Course 2S620, Lecture 1
Introduction to Gibbs measure

C. Giardinà *

May 25, 2007

1 Introduction

Statistical mechanics is concerned with the properties of matter in equilibrium in the empirical sense used in thermodynamic. The aim of statistical mechanics is to derive the equilibrium properties of a *macroscopic* system from the law of *microscopic* molecular dynamics.

Macroscopic systems are characterized by few parameters. Think for example of a gas whose thermodynamic behavior is fully specified by pressure, volume and temperature. Each macrostate associated to some fixed values of the thermodynamic variable is compatible, and hence is a summary of, many microstates.

To move from the detailed information contained in microstates to global information of macrostates, statistical mechanics use a probabilistic approach. The question is then what is the appropriate probability distribution to describe a physical macroscopic system made of many interacting components. The appropriate probability distribution (or *ensemble*) is determined by physical macroscopic constraints on the system. Isolated systems are studied with the microcanonical ensemble, for which the energy is fixed. Systems in contact with an heat reservoir are treated making use of the canonical ensemble for which the temperature is fixed. If the number of components in the system is very large, the two ensembles gives equivalent descriptions.

In this first lecture we introduce the general setting of statistical mechanics by looking at the simplest possible model, namely the ideal gas (non interacting particle). We will justify through large deviations the principle of maximum entropy that leads Boltzmann to develop the kinetic theory of gases. Then we will introduce the microcanonical ensemble and argue its equivalence with the canonical ensemble. Making use of the ideal gas system we will motivate the introduction of general Gibbs measure, which are the one to be studied in statistical mechanics. We conclude with

*Department of Mathematics and Computer Science, Eindhoven University, P.O. Box 513 - 5600 MB Eindhoven, The Netherlands, *c.giardina@tue.nl*

the statement of Gärtener-Ellis theorem which express the large deviation property for sequence of dependent random variables.

2 The ideal gas

Consider the following set-up. Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. random variables taking values in a finite set $\Omega = \{x_1, x_2, \ldots, x_r\}$ and having marginal law $\rho = (\rho_1, \ldots, \rho_r)$.We assume the $x_i's$ to the ordered, that is $x_1 < x_2 < \ldots < x_r$. We think of Ω as the set of possible outcomes of a random experiment in which each individual outcome x_i has probability ρ_i of occurring. For each positive integer n, the configuration space for n independent repetitions of the experiment is the product space

$$\Omega_n = \underbrace{\Omega \times \Omega \cdots \times \Omega}_{n \text{ times}}$$

A typical configuration of Ω_n is denoted by $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$. Of course, we have that the probability of each configuration $\omega \in \Omega_n$ is given by the product measure with one dimensional marginal ρ, that is

$$\mathbb{P}_n(\omega) = \prod_{i=1}^{n} \rho(\omega_i)$$

For a given realization ω of the i.i.d. random variables $(X_i)_{i=1,\ldots,n}$, we use the notation that $X_i(\omega) = \omega_i$.

The discrete ideal gas consists of n identical *non-interacting* particles, each having energy levels $x_1, x_2, \ldots, x_r$. For a configuration $\omega \in \Omega_n$ we define $H_n(\omega)$, the total energy in the configuration ω, as

$$H_n(\omega) = \sum_{i=1}^{n} \omega_i$$

In the absence of further information, one assigns the equal probabilities $\rho_i = 1/r$ to each of the x_i. This is one of the fundamental assumption made by Boltzmann in its original works on the ideal gas. Classical statistical mechanics is founded on the following postulate.

Postulate of Equal a Priori Probability. *When a macroscopic system is in thermodynamic equilibrium, its state is equally likely to be any state satisfying the macroscopic conditions of the system.*

In the following we want to understand how the equilibrium state is affected by macroscopic constraints in the so called *thermodynamic limit*, that is when the number of particles $n \to \infty$. For isolated systems the condition to be imposed is the conservation of the total energy H_n. Before treating this case we recall how the equilibrium state is determined in the absence of any constraint. This is given by Sanov's theorem and it will be a first application of the maximum entropy principle.

