A mini-course in large deviations

Tertuliano Franco (UFBA)
http://w3.impa.br/~tertu

Salvador - Brazil
March, 2015

A mini-course in large deviations

These are notes prepared for an one week mini-course presented at Universidad de La República, Facultad de Ingeniería, at the city of Montevideo, Uruguay in March of 2015. I appreciate the invitation of Professor Paola Bermolen, and also the audience for valuable ideas and corrections: Paola Bermolen, Diego Fernández, Valeria Goycoechea, Ernesto Mordecki, Eugenia Riaño, Andrés Sosa and Gastón Villarino.

I specially thank Leonardo De Carlo (Gran Sasso Science Institute, Italy) for improvements and finding a plenty of typos and mistakes.

The aim of this mini-course is a mild introduction to the subject of large deviations. A pre-requisite would be some knowledge at basic level in Probability and Measure Theory. The content is a mixture of small pieces of some texts [2, 4, 5, 6, 7, 8, 11] trying to introduce the subject in the simplest way as possible. At the end of each chapter, we include few exercises.

This text was typed in XƎLaTeX.

CONTENTS

CHAPTER 1

LARGE DEVIATIONS FOR I.I.D BERNOULLI

1.1 The idea

In the Weak Law of Large Numbers, we have convergence in probability. For instance,

Theorem 1.1.1 (Weak Law of Large Numbers). *Let $X_1, X_2, \ldots$ be i.i.d. random variables such that $\mathbb{E}|X_1| < \infty$. Then, for any $\varepsilon > 0$, holds*

$$\mathbb{P}\Big[\, \Big|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}X_1\Big| > \varepsilon \,\Big] \stackrel{n\to\infty}{\longrightarrow} 0.$$

The question that emerges is: how fast does converge to zero the probability above? In what sense should we formulate the respective speed?

We can roughly reformulate this question as follows. Denote $S_n = X_1 + \cdots + X_n$ the sum of those random variables above. Let P_n be the measure in $\mathbb{R}$ induced by the random variable $\frac{S_n}{n}$. If A is some (suitable) subset of $\mathbb{R}$, we have

$$P_n(A) \stackrel{n\to\infty}{\longrightarrow} \begin{cases} 1, & \text{if} \quad \mathbb{E}X_1 \in A, \\ 0 & \text{if} \quad \mathbb{E}X_1 \notin A. \end{cases}$$

Then, what should be the speed of such convergence?

The answer is that in general, the velocity is exponential, or between exponentials. Such question, called *large deviations*, takes place in different contexts of probability as Markov chains, fluid limit, hydrodynamic limit, diffusions, polymers, etc, etc, etc.

In the words of Dembo/Zeituni [2], the large deviations are sharp enough to be useful in applications and rough enough to be possible of being handled in a mathematical rigorous way.

We start the mini-course dealing with the simplest possible example. Sums of i.i.d. random variables with distribution Bernoulli($\frac{1}{2}$). Since a sum of such random variables is binomial, we are going to need information on the factorial. This is the content of next section.

1.2 Stirling's formula

Definition 1.2.1. Let $f, g : \mathbb{N} \to \mathbb{R}$. We say that f and g have same order, and write $f \sim g$ to denote it, if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} \;=\; 1.$$

Proposition 1.2.2 (Stirling's Formula). *We have that*

$$n! \;\sim\; \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}.$$

Proof. We follow Feller's book [4]. Notice that

$$\log n! \;=\; \log 1 + \log 2 + \cdots \log n.$$

Since

$$\int_{k-1}^{k} \log x\, dx \;<\; \log k \;<\; \int_{k}^{k+1} \log x\, dx,$$

we have that

$$\int_{0}^{n} \log x\, dx \;<\; \log n! \;<\; \int_{1}^{n+1} \log x\, dx.$$

Because $\int \log x\, dx = x \log x - x$, we get

$$n \log n - n \;<\; \log n! \;<\; (n+1)\log(n+1) - (n+1) + 1,$$

or else,

$$n \log n - n \;<\; \log n! \;<\; (n+1)\log(n+1) - n.$$

The inequality above suggests to compare $\log n!$ with some average of left and right sides. The simplest would be $(n + \frac{1}{2}) \log n - n$. Define

$$d_n \;=\; \log n! - (n + \tfrac{1}{2}) \log n + n$$

which represents the error. Provided we got the correct guess (the average above) we shall estimate[1] d_n. By simple calculations,

$$\begin{aligned} d_n - d_{n+1} \;&=\; \Big(n + \tfrac{1}{2}\Big) \log\Big(\frac{n+1}{n}\Big) - 1 \\ &=\; \frac{2n+1}{2} \log\Big(\frac{1 + \frac{1}{2n+1}}{1 - \frac{1}{2n+1}}\Big) - 1 \end{aligned}$$

[1] Indeed, we could forget the proof up to here. The aim of the discussion was to correctly guess the approximation.

Now we need an expansion for the logarithm. Integrating

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

we obtain

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

Therefore,

$$\log\Big(\frac{1+x}{1-x}\Big) = 2\Big(x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots\Big)$$

Applying this,

$$\log\Big(\frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}\Big) = \frac{2}{2n+1} + \frac{2}{3(2n+1)^3} + \frac{2}{5(2n+1)^5} + \cdots$$

which leads to

$$d_n - d_{n+1} = \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \frac{1}{7(2n+1)^6} + \cdots \tag{1.1}$$

Because $\sum_{n=1}^{\infty}(d_n - d_{n+1})$ is convergent, then d_n is convergent. Thus, for some c

$$\lim_{n\to\infty} e^{d_n} = e^c$$

which is the same as

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}}e^{-n}} = e^c.$$

For a proof that $e^c = \sqrt{2\pi}$, see Wallis formula or the Central Limit Theorem (see Feller [4] or Fernandez [9]). □

1.3 The large deviations

Theorem 1.3.1. *Let $X_1, X_2, X_3 \dots$ be i.i.d. random variables such that $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = \frac{1}{2}$. Then, for any $a > 1/2$,*

$$\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}\Big[\frac{S_n}{n} \geq a\Big] = -I(a), \tag{1.2}$$

where

$$I(a) = \begin{cases} a\log a + (1-a)\log(1-a) + \log 2\,, & \text{if } a \in [0,1], \\ \infty\,, & \text{otherwise.} \end{cases} \tag{1.3}$$

The function I is called *rate function*, see Figure 1.1 for an illustration. Observe that the statement above (1.2) roughly says that $P\Big[\frac{S_n}{n} \geq a\Big]$ goes to zero as $e^{-nI(a)}$.

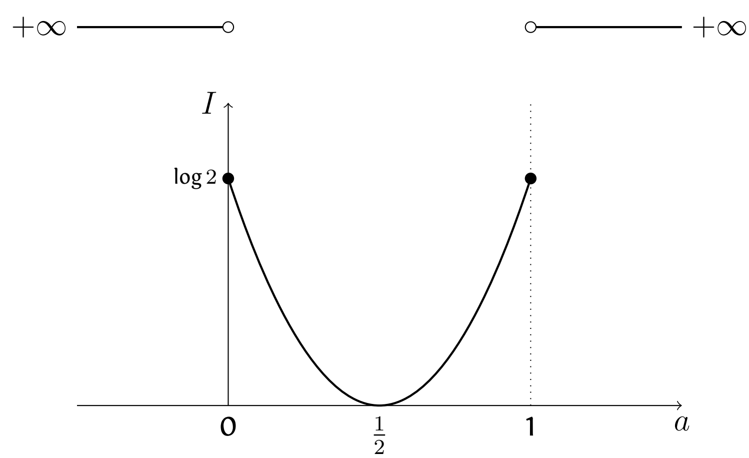

Figure 1.1: Rate function $I : \mathbb{R} \to [0, \infty]$

Proof. For $a > 1$, immediate. For $a \in (1/2, 1]$,

$$P\Big[\tfrac{S_n}{n} \geq a\Big] \;=\; P\Big[S_n \geq an\Big] \;=\; \sum_{k \geq an} P\Big[S_n = k\Big] \;=\; \sum_{k \geq an} \binom{n}{k} \frac{1}{2^n}.$$

Then,

$$\frac{1}{2^n} \max_{k \geq an} \binom{n}{k} \;\leq\; P\Big[S_n \geq an\Big] \;\leq\; \frac{n+1}{2^n} \max_{k \geq an} \binom{n}{k}.$$

For short, denote $q = \lfloor an \rfloor + 1$. Rewriting the above,

$$\frac{1}{2^n}\binom{n}{q} \;\leq\; \mathbb{P}\Big[S_n \geq an\Big] \;\leq\; \frac{n+1}{2^n}\binom{n}{q}.$$

Taking logarithms and dividing by n,

$$\begin{aligned}
\frac{1}{n}\log \frac{1}{2^n}\binom{n}{q} \;&=\; -\log 2 + \frac{1}{n}\log \frac{n!}{(n-q)!q!} \\
&\sim\; -\log 2 + \frac{1}{n}\log \frac{n^{n+\frac{1}{2}}e^{-n}}{(n-q)^{n-q+\frac{1}{2}}e^{-(n-q)}q^{q+\frac{1}{2}}e^{-q}} \\
&=\; -\log 2 + \frac{1}{n}\log \frac{n^{n+\frac{1}{2}}}{(n-q)^{n-q+\frac{1}{2}}q^{q+\frac{1}{2}}} \\
&\sim\; -\log 2 + \frac{1}{n}\log \Big(\frac{n}{n-q}\Big)^{n-q} + \frac{1}{n}\log \Big(\frac{n}{q}\Big)^{q}.
\end{aligned}$$

Because $q = \lfloor an \rfloor + 1$, the last expression above converges to

$$-\log 2 - a\log a - (1-a)\log(1-a),$$

which concludes the proof. $\square$

Remark 1.3.2. Notice that, for $a < \frac{1}{2}$, holds

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big[\tfrac{S_n}{n} \leq a\Big] \;=\; -I(a),$$

by symmetry. Moreover, we did not use the value of the constant $c = \sqrt{2\pi}$ appearing in the Stirling's formula.

The minimum of $I(a)$ is attained at $a = 1/2$, corresponding to the Strong Law of Large Numbers, see next corollary. When there are more than one minimum, the phenomena is associated to loss of uniqueness in the thermodynamic limit (in the Ising model, for instance). Also to the Curie-Weiss model.

Corollary 1.3.3. *Assuming the same hypothesis of Theorem 1.3.1,*

$$\frac{S_n}{n} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{2} \quad \textit{almost surely.}$$

Proof. For any $\varepsilon > 0$,

$$\mathbb{P}\Big[\Big|\tfrac{S_n}{n} - \tfrac{1}{2}\Big| \geq \varepsilon\Big] \;=\; \mathbb{P}\Big[\tfrac{S_n}{n} \geq \tfrac{1}{2} + \varepsilon\Big] + \mathbb{P}\Big[\tfrac{S_n}{n} \leq \tfrac{1}{2} - \varepsilon\Big]$$

For any $\delta > 0$, there exists n_0 such that

$$\mathbb{P}\Big[\tfrac{S_n}{n} \geq \tfrac{1}{2} + \varepsilon\Big] \leq e^{-n(I(\frac{1}{2}+\varepsilon)-\delta)},$$

for any $n \geq n_0$. Choose δ small enough such that

$$I\Big(\tfrac{1}{2} + \varepsilon\Big) - \delta \;>\; 0.$$

Then Borel-Cantelli's Lemma finishes the proof. □

1.4 Exercises

Exercise 1.4.1. Check that

$$d_n - d_{n+1} \;>\; \frac{1}{3(2n+1)^2} \;>\; \frac{1}{12n+1} - \frac{1}{12(n+1)+1}$$

to realize that $d_n - \frac{1}{12n+1}$ is decreasing.

