1 Introduction

The derivation of Boltzmann's distribution as an application of the central limit theorem is taken from [1]. For a short more recent discussion of the foundation of statistical mechanics and their relation to probability see [2]. The discussion of absolute continuity of probability measures is drawn from [3]. The scope of this lecture is to exemplify how some probabilistic tools can be used to construct physical arguments. It is also an invitation to read the classical Khinchin's work [1].

2 Background on Hamiltonian systems

We consider an Hamiltonian system with $d = 2\,N$ degrees of freedom. This means that we associate to a scalar function

$$H : \Omega \mapsto \mathbb{R} \tag{2.1}$$

of the *phase-space* $\Omega \subseteq \mathbb{R}^d$ the first order system of differential equations

$$\dot{\boldsymbol{x}}_t = \mathsf{J}\,\partial_{\boldsymbol{x}_t} H \tag{2.2a}$$

$$\boldsymbol{x}_t|_{t=t_o} = \boldsymbol{x}_o \tag{2.2b}$$

In (2.2) J stands for the linear map

$$\mathsf{J} = \begin{bmatrix} 0 & \mathsf{I}_d \\ -\mathsf{I}_d & 0 \end{bmatrix} \qquad \Rightarrow \qquad \mathsf{J}^{\dagger}\mathsf{J} = \mathsf{I} \tag{2.3}$$

associated to Darboux's coordinates

$$\boldsymbol{x} = \boldsymbol{q} \oplus \boldsymbol{p} \equiv \begin{bmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{bmatrix} \tag{2.4}$$

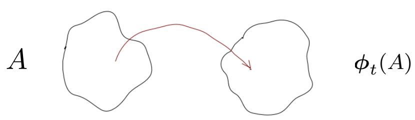
$\boldsymbol{q}$ and $\boldsymbol{p}$ individually being the coordinates of vectors in a $\mathbb{R}^{d/2} \equiv \mathbb{R}^N$ Euclidean space. An immediate consequence of the Hamiltonian structure of the differential equation (2.2) is

$$\dot{H} = \dot{\boldsymbol{x}}_t \cdot \partial_{\boldsymbol{x}_t} H = (\mathsf{J} \cdot \partial_{\boldsymbol{x}} H) \cdot \partial_{\boldsymbol{x}} H = \mathsf{J} : (\partial_{\boldsymbol{x}} H)(\partial_{\boldsymbol{x}} H) = 0 \tag{2.5}$$

Two further consequences are

- Liouville theorem stating that the flow $\boldsymbol{\phi}_t$ generated by (2.2) (i.e. the fundamental solution of (2.2)) is volume preserving

$$|A| = |\boldsymbol{\phi}_t(A)| \tag{2.6}$$



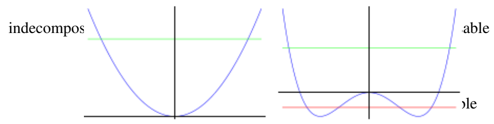
- Birkhoff theorem guaranteeing for any

$$f : \Omega \mapsto \mathbb{R} \tag{2.7}$$

  integrable over Ω ($f \in L^1(\Omega)$) the existence of the limit

$$\bar{f}\,(\boldsymbol{x}) := \lim_{T \uparrow \infty} \frac{1}{T} \int_0^T dt\, f(\boldsymbol{\phi}_t(\boldsymbol{x})) \qquad \forall \boldsymbol{x} \in \Omega \tag{2.8}$$

- if Ω is *metrically indecomposable* meaning that it cannot be represented as $\Omega = \Omega_1 \cup \Omega_2$ with Ω_i, $i = 1, 2$ invariant sets with respect to the flow



(2.9)

ϕ_t then

$$\bar{f}\,(\boldsymbol{x}) \equiv \lim_{T\uparrow\infty} \frac{1}{T} \int_0^T dt\, f(\boldsymbol{\phi}_t(\boldsymbol{x})) = \frac{1}{|\Omega|} \int_\Omega d^d x\, f\,(\boldsymbol{x}) = f_\star \tag{2.10}$$

