Higher-order moments and the maximum entropy inference: the thermodynamic limit approach

W Jaworski†

Institute of Physics, Nicholas Copernicus University, Grudziądzka 5, 87-100 Toruń, Poland

Abstract. According to Jaynes, statistical mechanics can be interpreted as a special type of statistical inference based on the principle of maximum entropy. The result of such an inference depends on the available information about a given physical system, but the principle itself does not decide what kind of information is essential and what is not. The well known Gibbs canonical state results from the principle when the statistical mean value of energy is supposed to be known. In our paper the maximum entropy inference employing higher-order moments of energy is examined from the point of view of the thermodynamic limit. It is shown that the contribution to the total entropy, arising from the extra information corresponding to the higher-order moments, is $o(N)$, when $N \to \infty$, $N/V = $ constant (N is the number of particles and V is the volume). Consequently, from a purely thermodynamic point of view, this extra information is non-essential and can be neglected in the maximum entropy procedure. These conclusions also characterise the maximum entropy inference employing higher-order moments of (almost) arbitrary extensive quantities. One can say that the maximum entropy inference has a certain 'stability' property with respect to information corresponding to higher-order moments of extensive physical quantities.

1. Introduction

The fundamental statistical mechanical distributions—the so-called Gibbs distributions—have a non-trivial common property: subject to certain constraints they maximise a functional known in statistical mechanics as entropy, and in information theory, probability theory and mathematical statistics as information. This fact enabled Jaynes to formulate a general *principle of maximum entropy* [1-5], as a criterion to single out these probability distributions and density operators that are best suited for a macroscopic description of physical systems. Statistical mechanics was interpreted as a special type of statistical inference based on this principle.

The result of the maximum entropy procedure is determined by the constraints that always accompany it. These constraints express available information about the considered physical system. In practice they depend on the actual experimental situation. In the case of thermodynamic equilibrium, statistical properties of energy, or of energy and the number of particles, are constrained in the well known manner and the Gibbs canonical, microcanonical or great canonical distributions are obtained [6-8].

Constraints that are most frequently applied in maximum entropy procedures are of the mean value type. They correspond to the situation when statistical mean values of some physical quantities are known. Constraints of this type yield probability distributions (density operators) of a convenient analytical form, which generalises the form of the canonical distribution. Such distributions are frequently applied in statistical mechanics (see, e.g. [6, 8-14, 35]).

† Present address: Department of Chemistry, Queen's University, Kingston, Ontario, Canada K7L 3N6.

A straightforward generalisation of the mean value constraints is obtained when not only mean values but also statistical moments of higher order of some physical quantities are taken into account [15, 16]. Maximum entropy distributions corresponding to such constraints can also be found in statistical mechanical applications, cf, e.g., [17-21, 35]. These applications, however, are, generally speaking, different from applications of the distributions described previously. Distributions maximising entropy, subject to constraints for only mean values, are used mainly to describe equilibrium or non-equilibrium states of a physical system as a whole (cf Gibbs distributions). On the contrary, probability distributions maximising entropy under conditions for higher-order moments are applied mainly to approximate 'exact' probability distributions for one or a few specified random variables, when a finite number of their moments is known (the problem of moments [17]).

Any available information about a physical system can be put in the form of some constraints and used in the maximum entropy inference. It depends on the available information whether results of this inference are correct or incorrect. When they are correct, the available information is sufficiently complete. When they are not, some important information is missing. The principle of maximum entropy does not *a priori* decide what information is actually essential and worth seeking and what is practically superfluous. This can be decided only *a posteriori*.

In phenomenological thermodynamics a thermodynamic equilibrium state is uniquely determined by the values of energy U, volume V, the number of particles N, and (maybe) some other extensive quantities $X_1, \ldots, X_n$. From the thermodynamic point of view, information I corresponding to the specified values of $U, V, N, X_1, \ldots, X_n$ is essential and complete. In statistical mechanics the particular values of $U, V, N, X_1, \ldots, X_n$ are interpreted rather as mean values. From the point of view of the Jaynes principle of maximum entropy, information I is also essential and complete enough to give predictions that agree with phenomenological thermodynamics (and with experiment). This fact is crucial. If this were not so, the principle of maximum entropy would be complete nonsense. So it now becomes clear that the question of what information is essential for the maximum entropy inference, and what is superfluous, is interesting and important not only from a purely practical point of view, but also from a theoretical one. This question concerns the principle of maximum entropy in itself.

Let us now analyse the case when statistical mean values of some physical quantities $A_1, \ldots, A_n$ are given, and additionally some of their higher-order moments (generally mixed) are known. In such a case the set of available data D is of the form $D = D_1 \cup D_2$, where $D_1 = \{$the mean values$\}$, $D_2 = \{$the higher-order moments$\}$. D_1 can be viewed as a basic and D_2 as a supplementary data set. Both D_1 and D_2 constrain the probability distribution of $A_1, \ldots, A_n$. Information contained in $D = D_1 \cup D_2$ is obviously more complete than information contained in D_1. Additional information contained in D_2 affects results of the maximum entropy inference. The problem is: in what manner and to what extent does it actually affect these results? In the present paper we study this problem from the point of view of the thermodynamic limit.

