*International Journal of Theoretical Physics, Vol. 29, No. 6, 1990*

# Level I Theory of Large Deviations in the Ideal Gas

**T. Lehtonen[1] and E. Nummelin[2]**

*Received April 5, 1989*

Let $X_1, X_2, \ldots$ be i.i.d. random elements (the states of the particles $1, 2, \ldots$). Let $f$ be an $\mathbb{R}^d$-valued, measurable function (an observable) and let $B \subset \mathbb{R}^d$ be a convex Borel set. Denote $S_n = f(X_1) + f(X_2) + \cdots + f(X_n)$. Using large-deviation theory, it may be shown that, under certain regularity conditions, there exists a point $v_B$ (the dominating point of $B$) so that, given $S_n/n \in B$, actually $S_n/n \to v_B$ in probability as $n \to \infty$. Having this conditional weak law of large numbers as our starting point, we consider physical systems of independent particles, especially the ideal gas. Given an observed energy level, we derive convergence results for empirical means, empirical distributions, and microcanonical distributions. Results are obtained for a closed system with a fixed number of particles as well as for an open particle system in the space (a Poisson random field). Our approach is elementary in the sense that we need not refer to the abstract "level II" theory of large deviations. However, the treatment is not restricted to the so-called discrete ideal gas, but we consider the continuous ideal gas.

## 1. INTRODUCTION

The ideal gas (i.e., a system of a large number of particles with no interaction energy) is a basic example of thermodynamical systems. Let $\Lambda \subset \mathbb{R}^3$ be a given container having the volume $|\Lambda| = V$. Suppose that we can observe the energy density $u$ ( = the internal energy/$V$) and the particle density $\nu$ of the ideal gas closed into $\Lambda$. Let $\Lambda_1 \subset \Lambda$ be a small (compared to $\Lambda$) subcontainer. According to the classical results of Boltzmann and Gibbs, the number of particles in $\Lambda_1$ is Poisson distributed with the parameter $\nu V_1$ and the moments of the particles are independent and identically distributed (i.i.d.) each having a centralized normal distribution with the variance $\sigma^2 = 2mu/3\nu$, where $m$ denotes the mass. The textbooks of thermodynamics usually derive this and corresponding results by simple

[1]Helsinki School of Economics, Department of Mathematics and Statistics, Runeberginkatu 22-24, 00100 Helsinki, Finland.
[2]University of Helsinki, Department of Mathematics, Hallituskatu 15, 00100 Helsinki, Finland.

combinatorial arguments using the Stirling approximation for the arising factorials $n!$ (see, e.g., Martin-Löf, 1979).

In the area of probability theory there has been for about 20 years a growing interest in the probabilistic foundations of thermodynamics. It has turned out that the basic laws of thermodynamics are manifestations of the so-called principle of large deviations. By now there exists an abundance of articles on this subject. An important early reference is the paper by Lanford (1973). The more recent development of the theory is extensively treated in the monograph by Ellis (1985).

As the starting point of this approach one can take the following scheme: Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random elements interpreted as the states of the particles. Considering an observable $f$ (i.e., an $\mathbb{R}^d$-valued measurable function), denote $\xi_i = f(X_i)$ and let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Denote $\mu = \mathbb{E}\xi_1$. According to the law of large numbers, $S_n/n \to \mu$ a.s. (and thus in probability) as $n \to \infty$. Let now $B \subset \mathbb{R}^d$ be a convex Borel set so that $\mu \notin \bar{B}$. Then, under certain regularity conditions, there exists a point $v_B \in \partial B$ (called the dominating point of $B$) so that, given the condition $S_n/n \in B$, actually

$$S_n/n \to v_B \qquad \text{in probability as } n \to \infty \tag{1.1}$$

Ellis does not formulate explicitly a conditional law of large numbers of this type in a general form, but derives results for the special case called the discrete ideal gas. The proofs are based on an abstract "level II" theory of large deviations concerning the empirical distributions of i.i.d. random variables [see Chapter III in Ellis (1985)].

In the present article we will formulate the above law of large numbers in a general form. As a special case we treat the (continuous) ideal gas and it turns out that we do not need to refer to the level II theory. Moreover, by considering a unit box in a Poisson random field having inside it $N$ particles with the states $X_1, X_2, \ldots, X_N$ and taking the state of the box as $\mathbf{X} = (X_1, X_2, \ldots, X_N; N)$, we can easily use our basic result to obtain the asymptotic equivalence of the microcanonical distribution and the grand canonical distribution (see also Aizenman *et al.*, 1978, and Dobrushin and Tirozzi, 1977).

## 2. MICROCANONICAL, CANONICAL, AND EMPIRICAL DISTRIBUTIONS

Let $(E, \mathscr{E})$ be a measurable space. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. $E$-valued random elements having the common law $\mathscr{L}(X_1) = \mathbb{P}X_1^{-1} = P$. We interpret $X_1, X_2, \ldots$ as the *states* of the *particles* $1, 2, \ldots$.

Let $f: E \to \mathbb{R}^d$ be a measurable mapping to be called an *observable*. We assume that $f$ splits into two observables $f = (g, u)$, where $g: E \to \mathbb{R}^{d_1}$ and $u: E \to \mathbb{R}^{d_2}$ with $d_1 + d_2 = d$. We think that information about the states $X_1, X_2, \ldots$ may be obtained through observing the sample mean

$$U_n/n = \sum_{i=1}^{n} u(X_i)/n$$

To be more precise, let $C$ be a (nonempty) convex Borel subset of $\mathbb{R}^{d_1}$. When considering the implications of an observation we shall be concerned with the conditional probabilities

$$\mathbb{P}_{n,C} := \mathbb{P}(\cdot \mid U_n/n \in C)$$

to be called the *microcanonical probabilities*. The induced conditional distributions

$$P_{n,C}(dx) := \mathbb{P}_{n,C} X_1^{-1}(dx) = \mathbb{P}(X_1 \in dx \mid U_n/n \in C)$$

are called the *microcanonical distributions* (of the states).

