Non-equilibrium macroscopic dynamics of chains of anharmonic oscillators

Cedric Bernardin
Stefano Olla

Introduction

> *The objective of statistical mechanics is to explain the macroscopic properties of matter on the basis of the behavior of the atom and molecules of which it is composed.*
>
> Oscar R. Lanford III [4]

If we want to make the above definition specific for the *non-equilibrium* statistical mechanics, we can refrase it as

The objective of non-equilibrium statistical mechanics is to explain the macroscopic evolution of matter on the basis of the dynamics of the atom and molecules of which it is composed.

This definition requires to be more specific about what we intend for *macroscopic evolution*. In most non-equilibrium problems it should be specified non only the space scale, but also the time scale of the macroscopic evolution. In fact, as we will see in this book, the same system can behave very differently at different space-time scaling.

This is also related to the choice of which observables we should follow in a macroscopic non-equilibrium evolution of a system. Here we are interested in the evolution of *conserved* quantities of the systems, like energy. A major concern is to distinguish these *slow* observables from the others (*fast*). In fact a deterministic hamiltonian dynamics with n degrees of freedom, may have other integrals of motion than energy. One could be total momentum, but there could be many others, and some systems are completely integrable (like a chain of harmonic oscillators, or the Toda lattice).

We are interested here in systems such that the only integrals of the motion that *survive* to the themodynamic limit $n \to \infty$ are given by energy, eventually momentum (if the infinite system has translation invariant properties), and number of particles. This is actually a very vague statement, since the corresponding infinite system typically has an infinite amount of energy, momentum and mass. So the precise definition of this property is that the only stationary translation invariant measures, enough regular so that are locally absolutely continuous with respect to Lebesgue measure, are given by Gibbs measures associated to the Hamiltonian of the system (see chapter **??** for a precise definition of all these notions). We call this property *ergodicity*, or *ergodicity of the infinite system* to distinguish it from the more classical definition of ergodicity of finite systems.

A well known conterexample to this ergodicity is given by the harmonic chain of oscillators, where the energy of each mode of vibration is conserved. This is a linear system, and this example suggests that ergodicity should be connected with some level of *chaoticity* induced by non-linearities in the interactions. The Toda lattice conterexample (non–linear) shows that the situation is not so simple, and in fact it is not easy to state necessary or sufficient conditions on the interaction between particles, that will imply this ergodicity property. This is one of the major open problem of statistical mechanics, since it is this property that allows, in a macroscopic (space-time) description, to separate an autonomous evolution of energy, momentum and density from the other observables.

In chapter **??** we give a proof that, if a stationary measure has an excheangeable distribution of velocities, then is a Gibbs measure, i.e. ergodicity follows from this excheageability of the velocities.

Since at the moment we are not able to prove ergodicity of the infinite system for any hamiltonian system, we consider stochastic perturbations of these hamiltonian dynamics. These stochastic perturbation exchange momentum between particles. They are local and conserve kinetic energy and eventually total momentum. Consequently the stationary measures for these infinite stochastic dynamics have excheangeable distributions of velocities, i.e. they are ergodic. All this is proven in chapter **??** for a one-dimensional chain of oscillators. One can think that these stochastic perturbations model the effect of the non-linearities, or of some other faster chaotic degree of freedom not included in the hamiltonian dynamics.

For these ergodic systems it is useful to define the concept of *local equilibrium*. This is not a property of a single probability measure on the configuration space of the finite of infinite system, but an asymptotic property of a sequence of probability measures. We define a sequence of probability measure a local equilibrium if *locally* they converge to a Gibbs measure for the infinite system corresponding to a given energy, momentum and density (cf. section 1.4 for a precise definition).

Thermodynamic entropy $S(r, \mathcal{E})$ is defined by formula (A.5.11) from the microcanonical ensemble, as the limit of the logarithm of the volume of the configurations with fixed total energy and volume of the system of n particles. Notice that this is actually a density of entropy, and it should be thought as the thermodynamic entropy of the *macroscopic* system of (macroscopic) length r at equilibrium with given value of energy $\mathcal{E}$. This is obtained from a one dimensional chain of n oscillators, with total length fixed to be nr and total energy fixed at $n\mathcal{E}$, as $n \to \infty$. It comes out from its definition that $S(r, \mathcal{E})$ is a concave function, and that

$$2S\left(\frac{r_1 + r_2}{2}, \frac{\mathcal{E}_1 + \mathcal{E}_2}{2}\right) \geq S(r_1, \mathcal{E}_1) + S(r_2, \mathcal{E}_2) \tag{0.0.1}$$

The quantity on the left in (0.0.1) is the thermodynamic entropy of a system of $2n$ oscillators with total length fixed at $n(r_1 + r_2)$ and energy fixed at $n(\mathcal{E}_1 + \mathcal{E}_2)$. This property of thermodynamic entropy has a classic interpretation. Suppose we have two systems of n oscillators in microcanonical equilibrium, with corresponding parameter $r_1, \mathcal{E}_1$ and $r_2, \mathcal{E}_2$, and we put them in contact fixing the two extremities

add picture here

We have then an inhomogeneous system of $2n$ oscillators with total length fixed to be $n(r_1 + r_2)$ and energy $n(\mathcal{E}_1 + \mathcal{E}_2)$. This system now is in non-equilibrium, and energy and density will evolve on a certain time scale depending on n that we will study later on. *If* this new system will reach equilibrium, then the thermodynamic entropy associated will be larger than the sum of the two initial thermodynamic entropies. This is the classic argument explaining increase with time of thermodynamic entropy, repeated probably thousand times by many authors. In principle it is correct even if we where not taking the limit as $n \to \infty$, i.e. using as entropy just the logarithm of the corresponding volume in the phase space. The usual objection to this argument is that the dynamics of the (*finite*) system may not reach equilibrium (and in fact typically it does not [1]). Here we insist in the inequality (0.0.1) for the thermodynamic entropy S defined in the limit $n \to \infty$, i.e. associated to the macroscopic (*infinite*) system in equilibrium. More precisely the sense of the entropy increase contained in (0.0.1) as to be understood in a *macroscopic* space-time limit, as we will make clear later on.

The above procedure can be generalized to k chains of n oscillators at different equilibrium parameters obtaining

$$S\left(\frac{1}{k}\sum_{i=1}^{k} r_i, \frac{1}{k}\sum_{i=1}^{k} \mathcal{E}_i\right) \geq \sum_{i=1}^{k} S(r_i, \mathcal{E}_i)\frac{1}{k}$$

where as before we identify the right hand side as the entropy of an inhomogeneous system where we have prepared each subsystem in equilibrium at different parameters. Going further we can rescale the (macroscopic) size of the k macroscopic

[1]Even in presence of non-linearities the finite system may have periodicities and other phenomena that prevent the system to converge to equilibrium, this was in fact the main point of the Fermi-Pasta-Ulam numerical experiment. These phenomenas should disappear in the limit as $n \to \infty$, see the clear discussion of this problem in [5]

systems as k^{-1} and obtain with the limit procedure, as $k \to \infty$, the thermodynamic entropy of a system in *local equilibrium*[2] with profiles of energy $\mathcal{E}(y)$ and inverse density $r(y)$ as

$$\int_0^1 S(r(y), \mathcal{E}(y))\, dy$$

that by concavity is bigger or equal than $S(\int r(y)dy, \int \mathcal{E}(y)dy)$. This definition of thermodynamic entropy has precise mathematical meaning for a *macroscopic* inhomogeneous system in *local equilibrium*. Homogeneous systems (systems in local equilibrium with flat profile of energy and density) maximize this entropy.

Assume now that we have prove that in a certain macroscopic scale energy and density evolve deterministically, following some profiles densities $r(y,t), \mathcal{E}(y,t)$, solution of certain conservative macroscopic equations (i.e. $\int r(y,t)dy, \int \mathcal{E}(y,t)dy$ are constant in t). It follows that the problem of the macroscopic increase of the entropy in time is related to the evolution of the profiles of density and energy in this macroscopic scale.

More precisely, if one looks at the hyperbolic macroscopic space-time scale, where space and time are rescaled in the same way (see chapter **??**), the momentum $\pi(y,t)$ is also a macroscopic observable, and the internal energy is given by $\mathcal{U}(y,t) = \mathcal{E}(y,t) - \pi(y,t)^2/2$, and the total thermodynamic entropy at time t is given by $\int_0^1 S(r(y,t), \mathcal{U}(y,t))\, dy$. The profiles triplet $r(y,t), \pi(y,t), \mathcal{E}(y,t)$ evolves in time as solution of the Euler non-linear hyperbolic system (**??**). If these solutions are smooth, then we prove in chapter **??**, under the assumption that the infinite dynamics is ergodic, that they describe the macroscopic evolution of the corresponding observables. It turns out that in this smooth regime

$$\partial_t S(r(y,t), \mathcal{U}(y,t)) = 0 \tag{0.0.2}$$

for any y. This means that if shock are not present, thermodynamic entropy remains constant. Correspondingly the sysstem is also macroscopically reversible in time (in the smooth regime Eler equations are time reversible).

[2]see the precise definition in section 1.4, where this notion is intended as a macroscopic asymptotic property.

Chapter 1

Statistical mechanics and thermodynamics of one dimensional chain of oscillators

1.1 The model: grand canonical formalism

We study a system of n anharmonic oscillators. The particles are denoted by $j = 1, \dots n$. We denote with $q_j, j = 1, \dots, n$ their positions, and with p_j the corresponding momentum (which is equal to its velocity since we assume that all particles have mass 1). We consider first the system attached to a *wall*, and we set $q_0 = 0, p_0 = 0$. Between each pair of consecutive particles $(i, i+1)$ there is an anharmonic spring described by its potential energy $V(q_{i+1} - q_i)$. We assume V is a positive smooth function such that $V(r) \to +\infty$ as $|r| \to \infty$ and such that

$$Z(\lambda, \beta) := \int e^{-\beta V(r) + \lambda r} dr < +\infty \tag{1.1.1}$$

for all $\beta > 0$ and all $\lambda \in \mathbb{R}$. Let a be the equilibrium interparticle spacing, where V attains its minimum that we assume is 0: $V(a) = 0$. It is convenient to work with interparticle distance as coordinates, rather than absolute particle position, so we define $\{r_j = q_j - q_{j-1} - a, \ j = 1, \dots, n\}$.