3 Boltzmann entropy principle and Sanov theorem

We recall the definition of the empirical measure for the model introduced in the previous section. The empirical measure L_n is the empirical probability vector given by

$$\begin{aligned} L_n(\omega) &= (L_{n,1}(\omega), \dots, L_{n,r}(\omega)) \\ &= \frac{1}{n} \left(\sum_{i=1}^{n} \delta_{X_i(\omega)}(x_1), \dots, \sum_{i=1}^{n} \delta_{X_i(\omega)}(x_r) \right) \end{aligned}$$

where $\delta_x(y)$ is the Kronecker delta function. In words, $L_n(\omega)$ counts the relative frequency with which x appears in the configuration ω. L_n is a random probability measure which takes values in the set of probability vectors

$$\mathcal{M}_1 = \left\{ \gamma = (\gamma_1, \gamma_2, \dots, \gamma_r) \in [0,1]^r : \sum_{s=1}^{r} \gamma_s = 1 \right\}$$

The limiting behavior of L_n as $n \to \infty$ is determined by the Sanov theorem, which we now recall in a heuristic formulation. Let $||\cdot||$ denote the Euclidean norm on $\mathbb{R}^r$. For any $\gamma \in \mathcal{M}$ and $\epsilon > 0$ define the open ball

$$B(\gamma, \epsilon) = \{\nu \in \mathcal{M} : ||\gamma - \nu|| < \epsilon\}$$

Since the X_j have common distribution ρ, then for each $i = 1, \dots, r$ we have

$$\mathbb{E}(L_{n,i}(\omega)) = \mathbb{E}\left(\frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}(x_i) \right) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{P}(X_j(\omega) = x_i) = \rho_i \, .$$

Hence by the weak law of large numbers we have, for any $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}_n(L_n \in B(\rho, \epsilon)) = 1$$

It follows that for any γ not equal to ρ and for any $\epsilon > 0$ satisfying $0 < \epsilon < ||\rho - \gamma||$

$$\lim_{n \to \infty} \mathbb{P}_n(L_n \in B(\gamma, \epsilon)) = 0$$

Sanov's theorem implies that the probability in the previous expression converges to 0 exponentially fast and the exponential decay rate is given by the relative entropy.

Theorem 3.1. *Define the relative entropy $I_\rho(\gamma)$ of γ with respect to ρ as*

$$I_\rho(\gamma) = \sum_{s=1}^{r} \gamma_s \log \frac{\gamma_s}{\rho_s} \tag{3.2}$$

Then the sequence of empirical vectors L_n satisfies the large deviation principle on $\mathcal{M}_1$ in the following sense.

(a) Large deviation upper bound.

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(L_n \in C) \leq -\inf_{\gamma\in C} I_\rho(\gamma) \qquad \forall \text{ closed } C \subseteq \mathcal{M}_1 \tag{3.3}$$

(b)Large deviation lower bound.

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(L_n \in O) \leq -\inf_{\gamma\in O} I_\rho(\gamma) \qquad \forall \text{ open } O \subseteq \mathcal{M}_1 \tag{3.4}$$

The relative entropy $I_\rho(\gamma)$ measure the discrepancy between γ and ρ, in the sense that $I_\rho(\gamma) \geq 0$ and $I_\rho(\gamma) = 0$ if and only if $\gamma = \rho$. Thus $I_\rho(\gamma) \geq 0$ attains its infimum of 0 over $\mathcal{M}$ at the unique measure $\gamma = \rho$. Also $I_\rho(\gamma)$ is strictly convex (see exercise 3).

We may heuristically express the large deviation using a formal notation. Then the theorem can be rewritten as

Theorem 3.5.

$$\mathbb{P}_n(L_n \in B(\gamma,\epsilon)) \approx \exp[-nI_\rho(\gamma)] \tag{3.6}$$

where $I_\rho(\gamma)$ is the relative entropy of γ with respect to ρ

Proof. In order to appreciate the original computations performed by Boltzmann we give here an heuristic proof of Sanov's theorem in the formal notation. This is just a simple computation concerning the asymptotic behavior of multinomial coefficient. We may write

$$\begin{aligned}
\mathbb{P}_n(L_n \in B(\gamma,\epsilon)) &= \mathbb{P}_n\left\{\omega \in \Omega_n : L_n(\omega) \sim \frac{1}{n}(n\gamma_1, n\gamma_2, \ldots, n\gamma_r)\right\} \\
&\approx \mathbb{P}_n\left\{\#(\omega_j's = x_1) \sim n\gamma_1, \ldots, \#(\omega_j's = x_r) \sim n\gamma_r\right\} \\
&\approx \frac{n!}{(n\gamma_1)!(n\gamma_2)!\ldots(n\gamma_r)!}\rho_1^{n\gamma_1}\rho_2^{n\gamma_2}\ldots\rho_r^{n\gamma_r}
\end{aligned}$$

