechanics, about 10^{24} ($\approx 6.02 \times 10^{23}$: "Avogadro constant") particles, of course, move hard in a box such as the following figure:



(4.3)

However it seems to be natural to think as follows:

(A_2^1) *All particles are even, or on a level.*

(A_2^2) *The motions of particles are (almost) independent of each other. In other words, the information about a subsystem composed of some particles is invalid for the inference of the state of another subsystem.*

This is our answer to the question (Q_2). In §4.2, the (A_2^1) and (A_2^2) will be represented in terms of PMT. Also, the (A_1) will be discussed in §4.3.

Summing up, we think that equilibrium statistical mechanics is formulated as follows:

$$\text{"equilibrium statistical mechanics"} = \underbrace{\underset{\text{(the probability such as in }(A_1))}{\text{"probabilistic rule"}} + \underset{\text{(the conditions }(A_2^1)\text{ and }(A_2^2))}{\text{"Newton equation"}}}_{\text{(+ "staying time interpretation")}} \tag{4.4}$$

in PMT. Or, equivalently,

- An equilibrium statistical system can be regarded as an urn containing about 10^{24} particles. Also, the motions of particles are dominated by the Newtonian equation with the conditions (A_2^1) and (A_2^2). Also, the "staying time interpretation" implies the common sense such as it is almost impossible to find a rare event.

And moreover, two conventional principles (i.e., "the principle of equal a priori probability" and "the ergodic hypothesis") will be completely clarified in our proposal (4.4).

The first attempt to understand equilibrium statistical mechanics in the framework of PMT was executed in [45]. The content in [45] will be slightly modified and improved in this chapter.

Note, for completeness, that our purpose is to understand equilibrium statistical mechanics as one of applications of PMT and not to derive equilibrium statistical mechanics from Newtonian mechanics (*cf.* [75]). That is, we are in theoretical informatics and not in theoretical physics.[4]

[4] We have no experimental evidence that the ergodic approach to statistical mechanics is proper. However, in theoretical informatics, it suffices to find a reason that many people do not doubt.

4.2 Dynamical aspects of equilibrium statistical mechanics

In this section we shall devote ourselves to the mathematical description of the answers (A_2^1) and (A_2^2) mentioned in Section 4.1. Readers should note that all arguments in this section are within Newtonian mechanics. Namely, it should be noted that it is prohibited to use the term "probability" in this section. For example, Lemma 4.9 ("the law of large numbers" in §4.5 Appendix) is not only most important in Kolmogorov's probability theory but also in this section (i.e., the derivation of the ergodic hypothesis (= Theorem 4.6)). Therefore, readers will see that Lemma 4.9 is used independently of the concept of "probability". This is the reason that the term "normalized measure" (and not "probability measure") is used in Lemma 4.9.

Now let us begin with the well-known ergodic theorem (*cf.* [57, 83]). In Newtonian mechanics, any state of a system composed of $N(\approx 10^{24})$ particles is represented by a point (q,p) $\big(\equiv (q_{1n}, q_{2n}, q_{3n}, \, p_{1n}, p_{2n}, \, p_{3n})_{n=1}^N \big)$ in a phase (or state) space $\mathbf{R}^{6N}$ (*cf.* the formula (2.8)). Let $\mathcal{H}: \mathbf{R}^{6N} \to \mathbf{R}$ be a Hamiltonian, i.e., a positive continuous function on $\mathbf{R}^{6N}$. Define $V(E)$, $E \geq 0$, by "the volume of the set $\{(q,p) \in \mathbf{R}^{6N} \mid \mathcal{H}(q,p) \leq E\}$", and define the measure $\nu_{_E}$ on the energy surface $\mathcal{S}_{_E}$ $(\equiv \{(q,p) \in \mathbf{R}^{6N} \mid \mathcal{H}(q,p) = E\})$ such that

$$\nu_{_E}(B) = \int_B |\nabla \mathcal{H}(q,p)|^{-1} dm_{6N-1} \quad (\forall B \in \mathcal{B}_{\mathcal{S}_{_E}}, \text{ the Borel field of } \mathcal{S}_{_E})^5 \tag{4.5}$$

where dm_{6N-1} is the usual surface measure on $\mathcal{S}_{_E}$. Note that $\nu_{_E}(\mathcal{S}_{_E}) = \frac{dV(E)}{dE}$ holds. Let $\{\psi_t^E\}_{-\infty<t<\infty}$ be the flow on the energy surface $\mathcal{S}_{_E}$ induced by the Newton equation with the Hamiltonian $\mathcal{H}$. Liouville's theorem (*cf.* [11]) says that the measure $\nu_{_E}$ is invariant concerning the flow $\{\psi_t^E\}_{-\infty<t<\infty}$. Defining the normalized measure $\overline{\nu}_{_E}$ such that $\overline{\nu}_{_E} = \frac{\nu_E}{\nu_E(\mathcal{S}_E)}$, we have the normalized measure space $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$.