The configuration of the system is given by $\{p_j, r_j, j = 1, \dots, n\} \in \mathbb{R}^{2n}$, and energy function (Hamiltonian) defined on each configuration is given by

$$\mathcal{H} = \sum_{j=1}^{n} \mathcal{E}_j$$

where

$$\mathcal{E}_j = \frac{1}{2}p_j^2 + V(r_j), \quad j = 1, \dots, n$$

is the energy of each oscillator. This choice is a bit arbitrary, because we associate the potential energy of the bond $V(r_j)$ to the particle j. Different choices can be made, but this one is notationally convenient.

At the other end of the chain we apply a constant force $\tau \in \mathbb{R}$ on the particle n (tension). The position of the particle n is given by $q_n = \sum_{j=1}^n r_j$. We consider the Hamiltonian dynamics:

$$\begin{aligned} \dot{r}_j(t) &= p_j(t) - p_{j-1}(t), \qquad j = 1, \dots, n, \\ \dot{p}_j(t) &= V'(r_{j+1}(t)) - V'(r_j(t)), \qquad j = 1, \dots, n-1, \\ \dot{p}_n(t) &= \tau - V'(r_n(t)), \end{aligned} \tag{1.1.2}$$

It is easy to see that, for any $\beta > 0$, the grand canonical measure $\mu^{gc}_{\tau,\beta}$ defined by

$$d\mu^{n,gc}_{\tau,\beta} = \prod_{j=1}^n \frac{e^{-\beta(\mathcal{E}_j - \tau r_j)}}{\sqrt{2\pi\beta^{-1}}Z(\beta\tau, \beta)} dr_j dp_j \tag{1.1.3}$$

is stationary for this dynamics. The distribution $\mu^{n,gc}_{\tau,\beta}$ is called grand canonical Gibbs measure at temperature $T = \beta^{-1}$ and tension (or pressure) τ. Notice that $\{r_1, \dots, r_n, p_1, \dots, p_n\}$ are independently distributed under this probability measure.

Let us now fix a reference measure μ^{gc}_{0,β_0}, corresponding to a given temperature $T_0 = \beta_0^{-1}$ and with external force $\tau = 0$. If we consider the random vector $\mathbf{X}_j = (r_j, \mathcal{E}_j) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$, then applying the result of appendix A, we obtain that the sum $\frac{1}{n}\sum_1^n \mathbf{X}_j$ has a large deviation function given by

$$I(r, \mathcal{E}) = \sup_{\lambda, \eta < \beta_0} \{\lambda r + \eta\mathcal{E} - \Lambda(\lambda, \eta)\}$$

with

$$\Lambda(\lambda, \eta) = \log\left(\frac{Z(\lambda, \beta_0 - \eta)}{Z(0, \beta_0)}\sqrt{\frac{\beta_0}{\beta_0 - \eta}}\right)$$

and $\Lambda(\lambda, \eta) = +\infty$ if $\eta > \beta_0$.

Then we obtain

$$I(r, \mathcal{E}) = \sup_{\lambda, \beta > 0}\left\{\lambda r - \beta\mathcal{E} - \log\left(\sqrt{2\pi\beta^{-1}}Z(\lambda, \beta)\right)\right\} + \beta_0\mathcal{E} + \log\left(\sqrt{2\pi\beta_0^{-1}}Z(0, \beta_0)\right)$$

The function

$$S(r,u) = \inf_{\lambda,\beta>0}\left\{-\lambda r + \beta u + \log Z(\lambda,\beta) + \frac{1}{2}\log\frac{2\pi}{\beta}\right\} \tag{1.1.4}$$

is called *thermodynamic entropy*.

So we have obtained that

$$I(r,\mathcal{E}) = -S(r,\mathcal{E}) + \beta_0\mathcal{E} + \log Z(0,\beta_0) + \frac{1}{2}\log\frac{2\pi}{\beta_0}$$

The density of the distribution of $\frac{1}{n}\sum_{j=1}^n \mathbf{X}_j$ under $\mu_{0,\beta_0}^{n,gc}$ is given by

$$\begin{aligned}
&f_n(r,\mathcal{E})\\
&= \int_{\mathbb{R}^{2n}} \frac{e^{-\beta_0\sum_j \mathcal{E}_j}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\delta\left(\frac{1}{n}\sum_{j=1}^n \mathcal{E}_j - \mathcal{E}; \frac{1}{n}\sum_{j=1}^n r_j - r\right)\prod_j dr_j dp_j\\
&= \frac{e^{-n\beta_0\mathcal{E}}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\int_{\mathbb{R}^{2n}} \delta\left(\frac{1}{n}\sum_{j=1}^n \mathcal{E}_j - \mathcal{E}; \frac{1}{n}\sum_{j=1}^n r_j - r\right)\prod_j dr_j dp_j\\
&\qquad\qquad = \frac{e^{-n\beta_0\mathcal{E}}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\Gamma_n(r,\mathcal{E}).
\end{aligned}$$

Observe that $\Gamma_n(r,\mathcal{E})$ defined by the equation above, does not depend on β_0. It is clearly sub–multiplicative

$$\Gamma_{n+m}(r,\mathcal{E}) \geq \Gamma_n(r,\mathcal{E})\Gamma_m(r,\mathcal{E})$$

Consequently it exists the limit

$$\lim_{n\to\infty}\frac{1}{n}\log\Gamma_n(r,\mathcal{E}) = S(r,\mathcal{E})\ . \tag{1.1.5}$$

that applying (A.5.8) we identify as the thermodynamic entropy defined by (1.1.4). This is the fundamental relation that connects the microscopic system to its thermodynamic macroscopic description.

We can now define the other thermodynamic quantities from the entropy definition (1.1.4). From equation (1.1.4) we have

$$\lambda(r,u) = -\frac{\partial S(r,u)}{\partial r}, \qquad \beta(r,u) = \frac{\partial S(r,u)}{\partial u} \tag{1.1.6}$$

and we will always define the tension as $\tau(r,u) = \lambda(r,u)/\beta(r,u)$.

$$
\begin{aligned}
r(\lambda,\beta) &= \frac{\partial \log Z(\lambda,\beta)}{\partial \lambda} = \int r\, \frac{e^{\lambda r - \beta V(r)}}{Z(\lambda,\beta)}\, dr = \int r_j \; d\mu^{gc}_{\tau,0,\beta} \\
u(\lambda,\beta) &= -\frac{\partial \log\left(Z(\lambda,\beta)\sqrt{2\pi/\beta}\right)}{\partial \beta} = \int V(r) \frac{e^{\lambda r - \beta V(r)}}{Z(\lambda,\beta)}\, dr + \frac{1}{2\beta} = \int \mathcal{E}_j \; d\mu^{gc}_{\tau,0,\beta}
\end{aligned}
\tag{1.1.7}
$$

In thermodynamics is used the following terminology

- r is the *length*,
- u is the *internal energy*,
- $T = \beta^{-1}$ is the *temperature*,
- $\tau = \beta^{-1}\lambda$ is the pressure or the *tension* [6].

The above are the basic thermodynamics coordinates. Usually one choose two of these as independent variables, and express the others as functions of these.

Computing the total differential of $S(r,u)$ we have

$$
dS = -\beta\tau dr + \beta du = \frac{dQ}{T} \tag{1.1.8}
$$

where dQ is the (non-exact) differential

$$
dQ = -\tau dr + du \tag{1.1.9}
$$

and represents the energy gained (or lost) by the system under the infinitesimal change dr, du. Equation (1.1.9) is the differential form of the *first law of thermodynamics*, while (1.1.8) is the one corresponding to the *second law of thermodynamics*.

1.2 Microcanonical measure

Instead of applying a force (tension) to one side of the chain, one can fix the particle n to another wall at distance nr ($q_n = \sum_{j=1}^n r_j = nr$ and $p_n = \dot{p}_n = 0$). The corresponding dynamics is then

$$
\begin{aligned}
&\dot{r}_j(t) = p_j(t) - p_{j-1}(t), \qquad j = 1,\dots,n-1, \\
&\dot{p}_j(t) = V'(r_{j+1}(t)) - V'(r_j(t)), \qquad j = 1,\dots,n-1, \\
&r_n(t) = nr - \sum_{j=1}^{n-1} r_j(t)\ .
\end{aligned}
\tag{1.2.1}
$$

The dynamics now is conserving the total energy $\mathcal{H} = \sum_j \mathcal{E}_j$ and the total length $\sum_{j=1}^n r_j$. The microcanonical measures $\mu_{r,u}^{n,mc}$ are now stationary for this dynamics. These are defined in the following way:

Consider the vector valued i.i.d. random variables

$$\{\mathbf{X}_j = (r_j, \mathcal{E}_j), j = 1, \dots, n\},$$

distributed by $d\mu_{0,\beta_0}^{n,gc}$. Fix $\mathbf{x} = (r, \mathcal{E})$,and define $\mu_{\mathbf{x}}^{n,mc}$ the conditional distribution of $(r_1, p_1, \dots, r_n, p_n)$ on the manifold $\sum_{j=1}^n \mathbf{X}_j = n\mathbf{x}$. This is defined, for any bounded continuous function $G : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ and $H : \mathbb{R}^{2n} \to \mathbb{R}$, by

$$\begin{aligned}
&\int G(\hat{\mathbf{S}}_n) H(r_1, p_1, \dots, r_n, p_n)\, d\mu_{0,\beta_0}^{n,gc}(r_1, p_1, \dots, r_n, p_n) \\
&= \int_{\mathbb{R}\times\mathbb{R}_+} d\mathbf{x}\; G(\mathbf{x}) f_n(\mathbf{x}) \int H(r_1, p_1, \dots, r_n, p_n)\, d\mu_{\mathbf{x}}^{n,mc}
\end{aligned}$$

where $\hat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$. It is easy to see that $\mu_{\mathbf{x}}^{n,mc}$ does not depend on β_0. We call $\mu_{\mathbf{x}}^{n,mc}$ the *microcanonical measure.*

The multidimensional application of theorem A.5.4 gives the following *equivalence between microcanonical and grandcanonical measure*:

Theorem 1.2.1 *Given $\mathbf{x} = (r, \mathcal{E})$, let*

$$\beta = \beta(r, \mathcal{E}), \qquad \tau = \lambda(r, \mathcal{E})\beta^{-1}.$$

Then for any bounded continuous function $F : \mathbb{R}^{2k} \to \mathbb{R}$ we have

$$\begin{aligned}
\lim_{n\to\infty} &\int F(r_1, p_1, \dots, r_k, p_k)\, d\mu_{\mathbf{x}}^{n,mc}(r_1, p_1, \dots, r_n, p_n) \\
&= \int F(r_1, p_1, \dots, r_k, p_k)\, d\mu_{\tau,\beta}^{gc}(\dots, r_1, p_1, \dots, r_n, p_n, \dots)
\end{aligned}$$

It will be useful later the equivalence of ensembles in the following form:

Theorem 1.2.2 *Under the same conditions of Theorem 1.2.1, assume that*

$$\int F(r_1, p_1, \dots, r_k, p_k)\, d\mu_{\tau,\beta}^{k,gc}(r_1, p_1, \dots, r_k, p_k) = 0.$$

Then

$$\lim_{n\to\infty} \int \left| \frac{1}{n-k} \sum_{i=1}^{n-k} F(r_{i+1}, p_{i+1}, \dots, r_{i+k}, p_{k+i}) \right| d\mu_{\mathbf{x}}^{n,mc} = 0$$

The proof of these two theorems follows the argument used for Theorems A.5.4 and A.5.5.

