UNIVERSITAT DE BARCELONA

ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

Large Deviations

Author:
Oriol Zamora Font

Supervisor:
Marta Sanz-Solé

Facultat de Matemàtiques i Informàtica

June 28, 2020

Abstract

The objective of this project is to give an introduction to the theory of large deviations (LDP), a topic in stochastic analysis that can be described as the asymptotic evaluation of small probabilities at exponential scale. We start with the fundamental and initial result by Cramér (1938) and then, we formulate general LDP principles. A basic result in the field of large deviations for stochastic processes is Schilder's Theorem regarding Brownian motion. A proof of this result is given in Chapter 4. Finally, we develop part of the Freidlin-Wentzell theory and give an application to LDPs for stochastic differential equations.

Large deviations is a very active research area with many applications namely, in statistics, finance, engineering, statistical mechanics and applied probability. Nevertheless, because of time and space constrains applications are not considered in this work.

Acknowledgments

I would like to gratefully thank my supervisor, Professor Marta Sanz-Solé, both for proposing me such an interesting topic and for her very detailed help throughout the development of this project.

Contents

1 Introduction

1.1 Example

In order to motivate the large deviation principle, we give an example involving the most classical topic of probability theory, namely, the behaviour of the empirical mean of independent identically distributed random variables.

Let $\{X_i\}$ be a sequence of independent identically distributed random variables $X_i : \Omega \to \mathbb{R}$ with $X_1 \overset{\text{d}}{=} N(0,1)$. We denote, for $n \geq 1$, the empirical mean $\widehat{S}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then, by the strong law of large numbers

$$\widehat{S}_n \xrightarrow[n\to\infty]{\text{a.s.}} E(X_1) = 0.$$

Since almost sure convergence implies convergence in probability, we have for any $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\widehat{S}_n| \geq \delta) = 0.$$

So, as $n \to \infty$, the event $\{|\widehat{S}_n| \geq \delta\}$ is unlikely to occur. However, it can be interesting to have a more precise control of this unlikeliness. For instance, both sequences $\{\frac{1}{n}\}$ and $\{e^{-n}\}$ tend to 0 as $n \to \infty$ but the second one does it significantly faster. Hence, we are interested in determining the speed of convergence to 0 of the sequence $\{\mathbb{P}(|\widehat{S}_n| \geq \delta)\}$.

Since $\widehat{S}_n \overset{\text{d}}{=} N(0, 1/n)$, we have

$$\mathbb{P}(|\widehat{S}_n| \geq \delta) = 2\int_\delta^\infty \sqrt{\frac{n}{2\pi}} e^{\frac{-nx^2}{2}} \mathrm{d}x = \sqrt{\frac{2n}{\pi}} \int_\delta^\infty e^{\frac{-nx^2}{2}} \mathrm{d}x. \tag{1.1.1}$$

The following inequalities hold:

$$\left(\frac{1}{\delta} - \frac{1}{\delta^3}\right) e^{\frac{-n\delta^2}{2}} \overset{(1)}{\leq} \int_\delta^\infty e^{\frac{-nx^2}{2}} \mathrm{d}x \overset{(2)}{\leq} \frac{1}{\delta} e^{\frac{-n\delta^2}{2}}. \tag{1.1.2}$$

Indeed, to prove inequality (1) note that

$$\begin{aligned} &\int_\delta^\infty e^{\frac{-nx^2}{2}} \mathrm{d}x \geq \int_\delta^\infty \left(1 - \frac{n-1}{x^2} - \frac{3}{x^4}\right) e^{\frac{-nx^2}{2}} \mathrm{d}x \\ &= \int_\delta^\infty \frac{\mathrm{d}}{\mathrm{d}x}\left[-\left(\frac{1}{x} - \frac{1}{x^3}\right) e^{\frac{-nx^2}{2}}\right] \mathrm{d}x = \left(\frac{1}{\delta} - \frac{1}{\delta^3}\right) e^{\frac{-n\delta^2}{2}}. \end{aligned}$$

Similarly, to prove inequality (2) in (1.1.2)

$$\int_\delta^\infty e^{\frac{-nx^2}{2}} \mathrm{d}x \leq \int_\delta^\infty \left(1 + \frac{1}{x^2}\right) e^{\frac{-nx^2}{2}} \mathrm{d}x$$

$$= \int_\delta^\infty \frac{\mathrm{d}}{\mathrm{d}x}\left[-\frac{1}{x}e^{\frac{-nx^2}{2}}\right]\mathrm{d}x = \frac{1}{\delta}e^{\frac{-n\delta^2}{2}}.$$

So, by (1.1.1) and (1.1.2) we have

$$\sqrt{\frac{2n}{\pi}}\left(\frac{1}{\delta}-\frac{1}{\delta^3}\right)e^{\frac{-n\delta^2}{2}} \leq \mathbb{P}(|\widehat{S}_n| \geq \delta) \leq \sqrt{\frac{2n}{\pi}}\frac{1}{\delta}e^{\frac{-n\delta^2}{2}}.$$

We conclude that when n is very big the term $\mathbb{P}(|\widehat{S}_n| \geq \delta)$ behaves like $e^{\frac{-n\delta^2}{2}}$. One way to write this fact is by

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|\widehat{S}_n| \geq \delta) \leq \limsup_{n\to\infty} \frac{1}{n}\log\left(\sqrt{\frac{2n}{\pi}}\frac{1}{\delta}e^{\frac{-n\delta^2}{2}}\right) = -\frac{\delta^2}{2}$$

and

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|\widehat{S}_n| \geq \delta) \geq \liminf_{n\to\infty} \frac{1}{n}\log\left(\sqrt{\frac{2n}{\pi}}\left(\frac{1}{\delta}-\frac{1}{\delta^3}\right)e^{\frac{-n\delta^2}{2}}\right) = -\frac{\delta^2}{2}.$$

So,

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(|\widehat{S}_n| \geq \delta) = -\frac{\delta^2}{2}. \tag{1.1.3}$$

This is an example of a large deviations statement: we do not only know that the sequence $\{\mathbb{P}(|\widehat{S}_n| \geq \delta)\}$ goes to 0, but also that it does with this exponentially fast ratio.

If we write μ_n for the law of $\widehat{S}_n$, equation (1.1.3) is

$$\lim_{n\to\infty} \frac{1}{n}\log \mu_n((-\infty,-\delta]\cup[\delta,\infty)) = -\frac{\delta^2}{2}. \tag{1.1.4}$$

At this point, one can ask several questions:

1. What happens in equation (1.1.4) if we replace $(-\infty,-\delta]\cup[\delta,\infty)$ for a general set $A \in \mathcal{B}(\mathbb{R})$?

2. Is there a similar result as (1.1.4) for independent identically distributed random sequences non necessarily Gaussian?

3. Is there an analogue to (1.1.4) when μ_n is not the law of the empirical mean of independent identically real-valued random variables but rather a general family of probabilities $\{\mu_n\}$ that converges in distribution to a degenerate measure?

The two first questions will be solved in Chapter 2 and now we are going to introduce the general problem in large deviations theory to have a more precise understanding of the last question.

1.2 The general problem

We now describe the general problem that we will consider in this the project without giving the precise definitions, which will be given in the next section. The situation will be the following: we have a sequence $\{X_n\}$ of $\mathcal{X}$-valued random variables, where $(\mathcal{X}, d)$ is a separable metric space, that converge in probability to a fixed element $x \in \mathcal{X}$. That is, for all $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}(d(X_n, x) \geq \delta) = 0.$$

The element $x \in \mathcal{X}$ can be thought as the expected way to behave of X_n when n is very big. Since convergence in probability implies convergence in distribution, if we write μ_n for the law of X_n and δ_x for the degenerate measure concentrated at $x \in \mathcal{X}$, we have

$$\limsup_{n\to\infty} \mu_n(F) \leq \delta_x(F), \qquad \text{for any } F \subset \mathcal{X} \text{ closed.}$$
$$\liminf_{n\to\infty} \mu_n(G) \geq \delta_x(G), \qquad \text{for any } G \subset \mathcal{X} \text{ open.}$$

We write $\mathcal{B}(\mathcal{X})$ for the Borel σ-field on $\mathcal{X}$ and for any $A \subset \mathcal{X}$, $\overline{A}$ its closure and A^{o} its interior. Note that two previous inequalities are equivalent to the following ones for any $A \in \mathcal{B}(\mathcal{X})$

$$\delta_x(A^{\text{o}}) \leq \liminf_{n\to\infty} \mu_n(A^{\text{o}}) \leq \liminf_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\overline{A}) \leq \delta_x(\overline{A}). \quad (1.2.1)$$

Observe that for any $A \in \mathcal{B}(\mathcal{X})$ such that $x \notin \overline{A}$, we have

$$\limsup_{n\to\infty} \mu_n(A) \leq \limsup_{n\to\infty} \mu_n(\overline{A}) \leq \delta_x(\overline{A}) = 0 \Rightarrow \lim_{n\to\infty} \mu_n(A) = 0.$$

