Large Deviations

Stefano Olla

January 11, 2014

Chapter 1

Tilting

1.1 Legendre transforms and the rate function I

Let $\alpha(dx)$ a probability distribution on $\mathbb{R}$. We define the moment generating function

$$M(\lambda) = \int e^{\lambda x} \alpha(dx) \tag{1.1.1}$$

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

$$\mathcal{Z}(\lambda) = \log M(\lambda) \tag{1.1.2}$$

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

Lemma 1.1.1 *1. $\mathcal{Z}(\cdot)$ is convex.*

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

$$\mathcal{Z}'(\lambda) = \int x e^{\lambda x - \mathcal{Z}(\lambda)} \alpha(dx) \qquad \lambda \in \mathcal{D}^o_{\mathcal{Z}}.$$

Proof: For any $\gamma \in [0,1]$, it follows by Hölder inequality

$$M(\gamma\lambda_1 + (1-\gamma)\lambda_2) \le M(\lambda_1)^\gamma M(\lambda_2)^{1-\gamma}$$

and consequently

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

The function $f_\epsilon(x) = (e^{(\lambda+\epsilon)x} - e^{\lambda x})/\epsilon$ converges point-wise 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 $\lambda \in \mathcal{D}^o_{\mathcal{Z}}$, there exists a $\delta > 0$ small enough such that $\int h(x)d\alpha(x) \leq M(\lambda+\delta) + M(\lambda-\delta) < +\infty$. 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

$$\mathcal{Z}''(\lambda) = \int x^2 e^{\lambda x - \mathcal{Z}(\lambda)}\alpha(dx) - \left(\int xe^{\lambda x - \mathcal{Z}(\lambda)}\alpha(dx)\right)^2 \geq 0$$

Observe that $\alpha_\lambda(dx) := e^{\lambda x - \mathcal{Z}(\lambda)}\alpha(dx)$ is a probability measure, with average $\mathcal{Z}'(\lambda)$ and variance $\mathcal{Z}''(\lambda)$.

To avoid the trivial deterministic case, we assume that $\mathcal{Z}''(0) > 0$. 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 $\mathcal{Z}$

$$I(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - \mathcal{Z}(\lambda)\} \tag{1.1.3}$$

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

$$I(x) = \sup_{\lambda \geq 0} \{\lambda x - \mathcal{Z}(\lambda)\} \qquad x > m \tag{1.1.4}$$

Similarly one obtains

$$I(x) = \sup_{\lambda \leq 0} \{\lambda x - \mathcal{Z}(\lambda)\} \qquad x < m \tag{1.1.5}$$

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

Lemma 1.1.2 *$I(x) \to +\infty$ 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. □

1.1.1 Properties of Legendre transforms

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

Lemma 1.1.3

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 unique 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 $\bar{x}$. □

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 $I(x) = (x-m)^2/2\sigma^2$.

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 $x < 0$.

5. If $\mathcal{Z}(\lambda) = p^{-1}|\lambda|^p$, then $\mathcal{Z}^*(x) = q^{-1}|x|^q$, with $p^{-1} + q^{-1} = 1$.

1.2 A more general setup

We can extend the above setup to situation where α is only a positive measure on a measurable topological space Ω. We will be interested essentially to $\Omega = \mathbb{R}^d$ (eventually $d = 2$) and α as the Lebesgue measure on it. Other examples are $\Omega = \mathbb{S}^d$, the d-dimensional sphere, and α the corresponding uniform measure.

Let $g : \Omega \to \mathbb{R}$ a measurable function and define

$$\mathcal{Z}(\lambda) = \log \int_\Omega e^{\lambda g(\omega)} d\alpha(\omega) \tag{1.2.1}$$

on the corresponding domain $\mathcal{D}_\mathcal{Z}$. Notice that now this could be empty, and in particular $0 \notin \mathcal{D}_\mathcal{Z}$ if α is not of finite measure. We assume that $\mathcal{D}^0_\mathcal{Z} \neq \emptyset$, and that g is not constant. With similar proofs as above, $\mathcal{Z}$ is $\mathcal{C}^\infty$ on the interior of $\mathcal{D}_\mathcal{Z}$.

In any case, if $\lambda \in \mathcal{D}_\mathcal{Z}$, then $d\alpha_\lambda(\omega) = e^{\lambda g - \mathcal{Z}(\lambda)} d\alpha(\omega)$ is a probability distribution on Ω. With respect to α_λ, g can be seen as a random variable with average $\mathcal{Z}'(\lambda)$ and variance $\mathcal{Z}''(\lambda)$. Notice that $\mathcal{Z}''(\lambda) > 0$ just because g is not constant.

Examples are easily recovered in this set up. The gaussian distribution of example 1 is obtained by taking $\Omega = \mathbb{R}$, $d\alpha = dx$, and $g(x) = -(x-m)^2/2$. Then $\mathcal{Z}(\lambda) = \frac{1}{2}\log\left(\frac{2\pi}{\lambda}\right)$, defined only for $\lambda > 0$, and the tilted measure α_λ is the gaussian measure with average m and variance $\sigma^2 = \lambda^{-1}$. Notice now that the Legendre transform of $\mathcal{Z}$, that we now denote with $\mathcal{Z}^*$ is different from the rate function I, in particular $\mathcal{Z}^*(x) = -\frac{1}{2}\log(-\pi x) - 2$ for $x < 0$, and $+\infty$ if $x > 0$.