1.3 Canonical measure

Applying a Langevin's thermostat at temperature $T = \beta^{-1}$ to the particle n (or to any other particle), we obtain a dynamics that has the canonical measure $\mu^{n,c}_{r,\beta}$ as stationary measure:

$$\begin{aligned}
\dot r_j(t) &= p_j(t) - p_{j-1}(t), \qquad j = 1, \dots, n-1, \\
dp_j(t) &= \left(V'(r_{j+1}(t)) - V'(r_j(t))\right) dt \\
&\qquad\qquad + \delta_{j,n-1}\left(-p_j(t)dt + \sqrt{\beta}dw(t)\right), \qquad j = 1, \dots, n-1, \\
r_n(t) &= nr - \sum_{j=1}^{n-1} r_j(t) \ .
\end{aligned} \tag{1.3.1}$$

This is defined as follows:

If we condition the grand canonical measure $\mu^{n,gc}_{0,0,\beta}$ on the total length of the chain equal to $L = nr = \sum_j r_j = q_n - q_0$, we obtain the canonical measure that we denote by $\mu^{n,c}_{r,\beta}$. We can formally write

$$d\mu^{n,c}_{r,\beta} = \prod_j \frac{e^{-\beta p_j^2/2}}{\sqrt{2\pi\beta^{-1}}} dp_j \otimes \frac{e^{-\beta \sum_j V(r_j)}}{Z_{n,c}(r,\beta)} \delta\left(\sum_j r_j = nr\right) \prod_j dr_j$$

where $Z_{n,c}(r,\beta)$ is the normalization constant (canonical partition function).

Similar statements as theorems 1.2.1 and 1.2.2 holds, $\mu^{n,c}_{r,\beta}$ converging to the grand-canonical measure $\mu^{n,gc}_{\tau,\beta}$, with τ given by the thermodynamic relations (1.1.6).

Other boundary conditions can be made, like applying a tension τ and a Langevin thermostat at temperature β^{-1} to the n particle, obtaining a system with $\mu^{gc}_{\tau,\beta}$ as stationary measure.

1.4 Local equilibrium, local Gibbs measures

The Gibbs distributions defined in the above sections are also called equilibrium distributions for the dynamics. Studying the non-equilibrium behaviour we need the concept of local equilibrium distributions. These are probability distributions that have some asymptotic properties when the system became large ($n \to \infty$), vaguely speaking *locally* they look like Gibbs measure. We need a precise mathematical definition, that will be useful later for proving macroscopic behaviour of the system.

Definition 1.4.1 *Given two functions $\beta(y) > 0, \tau(y), y \in [0,1]$, we say that the sequence of probability measures μ_n on $\mathbb{R}^{2n}$ has the local equilibrium property (with respect to the profiles $\beta(\cdot), \tau(\cdot)$) if for any $k > 0$ and $y \in (0,1)$,*

$$\lim_{n\to\infty} \mu_n\big|_{([ny],[ny]+k)} = \mu^{k,gc}_{\tau(y),\beta(y)} \tag{1.4.1}$$

Sometimes we will need some weaker definition of local equilibrium (for example relaxing the pointwise convergence in y). It is important here to understand that *local equilibrium* is a property of a *sequence* of probability measures.

The most simple example of local equilibrium sequence is given by the local Gibbs measures:

$$\prod_{j=1}^{n} \frac{e^{-\beta(j/n)(\mathcal{E}_j - \tau(j/n) r_j)}}{\sqrt{2\pi\beta(j/n)^{-1}}Z(\beta(j/n)\tau(j/n), \beta(j/n))} dr_j dp_j = g^n_{\tau(\cdot),\beta(\cdot)} \prod_{j=1}^{n} dr_j dp_j \tag{1.4.2}$$

Of course are local equilibrium sequence also *small order* perturbation of this sequence like

$$e^{\sum_j F_j(r_{j-h}, p_j - h, \dots, r_{j+h}, p_j + h)/n} g^n_{\tau(\cdot),\beta(\cdot)} \prod_{j=1}^{n} dr_j dp_j \tag{1.4.3}$$

where F_j are local functions.

To a local equilibrium sequence we can associate a thermodynamic entropy, defined as

$$S(r(\cdot), u(\cdot)) = \int_0^1 S(r(y), u(y))\, dy \tag{1.4.4}$$

where $r(y), u(y)$ are computed from $\tau(y), \beta(y)$ using (1.1.7).

Appendix A

Large Deviations

A.1 Introduction

As Dembo and Zeitouni point out in the introduction to their monograph on the subject [1], there is no real theory of large deviations, but a variety of tools that allow analysis of small probability.

To give an idea of what we mean with *large deviations*, let us consider a sequence of independent identical distributed real valued random variables $X_1, X_2, \ldots, X_n$ such that $\mathbb{E}(X_j^2) = 1$, and $\mathbb{E}(X_j) = 0$. Let $\hat{S}_n = \frac{1}{n}\sum_i X_i$ the empirical sum. The weak law of large numbers says that for any $\delta > 0$,

$$\mathbb{P}(|\hat{S}_n| \geq \delta) \underset{n\to\infty}{\longrightarrow} 0 \tag{A.1.1}$$

The central limit theorem is a refinement that says

$$\mathbb{P}(\sqrt{n}\hat{S}_n \in [a,b]) \underset{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx \ . \tag{A.1.2}$$

In the case $X_j \sim \mathcal{N}(0,1)$, we have $\hat{S}_n \sim N(0,1/n)$, and we can compute explicitly

$$\mathbb{P}(|\hat{S}_n| \geq \delta) \ = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-x^2/2}dx \ .$$

therefore (**exercise**)

$$\frac{1}{n}\log \mathbb{P}(|\hat{S}_n| \geq \delta) \underset{n\to\infty}{\longrightarrow} -\frac{\delta^2}{2} \tag{A.1.3}$$

Equation (A.1.3) is an example of a large deviation statement.

A.2 Cramér's Theorem in $\mathbb{R}$

Let $\{X_n\}$ a sequence of i.i.d. random variables on $\mathbb{R}$ with common probability distribution $\alpha(dx)$. We define the moment generating function

$$M(\lambda) = \mathbb{E}\left[e^{\lambda X_1}\right] \tag{A.2.1}$$

and let us assume that there exists $\lambda^* > 0$ such that $M(\lambda) < \infty$ if $|\lambda| < \lambda^*$. Notice that, since $|x| \leq \lambda^{-1}(e^{\lambda x} + e^{-\lambda x})$ for any $\lambda > 0$, this condition implies that X_1 is integrable and we denote $m = \mathbb{E}(X_1) \in \mathbb{R}$. It is easy to see that $m = M'(0)$. We are interested in the *logarithmic moment generating function*

$$\mathcal{Z}(\lambda) = \log \mathbb{E}\left[e^{\lambda X_1}\right] \tag{A.2.2}$$

By Jensen's inequality, we have $\mathcal{Z}(\lambda) \geq \lambda m > -\infty$. Let $\mathcal{D}_{\mathcal{Z}} = \{\lambda : \mathcal{Z}(\lambda) < +\infty\}$. Under our hypothesis, $0 \in \mathcal{D}_{\mathcal{Z}}^o$ (the interior of $\mathcal{D}_{\mathcal{Z}}$).

Lemma A.2.1 *1. $\mathcal{Z}(\cdot)$ is convex.*

2. $\mathcal{Z}(\cdot)$ is continuously differentiable in $\mathcal{D}_{\mathcal{Z}}^o$ and

$$\mathcal{Z}'(\lambda) = \frac{\mathbb{E}(X_1 e^{\lambda X_1})}{M(\lambda)} \qquad \lambda \in \mathcal{D}_{\mathcal{Z}}^o.$$

Proof:

1. For any $\alpha \in [0,1]$, it follows by Hölder inequality

$$\mathbb{E}(e^{(\alpha\lambda_1 + (1-\alpha)\lambda_2)X_1}) \leq M(\lambda_1)^\alpha M(\lambda_2)^{1-\alpha}$$

and consequently

$$\mathcal{Z}(\alpha\lambda_1 + (1-\alpha)\lambda_2) \leq \alpha\mathcal{Z}(\lambda_1) + (1-\alpha)\mathcal{Z}(\lambda_2)$$

2. The function $f_\epsilon(x) = (e^{(\lambda+\epsilon)x} - e^{\lambda x})/\epsilon$ converges pointwise to $xe^{\lambda x}$, and $|f_\epsilon(x)| \leq e^{\lambda x}(e^{\delta|x|} - 1)/\delta \leq e^{\lambda x}(e^{\delta x} + e^{-\delta x})/\delta = h(x)$, for every $|\epsilon| \leq \delta$. For any $\lambda \in \mathcal{D}_{\mathcal{Z}}^o$, there exists a $\delta > 0$ small enough such that $\mathbb{E}(h(X_1)) \leq M(\lambda+\delta) + M(\lambda-\delta) < +\infty$. Then the result follows by the dominated convergence theorem.

□

Using the same argument one can prove that $\mathcal{Z}(\cdot) \in \mathcal{C}^\infty(\mathcal{D}^o_{\mathcal{Z}})$. Computing the second derivative we obtain

$$\mathcal{Z}''(\lambda) = \frac{\mathbb{E}(X_1^2 e^{\lambda X_1})}{M(\lambda)} - \left(\frac{\mathbb{E}(X_1 e^{\lambda X_1})}{M(\lambda)}\right)^2 \geq 0$$

Observe that $\mathcal{Z}''(0) = \mathrm{Var}(X_1)$. To avoid the trivial deterministic case, we assume that $\mathrm{Var}(X_1) > 0$. It follows that $\mathcal{Z}''(\lambda) > 0$ for any $\lambda \in \mathcal{D}^o_{\mathcal{Z}}$, i.e. $\mathcal{Z}(\cdot)$ is strictly convex.