In order that equilibrium statistical mechanics must hold, we first assume that the Hamiltonian $\mathcal{H}$ satisfies the following *ergodic hypothesis* (EH):

(EH) The flow $\{\psi_t^E\}_{-\infty<t<\infty}$ on the $\mathcal{S}_{_E}$ is *ergodic*. That is, there uniquely exists an normalized invariant measure $\overline{\nu}_{_E}$ on $\mathcal{S}_{_E}$ such that $\overline{\nu}_{_E}(B) = \overline{\nu}_{_E}(\psi_t(B))$ $(-\infty < \forall t <$

[5] Or usually, $\nu_{_E}(B) = \frac{1}{h^{3N}N!}\int_B |\nabla \mathcal{H}(q,p)|^{-1} dm_{6N-1}$, where h is the Plank constant. In this book, for simplicity, the constant $\frac{1}{h^{3N}N!}$ will be omitted.

∞, $\forall B \in \mathcal{B}_{\mathcal{S}_E}$).

The ergodic theorem (*cf.* [11, 57]) says that the normalized measure $\overline{\nu}_E$ represents the *normalized averaging staying time,* i.e., it holds that

$$\overline{\nu}_E(B) = \lim_{K\to\infty} \frac{\sharp[\{k \mid \psi_{\epsilon k}\omega \in B, k = 1,2,...,K\}]}{K} \quad (\forall \omega \in \mathcal{S}_E, \forall \epsilon > 0).$$

or generally,

$$\underbrace{\int_\Omega f(\omega)\overline{\nu}_E(d\omega)}_{\text{(space average)}} = \underbrace{\lim_{T\to\infty}\frac{1}{T}\int_0^T f(\psi_t(\omega_0))dt}_{\text{(time average)}} \quad (\forall f \in C(\Omega), \quad \forall \omega_0 \in \Omega), \tag{4.6}$$

which is equivalent to the (EH). Thus the normalized measure space $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$ is called *the normalized averaging staying time space* (*cf.* Remark 4.1 later).

We assert that

(STI) [Staying time interpretation of statistical mechanics]. *Let $\mathcal{N} \in \mathcal{B}_{\mathcal{S}_E}$ such that the normalized averaging staying time $\overline{\nu}_E(\mathcal{N})$ is quite small (i.e., $\overline{\nu}_E(\mathcal{N}) \ll 1$). Then it is almost impossible (or precisely, quite rare) to see that the state $(q(t), p(t))$ belongs to the $\mathcal{N}$.*

We think that this (STI) is a common sense rather than a principle. The concept of "time" (or precisely "non-relativistic time") is within Newtonian mechanics, and therefore the statement (STI) (or "staying time") can be understood within Newtonian mechanics.

Remark 4.1. [The probabilistic interpretation of $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$]. The probabilistic interpretation is as follows:

(PI) [Probabilistic interpretation of statistical mechanics]. *The normalized averaging staying time space $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$ is regarded as Kolmogorov's probability space.*

That is, the probabilistic interpretation, which is usually called *"the principle of equal a priori probability"*, means that the probability that the state of the system belongs to $\Xi(\in \mathcal{B}_{\mathcal{S}_E})$ is given by $\overline{\nu}_E(\Xi)$. If the probabilistic interpretation (PI) is assumed, the (STI) obviously holds. However, the concept of "normalized staying time" is clearly different from that of "probability". Note that:

- the former (i.e., "the staying time interpretation") is within Newtonian mechanics, but the latter (i.e., "the probabilistic interpretation") is not so.

Thus, in this chapter we choose a common sense (i.e., "the staying time interpretation") rather than a principle (i.e., "the probabilistic interpretation").[6] This is the reason that the $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$ is not called the probability space in this chapter. Again note that all arguments in this section are within Newtonian mechanics. In this chapter the (STI) will be used instead of the (PI).

■

We introduce the following notation:

Notation 4.2. [In the sense of (STI)]. *Let* $\mathbf{P}(q,p)$ *be a proposition concerning a state* (q,p) $(\in \mathcal{S}_{_E})$ *such that* $\mathbf{P}(q(t),p(t))$ *is true for every* $t \in \mathcal{S}_{_E} \setminus \mathcal{N}$ $(\equiv \{\omega \mid \omega \in \mathcal{S}_{_E}, \omega \notin \mathcal{N}\})$. *Assume that the normalized averaging staying time* $\overline{\nu}_{_E}(\mathcal{N})$ *is quite small (i.e.,* $\overline{\nu}_{_E}(\mathcal{N}) \ll 1$*). Then we write it as*

$$\begin{aligned}&\mathbf{P}(q(t),p(t)) \textit{ is true} \quad \textit{(almost every } t \textit{ in the sense of (STI))}, \qquad (4.7)\\ &\Big(\textit{Or, } \mathbf{P}(q(t),p(t)) \textit{ is almost always true} \Big).\end{aligned}$$

Also, when the probabilistic interpretation (cf. Remark 4.1) is added to the $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$, *we may write it as*

$$\mathbf{P}(q(t),p(t)) \textit{ is true} \quad \textit{(almost every } t \textit{ in the sense of (PR))}.^{7} \qquad (4.8)$$

■

As seen in Remark 4.1, it holds that (4.8)⇒(4.7). Throughout this chapter we, of course, focus on the (4.7) and not (4.8).