More generally we need a multidimensional setup. Let $\mathbf{g} : \boldsymbol{\Omega} \to \mathbb{R}^\mathbf{r}$ a vector valued measurable function and

$$\mathcal{Z}(\boldsymbol{\lambda}) = \log \int_\Omega e^{\boldsymbol{\lambda}\cdot\mathbf{g}(\omega)} d\alpha(\omega) \tag{1.2.2}$$

finite in the corresponding domain $\mathcal{D}_\mathcal{Z} \subset \mathbf{R}^r$. $\mathcal{Z}(\boldsymbol{\lambda})$ is convex and lower semicontinuous (it maybe not continuous). Again strict convexity follows by assuming that every component of $\mathbf{g}$ is not constant. Furthermore $\mathcal{D}_\mathcal{Z}$ is convex.

The Fenchel-Legendre transform is now defined by

$$\mathcal{Z}^*(\boldsymbol{x}) = \sup_{\boldsymbol{\lambda}} \left\{\boldsymbol{\lambda}\cdot\mathbf{x} - \mathcal{Z}(\boldsymbol{\lambda})\right\}. \tag{1.2.3}$$

which, in the domain $\mathcal{D}_{\mathcal{Z}^*} = \{\boldsymbol{x} \in \mathbb{R}^r : \mathcal{Z}^*(\boldsymbol{x}) < +\infty\}$, is also convex and lower-semicontinuous as supremum of linear functionals. As before, there is a unique correspondence from $\mathcal{D}_\mathcal{Z}$ to $\mathcal{D}_{\mathcal{Z}^*}$ such that

$$\begin{aligned} \bar{\boldsymbol{x}} &= \nabla\mathcal{Z}(\bar{\boldsymbol{\lambda}}) \\ \bar{\boldsymbol{\lambda}} &= \nabla\mathcal{Z}^*(\bar{\boldsymbol{x}}) \end{aligned} \tag{1.2.4}$$

For $\boldsymbol{\lambda} \in \mathcal{D}_\mathcal{Z}$, we have the probability measure $d\alpha_{\boldsymbol{\lambda}}(\omega) = e^{\boldsymbol{\lambda}\cdot\mathbf{g} - \mathcal{Z}(\boldsymbol{\lambda})} d\alpha(\omega)$ on Ω. With respect to $\alpha_{\boldsymbol{\lambda}}$, $\mathbf{g}$ can be seen as a vector valued random variable with average $\nabla\mathcal{Z}(\boldsymbol{\lambda})$ and covariance matrix $\text{Hess}\mathcal{Z}(\boldsymbol{\lambda}) = \nabla^2\mathcal{Z}(\boldsymbol{\lambda})$. Also now $\text{Hess}\mathcal{Z}(\boldsymbol{\lambda}) > 0$ just because $\mathbf{g}$ is not constant.

The gaussian example can be recovered with the choice $d = 1$, $r = 2$, $\mathbf{g} = (x^2/2, x)$, $\alpha = dx$. Then

$$\mathcal{Z}(\lambda_1, \lambda_2) = \frac{1}{2}\log\left(\frac{2\pi}{\lambda_1}\right) + \frac{1}{2}\left(\frac{\lambda_2}{\lambda_1}\right)^2$$

and $\alpha_{\boldsymbol{\lambda}}$ is the gaussian measure on $\mathbb{R}$ with variance λ_1^{-1} and average $\lambda_2\lambda_1^{-1}$.

We will be interested mostly in the following example.

Let $\Omega = \mathbb{R}^2$, we will denote $\omega = (r, p)$, and α the usual Lebesgue measure $drdp$. Let $V : \mathbb{R} \to \mathbb{R}_+$ a smooth function, such that $V(r) \to +\infty$ as $|r| \to \infty$, and such that

$$\int e^{-\beta V(r)} dr < +\infty \qquad \forall \beta > 0. \tag{1.2.5}$$

Then we choose $g = (-[p^2/2 + V(r)], r)$. We will make sure that all conditions we will assume in the following will be satisfied by this example.

Observe that

$$\mathcal{Z}(\boldsymbol{\lambda} + \boldsymbol{\delta}) - \mathcal{Z}(\boldsymbol{\lambda}) = \log \int_\Omega e^{\boldsymbol{\delta}\cdot\mathbf{g}} d\alpha_{\boldsymbol{\lambda}} \tag{1.2.6}$$

So in the following we will work with a *probability* measure α that will indicate $\alpha_{\boldsymbol{\lambda}_0}$ for some $\boldsymbol{\lambda}_0$.

In particular notice that the rate function $I_{\boldsymbol{\lambda}}(\mathbf{x})$ corresponding to the tilted measure $\alpha_{\boldsymbol{\lambda}}$ is given by

$$I_{\boldsymbol{\lambda}}(\mathbf{x}) = \mathcal{Z}^*(\mathbf{x}) - \boldsymbol{\lambda}\cdot\mathbf{x} + \mathcal{Z}(\boldsymbol{\lambda}) \tag{1.2.7}$$

1.3 Local Central Limit Theorem

Theorem 1.3.1 ***Local central limit theorem.*** *Let $\phi(\mathbf{k})$ the characteristic function of a centered probability measure $\nu(d\mathbf{x})$ on $\mathbb{R}^r$ with finite covariance matrix σ^2, and assume that $|\phi(\mathbf{k})| < \mathbf{1}$ if $\mathbf{k} \neq \mathbf{0}$ and that there exists an integer $n_0 \geq 1$ such that $|\phi|^{n_0}$ 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 ν. Then*

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

Proof. This is a standard proof, we will consider here the one dimensional case, the multidimensional case is straighforward.