With a similar objective as in the previous example, we are interested to study the possible exponential velocity of the convergence to 0 of the sequence $\{\mu_n(A)\}$. We would like to find $r(A) \in (0, \infty)$ such that $\mu_n(A)$ behaves like $e^{-nr(A)}$ for n very big, namely,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(A) = -r(A). \quad (1.2.2)$$

Now we discuss what properties should the set function r have. Let $A_1, ..., A_n \in \mathcal{B}(\mathcal{X})$ be disjoints sets, so that $\mu_n(A_1 \cup ... \cup A_n) = \mu_n(A_1) + ... + \mu_n(A_n)$. Then, for n very by, we expect that $\mu_n(A_1 \cup ... \cup A_n)$ behaves like $e^{-nr(A_1 \cup ... \cup A_n)}$ at the same time that we expect that $\mu_n(A_1) + ... + \mu_n(A_n)$ behaves like $e^{-nr(A_1)} + ... + e^{-nr(A_n)}$. Note that in fact, $e^{-nr(A_1)} + ... + e^{-nr(A_n)}$ is asymptotically $e^{-n\min(r(A_1),...,r(A_n))}$. In summary, we should have

$$r(A_1 \cup ... \cup A_n) = \min(r(A_1), ..., r(A_n)).$$

This suggests to consider r of the form

$$r(A) = \inf_{x \in A} I(x), \tag{1.2.3}$$

where I is a function on $\mathcal{X}$. Then, motivated by (1.2.1), (1.2.2) and (1.2.3), in order to conclude that the sequence $\{\mu_n(A)\}$ have an exponential decay we should find a function $I : \mathcal{X} \to [0, \infty]$ such that

$$-\inf_{x \in A^{\mathrm{o}}} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(A) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(A) \leq -\inf_{x \in \overline{A}} I(x). \tag{1.2.4}$$

So, in general we do not expect that the limit of $\frac{1}{n} \log \mu_n(A)$ exists but rather to have a lower and upper bound of the exponential velocity of the convergence of $\{\mu_n(A)\}$ to 0. In addition, we allow the function I to take the values 0 and ∞ in order to no exclude the sets $A \in \mathcal{B}(\mathcal{X})$ that satisfy $\lim_{n \to \infty} \mu_n(A) = 1$ because $x \in A^{\mathrm{o}}$ or because the exponential decay of $\mu_n(A)$ is faster than $e^{-nr(A)}$.

Moreover, we can assume that the function I is lower semicontinuous, that is, for every $\alpha \in [0, \infty)$, the level set

$$\psi_I(\alpha) := \{x \in \mathcal{X} : \ I(x) \leq \alpha\}$$

is a closed subset of $\mathcal{X}$. Namely, if I satisfies (1.2.4) for every $A \in \mathcal{B}(\mathcal{X})$ there exists a lower semicontinuous function I_{lsc} that also satisfies (1.2.4) for every $A \in \mathcal{B}(\mathcal{X})$ and that $I_{\mathrm{lsc}} \leq I$. This is interesting because we will be able to use properties of lower semicontinuous functions. To prove this, we define

$$I_{\mathrm{lsc}}(x) := \sup_{\{G \text{ neighbourhood of } x\}} \inf_{y \in G} I(y).$$

Then, by this definition, $I_{\mathrm{lsc}} \leq I$. Consider $x \in \psi_{I_{\mathrm{lsc}}}(\alpha)^{\mathrm{c}} = \{x \in \mathcal{X} : \ I_{\mathrm{lsc}}(x) > \alpha\}$. By the definition of the supremum, there exists a neighbourhood G_x of x such that $\inf_{y \in G_x} I(y) > \alpha$. Since for every $y \in G_x$, G_x is also a neighbourhood of y, we have that

$$I_{\mathrm{lsc}}(y) \geq \inf_{y \in G_x} I(y) > \alpha \text{ for every } y \in G_x \Rightarrow x \in G_x \subset \psi_{I_{\mathrm{lsc}}}(\alpha)^{\mathrm{c}} \Rightarrow \psi_{I_{\mathrm{lsc}}}(\alpha) \text{ is closed.}$$

This proves that I_{lsc} is lower semicontinuous. Finally, since $I_{\mathrm{lsc}} \leq I$ and by a similar argument to the previous one, if $A \in \mathcal{B}(\mathcal{X})$ we have

$$\inf_{x \in \overline{A}} I_{\mathrm{lsc}}(x) \leq \inf_{x \in \overline{A}} I(x) \quad \text{ and } \quad \inf_{x \in A^{\mathrm{o}}} I_{\mathrm{lsc}}(x) = \inf_{x \in A^{\mathrm{o}}} I(x).$$

Therefore, I_{lsc} satisfies (1.2.4) for every $A \in \mathcal{B}(\mathcal{X})$. Furthermore, in Chapter 3 we are going to prove that there exists at most one lower semicontinuous functions satisfying (1.2.4) for every $A \in \mathcal{B}(\mathcal{X})$.

1.3 Rate functions

We now give the precise definition of a rate function and some properties without proofs about them which will be used throughout this project. In general, some auxiliary results will not be proved in this project, but the ones involving the large deviations topic will be proved in detail.

For the rest of this chapter, unless otherwise specified, $\mathcal{X}$ will denote a topological space. So, open and closed subsets of $\mathcal{X}$ are well-defined and for any $\Gamma \subset \mathcal{X}$ we denote $\overline{\Gamma}$ for its closure and Γ^o for its interior. For the rest of this project, the infimum of a function over an empty set is interpreted as ∞.

Definition 1.3.1. A function $f : \mathcal{X} \to [-\infty, \infty]$ is lower semicontinuous if, for all $\alpha \in \mathbb{R}$, the level set

$$\psi_I(\alpha) := \{x \in \mathcal{X} : \; I(x) \leq \alpha\}$$

is a closed subset of $\mathcal{X}$.

Definition 1.3.2. A rate function $I : \mathcal{X} \to [0, \infty]$ is a lower semicontinuous function. A good rate function is a rate function whose level sets $\psi_I(\alpha)$ are compact subsets of $\mathcal{X}$ for all $\alpha \in [0, \infty)$.

Proposition 1.3.3. *Let $f : \mathcal{X} \to [-\infty, \infty]$ be a lower semicontinuous function. Then, for all $x \in \mathcal{X}$,*

$$f(x) = \sup_{\{G \text{ neighborhood of } x\}} \inf_{y \in G} f(y).$$

Proof. For a reference, see [1], line 3 in page 117. ∎

When working with metric spaces the following equivalence will be useful.

Proposition 1.3.4. *Let $(\mathcal{X}, d)$ be a metric space and $f : \mathcal{X} \to [-\infty, \infty]$ a function. Then, f is lower semicontinuous if and only if for every $x \in \mathcal{X}$ and $\{x_n\} \subset \mathcal{X}$ with $\lim_{n\to\infty} x_n = x$, we have $\liminf_{n\to\infty} f(x_n) \geq f(x)$.*

Proof. For a reference, see [1], line -4 in page 4. ∎

Lemma 1.3.5. Let I be a rate function. Then, for any set $\Gamma \subset \mathcal{X}$

$$\lim_{\delta\to 0} \inf_{x\in\Gamma} I^\delta(x) = \inf_{x\in\Gamma} I(x),$$

where I^δ is the δ-rate function defined by $I^\delta(x) := \min\left\{I(x) - \delta, \frac{1}{\delta}\right\}$.

Proof. For a reference, see [1] page 6, equation (1.2.10). ■

In fact, most of the time we will work with good rate functions. The following results show some advantages to do it.

Lemma 1.3.6. Let $F \subset \mathcal{X}$ be a non-empty closed set and I a good rate function. Then, I achieves its infimum over F.

Proof. For a reference, see [1], line -2 in page 4, . ■

Lemma 1.3.7. Let I be a good rate function.

1. Let $\{F_\delta\}_{\delta>0}$ be a nested family of closed sets, that is, $F_\delta \subset F_{\delta'}$ if $\delta < \delta'$.

   Define $F_0 := \bigcap_{\delta>0} F_\delta$. Then,

$$\inf_{y\in F_0} I(y) = \lim_{\delta\to 0} \inf_{y\in F_\delta} I(y).$$

2. Suppose that $(\mathcal{X}, d)$ is a metric space. Then, as a consequence of 1, for any set $A \subset \mathcal{X}$,

$$\inf_{y\in \overline{A}} I(y) = \lim_{\delta\to 0} \inf_{y\in A^\delta} I(y),$$

   where

$$A^\delta := \left\{ y \in \mathcal{X} : d(y, A) := \inf_{z\in A} d(y, z) \le \delta \right\}$$

   denotes the closed blowup of A.

Proof. For a reference, see [1], Lemma 4.1.6 in page 119. ■

1.4 The Large Deviation Principle

We finally give the formal definition of the large deviation principle. Let $\{\mu_n\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{B}(\mathcal{X})$ is the Borel σ-Borel field on $\mathcal{X}$. Recall that our objective is to study the possible exponential decay of the sequences $\{\mu_n(\Gamma)\}$ where $\Gamma \in \mathcal{B}(\mathcal{X})$. Following the ideas that lead us to write equation (1.2.4) we have the following definition.

