

# STATISTICAL ANALYSIS OF LARGE SYSTEMS BASED ON THE EQUIVALENCE OF ENSEMBLES

## Contents

<b>1 General method</b>	<b>1</b>
<b>2 Examples and Applications</b>	<b>3</b>
2.1 Coin tossing and independent spins . . . . .	3
2.2 Information theory . . . . .	5
2.3 Gases . . . . .	6
2.4 Ising models . . . . .	6
2.5 Lattice gases . . . . .	6

## 1 General method

What we have learned so far can be summarized and generalized as follows.

**Microcanonical distribution** Given a phase space  $\Gamma$  of points  $x = (x_1, \dots, x_N)$  with volume element  $d\Gamma$ , one wishes to study the statistical regularities ( $N$  large) of the set of points which satisfies the constraints

$$\mathbf{E} \leq \mathbf{H}(x) \leq \mathbf{E} + \delta\mathbf{E} \tag{1}$$

where  $\mathbf{H}(x) = (H_0(x), H_1(x), \dots)$  is a vector of  $J$  variables ( $J \ll N$ ), each of which grows like  $N$  and  $\delta\mathbf{E} \ll \mathbf{E}$ . In the applications to a gas,  $H_0 = H$  is the Hamiltonian,  $H_1 = N$  the number of particles, etc.. This is the **microcanonical** setting of the problem.

**Thermodynamic limit** Let  $\Gamma_{\mathbf{E}, \delta\mathbf{E}}$  be the set of points that satisfy the constraints and denote by  $|\Gamma_{\mathbf{E}, \delta\mathbf{E}}|$  its volume. The **TL** is so defined:

$$\text{allow } N \rightarrow \infty \text{ in such a way that } \frac{\mathbf{E}}{N} \rightarrow \mathbf{e}, \frac{\delta\mathbf{E}}{N} \rightarrow \delta\mathbf{e}.$$

Under suitable conditions on the constraints, we have the convergence

$$\frac{1}{N} \log |\Gamma_{\mathbf{E}, \delta\mathbf{E}}| \rightarrow s(\mathbf{e}, \delta\mathbf{e}). \tag{2}$$

After that we take the the “thin phase space” limit  $\delta\mathbf{e} \rightarrow 0$ .

**Question:** Why should anyone care about results in the thermodynamic limit, when every real system is finite? **Answer:** Because real bulk systems have so many particles that they can be considered to exist in the thermodynamic limit. For example, as the system size increases, the free energy density  $f_N = F_N/N$  approaches the limiting value  $f_\infty$ , and typically the difference between  $f_{10^{23}}$  and  $f_\infty$  is much smaller than experimental error. **So we use the limit to provide an extremely good approximation for real systems.** In particular,

$$|\Gamma_{\mathbf{E}, \delta\mathbf{E}}| \approx e^{Ns(\mathbf{e})}, \quad \mathbf{e} = \frac{\mathbf{E}}{N}, \tag{3}$$

is a good asymptotic approximation of the volume of the set of points that satisfy the constraints (assuming the “thin phase space” limit  $\delta\mathbf{e} \ll \mathbf{e}$ ).

We can then use the usual microcanonical rules for finite system to compute the statistics of variables on phase space, e.g., for the mean value of  $Y = Y(x)$  we have

$$\mathbb{E}\{Y\} = \frac{1}{|\Gamma_{\mathbf{E}, \delta \mathbf{E}}|} \frac{\partial}{\partial \mathbf{E}} \int_{\mathbf{H}(x) \leq \mathbf{E}} Y(x) d\Gamma \quad (4)$$

Hereafter we shall also assume that  $s(\mathbf{e})$  is **concave** (stability) and smooth.

**Canonical distribution** For any vector  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)$ , the canonical distribution is defined as

$$\rho_{\boldsymbol{\lambda}}(x) = \frac{1}{Z(\boldsymbol{\lambda})} e^{-\boldsymbol{\lambda} \cdot \mathbf{H}(x)} d\Gamma, \quad (5)$$

$$Z_N(\boldsymbol{\lambda}) = \int_{\Gamma} e^{-\boldsymbol{\lambda} \cdot \mathbf{H}(x)} d\Gamma. \quad (6)$$

$Z_N(\boldsymbol{\lambda})$  is called the **partition function** and  $\Psi_N(\boldsymbol{\lambda}) = \log Z_N(\boldsymbol{\lambda})$  the **Massieu function**.