The characteristic function of ν is defined by

$$\phi(k) = \int e^{ixk} \nu(dx) \tag{1.3.1}$$

The characteristic function of the distribution of $X_1 + \cdots + X_n$ is $\phi^n(k)$ that is integrable for $n \geq n_0$. It follows that the probability density $\tilde{g}_n(x)$ exists for any $n \geq n_0$ (cf. Feller theorem XV.3.3). Then

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

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 $|\phi|^n{}_0$ 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

$$\begin{aligned}\int_{\delta\sqrt{n}}^{+\infty} \left| \phi\left(\frac{k}{\sqrt{n}}\right)^n - e^{-k^2\sigma^2/2} \right| dk &\leq \eta^{n-n_0} \int_{-\infty}^{+\infty} \left| \phi\left(\frac{k}{\sqrt{n}}\right) \right|^{n_0} dk + \int_{\delta\sqrt{n}}^{+\infty} e^{-k^2\sigma^2/2} dk \\ &= \eta^{n-n_0}\sqrt{n} \int_{-\infty}^{+\infty} |\phi(k)|^{n_0} dk + \int_{\delta\sqrt{n}}^{+\infty} e^{-k^2\sigma^2/2} dk\end{aligned}$$

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

□

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

1.4 A Local Large Deviation Theorem

We assume now that the probability measure $\nu(dx)$ satisfies all the assumptions made in section 1, and furthermore its characteristic function satisfies the conditions of the local central limit theorem 1.3.1. Then, for $n \geq n_0$, the distribution of $\hat{S}_n$ on $\mathbb{R}^r$ has a density that we denote by $f_n(\mathbf{x})$.

Theorem 1.4.1 *For any* $\mathbf{y} \in \mathcal{D}_I^o$ *we have*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(\mathbf{y}) = -I(\mathbf{y}) \ . \tag{1.4.1}$$

Proof.

Again we will prove it for $r = 1$, the generalization is straightforward.

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

Let $y \in \mathcal{D}_I{}^o$. Then by lemma 1.1.3 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

$$\tilde{\nu}(y, dx) = e^{(x+y)\lambda - \mathcal{Z}(\lambda)} \tau_y \nu(dx)$$

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

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

and

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

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

For $n \geq n_0$ 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

$$f_n(x, y) = e^{n((x+y)\lambda - \mathcal{Z}(\lambda))} f_n(x + y) = e^{n(I(y) + \lambda x)} f_n(x + y)$$

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

$$\begin{aligned}\mathbb{E}\left(G((X_1^y + \cdots + X_n^y)/n)\right) &= \int_{\mathbb{R}^n} G(\hat{s}_n) e^{n(I(y)+\lambda \hat{s}_n)} \tau_y \alpha(dx_1) \ldots \tau_y \alpha(dx_n) \\ &= \int_{\mathbb{R}} G(\hat{s}) e^{n(I(y)+\lambda \hat{s})} f_n(\hat{s}+y) d\hat{s}\end{aligned} \tag{1.4.2}$$

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 1.3.1, the result follows immediately. □

1.5 Large deviations probabilities

Corollary 1.5.1 *For any closed set* C

$$\lim_{n\to\infty} \frac{1}{n} \log \int_C f_n(\mathbf{x}) d\mathbf{x} \leq - \inf_{\mathbf{x}\in C} I(\mathbf{x})$$

and for any open set A

$$\lim_{n\to\infty} \frac{1}{n} \log \int_A f_n(\mathbf{x}) d\mathbf{x} \geq - \inf_{\mathbf{x}\in A} I(\mathbf{x})$$

Proof Under the condition of the previous section, let $C \subset \mathbb{R}^r$ a compact set such that $C \cap \mathcal{D}_I^0 \neq \emptyset$. Then it is immediate that

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_C f_n(\mathbf{x})\, d\mathbf{x} \leq - \inf_{x\in C} I(\mathbf{x}). \tag{1.5.1}$$

In fact let $\mathbf{x}_0$ the minimum attained by $I(\cdot)$ on the compact set C (under our assumptions I is continuous). Then

$$\int_C f_n(\mathbf{x})\, d\mathbf{x} \leq e^{-nI(\mathbf{x}_0)} \int_C f_n(0, \mathbf{x})\, d\mathbf{x} = e^{-nI(\mathbf{x}_0)} \sqrt{n} \int_C \tilde{g}_n(0, \mathbf{x})\, d\mathbf{x}$$

and it is not hard to prove that one can prove the local CLT page 6 with enough uniformity such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_C \tilde{g}_n(0, \mathbf{x})\, d\mathbf{x} = 0$$

If C is closed but unbounded, then by Lemma 1.1.2 we have

$$\lim_{k\to\infty}\lim_{n\to\infty}\frac{1}{n}\log\int_{C\cap\{\mathbf{x}>k\}} f_n(\mathbf{x})\, d\mathbf{x} = -\infty$$

and since

$$\log\int_C f_n(\mathbf{x})\, d\mathbf{x} \leq \max\left\{\log\int_{C\cap\{\mathbf{x}\leq k\}} f_n(\mathbf{x})\, d\mathbf{x}, \log\int_{C\cap\{\mathbf{x}>k\}} f_n(\mathbf{x})\, d\mathbf{x}\right\} + \log 2$$

by taking the logarithm and dividing by n, we are reduced to the compact case.