Definition 1.4.1. The family $\{\mu_n\}$ satisfies a LDP (large deviation principle) with a rate function I if, for all $\Gamma \in \mathcal{B}(\mathcal{X})$,

$$-\inf_{x\in\Gamma^\circ} I(x) \le \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le -\inf_{x\in\overline{\Gamma}} I(x). \tag{1.4.1}$$

We will refer to the right- and left hand sides of (1.4.1) as the upper and lower bound of the LDP, respectively. To have a more clear understanding of the implications of the LDP we make the following observation.

Observation 1.4.2. Suppose that the family $\{\mu_n\}$ satisfies a LDP with a rate function I and let $\Gamma \in \mathcal{B}(\mathcal{X})$. The three main situations are

1. $\inf_{x\in\overline{\Gamma}} I(x) > 0$. Then, $\mu_n(\Gamma)$ behaves, for big n, like a sequence between

$$e^{-n \inf_{x\in\Gamma^{o}} I(x)} \quad \text{and} \quad e^{-n \inf_{x\in\overline{\Gamma}} I(x)}.$$

2. $\inf_{x\in\Gamma^{o}} I(x) = 0$. Then, we can not deduce an exponential decay of $\mu_n(\Gamma)$. Typically, this will happen when $x \in \Gamma^{o}$ where δ_x is the limit in distribution of μ_n. In that case, $\lim_{n\to\infty} \mu_n(\Gamma) = 1$.

3. $\inf_{x\in\overline{\Gamma}} I(x) = \infty$. Then, the exponential decay of $\mu_n(\Gamma)$ is faster than the first case. For example, it can be $e^{-n^2 r(\Gamma)}$, where $r(\Gamma) > 0$ is some constant depending on Γ.

Typically, for the LDP that we are going to prove in this project the rate function will be a good rate function strictly positive except for an element $x_0 \in \mathcal{X}$. In such case, we have the following nice consequence of the LDP.

Proposition 1.4.3. *Suppose that the family $\{\mu_n\}$ satisfies a LDP with a good rate function I satisfying $I(x) > 0$ for all $x \neq x_0$ and $I(x_0) = 0$ for some element $x_0 \in \mathcal{X}$. Then, $\{\mu_n\}$ converge in distribution to δ_{x_0} and for any $\Gamma \in \mathcal{B}(\mathcal{X})$ with $x_0 \notin \overline{\Gamma}$ we are in case 1 of Observation 1.4.2, that is, we have a control of the exponential decay of $\{\mu_n(\Gamma)\}$.*

Proof.
(1) In order to prove that $\{\mu_n\}$ converge in distribution to δ_{x_0} we check that for any closed set $F \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \mu_n(F) \leq \delta_{x_0}(F).$$

If $x_0 \in F$, the previous inequality is clear. If $x_0 \notin F$, then $\inf_{x\in F} I(x) > 0$. Otherwise, by Lemma 1.3.6 we would have $x_0 \in F$. Then, by the upper bound of the LDP,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} I(x) < 0 \Rightarrow \limsup_{n\to\infty} \mu_n(F) = 0 = \delta_{x_0}(F).$$

(2) By the second part, if $\Gamma \in \mathcal{B}(\mathcal{X})$ is such that $x_0 \notin \overline{\Gamma}$, repeating the same argument we have that $\inf_{x\in\overline{\Gamma}} I(x) > 0$. So, we are in case 1 of Observation 1.4.2. ■

Observation 1.4.4. The large deviation principle can be formulated for a continuous indexed family of probabilities $\{\nu_\varepsilon\}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ where $\varepsilon > 0$. The definitions and results in this chapter are analogous for this case after replacing the factor n by $\frac{1}{\varepsilon}$. For instance, the family $\{\nu_\varepsilon\}$ satisfies a LDP with a rate function I if, for all $\Gamma \in \mathcal{B}(\mathcal{X})$,

$$-\inf_{x\in\Gamma^o} I(x) \leq \liminf_{\varepsilon\to 0} \varepsilon \log \nu_\varepsilon(\Gamma) \leq \limsup_{\varepsilon\to 0} \varepsilon \log \nu_\varepsilon(\Gamma) \leq -\inf_{x\in\overline{\Gamma}} I(x).$$

1.4.1 Weak LDP and exponential tightness

In fact, when proving a LDP we will use the following equivalent version.

Observation 1.4.5. The condition in (1.4.1) is equivalent to

1. For any closed set $F \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} I(x).$$

2. For any open set $G \subset \mathcal{X}$,

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

A natural strategy to prove to upper bound of the LDP is to first try with compact subsets. To avoid measurability issues, we consider that the topological space $\mathcal{X}$ is Hausdorff.

Definition 1.4.6. A topological space $\mathcal{X}$ is Hausdorff if every $x, y \in \mathcal{X}$, $x \neq y$, have disjoint neighborhoods.

Then, compacts subsets of $\mathcal{X}$ are closed. We have the following weaker version of the LDP, which under the condition of exponential tightness can be strengthened to the standard LDP.

Definition 1.4.7. The family $\{\mu_n\}$ satisfies a weak LDP with a rate function I if,

1. For any compact set $K \subset \mathcal{X}$,

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

2. For any open set $G \subset \mathcal{X}$,

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

Definition 1.4.8. The family $\{\mu_n\}$ is exponentially tight if for every $\alpha < \infty$, there exists a compact subset $K_\alpha \subset \mathcal{X}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha.$$

Before proving the main result of this subsection we give the following Lemma which will be very useful during this project.

Lemma 1.4.9. Let $N \geq 1$ and $\{a_{\varepsilon,i}\}$ with $a_{\varepsilon,i} \geq 0$ for $1 \leq i \leq N$ and $\varepsilon > 0$. Then,

$$\limsup_{\varepsilon\to 0} \varepsilon \log \left(\sum_{i=1}^{N} a_{\varepsilon,i} \right) = \max_{1\leq i\leq N} \limsup_{\varepsilon\to 0} \varepsilon \log a_{\varepsilon,i}, \tag{1.4.2}$$

and

$$\liminf_{\varepsilon\to 0} \varepsilon \log \left(\sum_{i=1}^{N} a_{\varepsilon,i} \right) \leq \max_{1\leq i\leq N} \liminf_{\varepsilon\to 0} \varepsilon \log a_{\varepsilon,i}. \tag{1.4.3}$$

Proof. For a reference, see [1], Lemma 1.2.15 in page 7. ■

Lemma 1.4.10. Suppose that $\{\mu_n\}$ is an exponentially tight family of probability measures and satisfies a weak LDP with a rate function I. Then, I is a good rate function and the LDP holds.

Proof.
(1) We first prove that I is a good rate function. We want to prove that the level set $\psi_I(\alpha)$ is compact for every $\alpha \in [0, \infty)$. Consider the compact subset $K_\alpha \subset \mathcal{X}$ such that $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha$. Since $K_\alpha^c \subset \mathcal{X}$ is open, by the lower bound of the weak LDP,

$$-\inf_{x\in K_\alpha^c} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha.$$

This implies that,

$$\inf_{x\in K_\alpha^c} I(x) > \alpha \Rightarrow \psi_I(\alpha) \subset K_\alpha.$$

Since $\psi_I(\alpha)$ is closed and K_α is compact, we conclude that I is a good rate function.

(2) We now prove the upper bound of the LDP. Let $F \subset \mathcal{X}$ be a closed set, $\alpha < \infty$ and $K_\alpha \subset \mathcal{X}$ a compact subset such that $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha$. Then,

$$\mu_n(F) \leq \mu_n(F \cap K_\alpha) + \mu_n(K_\alpha^c). \tag{1.4.4}$$

Applying (1.4.2) of Lemma 1.4.9, the fact that $F \cap K_\alpha \subset \mathcal{X}$ is a compact subset and (1.4.4) we obtain,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) &\leq \max\left(\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F \cap K_\alpha), \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^{\mathrm{c}})\right) \\ &\leq \max\left(-\inf_{x\in F\cap K_\alpha} I(x), -\alpha\right) \leq \max\left(-\inf_{x\in F} I(x), -\alpha\right).\end{aligned}$$

Since the last inequality is true for all $\alpha < \infty$ we conclude by letting $\alpha \to \infty$ that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} I(x).$$

■

2 Cramér's Theorem in $\mathbb{R}$

2.1 Introduction

In this section we are going to prove Cramér's Theorem in $\mathbb{R}$: the LDP of the empirical mean of independent indentically distributed random variables in $\mathbb{R}$. We start giving a motivation to study this theorem.