Exercise 1.4.2. Comparing (1.1) with geometric series of ratio $(2n+1)^{-2}$, show that

$$d_n - d_{n+1} \;<\; \frac{1}{3[(2n+1)^2 - 1]} \;=\; \frac{1}{12n} - \frac{1}{12(n+1)}$$

to realize that $d_n - \frac{1}{12n}$ is increasing.

Exercise 1.4.3. Show that

$$c + \frac{1}{12n+1} \;<\; d_n \;<\; c + \frac{1}{12n},$$

where $c = \lim_{n\to\infty} d_n$.

Exercise 1.4.4. Show that

$$\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}\, e^{\frac{1}{12n+1}} \;<\; n! \;<\; \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}\, e^{\frac{1}{12n}}$$

improving the version of Stirling's formula proved in the text.

Exercise 1.4.5. Explain why our proof of Stirling would be a mess if we have done

$$\frac{2n+1}{2}\log\Big(1+\frac{1}{n}\Big) \;=\; \frac{2n+1}{2}\Big(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - \cdots\Big)$$

(or else, the innocent step $\frac{n+1}{n} = \frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}$ is relevant!).

Exercise 1.4.6. Let $F : \mathbf{X} \to \mathbb{R}$ be a continuous bounded function, where $(\mathbf{X}, d)$ is a separable complete metric space. Fix some $\delta > 0$. Show that $\mathbf{X}$ can be written as a finite union of closed sets where the oscillation of F in each set is smaller than δ.

Exercise 1.4.7. We say that $f : (\mathbf{X}, d) \to \mathbb{R}$ is lower semi-continuous if

$$\liminf_{x\to x_0} f(x) \;\geq\; f(x_0),$$

for any x_0. Prove that f is l.s.c. if, and only if, for any $b \in \mathbb{R}$, the set

$$[f \leq b] := \{x \in \mathbf{X}\,;\; f(x) \leq b\}$$

is closed.

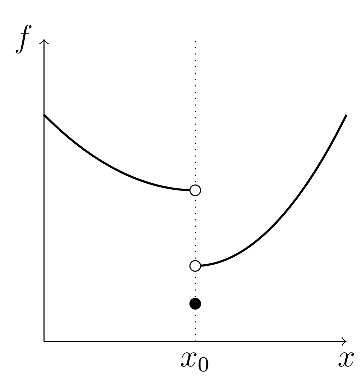

Figure 1.2: Illustration of a lower semi-continuous function

CHAPTER 2

GENERAL SETTING

2.1 The statement

Let $(\mathbf{X}, d)$ be a complete separable metric space and P_n be a sequence of measures, all of them defined in the Borelian sets of $(\mathbf{X}, d)$.

Definition 2.1.1. We say that $\{P_n\}$ satisfies a Large Deviation Principle (LDP) with rate function $I : \mathbf{X} \to [0, \infty]$ and speed $a_n \in \mathbb{R}$, $a_n \to \infty$, if

(a) $0 \leq I(x) \leq +\infty$, for all $x \in \mathbf{X}$,

(b) I is lower semi-continuous,

(c) for any $\ell < \infty$, the set $\{x \in \mathbf{X} \, : \, I(x) \leq \ell\}$ is compact in $(\mathbf{X}, d)$,

(d) for any closed set $C \subset \mathbf{X}$,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(C) \;\leq\; -\inf_{x\in C} I(x),$$

(e) for any open set $A \subset \mathbf{X}$,

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

Remark 2.1.2. Tipically, $P_n \xrightarrow{d} \delta_{x_0}$, $I(x_0) = 0$ and $I(x) > 0$ if $x \neq x_0$. Which implies the Strong Law of Large Numbers. For instance, the speed a_n can be n, n^2 or can be also ε^{-1}, and $\{P_\varepsilon\}$ a family of measures indexed in real $\varepsilon \to 0$. The inequalities may remember you of Pormanteau's Theorem. The topological issues are the same in this scenario, that's why the resemblance. But nothing more than that. Besides, some authors say that if (c) is satisfied, then I is a *good rate function*. Also see that (c) implies (b). We wrote both here despite the redundancy.

Let us discuss some aspects of the statement above. The name Principle is historical. Notice that if

$$\inf_{x\in\overset{\circ}{G}} I(x) \;=\; \inf_{x\in G} I(x) \;=\; \inf_{x\in \overline{G}} I(x), \tag{2.1}$$

then we will have convergence indeed:

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

Now, observe that there is an infimum of I that there was not in the last chapter, where the rate function was given by (1.3). In fact, there was an infimum hidden. Just notice that the rate function given by (1.3) is increasing in $[\frac{1}{2},\infty)$ and decreasing in $(-\infty,\frac{1}{2}]$.

Question: why the LDP is written in terms of an infimum $\inf_{x\in G} I(x)$? To explain this, we begin with a simple result of analysis, but very important in large deviations.

Proposition 2.1.3. *Let $a_n\to\infty$ and $b_n, c_n>0$. Then*

$$\limsup_{n\to\infty}\frac{1}{a_n}\log(b_n+c_n) \;=\; \max\Big\{\limsup_{n\to\infty}\frac{1}{a_n}\log b_n\,,\;\limsup_{n\to\infty}\frac{1}{a_n}\log c_n\Big\}.$$

In words, the bigger sequence wins, and the smaller sequence has no effect.

Proof. We have that

$$\frac{1}{a_n}\log(b_n+c_n)\leq\frac{1}{a_n}\log\Big(2\max\{a_n,b_n\}\Big)=\max\Big\{\frac{1}{a_n}\log b_n\,,\;\frac{1}{a_n}\log c_n\Big\}+\frac{1}{a_n}\log 2,$$

which implies

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{a_n}\log(b_n+c_n)&\leq\limsup_{n\to\infty}\max\Big\{\frac{1}{a_n}\log b_n\,,\;\frac{1}{a_n}\log c_n\Big\}\\&=\max\Big\{\limsup_{n\to\infty}\frac{1}{a_n}\log b_n\,,\;\limsup_{n\to\infty}\frac{1}{a_n}\log c_n\Big\}.\end{aligned}$$

The reverse inequality is obvious. □

Remark 2.1.4. A common application of above is "to throw away" superexponentially small events.

Returning to the discussion of why the rate functions appears in terms of an infimum, given disjoint sets A and B satisfying (2.1), then:

$$\begin{aligned}I(A\cup B)=-\lim\frac{1}{a_n}\log P_n(A\cup B)&=-\max\{\lim\frac{1}{a_n}\log P_n(A)\,,\;\lim\frac{1}{a_n}\log P_n(B)\}\\&=\min\{I(A)\,,\;I(B)\}.\end{aligned}$$

We therefore guess that $I(A)$ must be of the form

$$I(A) \;=\; \min_{x\in A} I(x),$$

where we have abused of notation, of course. Citing den Hollander, *"Any large deviation is done in the least unlikely of all unlikely ways!"*

Since there is a minus in $-\inf_{x\in A} I(x)$, where $I(x)$ is smaller it means that x is **less** improbable. Next, we are going to see two consequences of the general statement.

2.2 Laplace-Varadhan's Lemma

Theorem 2.2.1 (Laplace-Varadhan's Lemma). *Let* $(\mathbf{X}, d)$ *be a complete separable metric space and* $\{P_n\}$ *sequence of probability measures on the Borelian sets of* $\mathbf{X}$ *satisfying a LDP with rate function* I*. Then, for any bounded continuous function* $F : \mathbf{X} \to \mathbb{R}$*, holds*

$$\lim_{n\to\infty} \frac{1}{a_n} \log\Big(\int_X e^{a_n F(x)}\, P_n(dx)\Big) \;=\; \sup_{x\in X}\Big\{F(x) - I(x)\Big\}.$$

An interpretation of the result above would be: $\mathbb{E}_n[e^{a_n F}]$ increases (or decreases) exponentially as $\sup[F(x) - I(x)]$. Or else, the functions F and I compete together. As in Varadhan's words, the idea behind the result above relies in the simple fact that

$$\lim_{n\to\infty} \frac{1}{n} \log\Big(\sum_{i=1}^{k} e^{n x_i}\Big) \;=\; \max\{x_1, \ldots, x_k\}.$$

An often application of the Laplace-Varadhan's Lemma is to obtain a new LDP by a changing of measures:

$$\nu_n(dx) \;=\; \frac{e^{a_n F(x)} P_n(dx)}{\int_{\mathbf{X}} e^{a_n F(x)} P_n(dx)},$$

and $\tilde{I}(x) = I(x) - F(x) - \inf_{y\in X}\Big[I(y) - F(y)\Big]$ with $\tilde{a}_n = a_n$, see the exercises. Let us prove the Laplace-Varadhan's Lemma now.

Proof. Given $\delta > 0$, there exist finite closed sets covering $\mathbf{X}$ such that the oscillation of F in each these sets is smaller than δ (to see this, divide the image of F is intervals of size δ and take pre-images). Then,

$$\int_{\mathbf{X}} e^{a_n F(x)}\, P_n(dx) \;\leq\; \sum_{j=1}^{M} \int_{C_j} e^{a_n F(x)}\, P_n(dx) \;\leq\; \sum_{j=1}^{M} \int_{C_j} e^{a_n (F_j + \delta)}\, P_n(dx),$$

where F_j is the infimum of F on the set C_j. Thus,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{a_n} \log \int_{\mathbf{X}} e^{a_n F(x)}\, P_n(dx) &\leq \max_{j=1,\dots,M} \Big\{ \limsup_{n\to\infty} \frac{1}{a_n} \log \int_{C_j} e^{a_n (F_j+\delta)}\, P_n(dx) \Big\} \\
&= \limsup_{n\to\infty} \max_{j=1,\dots,M} \Big\{ F_j + \delta + \frac{1}{a_n} \log P_n(C_j) \Big\} \\
&\leq \max_{j=1,\dots,M} \Big\{ F_j + \delta - \inf_{x\in C_j} I(x) \Big\} \\
&\leq \max_{j=1,\dots,M} \sup_{x\in C_j} \Big\{ F_j - I(x) \Big\} + \delta \\
&= \sup_{x\in X} \Big\{ F(x) - I(x) \Big\} + \delta,
\end{aligned}$$

and since δ is arbitrary small, we get the upper bound. It is missing the reverse inequality. Given $\delta > 0$, there exists $y \in \mathbf{X}$ such that

$$F(y) - I(y) \;\geq\; \sup_{x\in\mathbf{X}} \Big\{ F(x) - I(x) \Big\} - \frac{\delta}{2}.$$

We can also find an open neighbourhood $U \ni y$ such that

$$F(x) \;\geq\; F(y) - \frac{\delta}{2}\,,\; \forall x \in U.$$

Then,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{a_n} \log \int_{\mathbf{X}} e^{a_n F(x)}\, P_n(dx) &\geq \liminf_{n\to\infty} \frac{1}{a_n} \log \int_{U} e^{a_n F(x)}\, P_n(dx) \\
&\geq \liminf_{n\to\infty} \frac{1}{a_n} \log \int_{U} e^{a_n (F(y)-\frac{\delta}{2})}\, P_n(dx) \\
&= F(y) - \frac{\delta}{2} + \liminf_{n\to\infty} \frac{1}{a_n} \log P_n(U) \\
&= F(y) - \frac{\delta}{2} - \inf_{x\in U} I(x) \\
&\geq F(y) - \frac{\delta}{2} - I(y) \\
&\geq \sup_{x\in\mathbf{X}} \Big\{ F(x) - I(x) \Big\} - \delta.
\end{aligned}$$

Since δ is arbitrary, we obtain the reverse inequality, finishing the proof. □

The next modification of the previous result is sometimes useful.