*Useful formulas.* Mean:

$$\begin{aligned} \langle \mathbf{H} \rangle_{\boldsymbol{\lambda}} &= \frac{1}{Z_N(\boldsymbol{\lambda})} \int_{\Gamma} \mathbf{H}(x) e^{-\boldsymbol{\lambda} \cdot \mathbf{H}(x)} d\Gamma = \frac{1}{Z_N(\boldsymbol{\lambda})} \left[ -\frac{\partial}{\partial \boldsymbol{\lambda}} \int_{\Gamma} e^{-\boldsymbol{\lambda} \cdot \mathbf{H}(x)} d\Gamma \right] \\ &= -\frac{1}{Z_N(\boldsymbol{\lambda})} \frac{\partial Z_N(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = -\frac{\partial \log Z_N(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = -\frac{\partial \Psi_N(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \end{aligned} \quad (7)$$

Correlation matrix:

$$\begin{aligned} \langle H_i H_j \rangle_{\boldsymbol{\lambda}} &= \frac{1}{Z_N(\boldsymbol{\lambda})} \int_{\Gamma} H_i(x) H_j(x) e^{-\boldsymbol{\lambda} \cdot \mathbf{H}(x)} d\Gamma = \frac{1}{Z_N(\boldsymbol{\lambda})} \frac{\partial^2 Z_N(\boldsymbol{\lambda})}{\partial \lambda_i \partial \lambda_j} \\ &= \frac{\partial}{\partial \lambda_i} \frac{\partial \log Z_N(\boldsymbol{\lambda})}{\partial \lambda_j} + \frac{1}{Z_N(\boldsymbol{\lambda})^2} \frac{\partial Z_N(\boldsymbol{\lambda})}{\partial \lambda_i} \frac{\partial Z_N(\boldsymbol{\lambda})}{\partial \lambda_j} \\ &= \frac{\partial^2 \Psi_N(\boldsymbol{\lambda})}{\partial \lambda_i \partial \lambda_j} + \frac{\partial \Psi_N(\boldsymbol{\lambda})}{\partial \lambda_i} \frac{\partial \Psi_N(\boldsymbol{\lambda})}{\partial \lambda_j} \\ &= -\frac{\partial \langle H_j \rangle_{\boldsymbol{\lambda}}}{\partial \lambda_i} + \langle H_i \rangle_{\boldsymbol{\lambda}} \langle H_j \rangle_{\boldsymbol{\lambda}} \end{aligned} \quad (8)$$

Covariance matrix:

$$\begin{aligned} C_{ij}^{\boldsymbol{\lambda}} &= \langle (H_i - \langle H_i \rangle) (H_j - \langle H_j \rangle) \rangle_{\boldsymbol{\lambda}} = \langle H_i H_j \rangle_{\boldsymbol{\lambda}} - \langle H_i \rangle_{\boldsymbol{\lambda}} \langle H_j \rangle_{\boldsymbol{\lambda}} \\ &= -\frac{\partial \langle H_j \rangle_{\boldsymbol{\lambda}}}{\partial \lambda_i} \end{aligned} \quad (9)$$

**Equivalence** A way to compute  $\mathbb{E}\{Y\}$  equivalents to (4) is by means of the canonical distribution with suitable  $\boldsymbol{\lambda}$ . More precisely,

(a)  $\boldsymbol{\lambda}$  has to be such that the canonical mean of  $\mathbf{H}$  is equal to micro-canonical constrained value  $\mathbf{E}$ , i.e.,  $\boldsymbol{\lambda}$  has to be the solution of the equation

$$\mathbf{E} = -\frac{\partial \Psi_N(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}. \quad (10)$$

(b) The solution of (10) has to be unique.

These conditions are met if the **TL**

$$\lim_{N \rightarrow \infty} \frac{\Psi_N(\boldsymbol{\lambda})}{N} = \psi(\boldsymbol{\lambda}) \quad (11)$$

exists and  $\psi(\boldsymbol{\lambda})$  is a “nice” function, that is, smooth and **concave**. Under these conditions we have:

(1) **Concentration of  $\mathbf{H}/N$  around its asymptotic mean**

$$-\frac{\partial \psi(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \mathbf{e} \quad (12)$$

This is so because the covariance matrix of  $\mathbf{H}$  grows like  $N$  and thus that of  $\mathbf{h}$  tends to 0.

