Large Deviations

Stefano Olla

April 3, 2008

Contents

Chapter 1

Large Deviations

1.1 Introduction

As Dembo and Zeitouini 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 $\hat{S}_n = \frac{1}{n}\sum_n X_i$ the empirical sum. The weak law of large numbers says that for any $\delta > 0$,

$$\mathbb{P}(|\hat{S}_n| \geq \delta) \underset{n\to\infty}{\longrightarrow} 0 \tag{1.1.1}$$

The central limit theorem is a refinement that says

$$\mathbb{P}(\sqrt{n}\hat{S}_n \in [a,b]) \underset{n\to\infty}{\longrightarrow} \frac{1}{2\pi}\int_a^b e^{-x^2/2}dx\ . \tag{1.1.2}$$

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

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

therefore (**exercise**)

$$\frac{1}{n}\log\mathbb{P}(|\hat{S}_n| \geq \delta) \underset{n\to\infty}{\longrightarrow} -\frac{\delta^2}{2} \tag{1.1.3}$$

Equation (1.1.3) is an example of a large deviation statement.

1.2 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

$$M(\lambda) = \mathbb{E}\left[e^{\lambda X_1}\right] \tag{1.2.1}$$

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*

$$\Lambda(\lambda) = \log \mathbb{E}\left[e^{\lambda X_1}\right] \tag{1.2.2}$$

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

Lemma 1.2.1 *1. $\Lambda(\cdot)$ is convex.*

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

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

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

$$\Lambda(\alpha\lambda_1 + (1-\alpha)\lambda_2) \leq \alpha\Lambda(\lambda_1) + (1-\alpha)\Lambda(\lambda_2)$$

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 $\lambda \in \mathcal{D}_\Lambda^o$, 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 wone can prove that $\Lambda(\cdot) \in \mathcal{C}^\infty(\mathcal{D}_\Lambda^o)$. Computing the second derivative we obtain

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

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

We define the rate function as the Fenchel-Legendre transform of Λ

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} \tag{1.2.3}$$

It is immediate to see that I is convex (as supremum of linear functions) 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

$$\lambda m - \Lambda(\lambda) \leq 0$$

and it is $= 0$ 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 $\lambda < 0$

$$\lambda x - \Lambda(\lambda) \leq \lambda m - \Lambda(\lambda)$$

that implies

$$I(x) = \sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\} \qquad x > m \tag{1.2.4}$$

Similarly one obtains

$$I(x) = \sup_{\lambda \leq 0} \{\lambda x - \Lambda(\lambda)\} \qquad x < m \tag{1.2.5}$$

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

Lemma 1.2.2 *$I(x) \to +\infty$ as $|x| \to \infty$, and its level sets are compact.*

Proof: If $x > m$, for any $0 < \lambda \in \mathcal{D}_\Lambda$,

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

and $\lim_{x \to +\infty} \Lambda(\lambda)/x = 0$, 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 1.2.3 *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.

1.2.1 Properties of Legendre transforms

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

Lemma 1.2.4

$I \in C^\infty(\mathcal{D}_I)$ and strictly convex in $\mathcal{D}_{\Lambda^*}$. Furthermore for any $\bar{x} \in \mathcal{D}_I$ there exists a unique $\bar{\lambda} \in \mathcal{D}_\Lambda$ such that

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

and

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

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

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

Proof: The function $F_x(\lambda) = \lambda x - \Lambda(\lambda)$ has a maximum for $\lambda = \bar{\lambda}$. In fact $\partial_\lambda F_x(\bar{\lambda}) = 0$ and $\partial^2_\lambda F_x(\bar{\lambda}) < 0$. It follows that $I(\bar{x}) = \bar{\lambda}\bar{x} - \Lambda(\bar{\lambda})$ and that $\Lambda(\lambda) = \sup_x \{\lambda x - I(x)\}$. By the same argument $G_\lambda(x) = \lambda x - I(x)$ is maximized by $\bar{x}$. □

1.2.2 Proof of Cramer's theorem

Upper bound

Let us start with 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 $x > m$

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \geq x) \leq -\sup_{\lambda>0}\{\lambda x - \Lambda(\lambda)\} = I(x) \qquad (1.2.6)$$

where we use (1.2.4) in the last equality. Similarly for $x < m$ we obtain

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \leq x) \leq -\sup_{\lambda<0}\{\lambda x - \Lambda(\lambda)\} = I(x) \qquad (1.2.7)$$

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) 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 $C \subseteq (-\infty, x_1]$). 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 (1.2.6) and (1.2.7)

$$\frac{1}{n}\log \mathbb{P}(\hat{S}_n \in C) \leq -\min\{I(x_2), I(x_1)\} + \frac{1}{n}\log 2 \tag{1.2.8}$$

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 $I(x) < \infty$. This implies there exists a unique λ_0 such that

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

Assuming $x > m$, we have that $\lambda_o > 0$.