Theorem 2.2.2. *Let $\{P_n\}$ satisfying a LDP with rate function I. Let $F_n : \mathbf{X} \to \mathbb{R}$ be functions $F_n \geq 0$ such that*

$$\liminf_{\substack{n\to\infty\\ y\to x}} F_n(y) \;\geq\; F(x)\,, \quad \forall x \in \mathbf{X},$$

for some lower semi-continuous function $F \geq 0$. Then,

$$\limsup_{n\to\infty} \frac{1}{a_n} \log\Big(\int_{\mathbf{X}} e^{-a_n F_n(x)}\, P_n(dx)\Big) \;\leq\; -\inf_{x\in\mathbf{X}} \Big\{ F(x) + I(x) \Big\}.$$

Proof. Let

$$\ell \;=\; \inf_{x\in\mathbf{X}} \Big\{ F(x) + I(x) \Big\}.$$

For any $\delta > 0$ and for any $x \in \mathbf{X}$, there exists a neighbourhood U_δ of x such that

$$\inf_{y\in\overline{U_\delta}} I(y) \;\geq\; I(x) - \delta$$

and

$$\liminf_{n\to\infty} \Big\{ \inf_{y\in U_\delta} F_n(y) \Big\} \;\geq\; F(x) - \delta.$$

Thus,

$$\begin{aligned} \int_{U_\delta} e^{-a_n F_n(y)}\, P_n(dy) \;&\leq\; \Big(\sup_{y\in U_\delta} e^{-a_n F_n(y)} \Big) P_n(U_\delta) \\ &\leq\; \exp\Big(-a_n \inf_{y\in U_\delta} F_n(y) \Big) \exp\Big(-a_n \big(\inf_{y\in U_\delta} I(y) - o(1) \big) \Big) \\ &\leq\; \exp\Big(-a_n [\ell - 2\delta + o(1)] \Big). \end{aligned}$$

Each compact set $K \subset \mathbf{X}$ can be writed as a subset of some open set U which is a finite union of U_δ's. Then

$$\int_U e^{-a_n F_n(y)}\, P_n(dy) \;\leq\; M \exp\Big\{ -a_n \big[\ell - 2\delta + o(1) \big] \Big\}.$$

On the other hand, since $F_n \geq 0$,

$$\begin{aligned} \int_{\mathbf{X}-U} e^{-a_n F_n(y)}\, P_n(dy) \;\leq\; P_n(X - U) \;&\leq\; \exp\Big\{ -a_n \big[\inf_{X-U} I(x) + o(1) \big] \Big\} \\ &\leq\; \exp\Big\{ -a_n \big[\inf_{X-K} I(x) + o(1) \big] \Big\}. \end{aligned}$$

Choosing the compact $K = \{x \,;\, I(x) \leq M\}$, for M much bigger than ℓ, the term above is small in comparison with the previous term, which gives us the result because δ is arbitrary small. □

2.3 Contraction Principle

Theorem 2.3.1 (Contraction Principle). *Let $\{P_n\}$ satisfying a LDP with rate function $I : \mathbf{X} \to [0, +\infty]$, where $(\mathbf{X}, d)$ is a Polish space. Let $T : \mathbf{X} \to \tilde{\mathbf{X}}$ be a continuous function, where $(\tilde{\mathbf{X}}, \tilde{d})$ is also a Polish space. Define $\tilde{P}_n = P_n \circ T^{-1}$, or else,*

$$\tilde{P}_n(A) \;=\; P_n(T^{-1}(A)).$$

Then, the sequence of measures $\{\tilde{P}_n\}$ satisfies a LDP with same speed and rate function

$$\tilde{I}(y) \;=\; \inf_{x\,;\,T(x)=y} I(x).$$

Proof. (a) $0 \leq \tilde{I}(y) \leq +\infty$ is ok.

(b) Since T is continuous and $[I \leq b]$ is compact, then $[\tilde{I} \leq b] = T([I \leq b])$ is compact, therefore closed. Thus I is lower semi-continuous.

(c) Same as above.

(d) For any closed $C \subset \tilde{\mathsf{X}}$, we have that $T^{-1}(C)$ is also closed. Then,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{a_n} \log \tilde{P}_n(C) \;=\; \limsup_{n\to\infty} \frac{1}{a_n} \log P_n(T^{-1}(C)) \;&\leq\; -\inf_{x\in T^{-1}(C)} I(x)\\ &=\; -\inf_{y\in C} \tilde{I}(y).\end{aligned}$$

(e) Same as above.

□

2.4 Exercises

Exercise 2.4.1. Verify the observation in the text about the *tilting*, or else, given measures P_n satisfying a LDP with rate I and speed a_n, then the measures

$$\nu_n(dx) \;=\; \frac{e^{a_n F(x)} P_n(dx)}{\int_{\mathsf{X}} e^{a_n F(x)} P_n(dx)},$$

satisfies a LDP with rate function $\tilde{I}(x) = I(x) - F(x) - \inf_{y\in X} \Big[I(y) - F(y)\Big]$ and speed $\tilde{a}_n = a_n$.

Exercise 2.4.2. Suppose that $a_n \to \infty$, $b_n, c_n > 0$ and there exists the limits

$$\lim_{n\to\infty} \frac{1}{a_n} \log b_n \qquad \text{and} \qquad \lim_{n\to\infty} \frac{1}{a_n} \log c_n.$$

Show that

$$\lim_{n\to\infty} \frac{1}{a_n} \log(b_n + c_n) \;=\; \max\Big\{ \lim_{n\to\infty} \frac{1}{a_n} \log b_n \,,\; \lim_{n\to\infty} \frac{1}{a_n} \log c_n \Big\}.$$

Exercise 2.4.3. Show that the Laplace-Varadhan's Lemma remains true if we change the assumption of boundedness of F by the superexponential assumption

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{a_n} \log \int_{\{x\,;\,F(x)\geq L\}} e^{a_n F(x)}\, P_n(dx) \;=\; -\infty.$$

Hint: put $F_L = F \wedge L$ and then observe that

$$\int_{\mathsf{X}} e^{a_n F}\, P_n \;\leq\; \int_{\mathsf{X}} e^{a_n F_L}\, P_n + \int_{[F\geq L]} e^{a_n F}\, P_n.$$

Exercise 2.4.4. Verify that the rate function obtained in previous chapter (for sums of Bernoulli's) satisfies (a) and (b) of the general statement for large deviations.

Exercise 2.4.5. Given a family $f_\lambda : \mathbb{R} \to \mathbb{R}$ of continuous functions, show that the function defined by $f(x) = \sup_\lambda f_\lambda(x)$ is lower semi-continuous.

Exercise 2.4.6. A sequence of measures $\{P_n\}$ is said *exponentially tight* if, for any $b < \infty$, there exists some compact $K_b \subset \mathbf{X}$ such that

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(K_b^{\complement}) \;\leq\; -b.$$

Suppose that $\{P_n\}$ is exponentially tight and, for each compact set $K \subset \mathbf{X}$, holds

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(K) \;\leq\; -\inf_{x\in K} I(x).$$

Show that these two assumptions implies the item (d) of the general statement, or else, for any closed $C \subset \mathbf{X}$, we will have

$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(C) \;\leq\; -\inf_{x\in C} I(x).$$

Hint: use that $P_n(C) \leq P_n(C \cap K_b) + P_n(K_b^{\complement})$.

Exercise 2.4.7. Suppose that for any $\delta > 0$ and any $y \in \mathbf{X}$, holds

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(U_\delta) \;\geq\; -I(y).$$

where $U_\delta = B(y, \delta)$ is the ball of center y and radius $\delta > 0$. Show that this implies the item (e) of the general statement, or else, for any open set $A \subset \mathbf{X}$,

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(A) \;\geq\; -\inf_{y\in A} I(y).$$

CHAPTER 3

LARGE DEVIATIONS FOR SUMS OF I.I.D. RANDOM VARIABLES

3.1 Crámer's Theorem

We come back to sums of i.i.d. random variables. Let $X_1, X_2 \ldots$ independent identically distributed random variables such that

(A1) X_i has all exponential moments finite, or else,

$$M(\theta) \;=\; \mathbb{E}[e^{\theta X_i}] \;=\; \int e^{\theta x}\,\alpha(dx) \;<\; +\infty, \qquad \forall\,\theta \in \mathbb{R},$$

where α is the measure induced by X_i in the line.

(A2) The random variables are not bounded from above neither from below, or else, for all $m \in \mathbb{N}$,

$$\alpha((-\infty, -m)) > 0 \qquad \text{and} \qquad \alpha((m, \infty)) > 0.$$

In Olivieri/Vares book [8], it is supposed only finite first moment. In den Hollander, only assumption (**A1**) above. We follow here Varadhan's book [7], which assumes the two conditions above, which makes the proof easier and points out some important idea to be recalled ahead. Define $S_n = X_1 + \cdots + X_n$ and

$$I(x) \;=\; \sup_{\theta\in\mathbb{R}} \Big\{ \theta x - \log \mathbb{E}[e^{\theta X}] \Big\}, \tag{3.1}$$

which is the Legendre transform of $\log \mathbb{E}[e^{\theta X}]$. Observe that I is convex because

$$\begin{aligned} I(px_1 + (1-p)x_2) \;&=\; \sup_{\theta} \Big\{ px_1 + (1-p)x_2 - p\mathbb{E}[e^{\theta X}] - (1-p)E[e^{\theta X}] \Big\} \\ &\leq\; p \sup_{\theta} \Big\{ x_1 - \mathbb{E}[e^{\theta X}] \Big\} + (1-p) \sup_{\theta} \Big\{ x_2 - \mathbb{E}[e^{\theta X}] \Big\} \\ &=\; p\, I(x_1) + (1-p)\, I(x_2). \end{aligned}$$

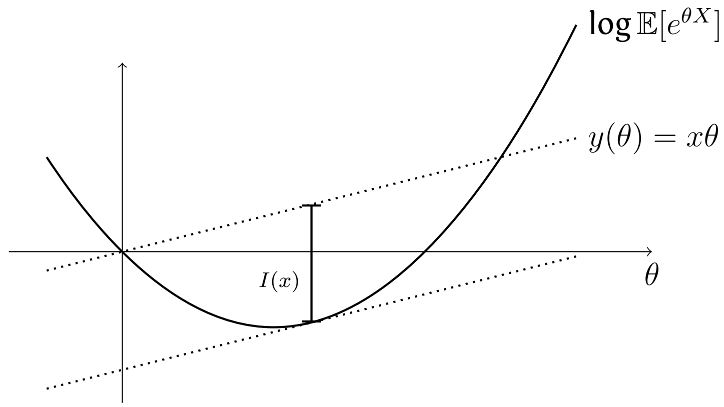

Figure 3.1: Geometric view of Legendre transform of $\log \mathbb{E}[e^{\theta X}]$, where x is the slope.