Abouth the lower bound, if A is an open set such that $A \cap \mathcal{D}_I \neq \emptyset$, then take any $x_0 \in A$ such that $I(x_0) < +\infty$. For any $\epsilon > 0$ let $\delta > 0$ such that $I(\mathbf{x}) < I(\mathbf{x}_0)+\epsilon$ if $|\mathbf{x} - \mathbf{x}_0| < \delta$. Then

$$\int_A f_n(\mathbf{x})d\mathbf{x} \geq \int_{J(\mathbf{x}_0,\delta)} f_n(\mathbf{x})d\mathbf{x} \geq e^{-n(I(\mathbf{x}_0)+\epsilon)}\int_{A\cap J(\mathbf{x}_0,\delta)} f_n(\mathbf{0},\mathbf{x})d\mathbf{x}$$

Where $J(\mathbf{x}_0, \delta) = \{|\mathbf{x} - \mathbf{x}_0| < \delta\}$. Since $A \cap J(\mathbf{x}_0, \delta) \neq \emptyset$, it follows that

$$\lim_{n\to\infty}\frac{1}{n}\log\int_A f_n(\mathbf{x})d\mathbf{x} \geq -I(\mathbf{x}_0) - \epsilon.$$

Since ϵ is arbitrarily small at this point, we have obtained the lower bound. □

1.6 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 1.6.1 ***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 1.6.2 ***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. □

Chapter 2

Conditioning

2.1 Conditional measures

Let Ω a d–dimensional manifold, as in the previous chapter, typically $\mathbb{R}^d, \mathbb{S}^d$ etc. Let α a positive σ-finite measure on Ω, typically the corresponding Lebesque measure.

Let $g : \Omega \to \mathbb{R}^r$ a measurable function. We assume that it is not constant and that has compact level sets.

On the product space Ω^n, with the product measure $d\alpha^n = \otimes_j d\alpha(\omega_j)$, we define the function

$$\mathbf{g}^{(n)} = \sum_{j=1}^{n} \mathbf{g}(\omega_j)/n.$$

For any $\mathbf{y} \in \mathbb{R}^r$ consider the set

$$\Sigma_n(\mathbf{y}) = \left\{(\omega_1, \ldots, \omega_n) \in \Omega^n : \mathbf{g}^{(n)} = \mathbf{y}\right\}. \tag{2.1.1}$$

This is bounded set since it is the boundary of a compact set.

We denote the *projection* of α^n on $\Sigma_n(\mathbf{y})$ as the positive measure $d\gamma_n(\omega_1, \ldots, \omega_n; \mathbf{y})$ on $\Sigma_n(\mathbf{y})$ defined by

$$\int_{\Omega^n} F(\mathbf{g}^{(n)}) G(\omega_1, \ldots, \omega_n) d\alpha^n = \int_{\mathbb{R}^r} F(\mathbf{y}) \int_{\Sigma_n(\mathbf{y})} G(\omega_1, \ldots, \omega_n)\, d\gamma_n(\omega_1, \ldots, \omega_n, \mathbf{y}). \tag{2.1.2}$$

for any bounded measurable functions $F : \mathbb{R}^r \to \mathbb{R}, G : \Omega \to \mathbb{R}$. In general $\gamma_n(\cdot, \mathbf{y})$ is a finite (not normalized) measure on $\Sigma_n(\mathbf{y})$. The total volume is given by

$$W_n(\mathbf{y}) = \int_{\Sigma_n(\mathbf{y})} d\gamma(\omega_1, \ldots, \omega_n; \mathbf{y}) \tag{2.1.3}$$

In the case $r = 1$, we have that

$$W_n(y) = \frac{d}{dy} \int_{g^{(n)} \leq y} d\alpha^n(\omega_1, \ldots, \omega_n). \tag{2.1.4}$$

Recall the definition of $Z(\boldsymbol{\lambda})$ and its Legendre transform $Z^*(\mathbf{y})$:

$$Z(\boldsymbol{\lambda}) = \log \int_\Omega e^{\boldsymbol{\lambda} \cdot g(\omega)} d\alpha(\omega), \qquad Z^*(\mathbf{y}) = \sup_{\boldsymbol{\lambda}} \left[\boldsymbol{\lambda} \cdot \mathbf{y} - Z(\boldsymbol{\lambda}) \right].$$

Theorem 2.1.1 *For any* $\mathbf{y} \in \mathcal{D}_{Z^*}$*:*

$$\lim_{n \to \infty} \frac{1}{n} \log W_n(\mathbf{y}) = -\mathcal{Z}^*(\mathbf{y}) \tag{2.1.5}$$

Proof.

Let $\boldsymbol{\lambda} = \nabla \mathcal{Z}^*(\mathbf{y})$, and consider the tilted probability measure $d\alpha_{\boldsymbol{\lambda}} = e^{\boldsymbol{\lambda} \cdot \mathbf{g}(\omega) - \mathcal{Z}(\boldsymbol{\lambda})} d\alpha(\omega)$. Then the product probability measure on Ω^n is given by

$$d\alpha^n_{\boldsymbol{\lambda}} = e^{n[\mathbf{g}^{(n)} \cdot \boldsymbol{\lambda} - \mathcal{Z}(\boldsymbol{\lambda})]} d\alpha^n(\omega_1, \ldots, \omega_n).$$