Using Stirling's formula $\log(n!) = n\log(n) - n + O(\log n)$, we have

$$\begin{aligned}
\frac{1}{n}\log \mathbb{P}_n(L_n \in B(\gamma,\epsilon)) &\approx \frac{1}{n}\log\left(\frac{n!}{(n\gamma_1)!(n\gamma_2)!\ldots(n\gamma_r)!}\right) + \sum_{s=1}^{r}\gamma_s \log\rho_s \\
&= -\sum_{s=1}^{r}\gamma_s\log\gamma_s + O\left(\frac{\log n}{n}\right) + \sum_{s=1}^{r}\gamma_s\log\rho_s \\
&= -I_\rho(\gamma) + O\left(\frac{\log n}{n}\right)
\end{aligned}$$

□

From the Sanov theorem and the fact that when $\gamma \neq \rho$ then $I_\rho(\gamma) > 0$, while $I_\rho(\rho) = 0$, we deduce that the weak limit of L_n is ρ as $n \to \infty$. In the context of statistical mechanics we call ρ the *equilibrium value* of L_n with respect to the measure $\mathbb{P}_n$. This limit is the simplest example of what is commonly called a maximum entropy principle. We have thus proved the following:

Maximum entropy principle $\gamma_0 \in \mathcal{M}_1$ *is an equilibrium value of* L_n *with respect to* $\mathbb{P}_n$ *if and only if* γ_0 *minimize* $I_\rho(\gamma)$ *over* $\mathcal{M}_1$. *This occurs if and only if* $\gamma_0 = \rho$.

Why is called "maximum entropy principle" if the equilibrium state is obtained by a minimization (not a maximization!) procedure for the relative entropy? This is due to a usual convention in the physics literature and it is explained as follow. If ρ is the uniform measure on Γ, that is $\rho_i = \frac{1}{r}\ \forall i = 1, \ldots, r$ then $I_\rho(\gamma) = \log r + \sum_{i=1}^{r} \gamma_i \log \gamma_i$. The quantity $S(\gamma) = -\sum_{i=1}^{r} \gamma_i \log \gamma_i$ is called the Shannon entropy of γ. Since $-\gamma_i \log \gamma_i \geq 0$, $S(\gamma)$ is non-negative. $S(\gamma)$ is a measure of the randomness in γ. $S(\gamma) = \log r - I_\rho(\gamma) \leq \log r$; $S(\gamma) = \log r$ iff $I_\rho(\gamma) = 0$ and this hold if $\gamma = \rho$. Hence $S(\gamma)$ attains its maximum value of $\log r$ if γ equals the uniform measure ρ. The measure ρ is in a sense the most random probability measure on $\mathcal{M}_1$. At the other extreme $S(\gamma)$ equals 0 if one of the $\rho_i's$ is 1 and the others are all 0. Then the corresponding measures are the least random probability measure on $\mathcal{M}_1$.

4 Microcanonical and canonical ensemble

We now want to consider the equilibrium state for the discrete ideal gas model introduced before, subject to relevant physical constraint. The macroscopic constraint we want to impose is that the total energy per particle

$$\frac{H_n(\omega)}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i(\omega)$$

is kept fixed. In this section we motivate a limit theorem for the empirical distribution L_n conditioned to the total energy conservation. This limit theorem has the added bonus of giving insight into a basic construction in statistical mechanics. As we will see it motivates the form of the Gibbs canonical distribution for the discrete ideal gas and, by extension, for any statistical mechanics model characterized by conservation of energy.

In the absence of conditioning, the weak law of large numbers tell us that - as $n \to \infty$ - the sample mean $H_n(\omega)/n$ should equal approximately the theoretical mean

$$\mathbb{E}(X_1) = \sum_{s=1}^{r} \rho_r x_r$$

and Sanov theorem tell us that $L_n \to \rho$. Let us now assume that $H_n/n \in [z - a, z]$, where a is a small positive number and $x_1 \leq z - a < z < \mathbb{E}(X_1)$ (a similar result would hold if we assumed that $H_n/n \in [z, z+a]$ where $\mathbb{E}(X_1) < z < z+a < x_r$). The question we are after is: determine the probability vector $\rho^* = (\rho_1^*, \ldots, \rho_r^*)$ such that

$$\rho_i^* = \lim_{n\to\infty} \mathbb{P}_n\{X_1 = x_i | H_n/n \in [z-a, z]\}$$

In other words we want ρ^* such that for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}_n(L_n \in B(\rho^*, \epsilon) | H_n/n \in [z-a, z]) = 1$$

Define the closed convex set

$$\Gamma(z) = \left\{ \gamma \in \mathcal{M}_1 : \sum_{i=1}^r x_i \gamma_i \in [z-a, z] \right\}$$