Moreover, by Jensen's inequality we have $\mathbb{E}[e^{\theta X}] \geq e^{\theta \mathbb{E} X}$, which implies the inequality $\theta\, \mathbb{E}X - \log \mathbb{E}[e^{\theta X}] \leq 0$. Or else,

$$I(\mathbb{E}X) \;=\; 0.$$

Since $I \geq 0$, we get that $x = \mathbb{E}X$ is a minimum for I. Since I is convex and has a minimum at $x = \mathbb{E}X$, we conclude that I is non increasing for $x < \mathbb{E}X$ and non decreasing for $x > \mathbb{E}X$.

Theorem 3.1.1. *Let $\{P_n\}$ be the sequence of measures in $\mathbb{R}$ induced by S_n/n, or else,*

$$P_n(A) \;=\; \mathbb{P}\Big[\tfrac{S_n}{n} \in A\Big].$$

Then $\{P_n\}$ satisfies a LDP with rate function I given by (3.1).

Proof. (a) $0 \leq I(x) \leq +\infty$ is immediate.

(b) Since I is the supremum of continuous functions then I is a lower semi-continuous function.

(c) Since $I \geq 0$ and I is lower semi-continuous, convex and has a minumum, then $I^{-1}((\infty, \ell])$ is closed and bounded in the line, therefore compact.

(d) For $x > \mathbb{E}X$ and $\theta < 0$, we have

$$\theta x - \log \mathbb{E}[e^{\theta X}] \;\leq\; \theta \mathbb{E}X - \log \mathbb{E}[e^{\theta X}] \;\leq\; 0.$$

Therefore, for $x > \mathbb{E}X$,

$$I(x) \;=\; \sup_{\theta > 0} \Big\{ \theta x - \log \mathbb{E}[e^{\theta X}] \Big\}.$$

Analogously, if $x < \mathbb{E}X$ and $\theta > 0$, we have

$$\theta x - \log \mathbb{E}[e^{\theta X}] \;\leq\; \theta \mathbb{E}X - \log \mathbb{E}[e^{\theta X}].$$

Therefore, for $x < \mathbb{E}X$,

$$I(x) \;=\; \sup_{\theta<0}\Big\{\theta x - \log \mathbb{E}[e^{\theta X}]\Big\}.$$

Let $y > \mathbb{E}X$ and $J_y = [y, +\infty)$. For $\theta > 0$, we have

$$\begin{aligned} P_n(J_y) \;&=\; \int_{J_y} P_n(dx) \;\leq\; e^{-\theta y}\int_{J_y} e^{\theta x} P_n(dx) \;\leq\; e^{-\theta y}\int_{\mathbb{R}} e^{\theta x} P_n(dx)\\ &=\; e^{-\theta y}\,\mathbb{E}\big[e^{\theta(\frac{X_1+\cdots+X_n}{n})}\big]\\ &=\; e^{-\theta y}\,\Big(\mathbb{E}\big[e^{\frac{\theta}{n}X_1}\big]\Big)^n. \end{aligned}$$

Thus,

$$\frac{1}{n}\log P_n(J_y) \;\leq\; -\frac{\theta y}{n} + \log \mathbb{E}[e^{\frac{\theta}{n}X}].$$

Since the above holds for any θ, we replace θ/n by θ, obtaining

$$\frac{1}{n}\log P_n(J_y) \;\leq\; -\theta\, y + \log \mathbb{E}[e^{\theta X}].$$

Notice that the inequality above holds for any $n \in \mathbb{N}$, having importance itself. In particular,

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(J_y) \;\leq\; -\theta\, y + \log \mathbb{E}[e^{\theta X}].$$

Optimizing over θ,

$$\begin{aligned} \limsup_{n\to\infty}\frac{1}{n}\log P_n(J_y) \;&\leq\; \inf_{\theta>0}\Big\{-\theta\, y + \log \mathbb{E}[e^{\theta X}]\Big\}\\ &=\; -\sup_{\theta>0}\Big\{\theta\, y - \log \mathbb{E}[e^{\theta X}]\Big\} \;=\; -I(y). \end{aligned}$$

Analogously, for sets of the form $\tilde{J}_y = (-\infty, y]$, we get

$$\begin{aligned} \limsup_{n\to\infty}\frac{1}{n}\log P_n(\tilde{J}_y) \;&\leq\; \inf_{\theta<0}\Big\{-\theta\, y + \log \mathbb{E}[e^{\theta X}]\Big\}\\ &=\; -\sup_{\theta<0}\Big\{\theta\, y - \log \mathbb{E}[e^{\theta X}]\Big\} \;=\; -I(y). \end{aligned}$$

Let us consider an arbitrary closed set $C \subset \mathbb{R}$ now. If $\mathbb{E}X \in C$, then $\inf_{x\in C} I(x) = 0$ and trivially

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(C) \;\leq\; 0.$$

If $\mathbb{E}X \notin C$, let (y_1, y_2) be the greatest interval around $\mathbb{E}X$ for which we have that $C \cap (y_1, y_2) = \varnothing$. Then $C \subset \tilde{J}_{y_1} \cup J_{y_2}$, yielding

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(C) \;\leq\; -\min\{I(y_1), I(y_2)\} \;=\; \inf_{y\in C} I(y),$$

by the monotonicity of I.

(e) It is enough to prove that, for all $\delta>0$, denoting $U_\delta=(y-\delta,y+\delta)$,

$$\liminf_{n\to\infty}\frac{1}{n}\log P_n(U_\delta)\;\geq\;-I(y).$$

(see Exercise 2.4.7). We claim now that

$$\lim_{|\theta|\to+\infty}\frac{\log\mathbb{E}[e^{\theta X}]}{|\theta|}\;=\;\infty. \tag{3.2}$$

Since $\mathbb{E}[e^{\theta X}]\geq e^{\theta M}\mathbb{P}[X\geq M]$, then

$$\frac{\log\mathbb{E}[e^{\theta X}]}{|\theta|}\;\geq\;\frac{M\theta}{|\theta|}+\frac{\log\mathbb{P}[X\geq M]}{|\theta|}.$$

By assumption **(A2)**, we have that $\mathbb{P}[X\geq M]>0$, which implies the claim with $\theta\to+\infty$. The case $\theta\to-\infty$ is analogous. This proves the claim.

By (3.2), for each fixed y,

$$I(y)\;=\;\sup_{\theta\in\mathbb{R}}\Big\{\theta y-\log M(\theta)\Big\}$$

is assumed in some θ_0. Then

$$I(y)\;=\;\theta_0 y-\log M(\theta_0)\qquad\text{and}\qquad y-\frac{M'(\theta_0)}{M(\theta_0)}\;=\;0. \tag{3.3}$$

We define now a new probability distribution α_{θ_0} by

$$\alpha_{\theta_0}(A)\;=\;\frac{1}{M(\theta_0)}\int_A e^{\theta_0 x}\,\alpha(dx),$$

which is the so-called *Crámer transform* of α. Notice that, by (3.3), the mean under α_{θ_0} is y, or else, $y=\int x\,\alpha_{\theta_0}(dx)$. Therefore, by the Weak Law of Large Numbers, for any δ we have that

$$\lim_{n\to\infty}\int_{\left|\frac{x_1+\cdots+x_n}{n}-y\right|<\delta}\alpha_{\theta_0}(dx_1)\cdots\alpha_{\theta_0}(dx_n)\;=\;1.$$

Besides, for $\delta_1<\delta$,

$$\begin{aligned}
P_n(U_\delta)\;&=\;\int_{\left|\frac{x_1+\cdots+x_n}{n}-y\right|<\delta}\alpha(dx_1)\cdots\alpha(dx_n)\\
&\geq\;\int_{\left|\frac{x_1+\cdots+x_n}{n}-y\right|<\delta_1}\alpha(dx_1)\cdots\alpha(dx_n)\\
&=\;\int_{\left|\frac{x_1+\cdots+x_n}{n}-y\right|<\delta_1}e^{-\theta_0(x_1+\cdots+x_n)}e^{\theta_0(x_1+\cdots+x_n)}\,\alpha(dx_1)\cdots\alpha(dx_n)\\
&\geq\;e^{-n\theta_0 y-n\delta_1|\theta_0|}\,M(\theta_0)^n\int_{\left|\frac{x_1+\cdots+x_n}{n}-y\right|<\delta_1}\alpha_{\theta_0}(dx_1)\cdots\alpha_{\theta_0}(dx_n).
\end{aligned}$$

Taking logarithms and dividing by n,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log P_n(U_\delta) &\geq -\theta_0 y - \delta_1|\theta_0| + \log M(\theta_0)\\ &= -\Big[\theta_0 y - \log M(\theta_0)\Big] - \delta_1|\theta_0|\\ &= -I(y) - \delta_1|\theta_0|.\end{aligned}$$

Since $\delta_1 < \delta$ is arbitrary, the proof is finished. □

Remark 3.1.2. We can interpret $I(y)$ as the smallest cost (exponential cost) in observing y instead of $\mathbb{E}X$ as the average of the random variables. Many others changing of measures would lead the system from $\mathbb{E}X$ (under α) to y. But the cost would be bigger. Again, the citation of den Hollander is suitable to the occasion:

"Any large deviation is done in the least unlikely of all unlikely ways!"

3.2 Exercises

Exercise 3.2.1. Check directly that if $\mathbb{P}[X_i = b] = 1$, then $\{P_n\}$ satisfies a LDP with rate function

$$I(x) = \begin{cases} 0\,, & \text{if } x = b,\\ \infty\,, & \text{if } x \neq b.\end{cases}$$

Exercise 3.2.2. Use Hölder's inequality to show that $\log \mathbb{E}[e^{\theta X}]$ is convex (being finite or not).

Exercise 3.2.3. Use Fatou's Lemma to show that $\log \mathbb{E}[e^{\theta X}]$ is lower semi-continuous (being finite or not).

Exercise 3.2.4. The goal of this exercise is to prove

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big[\tfrac{S_n}{n}\geq a\Big] = -I(a) \tag{3.4}$$

for any $a > \mathbb{E}X_1$ under only assumption **(A1)**. Notice that we did not make use of the assumption **(A2)** in the proof of the upper bound, hence we shall prove the lower bound without invoking **(A2)**.

(i) Show that we can suppose, without loss of generality, that $\mathbb{E}X_1 < 0$ and $a = 0$ in (3.4). We aim to prove

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big[\tfrac{S_n}{n}\geq 0\Big] = -I(0)$$

in the next itens, being $\mathbb{E}X_i < 0$.

(ii) Show that

$$I(0) \;=\; -\log \rho$$

where $\rho = \inf_{\theta\in\mathbb{R}}\big\{M(\theta)\big\}$.