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

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

Notice that

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

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

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

If $x < m$, we have $\lambda_o < 0$, and in the last step of the above we will have $x - \delta_1$ instead.

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

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

so that

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

Since $\delta_1 < \delta$ is arbitrary, we can let $\delta \to 0$, and this finish the proof of the lower bound.

Remark 1.2.5 *Notice that the proof contains the non-asymptotic bound* (1.2.8), *i.e.*

$$\mathbb{P}(\hat{S}_n \in C) \leq 2e^{-n \inf_{x\in C} I(x)} \qquad \forall n \tag{1.2.9}$$

also called Chernoff's bound.

Remark 1.2.6 *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 1.2.7 *For any $x > m$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \geq x) &= -I(x) \qquad \text{if } x > m \\ \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{S}_n \leq x) &= -I(x) \qquad \text{if } x < m \end{aligned} \tag{1.2.10}$$

Proof: By the central limit theorem

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

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 1.2.3.

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

$$I(x) = \qquad \text{if } x > 0$$

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

1.3 Cramér's Theorem in $\mathbb{R}^d$

Let $\{\mathbf{X}_n\}$ 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

$$M(\mathbf{u}) = \int_{\mathbb{R}^d} e^{\mathbf{u}\cdot\mathbf{x}}\, \alpha(d\mathbf{x}), \qquad \Lambda(\mathbf{u}) = \log M(\mathbf{u}) \tag{1.3.1}$$

as before we denote $\mathcal{D}_\Lambda = \{\mathbf{u} \in \ R^d : \Lambda(\mathbf{u})\} < \infty$ and we assume that $0 \in \mathcal{D}_\Lambda^o$. 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 Λ:

$$I(\mathbf{x}) = \sup_{\mathbf{u}\in\mathbb{R}^d} \{\mathbf{u}\cdot\mathbf{x} - \Lambda(\mathbf{u})\} \tag{1.3.2}$$

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 ne dimensional case:

Lemma 1.3.1 *$I(\mathbf{x}) \in \mathcal{C}^\infty(\mathcal{D}_I)^o$, and $\mathbf{m} \in (\mathcal{D}_I)^o$. There exists a diffeomorphism between $(\mathcal{D}_I)^o$ and $(\mathcal{D}_\lambda)^o$ defined by*

$$\mathbf{u}^* = (\nabla\Lambda)(\mathbf{u}), \quad \mathbf{u} = \nabla I(\mathbf{u}^*) \tag{1.3.3}$$

and

$$(Hess\Lambda)(\mathbf{u}) = [(HessI)(\mathbf{u}^*)]^{-1} \tag{1.3.4}$$

Theorem 1.3.2 *For any 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 the same way as in $d = 1$. Consider $\mathbf{u}^*$ such that $I(\mathbf{u}^*) < +\infty$. Then there exists a unique $\mathbf{u}$ such that

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

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

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

Observe that

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

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

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

The law of large numbers now says that

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

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

$$\frac{1}{n} \log \mathbb{P}\left(\hat{\mathbf{S}}_n \in C\right) \leq -\sup_{\mathbf{u}\in\mathbb{R}^d} \inf_{\mathbf{x}\in C} [\mathbf{u}\cdot\mathbf{x} - \Lambda(\mathbf{u})] \tag{1.3.5}$$

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 1.3.3 *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*

$$\inf_{\mathbf{x}\in C} \sup_{\mathbf{u}\in\mathbb{R}^d} g(\mathbf{u}, \mathbf{x}) = \sup_{\mathbf{u}\in\mathbb{R}^d} \inf_{\mathbf{x}\in C} g(\mathbf{u}, \mathbf{x}) \tag{1.3.6}$$

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.

$$\lim_{\rho\to\infty} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\hat{\mathbf{S}}_n \notin H_\rho\right) = -\infty \tag{1.3.7}$$

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 (1.3.7) follows by applying lemma 1.2.2.