We define the rate function as the Fenchel-Legendre transform of $\mathcal{Z}$

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \mathcal{Z}(\lambda)\} \tag{A.2.3}$$

It is immediate to see that I is convex (as supremum of linear functions), hence continuous, and that $I(x) \geq 0$. Furthermore we have that $I(m) = 0$. In fact by Jensen's inequality $M(\lambda) \geq e^{\lambda m}$ for any $\lambda \in \mathbb{R}$, so that

$$\lambda m - \mathcal{Z}(\lambda) \leq 0$$

and it is equal to 0 for $\lambda = 0$. We conclude that $I(m) = 0$.

Consequently m is a minimum of the convex positive function $I(x)$. It follows that $I(x)$ is nondecreasing for $x \geq m$ and nonincreasing for $x \leq m$.

Observe that if $x > m$ and $\lambda < 0$

$$\lambda x - \mathcal{Z}(\lambda) \leq \lambda m - \mathcal{Z}(\lambda)$$

that implies

$$I(x) = \sup_{\lambda \geq 0} \{\lambda x - \mathcal{Z}(\lambda)\} \qquad x > m \tag{A.2.4}$$

Similarly one obtains

$$I(x) = \sup_{\lambda \leq 0} \{\lambda x - \mathcal{Z}(\lambda)\} \qquad x < m \tag{A.2.5}$$

Here are other important properties of $I(\cdot)$:

Lemma A.2.2 *$I(x) \to +\infty$ as $|x| \to \infty$, and its level sets are compact.*

Proof: If $x > m \vee 0$, for any positive $\lambda \in \mathcal{D}_{\mathcal{Z}}$,

$$\frac{I(x)}{x} \geq \lambda - \frac{\mathcal{Z}(\lambda)}{x}$$

and $\lim_{x\to+\infty} \mathcal{Z}(\lambda)/x = 0$, so we have $\lim_{x\to+\infty} I(x)/x \geq \lambda$. Consequently its level sets $\{x : I(x) \leq a\}$ are bounded, and closed by continuity of I. $\square$

We want to prove the following theorem:

Theorem A.2.3 (Cramer) *For any set $A \subset \mathbb{R}$,*

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in A) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in A) \leq -\inf_{x\in \bar{A}} I(x)$$

were A^o is the interior of A and $\bar{A}$ is the closure of A.

A.2.1 Properties of Legendre transforms

We denote $\mathcal{D}_I = \{x \in \mathbb{R} : I(x) < \infty\}$.

Lemma A.2.4

The function I is convex in $\mathcal{D}_I$, strictly convex in $\mathcal{D}_I^0$ and $I \in C^\infty(\mathcal{D}_I^o)$. Furthermore for any $\bar{x} \in \mathcal{D}_I^o$ there exists a unique $\bar{\lambda} \in \mathcal{D}_\mathcal{Z}^o$ such that

$$\bar{x} = \mathcal{Z}'(\bar{\lambda})$$

and

$$\bar{\lambda} = I'(\bar{x})$$

Furthermore $I(\bar{x}) = \bar{\lambda}\bar{x} - \mathcal{Z}(\bar{\lambda})$.

We will say that $\bar{x}$ and $\bar{\lambda}$ are in duality if the conditions of the above lemma are satisfied.

Proof: The function $F_x(\lambda) = \lambda x - \mathcal{Z}(\lambda)$ has a maximum for $\lambda = \bar{\lambda}$. This is because it is concave and $\partial_\lambda F_x(\bar{\lambda}) = 0$. It follows that $I(\bar{x}) = \bar{\lambda}\bar{x} - \mathcal{Z}(\bar{\lambda})$ and that $\mathcal{Z}(\lambda) = \sup_x \{\lambda x - I(x)\}$. By the same argument $G_\lambda(x) = \lambda x - I(x)$ is maximized by $\bar{x}$. $\square$

A.2.2 Proof of Cramer's theorem

Upper bound

Let us start with A a closed interval of the form $J_x = [x, +\infty)$ and let $x > m$. Then the exponential Chebycheff's inequality gives for any $\lambda > 0$

$$\mathbb{P}(\hat{S}_n \geq x) \leq e^{-n\lambda x}\mathbb{E}[e^{\sum_{i=1}^n \lambda X_i}] = e^{-n\lambda x}M(\lambda)^n$$

Since $\lambda > 0$ is arbitrary, we can optimize the bound and obtain for $x > m$

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \geq x) \leq -\sup_{\lambda>0}\{\lambda x - \mathcal{Z}(\lambda)\} = -I(x) \tag{A.2.6}$$

where we use (A.2.4) in the last equality. Similarly for $x < m$ we obtain

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \leq x) \leq -\sup_{\lambda<0}\{\lambda x - \mathcal{Z}(\lambda)\} = I(x) \tag{A.2.7}$$

Consider now an arbitrary closed set $C \subset \mathbb{R}$. If $m \in C$, then $\inf_{x\in C} I(x) = 0$ and the upper bound is trivial.

If $m \notin C$, let (x_1, x_2) be the largest open interval around m such that $C \cap (x_1, x_2) = \emptyset$, i.e.

$$C \subseteq (-\infty, x_1] \cup [x_2, +\infty)$$

(if $x_1 = -\infty$ then $C \subseteq [x_2, +\infty)$ and if $x_2 = +\infty$ then $C \subseteq (-\infty, x_1]$). Observe that $x_1 < m < x_2$. Consequently

$$\mathbb{P}(\hat{S}_n \in C) \leq \mathbb{P}(\hat{S}_n \geq x_2) + \mathbb{P}(\hat{S}_n \leq x_1) \leq 2\max\{\mathbb{P}(\hat{S}_n \geq x_2), \mathbb{P}(\hat{S}_n \leq x_1)\}$$

and using (A.2.6) and (A.2.7)

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \in C) \leq -\min\{I(x_2), I(x_1)\} + \frac{1}{n}\log 2 \tag{A.2.8}$$

and from the monotonicity of $I(x)$ on $(-\infty, x_1]$ and $[x_2, +\infty)$

$$\inf_{x\in C} I(x) \geq \min\{I(x_2), I(x_1)\}$$

which concludes the upper bound.

Lower bound

Given an open set G, it is enough to prove that for any $x \in G$

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\hat{S}_n \in G) \geq -I(x)\ .$$

To this end, it is enough to prove that for any x and any $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in (x-\delta, x+\delta)) \geq -I(x) \ .$$

Clearly it is enough to consider x such that $I(x) < \infty$. Assume α has finite support and that $\alpha((-\infty, 0)) > 0, \alpha((0,\infty)) > 0$. Then $\mathcal{Z}$ is finite everywhere ($\mathcal{D}_{\mathcal{Z}} = \mathcal{D}^o_{\mathcal{Z}} = \mathbb{R}$) and there exists a unique $\lambda_0 \in \mathcal{D}^o_{\mathcal{Z}}$ such that

$$I(x) = \lambda_0 x - \mathcal{Z}(\lambda_0) \qquad \text{and} \qquad x = \mathcal{Z}'(\lambda_0)$$

Assuming $x \geq m$, we have that $\lambda_0 \geq 0$.

Let us define the probability law on $\mathbb{R}$

$$\alpha_{\lambda_0}(dy) = \frac{e^{\lambda_0 y}}{M(\lambda_0)} \alpha(dy)$$

Notice that

$$\int y \ \alpha_{\lambda_0}(dy) = \mathcal{Z}'(\lambda_0) = x$$

Noting $A_{n,\delta} = \{(x_1, \dots, x_n) : (x_1 + \dots + x_n)/n \in (x-\delta, x+\delta)\} \subset \mathbb{R}^n$, then for $\delta_1 < \delta$

$$\begin{aligned} \mathbb{P}(\hat{S}_n \in (x-\delta, x+\delta)) &\geq \int_{A_{n,\delta_1}} \alpha(dx_1)\dots\alpha(dx_n) \\ = M(\lambda_0)^n &\int_{A_{n,\delta_1}} e^{-\lambda_0(x_1+\dots+x_n)} \alpha_{\lambda_0}(dx_1)\dots\alpha_{\lambda_0}(dx_n) \\ &\geq M(\lambda_0)^n e^{-n\lambda_0(x+\delta_1)} \int_{A_{n,\delta_1}} \alpha_{\lambda_0}(dx_1)\dots\alpha_{\lambda_0}(dx_n) \end{aligned}$$

By the law of large numbers, for any $\delta_1 > 0$

$$\int_{A_{n,\delta_1}} \alpha_{\lambda_0}(dx_1)\dots\alpha_{\lambda_0}(dx_n) \underset{n\to\infty}{\longrightarrow} 1$$

so that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in (x-\delta, x+\delta)) \geq -\left[\lambda_0(x+\delta_1) - \mathcal{Z}(\lambda_0)\right] = -I(x) - \lambda_0\delta_1$$

Since $\delta_1 < \delta$ is arbitrary, we can let $\delta_1 \to 0$ and it gives the result. If $x < m$, we have $\lambda_0 < 0$, and in the steps of the above we will have $x - \delta_1$ instead of $x + \delta_1$.