We do not consider the most general case, but only a special one, when statistical moments of the Hamiltonian of a physical system are given†. Our approach, however, can be formally extended to more general situations as well (see § 5).

---

† Some properties of maximum entropy distributions corresponding to this situation have been already studied by the author in [22, 23], but not in this context.

2. The maximum entropy formalism

In this section we give a concise introduction to the maximum entropy formalism in classical and quantum mechanics. More complete treatments can be found, e.g., in [6-9].

We consider a physical system of $f < \infty$ degrees of freedom. In the classical description Γ denotes its phase space (of dimension $2f$). In the quantum one $\mathcal{H}$ is the corresponding Hilbert space. A state, an observable, the mean value and the entropy are defined as follows.

A state

(i) In the classical case: a state is an arbitrary probability density on Γ, i.e. a non-negative function ρ, such that $\int \rho \, d\Gamma = 1$.

(ii) In the quantum one: it is an arbitrary density operator, i.e. a non-negative operator ρ acting in $\mathcal{H}$, such that $\text{Tr } \rho = 1$.

An observable

(i) In the classical case it is a real-valued (measurable) function A defined on Γ, i.e. $A : \Gamma \to \mathbb{R}$.

(ii) In the quantum one it is a self-adjoint operator acting in $\mathcal{H}$.

The statistical mean value of an observable A in a state ρ is a real number $\langle A \rangle_\rho$:

$$\langle A \rangle_\rho = \begin{cases} \int A\rho \, d\Gamma & \text{in the classical case} \\ \text{Tr } A\rho & \text{in the quantum case.} \end{cases} \tag{2.1}$$

The entropy (information) of a state ρ is given by

$$S_\rho = -k \langle \ln \rho \rangle_\rho \tag{2.2}$$

where k is the Boltzmann constant†. Throughout this paper we put $k = 1$, to simplify formulae.

Now let $A_1, \ldots, A_n$ be some observables and let $U_1, \ldots, U_n$ denote their mean values in a state ρ_0. Let I be a trivial observable.

(i) In the classical case it is a function such that $I(\gamma) = 1$ for all $\gamma \in \Gamma$.

(ii) In the quantum one it is the identity operator.

We assume that the observables $I, A_1, \ldots, A_n$ are linearly independent.

Suppose that ρ_0 is not explicitly known and that the mean values $U_1, \ldots, U_n$ constitute the only available information about it. So ρ_0 belongs to the *macrostate*, which is defined as the set of all these states that are compatible with this information:

$$\Phi = \{ \rho \,|\, \langle A_i \rangle_\rho = U_i, \ i = 1, 2, \ldots, n \}. \tag{2.3}$$

This is all we know. It is clear that this knowledge becomes, in general, quite insufficient to infer anything about observables that are not linear combinations of $I, A_1, \ldots, A_n$. Any such inference must be based on some additional postulates.

---

† Generally speaking, information I_ρ is defined by $I_\rho = -a \langle \ln \rho \rangle_\rho$, where a is a constant depending on the information unit chosen.

The principle of maximum entropy constitutes such a postulate. According to it, a state $\rho^* \in \Phi$ such that

$$S_{\rho^*} = \max_{\rho \in \Phi} S_\rho \tag{2.4}$$

is best suited for any statistical inference based on the available data (i.e. $U_1, \ldots, U_n$). For instance, the statistical mean value of an observable X is estimated by the expression

$$\langle X \rangle_{\rho^*}. \tag{2.5}$$

When certain regularity conditions [24, 25] are satisfied, ρ^* exists, is unique, and has the form

$$\rho^* = \Omega^{-1}(\alpha_1, \ldots, \alpha_n) \exp\left(-\sum_{i=1}^n \alpha_i A_i \right) \tag{2.6}$$

where

$$\Omega(\alpha_1, \ldots, \alpha_n) = \begin{cases} \int \exp\left(-\sum_{i=1}^n \alpha_i A_i \right) d\Gamma & \text{in the classical case} \\[3mm] \mathrm{Tr} \exp\left(-\sum_{i=1}^n \alpha_i A_i \right) & \text{in the quantum case} \end{cases} \tag{2.7}$$

and the parameters $\alpha_1, \ldots, \alpha_n$ are determined by the equations

$$U_i = -\partial \ln \Omega / \partial \alpha_i \qquad i = 1, 2, \ldots, n \tag{2.8}$$

(when the variational method is used to find ρ^*, then $\alpha_1, \ldots, \alpha_n$ appear to be Lagrange multipliers). The entropy of the state ρ^*, also called the entropy of the macrostate Φ, is given by

$$S = S_{\rho^*} = \ln \Omega + \sum_{i=1}^n \alpha_i U_i. \tag{2.9}$$

When expressed as a function of $U_1, \ldots, U_n$, it satisfies

$$\alpha_i = \partial S / \partial U_i \qquad i = 1, 2, \ldots, n. \tag{2.10}$$

In the particular case when $A_1 = A$, $A_2 = A^2, \ldots, A_n = A^n$, the mean values $U_1, \ldots, U_n$ are statistical moments of the first, second, ..., nth order of the observable A. (Of course, in this case central moments, or cumulants can be used equally well.) When $A = H$ is the Hamiltonian of our system, then the maximum entropy state ρ^* has the form

$$\rho^* = \Omega^{-1}(\alpha_1, \ldots, \alpha_n) \exp\left(-\sum_{i=1}^n \alpha_i H^i \right). \tag{2.11}$$

This is a simple generalisation of the well known Gibbs canonical distribution. Now we proceed to investigate this generalisation.