(2) The solution  $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\mathbf{e})$  of

$$\mathbf{e} = -\frac{\partial \psi(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \quad (13)$$

is unique (because of smoothness and concavity). So we the canonical distribution with  $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\mathbf{e})$  is asymptotically equivalent to the micro-canonical one with constraint  $\mathbf{H}(x) = N\mathbf{e}$ .

(3)  $s(\mathbf{e})$  and  $\psi(\boldsymbol{\lambda})$  are one-to-one related by Legendre transformations:

$$\psi(\boldsymbol{\lambda}) = \sup_{\mathbf{e}} [s(\mathbf{e}) - \boldsymbol{\lambda} \cdot \mathbf{e}] \quad (14)$$

$$s(\mathbf{e}) = \inf_{\boldsymbol{\lambda}} [\boldsymbol{\lambda} \cdot \mathbf{e} + \psi(\boldsymbol{\lambda})] \quad (15)$$

This is so (heuristically) because of the asymptotic evaluation of the partition function by Laplace’s method

$$Z(\boldsymbol{\lambda}) \asymp e^{N \inf_{\mathbf{e}} \{s(\mathbf{e}) - \boldsymbol{\lambda} \cdot \mathbf{e}\}}$$

and appealing to concavity of  $s(\mathbf{e})$ .

(4) The minimizer  $\mathbf{e}(\boldsymbol{\lambda})$  of  $s(\mathbf{e}) - \boldsymbol{\lambda} \cdot \mathbf{e}$  is the solution of

$$\boldsymbol{\lambda} = \frac{\partial s(\mathbf{e})}{\partial \mathbf{e}},$$

and thus, by Legendre duality,  $\mathbf{e}(\boldsymbol{\lambda}(\mathbf{e})) = \mathbf{e}$ .

## 2 Examples and Applications

### 2.1 Coin tossing and independent spins

**Coin tossing** The phase space is the set of vectors  $x = (x_1, \dots, x_N)$  with  $x_i = 1$  or 0 (and thus contains  $2^N$  states). Constraint:

$$M(x) = \sum_{i=1}^N x_i = M \quad (\text{some fixed integer})$$

“Volume” of the constraint (number of points):

$$|\Gamma_{M=M}| = \binom{N}{M} \quad (16)$$

## Canonical distribution

$$P_\lambda(x_1, \dots, x_N) = \frac{1}{Z(\lambda)} e^{-\lambda M(x)}$$

Partition function:

$$Z(\lambda) = \sum_{x_1, \dots, x_N} e^{-\lambda \sum_{i=1}^N x_i} = \prod_{i=1}^N \underbrace{\sum_{x_i} e^{-\lambda x_i}}_{z(\lambda)} = \underbrace{(1 + e^{-\lambda})^N}_{z(\lambda)^N}$$

Massieu function

$$\psi(\lambda) = \log(1 + e^{-\lambda}) \quad (17)$$

Equation for  $\lambda$ :

$$p = \frac{M}{N} = -\psi'(\lambda) = \frac{e^{-\lambda}}{1 + e^{-\lambda}} \quad (18)$$

whence

$$e^{-\lambda} = \frac{p}{1-p}, \quad \lambda = -\log p + \log(1-p) \quad (19)$$

Thus

$$Z(p) = (1 + e^{-\lambda})^N = \frac{1}{(1-p)^N}$$

whence the canonical distribution

$$\begin{aligned} P(x) &= (1-p)^N e^{-\lambda M(x)} = (1-p)^N p^{M(x)} (1-p)^{-M(x)} \\ &= p^{M(x)} (1-p)^{N-M(x)} \end{aligned} \quad (20)$$

Then the distribution of  $M(x)$  is

$$P(M = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

which is, as expected, the binomial distribution with parameters  $N$  and  $p = M/N$ . Note that (20) could be the distribution of  $N$  independent two-level atoms with energy zero and  $\epsilon$  (in suitable units) and total energy  $M\epsilon$ . Then  $p\epsilon$  is the energy per atom.