Let $\epsilon > 0$, $f_1, f_2, ..., f_K \in C_0(\mathbf{R}^6)$. Define the 0-neighborhood U in $\mathcal{M}(\mathbf{R}^6)$ (in the sense of weak* topology of $\mathcal{M}(\mathbf{R}^6)$) such that:

$$U(= U^{\epsilon}_{f_1, \cdots f_K}) = \{\rho \in \mathcal{M}(\mathbf{R}^6)(= C_0(\mathbf{R}^6)^*) \ : \ |_{\mathcal{M}(\mathbf{R}^6)}\langle \rho, f_k \rangle_{C_0(\mathbf{R}^6)}| < \epsilon, k = 1,2,...,K\}. \qquad (4.9)$$

---

[6] What is the most important is to recognize that statistical mechanics belongs to the category of theoretical informatics and not that of theoretical physics. (*cf.* Table (1.7)). Thus, the present situation is the same as the following situation. Two ready-made suits (A) and (B) are on sale. The (A) is somewhat big, and the (B) is somewhat small. Which do you choose, (A) or (B)? Cf. (I_{15}) in §1.3. We must choose one from "the staying time interpretation" and "the probabilistic interpretation". In theoretical informatics, it can not be decided by experimental test. What we can say is we believe that "the staying time interpretation" will win more popularity than "the probabilistic interpretation".

[7] Note that this notation is different from that of Kolmogorov's probability theory, in which we use the phrase "almost every t in the sense of (PR)" when $\bar{\nu}(\mathcal{N}) = 0$.

Put $D_N = \{1, 2, ..., N(\approx 10^{24})\}$. For each k ($\in D_N \equiv \{1, 2, ..., N(\approx 10^{24})\}$), define the map $X_k : \mathcal{S}_E(\subset \mathbf{R}^{6N}) \to \mathbf{R}^6$ such that

$$X_k((q_{1n}, q_{2n}, q_{3n}, p_{1n}, p_{2n}, p_{3n})_{n=1}^N) = (q_{1k}, q_{2k}, q_{3k}, p_{1k}, p_{2k}, p_{3k}) \tag{4.10}$$

for all $(q, p) = (q_{1n}, q_{2n}, q_{3n}, p_{1n}, p_{2n}, p_{3n})_{n=1}^N$ in $\mathcal{S}_E(\subset \mathbf{R}^{6N})$. For any subset D ($\subseteq D_N \equiv \{1, 2, ..., N(\approx 10^{24})\}$), define the map $R_D^{(\cdot)} : \mathcal{S}_E(\subset \mathbf{R}^{6N}) \to \mathcal{M}_{+1}^m(\mathbf{R}^6)$ ($\equiv \{\rho \in \mathcal{M}(\mathbf{R}^6) : \rho \geq 0,\ \rho(\mathbf{R}^6) = 1\}$) such that

$$R_D^{(q,p)} = \frac{1}{\sharp[D]} \sum_{k \in D} \delta_{X_k(q,p)} \qquad (\forall (q, p) \in \mathcal{S}_E(\subset \mathbf{R}^{6N})), \tag{4.11}$$

where $\sharp[D]$ is the number of the elements of D and δ_x is a point measure at x ($\in \mathbf{R}^6$).

Let U be a 0-neighborhood in $\mathcal{M}(\mathbf{R}^6)$ such as defined in (4.9). For any (p, q) ($\in \mathcal{S}_E$), put

$$H_U(p, q) = k_B \log \left[\nu_E(\{(p', q') \in S_E \mid R_{D_N}^{(p,q)} - R_{D_N}^{(p',q')} \in U\})\right], \tag{4.12}$$

(k_B is the Boltzmann constant, i.e., $k_B = 1.381 \times 10^{-23}$J/K), which is called the *U-entropy* of a state (p, q).

Let $D_0 \subseteq D_N$. Define $\overline{\nu}_E \circ ((X_k)_{k \in D_0})^{-1}$ ($\in \mathcal{M}_{+1}^m(\mathbf{R}^{6 \times \sharp[D_0]})$) by the image measure concerning the map $(X_k)_{k \in D_0} : \mathbf{R}^{6N} \to \mathbf{R}^{6 \times \sharp[D_0]}$, that is,

$$\overline{\nu}_E \circ ((X_k)_{k \in D_0})^{-1}(\mathop{\times}_{k \in D_0} A_k) = \overline{\nu}_E(\{(p, q) \in \mathcal{S}_E \mid X_k(p, q) \in A_k \ (k \in D_0)\}) \tag{4.13}$$

for any open set A_k ($\subseteq \mathbf{R}^6$) ($k \in D_0$).

In what follows we shall represent the conditions (A_2^1) and (A_2^2) (mentioned in §4.1) in terms of mathematics. Cf. [45].