3. The problem of the thermodynamic limit

We shall consider quite general classical and quantum systems composed of identical particles. N will denote the number of particles in a system, V_N is its volume and H_N the Hamiltonian.

We are going to investigate a generalised Gibbs state of the form

$$\rho_N = \Omega_N^{-1}(\alpha_{1N}, \ldots, \alpha_{nN}) \exp\left(- \sum_{i=1}^{n} \alpha_{iN} H_N^i \right) \tag{3.1}$$

in the thermodynamic limit

$$N \to \infty \qquad V_N \to \infty \qquad V_N/N = v = \text{constant}. \tag{3.2}$$

The state ρ_N depends on the parameters $\alpha_{1N}, \ldots, \alpha_{nN}$. In order to pass to the thermodynamic limit it is necessary first to decide whether and how these parameters depend on N. According to § 2, ρ_N is the maximum entropy state corresponding to the macrostate

$$\Phi_N = \{\rho' \in \mathscr{P}_N \,|\, \langle H_N^j \rangle_{\rho'} = U_{jN}, j = 1, 2, \ldots, n\} \tag{3.3}$$

where $\mathscr{P}_N$ denotes the set of all states of the N-particle system and $U_{1N}, \ldots, U_{nN}$ are the given statistical moments of H_N. It is clear that these moments must depend on N. From phenomenological thermodynamics it is well known that energy is an extensive quantity. Therefore it is reasonable to assume that

$$U_{jN} = N^j u_j + o(N^j) \qquad j = 1, 2, \ldots, n \tag{3.4}$$

where the symbol $a_N = o(b_N)$ means that $\lim_{N \to \infty} a_N/b_N = 0$ and $u_1, \ldots, u_n$ are constants:

$$u_j = \lim_{N \to \infty} N^{-j} U_{jN}. \tag{3.5}$$

The parameters $\alpha_{1N}, \ldots, \alpha_{nN}$ are functions of $U_{1N}, \ldots, U_{nN}$. These functions are given implicitly by (2.8), i.e.

$$U_{jN} = -\frac{\partial \ln \Omega_N}{\partial \alpha_{jN}}. \tag{3.6}$$

When we know the dependence of the moments on N, then, in principle, by solving these equations the dependence of $\alpha_{1N}, \ldots, \alpha_{nN}$ on N can be determined. In practice, however, this programme is unrealisable.

The dependence of $\alpha_{1N}, \ldots, \alpha_{nN}$ on N can be determined using the following argument. Phenomenological thermodynamics suggests that the entropy S_N of the state ρ_N satisfies

$$S_N = Ns + o(N) \tag{3.7}$$

where s is a constant:

$$s = \lim_{N \to \infty} N^{-1} S_N \tag{3.8}$$

(extensivity). Now, using relations (2.10) we have

$$\alpha_{jN} = \frac{\partial S_N}{\partial U_{jN}} = \lim_{\Delta U_{jN} \to 0} \frac{\Delta S_N}{\Delta U_{jN}} = \frac{1}{N^{j-1}} \lim_{\Delta U_{jN} \to 0} \frac{\Delta S_N/N}{\Delta U_{jN}/N^j}. \tag{3.9}$$

So taking into account (3.4) and (3.7) we can expect that

$$\alpha_{jN} = \frac{1}{N^{j-1}} (\beta_j + o(1)) \tag{3.10}$$

where β_j $(j = 1, \ldots, n)$ are constants. In our calculation we shall simply put

$$\alpha_{jN} = \frac{\beta_j}{N^{j-1}} \qquad \beta_j = \text{constant } (j = 1, \ldots, n) \tag{3.11}$$

and calculate the thermodynamic limit for

$$\rho_N = Z_N^{-1}(\beta_1, \ldots, \beta_n) \exp\left(-\sum_{j=1}^n \frac{\beta_j}{N^{j-1}} H_N^j\right) \tag{3.12}$$

where

$$Z_N(\beta_1, \ldots, \beta_n) = \Omega_N(\beta_1, \beta_2/N, \ldots, \beta_n/N^{n-1}) \qquad \beta_n > 0.$$

We shall calculate the limits (3.5), (3.8) and

$$\psi = \lim_{N \to \infty} N^{-1} \ln Z_N \tag{3.13}$$

and investigate some asymptotic properties of the probability distribution for the energy. The results will show that our substitution (3.11) is correct, i.e. it is in agreement with the assumptions (3.4) and (3.7).

4. The calculation

In the case of a classical ideal gas the thermodynamic limit for the states ρ_N (3.12) can be calculated rigorously. This calculation can be found in [23]. Here we are concerned with quite general classical and quantum systems composed of identical particles. It turns out that the method used in [23] can be formally generalised to this case. Our present calculation is a formal one. It becomes rigorous in the case of the ideal gas†.