Let $\beta \in \mathbb{R}^{d_2}$ be fixed. Suppose that the Laplace transform $Z(\beta) = \mathbb{E}e^{\langle \beta, u(X_1) \rangle}$ is finite. Then the probability distribution

$$P_\beta(dx) = Z(\beta)^{-1} e^{\langle \beta, u(x) \rangle} P(dx) \tag{2.1}$$

is called a *canonical distribution*.

We shall also be concerned with empirical distributions: Let $\varepsilon_x$ denote the unit mass at $x \in E$. Then the (random) probability measure

$$\hat{P}_n(dx) = n^{-1} \sum_{i=1}^{n} \varepsilon_{X_i}(dx)$$

is called the *empirical distribution* of the states $X_1, X_2, \ldots, X_n$. Our main result (Theorem 1) states that under certain regularity conditions the empirical distributions converge, as $n \to \infty$, in the sense of exponential convergence with respect to the microcanonical probabilities (to be explained below) to a certain canonical distribution. As an easy corollary we obtain convergence of the microcanonical distributions to the same canonical distribution.

Throughout Sections 3–5 we consider the limiting behavior of empirical and microcanonical distributions, and the spatial distributions do not play any role here. Finally, in Section 6 we discuss the more general situation where the number of particles in a given container is also allowed to be a random variable. This leads us to consider the convergence of the microcanonical distribution of the random collection of particles in a container toward a limiting distribution to be called the *grand canonical distribution*.

## 3. PRELIMINARIES FROM LARGE-DEVIATION THEORY

In order to formulate our hypotheses and results in precise terms, we have to introduce some terminology and preliminary results from the large-deviation theory concerning the sums of i.i.d. $\mathbb{R}^d$-valued random variables.

Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d. $\mathbb{R}^d$-valued random variables with the common law $\mathscr{L}(\xi_1) = \rho$, and let $\mu = \mathbb{E}\xi_1 = \int s\,\rho(ds)$. Denote by $\phi$ the Laplace transform $\phi(t) = \int e^{\langle t,s\rangle}\,\rho(ds)$. We call $c(t) = \log\phi(t)$ the *free energy* function. Let

$$\mathscr{D} = \{t \in \mathbb{R}^d;\ \phi(t) < \infty\} = \{t \in \mathbb{R}^d;\ c(t) < \infty\}$$

and let $\mathscr{S} = \overline{\mathrm{co}}(\mathrm{supp}\,\rho)$ be the closed convex hull of the support of $\rho$. Let $\mathrm{ri}(\mathscr{S})$ denote the relative interior of $\mathscr{S}$.

Suppose that the domain $\mathscr{D}$ is nonempty and open (or, more generally, that the free energy function $c$ is essentially smooth). Then $m(t) = \nabla c(t)$ defines a $C^\infty$-homeomorphism $m: \mathscr{D} \to \mathrm{ri}(\mathscr{S})$ (see, e.g., Rockafellar, 1970, Theorem 26.5).

The conjugate distributions $\rho_t$, $t \in \mathscr{D}$, of $\rho$ are defined by

$$\rho_t(ds) = e^{\langle t,s\rangle - c(t)}\,\rho(ds)$$

and we have

$$m(t) = \mathbb{E}_t\xi_1 = \int s\,\rho_t(ds)$$

[Here $\rho_0 = \rho$ and $m(0) = \mu$.]

The convex conjugate function $c^*$ of $c$ is defined by the formula

$$\begin{aligned} c^*(v) &= \sup_{t \in \mathbb{R}^d}\{\langle t, v\rangle - c(t)\} \qquad \text{for} \quad v \in \mathbb{R}^d \\ &= \langle m^{-1}(v), v\rangle - c(m^{-1}(v)) \qquad \text{when} \quad v \in \mathrm{ri}(\mathscr{S}) \end{aligned}$$

(see Rockafellar, 1970, Section 12). We call it the *entropy* function.

Consider a Borel set $B \in \mathbb{R}^d$ and denote $c^*(B) = \inf_{v \in B} c^*(v)$. Suppose that $\mathscr{D}$ is open. Let $B$ be convex and such that $\mathrm{ri}(\mathscr{S}) \cap B^0 \neq \varnothing$ and $\mu \notin \bar{B}$. Then there exists a unique point $v_B \in \bar{B}$, the *dominating point* of $B$, so that $c^*(B) = c^*(v_B)$ and actually $v_B \in \partial B \cap \mathrm{ri}(\mathscr{S})$ (see, e.g., Ney, 1983). Denote $t_B = m^{-1}(v_B)$.