Since for each $\omega \in \Omega_n$

$$\frac{1}{n} H_n(\omega) = \sum_{i=1}^r x_i L_{n,i}(\omega)$$

it follows that

$$\{\omega \in \Omega_n : \frac{1}{n} H_n(\omega) \in [z-a, z]\} = \{\omega \in \Omega_n : L_n(\omega) \in \Gamma(z)\}$$

Thus we want ρ^* such that for any $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}_n(L_n \in B(\rho^*, \epsilon) | L_n \in \Gamma(z)) = 1$$

It follows that for any γ not equal to ρ^* and for any $\epsilon > 0$ satisfying $0 < \epsilon < ||\rho^* - \gamma||$

$$\lim_{n\to\infty} \mathbb{P}_n(L_n \in B(\gamma, \epsilon) | L_n \in \Gamma(z)) = 0$$

The next theorem implies that the probability in the previous expression converges to 0 exponentially fast and the exponential decay rate is again related to the relative entropy (as in Sanov's theorem)

Theorem 4.1.

$$\mathbb{P}_n(L_n \in B(\gamma, \epsilon) | L_n \in \Gamma(z)) \approx \exp[-nI(\gamma)] \tag{4.2}$$

where the rate function $I(\gamma)$ is given by

$$I(\gamma) = \begin{cases} I_\rho(\gamma) - I_\rho(\rho^{(\beta)}) & \textit{if } \gamma \in \Gamma(z) \\ \infty & \textit{if } \gamma \in \mathcal{M}_1 \setminus \Gamma(z) \end{cases} \tag{4.3}$$

and the probability vector $\rho(\beta)$ has the form

$$\rho_i^{(\beta)} = \frac{e^{-\beta x_i} \rho_i}{\sum_{j=1}^r e^{-\beta x_j} \rho_j} \tag{4.4}$$

with $\beta = \beta(z)$ the unique value of β satisfying

$$\sum_{i=1}^r x_i \rho_i^{(\beta)} = z\,. \tag{4.5}$$

We will not prove this theorem. However we will motivate it with the maximum entropy principle. In order to do this we will first prove the next proposition.

Proposition 4.6. *$I_\rho(\gamma)$ attains its infimum over $\Gamma(z)$ at the unique point $\rho^{(\beta)}$ defined in the previous theorem.*

Proof. We may write, for each $i = 1, \ldots, r$

$$\frac{\rho_i^{(\beta)}}{\rho_i} = \frac{e^{-\beta x_i}}{\sum_{j=1}^r e^{-\beta x_j} \rho_j} = \frac{e^{-\beta x_i}}{e^{\varphi(-\beta)}}$$

where we have defined, for $t \in \mathbb{R}$, $\varphi(t) = \log(\sum_{i=1}^r e^{tx_i} \rho_i)$. Hence for any $\gamma \in \Gamma(z)$

$$\begin{aligned} I_\rho(\gamma) &= \sum_{i=1}^r \gamma_i \log \frac{\gamma_i}{\rho_i} = \sum_{i=1}^r \gamma_i \log \frac{\gamma_i}{\rho_i^{(\beta)}} + \sum_{i=1}^r \gamma_i \log \frac{\rho_i^{(\beta)}}{\rho_i} \\ &= I_{\rho^{(\beta)}}(\gamma) - \beta \sum_{i=1}^r x_i \gamma_i - \varphi(-\beta) \end{aligned}$$

If $\gamma = \rho^{(\beta)}$ then, recalling Eq. (4.5) and the fact that $I_{\rho^{(\beta)}}(\rho^{(\beta)}) = 0$, from the previous equation we find

$$I_\rho(\rho^{(\beta)}) = -\beta z - \varphi(-\beta)$$

Consider instead $\gamma \in \Gamma(z)$ with $\gamma \neq \rho^{(\beta)}$. Since $I_{\rho^{(\beta)}}(\gamma) \geq 0$ with equality if and only if $\gamma = \rho^{(\beta)}$, we obtain

$$\begin{aligned} I_\rho(\gamma) &= I_{\rho^{(\beta)}}(\gamma) - \beta \sum_{i=1}^r x_i \gamma_i - \varphi(-\beta) \\ &> -\beta \sum_{i=1}^r x_i \gamma_i - \varphi(-\beta) \\ &\geq -\beta z - \varphi(-\beta) = I_\rho(\rho^{(\beta)}) \end{aligned}$$

where in the last inequality we used the definition of $\Gamma(z)$. We conclude that for any $\gamma \in \Gamma(z)$, $I_\rho(\gamma) \geq I_\rho(\rho^{(\beta)})$ with equality if and only if $\gamma = \rho^{(\beta)}$. Thus $I_\rho(\gamma)$ attains its infimum over $\Gamma(z)$ at the unique point $\rho^{(\beta)}$. □