Let us introduce the following auxiliary function:

$$J_{\nu Nt}(\beta_1, \ldots, \beta_n)$$

$$= \begin{cases} \displaystyle\int H_N^\nu \exp\left(it\frac{H_N}{N} - \sum_{j=1}^n \frac{\beta_j H_N^j}{N^{j-1}}\right) d\Gamma_N & \text{in the classical case} \\[4mm] \displaystyle\operatorname{Tr} H_N^\nu \exp\left(it\frac{H_N}{N} - \sum_{j=1}^n \frac{\beta_j H_N^j}{N^{j-1}}\right) & \text{in the quantum case} \end{cases} \tag{4.1}$$

where $\nu = 0, 1, \ldots$; $t \in \mathbb{R}$, $i = \sqrt{-1}$. We have

$$Z_N = J_{0N0} \qquad U_{jN} = J_{jN0}/J_{0N0} \qquad \chi_N(t) = J_{0Nt}/J_{0N0} \tag{4.2}$$

where χ_N denotes the characteristic function of the observable H_N/N. So the thermodynamic limit problem is equivalent to the problem of finding an asymptotic form for $J_{\nu Nt}$, when $N \to \infty$ ($V_N/N = v = \text{constant}$).

---

† We hope that in a general case the thermodynamic limit for (3.12) can also be calculated rigorously using the approximating Hamiltonian method [26–28], under the assumption that the thermodynamic limit for the canonical distribution exists. We know that this idea works at least in the case $n = 2$, cf [21], lemma 3.1 and 3.2.

Let $g_N(E, Nv)$ denote the density of states for our system with energy E and volume $V_N = Nv$, i.e.

$$g_N(E, Nv) = \begin{cases} \iint \delta(H_N - E) \, \mathrm{d}\Gamma_N = \dfrac{\partial}{\partial E} \displaystyle\int_{H_N \leqslant E} \mathrm{d}\Gamma_N & \text{in the classical case} \\[2ex] \text{the number of linearly independent} \\ \text{solutions of the equation } H_N\varphi = E\varphi & \text{in the quantum case.} \end{cases} \quad (4.3)$$

For simplicity we shall assume that the ground-state energy is zero†. Using the density of states, $J_{\nu Nt}$ can be written as follows:

$$J_{\nu Nt} = \begin{cases} \displaystyle\int_0^\infty g_N(E, Nv)E^\nu \exp\left(\dfrac{\mathrm{i}tE}{N} - \sum_{j=1}^n \dfrac{\beta_j E^j}{N^{j-1}}\right) \mathrm{d}E & \text{in the classical case} \\[3ex] \displaystyle\sum_{k=0}^\infty g_N(E_{kN}, Nv)E_{kN}^\nu \exp\left(\dfrac{\mathrm{i}tE_{kN}}{N} - \sum_{j=1}^n \dfrac{\beta_j E_{kN}^j}{N^{j-1}}\right) & \text{in the quantum case} \end{cases} \quad (4.4)$$

where $\{E_{kN}\}_{k=0}^\infty$ is the spectrum of the operator H_N. It is well known that, when $N \to \infty$, $V_N/N = v = \text{constant}$, this spectrum is usually getting more and more dense. We shall use the continuous spectrum approximation, i.e. the sum $\Sigma_{k=0}^\infty$ in (4.4) will be replaced by the integral $\int_0^\infty \mathrm{d}E$. We assume that this replacement does not affect thermodynamic limit results. When the continuous spectrum approximation is applied, the quantum formula in (4.4) takes the same form as the classical one. Finally we can write

$$J_{\nu Nt} = N^{\nu+1} \int_0^\infty g_N(Nx, Nv)x^\nu \exp\left(\mathrm{i}tx - N\sum_{j=1}^n \beta_j x^j\right) \mathrm{d}x. \quad (4.5)$$

Let us now consider, for a moment, the microcanonical distribution (density operator) $\tilde{\rho}_{EN}$. For $E \geqslant 0$, and $\Delta > 0$ we define it by

$$\tilde{\rho}_{EN} = \tilde{\Xi}_N^{-1}(E, Nv)\delta_\Delta(H_N - E) \quad (4.6)$$

where δ_Δ denotes the function

$$\delta_\Delta(x) = \begin{cases} 1 & \text{for } |x| \leqslant \Delta \\ 0 & \text{for } |x| > \Delta \end{cases} \quad (4.7)$$

and

$$\tilde{\Xi}_N(E, Nv) = \begin{cases} \iint \delta_\Delta(H_N - E) \, \mathrm{d}\Gamma_N & \text{in the classical case} \\[1.5ex] \mathrm{Tr}\, \delta_\Delta(H_N - E) & \text{in the quantum case.} \end{cases} \quad (4.8)$$

Using the density of states $g_N(E, Nv)$ and—in the quantum case—the continuous spectrum approximation, we obtain

$$\tilde{\Xi}_N(E, Nv) = \int_{E-\Delta}^{E+\Delta} g_N(E', Nv) \, \mathrm{d}E' \approx 2\Delta g_N(E, Nv). \quad (4.9)$$

The entropy $S_N^0(E, Nv)$ of the microcanonical state (4.6) is given by

$$S_N^0(E, Nv) = \ln \tilde{\Xi}_N(E, Nv) \approx \ln 2\Delta + \ln g_N(E, Nv). \quad (4.10)$$

---

† We also assume, in order to simplify the notation, that g_N is a function of E and V_N only. In general, it may depend on some other parameters characterising the system, i.e. electric or magnetic fields.