Assume now α is of unbounded support with $\alpha((-\infty, x)) > 0, \alpha((x,\infty)) > 0$. Let $A_0 > 0$ be such that $\alpha([-A_0, x)) > 0, \alpha((x, A_0]) > 0$. For any $A \geq A_0$ let β

be the law of X_1 conditioned on $\{|X_1| \leq A\}$, and β_n the law of $\hat{S}_n$ conditioned on $\{|X_i| \leq A,\, i = 1, \ldots, n\}$. Then, for all $n \geq 1$ and every $\delta > 0$,

$$\alpha_n((x-\delta, x+\delta)) = \beta_n((x-\delta, x+\delta))\, \{\alpha([-A, A])\}^n$$

The preceding result applies for β_n. Note

$$\mathcal{Z}^A(\lambda) = \log \int_{-A}^{A} e^{\lambda y} \alpha(dy),$$

and observe that the logarithmic generating function of β is given by

$$\mathcal{Z}^A(\lambda) - \log \alpha([-A, A]) \;\geq\; \mathcal{Z}^A(\lambda)$$

It follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log \alpha_n((x-\delta, x+\delta)) \geq \inf_{\lambda\in\mathbb{R}} \left\{\mathcal{Z}^A(\lambda) - \lambda x\right\} \tag{A.2.9}$$

Let us define

$$I^*(x) = \limsup_{A\to\infty} \left[\sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \mathcal{Z}^A(\lambda)\right\}\right]$$

then we have

$$\liminf_{n\to\infty} \frac{1}{n} \log \alpha_n((x-\delta, x+\delta)) \geq -I^*(x) \tag{A.2.10}$$

Observe that $\mathcal{Z}^A(\cdot)$ is nondecreasing in A, $\mathcal{Z}^A(0) \leq \mathcal{Z}(0) = 0$, and thus $-I^*(x) \leq 0$. Moreover the assumption $\alpha([-A_0, x)) > 0, \alpha((x, A_0]) > 0$ implies

$$\lambda x - \mathcal{Z}^A(\lambda) \geq -\inf \{\log \alpha[-A_0, x)\,,\, \log \alpha(x, A_0]\}$$

Therefore we have $-I^*(x) > -\infty$. The level sets $\{\lambda;\, \mathcal{Z}^A(\lambda) - \lambda x \leq -I^*(x)\}$ are non-empty, compact sets that are nested with respect to A. Then it exists λ_0 in their intersection and $-I(x) \leq \mathcal{Z}(\lambda_0) - \lambda_0 x = \lim_{A\to\infty} \mathcal{Z}^A(\lambda_0) - \lambda_0 x \leq -I^*(x)$. By (A.2.10) we get

$$\liminf_{n\to\infty} \frac{1}{n} \log \alpha_n((x-\delta, x+\delta)) \geq -I(x)$$

The proof for an arbitrary probability law α is completed by observing that if either $\alpha((-\infty, x))$ or $\alpha((x, \infty)) = 0$ then $\mathcal{Z}(\cdot)$ is a monotone function with $\inf_{\lambda\in\mathbb{R}}\{\mathcal{Z}(\lambda) - \lambda x\} = \log \alpha(\{x\})$. Then we have

$$\alpha_n((x-\delta, x+\delta)) \geq \alpha_n(\{x\}) = \alpha(\{x\})^n$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in (x-\delta, x+\delta)) \geq -I(x)$$

$\square$

Remark A.2.5 *Notice that the proof contains the non-asymptotic bound* (A.2.8), *i.e.*

$$\forall n \geq 1, \quad \mathbb{P}(\hat{S}_n \in C) \leq 2e^{-n \inf_{x \in C} I(x)} \tag{A.2.11}$$

also called Chernoff's bound.

Remark A.2.6 *The lower bound was obtained by using the change of variable in conjunction with the law of large numbers for the new probabilities. One can get better bound by using the central limit theorem, and obtain the following corollary*

Corollary A.2.7 *For any $x > m$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \geq x) &= -I(x) \qquad \textit{if } x > m \\ \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \leq x) &= -I(x) \qquad \textit{if } x < m \end{aligned} \tag{A.2.12}$$

Proof: By the central limit theorem

$$\int_{\{x_1+\cdots+x_n/n \in [x, x+\delta_1)\}} \alpha_{\lambda_0}(dx_1) \ldots \alpha_{\lambda_0}(dx_n) \underset{n\to\infty}{\longrightarrow} \frac{1}{2}$$

So in the proof of the lower bound one can substitute $(x - \delta, x + \delta)$ with $[x, x + \delta)$. Since $\mathbb{P}(\hat{S}_n \geq x) \geq \mathbb{P}(\hat{S}_n \in [x, x + \delta))$ one obtains

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \geq x) \geq -I(x)$$

The upper bound follows from the one in theorem A.2.3.

Examples in $\mathbb{R}$

1. Let α be the gaussian distribution

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2}\, dx$$

then $I(x) = (x - m)^2/2\sigma^2$. In this case one can compute it directly, since $\hat{S}_n - nm$ has law $\mathcal{N}(0, \sigma^2/n)$.

2. $\alpha = \frac{1}{2}(\delta_0 + \delta_1)$ (Bernoulli). Then $M(\lambda) = \frac{1}{2}(1 + e^\lambda)$ and

$$I(x) = x \log x + (1 - x) \log(1 - x) + \log 2 \qquad \text{if } x \in [0, 1]$$

and $I(x) = +\infty$ otherwise.

3. For the exponential law $\alpha(dx) = \beta e^{-\beta x} 1_{x \geq 0}\, dx$, we have $M(\lambda) = \beta/(\beta - \lambda)$ for $-\infty < \lambda < \beta$, otherwise $M(\lambda) = +\infty$. Then

$$I(x) = \beta x - 1 - \log(\beta x) \qquad \text{if } x > 0$$

and $I(x) = +\infty$ if $x \leq 0$.

4. If ξ in a random variable with law $\mathcal{N}(0, 1/\beta)$, then ξ^2 has law $\chi^2(1)$, i.e. a gamma law $\Gamma(1/2, \beta/2)$, which has density

$$\frac{\beta^{1/2}}{\sqrt{2}\Gamma(1/2)} x^{-1/2} e^{-\beta x}$$

Its moment generating function is $M(\lambda) = (\beta/(\beta - 2\lambda))^{1/2}$ if $\lambda < \beta/2$, otherwise equal to $+\infty$. The rate function results

$$I(x) = \frac{1}{2}\{\beta x - \log(\beta x) - 1\} \qquad \text{if } x > 0$$

and $+\infty$ if $x < 0$.

A.3 Cramér's Theorem in $\mathbb{R}^d$

Let $\{\mathbf{X}_n\}$ be a sequence of i.i.d. random variables in $\mathbb{R}^d$, and denote $\alpha(d\mathbf{x})$ the common law. We define as before, for $\mathbf{u} \in \mathbb{R}^d$, the moment generating function and its logarithm

$$M(\mathbf{u}) = \int_{\mathbb{R}^d} e^{\mathbf{u} \cdot \mathbf{x}}\, \alpha(d\mathbf{x}), \qquad \mathcal{Z}(\mathbf{u}) = \log M(\mathbf{u}) \tag{A.3.1}$$

and we denote $\mathcal{D}_{\mathcal{Z}} = \{\mathbf{u} \in \mathbb{R}^d : \mathcal{Z}(\mathbf{u}) < +\infty\}$. We assume that $0 \in \mathcal{D}^o_{\mathcal{Z}}$. Then $M(\mathbf{u})$ is smooth in this open set and $\nabla M(0) = \mathbf{m} = \mathbb{E}(\mathbf{X}_1)$.

The rate function is the Legendre-Fenchel transform of $\mathcal{Z}$:

$$I(\mathbf{x}) = \sup_{\mathbf{u} \in \mathbb{R}^d} \{\mathbf{u} \cdot \mathbf{x} - \mathcal{Z}(\mathbf{u})\} \tag{A.3.2}$$

As in the one dimensional case, it follows immediately from the definition that I is non negative, convex, lower semicontinuous and $I(\mathbf{m}) = 0$. Denoting $\mathcal{D}_I = \{\mathbf{x} : I(\mathbf{x}) < +\infty\}$ we have similar properties as in the one dimensional case:

Lemma A.3.1 *$I(\mathbf{x}) \in \mathcal{C}^\infty({\mathcal{D}_I}^o)$, and $\mathbf{m} \in ({\mathcal{D}_I}^o)$. There exists a diffeomorphism between ${\mathcal{D}_I}^o$ and $\mathcal{D}_\lambda^o$ defined by*

$$\mathbf{u}^* = (\nabla \mathcal{Z})(\mathbf{u}), \quad \mathbf{u} = (\nabla I)(\mathbf{u}^*) \tag{A.3.3}$$

and

$$(\nabla^2 \mathcal{Z})(\mathbf{u}) = \left[\nabla^2 I)(\mathbf{u}^*)\right]^{-1} \tag{A.3.4}$$

Theorem A.3.2 *For any Borel set $A \subset \mathbb{R}^d$,*

$$-\inf_{\mathbf{x}\in A^o} I(\mathbf{x}) \le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in A) \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \in A) \le -\inf_{\mathbf{x}\in \bar{A}} I(\mathbf{x})$$

were A^o is the interior of A and $\bar{A}$ is the closure of A.

Proof:

The lower bound is proven in the same way as in $d = 1$. Consider $\mathbf{u}^*$ such that $I(\mathbf{u}^*) < +\infty$. To simplify we assume there exists a unique $\mathbf{u} \in \mathcal{D}_I^o$ such that

$$I(\mathbf{u}^*) = \mathbf{u}^* \cdot \mathbf{u} - \mathcal{Z}(\mathbf{u}) \qquad \mathbf{u} = (\nabla I)(\mathbf{u}^*)$$

Then we consider the new probability law on $\mathbb{R}^d$, absolutely continuous with respect to α, defined by

$$\alpha_{\mathbf{u}}(d\mathbf{x}) = e^{\mathbf{u}\cdot\mathbf{x} - \mathcal{Z}(\mathbf{u})}\, \alpha(d\mathbf{x})$$

Observe that

$$\int \mathbf{x}\alpha_{\mathbf{u}}(d\mathbf{x}) \;=\; \mathbf{u}^*$$

Noting $A_{n,\delta} = \{(\mathbf{x}_1, \dots, \mathbf{x}_n) : |(\mathbf{x}_1 + \cdots + \mathbf{x}_n)/n - \mathbf{u}^*| \le \delta\} \subset \mathbb{R}^n$, then for any $\delta_1 < \delta$

$$\begin{aligned}
\mathbb{P}(|\hat{\mathbf{S}}_n - \mathbf{u}^*| < \delta) &\ge \int_{A_{n,\delta_1}} \alpha(d\mathbf{x}_1) \dots \alpha(d\mathbf{x}_n) \\
&= M(\mathbf{u})^n \int_{A_{n,\delta_1}} e^{-\mathbf{u}\cdot(\mathbf{x}_1 + \cdots + \mathbf{x}_n)} \alpha_{\mathbf{u}}(d\mathbf{x}_1) \dots \alpha_{\mathbf{u}}(d\mathbf{x}_n) \\
&= M(\mathbf{u})^n e^{-n\mathbf{u}\cdot\mathbf{u}^*} \int_{A_{n,\delta_1}} e^{-\mathbf{u}\cdot[(\mathbf{x}_1 + \cdots + \mathbf{x}_n) - n\mathbf{u}^*]} \alpha_{\mathbf{u}}(d\mathbf{x}_1) \dots \alpha_{\mathbf{u}}(d\mathbf{x}_n) \\
&\ge e^{-nI(\mathbf{u}^*)} e^{-n|\mathbf{u}|\delta_1} \int_{A_{n,\delta_1}} \alpha_{\mathbf{u}}(d\mathbf{x}_1) \dots \alpha_{\mathbf{u}}(d\mathbf{x}_n)
\end{aligned}$$

The law of large numbers now says that

$$\lim_{n\to\infty} \int_{A_{n,\delta_1}} \alpha_{\mathbf{u}}(d\mathbf{x}_1) \dots \alpha_{\mathbf{u}}(d\mathbf{x}_n) = 1$$

and we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|\hat{\mathbf{S}}_n - \mathbf{u}^*| < \delta) \geq -I(\mathbf{u}^*) - |\mathbf{u}|\delta_1$$

and letting $\delta_1 \to 0$ we conclude that for any $\delta > 0$ we have the lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|\hat{\mathbf{S}}_n - \mathbf{u}^*| < \delta) \geq -I(\mathbf{u}^*)$$

The upper bound requires a little more work. Convexity plays a role here.