Form the previous proposition, we see that theorem (4.1) is consistent with the following version of the maximum entropy principle

Maximum entropy principle *Conditioned on the event $S_n/n \in [z - a, z]$, the asymptotically most likely configuration of L_n is $\rho^{(\beta)}$, which is the unique $\gamma \in \mathcal{M}_1$ that minimize $I_\rho(\gamma)$ subject to the constraint that $\gamma \in \Gamma(z)$. In statistical mechanics terminology, $\rho^{(\beta)}$ is the equilibrium macrostate of L_n with respect to the conditional measure $\mathbb{P}_n(\cdot | H_n/n \in [z - a, z])$.*

Let us rederive the expression (4.4) for the equilibrium state under the condition of total energy in the case ρ is the uniform measure. Then we have that $I_\rho(\gamma) = \log r + \sum_{i=1}^r \gamma_i \log \gamma_i = \log r - S(\gamma)$. The maximum of the entropy $S(\gamma)$ under the constraint $\gamma \in \Gamma(z)$ can be obtained by introducing Lagrange multipliers β and μ. We consider the function

$$F(\gamma, \lambda, \mu) = S(\gamma) - \beta \left[\sum_{i=1}^r x_i \gamma_i - z \right] - \mu \left[\sum_{i=1}^r \gamma_i - 1 \right]$$

The stationarity conditions read:

$$\begin{cases} -\log \gamma_i - 1 - \beta x_i - \mu = 0 \\ \sum_i \gamma_i = 1 \\ \sum_i x_i \gamma_i = z \end{cases}$$

From the first equation we deduce that $\gamma_i = e^{-\beta x_i} e^{-(1+\mu)}$. Imposing the two constraints we find

$$\gamma_i = \frac{e^{-\beta x_i}}{\sum_j e^{-\beta x_j}}$$

where β is determined by $\sum_i x_i \gamma_i = z$. We deduce that, in the case ρ is the uniform measure, the maximum of the Shannon entropy $S(\gamma)$ (or equivalently the minimum of the relative entropy $I_\rho(\gamma)$) for $\gamma \in \Gamma(z)$ is attained at $\gamma = \rho^{(\beta)}$.

4.1 Equivalence of ensembles

In the previous section we found that

$$\lim_{n\to\infty} \mathbb{P}_n\{X_1 = x_i | H_n/n \in [z-a, z]\} = \rho_i^{(\beta)}$$

where

$$\rho_i^{(\beta)} = \frac{e^{-\beta x_i} \rho_i}{\sum_{j=1}^r e^{-\beta x_j} \rho_j}$$

with $\beta = \beta(z)$ the unique value of β satisfying $\sum_{i=1}^r x_i \rho_i^{(\beta)} = z$. The next natural question is the following: conditioned on $H_n/n \in [z-a, z]$, as $n \to \infty$ what is the limiting conditional distribution of the random variables $X_1, \ldots, X_k$ which represent the energy levels of the first k particles? Although $X_1, \ldots, X_k$ are independent with respect to the original product measure $\mathbb{P}_n$, this independence is lost when $\mathbb{P}_n$ is replaced by the conditional distribution $\mathbb{P}_n\{\cdot | S_n/n \in [z-a, z]\}$. The answer (somewhat surprising!) is that with respect to $\mathbb{P}_n\{\cdot | S_n/n \in [z-a, z]\}$, the limiting distribution is the product measure on Ω_k with one-dimensional marginal $\rho^{(\beta)}$. In other words, in the limit $n \to \infty$ the independence of $X_1, \ldots, X_k$ is regained. This is expressed by the following

Theorem 4.7. *Given $k \in \mathbb{N}$, $(x_{i_1}, \ldots, x_{i_k}) \in \Omega^k$ we have*

$$\lim_{n\to\infty} \mathbb{P}_n\{X_1 = x_{i_1}, \ldots, X_k = x_{i_k} | H_n/n \in [z-a, z]\} = \prod_{j=1}^k \rho_{i_j}^{(\beta)}$$