(iii) Suppose that $\mathbb{P}[X_1 < 0] = 1$. Show that $M(\theta)$ is strictly decreasing, that $\lim_{\theta\to-\infty} M(\theta) = \rho = 0$ and then conclude (3.4).

(iv) Suppose that $\mathbb{P}[X_1 \leq 0] = 1$ and $\mathbb{P}[X_1 = 0] > 0$. Show that $M(\theta)$ is strictly decreasing, that $\lim_{\theta\to-\infty} M(\theta) = \rho = \mathbb{P}[X_1 = 0]$ and then conclude (3.4).

(v) Suppose that $\mathbb{P}[X_1 < 0] > 0$ and $\mathbb{P}[X_1 > 0] > 0$. Show that $\lim_{\theta\to\pm\infty} M(\theta) = \infty$. Since $M(\theta)$ is convex, conclude that there exists θ_0 such that $M(\theta_0) = \rho$ and $M'(\theta_0) = 0$.

(vi) Let $\hat{X}_1, \hat{X}_2, \dots$ be a sequence of i.i.d random variables with distribution given by α_{θ_0}, the Cramer transform of α. Show that $\hat{M}(\theta) = \mathbb{E}[e^{\theta \hat{X}_i}] \in C^\infty(\mathbb{R})$.

Hint: relate $\hat{M}(\theta)$ and $M(\theta)$.

(vii) From now on, suppose that $\mathbb{P}[X_1 < 0] > 0$ and $\mathbb{P}[X_1 > 0] > 0$ and choose θ_0 as in item (v). Show that

   (i) $\mathbb{E}[\hat{X}_i] = 0$

   (ii) $\mathrm{Var}(\hat{X}_i) = \hat{\sigma} \in (0,\infty)$.

(viii) Define $\hat{S}_n = \hat{X}_1 + \cdots + \hat{X}_n$. Show that

$$\mathbb{P}[S_n \geq 0] \;=\; M(\theta_0)^n\, \mathbb{E}\big[e^{-\theta_0\, \hat{S}_n} \mathbf{1}_{[\hat{S}_n \geq 0]}\big]$$

(ix) Show that

$$\mathbb{E}\Big[e^{-\theta_0 \hat{S}_n} \mathbf{1}_{[\hat{S}_n \geq 0]}\Big] \;\geq\; \exp\Big\{-\theta_0\, C\, \hat{\sigma}\sqrt{n}\Big\}\, \mathbb{P}\Big[\frac{\hat{S}_n}{\hat{\sigma}\sqrt{n}} \in [0, C)\Big],$$

and conclude (3.4).

Exercise 3.2.5. Evaluate the rate function I for

(a) Bernoulli(p),

(b) Exponential(λ),

(c) Poisson(λ),

(d) Normal($0,1$).

Exercise 3.2.6. For i.i.d. $X_i \sim$ Exponential(λ), estimate $\mathbb{P}[\frac{S_n}{n} \geq 2\lambda]$.

Exercise 3.2.7. By the following changing of measure $\alpha \mapsto \tilde{\alpha}$, where $\tilde{\alpha}$ is defined through

$$\tilde{\alpha}(A) \;=\; \int_{A-y} \alpha(dx),$$

why the proof of large deviations wouldn't work? Give an example. Interpret this changing of measure in terms of the random variable.

CHAPTER 4

LARGE DEVIATIONS FOR MARKOV CHAINS

4.1 Sanov's Theorem

Beforing entering into the subject of Markov Chains, we discuss Sanov's Theorem for finite alphabets with the aim of enlightening the idea of frequencies.

Let $X_1, X_2, \ldots$ be i.i.d random variables taking values on the finite set $S = \{x_1, \ldots, x_r\}$, which is not necessarily a subset of $\mathbb{R}$. Hence, it makes no sense in speaking about expectation of those random variables. We shall henceforth look for an appropriated metric space $(\mathbf{X}, d)$.

Remark 4.1.1. As a motivation, think for instance of a questionnaire having answers as "I agree", "I partially disagree", "I completely agree", etc. We can not sum those results. It is therefore necessary to work with frequencies rather then averages.

Definition 4.1.2. Define

$$L_n \;=\; \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j}\,,$$

which is called *empirical measure*.

Notice that:

(a) L_n is a random element (it is a function of the random variables $X_1, \ldots, X_n$),

(b) $L_n(\omega)$ represents the fraction that each x_i has appeared until time n,

(c) $L_n(\omega)$ is a probability measure on S.

In other words,

$$\begin{aligned} L_n : (\Omega, \mathbb{A}, \mathbb{P}) &\longrightarrow \mathcal{M}_1(S) \\ \omega &\longmapsto \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j(\omega)} \end{aligned}$$

where $\mathcal{M}_1(S)$ denotes the set of probability measure on S with the distance of total variation given by

$$||\mu-\alpha|| \;=\; \frac{1}{2}\sum_{x\in S}|\mu(x)-\alpha(x)|$$

which is a Polish space. By the Strong Law of Large Numbers, it holds the convergence

$$L_n \stackrel{n\to\infty}{\longrightarrow} \alpha \qquad \text{almost surely,}$$

where $\alpha(x_j)=\mathbb{P}[X_1=x_j]$. Assume $\alpha(x_j)>0$ for any j.

Theorem 4.1.3 (Sanov for finite alphabets). *Let $X_1, X_2, \dots$ i.i.d random variables taking vales on the finite set $S=\{x_1,x_2,\dots,x_r\}$. Let P_n be the measure induced in $\mathcal{M}_1(S)$ by the random element L_n. Then $\{P_n\}$ satisfies a LDP with rate function $I:\mathcal{M}_1(S)\to[0,\infty]$ given by*

$$I(\mu) \;=\; \sum_{j=1}^{r}\mu(x_j)\log\frac{\mu(x_j)}{\alpha(x_j)} \tag{4.1}$$

where $\alpha(x_j)=\mathbb{P}[X_1=x_j]$.

Since the distribution of how many times each x_j has appeared until time n is multinomial, which depends on the factorial, the proof is quite similar to that one presented in Chapter 1 for sums of Bernoulli. For this reason, we leave the proof for the Exercise 4.4.10.

Remark 4.1.4. We have that

$$I(\mu) \;=\; \int_S \log\Big(\frac{d\mu}{d\alpha}\Big)\,d\mu \;=:\; H(\mu|\alpha)$$

is called the *relative entropy* of μ with respect to α. In the large deviations scenario for a general rate function I holds (sometimes rigorously, sometimes in a heuristic sense) that

$$I(y) \;=\; \inf_{\substack{\mu \text{ such that}\\ \int x\,\mu(dx)=y}} H(\mu|\alpha). \tag{4.2}$$

In the Sanov's Theorem, there is no infimum, because the observable is the measure itself. In the Crámer's case, there is. This is a way of deducing the Crámer transform. The Crámer transform of μ is the measure that minimizes the relative entropy (with respect to μ) among all the measures that have mean y (in the setting of Crámer's Theorem).

4.2 LDP for Markov chains first level

We turn now to Markov chains in a finite state $S=\{x_1\,\dots,x_r\}$ and discrete time. Let $p(x,y)$ be a transition probability and suppose additionally that $p(x,y)>0$ for all $x,y\in S$. We will be always starting the chain from a fixed state x.

Proposition 4.2.1. *Let $V : S \to \mathbb{R}$. Then*

$$\mathbb{E}_x\big[e^{V(X_1)+\cdots+V(X_n)}\big] \;=\; \sum_{y\in S} p_V^n(x,y)\,,$$

where p_V^n is the n^{th}-power of the matrix $p(x,y)e^{V(y)}$.

Proof. For $n=2$.

$$\begin{aligned}\mathbb{E}_x\big[e^{V(X_1)+V(X_2)}\big] \;&=\; \sum_y\sum_z e^{V(y)+V(z)}\,p(x,y)\,p(y,z)\\ &=\; \sum_z\Big[\sum_y p(x,y)\,e^{V(y)}\,p(y,z)\,e^{V(z)}\Big] \;=\; p_V^2(x,z).\end{aligned}$$

The remaining follows by induction. □

Fact of life: since p_V is a matrix of positive entries, by Perron-Frobenius,

$$\frac{1}{n}\log\sum_y p_V(x,y) \longrightarrow \log\lambda_p(V),$$

where $\lambda_p(V)$ is the greatest eigenvalue of the matrix $p_V(x,y)$. As we did before, define

$$L_n \;=\; \frac{1}{n}\sum_{j=1}^n \delta_{X_j}\,,$$

which is the *empirical measure*, and counts how many times the chain has been in each state (until time n). Let us discuss the large deviation's upper and lower bound in a informal way.

In some places ahead, we will commit (a strong) abuse of notation, writing $L_n(x_1,\dots,x_n) := \frac{1}{n}\sum_{j=1}^n \delta_{x_j}$.

Upper bound. Let $C\subset\mathcal{M}_1(S)$.

$$P_n(C) \;=\; \mathbb{P}_x[L_n\in C] \;=\; \mathbb{E}_x[\mathbf{1}_{A_n}],$$

where $A_n=\{\omega\in\Omega\,;\;L_n(\omega)\in C\}$. We have then

$$\begin{aligned}\mathbb{E}_x[\mathbf{1}_{A_n}] \;&=\; \mathbb{E}_x\Big[e^{-[V(X_1)+\cdots+V(X_n)]}\,e^{[V(X_1)+\cdots+V(X_n)]}\,\mathbf{1}_{A_n}\Big]\\ &\leq\; \Big\{\max_{\substack{x_1,\dots,x_n\\ L_n(x_1,\dots,x_n)\in C}} e^{-[V(x_1)+\cdots+V(x_n)]}\Big\}\,\mathbb{E}_x\Big[e^{[V(X_1)+\cdots+V(X_n)]}\Big]\,.\end{aligned}$$

Taking logarithms and dividing by n,

$$\begin{aligned}\frac{1}{n}\log P_n(C) \;&\leq\; \frac{1}{n}\max_{\substack{x_1,\dots,x_n\\ L_n(x_1,\dots,x_n)\in C}} -[V(x_1)+\cdots+V(x_n)] + \frac{1}{n}\log\sum_y p_V^n(x,y)\\ &=\; -\frac{1}{n}\min_{\substack{x_1,\dots,x_n\\ L_n(x_1,\dots,x_n)\in C}}[V(x_1)+\cdots+V(x_n)] + \frac{1}{n}\log\sum_y p_V^n(x,y).\end{aligned}$$

It is not hard to figure out that the above leads to

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(C)\;\leq\;-\inf_{q\in C}\;\Big[\sum_{y\in S}V(y)q(y)-\log\lambda_p(V)\Big].$$

Optimizing over V gives us

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log P_n(C)\;&\leq\;\inf_{V}\Big\{-\inf_{q\in C}\;\Big[\sum_{y\in S}V(y)q(y)-\log\lambda_p(V)\Big]\Big\}\\&=\;-\inf_{q\in C}\;\sup_{V}\;\Big[\sum_{y\in S}V(y)q(y)-\log\lambda_p(V)\Big],\end{aligned}$$

where the last equality is something delicate, to be comment in the next chapter, and has to do with the Minimax Lemma. Therefore, we have obtained

$$\limsup_{n\to\infty}\frac{1}{n}\log P_n(C)\;\leq\;-\inf_{q\in C}\;\sup_{V}\;I(q),$$

where

$$I(q)\;=\;\sum_{y\in S}V(y)q(y)-\log\lambda_p(V). \tag{4.3}$$

Remark 4.2.2. The set C above in fact should be an open set. After applying the Minimax Lemma we get the upper bound for closed sets, and then using exponential tightness we get the upper bound for closed sets. More details only in the next chapter. For the moment, let us just accept the idea.