□

1.4 Generalities on Large Deviations

Let X a complete separable metric space (we will be interetsed only to application in finite dimensions), 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 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 $l < \infty$ the set $\{x : I(x) \leq l\}$ 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 $G \subset X$

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G} I(x).$$

Theorem 1.4.1 ***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 $\delta>0$, we 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. Since $F-I$ is upper semicontinuous, for any $\delta>0$ we can find $y\in X$ such that $F(y)-I(y)\ge\sup_x[F(x)-I(x)]-\delta/2$. Correspondingly 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

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

and we conclude from the arbitrariness of δ. □

Theorem 1.4.2 ***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*

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

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. □

1.5 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 1.5.1 ***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 $\tilde{g}(x)_n$ the probability density of $(X_1 + \cdots + X_n)/\sqrt{n}$, where X_j are i.i.d. with common law α. Then*

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

Proof. The characteristic function of α is defined by

$$\phi(k) = \int e^{i2\pi xk} \alpha(dx) \tag{1.5.1}$$

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. [4], theorem XV.3.3). Then

$$\tilde{g}(x)_n = \int_{-\infty}^{+\infty} e^{-i2\pi xk} \left[\hat{\phi}\left(\frac{k}{\sqrt{n}}\right)\right]^n dk$$

and therefore

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

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

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

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

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

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 $k \neq 1$, 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 $\phi(k) < 1$ if $k \neq 1$ 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 1.2, and furthermore its characteristic function satisfies conditions of the local central limit theorem 1.5.1. 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 1.5.2 *For any $y \in \mathbb{R}$ such that $I(y) < +\infty$*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(y) = -I(y) \ . \tag{1.5.2}$$

Proof.

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

Let $y \in \mathbb{R}$ such that $I(y) < +\infty$. Then by lemma 1.2.4 there exists a unique $\lambda \in \mathcal{D}_\Lambda$ such that $y = \Lambda'(\lambda)$, $\lambda = I'(y)$, and $I(y) = \lambda y - \Lambda(\lambda)$. Define

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

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

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

and

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

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

For $n \geq r$ it exists the density for the distribution of $(X_1^y + \dots + X_n^y)/n$ 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 see this fact compute, for a given bounded measurable function $G(\cdot)$:

$$\begin{aligned}\mathbb{E}\left(G((X_1^y + \dots + X_n^y)/n)\right) &= \int_{\mathbb{R}^n} G(\hat{s}_n) e^{n(I(y)+\lambda\hat{s}_n)} f(x_1+y)\dots f(x_n+y)dx_1,\dots dx_n \\ &= \int_{\mathbb{R}} G(\hat{s}) e^{n(I(y)+\lambda\hat{s})} f_n(\hat{s}+y)d\hat{s}\end{aligned} \tag{1.5.3}$$

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 $(X_1^y + \dots + X_n^y)/\sqrt{n}$. Then $f_n(x,y) = \sqrt{n}\tilde{f}_n(\sqrt{n}x, y)$. By the local central limit theorem 1.5.1, the result follows immediately. □

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

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

Lemma 1.5.3 *For every $\theta \in \mathbb{R}$, the limit*

$$\lim_{n\to\infty} \frac{1}{n} \log \int e^{\theta(F(x_1)+\dots+F(x_n))}\nu_y^{(n)}(dx_1,\dots,dx_n) = G(\theta,y) \tag{1.5.4}$$

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

$$\left.\frac{\partial G(\theta,y)}{\partial\theta}\right|_{\theta=0} = \int F(x)\alpha_y(dx). \tag{1.5.5}$$

Proof.

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

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

Note that, by (1.5.3),

$$\int e^{\theta(F(x_1)+....+F(x_n))}\nu_y^{(n)}(dx_1,\ldots,dx_n) = \frac{H_n(\theta,y)}{f_n(y)} \tag{1.5.6}$$

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

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

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 $a(\theta)^{-n}H_n(\theta,y)$. Then by applying 1.5.2 to this law we obtain

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

Consequently we have, applying again 1.5.2

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

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

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

In order t compute this last expression let us set $\lambda^*(y,\theta) = \partial_y I(y,\theta)$, so that

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

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

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

So we have

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

and sending $\theta\to 0$ we obtain

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

□

Theorem 1.5.4 *For any $y \in R$, and any $\epsilon > 0$*

$$\lim_{n\to\infty} \nu_y^{(n)} \left(\left| \frac{1}{n} \sum_{j=1}^n F(x_j) - \int F(x)\alpha_y(dx) \right| \geq \epsilon \right) = 0 \tag{1.5.7}$$