Definition 4.3. [Thermodynamical condition, equilibrium state][8] *Let D_N be a set $\{1, 2, ..., N(\approx 10^{24})\}$. And let $\mathcal{H}$, E, ν_E, $\overline{\nu}_E$, $X_k : \mathcal{S}_E \to \mathbf{R}^6$ be as in the above. A Hamiltonian $\mathcal{H}$ on $\mathbf{R}^{6N}$ ($N \approx 10^{24}$) is said to be thermodynamical (concerning energy E) if the following condition (T) is satisfied:*

(T) $\{X_k : \mathcal{S}_E \to \mathbf{R}^6\}_{k=1}^N$ *is an almost independent sequence with the identical distribution.*

[8] Although this condition may be superficial and not fundamental, we believe, from the measurement theoretical point of view, that equilibrium statistical mechanics should start from this condition. Again note that our purpose is to understand equilibrium statistical mechanics as one of applications of PMT and not to derive equilibrium statistical mechanics from Newtonian mechanics.

In other words, there exists a normalized measure $\rho_{_E}$ *on* $\mathbf{R}^6$ *(i.e.,* $\rho_{_E} \in \mathcal{M}_{+1}^m(\mathbf{R}^6)$*) such that:*

(T^1) $\big[$*identical distribution, cf.* (A_2^1) *in* §4.1$\big]$ *it holds that*

$$\rho_{_E} \approx \overline{\nu}_{_E} \circ X_k^{-1} \qquad (\forall k = 1, 2, ..., N(\approx 10^{24})), \tag{4.14}$$

(T^2) $\big[$*independence, cf.* (A_2^2) *in* §4.1$\big]$ *it holds that*

$$\bigotimes_{k \in D_N} \rho_{_E} (\text{: product measure}) \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_N})^{-1}, \tag{4.15}$$

though the condition (T^2) *is too strong to assume it literally, (see Remark 4.4).*

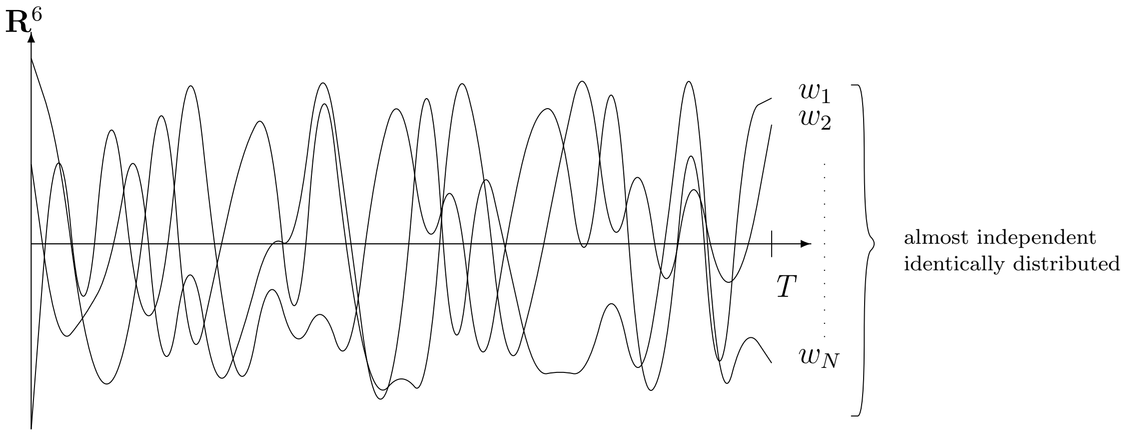
Here, a state (q, p) $(\in \mathcal{S}_E)$ *is called an equilibrium state if* $R_{D_N}^{(q,p)} \approx \rho_E$.[9]

■

Let T be a sufficiently large number. Assume that the closed interval $[0, T]$ has the measure: dt/T (thus, the total measure of $[0, T]$ is equal to 1). For each k $(\in D_N \equiv \{1, 2, ..., N(\approx 10^{24})\})$, define the map $w_k : [0, T] \to \mathbf{R}^6$ such that $w_k(t) = (q_{1k}(t), q_{2k}(t), q_{3k}(t), p_{1k}(t), p_{2k}(t), p_{3k}(t))$ for all t $(\in [0, T])$. Assume that

($\sharp$) $\{w_k \mid k \in D_N\}$ is a set composed of almost independent functions with the identical distribution.

This assumption ($\sharp$) is essentially the same as (T) in Definition 4.3.



[9]In our formulation, we do not assume that the "equilibrium state" is defined by $\overline{\nu}_{_E}$ since $\overline{\nu}_{_E}$ is not assumed to have the probabilistic interpretation (*cf.* Remark 4.1).

Remark 4.4. As mentioned in Definition 4.3, the condition (T^2) is too strong. Thus, it should be understood symbolically and not literally. Therefore, we actually assume some hypotheses, which are weaker than the (T_2). For example we assume the following conditions $(T^2)'$ and $(T^2)''$:

$(T^2)'$ $\big[$*independence*$\big]$ *it holds that*

$$\bigotimes_{k \in D_0} \rho_{_E} \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_0})^{-1}, \tag{4.16}$$

$(\forall D_0 \subset \{1, 2, ..., N(\approx 10^{24})\}$ *such that* $1 \ll \sharp[D_0] \ll N)$.