3 Time and phase-space averages, a (very!) short discussion

Thus from the mathematical point of view, if the energy is the only conserved quantity and the manifold $H = E$ is metrically indecomposable we can replace *time averages* over an infinite time horizon with *phase–space averages* with respect to the uniform measure thereby defined by

$$|\Omega| := \int_\Omega d^d x\, \delta(H - E) \equiv Z(E) \tag{3.1a}$$

$$f_\star = \prec f(\boldsymbol{x}) \succ := \frac{1}{Z(E)} \int_\Omega d^d x\, f\,(\boldsymbol{x})\ \delta(H - E) \tag{3.1b}$$

The function $Z(E)$ expressing the volume of the invariant manifold as a function of the energy is called the partition function of the micro-canonical ensemble. From the properties of the Dirac δ-distribution we can write

$$d\boldsymbol{x} \cdot \partial_{\boldsymbol{x}} H|_{H=E} = dx_\perp \parallel \partial_{\boldsymbol{x}} H \parallel |_{H=E} \tag{3.2}$$

in a suitable system of local coordinates where $\boldsymbol{e}_\perp$ is the unit vector perpendicular to the manifold. With these conventions

$$\int_\Omega d^d x\, \delta(H - E) = \int_\Omega dx_\perp d^{d-1} x_\parallel\, \delta(x_\perp \parallel \partial_{\boldsymbol{x}} H \parallel_{H=E}) = \int_{H=E} \frac{d^{d-1} x_\parallel}{\parallel \partial_{\boldsymbol{x}} H \parallel_{H=E}} \tag{3.3}$$

Example 3.1. Let be the quadratic function

$$H = \sum_{i=1}^{d} x_i^2 \tag{3.4}$$

then we have

$$Z(E) = \int_{\mathbb{R}^d} d^d x\, \delta^{(d)}(E - \sum_{i=1}^{d} x_i^2) = \frac{2\,\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_0^\infty \frac{dr}{r}\, r^d \delta^{(d)}(E - r^2) = \frac{2\,\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \frac{E^{\frac{d-2}{2}}}{2} \tag{3.5}$$

The *partition function* $Z(E)$ in (3.5) is the product of

$$\frac{2\,\pi^{d/2}\, E^{\frac{d-1}{2}}}{\Gamma\left(\frac{d}{2}\right)} = \text{measure of the hypersphere of radius } r = \sqrt{E} \tag{3.6a}$$

$$\parallel \partial_{\boldsymbol{x}} H \parallel = 2\, \parallel \boldsymbol{x} \parallel = 2\sqrt{E} \tag{3.6b}$$

From the physical point of view we can advocate collating *time averages* with experimental measures as the typical time of the measurement process is large in comparison with the typical time-scale of the microscopic dynamics. In order to invoke Birkhoff's theorem further considerations must be added.

- The system may have further integrals of motion (known or unknown) beside the energy. In principle, we should fix the value of each of them to specify the invariant manifold. In practice, we should restrict the attention only to "*controllable*" ones (in Khinchin's [1] terminology) to define the support of phase–space averages. Controllable means that the value of the integral can be experimentally fixed a priori as in the case of the energy. Below we will always assume (as is often the case in statistical mechanics) that the energy is the only controllable integral of motion.
- Physically relevant indicators I are extensive (i.e. additive for independent, non-interacting subsystems) phase-space functions, the typical values whereof furthermore *do not typically deviate from their time average* $I \approx \bar{I}$.
- For non-controllable integrals of the motion the replacement $I \mapsto \bar{I}$ it is physically meaningful also in consequence of the coarse-graining brought up by limitations in experimental resolution.
- In general, we can use $I \approx \bar{I}$ as a crucial argument to advocate the replacement of time averages with phase-space averages even when the typical time required by the dynamics to explore the invariant manifold is much larger than the typical time associated to the measurement process.