Then, under the $\alpha^n_{\boldsymbol{\lambda}}$ probability, we can see $g^{(n)}$ as a normalized sum of independent random variables. We denote $f_n(\mathbf{x}, \boldsymbol{\lambda})$ the density of its probability distribution, i.e. for any $F : \mathbb{R}^r \to \mathbb{R}$,

$$\begin{aligned} \int_{\mathbb{R}^r} F(\mathbf{x}) f_n(\mathbf{x}, \boldsymbol{\lambda}) d\mathbf{x} &= \int_{\Omega^n} F(\mathbf{g}^{(n)}) d\alpha^n_{\boldsymbol{\lambda}} \\ &= \int_{\Omega^n} F(\mathbf{g}^{(n)}) e^{n[\mathbf{g}^{(n)} \cdot \boldsymbol{\lambda} - \mathcal{Z}(\boldsymbol{\lambda})]} d\alpha^n \\ &= \int_{\mathbb{R}^r} F(\mathbf{x}) e^{n[\mathbf{x} \cdot \boldsymbol{\lambda} - \mathcal{Z}(\boldsymbol{\lambda})]} W_n(\mathbf{x}) d\mathbf{x} \end{aligned}$$

i.e. $f_n(\mathbf{y}, \boldsymbol{\lambda}) = e^{n[\mathbf{x} \cdot \boldsymbol{\lambda} - \mathcal{Z}(\boldsymbol{\lambda})]} W_n(\mathbf{x})$. Applying proposition 1.4.1, we have that

$$\lim_{n \to \infty} \frac{1}{n} \log f_n(\mathbf{y}, \boldsymbol{\lambda}) = -I_\lambda(\mathbf{y}) = -\left[Z^*(\mathbf{y}) + Z(\boldsymbol{\lambda}) - \boldsymbol{\lambda} \cdot \mathbf{y} \right]$$

and (2.1.5) follows. □

The conditional probability distribution of $\alpha^n_{\boldsymbol{\lambda}}$ on $\Sigma_n(\mathbf{y})$ is defined by

$$\begin{aligned} &\int_{\Omega^n} F(\mathbf{g}^{(n)}) G(\omega_1, \ldots, \omega_n) e^{n[\boldsymbol{\lambda} \mathbf{g}^{(n)} - \mathcal{Z}(\boldsymbol{\lambda})]} d\alpha^n_{\boldsymbol{\lambda}}(\omega_1, \ldots, \omega_n) \\ &\quad = \int_{\mathbb{R}^r} F(\mathbf{y}) f_n(\mathbf{y}, \boldsymbol{\lambda}) \int_{\Sigma_n(\mathbf{y})} G(\omega_1, \ldots, \omega_n) d\alpha^n_{\boldsymbol{\lambda}}(\omega_1, \ldots, \omega_n | y) \end{aligned}$$

Since $f_n(\mathbf{y}) = e^{n[\boldsymbol{\lambda}\cdot\mathbf{y}-\mathcal{Z}(\boldsymbol{\lambda})]}W_n(\mathbf{y})$ we have the relation

$$d\alpha^{(n)}_{\boldsymbol{\lambda}}(\cdot|\mathbf{y}) = d\gamma(\cdot,\mathbf{y})/W_n(\mathbf{y})$$

that in particular imply that it does not depend on $\boldsymbol{\lambda}$[1].

Lemma 2.1.2 *Let F be a bounded continuous function on Ω, and $\mathbf{g} : \Omega \to \mathbb{R}^r$ as above. Then for any $\theta \in \mathbb{R}$*

$$\lim_{n\to\infty}\frac{1}{n}\log\int_{\Sigma_n(\mathbf{y})} e^{\theta nF^{(n)}}d\alpha^{(n)}(\omega_1,\ldots,\omega_n|\mathbf{y}) = -\mathcal{Z}^*_\theta(\mathbf{y}) + Z^*(\mathbf{y}) \tag{2.1.6}$$

where $Z^*_\theta(\mathbf{y})$ is the Fenchel-Legendre transform of $Z_\theta(\boldsymbol{\lambda}) = \log\int_\Omega e^{\theta F+\boldsymbol{\lambda}\cdot\mathbf{g}}d\alpha$:

$$Z^*_\theta(\mathbf{y}) = \sup_{\boldsymbol{\delta}}\left\{\boldsymbol{\delta}\cdot\mathbf{y} - Z_\theta(\boldsymbol{\delta})\right\}.$$

Furthermore

$$\partial_\theta Z^*_\theta(\mathbf{y})\Big|_{\theta=0} = -\int_\Omega F e^{\boldsymbol{\delta}^*\cdot\mathbf{g}-\mathcal{Z}(\boldsymbol{\delta}^*)}d\alpha = -\int_\Omega F d\alpha_{\boldsymbol{\delta}^*(\mathbf{y})} \tag{2.1.7}$$

for $\delta^* = \nabla Z^*(\mathbf{y})$.

Proof:

Consider the doubly tilted probability measure

$$d\alpha_{\boldsymbol{\lambda},\theta}(\omega) = e^{\theta F(\omega)+\boldsymbol{\lambda}\cdot\mathbf{g}(\omega)-Z_\theta(\boldsymbol{\lambda})}d\alpha(\omega).$$

Under the product measure $\alpha^n_{\boldsymbol{\lambda},\theta}$ on Ω^n, the probability distribution of $\mathbf{g}^{(n)}$ is given by:

$$f_n(\mathbf{y};\boldsymbol{\lambda},\theta) = e^{n[\boldsymbol{\lambda}\cdot\mathbf{y}-\mathcal{Z}_\theta(\boldsymbol{\lambda})]}W_n(\mathbf{y})\int_{\Sigma_n(\mathbf{y})} e^{n\theta F^{(n)}}d\alpha^{(n)}(\omega_1,\ldots,\omega_n|\mathbf{y})$$

Then, applying theorem 1.4.1 and (1.2.7) we have

$$\lim_{n\to\infty}\frac{1}{n}\log f_n(\mathbf{y};\boldsymbol{\lambda},\theta) = -I(y;\boldsymbol{\lambda},\theta) = -\mathcal{Z}^*_\theta(\mathbf{y}) + \boldsymbol{\lambda}\cdot\mathbf{y} - \mathcal{Z}_\theta(\boldsymbol{\lambda}).$$

and (2.1.6) follows directly.