Proof. We give just the idea of the proof. We consider $k = 2$; arbitrary $k \in \mathbb{N}$ can be handled similarly. For $\omega \in \Omega_n$ define the pair empirical measure

$$L_n^2(\omega) = \{L_{n,i,j}^2(\omega) : i, j = 1, \dots r\}$$

where

$$L_{n,i,j}^2(\omega) = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k(\omega), X_{k+1}(\omega)}(x_i, x_j)$$

with periodic boundary conditions $X_{n+1}(\omega) = X_1(\omega)$. The empirical pair vector L_n^2 counts the relative frequency with which the pair (x_i, x_j) appears in the configuration ω. L_n^2 takes values in the set

$$\mathcal{M}_2 = \left\{ \tau = (\tau_{i,j})_{i,j=1,\dots,r} \in [0,1]^{r^2} : \sum_{i,j=1}^{r} \tau_{i,j} = 1 \right\}$$

Suppose one can show that $\tau^* = (\rho_i^{(\beta)} \rho_j^{(\beta)})_{i,j=1,\dots,r}$ has the property that for every $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}_n(L_n^2 \in B(\tau^*, \epsilon) | H_n/n \in [z-a, z]) = 1 \ .$$

Then it follows that

$$\lim_{n\to\infty} \mathbb{P}_n\{X_1 = x_i, X_2 = x_j | H_n/n \in [z-a, z]\} = \rho_i^{(\beta)} \rho_j^{(\beta)} \ .$$

Like the analogous limit for the empirical measure, the limit for the pair empirical measure can be proved by showing that the sequence L_n^2, $n \in \mathbb{N}$ satisfies a large deviation principle on $\mathcal{M}_2$ and the rate function attains its infimum at the unique point τ^*. The details are omitted. □

Let us summarize the discussion up to now. For the ideal gas model considered above we have argued that the equilibrium macrostate conditioned on conservation of total energy is given by

$$\lim_{n\to\infty} \mathbb{P}_n\{X_1 = x_{i_1}, \dots, X_k = x_{i_k} | H_n/n \in [z-a, z]\} = \prod_{j=1}^{k} \rho_{i_j}^{(\beta)} \tag{4.8}$$

On the left side we have the microcanonical measure (fixed energy); on the right hand side we have the canonical measure (only the average value of the total energy is fixed through the parameter β, which will have the physical meaning of the inverse temperature). Being the ideal gas a non interacting system, the quantity on the right hand side defines a probability measure $\mathbb{P}_{k,\beta}$ on Ω_k that equals the product measure with one dimensional marginals $\rho^{(\beta)}$.

We now observe that $\mathbb{P}_{k,\beta}$ can be rewritten in terms of the total energy $H_k(\omega) = \sum_{i=1}^{k} \omega_i$ as follows. For $\omega \in \Omega_k$ we have

$$\mathbb{P}_{k,\beta}(\omega) = \prod_{j=1}^{k} \rho^{(\beta)}(\omega_j) = \prod_{j=1}^{k} \frac{e^{-\beta\omega_j} \rho(\omega_j)}{\sum_{\omega_l \in \Omega} e^{-\beta\omega_l} \rho(\omega_l)} = \frac{e^{-\beta H_k(\omega)}}{Z_k(\beta)} \mathbb{P}_k(\omega)$$

where

$$\mathbb{P}_k(\omega) = \prod_{j=1}^{k} \rho(\omega_j)$$

and

$$Z_k(\beta) = \sum_{\omega\in\Omega_k} e^{-\beta H_k(\omega)}\mathbb{P}_k(\omega)$$

with $\beta = \beta(z)$ the unique value of β satisfying $\sum_{i=1}^{r} x_i\rho_i^{(\beta)} = z$. Since $\sum_{i=1}^{r} x_i\rho_i^{(\beta)} = \sum_{\omega\in\Omega_k}[H_k(\omega)/k]\mathbb{P}_{k,\beta}(\omega)$, the constraint on $\beta = \beta(z)$ can be expressed as a constraint on $\mathbb{P}_{k,\beta}$. Namely, choose β so that $\sum_{\omega\in\Omega_k}[H_k(\omega)/k]\mathbb{P}_{k,\beta}(\omega) = z$. The limit in Eq.(4.8) express the equivalence of the microcanonical and canonical ensemble provided β is chosen as above. Since the canonical ensemble has a much simpler form than the microcanonical ensemble, one usually prefer to work with the former.

5 Gibbs states for models in statistical mechanics

The previous discussion motivates the definition of the Gibbs state for a wide class of statistical mechanics models that are defined in terms of an arbitrary energy function, the Hamiltonian H.