Let $(W_n)$ be a sequence of $\mathbb{R}^d$-valued random variables and let $(\mathbb{P}_n)$ be a sequence of probability measures. Following Ellis (1985), we say that *the random variables* $W_n$ *converge to a random variable* $W_\infty$ *exponentially with respect to the probabilities* $\mathbb{P}_n$ if, for each $\varepsilon > 0$, there exists a constant $I_\varepsilon > 0$ so that

$$\mathbb{P}_n(|W_n - W_\infty| > \varepsilon) < e^{-I_\varepsilon n} \quad \text{eventually}$$

This is denoted by $W_n \to_{\exp} W_\infty$ [wrt $(\mathbb{P}_n)$]. Note that, if, additionally $W_1, W_2, \ldots$ are uniformly bounded, then also

$$\mathbb{E}_n|W_n - W_\infty| \to 0 \qquad \text{as} \quad n \to \infty \tag{3.1}$$

Let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Consider the distribution of the mean $S_n/n$ given that $S_n/n \in B$. Under the assumptions mentioned above, i.e., (i) $\mathscr{D}$ is open and (ii) $B$ is a convex Borel set so that $\mathrm{ri}(\mathscr{S}) \cap B^0 \neq \varnothing$ and $\mu \notin \bar{B}$, it was shown in Nummelin (1987) that

$$S_n/n \xrightarrow[\exp]{} v_B$$

with respect to the conditional probabilities $\mathbb{P}_{n,B} = \mathbb{P}(\cdot\,|S_n/n \in B)$. This result will be used in the sequel and we call it the *conditional weak law of large numbers.* [A more general conditional law of large numbers of level II type can be found in Csiszar (1984).]

## 4. CONVERGENCE TO THE CANONICAL DISTRIBUTION

In this section we shall formulate and prove the convergence results mentioned in Section 2.

We consider the scheme introduced at the beginning of Section 2. Thus, $X_1, X_2, \ldots$ denote states of particles and $f = (g, u)$ is an $\mathbb{R}^d$-valued observable. For any $v \in \mathbb{R}^d$, $v_1$ and $v_2$ refer to the splitting $v = (v_1, v_2)$ corresponding to the decomposition $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We apply the results of Section 3 to the random variables $\xi_i = f(X_i)$ and so, e.g., $c(t) = \log \mathbb{E} e^{\langle t, f(X_1)\rangle}$, $t \in \mathbb{R}^d$. On the other hand, in the case of the random variables $u(X_i)$ we use the subscript $u$, so, e.g., $c_u(\beta) = \log \mathbb{E} e^{\langle \beta, u(X_1)\rangle}$, $\beta \in \mathbb{R}^{d_2}$.

Suppose that $\mathscr{D}_u$ is open. Suppose further that $g$ is bounded. Then the domain of $c$ is clearly the open set $\mathscr{D} = \mathbb{R}^{d_1} \times \mathscr{D}_u$.

Let $\alpha \in \mathrm{ri}(\mathscr{S}_u)$. Denote $\beta = m_u^{-1}(\alpha)$ $(\in \mathbb{R}^{d_2})$, $t_\alpha = (0, \beta)$ $(\in \mathbb{R}^d)$, and $v_\alpha = m(t_\alpha)$. Note that the following "contraction principle" holds:

$$c^*(v_\alpha) = \inf_{v: v_2 = \alpha} c^*(v) \tag{4.1}$$

$$= c_u^*(\alpha) \tag{4.2}$$

This is seen because for any $v \in \mathbb{R}^d$ with $v_2 = \alpha$

$$c^*(v) \geq \langle t_\alpha, v\rangle - c(t_\alpha)$$
$$= \langle \beta, \alpha\rangle - c_u(\beta) = c_u^*(\alpha)$$

with the equality in the case $v = v_\alpha$.

Suppose that $C \subset \mathbb{R}^{d_2}$ is a convex Borel set so that $\mathrm{ri}(\mathscr{S}_u) \cap C^0 \neq \varnothing$ and $m_u(0) = \mathbb{E} u(X_1) \notin \bar{C}$. Then $C$ has a dominating point (for $c_u^*$) and we denote it by $\alpha_C$. Denote $B = \mathbb{R}^{d_1} \times C$ and consider the dominating point of $B(=v_B)$. We have

$$c^*(B) = \inf_{v_1 \in \mathbb{R}^{d_1}, \alpha \in C} c^*(v_1, \alpha)$$
$$= \inf_{\alpha \in C} c^*(v_\alpha) \qquad \text{[by (4.1)]}$$
$$= \inf_{\alpha \in C} c_u^*(\alpha) \qquad \text{[by (4.2)]}$$

and thus uniquely

$$v_B = v_\alpha = m(t_\alpha) = m(0, \beta) \tag{4.3}$$

where

$$\beta = m_u^{-1}(\alpha) \qquad \text{and} \qquad \alpha = \alpha_C \tag{4.4}$$

The connection of $\alpha$ and $\beta$ may also be expressed as

$$\alpha = m_u(\beta) = \mathbb{E}_\beta u(X_1) = \int e^{\langle \beta, u(x) \rangle - c_u(\beta)} u(x) \, P(dx) = \int u(x) \, P_\beta(dx)$$

where $P_\beta$ is the canonical distribution (2.1) with $\beta$ chosen according to (4.4). Finally note that $v_B$ is split as

$$(v_B)_2 = \mathbb{E}_\beta u(X_1) = \alpha$$

and

$$(v_B)_1 = \mathbb{E}_\beta g(X_1) = \int g(x) \, P_\beta(dx)$$

Considering the observable $f$, we write $F_n = \sum_{i=1}^n f(X_i)$ and similarly for $g$ and $u$. Note that $F_n/n = (G_n/n, U_n/n)$ and $F_n/n = S_n/n$ in terms of Section 3. Now we are ready to state our basic result.