We will assume that the *thermodynamic limit for the microcanonical distribution exists,* i.e.

$$\lim_{N \to \infty} N^{-1} S_N^0(Nu, Nv) = s^0(u, v) \tag{4.11}$$

(independently of Δ). If so, then taking into account (4.9) and (4.10) we can expect that

$$\ln g_N(Nu, Nv) = Ns^0(u, v) + q_N(u, v)$$

$$q_N(u, v) = o(N) \qquad N \to \infty. \tag{4.12}$$

Substituting now (4.12) into (4.5), we obtain

$$J_{\nu Nt} = N^{\nu+1} \int_0^\infty x^\nu \exp\left[N\left(s^0(x, v) - \sum_{j=1}^n \beta_j x^j \right) \right] \exp(itx + q_N(x, v)) \, dx. \tag{4.13}$$

Under certain regularity conditions for s^0 and q_N, the well known Laplace method (cf, e.g., [29–31]) can be used to evaluate asymptotically the integral (4.13). Assuming these regularity conditions to be satisfied, we obtain

$$J_{\nu Nt} \approx \exp\left[N \max_{x \geq 0} \left(s^0(x, v) - \sum_{j=1}^n \beta_j x^j \right) \right] \sum_{k=1}^m A_N^{(k)} x_k \exp(itx_k) \tag{4.14}$$

where $A_N^{(k)} > 0$, and $x_1, \ldots, x_m \geq 0$ denote all these points, which maximise the function $f(x) = s^0(x, v) - \sum_{j=1}^n \beta_j x^j$. Note that β_n must be positive to ensure convergence of expressions (4.1) and that the entropy $s^0(x, v)$ is theoretically equal to the entropy defined in phenomenological thermodynamics. It follows that $f(x)$ assumes its absolute maximum in the open interval $(0, \infty)$, so $x_1, \ldots, x_m > 0$.

Proof. The function $s^0(x, v)$ is a convex function of x, i.e. $\partial^2 s^0/\partial x^2 \leq 0$. So for $x_1, x_2 > 0$, we have $s^0(x_2, v) \leq s^0(x_1, v) + (x_2 - x_1)(\partial s^0/\partial x)(x_1, v) = s^0(x_1, v) + (x_2 - x_1)/T(x_1, v)$, where T is the temperature. Since $\lim_{x \to \infty} T(x, v) = \infty$, the function $f(x) = s^0(x, v) - \sum_{j=1}^n \beta_j x^j$ must remain bounded above when $x \to \infty$ ($\beta_n > 0$). For $x \to +0$, we have $\partial s^0/\partial x = 1/T \to \infty$ (zero is the lower bound for the energy), so that $f(x)$ is increasing for sufficiently small $x > 0$.

A typical behaviour of the numbers $A_N^{(k)}$ is such that $\ln A_N^{(k)} = o(N)$.

Using the convexity of s^0, it is easy to see that for $n = 1, 2, f(x)$ assumes its absolute maximum at a unique point x_1, which solves the equation

$$\beta_1 + 2\beta_2 x_1 = (\partial s^0/\partial x)(x_1, v). \tag{4.15}$$

For $n \geq 3$, the number of points maximising $f(x)$ can be greater than one. The example of the ideal gas ($s^0(x, v) = \frac{3}{2} \ln x + \ln v + \text{constant}$) shows this explicitly [23].

Formula (4.14) yields the following conclusions:

$$\psi = \lim_{N \to \infty} N^{-1} \ln Z_N = \max_{x > 0} \left(s^0(x, v) - \sum_{j=1}^n \beta_j x^j \right) \tag{4.16}$$

$$u_j = \lim_{N \to \infty} N^{-j} U_{jN} = \sum_{k=1}^m a_k x_k^j \qquad j = 1, 2, \ldots \tag{4.17}$$

where

$$a_k = \lim_{N \to \infty} \frac{A_N^{(k)}}{\sum_{k=1}^m A_N^{(k)}} \tag{4.18}$$

$$s = \lim_{N \to \infty} N^{-1} S_N = \psi + \sum_{j=1}^n \beta_j u_j = \sum_{k=1}^m a_k s^0(x_k, v) \tag{4.19}$$

$$\lim_{N \to \infty} \chi_N(t) = \sum_{k=1}^m a_k \exp(\mathrm{i} t x_k). \tag{4.20}$$

We assumed (when $m > 1$) that the limits (4.18) exist. It is always so in the case of the ideal gas [23]. This case also shows explicitly that more than one of the numbers a_k can be non-zero. Note that a_k, $k = 1, \dots, m$, form a probability distribution, since $\sum_{k=1}^m a_k = 1$ and $a_k \geq 0$. Formula (4.20) gives an interpretation to this probability distribution. It means that the probability distributions P_N of the observables H_N / N (energy per particle) converge weakly to the probability distribution P concentrated at the points $x_1, \dots, x_m$ with $P(\{x_k\}) = a_k$. Weak convergence (cf, e.g., [32, 33]) means that $P_N(A) \to_{N \to \infty} P(A)$ for any (Borel) set $A \subset \mathbb{R}$ such that $P(\partial A) = 0$, where ∂A denotes the boundary of A. Formula (4.17) for the moments of H_N / N agrees with the form of the asymptotic distribution P.