**Independent spins** By the same means we obtain the distribution of  $N \gg 1$  independent spins. Now the phase space points are  $y = (y_1, \dots, y_N)$  with  $y_i = 1$  or  $-1$  and the global constraint is

$$L(y) = \sum_{i=1}^N y_i = L$$

(e.g., fixed total spin, that is, total magnetization). Since  $y = 2x - 1$ , the canonical distribution now will be obtained from the above just by change of variable  $x = \frac{1}{2}(y + 1)$ , so that

$$M(x) = M\left(\frac{y+1}{2}\right) = \sum_{i=1}^N \left(\frac{y_i+1}{2}\right) = \frac{L(y) + N}{2}.$$

Therefore the canonical distribution equivalent to the micro-canonical one for  $N \gg 1$  is

$$P(y) = p^{\frac{1}{2}(L(y)+N)} (1-p)^{\frac{1}{2}(L(y)-N)} \quad (21)$$

In these case, we have

$$Z(\lambda) = (e^\lambda + e^{-\lambda})^N, \quad \Psi(\lambda) = N \log(e^\lambda + e^{-\lambda}), \quad \frac{M}{N} \equiv p = \frac{e^{-\lambda} - e^\lambda}{e^\lambda + e^{-\lambda}}$$

## 2.2 Information theory

The basic asymptotic problem in coding theory treated by Shannon is that of estimating how many messages have effectively to be considered when coding the output of an information source.

The messages are long sequences  $x = (x_1, \dots, x_N)$  of symbols from a finite alphabet. For simplicity we shall assume that the alphabet contains only 2 symbols, say 1 and 0. One wishes to estimate the number of different messages of length  $N$  that can be written under the constraint that the relative frequencies of the symbols are given, say  $p$  for 1 and  $q = 1 - p$  for 0. <sup>1</sup> This is easily obtained by estimating (16) for  $M = pN$  by using Stirling's formula.

We wish however to obtain the estimate by using the method of equivalence of ensembles, because this method extends naturally to the more general case in which the alphabet contains any finite number of letters and the macroscopic variables are more complicated, say the relative frequencies of pair of consecutive letters (these variables allows for a finer description of a language).

The quantity to be estimated is  $|\Gamma_p|$ , which in the microcanonical ensemble is given by

$$|\Gamma_p| \asymp e^{Ns(p)}$$

To compute  $s(p)$ , we use the canonical ensemble, in particular equation (15). We obtain

$$s(p) = \lambda(p)p + \psi(\lambda(p)).$$

Using (17) and (18), we arrive at

$$s(p) = -p \log p + p \log(1 - p) - \log(1 - p) = -[p \log p + (1 - p) \log(1 - p)]$$

Not surprisingly,  $s(p)$  is related to the rate function for coin tossing that we have computed in a previous lecture.

If the alphabet contains  $J$  letters with relative frequency  $p_k$ ,  $\sum p_k = 1$ , the formula one obtains by the method of equivalence of ensembles is (exercise)

$$s(p_1, \dots, p_J) = - \sum_{k=1}^J p_k \log p_k \quad (22)$$

(Note that now the canonical distribution is parametrized by the vector  $\mathbf{p} = (p_1, \dots, p_J)$ .)

The more difficult case of given relative frequencies  $P_{ik}$  of consecutive pairs of symbols is given by (exercise<sup>2</sup>)

$$s(\{P_{ik}\}) = - \sum_{ik} p_i P_{ik} \log P_{ik} \quad (23)$$

where  $p_i$  are the stationary probabilities of Markov chain defined by the transition matrix  $P_{ik}$ . This is **Shannon** famous formula.

<sup>1</sup>In a natural language, the relative frequencies of the different letters (and combinations of them) occur with varying frequencies. Moreover, there is a characteristic distribution of letters that is roughly the same for almost all samples of a given language.

<sup>2</sup>To prove Shannon formula is a difficult exercise. Yet, the method of the equivalence of ensembles is one of the slickest ones to obtain it.

## 2.3 Gases

$$\mathbf{H} = (H, N) \tag{24}$$

$$\boldsymbol{\lambda} = (\beta, -\beta\mu) \tag{25}$$

Microcanonical constraints:

$$H(x) = \sum = \sum_{i: \mathbf{q}_i \in \Lambda} \frac{\mathbf{p}_i^2}{2m_i} + \frac{1}{2} \sum_{\mathbf{q}_i, \mathbf{q}_j \in \Lambda} V_{ij}(\mathbf{q}_i - \mathbf{q}_j) = E \tag{26}$$

$$N(x) = \sum_{i: \mathbf{q}_i \in \Lambda} 1 = N \tag{27}$$

...

## 2.4 Ising models

...

## 2.5 Lattice gases

...