Ergodic theory is the mathematical theory investigation the conditions under which the replacement of time averages with phase–space averages. The above considerations should serve as warning that ergodic theory plays a role in the foundations of statistical mechanics only in consequence of the special nature of the physically relevant indicators on which the theory focuses. A mathematically more precise formulation of the concept of "typical value" of the indicators calls for the concept of *absolute continuity* of a probability measure. We are in fact assuming that if the region of phase space where I deviates from the time average $\bar{I}$ is small with respect to the Lebesgue measure on the $H = E$ manifold, then the probability of observing such deviations must be small.

4 Absolute continuity

Consider two (probability) measures defined on $(\Omega, \mathcal{F})$ we have [3]

Definition 4.1. *A (probability) measure* Q *is said to be absolutely continuous with respect to a (probability) measure* P*, denoted as* $\mathrm{Q} \ll \mathrm{P}$*, if for any* $F \in \mathcal{F}$ $\mathrm{P}(F) = 0$ *implies* $\mathrm{Q}(F) = 0$

Suppose

$$\xi \colon \Omega \mapsto \mathbb{R}_{+} \tag{4.1}$$

is a positive definite random variable on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ such that

$$\mathrm{E}_{\mathrm{P}}\xi = 1 \tag{4.2}$$

where the notation E_{P} emphasizes that the expectation is taken respect to the probability measure P. Then for any $F \in \mathcal{F}$ we have readily

$$\mathrm{Q}(F) = \mathrm{E}_{\mathrm{P}}\left\{\xi\, \chi_{F}\right\} \, \geq \, 0 \tag{4.3a}$$

$$\mathrm{Q}(\Omega) = \mathrm{E}_{\mathrm{P}}\left\{\xi \,\right\} = 1 \tag{4.3b}$$

for

$$\chi_{\mathrm{F}}(\omega)=\begin{cases}1 & \text{if } \omega\in F\\ 0 & \text{if } \omega\notin F\end{cases}\tag{4.4}$$

the characteristic function of the set F. The $\mathrm{Q}(\cdot)$ constructed in this way is also a probability measure on $(\Omega,\mathcal{F})$ with density

$$\frac{d\mathrm{Q}}{d\mathrm{P}}(\omega)=\xi(\omega)\tag{4.5}$$

with respect to P. Since $\mathrm{P}(F)=0$ implies $\mathrm{Q}(F)=0$ Q is also absolutely continuous with respect to P. An important theorem of measure theory also states the equivalence of the existence of a density to absolute continuity of the measures:

Theorem 4.1 (Radon-Nikodym)**.** *Suppose that* Q *and* P *are two probability measures on the space* $(\Omega,\mathcal{F})$*. Then there exists a non negative* $\mathcal{F}$*-measurable function* ξ *with* $\mathrm{E}_{\mathrm{P}}\xi=1$*, such that* $\mathrm{Q}(F)=\mathrm{E}_{\mathrm{P}}\left\{\xi\,\chi_F\right\}$ *for every* $F\in\mathcal{F}$*. Moreover,* ξ *is unique in the sense that if* $\tilde{\xi}$ *is another* $\mathcal{F}$*-measurable function with this property, then* $\xi-\tilde{\xi}=0$ *with* $\mathrm{P}=1$*. Hence (4.5) is well-defined.*

5 On the meaning of typical

Let Ω be the manifold of constant energy E for the Hamiltonian system (2.2)

$$\Omega=\left\{\boldsymbol{x}\in\mathbb{R}^d\,|\,H(\boldsymbol{x})=E\right\}\tag{5.1}$$

If we denote by P the probability measure on $(\Omega,\mathcal{B})$ ($\mathcal{B}$ Borel σ-algebra on the constant energy manifold) specified by the micro-canonical ensemble. More precisely for any $B\in\mathcal{B}$

$$\mathrm{P}(B)=\frac{1}{Z(E)}\int_{\mathbb{R}^d}d^dx\,\delta(E-H(\boldsymbol{x}))\,\chi_B(\boldsymbol{x})\tag{5.2}$$