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

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

and by (1.5.4)

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

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

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

Theorem 1.5.5 *Let $F(x_1, \dots, x_k)$ a bounded continuous function on $\mathbb{R}^k$, then*

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

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 $\int F_j(x)\alpha_y(dx) = 0$. By the exchange symmetry of $\nu_y^{(n)}$ we have

$$\begin{aligned}\int F_1(x_1)F_2(x_2)\nu_y^{(n)}(dx_1, \dots, dx_n) &= \int \frac{1}{n(n-1)} \sum_{i\neq j} F_1(x_i)F_2(x_j)\nu_y^{(n)}(dx_1, \dots, dx_n) \\ &= \int \frac{n^2}{n(n-1)} \left(\frac{1}{n} \sum_i F_1(x_i) \right) \left(\frac{1}{n} \sum_j F_2(x_j) \right) \nu_y^{(n)}(dx_1, \dots, dx_n) + O\left(\frac{1}{n}\right)\end{aligned}$$

and this last expression converges to 0 an $n \to \infty$ by (1.5.7) .
□

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 1.3. 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 1.5.6 *For any $\mathbf{y} \in \mathbb{R}^d$ such that $I(\mathbf{y}) < +\infty$*

$$\lim_{n\to\infty} \frac{1}{n} \log f_n(\mathbf{y}) = -I(\mathbf{y}) \ . \tag{1.5.8}$$

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 $\mathbb{R}$)

$$f_{\lambda,\beta}(y) = \frac{e^{-\beta V(y)+\lambda y}}{Z(\lambda, \beta)}$$

Let $\{Y_j\}$ a suite of i.i.d. rv with common law given by $f_{0,\beta}(y)dy$. Then the vector valued random variables $\mathbf{X}_j = (Y_j, V(Y_j))$ clearly has a law $\alpha(d\mathbf{x})$ in $\mathbb{R}^2$ that is degenerate, but $\alpha * \alpha$ has already a density. Its logarithmic moment generating function is given by

$$\Lambda(\lambda, \eta) = \log \int e^{\lambda y + \eta V(y)} f_{0,\beta}(y) dy = \log \frac{Z(\lambda, \beta - \eta)}{Z(0, \beta)}$$

for $\eta < \beta$. The corresponding Legendre transform, for $\lambda \in \mathbb{R}$ and $e > o$, is given by

$$\begin{aligned} I(r, e) &= \sup_{\eta<\beta,\lambda} \{\lambda r + \eta e - \log \Lambda(\lambda, \eta)\} \\ &= \sup_{\beta'>0,\lambda} \{\lambda r - \beta' e - \log Z(\lambda, \beta')\} + \beta e + \log Z(0, \beta) \end{aligned}$$

The function defined by

$$S(r, e) = \inf_{\lambda,\beta'>o} \{-\lambda r + \beta' e - \log Z(\lambda, \beta')\} \tag{1.5.9}$$

is called *thermodynamic entropy*. So we have obtained $I(r, e) = -S(r, e) + \beta e + \log Z(\lambda, \beta)$. Observe that S does not depend on β.

For $n \geq 2$, the density of the distribution of $(\mathbf{X}_1 + \cdots + \mathbf{X}_n)/n$ is given by

$$\begin{aligned} f_n(r,e) &= \int_{\mathbb{R}^n} \frac{e^{-\beta \sum_{j=1}^n V(y_j)}}{Z(0,\beta)^n} \delta\left(\frac{1}{n}\sum_{j=1}^n V(y_j) - e; \frac{1}{n}\sum_{j=1}^n y_j - r\right) dy_1 \ldots dy_n \\ &= \frac{e^{-n\beta e}}{Z(0,\beta)^n} \int_{\mathbb{R}^n} \delta\left(\frac{1}{n}\sum_{j=1}^n V(y_j) - e; \frac{1}{n}\sum_{j=1}^n y_j - r\right) dy_1 \ldots dy_n \\ &= \frac{e^{-n\beta e}}{Z(0,\beta)^n} \Gamma_n(r,e) \end{aligned}$$

where $\Gamma_n(e,r)$ is the volume of the corresponding $n-2$ dimensional surface on $\mathbb{R}^n$. So by applying (1.5.8) we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log \Gamma_n(r,e) = S(r,e) \tag{1.5.10}$$

1.6 Applications to statistical mechanics: thermodynamics

We consider here a system of anharmonic oscillators. The particles are denoted by $j = 0, 1, \ldots n$, $\{q_j, j = 0, \ldots, n\}$ are their position, and $\{r_j = q_{j+1} - q_j, j = 1, \ldots, n\}$ is the distance between subsequent particles. To each particle is also assigned a momentum (which is equal to its velocity since we assume here that all particles have the same mass) v_x. We look at this system from a reference point solidal to the particle $x = 0$, so we define the relative velocity $p_j = v_j - v_0$. Equivalently we can think that the particle $x = 0$ is attached to a wall, so that $q_0 = v_0 = 0$. The configuration of the system is then given by $\{r_j, p_j, j = 1, \ldots, n\} \in (\mathbb{R} \times \mathbb{R})^n$.