Let $H_n : \Omega_n \to \mathbb{R}$ be the Hamiltonian function. H_n defines the microscopic energy of a configuration $\omega \in \Omega_n$. Let β be a parameter proportional to the inverse temperature. Then the canonical ensemble, or *Gibbs state*, is the probability measure that to each each $\omega \in \Omega_n$ assigns the weight

$$\mathbb{P}_{n,\beta}(\omega) = \frac{e^{-\beta H_n(\omega)}}{Z_n(\beta)}\mathbb{P}_n(\omega) \tag{5.1}$$

where $\mathbb{P}_n$ is the apriori product measure on Ω_n with one dimensional marginal ρ. $Z_n(\beta)$ is the normalizing factor that makes $\mathbb{P}_{n,\beta}$ a probability measure. That is

$$Z(\beta) = \sum_{\omega\in\Omega_n} e^{-\beta H_n(\omega)}\mathbb{P}_n(\omega) \tag{5.2}$$

For future use we introduce the following quantities

1. $Z_n(\beta)$ is called the partition function
2. $F_n(\beta) = -\frac{1}{\beta}\ln Z_n(\beta)$ is called the free energy function
3. $U_n(\beta) = \sum_{\omega\in\Omega_n} H_n(\omega)\mathbb{P}_{n,\beta}(\omega)$ is called the internal energy
4. $S_n(\beta) = -\sum_{\omega\in\Omega_n} \mathbb{P}_{n,\beta}(\omega)\log\mathbb{P}_{n,\beta}(\omega)$ is called the thermodynamic entropy

Simple algebra shows that they are related by the following thermodynamic relation:

$$F_n(\beta) = U_n(\beta) - \frac{S_n(\beta)}{\beta}$$

Also, the internal energy and entropy can be derived from the free energy as follows:

$$\frac{\partial}{\partial\beta}(\beta F_n(\beta)) = U_n(\beta)$$

$$\beta^2 \frac{\partial}{\partial\beta}(F_n(\beta)) = S_n(\beta)$$

6 Gärtner-Ellis theorem

We finish this lecture by considering a powerful extension of large deviation property to sequences of dependent random variables. This is due to Gärtner and Ellis, who generalized Cramer's theorem for the sum of i.i.d. random variables. This extension will be relevant for future applications in statistical mechanics.

Let $(Y_n)_{n\geq 1}$ be a sequence of random variables which are defined on a probability space $(\mathbb{R}^d, \mathcal{F}, \mathbb{P}_n)$. We define the cumulant generating functions

$$c_n(t) = \frac{1}{n}\log \mathbb{E}_n(e^{<t,Y_n>}) \qquad n = 1, 2, \dots \qquad t \in \mathbb{R}^d \tag{6.1}$$

where $\mathbb{E}_n$ denotes expectation with respect to $\mathbb{P}_n$ and $\langle\cdot,\cdot\rangle$ denotes the standard Euclidean inner product on $\mathbb{R}^d$.

Theorem 6.2. *The following hypothesis are assumed to hold:*

1. *each function $c_n(t)$ is finite for all $t \in \mathbb{R}^d$*
2. *the limit $c(t) = \lim_{n\to\infty} c_n(t)$ exists for all $t \in \mathbb{R}^d$ and is finite*

We call $c(t)$ the free energy function of $(Y_n)_{n\geq 1}$. Let Q_n be the distribution of Y_n/n on $\mathbb{R}^d$ and define the Legendre transform

$$I(y) = \sup_{t\in\mathbb{R}^d}\{\langle t, y\rangle - c(t)\} \qquad \text{for } y \in \mathbb{R}^d \tag{6.3}$$

Then the following holds:

(a) The upper large deviation bound is valid:

$$\limsup_{n\to\infty} \frac{1}{n}\log Q_n(C) \leq -\inf_{y\in C} I(y) \qquad \forall C \subseteq \mathbb{R}^d \text{closed} \tag{6.4}$$

(b) Assume in addition that $c(t)$ is differentiable $\forall t \in \mathbb{R}^d$. Then the lower bound is valid:

$$\liminf_{n\to\infty} \frac{1}{n}\log Q_n(O) \leq -\inf_{y\in O} I(y) \qquad \forall O \subseteq \mathbb{R}^d \text{open} \tag{6.5}$$

Hence, if $c(t)$ is differentiable for all t, then Q_n satisfies a large deviation property with rate n and rate function $I(y)$.

Remark 1 We may heuristically express the large deviation property by the formal notation

$$Q_n(Y_n/n \in dy) \approx e^{-nI(y)}dy \tag{6.6}$$

which is valid to exponential order for large n.