Let $\{X_i\}$ be a sequence of independent identically distributed random variables $X_i : \Omega \to \mathbb{R}$. For $n \geq 1$, we denote by μ_n the law of the empirical mean $\widehat{S}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Suppose that $X_1 \in L^1(\Omega)$, then by the strong law of large numbers

$$\widehat{S}_n \xrightarrow[n\to\infty]{\text{a.s.}} E(X_1).$$

Let $F \subset \mathbb{R}$ be a closed set such that $E(X_1) \notin F$. Then, there exists $\delta > 0$ such that

$$\inf_{x\in F} |x - E(X_1)| = \delta > 0.$$

Since almost sure convergence implies convergence in probability, we have

$$\lim_{n\to\infty} \mu_n(F) = \lim_{n\to\infty} \mathbb{P}\left(\widehat{S}_n \in F\right) \leq \lim_{n\to\infty} \mathbb{P}\left(|\widehat{S}_n - E(X_1)| \geq \delta\right) = 0 \Rightarrow \lim_{n\to\infty} \mu_n(F) = 0.$$

Our objective is to study how fast is this convergence. Cramér's theorem characterizes the exponential velocity of this convergence without assuming that $X_1 \in L^1(\Omega)$.

Since the version in $\mathbb{R}^d$ requires some partial results proved in this case and the ideas are quite similar, we restrict ourselves to the case in $\mathbb{R}$. In fact, there are stronger and more general versions of Crámer's Theorem which are proved using sub-additivity arguments.

2.2 The moment generating function and its Fenchel-Legendre transform

We introduce and study the properties of the logarithmic moment generating function of a random variable and its Fenchel-Legendre transform, which will be constantly used during this chapter. For this section, $X : \Omega \to \mathbb{R}$ will denote a random variable.

Definition 2.2.1. The moment generating function of X is the function

$$M(\lambda) := E\left(e^{\lambda X}\right), \qquad \lambda \in \mathbb{R}$$

and the logarithmic moment generating function is

$$\Lambda(\lambda) := \log M(\lambda) = \log E\left(e^{\lambda X}\right), \qquad \lambda \in \mathbb{R}.$$

Note that the expectation $E\left(e^{\lambda X}\right)$ is well defined (possibly $E\left(e^{\lambda X}\right) = \infty$) because the random variable $e^{\lambda X}$ is strictly positive. So, $\Lambda(\lambda) \in (-\infty, \infty]$ for all $\lambda \in \mathbb{R}$ and $\Lambda(0) = 0$.

Definition 2.2.2. The Fenchel-Legendre transform of Λ is defined as

$$\Lambda^*(x) := \sup_{\lambda\in\mathbb{R}} \{\lambda x - \Lambda(\lambda)\}, \qquad \lambda \in \mathbb{R}.$$

We denote $\mathcal{D}_\Lambda := \{\lambda \in \mathbb{R} : \ \Lambda(\lambda) < \infty\}$ and $\mathcal{D}_{\Lambda^*} := \{\lambda \in \mathbb{R} : \ \Lambda^*(\lambda) < \infty\}$.

The following Lemma gathers some fundamental properties of the logarithmic moment generating function and its Fenchel-Legendre transform.

Lemma 2.2.3.

1. Λ is a convex function.
2. Λ^* is a convex rate function.
3. If there exist $\lambda_- < 0$ and $\lambda_+ > 0$ with $\lambda_-, \lambda_+ \in \mathcal{D}_\Lambda$, then Λ^* is a good rate function.
4. $\{0\} \subset \mathcal{D}_\Lambda$. If $\mathcal{D}_\Lambda = \{0\}$, then Λ^* is identically zero.
5. If there exists $\lambda_+ > 0$ with $\lambda_+ \in \mathcal{D}_\Lambda$, then $E(X) < \infty$ (possibly $E(X) = -\infty$), and for all $x \geq E(X)$,
$$\Lambda^*(x) = \sup_{\lambda\geq 0} \{\lambda x - \Lambda(\lambda)\} \tag{2.2.1}$$
is, for $x > E(X)$, a nondecreasing function.
6. If there exist $\lambda_- < 0$ with $\lambda_- \in \mathcal{D}_\Lambda$, then $E(X) > -\infty$ (possibly $E(X) = \infty$), and for all $x \leq E(X)$,
$$\Lambda^*(x) = \sup_{\lambda\leq 0} \{\lambda x - \Lambda(\lambda)\} \tag{2.2.2}$$
is, for $x < E(X)$, a nonincreasing function.
7. If $X \in L^1(\Omega)$, $\Lambda^*(E(X)) = 0$.
8. $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$.
9. Suppose that $\mathcal{D}_\Lambda^o$ is nonempty. For each $\lambda_0 \in \mathcal{D}_\Lambda^o$, Λ is differentiable in λ_0, $Xe^{\lambda_0 X} \in L^1(\Omega)$,
$$\Lambda'(\lambda_0) = \frac{1}{M(\lambda_0)} E(Xe^{\lambda_0 X})$$
and
$$\Lambda'(\lambda_0) = y \Leftrightarrow \Lambda^*(y) = \lambda_0 y - \Lambda(\lambda_0).$$

Proof.
(1) Let $\lambda_1, \lambda_2 \in \mathbb{R}$ and $\theta \in [0,1]$. Using Hölder's inequality with $p = \frac{1}{\theta} \in [1, \infty)$ and $q = \frac{1}{1-\theta} \in [1, \infty)$ we have

$$\Lambda(\theta\lambda_1 + (1-\theta)\lambda_2) = \log E\left(e^{(\theta\lambda_1+(1-\theta)\lambda_2)X}\right) = \log E\left(\left(e^{\theta\lambda_1 X}\right)\left(e^{(1-\theta)\lambda_2 X}\right)\right)$$
$$\leq \log \left(E\left(e^{\lambda_1 X}\right)\right)^{\theta}\left(E\left(e^{\lambda_2 X}\right)\right)^{1-\theta} = \theta\Lambda(\lambda_1) + (1-\theta)\Lambda(\lambda_2).$$

Hence, Λ is a convex function.

(2) Let $x_1, x_2 \in \mathbb{R}$ and $\theta \in [0,1]$. Then

$$\theta\Lambda^*(x_1) + (1-\theta)\Lambda^*(x_2) = \sup_{\lambda\in\mathbb{R}}\{\theta\lambda x_1 - \theta\Lambda(\lambda)\} + \sup_{\lambda\in\mathbb{R}}\{(1-\theta)\lambda x_2 - (1-\theta)\Lambda(\lambda)\}$$
$$\geq \sup_{\lambda\in\mathbb{R}}\{(\theta x_1 + (1-\theta)x_2)\,\lambda - \Lambda(\lambda)\} = \Lambda^*\left(\theta x_1 + (1-\theta)x_2\right).$$

Therefore, Λ^* is a convex function. Moreover, Λ^* is a nonnegative function because

$$\Lambda(0) = 0 \Rightarrow \Lambda^*(x) \geq 0 \cdot x - \Lambda(0) = 0 \Rightarrow \Lambda^*(\lambda) \in [0,\infty], \qquad \forall\lambda \in \mathbb{R}.$$

Now we check that Λ^* is lower semicontinuous. Let $x \in \mathbb{R}$ and $\{x_n\} \subset \mathbb{R}$ a sequence such that $\lim_{n\to\infty} x_n = x$. Then, for every $\lambda \in \mathbb{R}$,

$$\liminf_{n\to\infty} \Lambda^*(x) \geq \liminf_{x\to x_0} [\lambda x - \Lambda(\lambda)] = \lambda x - \Lambda(\lambda).$$

Finally,

$$\liminf_{n\to\infty} \Lambda^*(x) \geq \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\} = \Lambda^*(x).$$

We conclude that Λ^* is a convex rate function.

(3) Since for any $\lambda \in \mathbb{R}$,

$$\frac{\Lambda^*(x)}{|x|} \geq \lambda \ \mathrm{sign}(x) - \frac{\Lambda(\lambda)}{|x|},$$

it implies that

$$\liminf_{|x|\to\infty} \frac{\Lambda^*(x)}{|x|} \geq \min\{\lambda_+, -\lambda_-\} > 0 \Rightarrow \lim_{|x|\to\infty} \Lambda^*(x) = \infty.$$

Since Λ^* is a rate function, its level sets are closed. Moreover, by the last limit the level sets are also bounded. So, they are compact and Λ^* is a good rate function.

(4) Since $\Lambda(0) = 0$, we always have $\{0\} \subset \mathcal{D}_\Lambda$. Suppose that $\mathcal{D}_\Lambda = \{0\}$. Then, for all $x \in \mathbb{R}$,

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\} = 0 \cdot x - \Lambda(0) = 0.$$

(5) Suppose that $\Lambda(\lambda_+) < \infty$ for some $\lambda_+ > 0$. Write μ for the law of X and μ^+ for the law of $X^+ := \max(0, X)$. Using that $\mu = \mu^+$ in $\mathcal{B}((0,\infty))$ and that $\lambda_+ x < e^{\lambda_+ x}$ for all $x \in \mathbb{R}$ we have

$$\begin{aligned} E(X^+) &= \int_{\mathbb{R}} x\mu^+(\mathrm{d}x) = \frac{1}{\lambda_+}\int_{(0,\infty)} \lambda_+ x\mu^+(\mathrm{d}x) = \frac{1}{\lambda_+}\int_{(0,\infty)} \lambda_+ x\mu(\mathrm{d}x) \\ &\le \frac{1}{\lambda_+}\int_{(0,\infty)} e^{\lambda_+ x}\mu(\mathrm{d}x) \le \frac{1}{\lambda_+}\int_{\mathbb{R}} e^{\lambda_+ x}\mu(\mathrm{d}x) = \frac{E(e^{\lambda_+ X})}{\lambda_+} = \frac{M(\lambda_+)}{\lambda_+} < \infty. \end{aligned}$$

Hence, $E(X) < \infty$ (possibly $E(X) = -\infty$).