*Theorem 1.* (Convergence of the empirical mean of an observable with respect to the microcanonical probabilities.) Suppose that (i) the domain $\mathscr{D}_u$ is open, (ii) the function $g$ is bounded, and (iii) $C$ is a convex Borel set with $\mathrm{ri}(\mathscr{S}_u) \cap C^0 \neq \varnothing$ and $m_u(0) = \mathbb{E}u(X_1) \notin \bar{C}$. Let $P_\beta$ be the canonical distribution (2.1) where $\beta = m_u^{-1}(\alpha_C)$. Then

$$G_n/n \xrightarrow[\exp]{} \int g(x) \, P_\beta(dx) \qquad \text{as} \quad n \to \infty$$

with respect to the microcanonical probabilities $\mathbb{P}_{n,C} = \mathbb{P}(\cdot \mid U_n/n \in C)$.

*Proof.* The condition $U_n/n \in C$ is equivalent to $F_n/n \in B = \mathbb{R}^{d_1} \times C$. By the conditional weak law of large numbers $F_n/n \to_{\exp} v_B$ wrt the probabilities

$\mathbb{P}(\cdot|F_n/n \in B)$. But then it is seen that $G_n/n \to_{\exp} (v_B)_1$ with respect to the probabilities $\mathbb{P}(\cdot|U_n/n \in C)$. ■

Now we apply Theorem 1 to empirical distributions. Let $\pi_1, \pi_2, \ldots$ be a sequence of random probability measures and let $P$ be a (fixed) probability measure on $(E, \mathscr{E})$. We say that the *random probabilities* $\pi_n$ *converge to* $P$ *exponentially with respect to the probabilities* $\mathbb{P}_n$ if, for all bounded, measurable functions $g: E \to \mathbb{R}$ the random variables $\int g\, d\pi_n$ converge to the limit $\int g\, dP$ exponentially with respect to the probabilities $\mathbb{P}_n$. This is denoted by $\pi_n \to_{\exp} P$ [wrt $(\mathbb{P}_n)$].

Recall that $\hat{P}_n = n^{-1} \sum_{i=1}^n \varepsilon_{X_i}$ denotes the empirical distribution of the sample $X_1, X_2, \ldots, X_n$. Let $g$: $E \to \mathbb{R}$ be bounded and measurable. Then

$$\int g\, \hat{P}_n(dx) = \sum_{i=1}^{n} g(X_i)/n = G_n/n$$

Applying Theorem 1 (and using its notations), we obtain the following result.

*Corollary 1.* (Conditional convergence of the empirical distributions to a canonical distribution.) Under the assumptions (i) and (iii) of Theorem 1,

$$\hat{P}_n \xrightarrow[\exp]{} P_\beta \quad [\text{with } \beta = m_u^{-1}(\alpha_C)] \qquad \text{as} \quad n \to \infty$$

with respect to the microcanonical probabilities $\mathbb{P}_{n,C} = \mathbb{P}(\cdot|U_n/n \in C)$.

Let now $P, P_1, P_2, \ldots$ be a sequence of (nonrandom) probability measures on $(E, \mathscr{E})$. We say that the probabilities $P_n$ *b-converge* to $P$ if

$$\int g\, dP_n \to \int g\, dP \qquad \text{as} \quad n \to \infty$$

for all bounded, measurable functions $g: E \to \mathbb{R}$. This is denoted by $P_n \to_b P$.

It follows from (4.1) that exponential convergence of the random probabilities $\pi_n$ to the probability $P$ implies the $b$-convergence of the nonrandom probability measures

$$\mathbb{E}_n\, \pi_n(dx) := \int \pi_n(\omega, dx)\, \mathbb{P}_n(d\omega)$$

to the same limit $P$. This observation leads us to the following corollary.

*Corollary 2.* (Convergence of the microcanonical distributions to a canonical distribution.) Under the assumptions (i) and (iii) of Theorem 1 the microcanonical distributions $P_{n,C} = \mathbb{P}(X_1 \in \cdot\, |U_n/n \in C)$ $b$-converge to the canonical distribution $P_\beta$ [with $\beta = m_u^{-1}(\alpha_C)$] as $n \to \infty$.

## 5. AN APPLICATION: THE CANONICAL DISTRIBUTION OF THE IDEAL GAS

As an application of the above theory, consider a stationary physical system which consists of a large number of identical particles having the states $X_1, X_2, \ldots, X_n$. We assume especially that the observable $u$ is nonnegative and scalar valued. We call $u(X_i)$ the *energy* of particle $i$. About $u$ we have an observation $U_n/n \in (\alpha_1, \alpha_2)$, say, where $(\alpha_1, \alpha_2)$ is a (small) subinterval of $(\mathscr{S}_u)^0 \subset (0, \infty)$.

The choice of the unconditional distribution $P$ of $X_1$ should be consistent with the assumed stationarity. In the case of the ideal gas the state space $E$ is a subset of $\mathbb{R}^6$ and $P$ is (essentially) the Lebesgue measure (see below). Then $P$ is $\sigma$-finite, but it may be interpreted probabilistically via uniform conditional distributions on subsets of finite measure. This situation is similar to the use of the Lebesgue measure as an *a priori* measure in the Bayesian statistical analysis. Note that the earlier results do not demand $P$ (or $\rho$) to be a probability measure, namely all the conjugate measures $P_t$ (or $\rho_t$) are always probability measures.

We suppose that $\mathscr{D}_u \subset (-\infty, 0]$. This holds in the case where $P$ has an infinite total mass, because then

$$\int e^{\beta u(x)}\, P(dx) = \infty \qquad \text{for all} \quad \beta \geq 0$$

It follows that the entropy function is decreasing, because its derivative is

$$(c_u^*)'(\alpha) = m_u^{-1}(\alpha) \leq 0 \qquad \text{for all} \quad \alpha \in (\mathscr{S}_u)^0 \tag{5.1}$$

Consider $C = (\alpha_1, \alpha_2)$. Suppose that $C \subset (\mathscr{S}_u)^0$ and $m_u(0) = \int u(x)\, P(dx) \notin \bar{C}$. Then $C$ has the dominating point $\alpha = \alpha_2$ and we write henceforth $C = (\alpha - \delta, \alpha)$.