There is an energy function defined on each configuration as

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

where

$$e_j = \frac{1}{2} p_j^2 + V(r_j)$$

is the energy of each oscillator. Here V is a positive smooth function such that $V(r) \to +\infty$ as $|r| \to \infty$ and such that $Z(\lambda,\beta) = \int e^{-\beta V(r) + \lambda r} dr < +\infty$ for all $\beta > 0$ and all $\lambda \in \mathbb{R}$.

For any $\beta > 0$, $\bar{p} \in R$, $T \in \mathbb{R}$, consider the probability measure on the configuration space given by

$$d\mu^{gc}_{\tau,\bar{p},\beta} = \frac{e^{-\beta \sum_x (e_x - \bar{p} p_x - \tau r_x)}}{\mathcal{Z}(\beta\tau, \beta\bar{p}, \beta)} \prod_{x=1}^n dr_x dp_x \tag{1.6.1}$$

here $\mathcal{Z}$ is the normalization factor defined by

$$\mathcal{Z}(\lambda,\zeta,\beta) = \int e^{\lambda r+\zeta p-\beta(p^2/2+V(r))} dr\ dp = Z(\lambda,\beta)\sqrt{\frac{2\pi}{\beta}} e^{\zeta^2/(2\beta)}$$

The distribution $\mu^{gc}_{\tau,\bar{p},\beta}$ is called gran-canonical Gibbs measure at temperature $T=\beta^{-1}$, velocity $\bar{p}$ and tension (or pressure) τ, and it is a product measure.

Define the random vectors $\mathbf{X}_j = (r_j,p_j,e_j) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$, where (r_j,p_j) are i.i.d. distributed by $\mu^{gc}_{0,0,\beta}$. We also define the vector $\mathbf{v} = (\beta\tau, \beta\bar{p}, -\beta)$. Under $d\mu^{gc}_{\tau,\bar{p},\beta}$, $\{\mathbf{X}_j\}_{j=1}^n$ are i.i.d. with law given by

$$\alpha(d\mathbf{x}) = \frac{e^{\mathbf{v}\cdot\mathbf{x}}}{\mathcal{Z}(\mathbf{v})}\delta\left(e - V(r) - p^2/2\right) dr\ dp\ de$$

This is just a formal writing, since energy is a function of r and p, this law does not have a density in $\mathbb{R}^3$.

The logarithmic moment generating function of α is given now by

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

We define by $I(\mathbf{x}) = I(r,p,e)$ the Legendre transform of $\Lambda(\mathbf{v}) = \log\mathcal{Z}(\mathbf{v})$

$$\begin{aligned}
I(\mathbf{x}) &= \sup_{\mathbf{v}}\{\mathbf{v}\cdot\mathbf{x} - \Lambda(\mathbf{v})\} = \sup_{\lambda,\zeta,\eta<\beta}\{\lambda r+\zeta p+\eta e - \Lambda(\lambda,\zeta,\eta)\} \\
&= \sup_{\lambda,\zeta,\beta'>0}\{\lambda r+\zeta p-\beta' e - \log\mathcal{Z}(\lambda,\zeta,\beta')\} + \beta e + \log\mathcal{Z}(0,0,\beta) \\
&= \sup_{\lambda,\zeta,\beta'>0}\left\{\lambda r+\zeta p-\beta' e-\log Z(\lambda,\beta') - \frac{1}{2}\left(\zeta^2/\beta' + \log\frac{2\pi}{\beta'}\right)\right\} + \beta e + \log Z(0,\beta) + \frac{1}{2}\log\frac{2\pi}{\beta} \\
&= \sup_{\lambda,\beta'>0}\left\{\lambda r - \beta'\left(e-\frac{p^2}{2}\right) - \log\left(Z(\lambda,\beta')\sqrt{\beta'}\right)\right\} + \beta e + \log\left(Z(0,\beta)\sqrt{\beta}\right)
\end{aligned}$$