About (2.1.7), observe that

$$\mathcal{Z}^*_\theta(\mathbf{y}) = \boldsymbol{\delta}^*\cdot\mathbf{y} - Z_\theta(\boldsymbol{\delta}^*)$$

where $\boldsymbol{\delta}^* = \nabla_{\mathbf{y}}Z^*_\theta(\mathbf{y})$ depends on $\mathbf{y}$ and θ, and $\boldsymbol{y} = \nabla_{\boldsymbol{\delta}}Z_\theta(\boldsymbol{\delta}^*)$. Differentiating in θ we have:

$$\partial_\theta\mathcal{Z}^*_\theta(\mathbf{y}) = \mathbf{y}\cdot\partial_\theta\boldsymbol{\delta}^* - \partial_\theta\mathcal{Z}_\theta(\boldsymbol{\delta}^*) - \nabla_{\boldsymbol{\delta}}Z_\theta(\boldsymbol{\delta}^*)\cdot\partial_\theta\boldsymbol{\delta}^* = -\partial_\theta\mathcal{Z}_\theta(\boldsymbol{\delta}^*) = -\int_\Omega F(\omega)d\alpha_{\boldsymbol{\delta}^*,\theta}(\omega).$$

and (2.1.7) follows after taking $\theta \to 0$. □

[1]In statistics we say that $\mathbf{g}^{(n)}$ is a sufficient statistics for estimating $\boldsymbol{\lambda}$.

2.2 Equivalence of ensembles

In the following $\mathbf{y} \in \mathcal{D}^0_{Z^*}$.

Theorem 2.2.1 *There exists a constant $C > 0$ such that for any $\epsilon > 0$*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\Sigma_n(\mathbf{y})} \mathbf{1}_{\left[\left|F^{(n)} - \int F d\alpha_{\boldsymbol{\lambda}}\right| \geq \epsilon\right]} d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) \leq -C\epsilon^2. \tag{2.2.1}$$

Proof:

Without loosing any generality we can here assume that $\int F d\alpha_{\boldsymbol{\lambda}} = 0$. Consequently $Z^*_\theta(\mathbf{y}) - Z^*(\mathbf{y}) = O(\theta^2)$. Then for any $\theta > 0$ by exponential Chebichef inequality:

$$\begin{aligned}\int_{\Sigma_n(\mathbf{y})} \mathbf{1}_{\left[\left|F^{(n)}\right| \geq \epsilon\right]} d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) &\leq e^{-n\theta\epsilon} \int_{\Sigma_n(\mathbf{y})} e^{\theta\left|nF^{(n)}\right|} d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) \\ &\leq e^{-n\theta\epsilon} \int_{\Sigma_n(\mathbf{y})} \left[e^{\theta n F^{(n)}} + e^{-\theta n F^{(n)}}\right] d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y})\end{aligned}$$

and by (2.1.6)

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{n} \log \int_{\Sigma_n(\mathbf{y})} \mathbf{1}_{\left[\left|F^{(n)}\right| \geq \epsilon\right]} d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) \leq \\ &\qquad - \theta\epsilon + \max\{-Z^*_\theta(\mathbf{y}) + Z^*(\mathbf{y}), -Z^*_{-\theta}(\mathbf{y}) + Z^*(\mathbf{y})\}\end{aligned}$$

and since $Z^*_\theta(\mathbf{y}) = O(\theta^2)$, we can choose θ such that the right hand side is bounded by $-C\epsilon^2$ for some positive constant C. □

Since $d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y})$ is a symmetric measure:

$$\int_{\Sigma_n(\mathbf{y})} F(\omega_1) d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) = \int_{\Sigma_n(\mathbf{y})} F^{(n)} d\alpha^{(n)}(\omega_1, \ldots, \omega_n|\mathbf{y}) \underset{n\to\infty}{\longrightarrow} \int_\Omega F d\alpha_{\boldsymbol{\lambda}}$$

More generally we have

Theorem 2.2.2 *Let $F(\omega_1, \ldots, \omega_k)$ a bounded continuous function on Ω^k and $y \in \mathcal{D}^o_{\mathcal{Z}^*}$, then*

$$\lim_{n\to\infty} \int_{\Sigma_n(\mathbf{y})} F(\omega_1, \ldots, \omega_k) d\alpha^{(n)}(\omega_1, \ldots, d\omega_n|\mathbf{y}) = \int_{\Omega^k} F(\omega_1, \ldots, \omega_k) \alpha_{\boldsymbol{\lambda}}(d\omega_1) \ldots \alpha_{\boldsymbol{\lambda}}(d\omega_k)$$