Using Corollary 1, it is seen that, given $U_n/n \in C$, the empirical distribution of the particles over states converges exponentially, as $n \to \infty$, to the canonical distribution

$$P_\beta(dx) = (1/Z)\, e^{\beta u(x)}\, P(dx)$$

where $\beta = m_u^{-1}(\alpha)$ and $Z = Z(\beta) = e^{c_u(\beta)}$ is the normalizing constant, usually called the *partition function.* Similarly, by Corollary 2 the microcanonical distributions $P_{n,C}$ $b$-converge to the same limit as $n \to \infty$. The limit $P_\beta$ is called a *Boltzmann distribution.*

The absolute *temperature* $T_\alpha$ corresponding to an energy level $\alpha \in (\mathscr{S}_u)^0$ is defined by

$$T_\alpha = -1/\beta_\alpha \qquad \text{where} \quad \beta_\alpha = m_u^{-1}(\alpha)$$

The usual *thermodynamic entropy* function $s$ is given by $s(\alpha) = -c_u^*(\alpha)$. According to (5.1),

$$-ds = dc_u^* = \beta_\alpha\, d\alpha = (-1/T_\alpha)\, d\alpha$$

and thus one obtains the well-known formula ($T_\alpha = T$)

$$T\, ds = d\alpha$$

## The Ideal Gas

The ideal gas is a physical system which consists of homogeneous noninteracting particles called molecules and having only kinetic energy. We think of the case that the number of molecules is fixed and large and the gas is bounded to a container $\Lambda \subset \mathbb{R}^3$, with $|\Lambda| < \infty$. The state of a molecule is described by $x = (q, p) \in \Lambda \times \mathbb{R}^3 \subset \mathbb{R}^6$, where $q$ and $p$ are the position and the momentum, respectively, of the molecule concerned. As the common distribution of the states of the molecules $X_1, X_2, \ldots$ in $E = \Lambda \times \mathbb{R}^3$ we choose, in accordance with *Liouville's theorem*,

$$P(dx) = dq\, dp/|\Lambda| \qquad \text{for } x = (q, p) \in E,$$

(Martin-Löf, 1979).

The mean energy of $n$ molecules is $U_n/n = n^{-1} \sum_{i=1}^n |p_i|^2/2m$, where $m$ denotes the *mass*. We shall consider the case $U_n/n \in C = (\alpha - \delta, \alpha)$ discussed above. By applying the earlier notation and by calculating, it is seen that the following holds.

Consider the energy observable $u(x) = u(p, q) = |p|^2/2m$. We obtain

$$\begin{aligned} c_u(\beta) &= \log \int e^{\beta u(x)}\, P(dx) \\ &= (-3/2) \log(-\beta/2\pi m) \end{aligned}$$

and so

$$m_u(\beta) = c_u'(\beta) = -3/2\beta$$

for $\beta \in \mathcal{D}_u = (-\infty, 0)$. Then for $\alpha \in (\mathcal{S}_u)^0 = (0, \infty)$

$$\beta = \beta_\alpha = -3/2\alpha$$

and

$$\begin{aligned} c_u^*(\alpha) &= \alpha\beta_\alpha - c_u(\beta_\alpha) \\ &= (-3/2) \log(4\pi e m \alpha/3) \end{aligned}$$

and thus the well-known formula for the thermodynamic entropy is

$$\begin{aligned} s(\alpha) &= -c_u^*(\alpha) \\ &= (3/2) \log(4\pi e m \alpha/3) \end{aligned}$$

Consider $\hat{P}_n$ = the empirical distribution of the molecules, given $U_n/n \in (\alpha-\delta, \alpha)$. Corollary 1 shows that $\hat{P}_n \to_{\exp} P_\beta$ wrt the microcanonical probabilities $\mathbb{P}(\cdot \mid U_n/n \in (\alpha-\delta, \alpha))$ as $n \to \infty$. Similarly, by Corollary 2 the microcanonical distributions $P(X_1 \in \cdot \mid U_n/n \in (\alpha-\delta, \alpha))$ $b$-converge to the same limit $P_\beta$. Here the limiting distribution is the canonical distribution

$$P_\beta(dq, dp) = (1/Z)\, e^{-|p|^2/2mT}\, dq\, dp/|\Lambda|, \qquad (q, p) \in E$$

where $Z = Z(\beta)$ is the partition function and $T = -1/\beta$ is the temperature, with $\beta = \beta_\alpha = -3/2\alpha$ corresponding to the observed energy level $\alpha$.