Example 5.1. Let $H=\sum_{i=1}^d x_i^2$, Ω as in (5.1) and $B=\left\{\boldsymbol{x}\in\Omega\,|\,\sum_{i=1}^2 x_i^2\leq E_1\ <E\right\}$. Then we have for $Z(E)$ specified by (3.5)

$$\mathrm{P}(B)=\frac{1}{Z(E)}\int_{\mathbb{R}^d}d^dx\,\delta^{(d)}(E-\sum_{i=1}^d x_i^2)\,\chi_B(\boldsymbol{x})=\frac{\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d-2}{2}\right)\,E^{\frac{d-2}{2}}}\int_0^{\sqrt{E}_1}\frac{dr}{r}\,r^2\,(E-r^2)^{\frac{d-4}{2}}\tag{5.3}$$

Since

$$\int_0^{\sqrt{E}_1}\frac{dr}{r}\,r^2\,(E-r^2)^{\frac{d-4}{2}}=\frac{E^{\frac{d-2}{2}}-(E-E_1)^{\frac{d-2}{2}}}{\frac{d}{2}-1}\tag{5.4}$$

we conclude

$$\mathrm{P}(B)=1-\left(1-\frac{E_1}{E}\right)^{\frac{d-2}{2}}\ \leq\ 1\tag{5.5}$$

We expect that if

$$I\colon \Omega \mapsto \mathbb{R} \tag{5.6}$$

describes a physically relevant indicator then for any reasonably small $\varepsilon > 0$

$$\lim_{d:=2\,N\uparrow\infty} \mathrm{P}\left(\boldsymbol{x}\in\Omega\,|\,|I(\boldsymbol{x})-\mathrm{E_P}I| \geq \varepsilon\right)=0 \tag{5.7}$$

where from now on

$$\mathrm{E_P}I := \frac{1}{Z(E)}\int_{\mathbb{R}^{2d}} d^d x\,\delta(E-H(\boldsymbol{x}))\,I\,(\boldsymbol{x}) \tag{5.8}$$

A result in this direction is the following. By Čebyšev inequality we know that for any positive K

$$\mathrm{P}\left(\boldsymbol{x}\in\Omega\,|\,|I(\boldsymbol{x})-\mathrm{E_P}I| \geq K\right) \leq \frac{\mathrm{E}_P|I-\mathrm{E_P}I|}{K} \tag{5.9}$$

Using Lyapunov inequality we can also write

$$\mathrm{P}\left(\boldsymbol{x}\in\Omega\,|\,|I(\boldsymbol{x})-\mathrm{E_P}I| \geq K\right) \leq \frac{(\mathrm{E}_P|I-\mathrm{E_P}I|^2)^{1/2}}{K} \tag{5.10}$$

As I is by hypothesis an extensive quantity we should have

$$\mathrm{E}_P|I| = O(N) \tag{5.11a}$$

$$\mathrm{E}_P|I-\mathrm{E_P}I|^2 = O(N) \tag{5.11b}$$

If we choose $K = O(N^\alpha)$ for some $0 < \alpha < 1$ it follows immediately that

$$\mathrm{P}\left(\boldsymbol{x}\in\Omega\,|\,|O(\boldsymbol{x})-\mathrm{E_P}O| \geq O(N^\alpha)\right) \leq O(N^{\frac{1-2\alpha}{2}}) \tag{5.12}$$

Thus from extensive indicators hypothesis it follows that deviation from the micro-canonical average cannot, in probability, grow faster with N than the micro-canonical dispersion (i.e. the square root of the variance). In other words, the relative error done in replacing fluctuating quantities with their average vanishes in probability

$$\mathrm{P}\left(\boldsymbol{x}\in\Omega\,|\,\left|\frac{I(\boldsymbol{x})}{\mathrm{E_P}I}-1\right| \geq O(N^{\alpha-1})\right) \leq O(N^{\frac{1-2\alpha}{2}}) \tag{5.13}$$

for $1 > \alpha \geq 1/2$.