Proof. It is enough to consider functions of the form $F(\omega_1, \ldots, \omega_k) = F_1(\omega_1) \ldots F_k(\omega_k)$. For simplicity let us prove the case $k = 2$, the generalization to any k is straightforward. Without loosing generality, let us assume that $\int F_j(\omega)\alpha_{\boldsymbol{\lambda}}(d\omega) = 0$. By the exchange symmetry of $d\alpha^{(n)}(\cdot|\mathbf{y})$ we have

$$\int F_1(\omega_1)F_2(\omega_2)\alpha^{(n)}(d\omega_1, \ldots, d\omega_n|\mathbf{y}) = \int \frac{1}{n(n-1)} \sum_{i \neq j} F_1(\omega_i)F_2(\omega_j)\alpha^{(n)}(d\omega_1, \ldots, d\omega_n|\mathbf{y})$$

$$= \int \frac{n^2}{n(n-1)} F_1^{(n)} F_2^{(n)} \alpha^{(n)}(d\omega_1, \ldots, d\omega_n|\mathbf{y}) + O\left(\frac{1}{n}\right)$$

and this last expression converges to 0 an $n \to \infty$ by (2.2.1) . □

Examples

1. Choose $\Omega = \mathbb{R}$ and $g(x) = x^2$. It follows from the above equivalence the so called Poincare lemma[2]: the uniform measure on the n-dimensional sphere with radius $\sqrt{n}$ converges, in terms of the finite dimensional distributions, to the product of gaussian measures $e^{-x_i^2/2}/\sqrt{2\pi}$.

[2] Apparently Poincare has nothing to do with such statement, that should be attributed to Maxwell.

Chapter 3

Statistical mechanics and thermodynamics of one dimensional chain of oscillators

3.1 Grand canonical formalism

We study a system of $m = [nM]$ anharmonic oscillators, where $M > 0$ is a positive parameter corresponding to the macroscopic mass of the total system. The particles are denoted by $j = 1, \dots m$. We denote with $q_j, j = 1, \dots, m$ 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 *point*, 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

$$Z(\lambda, \beta) := \int e^{-\beta V(r) + \lambda r} dr < +\infty \tag{3.1.1}$$

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 $\{r_j = q_j - q_{j-1} - a,\ j = 1, \dots, m\}$. Without loosing any generality, we will choose $a = 0$ for the sequence.

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

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

where

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

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 m is given by $q_n = \sum_{j=1}^m r_j$. We consider the Hamiltonian dynamics:

$$\begin{aligned} \dot{r}_j(t) &= p_j(t) - p_{j-1}(t), \qquad j = 1, \dots, m, \\ \dot{p}_j(t) &= V'(r_{j+1}(t)) - V'(r_j(t)), \qquad j = 1, \dots, m-1, \\ \dot{p}_m(t) &= \tau - V'(r_m(t)), \end{aligned} \tag{3.1.2}$$

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

$$d\mu_{\tau,\beta}^{n,gc} = \prod_{j=1}^m \frac{e^{-\beta(\mathcal{E}_j - \tau r_j)}}{\sqrt{2\pi\beta^{-1}} Z(\beta\tau, \beta)} dr_j dp_j \tag{3.1.3}$$

is stationary for this dynamics.

The distribution $\mu_{\tau,\beta}^{m,gc}$ is called grand canonical Gibbs measure at temperature $T = \beta^{-1}$ and tension (or pressure) τ. Notice that $\{r_1, \dots, r_m, p_1, \dots, p_m\}$ are independently distributed under this probability measure.

We can apply the result of chapters 1 and 2, with $\Omega = \mathbb{R}^2$, $\mathbf{g}(r,p) = \mathcal{E}(p,r)$, $\boldsymbol{\lambda} = (-\beta, -\tau/\beta)$, $\mathcal{Z}(\boldsymbol{\lambda}) = Z(\beta\tau, \beta)\sqrt{\frac{2\pi}{\beta}}$. Then, for $M > 0, U > 0, \mathcal{L} \in \mathbb{R}$, we have the microcanonical surface

$$\begin{aligned} \tilde{\Sigma}_m(M, MU, M\mathcal{L}) &:= \left\{ (r_1, p_1, \dots, r_m, p_m) : \frac{1}{n}\sum_{j=1}^m \mathcal{E}_j = MU, \frac{1}{n}\sum_{j=1}^m r_j = M\mathcal{L} \right\} \\ &= \Sigma_m(U, \mathcal{L}) = \left\{ (r_1, p_1, \dots, r_m, p_m) : \mathcal{E}^{(m)} = U, r^{(m)} = \mathcal{L} \right\}. \end{aligned} \tag{3.1.4}$$

Defining $W_m(U, \mathcal{L})$ as in (2.1.3), By theorem (2.1.5) the following limit exists

$$S(M, MU, M\mathcal{L}) := \lim_{n\to\infty} \frac{1}{n} \log W_m(U, \mathcal{L}) = MS(1, U, \mathcal{L}) \tag{3.1.5}$$

In fact by the definition follows that S is homogeneous of degree 1. In the following we just use the notation $S(U, \mathcal{L}) = S(1, U, \mathcal{L})$. This function is the *thermodynamic entropy* that by (2.1.5)

$$S(U, \mathcal{L}) = \inf_{\lambda,\beta>0} \left\{ -\lambda\mathcal{L} + \beta U - \log\left(Z(\lambda, \beta)\sqrt{2\pi\beta^{-1}} \right) \right\}. \tag{3.1.6}$$

This is the fundamental relation that connects the microscopic system to its thermodynamic macroscopic description.

The limit in (3.1.5) is intended for all values of the internal energy $U > 0$.

S is concave, since inf of linear functions.