Let C any Borel set in $\mathbb{R}^d$. Then the exponential Chebicheff inequality implies for any $\mathbf{u} \in \mathbb{R}^d$

$$\mathbb{P}\left(\hat{\mathbf{S}}_n \in C\right) \leq \exp\left[-n \inf_{\mathbf{x}\in C} \mathbf{u}\cdot\mathbf{x}\right] \mathbb{E}\left(e^{n\mathbf{u}\cdot\hat{\mathbf{S}}_n}\right) = \exp\left[-n \inf_{\mathbf{x}\in C} \mathbf{u}\cdot\mathbf{x}\right] M(\mathbf{u})^n$$

and optimizing in $\mathbf{u} \in \mathbb{R}^d$ we obtain

$$\frac{1}{n} \log \mathbb{P}\left(\hat{\mathbf{S}}_n \in C\right) \leq - \sup_{\mathbf{u}\in\mathbb{R}^d} \inf_{\mathbf{x}\in C} [\mathbf{u}\cdot\mathbf{x} - \mathscr{Z}(\mathbf{u})] \tag{A.3.5}$$

So to conclude we need to exchange "$\sup_{\mathbf{u}\in\mathbb{R}^d}$" with "$\inf_{\mathbf{x}\in C}$". This is immediate if C is a convex set by the following lemma (c.f. [3], chapter 6):

Lemma A.3.3 *Let $g(\mathbf{u}, \mathbf{x})$ be convex and lower semicontinuous in $\mathbf{x}$, concave and uppersemicontinuous in $\mathbf{u}$, then if C is compact and convex*

$$\inf_{\mathbf{x}\in C} \sup_{\mathbf{u}\in\mathbb{R}^d} g(\mathbf{u}, \mathbf{x}) = \sup_{\mathbf{u}\in\mathbb{R}^d} \inf_{\mathbf{x}\in C} g(\mathbf{u}, \mathbf{x}) \tag{A.3.6}$$

Consider now any compact set $K \subset \mathbb{R}^d$,there exists $l > 0$ such that $\inf_{\mathbf{x}\in K} I(\mathbf{x}) = l$. By the ower semicontinuity of $I(\cdot)$, for a fixed $\epsilon > 0$ and any $\mathbf{x}' \in K$, there exists a closed ball $C(\mathbf{x}')$ such that

$$I(\mathbf{x}) \geq l - \epsilon \qquad \forall \mathbf{x} \in C(\mathbf{x}')$$

Since K is compact, there exists a finite subcover $C(\mathbf{x}'_1), \ldots, C(\mathbf{x}'_N)$ extracted from these closed ball. Then

$$\begin{aligned}\mathbb{P}\left(\hat{\mathbf{S}}_n \in K\right) &\leq \sum_{j=1}^{N} \mathbb{P}\left(\hat{\mathbf{S}}_n \in C(\mathbf{x}'_j)\right) \leq N \max_{1\leq j\leq N} \mathbb{P}\left(\hat{\mathbf{S}}_n \in C(\mathbf{x}'_j)\right) \\ &\leq N \max_{1\leq j\leq N} \exp\left(-n \inf_{C(\mathbf{x}'_j)} I\right) \leq N e^{-n(l-\epsilon)}\end{aligned}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\hat{\mathbf{S}}_n \in K\right) \leq -(l-\epsilon)$$

Since ϵ is arbitrary, this proves the upper bound for compact sets.

To extend this bound from compact to closed sets, we need to prove the *exponential tightness* of the distribution of $\hat{\mathbf{S}}_n$, i.e.

$$\lim_{\rho\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\hat{\mathbf{S}}_n \notin H_\rho\right) = -\infty \tag{A.3.7}$$

where $H_\rho = [-\rho, \rho]^d$ is the centered hypercube of length 2ρ. To prove this observe that, denoting $\hat{S}_n^{(j)}$ is the average of $X_1^{(j)}, \ldots, X_n^{(j)}$, by applying the results obtained in the one-dimensional case, we have

$$\mathbb{P}\left(\hat{\mathbf{S}}_n \notin H_\rho\right) \leq \sum_{j=1}^{d} \mathbb{P}\left(\hat{S}_n^{(j)} \notin (-\rho, \rho)\right) \leq d \max_{j=1,\ldots,d} \exp\left(-n \min\{I^j(\rho), I^j(-\rho)\}\right)$$

where I^j is the rate function for the $j-$marginal distribution of the law α. Then (A.3.7) follows by applying lemma A.2.2.

□

A.4 Generalities on Large Deviations

Let X a complete separable metric space and P_n a family of probability distributions on X. In the previous sections $X = \mathbb{R}^d$ and P_n the distribution of $\hat{S}_n$. We says that $\{P_n\}$ satisfies a large deviation principle with good rate function $I(\cdot)$ if there exists a function $I : X \to [0, \infty]$ such that:

1. $I(\cdot)$ is lower semicontinuous.
2. For each $\ell < \infty$ the set $\{x : I(x) \leq \ell\}$ is compact in X.
3. For each closed set $C \subset X$
$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \leq -\inf_{x\in C} I(x).$$
4. For each open set $G \subset X$
$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G} I(x).$$

Here the adjective *good* refers to properties 1 and 2. The next lemma does not require the rate function I to be good.

Theorem A.4.1 ***Varadhan's Lemma.*** *Let P_n satisfy the large deviation principle with rate function I. Then for any bounded continuous function $F(x)$ on X*

$$\lim_{n\to\infty}\frac{1}{n}\log\int e^{nF(x)}dP_n(x)=\sup_{x\in X}\{F(x)-I(x)\}.$$

Proof.

Upper bound. For any given $\delta>0$, since F is bounded and continuous, we can find a finite number of closed sets covering X such that the oscillation of $F(\cdot)$ on each of these closed sets is less or equal δ. Then

$$\int e^{nF(x)}dP_n(x)\le\sum_{j=1}^{m}\int_{C_j}e^{nF(x)}dP_n(x)\le\sum_{j=1}^{m}e^{nF_j+\delta}P_n(C_j)$$

where $F_j=\inf_{C_j}F(x)$. It follows

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\int e^{nF(x)}dP_n(x)&\le\sup_{1\le j\le m}[F_j+\delta-\inf_{C_j}I(x)]\\&\le\sup_{1\le j\le m}\sup_{C_j}[F(x)-I(x)]+\delta\\&=\sup_{x\in X}[F(x)-I(x)]+\delta\end{aligned}$$

Since δ is arbitrary, we can let it go to 0.

Lower bound. By definition of a supremum for any $\delta>0$ we can find $y\in X$ such that $F(y)-I(y)\ge\sup_x[F(x)-I(x)]-\delta/2$. Since F is continuous we can find an open neighborhood U of y such that $F(x)\ge F(y)-\delta/2$ for any $x\in U$. Then we obtain

$$\liminf_{n\to\infty}\frac{1}{n}\log\int e^{nF(x)}dP_n(x)\ge\liminf_{n\to\infty}\frac{1}{n}\log\int_U e^{nF(x)}dP_n(x)$$
$$\ge F(y)-\frac{\delta}{2}-\inf_{x\in U}I(x)\ge F(y)-I(y)-\frac{\delta}{2}\ge\sup_x[F(x)-I(x)]-\delta$$

and we conclude from the arbitrariness of δ. $\square$

Theorem A.4.2 ***Contraction Principle.*** *Let P_n satisfy the large deviation principle with rate function I, and $\pi : X \to Y$ a continuous mapping from X to another complete separable metric space Y. Then $\tilde{P}_n = P_n \pi^{-1}$ satisfies a large deviation principle with rate function*

$$\tilde{I}(y) = \inf_{x:\pi(x)=y} I(x),$$
$$\tilde{I}(y) = +\infty \qquad \text{if } \{x : \pi(x) = y\} = \emptyset$$

Proof. Since π is continuous, given any closed set $\tilde{C} \subset Y$, the subset $C = \pi^{-1}(\tilde{C})$ is closed in X. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log \tilde{P}_n(\tilde{C}) = \limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \le -\inf_{x\in C} I(x) = -\inf_{y\in\tilde{C}} \inf_{x:\pi(x)=y} I(x).$$

and similarly for the lower bound. □

A.5 Large deviations for densities

We deal first with the one-dimensional case. If the distribution of $\hat{S}_n$ on $\mathbb{R}$ has a density that we denote by $f_n(x)$, from Cramers theorem we have the intuition that $f_n(x) \sim e^{-nI(x)}$ for large n. We will prove this under some condition on the probability $\alpha(dx)$. It is interesting to notice that we will not use Cramer's theorem in the proof, but the following *local central limit theorem.*

Theorem A.5.1 ***Local central limit theorem.*** *Let $\phi(k)$ the characteristic function of a centered probability measure $\alpha(dx)$ with finite variance σ^2, and assume that $|\phi(k)| < 1$ if $k \neq 0$ and that there exists an integer $r \geq 1$ such that $|\phi|^r$ is integrable. Let $\tilde{g}_n(x)$ the probability density of $(X_1 + \cdots + X_n)/\sqrt{n}$, where X_j are i.i.d. with common law α. Then*

$$\lim_{n\to\infty} \tilde{g}_n(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}.$$

Proof. The characteristic function of α is defined by

$$\phi(k) = \int e^{ixk} \alpha(dx) \tag{A.5.1}$$

The characteristic function of the distribution of $X_1 + \cdots + X_r$ is $\phi^r(k)$ that is integrable. It follows that the probability density $\tilde{g}_n(x)$ exists for any $n \geq r$ (cf. [?], theorem XV.3.3). Then

$$\tilde{g}_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixk} \left[\phi\left(\frac{k}{\sqrt{n}} \right) \right]^n dk$$

and therefore

$$\left|\tilde{g}_n(x) - \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}\right| \leq \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left|\phi\left(\frac{k}{\sqrt{n}}\right)^n - e^{-k^2\sigma^2/2}\right| dk$$

Given $a > 0$, we split the integral in three parts.