6 Criterion for ergodicity

This section closely follows § 13 of [1]. Let us suppose

$$\mathrm{E_P}I = 0 \tag{6.1}$$

then by Lyapunov inequality

$$\mathrm{E_P}\left|\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\,I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right| \leq \left\{\mathrm{E_P}\left[\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\,I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2\right\}^{1/2} \tag{6.2}$$

The inequality implies that if the *correlation function*

$$C(t) := \mathrm{E_P}\left\{I\left(\boldsymbol{\phi}_{t_o}(\boldsymbol{x})\right)\, I\left(\boldsymbol{\phi}_{t_o+t}(\boldsymbol{x})\right)\right\} \tag{6.3}$$

decays sufficiently fast then

$$\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\, I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right) = 0 \tag{6.4}$$

in $\mathbb{L}^1(\Omega)$ sense and also with with probability one over Ω. In order to prove this statement observe first that from the definition of flow

$$\boldsymbol{\phi}_{t+t_o}(\boldsymbol{x}) = \boldsymbol{\phi}_t(\boldsymbol{\phi}_{t_o}(\boldsymbol{x})) \tag{6.5}$$

Then it follows that

$$\int_{\mathbb{R}^d} d^d x\, I\left(\boldsymbol{\phi}_{t_o}(\boldsymbol{x})\right)\, I\left(\boldsymbol{\phi}_{t_o+t}(\boldsymbol{x})\right)\, \delta\left(E - H(\boldsymbol{x})\right) = \int_{\mathbb{R}^d} d^d x\, I\left(\boldsymbol{x}\right)\, I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\, \delta\left(E - H(\boldsymbol{x})\right) \tag{6.6}$$

since by Liouville theorem the Jacobian matrix of the change of variable is equal to the unit and the Hamiltonian is an integral of the motion. Turning to the left hand side of (6.2) we can write

$$\mathrm{E_P}\left[\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\, I_{\tilde{T}}\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 = \mathrm{E_P}\tilde{I}_{\tilde{T}} + \frac{1}{\tilde{T}^2}\mathrm{E_P}\left[\int_0^{\tilde{T}} dt\, I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 \tag{6.7a}$$

$$\tilde{I}_{\tilde{T}}\left(\boldsymbol{x}\right) := \left[\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\, I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 - \left[\frac{1}{\tilde{T}}\int_0^{\tilde{T}} dt\, I\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 \tag{6.7b}$$

The we observe that

$$\mathrm{E_P}\tilde{I}_{\tilde{T}} = \mathrm{E_P}\tilde{I}_{\tilde{T}}\chi_{|\tilde{I}_{\tilde{T}}|\leq\varepsilon} + \mathrm{E_P}\tilde{I}_{\tilde{T}}\chi_{|\tilde{I}_{\tilde{T}}|>\varepsilon} \tag{6.8}$$

Since

$$\lim_{\tilde{T}\uparrow\infty}\tilde{I}_{\tilde{T}} = 0 \qquad\qquad \text{a.s.} \tag{6.9}$$

and for Ω compact there is a $K > 0$ such that $|\tilde{I}_{\tilde{T}}| \leq K$ we have

$$\mathrm{E_P}\tilde{I}_{\tilde{T}}\chi_{|\tilde{I}_{\tilde{T}}|\leq\varepsilon} \leq \varepsilon \tag{6.10}$$

and

$$\lim_{\tilde{T}\uparrow\infty}\mathrm{E_P}\tilde{I}_{\tilde{T}}\chi_{|\tilde{I}_{\tilde{T}}|>\varepsilon} \leq \lim_{\tilde{T}\uparrow\infty} K\,\mathrm{P}\left(|\tilde{I}_{\tilde{T}}| > \varepsilon\right) \leq K\,\mathrm{P}\left(\lim_{\tilde{T}\uparrow\infty}|\tilde{I}_{\tilde{T}}| > \varepsilon\right) = 0 \tag{6.11}$$