The convexity of the entropy s^0 implies the following inequality:

$$s = \sum_{k=1}^m a_k s^0(x_k, v) \leq s^0 \left(\sum_{k=1}^m a_k x_k, v \right) = s^0(u_1, v). \tag{4.21}$$

In the case when at least two of the probabilities a_k are non-zero, and when s^0 is strictly convex (no phase transitions), this inequality is strict.

In the case when only one of the probabilities $a_1, \dots, a_m$ is non-zero (let $a_1 = 1$), in particular when f assumes its absolute maximum at a unique point x_1, equations (4.17)–(4.20) are identical to those obtained using the ordinary canonical distribution (with temperature $T = [(\partial s^0/\partial x)(x_1, v)]^{-1}$) or microcanonical (with energy $E_N = N x_1$). In this case let us consider fluctuations of energy. It is easy to verify that for the generalised canonical state ρ_N (3.12) we have

$$\langle (H_N - \langle H_N \rangle_{\rho_N})^2 \rangle_{\rho_N} = \partial^2 \ln Z_N / \partial \beta_1^2. \tag{4.22}$$

So when $\partial^2 \psi / \partial \beta_1^2$ exists, we can expect that

$$\lim_{N \to \infty} N^{-1} \langle (H_N - \langle H_N \rangle_{\rho_N})^2 \rangle_{\rho_N} = \partial^2 \psi / \partial \beta_1^2. \tag{4.23}$$

The derivative $\partial^2 \psi / \partial \beta_1^2$ can be evaluated using (4.16):

$$\lim_{N \to \infty} N^{-1} \langle (H_N - \langle H_N \rangle_{\rho_N})^2 \rangle_{\rho_N} = \left(\sum_{j=2}^n \beta_j j(j-1) x_1^{j-2} + T^{-2} c_v^{-1} \right)^{-1} \tag{4.24}$$

where $T = [(\partial s^0/\partial x)(x_1, v)]^{-1}$ is the temperature and $c_v = T(\partial s^0/\partial T)_v$ is the specific heat at constant volume†. For $n = 1$ formula (4.24) is the well known formula describing fluctuations in the canonical ensemble. For $n = 2$, $\beta_2 > 0$ and the fluctuation (4.24) is always strictly smaller than it would be in the case of the corresponding canonical distribution (i.e. with temperature $T = ((\partial s^0/\partial x)(x_1, v))^{-1}$). It remains finite even in the case of a phase transition such that $c_v \to \infty$. For $n \geq 3$ the fluctuation can be as well smaller as bigger than for the corresponding canonical distribution, and it can become infinite even when $c_v < \infty$ (cf also [23]).

---

† Also constant other parameters characterising the system, cf footnote on p 921.

The derivatives $C_{kN} = (-\partial/\partial\beta_1)^k \ln Z_N(\beta_1, \ldots, \beta_n)$ are known as cumulants of the observable H_N. The first-order cumulant is simply the mean value of H_N, C_{2N} and C_{3N} are the central statistical moments of H_N and, in general, the kth-order cumulant C_{kN} can be expressed as a (non-linear) combination of the ordinary moments $U_{1N}, \ldots, U_{kN}$ of H_N [33]. In probability theory it is shown that cumulants are additive over independent random variables. In our case, when the derivatives $(-\partial/\partial\beta_1)^k$ exist, we can expect that

$$c_k = \lim_{N \to \infty} N^{-1} C_{kN} = \lim_{N \to \infty} N^{-1}\left(-\frac{\partial}{\partial\beta_1}\right)^k \ln Z_N = \left(-\frac{\partial}{\partial\beta_1}\right)^k \psi \qquad (4.25)$$

and this means that cumulants are extensive quantities. The generalised canonical state ρ_N (3.12) has been defined as maximising the entropy under constraints for the moments $U_{1N}, \ldots, U_{uN}$ of H_N. These moments determine our parameters $\beta_1, \ldots, \beta_n$: $-\partial \ln Z_N/\partial\beta_j = U_{jN}/N^{j-1}$, $j = 1, \ldots, n$. When in formulae (4.17)-(4.20) only one of the probabilities $a_1, \ldots, a_m$ is non-zero, the limit moments $u_j = \lim_{N\to\infty} U_{jN}/N^j$ cannot be used to determine $\beta_1, \ldots, \beta_n$ uniquely, since

$$u_j = -\partial\psi/\partial\beta_j = u_1^j \qquad j = 1, \ldots, n. \qquad (4.26)$$

However the limit cumulant densities $c_1, \ldots, c_n$ can serve this purpose. (This can be verified explicitly using (4.16).) The connection between the maximum entropy procedure and passing to the thermodynamic limit (cf (3.11)) now becomes more clear.