Now, for all $\lambda \in \mathbb{R}$, using Jensen's inequality

$$\Lambda(\lambda) = \log E(e^{\lambda X}) \ge E(\log e^{\lambda X}) = \lambda E(X).$$

If $E(X) = -\infty$, then $\Lambda(\lambda) = \infty$ for all $\lambda < 0$. Hence, $\Lambda^*(x) = \sup_{\lambda \ge 0} \{\lambda x - \Lambda(\lambda)\}$.

If $E(X) > -\infty$, since $\lambda E(X) - \Lambda(\lambda) \le 0$ for all $\lambda \in \mathbb{R}$ and Λ^* is nonnegative we deduce that

$$\Lambda^*(E(X)) = \sup_{\lambda \in \mathbb{R}} \{\lambda E(X) - \Lambda(\lambda)\} = 0.$$

Moreover, for every $x \ge E(X)$ and $\lambda < 0$

$$\lambda x - \Lambda(\lambda) \le \lambda E(X) - \Lambda(\lambda) \le \Lambda^*(E(X)) = 0,$$

this implies that $\Lambda^*(x) = \sup_{\lambda \ge 0} \{\lambda x - \Lambda(\lambda)\}$.

Finally, Λ^* is nondecreasing in $(E(X), \infty)$ because for every $\lambda \ge 0$, $\lambda x - \Lambda(\lambda)$ is nondecreasing as a function of x.

(6) Suppose that $\Lambda(\lambda_-) < \infty$ for some $\lambda_- < 0$. Consider the logarithmic moment generating function of $-X$, say Λ_{-X}. Note that $\Lambda_{-X}(-\lambda_-) < \infty$ and we can apply the previous part.

(7) The proof is the same as in part 5.

(8) This is already proved when $\mathcal{D}_\Lambda = \{0\}$ and when $X \in L^1(\Omega)$ because then $\Lambda^*(E(X)) = 0$. Hence, we only have the consider the two following cases.

(8.a) If $\Lambda(\lambda_+) < \infty$ for some $\lambda_+ > 0$ and $E(X) = -\infty$, using Chebycheff's inequality we have for all $\lambda \geq 0$

$$\log \mathbb{P}(X \geq x) \leq \log \frac{E\left(e^{\lambda X}\right)}{e^{\lambda x}} = -\lambda x + \Lambda(\lambda).$$

Then, taking infimum in λ we deduce using (2.2.1)

$$\log \mathbb{P}(X \geq x) \leq \inf_{\lambda \geq 0} \{-\lambda x + \Lambda(\lambda)\} = -\sup_{\lambda \geq 0} \{\lambda x - \Lambda(\lambda)\} = -\Lambda^*(x).$$

By continuity of $\mathbb{P}$ and using that Λ^* is a nondecreasing function in $\mathbb{R}$,

$$\inf_{x \in \mathbb{R}} \Lambda^*(x) = \lim_{x \to -\infty} \Lambda^*(x) \leq \lim_{x \to -\infty} -\log \mathbb{P}(X \geq x) = 0.$$

(8.b) Suppose that $\Lambda(\lambda_-) < \infty$ for some $\lambda_- < 0$ and $E(X) = \infty$. Then $-X$ satisfy the conditions of the previous case, then using (2.2.2)

$$\begin{aligned} 0 = \inf_{x \in \mathbb{R}} \Lambda^*_{-X}(x) &= \inf_{x \in \mathbb{R}} \sup_{\lambda \geq 0} \{\lambda x - \Lambda_{-X}(\lambda)\} = \inf_{x \in \mathbb{R}} \sup_{\lambda \leq 0} \{-\lambda x - \Lambda_{-X}(-\lambda)\} \\ &= \inf_{x \in \mathbb{R}} \sup_{\lambda \leq 0} \{-\lambda x - \Lambda(\lambda)\} = \inf_{x \in \mathbb{R}} \Lambda^*(-x) = \inf_{x \in \mathbb{R}} \Lambda^*(x). \end{aligned}$$

(9) Let $\lambda_0 \in \mathcal{D}^o_\Lambda$, $\varepsilon > 0$ such that $[\lambda_0 - 2\varepsilon, \lambda_0 + 2\varepsilon] \subset \mathcal{D}_\Lambda$ and $\{\lambda_n\} \subset \mathcal{D}^o_\Lambda$ such that $\lambda_n \xrightarrow[n \to \infty]{} \lambda_0$ and $0 < |\lambda_n - \lambda_0| \leq \varepsilon$ for all $n \geq 1$. For $n \geq 1$,

$$\frac{M(\lambda_n) - M(\lambda_0)}{\lambda_n - \lambda_0} = \frac{E(e^{\lambda_n X}) - E(e^{\lambda_0 X})}{\lambda_n - \lambda_0} = E(Y_n),$$

where

$$Y_n := \frac{e^{\lambda_n X} - e^{\lambda_0 X}}{\lambda_n - \lambda_0}.$$

Fix $\omega \in \Omega$. Note that

$$\lim_{n \to \infty} Y_n(\omega) = X(\omega) e^{\lambda_0 X(\omega)}.$$

If there exists a random variable $Z \in L^1(\Omega)$ such that $|Y_n| \leq Z$ for all $n \geq 1$, we can conclude by the dominated convergence theorem that $X e^{\lambda_0 X} \in L^1(\Omega)$ and that

$$M'(\lambda_0) = \lim_{n \to \infty} \frac{M(\lambda_n) - M(\lambda_0)}{\lambda_n - \lambda_0} = \lim_{n \to \infty} E(Y_n) = E\left(\lim_{n \to \infty} Y_n\right) = E\left(X e^{\lambda_0 X}\right).$$

This implies that

$$\Lambda'(\lambda_0) = \frac{1}{M(\lambda_0)} E\left(X e^{\lambda_0 X}\right).$$

We now check that such random variable Z exists. By the mean value theorem we have

$$e^{\lambda_n X} - e^{\lambda_0 X} = X e^{\widetilde{\lambda}_n} (\lambda_n - \lambda_0),$$

where $\widetilde{\lambda}_n$ is between λ_n and λ_0. Note that $\widetilde{\lambda}_n$ also depends on ω but we do not write it for convenience. Then, since $\lambda_n \in [\lambda_0 - \varepsilon, \lambda_0 + \varepsilon] \setminus \{\lambda_0\}$,

$$|Y_n| = \left| \frac{e^{\lambda_n X} - e^{\lambda_0 X}}{\lambda_n - \lambda_0} \right| = |X| \, e^{\widetilde{\lambda}_n} \leq |X| \left(e^{(\lambda_0 - \varepsilon)X} + e^{(\lambda_0 + \varepsilon)X} \right) \tag{2.2.3}$$

$$|X| = \frac{1}{\varepsilon} \varepsilon \, |X| \leq \frac{1}{\varepsilon} e^{\varepsilon |X|} \leq \frac{1}{\varepsilon} \left(e^{-\varepsilon X} + e^{\varepsilon X} \right) \tag{2.2.4}$$

Using (2.2.3) and (2.2.4), we obtain

$$\begin{aligned} |Y_n| &\leq \frac{1}{\varepsilon} \left(e^{-\varepsilon X} + e^{\varepsilon X} \right) \left(e^{(\lambda_0 - \varepsilon)X} + e^{(\lambda_0 + \varepsilon)X} \right) \\ &= \frac{1}{\varepsilon} \left(e^{(\lambda_0 - 2\varepsilon)X} + 2 e^{\lambda_0 X} + e^{(\lambda_0 + 2\varepsilon)X} \right) =: Z. \end{aligned}$$

Since $\lambda_0 - 2\varepsilon, \lambda_0, \lambda_0 + 2\varepsilon \in \mathcal{D}_\Lambda$, we conclude that

$$E(Z) = \frac{1}{\varepsilon} \left(M(\lambda_0 - 2\varepsilon) + 2M(\lambda_0) + M(\lambda_0 + 2\varepsilon) \right) < \infty.$$

Finally, let $\Lambda'(\lambda_0) = y \in \mathbb{R}$. Note that the function $g_y(\lambda) := \lambda y - \Lambda(\lambda)$ is concave and differentiable at λ_0 with $g_y'(\lambda_0) = 0$. Therefore,

$$\lambda_0 y - \Lambda(\lambda_0) = g_y(\lambda_0) = \sup_{\lambda \in \mathbb{R}} g_y(\lambda) = \sup_{\lambda \in \mathbb{R}} \{\lambda y - \Lambda(\lambda)\} = \Lambda^*(y)$$

So, $\Lambda^*(y) = \lambda_0 y - \Lambda(\lambda_0)$. ■

2.3 Proof of Cramér's Theorem

For the rest of this chapter, $\{X_i\}$ will denote a sequence of independent identically distributed random variables $X_i : \Omega \to \mathbb{R}$ with common law μ, Λ will be the logarithmic moment generating function of X_1 and, for $n \geq 1$, we will write μ_n for the law of the empirical mean $\widehat{S}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Our objective is to study the LDP for the family $\{\mu_n\}$.