## 6. CONVERGENCE TO THE GRAND CANONICAL DISTRIBUTION

In this section we consider a stationary particle system in $\mathbb{R}^3$. About the particles $1, 2, \ldots$ we assume that each particle $i$ has a *site* $Q_i$ (in $\mathbb{R}^3$) and a *state* $X_i$ (in a given state space $E$). We assume further that the sites $(Q_i)$ and the states $(X_i)$ are independent and $(Q_i)$ forms a homogeneous Poisson process with the parameter $\nu_0$ (the assumption of a spatial Poisson process is in accordance with the assumed stationarity, since a Poisson distribution is preserved under the Hamiltonian dynamics), while $X_1, X_2, \ldots$ are i.i.d. with the common distribution $\mathbb{P}X_1^{-1} = P$. Consider a Borel set $\Lambda \subset \mathbb{R}^3$ (a *container*). The number of particles in $\Lambda$ is

$$N_\Lambda = \sum_{i: Q_i \in \Lambda} 1$$

If the volume of $\Lambda$ is finite, then $N_\Lambda$ has a Poisson distribution with the parameter $\nu_0|\Lambda|$. For a while let $\Lambda$ be such that $|\Lambda| = 1$ and write for short $N_\Lambda = N$. Denote the states of the particles in $\Lambda$ simply by $X_1, X_2, \ldots, X_N$. We call the random element $\mathbf{X} = (X_1, X_2, \ldots, X_N; N)$ the *grand state* of the unit container $\Lambda$. The element $\mathbf{X}$ obtains values in the state space $\mathbf{E} = \{0\} \cup \bigcup_{n=1}^\infty (E^{\times n} \times \{n\})$ ($\mathbf{X} = 0$ when $N = 0$) and the distribution of $\mathbf{X}$ is given by

$$\begin{aligned}\mathbf{P}(d\mathbf{x}) &= \mathbf{P}(dx_1, dx_2, \ldots, dx_n; n)\\ &= (\nu_0^n/n!)e^{-\nu_0}\, P(dx_1)\, P(dx_2) \ldots P(dx_n)\end{aligned}$$

(In accordance with the earlier case, given $N = n$, the positions of the $n$ particles are i.i.d. and uniformly distributed in $\Lambda$ and the states $X_1, X_2, \ldots, X_n$ are i.i.d. with the distribution $P$.) Let $u: E \to \mathbb{R}$ be an energy observable [i.e., $u(X_i)$ is the energy of particle $i$; cf. Section 4]. In this section we think simultaneously of the energy and the number of the particles in $\Lambda$ and thus we define the *grand energy* observable $\mathbf{u}: \mathbf{E} \to \mathbb{R}^2$ by

$$\mathbf{u}(\mathbf{x}) = \mathbf{u}(x_1, x_2, \ldots, x_n; n) = \left(\sum_{i=1}^n u(x_i), n\right)$$

Henceforth let $\Lambda$ be a large container and let $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ be a partition of $\Lambda$ into unit containers. (In order to avoid unessential technicalities, we assume that $|\Lambda|$ is an integer $= n$.) Denote the state of $\Lambda_i$ by $\mathbf{X}_i$. Then $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are i.i.d. with the common distribution $\mathbf{P}$. Considering this "particle system", we may apply the results of the earlier sections to derive convergence results for hypothetical containers which can interchange particles with their surroundings.

Consider first the free energy function $c_{\mathbf{u}}$. By a direct calculation we obtain

$$\begin{aligned} c_{\mathbf{u}}(\beta, \gamma) &= \log \mathbb{E} \exp[\langle(\beta, \gamma), \mathbf{u}(\mathbf{X}_1)\rangle] \\ &= \log \mathbb{E} \exp\left[\beta \sum_{i=1}^{N} u(X_i) + \gamma N\right] \\ &= \log \sum_{n=0}^{\infty} \mathbb{P}(N=n) \mathbb{E} \exp\left[\beta \sum_{i=1}^{n} u(X_i) + \gamma n\right] \\ &= \log \sum_{n=0}^{\infty} (\nu_0^n/n!) e^{-\nu_0} [\phi_u(\beta) e^{\gamma}]^n \\ &= -\nu_0 + \nu_0 \phi_u(\beta) e^{\gamma} \end{aligned}$$

Suppose that $u$ is such that $\mathscr{D}_u$ is open. Then $\mathscr{D}_{\mathbf{u}} = \mathscr{D}_u \times \mathbb{R}$ is open, too. We obtain the derivative

$$m_{\mathbf{u}}(\beta, \gamma) = (\nu_0 \phi_u'(\beta) e^{\gamma}, \nu_0 \phi_u(\beta) e^{\gamma})$$

For $(\beta, \gamma) \in \mathscr{D}_{\mathbf{u}}$ the conjugate distribution $\mathbf{P}_{(\beta,\gamma)}$ of $\mathbf{P}$ is defined by

$$\mathbf{P}_{(\beta,\gamma)}(d\mathbf{x}) = \{\exp[-c_u(\beta, \gamma) + \langle(\beta, \gamma), \mathbf{u}(\mathbf{x})\rangle]\} \, \mathbf{P}(d\mathbf{x})$$

and it is called a *grand canonical distribution.*

Let now $(\alpha, \nu) \in \mathrm{ri}(\mathscr{S}_{\mathbf{u}})$ and consider the equation $m_{\mathbf{u}}(\beta, \gamma) = (\alpha, \nu)$. Let us solve $(\beta, \gamma)$. First

$$\nu_0 \phi_u'(\beta) e^{\gamma} = \alpha, \qquad \nu_0 \phi_u(\beta) e^{\gamma} = \nu$$

and thus necessarily

$$\alpha/\nu = \phi_u'(\beta)/\phi_u(\beta) = c_u'(\beta) = m_u(\beta)$$

whence

$$\beta = m_u^{-1}(\alpha/\nu)$$

Then

$$\begin{aligned} \gamma &= \log(\nu/\nu_0) - c_u(\beta) \\ &= \log(\nu/\nu_0) - c_u(m_u^{-1}(\alpha/\nu)) \end{aligned}$$

This connection of $(\beta, \gamma)$ and $(\alpha, \nu)$ is one to one.