5. Discussion of the results and conclusions

Any available information about a macroscopic physical system is never complete enough to determine uniquely its microscopic state. Among all these states that are compatible with the given information there is one which has maximum entropy. The principle of maximum entropy postulates that this state is best suited to make any statistical inference concerning the system. Using the principle of maximum entropy, the properties of a system in thermal equilibrium can be predicted correctly when the mean energy is known. Our problem was to study the effect of information corresponding to higher-order moments of energy on the results of the maximum entropy inference. We investigated the thermodynamic limit for the generalised canonical distribution (3.12), which has maximum entropy subject to constraints exactly for the moments of the energy.

It appears that two cases are possible.

(i) In the thermodynamic limit the energy per particle $N^{-1}H_N$ becomes a discrete random variable which can assume at least two different values with non-zero probabilities (cf (4.18) and (4.20)). Inequality (4.21) shows that in this, and only in this case, the limit entropy density $s = \lim_{N\to\infty} N^{-1}S_N$ is strictly smaller than it would be for the ordinary canonical distribution[†]. So only in this case the additional information corresponding to the higher-order moments is large enough to be able to decrease the limit entropy density s. It must be stressed at the same time that the case is artificial and does not correspond to any observable state of a real macroscopic system.

[†] In equations (4.19) and (4.21) the entropy s is expressed in terms of the entropy s^0, which is obtained in the thermodynamic limit from the microcanonical distribution. In our discussion we make a standard assumption that the canonical and the microcanonical distributions are equivalent from the point of view of the thermodynamic limit.

(ii) In the thermodynamic limit the energy per particle $N^{-1}H_N$ has a trivial probability distribution concentrated at a single point. In this case the limit entropy density $s = \lim_{N\to\infty} N^{-1}S_N$ is equal to the limit entropy density calculated using the ordinary canonical distribution (see footnote on p 924), i.e. neglecting the higher-order moments. This means that the contribution to the entropy S_N arising from the additional information corresponding to the higher-order moments is of the order $o(N)$ when $N \to \infty$.

In phenomenological thermodynamics the equation that expresses the entropy as a function of extensive parameters (energy, volume,...), or the entropy density as a function of the densities of the extensive parameters, is known as the fundamental equation (cf, e.g., [34]). Knowledge of this equation enables one to calculate any thermodynamic property of the system under consideration. The fact that the generalised canonical distribution (3.12) and the ordinary canonical distribution yield the same fundamental equation is fully satisfactory from the point of view of the information theory approach to thermodynamics and it confirms the reasonableness of the principle of maximum entropy. (It would be difficult to agree that the fundamental equation, and consequently all the thermodynamic properties, depend in some essential way on energy fluctuations.)

So in (ii) the generalised canonical distribution yields the same thermodynamics as the ordinary one. However, the fluctuations, in particular the energy fluctuations, are different from those calculated on the grounds of the canonical distribution (cf (4.22)-(4.24)). This can be observed particularly well in the case of a phase transition such that $\langle (H_N - \langle H_N\rangle)^2\rangle / N \to \infty$ for the canonical distribution, cf (4.24). Thus our results can also be interpreted as follows. *When energy fluctuations differ from those calculated from the canonical distribution, the entropy change corresponding to this difference is of the order $o(N)$, when $N \to \infty$.*

Finally consider the function $\psi(\beta_1,\ldots,\beta_n) = \lim_{N\to\infty} N^{-1}\ln Z_N(\beta_1,\ldots,\beta_n)$. The derivative $c_k = (-\partial/\partial\beta_1)^k\psi$ equals the limit density of the kth cumulant of energy, cf (4.25). The parameters $\beta_1,\ldots,\beta_n$ can be expressed as functions of $c_1,\ldots,c_n$. Then the limit entropy density $s = \psi + \sum_{j=1}^n \beta_j u_1^j$ also becomes a function of $c_1,\ldots,c_n$. Since cumulants of energy are extensive (cf (4.25)), the equation $s = s(c_1,\ldots,c_n,v)$ seems a legitimate generalisation of the thermodynamic fundamental equation (remember that $c_1 = u_1$). But our result is that $s(u_1,c_2,\ldots,c_n,v) = s^0(u_1,v)$, where $s = s^0(u_1,v)$ is the ordinary fundamental equation. So, in fact, there is no generalisation—the entropy of a macroscopic system is independent of $c_2,\ldots,c_n$, i.e. of energy fluctuations, and the thermodynamic forces $\partial s/\partial c_2,\ldots,\partial s/\partial c_n$ vanish. This is in full agreement with the fact that thermodynamic equilibrium state can be realised both in a completely isolated system and in a system exchanging energy with its surroundings.