The next Lemma is a partial result on the LDP for the family $\{\mu_n\}$, namely, is the LDP lower bound for the open sets $(-\delta, \delta)$ and will be used in the proof of the LDP lower bound in Cramér's Theorem. The essential step in the proof of this Lemma is to make an appropriate change of the measure μ together with an application of the law of large numbers.

Lemma 2.3.1. Let $\delta > 0$. Then,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n \left((-\delta, \delta)\right) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) = -\Lambda^*(0).$$

Proof.
(1) Suppose first that $\mu\left((-\infty, 0)\right) > 0$, $\mu\left((0, \infty)\right) > 0$ and that μ is supported on a bounded subset $A \subset \mathbb{R}$. Observe that

$$\mu\left((0, \infty)\right) > 0 \Rightarrow \int_{(0,\infty)} x\mu(\mathrm{d}x) > 0.$$

Then, if $\lambda > 0$

$$\Lambda(\lambda) = \log\left(\int_{(-\infty,0)} e^{\lambda x}\mu(\mathrm{d}x) + \mu(0) + \int_{(0,\infty)} e^{\lambda x}\mu(\mathrm{d}x)\right) \geq \log \int_{(0,\infty)} e^{\lambda x}\mu(\mathrm{d}x)$$
$$\geq \log\left(\lambda \int_{(0,\infty)} x\mu(\mathrm{d}x)\right) = \log \lambda + \log \int_{(0,\infty)} x\mu(\mathrm{d}x) \xrightarrow[\lambda\to\infty]{} \infty.$$

By a similar argument, $\Lambda(\lambda) \xrightarrow[\lambda\to-\infty]{} \infty$.

Note that $\mathcal{D}_\Lambda = \mathcal{D}_\Lambda^o = \mathbb{R}$ because $\sup\limits_{x\in A} e^{\lambda x} < \infty$ and

$$\Lambda(\lambda) = \log \int_A e^{\lambda x}\mu(\mathrm{d}x) \leq \log\left(\sup_{x\in A} e^{\lambda x}\mu(A)\right) < \infty.$$

By parts 1 and 9 in Lemma 2.2.3, Λ is a convex differentiable function in $\mathbb{R}$. In addition, since $\Lambda(\lambda) \xrightarrow[|\lambda|\to\infty]{} \infty$ there exist $\eta \in \mathbb{R}$ such that

$$\Lambda(\eta) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) \quad \text{and} \quad \Lambda'(\eta) = 0. \tag{2.3.1}$$

Now, define the following measure $\widetilde{\mu}$ by

$$\widetilde{\mu}(\mathrm{d}x) := e^{\eta x - \Lambda(\eta)}\mu(\mathrm{d}x).$$

Note that $\widetilde{\mu}$ is a probability measure because

$$\widetilde{\mu}\left(\mathbb{R}\right) = \int_{\mathbb{R}} \widetilde{\mu}(\mathrm{d}x) = e^{-\Lambda(\eta)} \int_{\mathbb{R}} e^{\eta x}\mu(\mathrm{d}x) = e^{-\Lambda(\eta)}M(\eta) = e^{-\Lambda(\eta)}e^{\Lambda(\eta)} = 1.$$

Consider $\{Y_i\}$ a sequence of independent identically distributed random variables $Y_i : \Omega \to \mathbb{R}$ with common law $\widetilde{\mu}$. For $n \geq 1$, denote by $\widetilde{\mu}_n$ the law of the empirical mean $\widehat{S}_n^Y = \frac{1}{n}\sum_{i=1}^n Y_i$ and for every $\varepsilon > 0$ write

$$B_\varepsilon := \left\{ (x_1, ..., x_n) \in \mathbb{R}^n : \left| \sum_{i=1}^n x_i \right| < n\varepsilon \right\}.$$

Then

$$\begin{aligned}
\mu_n\left((-\varepsilon,\varepsilon)\right) &= \mathbb{P}(|\widehat{S}_n| < \varepsilon) = \mathbb{P}\left(|X_1 + ... + X_n| < n\varepsilon\right) \\
&= \mathbb{P}\left((X_1, ..., X_n) \in B_\varepsilon\right) = \int_{B_\varepsilon} \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n) \\
&\geq e^{-n\varepsilon|\eta|} \int_{B_\varepsilon} \exp\left(\eta \sum_{i=1}^n x_i\right) \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n) \\
&= e^{-n\varepsilon|\eta|} e^{n\Lambda(\eta)} \int_{B_\varepsilon} \prod_{i=1}^n \exp\left(\eta x_i - \Lambda(\eta)\right) \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n) \\
&= e^{-n\varepsilon|\eta|} e^{n\Lambda(\eta)} \int_{B_\varepsilon} \widetilde{\mu}(\mathrm{d}x_1)\cdots\widetilde{\mu}(\mathrm{d}x_n) = e^{-n\varepsilon|\eta|} e^{n\Lambda(\eta)} \widetilde{\mu}_n\left((-\varepsilon,\varepsilon)\right). \qquad (2.3.2)
\end{aligned}$$

Now, by part 9 in Lemma 2.2.3,

$$E\left(Y_1\right) = \int_{\mathbb{R}} x\widetilde{\mu}(\mathrm{d}x) = \int_{\mathbb{R}} x e^{\eta x - \Lambda(\eta)} \mu(\mathrm{d}x) = \frac{1}{M(\eta)} E\left(X_1 e^{\eta X_1}\right) = \Lambda'(\eta) = 0,$$

and by the law of large numbers

$$\lim_{n\to\infty} \widetilde{\mu}_n\left((-\varepsilon,\varepsilon)\right) = \lim_{n\to\infty} \mathbb{P}\left(|\widehat{S}_n^Y| < \epsilon\right) = \lim_{n\to\infty} \left[1 - \mathbb{P}\left(|\widehat{S}_n^Y| \geq \epsilon\right)\right] = 1. \qquad (2.3.3)$$

Finally, by equation (2.3.2) we deduce that for every $0 < \varepsilon < \delta$,

$$\frac{1}{n}\log\mu_n\left((-\delta,\delta)\right) \geq \frac{1}{n}\log\mu_n\left((-\varepsilon,\varepsilon)\right) \geq \Lambda(\eta) - \varepsilon|\eta| + \frac{1}{n}\log\widetilde{\mu}_n\left((-\varepsilon,\varepsilon)\right). \qquad (2.3.4)$$

Taking $\liminf$ as $n \to \infty$ and using (2.3.3) and (2.3.4)

$$\liminf_{n\to\infty} \frac{1}{n}\log\mu_n\left((-\delta,\delta)\right) \geq \liminf_{n\to\infty}\left[\Lambda(\eta) - \varepsilon|\eta| + \frac{1}{n}\log\widetilde{\mu}_n\left((-\varepsilon,\varepsilon)\right)\right] = \Lambda(\eta) - \epsilon|\eta|,$$

and now taking limit as $\epsilon \to 0$ we obtain, by (2.3.1),

$$\liminf_{n\to\infty} \frac{1}{n}\log\mu_n\left((-\delta,\delta)\right) \geq \Lambda(\eta) = \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda).$$

(2) Suppose now that $\mu\left((-\infty,0)\right)>0$, $\mu\left((0,\infty)\right)>0$ but that μ is of unbounded support. Fix $M_0>0$ large enough such that $\mu\left([-M_0,0)\right)>0$ and $\mu\left((0,M_0]\right)>0$. Note that such M_0 exists because otherwise $\mu\left((-\infty,0)\right)=\mu\left((0,\infty)\right)=0$. Define

$$\Lambda_{M_0}(\lambda):=\log\int_{-M_0}^{M_0}e^{\lambda x}\mu(\mathrm{d}x),\qquad \lambda\in\mathbb{R}.$$

Let ν be the law of X_1 conditioned on $\{|X_1|\le M_0\}$. That is, for $B\in\mathcal{B}(\mathbb{R})$,

$$\nu(B)=\mathbb{P}\left(X_1\in B\mid |X_1|\le M_0\right)=\frac{\mu\left(B\cap[-M_0,M_0]\right)}{\mu\left([-M_0,M_0]\right)}\le\frac{\mu(B)}{\mu\left([-M_0,M_0]\right)}.$$

Let $\{Z_i\}$ be a sequence of independent identically distributed random variables $Z_i:\Omega\to\mathbb{R}$ with common law ν and for $n\ge 1$, denote by ν_n the law of the empirical mean $\widehat{S}_n^Z=\frac{1}{n}\sum_{i=1}^n Z_i$. The key point is that ν satisfies the hypothesis of the first part of the proof and we can use the results obtained in that part.