We can now define the other thermodynamic quantities from the entropy definition (3.1.6). From equation (3.1.6) we have

$$\lambda(\mathcal{L}, u) = -\frac{\partial S(\mathcal{L}, u)}{\partial \mathcal{L}}, \qquad \beta(\mathcal{L}, u) = \frac{\partial S(\mathcal{L}, u)}{\partial u} \tag{3.1.7}$$

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

$$\begin{aligned} \mathcal{L}(\lambda, \beta) &= \frac{\partial \log Z(\lambda, \beta)}{\partial \lambda} = \int r \, \frac{e^{\lambda r - \beta V(r)}}{Z(\lambda, \beta)} \, dr = \int r_j \, d\mu^{gc}_{\tau,\beta} \\ u(\lambda, \beta) &= -\frac{\partial \log \left(Z(\lambda, \beta) \sqrt{2\pi/\beta} \right)}{\partial \beta} = \int V(r) \frac{e^{\lambda r - \beta V(r)}}{Z(\lambda, \beta)} \, dr + \frac{1}{2\beta} = \int \mathcal{E}_j \, d\mu^{gc}_{\tau,\beta} \end{aligned} \tag{3.1.8}$$

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

$$dS = -\beta \tau d\mathcal{L} + \beta du = \frac{đQ}{T} \tag{3.1.9}$$

where $đQ$ is the (non-exact) differential

$$đQ = -\tau d\mathcal{L} + du. \tag{3.1.10}$$

3.2 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 constrained dynamics is

$$\begin{aligned} \dot{r}_j(t) &= p_j(t) - p_{j-1}(t), \qquad j = 1, \ldots, n-1, \\ \dot{p}_j(t) &= V'(r_{j+1}(t)) - V'(r_j(t)), \qquad j = 1, \ldots, n-1, \\ r_n(t) &= nr - \sum_{j=1}^{n-1} r_j(t) \; . \end{aligned} \tag{3.2.1}$$

The dynamics now is conserving the total energy $\mathcal{H} = \sum_j \mathcal{E}_j = nu$ and the total length $\sum_{j=1}^n r_j = nr$. The microcanonical measures $\mu^{n,mc}_{r,u}$ 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 $d\mu^{n,gc}_{\tau_0,\beta_0}$. Fix $\mathbf{x} = (r, u)$, and define $\mu^{n,mc}_{\mathbf{x}}$ 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

$$\begin{aligned} &\int G(\hat{\mathbf{S}}_n) H(r_1, p_1, \dots, r_n, p_n)\, d\mu^{n,gc}_{\tau_0,\beta_0}(r_1, p_1, \dots, r_n, p_n) \\ &= \int_{\mathbb{R}\times\mathbb{R}_+} d\mathbf{x}\ G(\mathbf{x}) f_n(\mathbf{x}) \int H(r_1, p_1, \dots, r_n, p_n)\, d\mu^{n,mc}_{\mathbf{x}} \end{aligned}$$

where $\hat{\mathbf{S}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$. It is easy to see that $\mu^{n,mc}_{\mathbf{x}}$ does not depend on τ_0, β_0. We call $\mu^{n,mc}_{\mathbf{x}}$ the *microcanonical measure.*

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

Theorem 3.2.1 *Given* $\mathbf{x} = (r, u)$*, let*

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

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

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

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

Theorem 3.2.2 *Under the same conditions of Theorem 3.2.1, assume that*

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

Then

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

The proof of these two theorems follows the argument used for Theorems 2.2.2.

3.3 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 $\mu_{r,\beta}^{n,c}$ as stationary measure:

$$\begin{aligned} \dot{r}_j(t) &= p_j(t) - p_{j-1}(t), \qquad j = 1, \dots, n-1, \\ dp_j(t) &= (V'(r_{j+1}(t)) - V'(r_j(t)))\, dt \\ &\qquad\qquad + \delta_{j,n-1} \left(-p_j(t) dt + \sqrt{\beta} dw(t) \right), \qquad j = 1, \dots, n-1, \\ r_n(t) &= nr - \sum_{j=1}^{n-1} r_j(t)\ . \end{aligned} \tag{3.3.1}$$

This is defined as follows:

If we condition the grand canonical measure $\mu_{0,0,\beta}^{n,gc}$ 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 $\mu_{r,\beta}^{n,c}$. We can formally write

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

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

Similar statements as theorems 3.2.1 and 3.2.2 holds, $\mu_{r,\beta}^{n,c}$ converging to the grand-canonical measure $\mu_{\tau,\beta}^{n,gc}$, with τ given by the thermodynamic relations (3.1.7).

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 $\mu_{\tau,\beta}^{gc}$ as stationary measure.

3.4 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 3.4.1 *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* $y \in (0,1)$*,*

$$\lim_{n\to\infty} \mu_n\big|_{([ny],[ny]+k)} = \mu^{k,gc}_{\tau(y),\beta(y)} \tag{3.4.1}$$

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:

$$\prod_{j=1}^{n} \frac{e^{-\beta(j/n)(\mathcal{E}_j - \tau(j/n) r_j)}}{\sqrt{2\pi\beta(j/n)^{-1}}Z(\beta(j/n)\tau(j/n), \beta(j/n))} dr_j dp_j = g^n_{\tau(\cdot),\beta(\cdot)} \prod_{j=1}^{n} dr_j dp_j \tag{3.4.2}$$

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

$$e^{\sum_j F_j(r_{j-h},p_{j-h},\dots,r_{j+h},p_{j+h})/n} g^n_{\tau(\cdot),\beta(\cdot)} \prod_{j=1}^{n} dr_j dp_j \tag{3.4.3}$$

where F_j are local functions.

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

$$S(r(\cdot), u(\cdot)) = \int_0^1 S(r(y), u(y))\, dy \tag{3.4.4}$$

where $r(y), u(y)$ are computed from $\tau(y), \beta(y)$ using (3.1.8).

Bibliography

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