The function

$$S(r,e) = \sup_{\lambda,\beta'>0}\left\{\lambda r - \beta' e - \log\left(Z(\lambda,\beta')\sqrt{\beta'/2\pi}\right)\right\} \tag{1.6.2}$$

is called thermodynamic entropy. So we have obtained that

$$I(r,p,e) = S(r,e-p^2/2) + \beta e + \log\left(Z(0,\beta)\sqrt{\beta/2\pi}\right)$$

For $n\geq 2$ the density of the distribution of $\frac{1}{n}\sum_{j=1}^n \mathbf{X}_j$ under $\mu^{gc}_{0,0,\beta}$ is given by

$$\begin{aligned}
&f_n(r,p,\beta) \\
&= \int_{\mathbb{R}^{2n}} \frac{e^{-\beta\sum_j e_j}}{Z(0,\beta)^n(2\pi\beta^{-1})^{n/2}}\delta\left(\frac{1}{n}\sum_{j=1}^n e_j - e; \frac{1}{n}\sum_{j=1}^n p_j - p; \frac{1}{n}\sum_{j=1}^n r_j - r\right) dr_1 dp_1\dots dr_n dp_n \\
&\qquad = \frac{e^{-n\beta e}}{Z(0,\beta)^n(2\pi\beta^{-1})^{n/2}}\Gamma_n(r,p,e)
\end{aligned}$$

where $\Gamma_n(r,p,e)$ is the volume of the corresponding $2n-3$-dimensional surface on $\mathbb{R}^{2n}$.

As consequence of (1.5.8) we have computed the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \Gamma_n(r,p,e) = S(r, e - p^2/2) \ . \tag{1.6.3}$$

We can now define the thermodynamic quantities from the entropy definition (1.6.2). The convex duality gives

$$\lambda(r,u) = \frac{\partial S(r,u)}{\partial r}, \qquad \beta(r,u) = \frac{\partial S(r,u)}{\partial u} \tag{1.6.4}$$

and

$$\begin{aligned} r(\lambda,\beta) &= \frac{\partial \log Z(\lambda,\beta)}{\partial \lambda} = \frac{\int r e^{\lambda r - \beta V(r)} dr}{Z(\lambda,\beta)} \\ u(\lambda,\beta) &= \frac{\partial \log \left(Z(\lambda,\beta)\sqrt{\beta/2\pi} \right)}{\partial \beta} = \frac{\int V(r) e^{\lambda r - \beta V(r)} dr}{Z(\lambda,\beta)} + \frac{1}{2\beta} \end{aligned} \tag{1.6.5}$$

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 *tension*.

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

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

$$dS = \beta\tau dr + \beta du = \frac{dQ}{T} \tag{1.6.6}$$

where dQ is the (non-exact) differential

$$dQ = \tau dr + du \tag{1.6.7}$$

represent the energy gained (or lost) by the system under the infinitesimal change dr, du.

The *heat capacity at constant tension* is defined as

$$
\begin{aligned}
C(\tau, T) = \frac{d}{dT}u = -\frac{1}{T^2}\frac{d}{d\beta}u(\beta\tau, \beta) = -\frac{1}{T^2}\frac{d}{d\beta}\frac{\int V(r)e^{\beta(\tau r - V(r))}dr}{Z(\tau\beta, \beta)} + 1 \\
= -\frac{1}{T^2}\frac{\int(\tau r - V(r))V(r)e^{\beta(\tau r - V(r))}dr}{Z(\tau\beta, \beta)} \\
+\frac{1}{T^2}\frac{(\int(\tau r - V(r))e^{\beta(\tau r - V(r))}dr)(\int V(r)e^{\beta(\tau r - V(r))}dr)}{Z(\tau\beta, \beta)^2} + 1 \\
= -\frac{1}{T^2}\frac{\tau\int rV(r)e^{\beta(\tau r - V(r))}dr - \int V(r)^2 e^{\beta(\tau r - V(r))}dr}{Z(\tau\beta, \beta)} + \frac{1}{T^2}\left(\tau r u - u^2\right) + 1 \\
= \frac{1}{T^2}\left(\mathrm{var}_{\tau,T}(V(r)) - \tau\mathrm{cov}_{\tau,T}(r, V(r))\right) + 1
\end{aligned}
\tag{1.6.8}
$$

Exercise: Prove that $C(\tau, T) > 0$.

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] William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.