This is needed for the derivation of the ergodic hypothesis (*cf.* Theorem 4.6 later). Also, we assume that

$(T^2)''$ $\big[$*independence*$\big]$ *it holds that*

$$\Big(\overline{\nu}_{_E} \circ ((X_k)_{k \in D_1})^{-1}\Big) \bigotimes \Big(\overline{\nu}_{_E} \circ ((X_k)_{k \in D_2})^{-1}\Big) \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_1 \cup D_2})^{-1}, \tag{4.17}$$

for any D_1, D_2 $(\subset D)$ *such that* $D_1 \cap D_2 = \emptyset$ *and* $1 \ll \sharp[D_1], \sharp[D_2] \leq N$.

That is because, in equilibrium statistical mechanics, we usually assume that the interaction between the subsystem composed of the particles D_1 and that of the particles D_2 can be neglected.

■

Remark 4.5. (i) If N_0 is arbitrarily large (and thus $N = \infty$) and if the approximation symbol "$\approx$" is interpreted by the equality "$=$", then (4.4) and (4.16) imply that the sequence $\{X_k\}_{k=1}^{\infty}$ on the normalized averaging staying time space $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$ is an independent sequence with the identical distribution $\rho_{_E}$. Thus, Lemma 4.9 (i.e., the law of large numbers) says that

$$\lim_{N_0(=\sharp[D]) \to \infty} R_D^{(q,p)} = \rho_{_E} \quad (\text{ in the sense of the weak}^* \text{ topology of } \mathcal{M}(\mathbf{R}^6)) \tag{4.18}$$

holds for almost every (q, p) in $(\mathcal{S}_{_E}, \mathcal{B}(\mathcal{S}_{_E}), \overline{\nu}_{_E})$. Note that Kolmogorov's probability theory [56] is mathematics, and therefore, it is valid even if the probabilistic interpretation (*cf.* Remark 4.1) is not added to the normalized averaging staying time measure space $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$. For completeness, again note that the terms: "identical distribution" in (T^1) and "independence" in (T^2) are not related to the concept of "probability" (but that of "staying time").

(ii) The reader may doubt if the concepts of "identical distribution" and "independence" are meaningful without the probabilistic interpretation. However, the following example shows that these concepts are not only meaningful on a measure space but also on a topological space. Let $f : \Omega \to \mathbf{R}$ be a continuous function on a topological space Ω. For each $n(= 1, 2, ..., N)$, define the function $f_n : \Omega^N (=$ product topological space$) \to \mathbf{R}$ such that $\Omega^N \ni (\omega_1, \omega_2, ..., \omega_n, ..., \omega_N) \mapsto f(\omega_n) \in \mathbf{R}$. Then we may say that $\{f_n\}_{n=1}^N$ is *"an independent sequence with the identical distribution"*. In fact we often say "The motions of two particles are independent" in Newtonian mechanics (and not in statistical mechanics).

■

By an analogy of the arguments (i.e., the derivation of (4.18)) in the above Remark 4.5(i), we can assert that (4.14) and (4.16) imply that, if $1 \ll N_0(\approx \sharp[D_0]) \ll N(\approx 10^{24})$,

$$R_{D_0}^{(q(t),p(t))} \approx \overline{\nu}_{_E} \circ X_k^{-1} \ (\approx \rho_{_E}) \qquad (\text{ almost every time } t \text{ in the sense of (STI) }) \tag{4.19}$$

holds for any k ($= 1, 2, ..., N(\approx 10^{24})$). Here consider the decomposition $\{D_{(1)}, D_{(2)}, ..., D_{(L)}\}$ of D_N ($\equiv \{1, 2, ..., N(\approx 10^{24})\}$) such that $\sharp[D_{(l)}] \approx N_0$ ($l = 1, 2, ..., L$). Then we see, by (4.19), that

$$R_{D_N}^{(q(t),p(t))} = \frac{\sum_{l=1}^{L}[\sharp[D_{(l)}] \times R_{D_{(l)}}^{(q(t),p(t))}]}{N} \approx \frac{\sum_{l=1}^{L}[\sharp[D_{(l)}] \times \rho_{_E}]}{N} = \overline{\nu}_{_E} \circ X_k^{-1} \ (\ \approx \rho_{_E} \ )$$
$$(\text{ almost every time } t \text{ in the sense of (STI) })$$

holds for any k ($= 1, 2, ..., N(\approx 10^{24})$).

Summing up, we have the following theorem.

Theorem 4.6. (Ergodic hypothesis). *Assume the thermodynamical condition (i.e., (T_1) in Definition 4.3 and $(T^2)'$ in Remark 4.4). Then it holds that*

$$R_{D_N}^{(q(t),p(t))} \approx \overline{\nu}_{_E} \circ X_k^{-1} \ (\ \approx \rho_{_E} \ ) \quad (k = 1, 2, ..., N(\approx 10^{24})) \tag{4.20}$$
$$(\textit{ almost every time } t \textit{ in the sense of (STI) })$$

Thus, the state of the system is almost always equal to the equilibrium state (cf. Definition 4.3). That is, we see:

- $R_{D_N}^{(q(t_1),p(t_1))} \approx R_{D_N}^{(q(t_2),p(t_2))}$ (*almost every time* t_1 *and* t_2 *in the sense of* (*STI*)).