Remark 2 The proof of the theorem can be found in [Ellis, Chap. VII] or in [Den Hollander, Chap. V] in a slightly different form. The theorem is proved by suitably generalizing the proof of Cramér's theorem. The assumption that $c(t)$ is finite can be relaxed to $c(t) \in [-\infty, \infty]$.

Remark 3 Like the rate function of the Cramér's theorem, the rate function $I(y)$ in Eq. (6.3) is a convex, lower semicountinuous functions and it has compact level sets. The infimum of $I(y)$ is 0 and the infimum is attained at some point. However, in contrast to the rate function of the Cramér's theorem, the minimum point need not to be unique (see example below). Whether or not $I(y)$ attains its infimum at a unique point has interesting consequences which will be explored later in this course (phase transitions).

Remark 3 One of the hypothesis of the theorem is the differentiability of the free energy function $c(t)$. An interesting problem is to investigate the existence of large deviation principles when this condition is violated. Unfortunately the situation is complicated and a general theory does not exists. In the example below, the free energy function is not differentiable at a single point and the lower bound fails for a whole class of open sets. Nevertheless, a large deviation property holds with a non-convex rate function. Let Y_n have distribution $\mathbb{P}(Y_n = n) = \mathbb{P}(Y_n = -n) = 1/2$. Then

$$c(t) = \lim_{n\to\infty} \frac{1}{n} \log\left(\frac{1}{2}(e^{nt} + e^{-nt})\right) = |t| \, .$$

Thus $c(t)$ is not differentiable at $t = 0$ and the Gärtner-Ellis theorem is not applicable. We also have from Eq.(6.3)

$$I(y) = \sup_{t\in\mathbb{R}}\{ty - |t|\} = \begin{cases} 0 & \text{if } |y| \leq 1 \\ \infty & \text{if } |y| > 1 \end{cases}$$

$I(y)$ attains its infimum not at a single point but in the whole interval $[-1, 1]$. On the other hand it is easy to check that the distributions Q_n of Y_n/n have a large deviation property with rate function

$$J(y) = \begin{cases} 0 & \text{if } |y| = 1 \\ \infty & \text{if } |y| \neq 1 \end{cases}$$

The function I is the largest convex function less than or equal to the rate function J.

7 Exercises

Exercise 1 Let $(X_i)_{i\geq 1}$ be i.i.d. random variables taking value in $\mathbb{R}$ satisfying

$$\varphi(t) = \mathbb{E}(e^{tX_1}) < \infty \qquad \forall t \in \mathbb{R}$$

Let $Y_n = \sum_{i=1}^{n} X_i$. By applying Th. (6.2) show that Y_n/n satisfies a LDP with rate funtion

$$I(y) = \sup_{t\in\mathbb{R}}\{ty - \log\varphi(t)\}$$

Exercise 2 Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. random variables taking values in a finite set $\Omega = \{x_1, x_2, \ldots, x_r\}$ and having marginal law $\rho = (\rho_1, \ldots, \rho_r)$. Deduce Sanov's theorem from the Gartner-Ellis theorem by considering $Y_n = nL_n$ where L_n denotes the empirical measure. Show that the the rate function is given by the relative entropy.

Exercise 3 Show that the relative entropy $I_\rho(\gamma)$ measure the discrepancy between γ and ρ, in the sense that $I_\rho(\gamma) \geq 0$ and $I_\rho(\gamma) = 0$ if and only if $\gamma = \rho$. Thus $I_\rho(\gamma) \geq 0$ attaind its infimum of 0 over $\mathcal{M}$ at the unique measure $\gamma = \rho$. Show also that $I_\rho(\gamma)$ is strictly convex.

(Hint: for $x \geq 0$ the graph of the strictly convex function $x\log x$ has the tangent line $y = x - 1$ at $x = 1$. Hence $x\log x \geq x - 1$ with equality if and only if $x = 1$.)

Exercise 4 Let X_i be a stationary Markov chain taking value in Ω and having transition matrix P. By applying the Gärtner-Ellis theorem show that the pair empirical measure $L_n^2 = \frac{1}{n}\sum_i \delta_{X_i, X_{i+1}}$ satisfy a large deviation principle with rate function

$$I(\nu) = \sum_{i,j} \nu_{i,j} \log\left(\frac{\nu_{i,j}}{\bar{\nu}_i P_{i,j}}\right)$$

where $\bar{\nu}_i = \sum_j \nu_{i,j}$.

References

[1] R. Ellis *Entropy, Large Deviations and Statistical Mechanics*, Springer-Verlag (1985).

[2] F. Den Hollander *Large Deviations*, Fields Institute monographs (2000)

[3] K. Huang *Statistical mechanics*, John Wiley & Sons (1987)