Observe that for all $n\ge 1$ and every $\delta>0$

$$\begin{aligned}\nu_n\left((-\delta,\delta)\right)&=\int_{B_\delta}\nu(\mathrm{d}y_1)\cdots\nu(\mathrm{d}y_n)\le\frac{1}{\mu\left([-M_0,M_0]\right)^n}\int_{B_\delta}\mu(\mathrm{d}y_1)\cdots\mu(\mathrm{d}y_n)\\&=\frac{\mu_n\left((-\delta,\delta)\right)}{\mu\left([-M_0,M_0]\right)^n}\Rightarrow\mu_n\left((-\delta,\delta)\right)\ge\nu_n\left((-\delta,\delta)\right)\mu\left([-M_0,M_0]\right)^n.\end{aligned}$$

Note that the logarithmic moment generating function associated with ν is

$$\begin{aligned}\Lambda_\nu(\lambda)&=\log\int_{\mathbb{R}}e^{\lambda x}\nu(\mathrm{d}x)=\log\int_{-M_0}^{M_0}e^{\lambda x}\nu(\mathrm{d}x)=\log\left(\frac{1}{\mu\left([-M_0,M_0]\right)}\int_{-M_0}^{M_0}e^{\lambda x}\mu(\mathrm{d}x)\right)\\&=\log\int_{-M_0}^{M_0}e^{\lambda x}\mu(\mathrm{d}x)-\log\mu\left([-M_0,M_0]\right)=\Lambda_{M_0}(\lambda)-\log\mu\left([-M_0,M_0]\right).\end{aligned}$$

Then, using that ν satisfy the hypothesis of the previous part we have

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log\mu_n\left((-\delta,\delta)\right)&\ge\log\mu\left([-M_0,M_0]\right)+\liminf_{n\to\infty}\frac{1}{n}\log\nu_n\left((-\delta,\delta)\right)\\&\ge\log\mu\left([-M_0,M_0]\right)+\inf_{\lambda\in\mathbb{R}}\Lambda_\nu(\lambda)=\inf_{\lambda\in\mathbb{R}}\Lambda_{M_0}(\lambda).\end{aligned}\tag{2.3.5}$$

Note that the previous argument is true for all $M\ge M_0$. So, equation (2.3.5) also holds for $M\ge M_0$. Write $I_M:=\inf_{\lambda\in\mathbb{R}}\Lambda_M(\lambda)$ for $M\ge M_0$ and $I^*:=\liminf_{M\to\infty}I_M$. Then, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n\left((-\delta,\delta)\right)\ge\liminf_{M\to\infty}\inf_{\lambda\in\mathbb{R}}\Lambda_M(\lambda)=\liminf_{M\to\infty}I_M=I^*.\tag{2.3.6}$$

Since $\Lambda_M(\lambda) = \Lambda_\nu(\lambda) + \log\mu([-M,M])$ we have

$$I_M = \inf_{\lambda\in\mathbb{R}} \Lambda_M(\lambda) = \inf_{\lambda\in\mathbb{R}} \Lambda_\nu(\lambda) + \log\mu([-M,M]) = \Lambda_\nu(\eta_\nu) + \log\mu([-M,M]),$$

where $\eta_\nu \in \mathbb{R}$ is such that $\Lambda_\nu'(\eta_\nu) = 0$. Then,

$$I^* = \liminf_{M\to\infty} I_M = \liminf_{M\to\infty} [\Lambda_\nu(\eta_\nu) + \log\mu([-M,M])] = \Lambda_\nu(\eta_\nu) \in \mathbb{R},$$

because $\mathcal{D}_{\Lambda_\nu} = \mathbb{R}$. Consider the following level sets for $M \geq M_0$

$$C_M := \psi_{\Lambda_M}(I^*) = \{\lambda \in \mathbb{R} : \ \Lambda_M(\lambda) \leq I^*\}$$

Observe the following:

(b.1) C_M are non-empty because

$$\Lambda_M(\lambda) \leq I^* \Leftrightarrow \Lambda_\nu(\lambda) \leq \Lambda_\nu(\eta_\nu) - \log\mu([-M,M]),$$

and such $\lambda \in \mathbb{R}$ exists because $\Lambda_\nu(\eta_\nu) = \inf_{\lambda\in\mathbb{R}} \Lambda_\nu(\lambda)$ and $\log\mu([-M,M]) < 0$.

(b.2) C_M are compact. They are closed because Λ_M is a continuous function since Λ_ν is a continuous function. They are bounded because

$$\lim_{|\lambda|\to\infty} \Lambda_\nu(\lambda) = \infty \Rightarrow \lim_{|\lambda|\to\infty} \Lambda_M(\lambda) = \infty.$$

(b.3) If $M_0 \leq M_1 \leq M_2$ then $\Lambda_{M_1}(\lambda) \leq \Lambda_{M_2}(\lambda)$ for all $\lambda \in \mathbb{R}$. Therefore, $C_{M_2} \subset C_{M_1}$.

Then, by Cantor's intersection theorem there exists $\lambda_0 \in \bigcap_{M\geq M_0} C_M$. Moreover, using Lebesgue's monotone convergence theorem, we have

$$\begin{aligned} I^* &\geq \lim_{M\to\infty} \Lambda_M(\lambda_0) = \log \lim_{M\to\infty} \int_{\mathbb{R}} e^{\lambda_0 x} \mathbb{1}_{[-M,M]}(x)\mu(\mathrm{d}x) \\ &= \log \int_{\mathbb{R}} \lim_{M\to\infty} e^{\lambda_0 x} \mathbb{1}_{[-M,M]}(x)\mu(\mathrm{d}x) = \log \int_{\mathbb{R}} e^{\lambda_0 x}\mu(\mathrm{d}x) = \Lambda(\lambda_0) \end{aligned}$$

Finally, by equation (2.3.6)

$$\liminf_{n\to\infty} \frac{1}{n} \log\mu_n((-\delta,\delta)) \geq I^* \geq \Lambda(\lambda_0) \geq \inf_{\lambda\in\mathbb{R}} \Lambda(\lambda).$$

(c) Suppose now that $\mu((-\infty,0)) = 0$. Then, if $\lambda_1 \leq \lambda_2$ we have

$$\Lambda(\lambda_1) = \int_0^\infty e^{\lambda_1 x}\mu(\mathrm{d}x) \leq \int_0^\infty e^{\lambda_2 x}\mu(\mathrm{d}x) = \Lambda(\lambda_2).$$

Then, using the Lebesgue's monotone convergence theorem for a decreasing sequences of functions we obtain

$$\begin{aligned}\inf_{\lambda\in\mathbb{R}} \Lambda(\lambda) &= \lim_{\lambda\to-\infty} \log \int_{[0,\infty)} e^{\lambda x}\mu(\mathrm{d}x) = \log\mu(\{0\}) + \lim_{\lambda\to-\infty}\int_{(0,\infty)} e^{\lambda x}\mu(\mathrm{d}x) \\ &= \log\mu(\{0\}) + \int_{(0,\infty)} \lim_{\lambda\to-\infty} e^{\lambda x}\mu(\mathrm{d}x) = \log\mu(\{0\}).\end{aligned} \tag{2.3.7}$$

Then,

$$\begin{aligned}\mu_n((-\delta,\delta)) &\geq \mu_n(\{0\}) = \mathbb{P}(X_1 + ... + X_n = 0) \\ &\geq \mathbb{P}(X_1 = 0, ..., X_n = 0) = \prod_{i=1}^{n} \mathbb{P}(X_i = 0) = \mu(\{0\})^n.\end{aligned} \tag{2.3.8}$$

Finally, by (2.3.7) and (2.3.8)

$$\liminf_{n\to\infty} \frac{1}{n}\log\mu_n((-\delta,\delta)) \geq \log\mu(\{0\}) = \inf_{\lambda\in\mathbb{R}}\Lambda(\lambda).$$

(d) Suppose that $\mu((0,\infty)) = 0$. The argument is similar to the previous case. ■

We finally have all the tools to prove Cramér's Theorem. Note that we do not need the random variables $\{X_i\}$ to be integrable.

The strategy to prove the upper bound of the LDP is to use the independence of the random variables together with the basic inequality $\mathbb{1}_{\{f(x)\geq 0\}}(x) \leq e^{f(x)}$ for any function $f:\mathbb{R}\to\mathbb{R}$. Then, the upper bound follows considering carefully different cases and using the properties of the Fenchel-Legendre transform of the logarithmic moment generating function of X_1.

On the other hand, in order to prove the upper bound of the LDP we are going to use Lemma 2.3.1 after making a suitable linear transformation of the random variables X_i.