(4.21)

■

This says that

$$\text{"the distribution of } N(\approx 10^{24}) \text{ particles at almost every time } t\text{" (in the sense of (STI))}$$
$$= \text{"normalized averaging staying time of the } k\text{-th particle } (\forall k = 1, 2, ..., N \approx 10^{24})\text{"} \tag{4.22}$$

We believe that this is just what should be represented by the *"ergodic hypothesis"* :[10]

$$\text{"population average of many particles"} = \text{"time average of one particle"}, \tag{4.23}$$

that is, we see that (4.20)=(4.22)=(4.23).

Remark 4.7. [Another formulation of equilibrium statistical mechanics]. For completeness, note that the condition $(T^2)'$ in (4.16) is assumed in order that (4.21) holds. Thus some may assert that it suffices to start from the $\mathcal{S}_E$ (with the measure ν_E which induces (STI)) and the (4.21). This formulation may be called the formulation without the ergodic hypothesis. Also, see the formula (4.29) later.

■

Remark 4.8. (i). If the probabilistic interpretation (i.e., the principle of equal a priori probability) is assumed, in (4.20) we can replace "almost every time t in the sense of (STI)" to "almost every time t in the sense of (PR)". However, if the (STI) is accepted as a common sense, we can do well without this replacement, that is, the replacement does not bring us any merit. Thus we think that the probabilistic interpretation is not needed. Cf. Remark 4.5

(ii). We may still have a question:

- *Why is the thermodynamical condition* (*i.e.,* (T^1) *and* (T^2)) *always satisfied in the usual circumstance of equilibrium statistical mechanics?*

Though we do not know the firm answer,[11] we can easily show, by (4.20), that the thermodynamical condition ((T^1) and (T^2)) explains the following law (i.e., *"the law of increasing*

[10] In this book, the term "ergodic hypothesis" has two meanings. One is used in the sense of the formula (4.6). And the other is used in the sense of the formula (4.23) (or, Theorem 4.6).

[11] If we think that statistical mechanics belongs to informatics and not physics (*cf.* in this book we consider so), the firm answer may not be needed. If the thermodynamical condition is useful, it is enough.

entropy").[12]

(IE) *the U-entropy $H_U(q(t), p(t))$, (cf. (4.12)), is increasing concerning t, that is*

$$H_U(q(t), p(t)) \uparrow \log[\nu(\mathcal{S}_E)] \quad (\text{ if } t \uparrow \infty) \quad \text{in the sense of (STI)} \tag{4.24}$$

for a suitable small 0-neighborhood U in $\mathcal{M}(\mathbf{R}^6)$.

That is because $H_U(q(t), p(t)) \approx \log[\nu(\mathcal{S}_E)]$ holds for almost every time t in the sense of (STI) if the neighborhood U is chosen suitably. (How to choose the U suitably is our future problem.) Therefore we consider that the law of increasing entropy is hidden behind the thermodynamical condition ((T^1) and (T^2)).

■

4.3 Probabilistic aspects of equilibrium statistical mechanics

In this section we shall study the probabilistic aspects of equilibrium statistical mechanics. Note that the (4.20) implies that the equilibrium statistical mechanical system at almost every time t (in the sense of the (STI)) can be regarded as:

(U) an urn including about 10^{24} particles such as the number of the particles whose states belong to Ξ ($\in \mathcal{B}_{\mathbf{R}^6}$) is given by $\rho_E(\Xi) \times 10^{24}$.

Recall the (A_1) in §4.1, that is, the probability appearing in classical systems (or particularly, in the probabilistic rule in (4.2)) is essentially the same as the probability appearing in urn problems. Therefore, we see, by the above (U),

(A_1') if we choose a particle at random from the urn (="box in Figure (4.3)") at time t, then the probability that the state of the particle belongs to Ξ ($\in \mathcal{B}_{\mathbf{R}^6}$) is given by $\rho_E(\Xi)$.

[12] If my memory serves me right, in some book A. Einstein says: *There is a possibility that someone will find his relativity theory is not true, but there is no possibility that someone will find that the law of increasing entropy is not true.* We can understand what he wants to say, if we think that statistical mechanics should be understood as an application of measurement theory, on the other hand, his relativity theory belongs to theoretical physics. That is, we think that the law of increasing entropy is as "true" as the statement "A cat is stronger than a mouse". (Cf. footnote[9] in Chapter 2.) It should be noted that the statement "A cat is stronger than a mouse" is ambiguous, fuzzy, vague, etc, though it is "almost experimentally true" (*cf.* (I_{14}) in §1.3).