1. Uniformly in $k \in [-a, a]$,

$$\phi\left(\frac{k}{\sqrt{n}}\right)^n = \left(1 - \frac{k^2\sigma^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \underset{n\to\infty}{\longrightarrow} e^{-k^2\sigma^2/2}$$

   so that

$$\int_{-a}^{+a}\left|\phi\left(\frac{k}{\sqrt{n}}\right)^n - e^{-k^2\sigma^2/2}\right| dk \to 0$$

2. Observe that it is possible to choose $\delta > 0$ such that

$$|\phi(k)| \leq e^{-k^2\sigma^2/4} \qquad \text{if} \quad |k| \leq \delta.$$

   Then for the interval $|k| \in (a, \delta\sqrt{n})$, we can estimate as

$$\int_{a}^{\delta\sqrt{n}}\left|\phi\left(\frac{k}{\sqrt{n}}\right)^n - e^{-k^2\sigma^2/2}\right| dk \leq \int_{a}^{\delta\sqrt{n}} 2e^{-k^2\sigma^2/4} dk \leq \int_{a}^{+\infty} 2e^{-k^2\sigma^2/4} dk$$

   that converge to 0 as $a \to \infty$.

3. It remains to estimate the contribution from the interval $(\delta\sqrt{n}, +\infty)$. Since we assumed that $|\phi(k)| < 1$ for $k \neq 0$, and since $|\phi|^k$ is integrable, we have $\phi(k) \to 0$ as $k \to \infty$. Consequently we must have $\sup_{|k|\geq\delta} |\phi(k)| = \eta < 1$, and we can estimate

$$\begin{aligned}\int_{\delta\sqrt{n}}^{+\infty}\left|\phi\left(\frac{k}{\sqrt{n}}\right)^n - e^{-k^2\sigma^2/2}\right| dk &\leq \eta^{n-r}\int_{-\infty}^{+\infty}\left|\phi\left(\frac{k}{\sqrt{n}}\right)\right|^r dk + \int_{\delta\sqrt{n}}^{+\infty} e^{-k^2\sigma^2/2} dk \\ &= \eta^{n-r}\sqrt{n}\int_{-\infty}^{+\infty}|\phi(k)|^r dk + \int_{\delta\sqrt{n}}^{+\infty} e^{-k^2\sigma^2/2} dk\end{aligned}$$

   that converges to 0 as $n \to \infty$.

□

Distributions such that their characteristic function $|\phi(k)| < 1$ for $k \neq 0$ are called *non-lattice* ([2], chapter 2). It does not imply they have density.

We assume now that the measure $\alpha(dx)$ satisfies all the assumptions made in section A.2, and furthermore its characteristic function satisfies conditions of the local central limit theorem A.5.1. Then, for $n \geq r$, the distribution of $\hat{S}_n$ on $\mathbb{R}$ has a density that we denote by $f_n(x)$.

Theorem A.5.2 *For any $y \in \mathcal{D}_I^o$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(y) = -I(y) \ . \tag{A.5.2}$$

Proof.

Let $\tau_y\alpha$ the translation of the measure α by y. Assume that $m = \int x\alpha(dx) = 0$, otherwise just recenter it and consider $\tau_m\alpha$.

Let $y \in \mathcal{D}_I{}^o$. Then by lemma A.2.4 there exists a unique $\lambda \in \mathcal{D}_{\mathcal{Z}}{}^o$ such that $y = \mathcal{Z}'(\lambda)$, $\lambda = I'(y)$, and $I(y) = \lambda y - \mathcal{Z}(\lambda)$. Define

$$\tilde{\alpha}(y, dx) = \frac{1}{M(\lambda)} e^{(x+y)\lambda} \tau_y\alpha(dx)$$

Observe that this is a probability distribution with 0 average. In fact

$$\int \tilde{\alpha}(y, dx) = \frac{1}{M(\lambda)} \int e^{z\lambda}\alpha(dz) = 1$$

and

$$\int x\tilde{\alpha}(y, dx) = -y + \frac{1}{M(\lambda)} \int z e^{z\lambda}\alpha(dz) = -y + \mathcal{Z}'(\lambda) = 0$$

So we treat here y as a parameter. Let $X_1^y, \ldots, X_n^y$ i.i.d. random variables with law given by $\tilde{\alpha}(y, dx)$.

For $n \geq r$ it exists the density for the distribution of $(X_1^y + \cdots + X_n^y)/n$ that we denote by $f_n(x, y)$, and it is equal to

$$f_n(x, y) = \frac{e^{n(x+y)\lambda}}{M(\lambda)^n} f_n(x + y) = e^{n(I(y)+\lambda x)} f_n(x + y)$$

To prove this formula, compute, for a given bounded measurable function $G(\cdot)$:

$$\begin{aligned} \mathbb{E}\left(G((X_1^y + \cdots + X_n^y)/n)\right) &= \int_{\mathbb{R}^n} G(\hat{s}_n) e^{n(I(y)+\lambda\hat{s}_n)} \tau_y\alpha(dx_1) \ldots \tau_y\alpha(dx_n) \\ &= \int_{\mathbb{R}} G(\hat{s}) e^{n(I(y)+\lambda\hat{s})} f_n(\hat{s} + y) d\hat{s} \end{aligned} \tag{A.5.3}$$

It follows that

$$f_n(y) = e^{-nI(y)} f_n(0, y)$$

To conclude we only need to prove that $(\log f_n(0, y))/n \to 0$ as $n \to \infty$.

Let $\tilde{f}_n(x, y)$ the density of $(X_1^y + \cdots + X_n^y)/\sqrt{n}$. Then $f_n(x, y) = \sqrt{n}\tilde{f}_n(\sqrt{n}x, y)$. By the local central limit theorem A.5.1, the result follows immediately. □

For $y \in \mathbb{R}$ define $\nu_y^{(n)}(dx_1, \ldots, dx_n)$ the conditional distribution of $(X_1, \ldots, X_n)$ on the hyperplane $x_1 + \cdots + x_n = ny$. This is defined as the probability measure on $\mathbb{R}^{n-1}$ satisfying the relation

$$\mathbb{E}\left(G(\hat{S}_n)H(X_1, \ldots, X_n)\right) = \int_{\mathbb{R}} dy f_n(y) G(y) \int H(x_1, \ldots, x_n) \nu_y^{(n)}(dx_1, \ldots, dx_n)$$

Lemma A.5.3 *Let F be a bounded continuous function on $\mathbb{R}$ and $y \in \mathcal{D}_I^o$, $\lambda = I'(y)$. For every $\theta \in \mathbb{R}$, the limit*

$$\lim_{n\to\infty} \frac{1}{n} \log \int e^{\theta(F(x_1)+\ldots.+F(x_n))} \nu_y^{(n)}(dx_1, \ldots, dx_n) = G(y, \theta) \qquad \text{(A.5.4)}$$

exists and G is differentiable at $\theta = 0$ with

$$\left.\frac{\partial G(y, \theta)}{\partial \theta}\right|_{\theta=0} = \int F(x) \alpha_\lambda(dx). \qquad \text{(A.5.5)}$$

Proof Denote by $H_n(y, \theta)$ the function

$$\int e^{\theta(F(x_1)+\ldots.+F(x_n))} \nu_y^{(n)}(dx_1, \ldots, dx_n) = \frac{H_n(y, \theta)}{f_n(y)} \qquad \text{(A.5.6)}$$

which, by (A.5.3), can be formally written as

$$H_n(y, \theta) = \int_{x_1+\cdots+x_n=ny} e^{\theta(F(x_1)+\ldots.+F(x_n))} \alpha(dx_1) \ldots \alpha(dx_n).$$

Let us denote

$$a(\theta) = \int e^{\theta F(x)} \alpha(dx), \qquad M(\lambda, \theta) = \frac{1}{a(\theta)} \int e^{\lambda x + \theta F(x)} \alpha(dx)$$

Then we can compute the Cramér rate function for the law $a(\theta)^{-1} e^{\theta F(x)} \alpha(dx)$, and this is given by

$$I_\theta(y) = I(y, \theta) = \sup_{\bar{\lambda}} \left\{\bar{\lambda} y - \log M(\bar{\lambda}, \theta)\right\}$$

Observe that $\mathcal{D}_{I_\theta} = \mathcal{D}_I$ because F is bounded. If $(Y_1, \ldots, Y_n)$ are i.i.d. distributed by $a(\theta)^{-1} e^{\theta F(x)} \alpha(dx)$, then the density of the distribution of $(Y_1 + \cdots + Y_n)/n$ is given by $a(\theta)^{-n} H_n(y, \theta)$. Then by applying A.5.2 to this law we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log H_n(y, \theta) = -I(y, \theta) + \log a(\theta).$$

Consequently we have, applying again A.5.2

$$\lim_{n\to\infty}\frac{1}{n}\log\int e^{\theta(F(x_1)+\dots+F(x_n))}\nu_y^{(n)}(dx_1,\dots,dx_n) = \lim_{n\to\infty}\frac{1}{n}\log H_n(y,\theta) - \lim_{n\to\infty}\frac{1}{n}\log f_n(y)$$
$$= \log a(\theta) - I(y,\theta) + I(y) \equiv G(y,\theta).$$

Differentiating $G(y,\theta)$ we have

$$\frac{\partial G(y,\theta)}{\partial\theta} = \frac{a'(\theta)}{a(\theta)} - \frac{\partial I(y,\theta)}{\partial\theta}$$

In order to compute this last expression let us set $\lambda^*(y,\theta) = \partial_y I(y,\theta)$. Existence of $\lambda^*(y,\theta)$ is provided by the assumption $y \in \mathcal{D}_I^o$ and the equality between the sets $\mathcal{D}_I$ and $\mathcal{D}_{I_\theta}$. We have

$$I(y,\theta) = \lambda^* y - \log M(\lambda^*,\theta).$$

Then, since $\partial_\lambda \log M(\lambda^*,\theta) = y$, we get

$$\partial_\theta I(y,\theta) = y\partial_\theta\lambda^* - M^{-1}\left(\partial_\theta M + \partial_\lambda M\partial_\theta\lambda^*\right) = -\partial_\theta \log M(\lambda^*,\theta)$$
$$= \partial_\theta \log a(\theta) - M^{-1}\partial_\theta\int e^{\lambda x+\theta F(x)}\alpha(dx) = \frac{a'(\theta)}{a(\theta)} - \int F(x)e^{\lambda^* x+\theta F(x)-\log M(\lambda^*,\theta)}\alpha(dx)$$