Lower bound. Let $A\subset\mathcal{M}_1(S)$ be an open set. We have

$$P_n(A)\;=\;\sum_{\substack{x_1,\dots,x_n\\L_n(x_1,\dots,x_n)\in A}}p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n).$$

Let $q\in A$ be a frequency. Pick any transition probability $\tilde{p}$ such that $q\tilde{p}=q$. Then,

$$\begin{aligned}P_n(A)\;&=\;\sum_{\substack{x_1,\dots,x_n\\L_n(x_1,\dots,x_n)\in A}}\tilde{p}(x,x_1)\tilde{p}(x_1,x_2)\cdots\tilde{p}(x_{n-1},x_n)\exp\Big\{-\sum_{j=1}^{n}\log\frac{\tilde{p}(x_{j-1},x_j)}{p(x_{j-1},x_j)}\Big\}\\&=\;\int_{A_n}\exp\Big\{-\sum_{j=1}^{n}\log\frac{\tilde{p}(X_{j-1},X_j)}{p(X_{j-1},X_j)}\Big\}\,d\tilde{P},\end{aligned}$$

where $A_n=\{(x_1,x_2,\dots)\;;\;L_n(x_1,\dots,x_n)\in A\}$. For short, denote

$$\Theta_n\;=\;-\sum_{j=1}^{n}\log\frac{\tilde{p}(X_{j-1},X_j)}{p(X_{j-1},X_j)}.$$

Then

$$\begin{aligned}\frac{1}{n}\log P_n(A) &= \frac{1}{n}\log\Big(\int e^{\Theta_n}\frac{\mathbf{1}_{A_n}}{\tilde{P}(A_n)}\,d\tilde{P}\cdot\tilde{P}(A_n)\Big)\\ &= \frac{1}{n}\log\Big(\int e^{\Theta_n}\frac{\mathbf{1}_{A_n}}{\tilde{P}(A_n)}\,d\tilde{P}\Big)+\frac{1}{n}\log\tilde{P}(A_n)\\ &= \frac{1}{n}\log\Big(\int e^{\Theta_n}\,d\hat{P}\Big)+\frac{1}{n}\log\tilde{P}(A_n),\end{aligned}$$

where

$$d\hat{P} = \frac{\mathbf{1}_{A_n}}{\tilde{P}(A_n)}\,d\tilde{P}.$$

Since the logarithm is a concave funtion, applying Jensen's inequality,

$$\frac{1}{n}\log P_n(A) \geq \int\frac{\Theta_n}{n}\,d\hat{P}+\frac{1}{n}\log\tilde{P}(A_n),$$

By the Ergodic Theorem,

$$\lim_{n\to\infty}\tilde{P}(A_n) = 1$$

because $q\tilde{p} = q$. Moreover, it is not immediate, but is also a consequence of the Ergodic Theorem that, under $\tilde{P}$,

$$\frac{\Theta_n}{n} \stackrel{n\to\infty}{\longrightarrow} -\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}$$

almost surely and in L^1. Therefore,

$$\liminf_{n\to\infty}\frac{1}{n}\log P_n(A) \geq -\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}.$$

Optimizing in $\tilde{p}$, we get

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log P_n(A) &\geq \sup_{\substack{\tilde{p}\text{ such that}\\ q\tilde{p}=q}} -\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}\\ &= -\inf_{\substack{\tilde{p}\text{ such that}\\ q\tilde{p}=q}}\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}\end{aligned}$$

Optimizing in $q\in A$ yields

$$\liminf_{n\to\infty}\frac{1}{n}\log P_n(A) \geq -\inf_{q\in A}\tilde{I}(q),$$

where

$$\tilde{I}(q) = \inf_{\substack{\tilde{p}\text{ such that}\\ q\tilde{p}=q}}\sum_{x,y} q(x)\tilde{p}(x,y)\log\frac{\tilde{p}(x,y)}{p(x,y)}.$$

We notice some resemblance of above with the heuristic equation (4.2). In order to conclude the large deviations, it is required to prove that $I(q)$ defined in (4.3) is equal to $\tilde{I}(q)$ above. This is true, but we are not going to treat this subject here. For a complete proof of a large deviations principle in this setting, see the book of Vares/Olivieri [8, page 126, Thm. 3.9].

4.3 LDP for Markov chains second level

In some sense, dealing with L_n is **not** the most natural way of proving a large deviations principle for Markov chains, due to the fact that

$$\mathbb{P}_x[X_1 = x_1, \dots, X_n = x_n] \tag{4.4}$$

is not a function of $L_n(x_1, \dots, x_n)$. In this way, define

$$L_n^{(2)} \;=\; \frac{1}{n-1}\sum_{i=0}^{n-1} \delta_{(X_i, X_{i+1})} \tag{4.5}$$

which is an empirical measure that counts the number that each *transition* has occurred until time n. Notice also that $L_n^{(2)}$ is a random element taking values on $\mathcal{M}_1(S \times S)$.

Exercise 4.3.1. Verify that (4.4) is a function of $L_n^{(2)}(x_1, \dots, x_n)$ modulo a small error.

It is a consequence of the Ergodic Theorem that

$$L_n^{(2)} \stackrel{n\to\infty}{\longrightarrow} \alpha \times p \qquad \text{almost surely,}$$

where α is the stationary distribution and $(\alpha \times p)(x, y) = \alpha(x)p(x, y)$. We state without proof the following result:

Theorem 4.3.2. *Let $P_n^{(2)}$ be measure induced by $L_n^{(2)}$ in $\mathcal{M}_1(S \times S)$. Then $\{P_n\}$ satisfies a LDP with rate function*

$$I_2(\nu) \;=\; \sum_{x,y} \nu(x, y) \log \frac{\nu(x, y)}{\hat{\nu}(x)p(x, y)}$$

where $\hat{\nu}(x) = \sum_y \nu(x, y)$ is the marginal of the measure $\nu(x, y)$ in the first coordinate.

Notice that the rate functional is an entropy. Provided by the second level large deviations, we obtain as a corollary the first level large deviations by means of the Contraction Principle, that is, Theorem 2.3.1.

Corollary 4.3.3. *Let P_n be the measure induced by L_n in $\mathcal{M}_1(S)$. Then $\{P_n\}$ satisfies a LDP with rate function*

$$I_1(\mu) \;=\; \inf_{\substack{\nu \in \mathcal{M}_1(S) \\ \hat{\nu} = \mu}} I_2(\nu).$$

4.4 Exercises

Exercise 4.4.1. Verify that for an alphabet with two letters of same probability, the rate function (4.1) coincide with the rate function (1.3) obtained in Chapter 1 for the large deviations of sums of i.i.d. Bernoulli(1/2).

Exercise 4.4.2. Let S be a finite set and μ and α be two probability measures on S. Define the entropy of μ with respect to α by

$$H(\mu|\alpha) \;=\; \sup_f \Big\{ \int f\, d\mu - \log \int e^f\, d\alpha \Big\},$$

where the supremum is taken over bounded functions $f : S \to \mathbb{R}$. Show that the supremum above can be taken only over bounded positive functions.

Exercise 4.4.3. Prove the entropy inequality: for all bounded f and for all $c > 0$,

$$\int f\, d\mu \;\leq\; \frac{1}{c}\Big(\log \int e^{cf}\, d\alpha + H(\mu|\alpha)\Big).$$

Exercise 4.4.4. Show that, for any $A \subset S$,

$$\mu(A) \;\leq\; \frac{\log 2 + H(\mu|\alpha)}{\log\Big(1 + \frac{1}{\alpha(A)}\Big)}\,.$$

Exercise 4.4.5. Show that the entropy is non-negative, convex and lower semi-continuous.

Exercise 4.4.6. Use Hölder's inequality to show that

$$\begin{aligned} \Phi :& \mathbb{R}^S \to \mathbb{R} \\ & f \mapsto \int f\, d\mu - \log \int e^f\, d\alpha \end{aligned}$$

is concave. Show that its maximum is attained in any f such that

$$e^{f(x_j)}\,\alpha(x_j) \;=\; \mu(x_j) \sum_{i=1}^{n} e^{f(x_i)}\,\alpha(x_i)\,, \qquad \forall j = 1, \dots, n.$$

Exercise 4.4.7. Invoking the previous item, check that

$$H(\mu|\alpha) \;=\; \int \log\Big(\frac{d\mu}{d\alpha}\Big) d\mu.$$

Exercise 4.4.8. Show that if $\mu \not\ll \alpha$, then

$$H(\mu|\alpha) \;=\; +\infty.$$

Exercise 4.4.9. Evaluate

$$p(n_1, \cdots, n_r) \;=\; \mathbb{P}\Big[\, L_n(\{x_1\}) = \frac{n_1}{n}, \dots, L_n(\{x_r\}) = \frac{n_r}{n}\,\Big],$$

where $n_1 + \cdots + n_r = n$.

Exercise 4.4.10. Using Stirling's formula, prove that

$$(2\pi)^{-\frac{r-1}{2}} n^{-\frac{r-1}{2}} e^{-\frac{r}{12}} \prod_{j=1}^{r} \Big(\frac{n\alpha(x_j)}{n_j}\Big)^{n_j} \;\leq\; p(n_1,\dots,n_r) \;\leq\; \prod_{j=1}^{r} \Big(\frac{n\alpha(x_j)}{n_j}\Big)^{n_j}.$$

Exercise 4.4.11. Compute I^1 for the Markov chain

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

where $p \in (0,1)$.

Exercise 4.4.12. Turning back to the context of the Chapter 3, show that the Cramer transform is the measure μ that minimizes the entropy $H(\mu|\alpha)$ among all the measures that have new mean $y \neq \mathbb{E}X$. In order to make life easier, suppose **(A1)** and **(A2)** for the measure α. You probably have to use variational calculus.

CHAPTER 5

PROOF'S SCHEME

5.1 The recipe: basic ingredients

Now, we describe the a general scheme of proof. Rather than a proof, this is a recipe to clarify ideas when proving large deviations for some model. Let P_n be a sequence of measures on $(\mathbf{X}, d)$ such that

$$P_n \xrightarrow{\text{d}} \delta_{x_0},$$

or else, the sequence of measures satisfy some Weak Law of Large Numbers. The first ingredient we need is a perturbed process P_n^β such that, for any x in the space, there is some perturbation β such that

$$P_n^\beta \xrightarrow{\text{d}} \delta_x.$$

Denote by $e^{a_n J_\beta}$ the Radon-Nikodym derivative between P_n^β and P_n. Note that a_n can be n, n^2 etc., depending on the scaling.