In what follows, we shall represent this (A_1') in terms of measurements. Define the observable $\mathbf{O} = (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F)$ in $C(\mathcal{S}_E)$ such that, (*cf.* (4.11)),

$$[F(\Xi)](q,p) = [R_{D_N}^{(q,p)}](\Xi)\Big(\equiv \frac{\sharp[\{k \mid X_k(q,p) \in \Xi\}]}{\sharp[D_N]}\Big) \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}^6}, \forall (q,p) \in \mathcal{S}_{_E}(\subset \mathbf{R}^{6N})). \tag{4.25}$$

Thus, we have the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(\mathbf{O} \equiv (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F), S_{[\delta_{\psi_t(q_0,p_0)}]})$. Then we see that

(B_1') the probability that the measured value obtained by the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(\mathbf{O} \equiv (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F), S_{[\delta_{\psi_t(q_0,p_0)}]})$ belongs to $\Xi (\in \mathcal{B}_{\mathbf{R}^6})$ is given by $\rho_E(\Xi)$. That is because Theorem 4.6 says that

$$[F(\Xi)](\psi_t(q_0,p_0)) = \rho_E(\Xi) \quad \text{(almost every time } t \text{ in the sense of (STI))} \tag{4.26}$$

which is just the measurement theoretical representation of the (A_1').

Also, we see that

(A_1'') if we choose N' particles at random from the urn (="box in Figure (4.3)"), then statistics say that the distribution of the states of these particles is almost surely expected to be approximately equal to ρ_E, where $1 \ll N' \le N(\approx 10^{24})$.

Here, consider the product observable $\mathbf{O}^{N'} = (\mathbf{R}^{6N'}, \mathcal{B}_{\mathbf{R}^{6N'}}, F^{N'})$ in $C(\mathcal{S}_E)$. For each k $(\in K_{N'} \equiv \{1, 2, ..., N'\})$, define the map $X_k : \mathbf{R}^{6N'} \to \mathbf{R}^6$ such that

$$X_k((x_{1n}, x_{2n}, x_{3n}, x_{4n}, x_{5n}, x_{6n})_{n=1}^{N'}) = (x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k}, x_{6k})$$

for all $x = (x_{1n}, x_{2n}, x_{3n}, x_{4n}, x_{5n}, x_{6n})_{n=1}^{N'}$ in $\mathbf{R}^{6N'}$). Define the map $G_{N'} : \mathbf{R}^{6N'} \to \mathcal{M}_{+1}^m(\mathbf{R}^6)$ $(\equiv \{\rho \in \mathcal{M}(\mathbf{R}^6) : \rho \ge 0,\ \rho(\mathbf{R}^6) = 1\})$ such that

$$G_{N'}(x) = \frac{1}{N'} \sum_{n=1}^{N'} \delta_{X_n(x)} \qquad (\forall x \in \mathbf{R}^{6N'}). \tag{4.27}$$

Then we have the image observable $G_{N'}(\mathbf{O}^{N'}) \equiv (\mathcal{M}_{+1}^m(\mathbf{R}^6), \mathcal{B}_{\mathcal{M}_{+1}^m(\mathbf{R}^6)}, G_{N'}(F^{N'}))$. And we see, by Theorem 4.6, that

(B_1'') the measured value obtained by the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(G_{N'}(\mathbf{O}^{N'}), S_{[\delta_{\psi_t(q_0,p_0)}]})$ is approximately equal to ρ_E.

which is just the measurement theoretical representation of the (A_1'').

4.4 Conclusions

In this chapter we assert that equilibrium statistical mechanics is formulated as follows:[13]

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{((B_1'')(=\text{Axiom } 1))}{\text{“probabilistic rule”}} + \underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{“Newton equation”}}}_{(+\text{ STI})} \qquad \underset{(=(4.4))}{(4.28)}$$

in the framework of PMT.

It may be generally believed that the principle of equal a priori probability and the ergodic hypothesis are two basic principles of statistical mechanics. However, our formulation (4.28) says that the principle of equal a priori probability is not needed (*cf.* Remark 4.5 and Remark 4.8(i)), and moreover, the ergodic hypothesis is a consequence of the thermodynamical condition (i.e., (T^1) and (T^2) under the (EH)), *cf.* the formulas (4.20)$\sim$(4.23).

However we may assert that the following formulation is also possible:

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{((B_1'')(=\text{Axiom } 1))}{\text{“probabilistic rule”}} + \underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{“Newton equation”}}}_{(+\text{ PI})} \qquad (4.30)$$

which is, strictly speaking, related to SMT (*cf.* Chapter 8, Statistical measurement theory).

Thus we have the question:

- Which should be chosen, (4.28) or (4.30)?[14]

[13]Or simply (*cf.* Remark 4.7), we may consider that

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{((B_1'')(=\text{Axiom } 1))}{\text{“probabilistic rule”}} + \underset{\nu_E\text{ (in (4.5)) and (4.21)}}{\text{“Newton equation”}}}_{(+\text{ STI})} \qquad (4.29)$$

We believe that the term “economical” is one of the most important key-words of theoretical informatics (*cf.* Table (1.8b)). In this sense, the (4.29) should be also admitted though we did not focus on the (4.29) in this chapter.

[14]This situation is the same as the following situation. Two ready-made suits “the staying time interpretation” and “the probabilistic interpretation” are on sale. The former is too weak, and so somewhat ambiguous. The latter may be too strong. However, we must choose one from “the staying time interpretation” and “the probabilistic interpretation”. In theoretical informatics, we believe that it can

The reason that we choose (4.28) is as follows: Recall quantum mechanics, in which it is often emphasized that the concept of "probability" is not related to "Schrödinger equation" but "Born's quantum measurements". Comparing quantum mechanics (1.3) and the above (4.28), we have the reason to emphasize that the concept of "probability" is not related to the thermodynamical condition but "probabilistic rule in (4.28)". That is because we want to believe in the spirit that the term of "probability" should be used commonly in both classical and quantum systems, or, that there is no probability without measurements. After all, we say that

- Our proposal (4.28) and quantum mechanics (1.3) are compatible.