The last inequality correspond to the statement that almost sure convergence implies convergence in probability (see [3] pag. 32). Thus we are left with

$$\mathrm{E_P}\left[\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\, I_{\tilde{T}}\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 \leq \varepsilon\left(1+K\right) + \lim_{\tilde{T}\uparrow\infty}\frac{2}{\tilde{T}^2}\int_0^{\tilde{T}} dt_1\int_0^{t_1} dt_2\, C(t_2) \tag{6.12}$$

If $C(t)$ converges sufficiently fast that

$$\int_0^\infty dt\, C(t) < \infty \tag{6.13}$$

then we are entitled to conclude

$$\mathrm{E_P}\left[\lim_{T\uparrow\infty}\frac{1}{T}\int_0^T dt\, I_{\tilde{T}}\left(\boldsymbol{\phi}_t(\boldsymbol{x})\right)\right]^2 = 0 \tag{6.14}$$

7 Boltzmann distribution and central limit theorem

Let us now suppose that we can write the Hamiltonian as

$$H(\boldsymbol{x}) \approx H_1(\boldsymbol{x}_1) + H_2(\boldsymbol{x}_2) \tag{7.1}$$

corresponding to the splitting

$$\boldsymbol{x} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \tag{7.2}$$

where we have

$$\boldsymbol{x}_1 = \begin{bmatrix} x_1 \\ \vdots \\ x_{2\,N_1} \end{bmatrix} \qquad\qquad \boldsymbol{x}_1 = \begin{bmatrix} x_{2\,N_1+1} \\ \vdots \\ x_{2N} \end{bmatrix} \tag{7.3}$$

The $\approx$ implies that we are neglecting an interaction term which is small in some mathematical sense we won't here explore. We suppose from now on that

$$N_1 \ll N \qquad \& \qquad N_2 := N - N_1 \sim O(N) \tag{7.4}$$

Within these approximations we find that for $E = H_1 + H_2$ if B is an event depending only upon $\boldsymbol{x}_1$

$$\mathrm{P}\,(B) = \frac{1}{Z(E)} \int d^{2\,N_1}x_1\, Z_2(E - H(\boldsymbol{x}_1))\, \chi_B(\boldsymbol{x}_1) \tag{7.5}$$

In applications it is often useful to obtain an asymptotic expression for for probabilities of the same type as (7.5). We can obtain the desired asymptotic expression if we assume that Z_2 consists of N_2 quasi independent components

$$Z_2(E - E_1) \sim \int_0^{\infty} \prod_{i=1}^{N_2} dE_i\, Z_{i+1}(E_{i+1}) \delta(E - E_1 - \sum_{j=1}^{N_2} E_{j+1}) \tag{7.6}$$

$E_2 = E - E_1$. To accomplish this program we should associate a probability distribution to the component partition function and then apply the central limit theorem to derive the asymptotic expression for the collective behavior. Following Khinchin we can proceed as follows.

7.1 Generating function and its properties

We expect each of the partition functions to be monotonically increasing functions of their arguments. We also suppose that for any $\beta > 0$ the *generating function*

$$W\,(\beta) = \int_0^{\infty} dE\, Z(E)\, e^{-\beta\, E} \tag{7.7}$$

is well defined.

Proposition 7.1. *Under the above hypotheses (7.7) has well defined derivatives of any order for any* $\beta > 0$

Proof. There exists a E_* such that

$$E^n < e^{\frac{\beta}{2} E} \tag{7.8}$$

for any $E \geq E_*$. Then we can write

$$\int_0^{\infty} dE\, E^n\, Z(E)\, e^{-\beta\, E} \leq \int_0^{E_*} dE\, E^n\, Z(E)\, e^{-\beta\, E} + \int_{E_*}^{\infty} dE\, Z(E)\, e^{-\frac{\beta}{2}\, E} \tag{7.9}$$