Consider the *grand entropy* function. For $(\alpha, \nu) \in \mathrm{ri}(\mathcal{S}_u)$ we obtain

$$c_{\mathbf{u}}^*(\alpha, \nu) = \langle(\alpha, \nu), (\beta, \gamma)\rangle - c_{\mathbf{u}}(\beta, \gamma)$$
$$= \alpha\beta + \gamma\nu + \nu_0 - \nu$$

It is also noteworthy that

$$c_{\mathbf{u}}^*(\alpha, \nu) = \nu_0 e^{\gamma} \phi_{\mathbf{u}}^*(\alpha/\nu_0 e^{\gamma}) + \gamma\nu + \nu_0$$

where $\phi_{\mathbf{u}}^*$ is the convex conjugate of the Laplace transform $\phi_{\mathbf{u}}$.

Now we are ready to state the convergence results for the present case. Suppose that an observation is made about the grand mean

$$\mathbf{U}_\Lambda/|\Lambda| := \sum_{i=1}^{n} \mathbf{u}(\mathbf{X}_i)/n \qquad (n = |\Lambda|)$$
$$= \left( \sum_{i:Q_i \in \Lambda} u(X_i)/|\Lambda|, N_\Lambda/|\Lambda| \right)$$

i.e., the observation is about the mean energy and the particle density in the container $\Lambda$. Let the result be $\mathbf{U}_\Lambda/|\Lambda| \in C$, where $C \subset \mathbb{R}^2$ is a convex Borel set, e.g., $C = (\alpha_1, \alpha_2) \times (\nu_1, \nu_2)$. Suppose further that $\mathrm{ri}(\mathcal{S}_{\mathbf{u}}) \cap C^0 \neq \varnothing$ and $\mathbb{E}\mathbf{u}(\mathbf{X}_1) = (\nu_0 \mathbb{E}(u(X_1), \nu_0) \notin \bar{C}$.

Let $g: \mathbf{E} \to \mathbb{R}^d$ be a measurable, bounded function and consider the grand mean of $g$,

$$G_\Lambda/|\Lambda| := \sum_{i=1}^{n} g(\mathbf{X}_i)/n$$
$$= \sum_{i:Q_i \in \Lambda} g(X_i)/|\Lambda|$$

As a counterpart of Theorem 1, we obtain the following corollary:

*Corollary 3.* (Conditional convergence of a grand mean.) Under the assumptions mentioned above,

$$G_\Lambda/|\Lambda| \xrightarrow[\exp]{} \int g(\mathbf{x})\, \mathbf{P}_{(\beta,\gamma)}(d\mathbf{x}) \qquad \text{as} \quad |\Lambda| \to \infty$$

with respect to the probabilities $\mathbb{P}_{\Lambda,C} := \mathbb{P}(\cdot\,|\mathbf{U}_\Lambda/|\Lambda| \in C)$. Here $\mathbf{P}_{(\beta,\gamma)}$ is the grand canonical distribution, where $(\beta, \gamma) = m_{\mathbf{u}}^{-1}(\alpha, \nu)$ and $(\alpha, \nu)$ is the dominating point of $C$.

A grand empirical distribution of $\Lambda$ is defined by

$$\hat{\mathbf{P}}_\Lambda(d\mathbf{x}) = |\Lambda|^{-1} \sum_{i=1}^{|\Lambda|} \varepsilon_{\mathbf{X}_i}(d\mathbf{x})$$

By Corollary 1 we obtain the following result.

*Corollary 4.* (Convergence of grand empirical distributions to a grand canonical distribution.) Under the assumptions mentioned above,

$$\hat{\mathbf{P}}_\Lambda \xrightarrow[\exp]{} \mathbf{P}_{(\beta,\gamma)} \qquad \text{as} \quad |\Lambda| \to \infty$$

with respect to the probabilities $\mathbb{P}_{\Lambda,C}$.

Finally from Corollary 2 we obtain the following.

*Corollary 5.* (Convergence of the microcanonical distributions to a grand canonical distribution.) Under the assumptions mentioned above, the microcanonical distributions $P_{\Lambda,C} := \mathbb{P}(\mathbf{X}_1 \in \cdot \mid U_\Lambda/|\Lambda| \in C)$ $b$-converge to the grand canonical distribution $\mathbf{P}_{(\beta,\gamma)}$ as $|\Lambda| \to \infty$.

Let us look at the grand canonical distribution $\mathbf{P}_{(\beta,\gamma)}$ more accurately. It can be written as

$$\begin{aligned} \mathbf{P}_{(\beta,\gamma)}(d\mathbf{x}) &= \{\exp[-c_{\mathbf{u}}(\beta,\gamma) + \langle(\beta,\gamma), \mathbf{u}(\mathbf{x})\rangle]\}\, \mathbf{P}(d\mathbf{x}) \\ &= \{\exp[-c_{\mathbf{u}}(\beta,\gamma) + \sum_{i=1}^{n} \beta u(x_i) + \gamma n]\}\, \mathbf{P}(dx_1, \ldots, dx_n; n) \\ &\;\;\vdots \\ &= (\nu^n/n!)[\exp(-\nu)]\, P_\beta(dx_1)\, P_\beta(dx_2) \ldots P_\beta(dx_n) \end{aligned}$$

where $P_\beta$ is the canonical distribution with $\beta = m_u^{-1}(\alpha/\nu)$. Thus, in the limit, the number of particles in a unit container is Poisson-distributed with the parameter $\nu$ and, given this number, the states of the particles of the container are i.i.d. and they are distributed accoording to the canonical distribution $P_\beta$ $[\beta = m_u^{-1}(\alpha/\nu)]$. Here $\alpha$ and $\nu$ may be interpreted as the observed mean energy and the observed particle density of the large container.