We investigated a special case of the maximum entropy inference employing higher-order moments. However, it can easily be seen that the method used in § 4 can be formally applied to some more general situations when statistical moments of some extensive physical quantities $A_1,\ldots,A_k$ are given. (In such a case a 'microcanonical' distribution $\tilde{\rho}_N = \tilde{\Xi}_N^{-1}(\mathscr{A}_1,\ldots,\mathscr{A}_k, Nv)\delta_\Delta(A_1 - \mathscr{A}_1)\delta_\Delta(A_2 - \mathscr{A}_2)\ldots\delta_\Delta(A_k - \mathscr{A}_k)$, and a density of states $g_N(\mathscr{A}_1,\ldots,\mathscr{A}_k, Nv)$ must be considered. It is necessary to assume that the thermodynamic limit for $\tilde{\rho}_N$ exists.) The results will be analogous to those derived above. In the analogue of (ii) we will obtain the following. *The contribution to the entropy S_N arising from the extra information corresponding to the higher-order moments is of the order $o(N)$ when $N \to \infty$, $N/V = v = constant$. The entropy density s is, in the thermodynamic limit, independent of fluctuations (cumulant densities of higher*

order) of the quantities $A_1, \ldots, A_k$; *thermodynamic forces corresponding to these fluctuations vanish.*

The main result of this paper is that from a purely thermodynamic point of view, the information corresponding to the higher-order moments of extensive physical quantities is not essential and can be neglected in the maximum entropy procedure. This is, of course, not an unexpected result. Our work, however, is not a proof of a triviality. It can be viewed rather as a test of the maximum entropy formalism and, in fact, of the formula $S_\rho = -\langle \ln \rho \rangle_\rho$. We showed that the maximum entropy inference has a certain 'stability' property with respect to information corresponding to higher-order moments of extensive quantities. This result is reasonable. It can serve as an argument in favour of the maximum entropy method in statistical physics. It also enables us to understand better why these methods are successful.

References

[1] Jaynes E T 1957 *Phys. Rev.* **106** 620
[2] Jaynes E T 1957 *Phys. Rev.* **108** 171
[3] Jaynes E T 1963 *Statistical Physics (1962 Brandeis Lectures)* ed K W Ford (New York: Benjamin)
[4] Ingarden R S and Urbanik K 1961 *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.* **9** 313
[5] Ingarden R S and Urbanik K 1962 *Acta Phys. Polon.* **21** 281
[6] Zubarev D N 1971 *Non-Equilibrium Statistical Thermodynamics* (Moscow: Nauka) (in Russian)
[7] Katz A 1967 *Principles of Statistical Mechanics. The Information Theory Approach* (San Francisco: Freeman)
[8] Hobson A 1971 *Concepts in Statistical Mechanics* (New York: Gordon and Breach)
[9] Ingarden R S and Kossakowski A 1975 *Ann. Phys., NY* **89** 451
[10] Kossakowski A 1969 *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.* **17** 263
[11] Ingarden R S 1973 *Acta Phys. Polon.* A **43** 3
[12] Levine R D and Tribus M (ed) 1981 *The Maximum Entropy Formalism* (Cambridge, MA: MIT Press) 2nd printing
[13] Robertson B 1966 *Phys. Rev.* **144** 151
[14] Robertson B 1967 *Phys. Rev.* **160** 175
[15] Ingarden R S 1963 *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.* **11** 541
[16] Ingarden R S 1965 *Fortschr. Phys.* **13** 755
[17] Mead L R and Papanicolaou N 1984 *J. Math. Phys.* **25** 2404
[18] Jaworski W and Ingarden R S 1981 *Acta Phys. Polon.* A **59** 635
[19] Czajkowski G Z 1981 *Z. Phys.* B **43** 87
[20] Czajkowski G Z and Kociszewski A 1983 *Z. Phys.* B **53** 307
[21] Jaworski W 1985 *Z. Phys.* B **59** 483
[22] Jaworski W 1981 *Acta Phys. Polon.* A **60** 645
[23] Jaworski W 1983 *Acta Phys. Polon.* A **63** 3
[24] Ochs W and Bayer W 1973 *Z. Naturf.* **28a** 693
[25] Ochs W and Bayer W 1973 *Z. Naturf.* **28a** 1571
[26] Tindemans P A J and Capel H W 1975 *Physica* A **79** 478
[27] Den Ouden L W J, Capel H W and Perk J H H 1976 *Physica* A **85** 51
[28] Bogoliubov N N Jr, Brankov J G, Zagrebnov V A, Kurbatov A M and Tonchev N S 1981 *The Approximating Hamiltonian Method in Statistical Physics* (Sofia: Publishing House of the Bulgarian Academy of Sciences)
[29] Fedoruk M V 1977 *The Saddle-Point Method* (Moscow: Nauka) (in Russian)
[30] Riekstynsh E J 1977 *Asymptotic Expansions of Integrals* (Riga: Zinatne) (in Russian)
[31] Olver F W J 1974 *Introduction to Asymptotics and Special Functions* (New York: Academic)
[32] Billingsley P 1977 *Convergence of Probability Measures* (Moscow: Nauka) (Russian transl.)
[33] Gnedenko B V and Kolmogorov A N 1957 *Limit Distributions for Sums of Independent Random Variables* (Warsaw: PWN) (in Polish)
[34] Callen H B 1960 *Thermodynamics* (New York: Wiley)
[35] de la Mota V and Hernandez E S 1985 *Phys. Rev.* A **31** 1095