5.2 The recipe: upper bound

For a while, consider $C \subset \mathbf{X}$ a closed set.

$$\begin{aligned} P_n(C) \;=\; E_n\big[\mathbf{1}_C\big] \;=\; E_n\big[e^{-a_n J_\beta} e^{a_n J_\beta} \mathbf{1}_C\big] \;&\leq\; \sup_{x\in C} \Big\{e^{-a_n J_\beta(x)}\Big\}\, E\big[e^{a_n J_\beta} \mathbf{1}_C\big] \\ &\leq\; \sup_{x\in C} \Big\{e^{-a_n J_\beta(x)}\Big\}\, E\big[e^{a_n J_\beta}\big] \\ &=\; \sup_{x\in C} \Big\{e^{-a_n J_\beta(x)}\Big\}, \end{aligned}$$

because $E\left[e^{a_n J_\beta}\right]=1$. Therefore,

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(C)\;\leq\;-\inf_{x\in C}J_\beta(x).$$

Optimizing over the perturbations, we get

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(C)\;&\leq\;\inf_{\beta}\Big\{-\inf_{x\in C}J_\beta(x)\Big\}\\&=\;-\sup_{\beta}\;\inf_{x\in C}J_\beta(x).\end{aligned}$$

The question now is how to interchange the supremum with the infimum. Sometimes we can do that directly (as in Crámer's Theorem) but in the general it is required the Minimax Lemma:

Proposition 5.2.1 (Minimax Lemma). *Let $K\subset \mathbf{X}$ compact, where $(\mathbf{X},d)$ is a Polish space. Given $\{-J_\beta\}_\beta$ a family of upper semi-continuous functions, it holds*

$$\inf_{\substack{\mathcal{O}_1,\dots,\mathcal{O}_M\\ \text{covering of }K}}\;\max_{1\leq j\leq M}\;\inf_{\beta}\;\sup_{x\in\mathcal{O}_j}\;-J_\beta(x)\;\leq\;\sup_{x\in K}\inf_{\beta}\;-J_\beta(x).$$

For a proof of above, see [11].

Now, we repeat the steps above considering an open set A in lieu of the closed set C, arriving as well at

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(A)\;\leq\;-\sup_{\beta}\;\inf_{x\in A}J_\beta(x)\,,$$

because in any moment we have used that C was closed. The ensuing step is to pass from open sets to compacts sets, see next exercise.

Exercise 5.2.2. Let J_β be a set of lower semi-continuous functionals indexed on β, and suppose that

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(A)\;\leq\;-\sup_{\beta}\;\inf_{x\in A}J_\beta(x).$$

for any open set $A\subset\mathbf{X}$. Prove that

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(K)\;\leq\;-\inf_{x\in K}\;\sup_{\beta}J_\beta(x),\tag{5.1}$$

for any $K\subset\mathbf{X}$ compact. Hint: start with a finite open covering of K.

In possess of (5.1), we prove exponential tightness[1] and then use Exercise 2.4.6. This leads to

$$\limsup_{n\to\infty}\frac{1}{a_n}\log P_n(C)\;\leq\;-\inf_{x\in C}\;\sup_{\beta}J_\beta(x).$$

for any closed $C\subset X$, concluding the upper bound.

[1]By some means. Each model requires a specific proof of exponential tightness. Remember this chapter is not a proof of anything, it is just the structure of proof suitable for many cases.

5.3 The recipe: lower bound

Let $A \subset \mathbf{X}$ be an open set and let $x \in A$. Choose β a perturbation such that

$$P_n^\beta \xrightarrow{\text{d}} \delta_x. \tag{5.2}$$

Hence

$$P_n^\beta(A) \longrightarrow 1$$

as n goes to infinity. As before, we write

$$P_n(A) \;=\; E_n\big[\mathbf{1}_A\big] \;=\; E_n\big[e^{-a_n J_\beta} e^{a_n J_\beta} \mathbf{1}_A\big],$$

which implies

$$\frac{1}{a_n} \log P_n(A) \;=\; \frac{1}{a_n} \log E_n\big[e^{-a_n J_\beta} e^{a_n J_\beta} \mathbf{1}_A\big]. \tag{5.3}$$

In this point, we would like to put the logarithm inside the expectation achieving an inequality from below (by Jensen). However, the indicator function vanishes in a set of positive probability, which would gives us minus infinity (or else, with respect to the lower bound, nothing). For this reason, we rewrite (5.3) in the form

$$\begin{aligned}
\frac{1}{a_n} \log P_n(A) \;&=\; \frac{1}{a_n} \log \Big\{ E_n\big[e^{-a_n J_\beta} \tfrac{e^{a_n J_\beta} \mathbf{1}_A}{P_n^\beta(A)}\big] \cdot P_n^\beta(A) \Big\}\\
&=\; \frac{1}{a_n} \log E_n\big[e^{-a_n J_\beta} \tfrac{e^{a_n J_\beta} \mathbf{1}_A}{P_n^\beta(A)}\big] + \frac{1}{a_n} \log P_n^\beta(A).
\end{aligned}$$

Notice that

$$\frac{e^{a_n J_\beta} \mathbf{1}_A}{P_n^\beta(A)}$$

is a Radon-Nikodym derivative of a measure with respect to P_n (it is a non-negative function and its mean with respect to P_n is one). Thus, since the logarithm function is concave, we can apply Jensen inequality to obtain

$$\begin{aligned}
\frac{1}{a_n} \log P_n(A) \;&\geq\; E_n\Big[-J_\beta \cdot \frac{e^{a_n J_\beta} \mathbf{1}_A}{P_n^\beta(A)} \Big] + \frac{1}{a_n} \log P_n^\beta(A)\\
&=\; \frac{1}{P_n^\beta(A)} E_n^\beta\Big[-J_\beta \, \mathbf{1}_A \Big] + \frac{1}{a_n} \log P_n^\beta(A).
\end{aligned}$$

Recalling (5.2), we have

$$\lim_{n\to\infty} P_n^\beta(A) \;=\; 1,$$

then

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(A) \;\geq\; -J_\beta(x).$$

Remember that β was chosen trough a given x. Or else $\beta = \beta(x)$. The question we address now is: fixed $x \in \mathbf{X}$, if does hold

$$I(x) := \sup_{\gamma} J_\gamma(x) = J_{\beta(x)}(x).$$

Intuitively, the question is if the perturbation β is the cheapest way to observe the point x. This is a purely analytic question. Once this question is checked, we may close the large deviations proof. See next exercise:

Exercise 5.3.1. Suppose that, for each $x \in \mathbf{X}$ and $\beta = \beta(x)$,

$$I(x) := \sup_{\gamma} J_\gamma(x) = J_{\beta(x)}(x),$$

that $J_{\beta(x)}(x)$ is lower semi-continuous, and also that

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(A) \geq -J_{\beta(x)}(x),$$

for any open set $A \subset \mathbf{X}$. Use Exercise 2.4.7 to show that

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

5.4 Example: large deviations for hydrodynamics of exclusion

What it have been presented at these notes is only the central idea fo large deviations: challenges and difficulties may emerge in specific problems, subjects of vast research. In this section we discuss a few of them.

We warn the reader that we will be very hand waving in this section. The aim is only to fix the main ideias from previous subsections.

The SSEP (symmetric simple exclusion process) is an standard continuous time Markov chain in the state space $\{0,1\}^{T_n}$, where $T_n = \mathbb{Z}/n\mathbb{Z}$ is the discrete torus with n sites. Or else, in each site of T_n it is allowed at most one particle. Let us describe the dynamics of SSEP.

To each *edge* of T_n we associate a Poisson Point Process of parameter one, all of them independent. We start at a configuration of particles η_0. If the site x is empty we say that $\eta_0(x) = 0$. If it is occupied, we say that $\eta_0(x) = 1$. For instance, in the Figure 5.1, we have $\eta_0(2) = 0$ and $\eta_0(3) = 1$.

When a clock rings, we interchange the occupations at the vertices of the corresponding edge. Of course, if both sites are empty or occupied, nothing happens. See Figure 5.1 again.

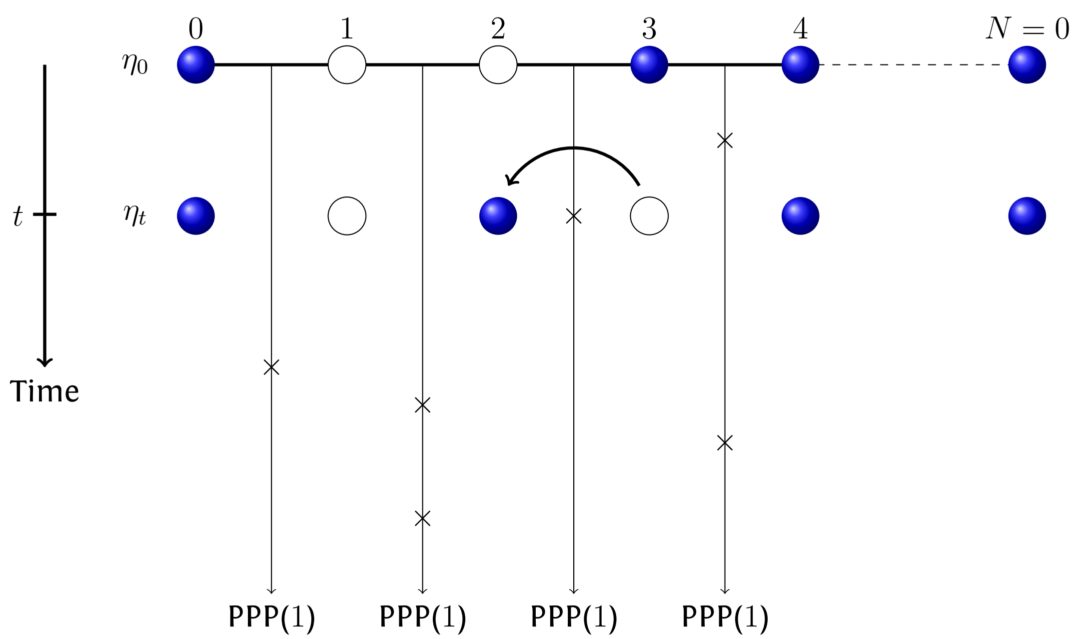

Figure 5.1: Symmetric Simple Exclusion Process

The spatial density of particles is usually represented by a measure in the continuous torus $[0,1]$, where we identify zero an one. This measure is called *empirical measure*, being defined by

$$\pi_t^n \;=\; \frac{1}{n}\sum_{x\in T_n} \eta_{n^2t}(x)\,\delta_{\frac{x}{n}}\,.$$

Or else, where there is a particle we put a Delta of Dirac with mass $1/n$. The time is taken as n^2t which is known as *diffusive scale*. It is a celebrated result the next theorem, which is called *hydrodynamic limit* of the SSEP.

Theorem 5.4.1. *Suppose that η_0, the initial configuration, is chosen in such a way*

$$\pi_0^n \xrightarrow{d} \gamma(u)\,du,$$

where γ is a fixed function, let us say, continuous. Then,

$$\{\pi_t^n, 0\leq t\leq T\} \xrightarrow{d} \{\rho(t,u)\,du\,,\,0\leq t\leq T\}, \tag{5.4}$$

where ρ is the solution of the heat equation in the torus:

$$\begin{cases}\partial_t\rho \;=\; \partial_{uu}\rho\,, & u\in[0,1]\,,\,t>0\,,\\ \rho(0,u) \;=\; \gamma(u)\,, & u\in[0,1]\,.\end{cases} \tag{5.5}$$

The result above is a law of large numbers. Therefore, the natural ensuing question is to prove large deviations. Denote by P_n the measure on the induced by π_t^n. Then, we can rewrite (5.4) as

$$P_n \xrightarrow{d} \delta_\rho,$$

where ρ is the solution of (5.5). What should be the perturbed process, as discussed before?