So we have

$$\partial_\theta G(y,\theta) = \int F(x)e^{\lambda^* x+\theta F(x)-\log M(\lambda^*,\theta)}\alpha(dx)$$

and sending $\theta \to 0$ we obtain

$$\partial_\theta G(y,0) = \int F(x)e^{\lambda^*(y,0)x-\log M(\lambda^*(y,0),0)}\alpha(dx) = \int F(x)\alpha_\lambda(dx)$$

□

Theorem A.5.4 *For any $y \in \mathcal{D}_I^o$, and any $\epsilon > 0$*

$$\lim_{n\to\infty}\nu_y^{(n)}\left(\left|\frac{1}{n}\sum_{j=1}^n F(X_j) - \int F(x)\alpha_\lambda(dx)\right| \geq \epsilon\right) = 0 \qquad \text{(A.5.7)}$$

Proof. Without loosing any generality, let us assume that $\int F(x)\alpha_\lambda(dx) = 0$. Consequently $G(\theta,y) = O(\theta^2)$. Then for any $\theta > 0$

$$\nu_y^{(n)}\left(\left|\frac{1}{n}\sum_{j=1}^n F(X_j)\right| \geq \epsilon\right) \leq e^{-n\theta\epsilon}\int e^{\theta|\sum_{j=1}^n F(x_j)|}\nu_y^{(n)}(dx_1,\dots,dx_n)$$
$$\leq e^{-n\theta\epsilon}\int e^{\theta\sum_{j=1}^n F(x_j)}\nu_y^{(n)}(dx_1,\dots,dx_n)$$
$$+e^{-n\theta\epsilon}\int e^{-\theta\sum_{j=1}^n F(x_j)}\nu_y^{(n)}(dx_1,\dots,dx_n)$$

and by (A.5.4)

$$\lim_{n\to\infty} \frac{1}{n} \log \nu_y^{(n)} \left(\left| \frac{1}{n} \sum_{j=1}^{n} F(x_j) \right| \geq \epsilon \right) \leq -\theta\epsilon + \max\{G(\theta, y), G(-\theta, y)\}$$

Optimizing the above bound in θ one obtains

$$\lim_{n\to\infty} \frac{1}{n} \log \nu_y^{(n)} \left(\left| \frac{1}{n} \sum_{j=1}^{n} F(x_j) \right| \geq \epsilon \right) \leq -C\epsilon^2$$

for some positive constant C. □

Observe that $\nu_y^{(n)}$ is a symmetric measure, so we have

$$\int F(x_1)\nu_y^{(n)}(dx_1, \dots, dx_n) = \int \frac{1}{n} \sum_{j=1}^{n} F(x_j)\nu_y^{(n)}(dx_1, \dots, dx_n) \underset{n\to\infty}{\longrightarrow} \int F(x)\alpha_\lambda(dx)$$

Theorem A.5.5 *Let $F(x_1, \dots, x_k)$ a bounded continuous function on $\mathbb{R}^k$ and $y \in \mathcal{D}_I^o$, then*

$$\lim_{n\to\infty} \int F(x_1, \dots, x_k)\nu_y^{(n)}(dx_1, \dots, dx_n) = \int F(x_1, \dots, x_k)\alpha_\lambda(dx_1) \dots \alpha_\lambda(dx_k)$$

Proof. It is enough to consider functions of the form $F(x_1, \dots, x_k) = F_1(x_1) \dots F(x_k)$. For simplicity let us prove the case $k = 2$, the generalization to any k is straightforward. Without loosing generality, let us assume that $\int F_j(x)\alpha_\lambda(dx) = 0$. By the exchange symmetry of $\nu_y^{(n)}$ we have

$$\int F_1(x_1)F_2(x_2)\nu_y^{(n)}(dx_1, \dots, dx_n) = \int \frac{1}{n(n-1)} \sum_{i\neq j} F_1(x_i)F_2(x_j)\nu_y^{(n)}(dx_1, \dots, dx_n)$$

$$= \int \frac{n^2}{n(n-1)} \left(\frac{1}{n} \sum_i F_1(x_i) \right) \left(\frac{1}{n} \sum_j F_2(x_j) \right) \nu_y^{(n)}(dx_1, \dots, dx_n) + O\left(\frac{1}{n}\right)$$

and this last expression converges to 0 an $n \to \infty$ by (A.5.7) .

□

The generalization to more dimensions of the above results is quite straightforward and can be left as exercise. Let us state here what the result is in this context.

Let $\alpha(d\mathbf{x})$ a probability measure on $\mathbb{R}^d$ that satisfies conditions used in section A.3. Let us assume that its characteristic function is such that $|\phi(\mathbf{k})| < 1$ for $\mathbf{k} \neq 0$, and such that $|\phi(\mathbf{k})|^r$ is integrable on $\mathbb{R}^d$ for some integer $r \geq 1$. Then, for $n \geq r$ the n-convolution of α has a density and we denote by $f_n(\mathbf{x})$ the density of the distribution of $(\mathbf{X}_1 + \cdots + \mathbf{X}_n)/n$, where $\{\mathbf{X}_j\}$ are i.i.d. with common distribution $\alpha(d\mathbf{x})$.

Theorem A.5.6 *For any* $\mathbf{y} \in \mathcal{D}_I^o$ *we have*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(\mathbf{y}) = -I(\mathbf{y}) \ . \tag{A.5.8}$$

Example Let $V : \mathbb{R} \to \mathbb{R}_+$ a positive function such that $V(y) \to +\infty$ for $|y| \to +\infty$, and such that

$$Z(\lambda, \beta) = \int e^{-\beta V(y) + \lambda y} dy < \infty \qquad \forall\ \lambda \in \mathbb{R}, \beta > 0.$$

Then we can define the probability density (on $\mathbb{R}^2$)

$$f_{\lambda,\beta}(r,p) = \frac{e^{-\beta(V(r)+p^2/2)+\lambda r}}{\sqrt{2\pi\beta^{-1}}Z(\lambda,\beta)} \tag{A.5.9}$$

Let $\{Y_j = (r_j, p_j)\}$ be a sequence of i.i.d. random variables with common law given by $f_{0,\beta_0}(r,p) dr\, dp$, $\beta_0 > 0$ fixed.

Then the vector valued random variables $\mathbf{X}_j = (r_j,\ (V(r_j) + p_j^2/2))$ clearly has a law $\alpha(d\mathbf{x})$ which is degenerate in $\mathbb{R}^2$ but $\alpha * \alpha$ has a density w.r.t. the Lebesgue measure. Its logarithmic moment generating function is given by

$$\mathcal{Z}(\lambda, \eta) = \log \int e^{\lambda r + \eta (V(r)+p^2/2)} f_{0,\beta}(r,p) dr dp = \log \left(\frac{Z(\lambda, \beta_0 - \eta)}{Z(0,\beta_0)} \sqrt{\frac{\beta_0}{\beta_0 - \eta}} \right)$$

for $\eta < \beta_0$ and $+\infty$ otherwise. The corresponding Legendre transform, for $r \in \mathbb{R}$ and $\mathcal{E} > 0$, is given by

$$\begin{aligned} I(r, \mathcal{E}) &= \sup_{\eta<\beta_0,\lambda} \{\lambda r + \eta\mathcal{E} - \log \mathcal{Z}(\lambda,\eta)\} \\ &= \sup_{\beta>0,\lambda} \left\{ \lambda r - \beta\mathcal{E} - \log\left(\sqrt{2\pi\beta^{-1}} Z(\lambda,\beta)\right)\right\} + \beta_0\mathcal{E} + \log\left(\sqrt{2\pi\beta_0^{-1}}Z(0,\beta_0)\right) \end{aligned}$$

The function defined by

$$S(r,\mathcal{E}) = \inf_{\lambda,\beta>0}\left\{-\lambda r + \beta\mathcal{E} - \log\left(Z(\lambda,\beta)\sqrt{2\pi\beta^{-1}}\right)\right\} \tag{A.5.10}$$

is called *thermodynamic entropy*. So we have obtained

$$I(r,\mathcal{E}) = -S(r,\mathcal{E}) + \beta_0\mathcal{E} + \log Z(0,\beta_0) + \frac{1}{2}\log\frac{2\pi}{\beta_0}$$

Observe that S does not depend on β_0.

The density of the distribution of $\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_j$ is given by

$$\begin{aligned}
&f_n(r,\mathcal{E}) \\
&= \int_{\mathbb{R}^{2n}} \frac{e^{-\beta_0\sum_j \mathcal{E}_j}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\delta\left(\frac{1}{n}\sum_{j=1}^{n}\mathcal{E}_j - \mathcal{E};\frac{1}{n}\sum_{j=1}^{n}r_j - r\right)\prod_j dr_j dp_j \\
&= \frac{e^{-n\beta_0\mathcal{E}}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\int_{\mathbb{R}^{2n}}\delta\left(\frac{1}{n}\sum_{j=1}^{n}\mathcal{E}_j - \mathcal{E};\frac{1}{n}\sum_{j=1}^{n}r_j - r\right)\prod_j dr_j dp_j \\
&\qquad = \frac{e^{-n\beta_0\mathcal{E}}}{(2\pi\beta_0^{-1})^{n/2}Z(0,\beta_0)^n}\Gamma_n(r,\mathcal{E}).
\end{aligned}$$

where $\Gamma_n(r,p,\mathcal{E})$ is the volume of the corresponding $2n-2$-dimensional surface on $\mathbb{R}^{2n}$ and does not depend on β_0. Applying (A.5.8) we obtain

$$\lim_{n\to\infty}\frac{1}{n}\log\Gamma_n(r,\mathcal{E}) = S(r,\mathcal{E})\ . \tag{A.5.11}$$

for any $(r,\mathcal{E}) \in \mathcal{D}_S^o$. A sufficient condition to have $\mathcal{D}_S^0 = \mathbb{R}\times(0,\infty)$ is $V(r) \geq cr^2$ for a positive constant c.

Bibliography

[1] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[2] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, third edition, 2005.

[3] Ivar Ekeland and Roger Temam. *Convex analysis and variational problems*. North-Holland Publishing Co., Amsterdam, 1976. Translated from the French, Studies in Mathematics and its Applications, Vol. 1.

[4] Oscar E Lanford, III. Entropy and equilibrium states in classical statistical mechanics. In A Lenard, editor, *Statistical Mechanics and Mathematical Problems*, volume 20 of *Lecture Notes in Physics*, pages 1–113. Springer Berlin / Heidelberg, 1973.

[5] Giorgio Parisi. On the approach to equilibrium of a hamiltonian chain of anharmonic oscillators. *Europhysics Letters*, 40:357–362, 1997.

[6] Mark W. Zemansky. *Heat and Thermodynamics*. McGraw-Hill, 1957.