□

Proposition 7.2. *The equation*

$$\frac{d}{d\beta}\ln W(\beta) + y = 0 \tag{7.10}$$

admits a unique solution for each $y > 0$

Proof. We observe that

$$U(\beta,y) := e^{\beta\, y}\int_0^\infty dE\, Z(E)\, e^{-\beta\, E} \geq e^{\beta\, y}\int_0^{\frac{y}{2}} dE\, Z(E)\, e^{-\beta\, y} \geq e^{\frac{\beta\, y}{2}}\int_0^{\frac{y}{2}} dE\, Z(E) \tag{7.11}$$

implying

$$\lim_{y\uparrow\infty} e^{\beta\, y}\int_0^\infty dE\, Z(E)\, e^{-\beta\, E} = \infty \tag{7.12}$$

Thus we have

$$\lim_{\beta\downarrow 0}\ln U(\beta,y) = +\infty \qquad \& \qquad \lim_{\beta\uparrow\infty}\ln U(\beta,y) = +\infty \tag{7.13}$$

and by the previous proposition

$$\frac{d^2 \ln U}{d\beta^2}(\beta,y) > 0 \tag{7.14}$$

Thus

$$\ln U(\beta,y) = \beta\, y + \ln W(\beta) \tag{7.15}$$

is a convex function of $\beta > 0$ and as such it must have a unique minimum for each $y > 0$. □

7.2 Probability distribution of individual components

The approximate product form of the partition function (7.6) suggests the energy of individual components should be distributed as

$$p_\beta(E) = \frac{Z(E)\, e^{-\beta\, E}}{W\,(\beta)} \tag{7.16}$$

Alternatively, we may conjecture the partition function to be related to the probability distribution of individua components as

$$Z(E) = W(\beta)\, e^{\beta\, E} p_\beta(E) \tag{7.17}$$

Let us suppose that the energies of the sub-components contributing to behave as independent identically distributed random variables with average and variance respectively specified by m and σ From these assumptions it follows that

$$Z_2(E - E_1) \overset{N_2 \gg 1}{\sim} e^{\beta\,(E-E_1)}\prod_{i=1}^{N_2} W_{i+1}(\beta)\frac{e^{-\frac{(E-E_1-N_2\, m)^2}{2\, N_2\, \sigma^2}}}{(2\,\pi\, N_2\, \sigma^2)^{1/2}} \tag{7.18}$$

It is natural then a similar form for $Z(E)$. The partition function ratio in (7.5) for $N \sim N_2 \gg N_1$ and $E_1 \sim O(N_1)$ yields

$$\frac{Z_2(E-E_1)}{Z(E)} = \frac{e^{-\beta\, H_1(\boldsymbol{x}_1)}}{W_1(\beta)} \left(\frac{N_2}{N}\right)^{1/2} e^{-\frac{(E-E_1-N_2\, m)^2}{2\, N_2\, \sigma^2} + \frac{(E-N\, m)^2}{2\, N\, \sigma^2}} \sim \frac{e^{\beta\, H_1(\boldsymbol{x})}}{W_1(\beta)} e^{-\frac{E_1\, (E-N\, m)}{N\, \sigma^2}} \tag{7.19}$$

By hypothesis energy is an extensive quantity $E = O(N)$ hence we can consistently hypothesize

$$\frac{E}{N\, m} - 1 \ll 1 \tag{7.20}$$

After this last step we arrive to

$$\frac{Z_2(E-E_1)}{Z(E)} \propto e^{-\beta\, H_1(\boldsymbol{x}_1)} \tag{7.21}$$

which is the desired asymptotic expression for the statistical weight of the subsystem of Hamiltonian $H_1(\boldsymbol{x}_1)$. This reduced expression is called the *canonical ensemble*. For a more detailed argumentation see [1].

References

[1] A. I. Khinchin. *Mathematical foundations of statistical mechanics*. Dover series in mathematics and physics. Dover Publications, 1949.

[2] D. B. Malament and S. L. Zabell. Why Gibbs Phase Averages Work – The Role of Ergodic Theory. *Philosophy of Science*, 47(3):339–349, September 1980.

[3] R. van Handel. Stochastic calculus and stochastic control. Lecture Notes, Caltech, 2007.