In this case, the family of perturbation will be taken as the Weakly Asymmetric Exclusion Process (WASEP). Denote by $\mathbb{T} = [0,1]$ the continuous torus. Let $H \in C^2(\mathbb{T} \times [0,T])$. Note that H do depend on time. The WASEP is the exclusion type process non-homogeneous in time which rate of jump from x to $x+1$ given by $\frac{1}{2}e^{H(\frac{x+1}{n})-H(\frac{x}{n})}$, and rate of jump from $x+1$ to x given by $\frac{1}{2}e^{H(\frac{x}{n})-H(\frac{x+1}{n})}$. For an illustration, see Figure 5.2.

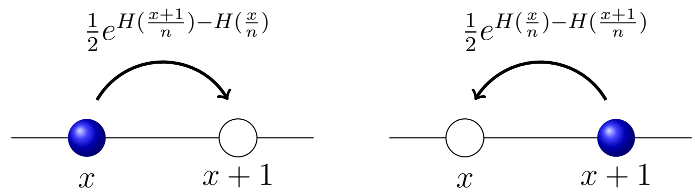

Figure 5.2: Weakly Assymmetric Exclusion Process: rates of jump

Denote by η_t^H the continuous time Markov Process on $\{0,1\}^{T_n}$ defined by the rates of jump above. Analogously, we define also

$$\pi_t^{n,H} \;=\; \frac{1}{n}\sum_{x\in T_n} \eta^H_{n^2t}(x)\,\delta_{\frac{x}{n}}$$

which represents the spatial density of particles in the diffusive scale. It is a fact of life that (see Kipnis/Landim book [11])

Theorem 5.4.2. *Suppose that η_0^H, the initial configuration, is chosen in such a way*

$$\pi_0^{n,H} \xrightarrow{d} \gamma(u)\,du,$$

where γ is a fixed function, let us say, continuous. Then,

$$\{\pi_t^{n,H}, 0\leq t\leq T\} \xrightarrow{d} \{\rho(t,u)\,du\,,\, 0\leq t\leq T\}, \tag{5.6}$$

where ρ is the solution of the heat equation in the torus:

$$\begin{cases} \partial_t\rho \;=\; \partial_{uu}\rho - 2\,\partial_u\big(\rho(1-\rho)\partial_u H\big)\,, & u\in[0,1]\,,\, t>0\,,\\ \rho(0,u) \;=\; \gamma(u)\,, & u\in[0,1]\,. \end{cases} \tag{5.7}$$

Denote by P_n^H the measure on the induced by $\pi_t^{n,H}$. Then, we can rewrite (5.6) as

$$P_n^H \xrightarrow{d} \delta_{\rho^H},$$

where ρ^H is the solution of (5.7). Here, H is perturbation. If we want to observe a profile ρ, we replace that function ρ in (5.7), then find the correct perturbation H.

In the case of Markov process, there is two incomings of deviations: from the initial measure, and from the dynamics. Since we do not have the ambition of completeness now, we suppose that the initial configurations have been chosen in a deterministic way. That forbids deviations from the initial measure. Denote by $\langle \cdot, \cdot \rangle$ the L^2 inner product in $\mathbb{T}$.

For $\pi = \rho\, du$, let us accept that

$$J_H(\pi) \;=\; \ell_H(\pi) - \int_0^T \langle \rho(1-\rho), (\partial_u H)^2 \rangle \, dt \tag{5.8}$$

where

$$\ell_H(\pi) \;=\; \langle \rho_T, H_T \rangle - \langle \rho_0, H_0 \rangle - \int_0^T \langle \rho_t, (\partial_t + \partial_{uu}) H \,\rangle \, dt.$$

To simplify the discussion, let us say that ρ is always smooth. Let us debate in the exercises the question addressed before: if hold the equality

$$I(\rho^H) \;:=\; \sup_G J_G(\rho^H) \;=\; J_H(\rho^H). \tag{5.9}$$

We finish the text with a few defiances one can face up to when dealing with large deviations:

- The set of perturbations does not lead the system to all possible points in metric space, but only in a dense subset of the metric space. In that case is necessary to show a I-density or to show that the set of not accessible points has probability super-exponentially small and use Remark 2.1.4. See [1, 3] on the question of I-density.

- The measures P_n are induced by some random element and the functional J_β is not a function of that random element. In general, asymptotically it is. That is the case in large deviations of the SSEP. It is also the case of Markov chains, see Exercise 4.4.1.

- The set of perturbations is large. Instead of helping, that may complicate the scenario because in that case, it would be harder to find the cheapest perturbation.

5.5 Exercises

Exercise 5.5.1. Suppose that ρ^H is the solution of (5.7). Use integration by parts to show that

$$\langle \rho^H_T, G_T \rangle - \langle \rho^H_0, G_0 \rangle - \int_0^T \langle \rho^H_t, (\partial_t + \partial_{uu}) G \rangle \, dt - \int_0^T \langle \rho^H(1-\rho^H), \partial_u G \, \partial_u H \rangle \, dt = 0, \tag{5.10}$$

for any $G \in C^2(\mathbb{T} \times [0,T])$.

Exercise 5.5.2. Assume and (5.8) and (5.10) to prove that

$$\sup_G J_G(\rho^H) \; = \; J_H(\rho^H)$$

where $H, G \in C^2(\mathbb{T} \times [0, T])$. Hint: replace (5.10) in the expression of J_G.

CHAPTER 6

SOLUTIONS

Solution Exercise 1.4.5: It would be harder to guarantee that $\sum_{n=1}^{\infty}(d_n - d_{n+1})$ is convergent because of the alternating signs.

Solution Exercise 1.4.6: Make a partition of the image in small intervals and take pre-images.

Solution Exercise 1.4.7: ($\Leftarrow$) By assumption, we have that

$$\{z\,;\ f(z) \leq \liminf_{x\to x_0} f(x) + \tfrac{1}{n}\}$$

is closed. Hence,

$$\bigcap_{n\in\mathbb{N}}\{z\,;\ f(z) \leq \liminf_{x\to x_0} f(x) + \tfrac{1}{n}\}$$

is closed. Taking a sequence $z_n \to x_0$ such that $f(z_n)$ converges to the $\liminf_{x\to x_0} f(x)$, we conclude that

$$x_0 \in \bigcap_{n\in\mathbb{N}}\{z\,;\ f(z) \leq \liminf_{x\to x_0} f(x) + \tfrac{1}{n}\}$$

i.e., $f(x_0) \leq \liminf_{x\to x_0} f(x)$.

Solution Exercise 2.4.5: Recall Exercise 1.4.7 and the fact that an arbitrary intersection of closed sets is closed.

Solution Exercise 2.4.7:

$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(A) \;\geq \liminf_{n\to\infty} \frac{1}{a_n} \log P_n(U_\delta) \;\geq\; -I(y).$$

Now, optimize over $y \in A$.

Solution Exercise 3.2.2: Let $\frac{1}{p}+\frac{1}{q}=1$ and $p>0$. By Hölder's inequality, we have that

$$\mathbb{E}\Big[e^{p\theta_1+q\theta_2}\Big] \;\leq\; \mathbb{E}\Big[\big(e^{p\theta_1}\big)^p\Big]^{\frac{1}{p}}\cdot\mathbb{E}\Big[\big(e^{q\theta_2}\big)^q\Big]^{\frac{1}{q}}.$$

Taking logarithms, we get the convexity. Notice that Hölder's inequality holds even if some L_p-norm in the right hand side is not finite.

Solution Exercise 3.2.5:

(a) Bernoulli(p).

The moment generating function is $M(\theta)=(1-p)+pe^\theta$, and the rate function is

$$I(x) \;=\; \begin{cases} x\log\frac{x}{p}+(1-x)\log\frac{1-x}{1-p}, & \text{if } x\in(0,1)\\ \infty, & \text{otherwise}\end{cases}$$

(b) Exponential(λ).

The moment generating function is

$$M(\theta) \;=\; \begin{cases} \frac{\lambda}{\lambda-\theta}, & \text{if } \theta<\lambda\\ \infty, & \text{otherwise}\end{cases}$$

(c) Poisson(λ). The moment generating function is $M(\theta)=e^{\lambda(e^\theta-1)}$, and the rate function is

(d) Normal(μ,σ). The rate function is

$$I(x) \;=\; \frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2$$

Solution Exercise 3.2.7: Notice that α and $\tilde{\alpha}$ can be mutually singular (give an example). Since the the changing of measure's price is related to the Radon-Nikodym derivative, we wouldn't get anything.

Solution Exercise 4.4.2: Observe that the expression $\int f\,d\mu-\log\int e^f\,d\alpha$ remains invariant if we replace f by f plus a constant.

Solution Exercise 4.4.4: Take $f=\mathbf{1}_A$ in the entropy inequality.

Solution Exercise 4.4.9: Being $n_1 + \cdots + n_r = n$,

$$p(n_1, \cdots, n_r) \;=\; \frac{n!}{n_1! \cdots n_r!}\, p_1^{r_1} \cdots p_r^{n_r}\,,$$

where $p_k = \mathbb{P}[X_1 = x_k]$.

BIBLIOGRAPHY

[1] Bertini, L.; Landim, C.; Mourragui, M.: *Dynamical large deviations of boundary driven weakly asymmetric exclusion processes*. Annals of Probability, v. 37, p. 2357-2403 (2009).

[2] A. Dembo, O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag. Series: Stochastic Modelling and Applied Probability, Vol. 38 (1998).

[3] Farfan, J.; Landim, C.; Mourragui, M.: *Hydrostatics and dynamical large deviations of boundary driven gradient symmetric exclusion*. Stochastic Processes and their Applications, Volume 121, Issue 4, Pages 725–758 (2011).

[4] W. Feller. *An Introduction to Probability Theory and Its Applications*, Vol. 1. Wiley Series in Mathematical Statistics (1968).

[5] F. den Hollander. *Large Deviations*. Fields Institute Monographs, American Mathematical Society (2000).

[6] S. R. S. Varadhan. *Large Deviations*. The Annals of Probability, Special Invited Paper, Vol. 36, No. 2, 397–419 (2008).

[7] S. R. S. Varadhan. *Large Deviations and Applications*. Courant Institute of Mathematical Sciences, New York University (1984).

[8] E. Olivieri, M. E. Vares, *Large Deviations and Metastability*. Cambridge University Press, (2005).

[9] P. Fernandez. *Introdução à Teoria das Probabilidades*. LTC, Livros Técnicos e Científicos. UNB (1973).

[10] L. Evans. *Partial Differential Equations*. [Graduate Studies in Mathematics], American Mathematical Society (1998).

[11] C. Kipnis, C. Landim. *Scaling limits of interacting particle systems*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 320. Springer-Verlag, Berlin (1999).