On the other hand, the part "Newton equation ((T^1) and (T^2)) under (EH))" in (4.30) is related to the concept of "probability" under the assumption "probabilistic interpretation of $\overline{\nu}_E$". Thus, we think that

- The (4.30) and quantum mechanics (1.3) are not compatible.

Thus, we do not choose the (4.30). However, we may choose the following (4.18):

$$\text{"equilibrium statistical mechanics"} = \underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{"probabilistic rule"}}}_{(+\text{PI})} + \underbrace{\underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{"Newton equation"}}}_{(+\text{STI})} \tag{4.31}$$

This (4.31)[15] and quantum mechanics (1.3) are compatible. Thus, the following question is meaningful in measurement theory.

- Which should be chosen, (4.28) or (4.31)?

This may be a matter of opinion (though it is not serious as statistical mechanics is assumed to belong to theoretical informatics in this chapter). If we are required to say something, we guess that the (4.28) will win more popularity than the (4.31). In fact,

---

not be decided by an experimental test. Or at least, we are convinced that it is not worthwhile deciding it by an experimental test. That is because we believe that nobody wants to challenge the following problem:

- Decide (4.28) or (4.30) (or (4.31)) by an experimental test!

Thus, "(4.28) or (4.30)" should be chosen from the philosophical point of view, if we are urged to choose one. Cf. (I_{15}) in §1.3.

[15]The part $\underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{"probabilistic rule"}}}_{(\text{PI})}$ in (4.31) is characterized as "Proclaim 1" in Chapter 8.

- we prefer (4.28) to (4.31),

since we do not want use the (PI) if possible.[16] This is our opinion, though, in theoretical informatics, we must admit the case that opinion is divided.

We hope that our proposal (4.28) (or, (4.29), (4.31)) will be accepted as the standard formulation of equilibrium statistical mechanics.

4.5 Appendix (The law of large numbers)

As a preparation of our main assertion (i.e., the derivation of the ergodic hypothesis (4.20)), we add the following well-known Lemma 4.9.

Lemma 4.9. [The strong law of large numbers, *cf.* [56]]. *Let* $(\mathcal{S}, \mathcal{B}_{\mathcal{S}}, \nu)$ *be a measure space such that* $\nu(\mathcal{S}) < \infty$. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of bounded measurable (or generally,* L^1*) maps* $X_n : \mathcal{S} \to \mathbf{R}^6$ *such that there exists a normalized measure* ρ *on* $\mathbf{R}^6$ *(i.e.,* $\rho(\mathbf{R}^6) = 1$, $\rho(\Gamma) \geq 0$ $(\forall \Gamma \in \mathcal{B}_{\mathbf{R}^6})$*) such that:*

- (identical distribution)

$$\frac{\nu(\{x \in \mathcal{S} \mid X_n(x) \in \Gamma\})}{\nu(\mathcal{S})} = \rho(\Gamma) \quad (\forall n = 1, 2, ..., \quad \forall \Gamma \in \mathcal{B}_{\mathbf{R}^6}),$$

- (independence) *for any positive integer* N*, it holds that*

$$\frac{\nu(\{x \in \mathcal{S} \mid X_n(x) \in \Gamma_n \ \ (\forall n = 1, 2, ..., N)\})}{\nu(\mathcal{S})} = \bigtimes_{n=1}^{N} \rho(\Gamma_n) \qquad (\forall \Gamma_n \in \mathcal{B}_{\mathbf{R}^6}).$$

Then, there exists a measurable set $\mathcal{N}(\in \mathcal{B}_{\mathcal{S}})$ *such that* $\nu(\mathcal{N}) = 0$ *and*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n(x)} = \rho \quad \text{in the sense of weak* topology of } \mathcal{M}(\mathbf{R}^6),$$

for all $x \in \mathcal{S} \setminus \mathcal{N}$ $(\equiv \{x \mid x \in \mathcal{S}, x \notin \mathcal{N}\})$. *Here* $\delta_w(\in \mathcal{M}_{+1}^m(\mathbf{R}^6))$ *is a point measure at* $w(\in \mathbf{R}^6)$*, i.e.,* $\delta_w(\Gamma) = 1$ *(if* $w \in \Gamma \in \mathcal{B}_{\mathbf{R}^6}$*),* $= 0$ *(if* $w \notin \Gamma \in \mathcal{B}_{\mathbf{R}^6}$*).*

■

In the formula (4.18), readers should see that Lemma 4.9 is used in the part "Newton equation" (and not "probability rule") in our proposal (4.28), that is, Lemma 4.9 (the law of large numbers) is used independently of the concept of "probability".

[16] Also, recall "Occam's razor", that is, *"Given two equally predictive theories, choose the simplest".*