## The Ideal Gas

As an application of the theory of this section, consider the case of an ideal gas. We choose the site and the state of particle $i$ as $q_i$ (the position in $\mathbb{R}^3$) and $p_i$ (the momentum $\in \mathbb{R}^3$), respectively. We assume that the sites $q_1, q_2, \ldots$ form a spatial Poisson process and, independently, the states $p_1, p_2, \ldots$ are i.i.d. with the common distribution described by the Lebesgue measure (cf. Section 4). With this starting point the theory above applies for the ideal gas.

Let us derive the limiting grand canonical distribution, which corresponds to the observation about the grand energy

$$\mathbf{U}_\Lambda/|\Lambda| \in C := (\alpha_1, \alpha_2) \times (\nu_1, \nu_2)$$

Consider the dominating point of $C$. Using the notation introduced above as well as the results of Section 4, we obtain

$$\nabla c_{\mathbf{u}}^{*}(\alpha, \nu) = (\beta, \gamma) \qquad (\alpha, \nu) \in \mathrm{ri}(\mathcal{S}_{\mathbf{u}})$$

and so

$$\partial c_{\mathbf{u}}^{*}/\partial\alpha = \beta = m_u^{-1}(\alpha/\nu) = -3\nu/2\alpha < 0$$

and

$$\partial c_{\mathbf{u}}^{*}/\partial\nu = \gamma = \log(\nu/\nu_0) + (3/2)\log(3\nu/4\pi m\alpha)$$

The latter derivative is $\gtreqless 0$ for $\nu \gtreqless \nu(\alpha) := \nu_0^{2/5}(4\pi m\alpha/3)^{3/5}$. Thus, the dominating point of $(\alpha_1, \alpha_2) \times (\nu_1, \nu_2)$ is $(\alpha, \nu)$, where

$$\alpha = \alpha_2$$

and

$$\nu = \begin{cases} \nu_1 & \text{if} \quad \nu_1 > \nu(\alpha) \\ \nu(\alpha) & \text{if} \quad \nu_1 \leq \nu(\alpha) \leq \nu_2 \\ \nu_2 & \text{if} \quad \nu_2 < \nu(\alpha) \end{cases}$$

Choose this solution $(\alpha, \nu)$ and let $\beta = -3\nu/2\alpha$.

The distribution $P_\beta$ is given by

$$P_\beta(dp) = (1/Z)e^{-|p|^2/2mT}\, dp$$

where

$$T = -1/\beta = 2\alpha/3\nu$$

and

$$Z = (2\pi mT)^{3/2}$$

Thus, the grand canonical distribution is given by the formula

$$\mathbf{P}_{(\beta,\gamma)}(dp_1, \ldots, dp_n; n) = (\nu^n/n!)[\exp(-\nu)](1/Z)^n \times \left[\exp\left(-\sum_{i=1}^{n} |p_i|^2/2mT\right)\right] dp_1 \ldots dp_n$$

Finally consider an arbitrary container $\Lambda_1$ with volume $V$. The grand canonical distribution of the state $\mathbf{X}$ of $\Lambda_1$ is clearly obtained by replacing above $\nu$ by $\nu V$. One may also be interested in the limiting conditional distribution of the pair $(\mathbf{Q}, \mathbf{X}) := (Q_1, \ldots, Q_N; X_1, \ldots, X_N; N)$ given $\mathrm{U}_\Lambda/|\Lambda| \in C$. With an obvious notation,

$$\mathbf{P}(dq_1, \ldots, dq_n; dx_1, \ldots, dx_n; n|\mathrm{U}_\Lambda/|\Lambda| \in C)$$

$$= \mathbf{P}(dx_1, \ldots, dx_n; n|\mathrm{U}_\Lambda/|\Lambda| \in C)(dq_1/V)\ldots(dq_n/V)$$

$$\xrightarrow{b} [(\nu V)^n/n!][\exp(-\nu V)](1/Z)^n$$

$$\times \left[\exp\left(-\sum_{i=1}^{n} |p_i|^2/2mT\right)\right] dp_1 \ldots dp_n\, (dq_1/V)\ldots(dq_n/V)$$

for $p_i \in \mathbb{R}^3$ and $q_i \in \Lambda_1$, $i = 1, \ldots, n$ and $n = (0), 1, \ldots$.

## REFERENCES


Aizenman, M., Goldstein, S., and Lebowitz, J. L. (1978). *Communications in Mathematical Physics*, **62**, 279–302.

Csiszar, I. (1984). *Annals of Probability*, **12**, 768–793.

Dobrushin, R. L., and Tirozzi, B. (1977). *Communications in Mathematical Physics*, **54**, 173–192.

Ellis, R. S. (1985). *Entropy, Large Deviations and Statistical Mechanics*, Springer, New York.

Lanford, O. E. (1973). Entropy and equilibrium states in classical statistical mechanics, in *Statistical Mechanics and Mathematical Problems*, Springer, Berlin, pp. 1–113.

Martin-Löf, A. (1979). *Statistical Mechanics and the Foundations of Thermodynamics*, Springer, Berlin.

Ney, P. (1983). *Annals of Probability* **11**, 158–167.

Nummelin, E. (1987). A conditional weak law of large numbers, in Proceedings of the Seminar on Stability Problems for Stochastic Models, Suhumi, USSR. Lecture Notes in Mathematics, **1142**, 259–262.

Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press, Princeton, New Jersey.