Theorem 2.3.2. Crámer. *The family $\{\mu_n\}$ satisfies a LDP with the convex rate function Λ^*, namely:*

1. *For any closed set $F\subset\mathbb{R}$,*

$$\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(F) \leq -\inf_{x\in F}\Lambda^*(x). \tag{2.3.9}$$

2. *For any open set $G\subset\mathbb{R}$,*

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(G) \geq -\inf_{x\in G}\Lambda^*(x). \tag{2.3.10}$$

Proof.
(1) Let $F \subset \mathbb{R}$ be a non-empty closed set and $I_F := \inf_{x\in F} \Lambda^*(x)$. If $I_F = 0$, then (2.3.9) holds.

Assume that $I_F > 0$. Then, Λ^* is not identically zero and by part 4 of Lemma 2.2.3, we deduce that there exists $\lambda_0 \neq 0$ such that $\Lambda(\lambda_0) < \infty$. By part 5 and 6 of Lemma 2.2.3, we obtain that $E(X_1)$ exists, possibly as an extended real number.

Let $x \in \mathbb{R}$ and $\lambda \geq 0$. Since $\mathbb{1}_{\{\widehat{S}_n - x \geq 0\}}(x) \leq e^{n\lambda(\widehat{S}_n - x)}$ we have

$$\begin{aligned}\mu_n([x,\infty)) &= \mathbb{P}\left(\widehat{S}_n \geq x\right) = E\left(\mathbb{1}_{\{\widehat{S}_n - x\geq 0\}}\right) \leq E\left(e^{n\lambda(\widehat{S}_n - x)}\right) = e^{-n\lambda x}\prod_{i=1}^{n} E\left(e^{\lambda X_i}\right)\\ &= e^{-n\lambda x} E\left(e^{\lambda X_1}\right)^n = e^{-n\lambda x} M(\lambda)^n = e^{-n\lambda x} e^{n\Lambda(\lambda)} = e^{-n(\lambda x - \Lambda(\lambda))}\end{aligned}$$

Then, if $\lambda_0 > 0$, by part 5 of Lemma 2.2.3 we know that $E(X) < \infty$ and that for $x \geq E(X)$, $\Lambda^*(x) = \sup_{\lambda\geq 0}\{\lambda x - \Lambda(\lambda)\}$. Then,

$$\mu_n([x,\infty)) \leq \inf_{\lambda\geq 0} \exp\left[-n\left(\lambda x - \Lambda(\lambda)\right)\right] = \exp[-n\sup_{\lambda\geq 0}\{\lambda x - \Lambda(\lambda)\}] = e^{-n\Lambda^*(x)} \tag{2.3.11}$$

By a similar argument, if $\lambda_0 < 0$, then $E(x) > -\infty$ and for $x \leq E(X)$ we have

$$\mu_n((-\infty, x]) \leq e^{-n\Lambda^*(x)} \tag{2.3.12}$$

(1.a) Consider first that $X_1 \in L^1(\Omega)$. By part 7 of Lemma 2.2.3, $\Lambda^*(E(X_1)) = 0$. Since $I_F > 0$, $E(X_1) \in F^c$. Consider

$$(x_-, x_+) := \bigcup_{\substack{a<E(X_1)<b\\(a,b)\subset F^c}} (a,b).$$

Note that the right-hand side is in fact an interval because $E(X_1)$ belongs in every interval (a,b) and that either x_- or x_+ is finite because F is non-empty.

If $x_- > -\infty$, then $x_- \in F$. Otherwise, if $x_- \in F^c$, we could enlarge the interval (x_-, x_+) because F^c is open. Hence, $\Lambda^*(x_-) \geq I_F$. Similarly, if $x_+ < \infty$, then $x_+ \in F$ and $\Lambda^*(x_+) \geq I_F$.

(1.a.i) Suppose that both x_- and x_+ are finite. Applying (2.3.11) for $x = x_+ \geq E(X_1)$, (2.3.12) for $x = x_- \leq E(X_1)$ and using that $F \cap (x_-, x_+) = \varnothing$ we have

$$\mu_n(F) \leq \mu_n((-\infty, x_-]) + \mu_n([x_+, +\infty)) \leq e^{-n\Lambda^*(x_-)} + e^{-n\Lambda^*(x_+)} \leq 2e^{-nI_F}$$

Then, by the previous inequality

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq \limsup_{n\to\infty} \frac{1}{n} (\log 2 - nI_F) = -I_F = -\inf_{x\in F} \Lambda^*(x).$$

(1.a.ii) The case when $x_- = -\infty$ or $x_+ = \infty$ is similar to the previous one.

(1.b) Suppose now that $E(X_1) = -\infty$. Then, we are in part 5 of Lemma 2.2.3. Since Λ^* is a nondecreasing function in all $\mathbb{R}$ and $\inf_{x\in\mathbb{R}} \Lambda^*(x) = 0$ we deduce that

$$\lim_{x\to-\infty} \Lambda^*(x) = 0.$$

Note that $x_+ := \inf F > -\infty$ because otherwise $I_F = 0$. Moreover, $x_+ \in F$ because F is closed and then $\Lambda^*(x_+) \geq I_F$. Applying (2.3.11) for $x = x_+ \geq E(X_1)$ and using that $F \cap (-\infty, x_+) = \varnothing$ we obtain

$$\mu_n(F) \leq \mu_n([x_+, \infty)) \leq e^{-n\Lambda^*(x_+)} \leq e^{-nI_F} \Rightarrow \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x).$$

(1.c) The case when $E(X_1) = \infty$ is solved analogously.

(2) Let $G \subset \mathbb{R}$ be a non-empty open set and $x \in G$. There exists $\delta > 0$ such that $(x - \delta, x + \delta) \subset G$.

Define the sequence of independent identically distributed random variables $\{Y_i\}$ by $Y_i := X_i - x$. If ν_n is the law of the empirical mean $\widehat{S}_n^Y = \frac{1}{n}\sum_{i=1}^n Y_i$, we have for $B \in \mathcal{B}(\mathbb{R})$,

$$\nu_n(B) = \mathbb{P}\left(\widehat{S}_n^Y \in B\right) = \mathbb{P}\left(\frac{(X_1 - x) + ... + (X_n - x)}{n} \in B\right)$$
$$= \mathbb{P}\left(\widehat{S}_n \in B + x\right) = \mu_n(B + x).$$

Moreover, note that

$$\Lambda_Y(\lambda) = \log E\left(e^{\lambda Y_1}\right) = \log E\left(e^{\lambda X_1} e^{-\lambda x}\right) = \Lambda(\lambda) - \lambda x,$$

and

$$\Lambda_Y^*(y) = \sup_{\lambda\in\mathbb{R}} \{\lambda y - \Lambda_Y(\lambda)\} = \sup_{\lambda\in\mathbb{R}} \{\lambda(y + x) - \Lambda(\lambda)\} = \Lambda^*(y + x).$$

Applying Lemma 2.3.1 to the sequence $\{Y_i\}$ we have

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

$$= \liminf_{n\to\infty} \frac{1}{n} \log \nu_n\left((-\delta,\delta)\right) \geq -\Lambda_Y^*(0) = -\Lambda^*(x).$$

Since the last inequality is true for all $x \in G$, we finally have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq \sup_{x\in G} -\Lambda^*(x) = -\inf_{x\in G} \Lambda^*(x).$$

■

We finish this chapter giving two examples of Cramér's Theorem.

Corollary 2.3.3. Suppose that $X_1 \overset{\mathrm{d}}{=} \text{Bernoulli}(p)$. Then, the family $\{\mu_n\}$ satisfies a LDP with the good rate function

$$\Lambda^*(x) = \begin{cases} x\log(\frac{x}{p}) + (1-x)\log(\frac{1-x}{1-p}) & \text{if } x \in [0,1] \\ \infty & \text{otherwise.} \end{cases}$$

Proof. Applying Crámer's Theorem (Theorem 2.3.2) we know that $\{\mu_n\}$ satisfies a LDP with the rate function Λ^*, where

$$\begin{aligned} \Lambda(\lambda) &= \log E(e^{\lambda X_1}) = \log(pe^\lambda + 1 - p). \\ \Lambda^*(x) &= \sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \log(pe^\lambda + 1 - p)\right\} \\ &= \begin{cases} x\log(\frac{x}{p}) + (1-x)\log(\frac{1-x}{1-p}) & \text{if } x \in [0,1] \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Since $\mathcal{D}_\Lambda = \mathbb{R}$, by part 3 in Lemma 2.2.3, Λ^* is a good rate function. Moreover, note that $\Lambda^*(p) = 0$ and $\Lambda^*(x) > 0$ for $x \neq p$. In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\mu_n(\Gamma)\}$ for $\Gamma \in \mathcal{B}(\mathbb{R})$ whenever $p \notin \overline{\Gamma}$. ■

Corollary 2.3.4. Suppose that $X_1 \overset{\mathrm{d}}{=} N(0,\sigma^2)$. Then, the family $\{\mu_n\}$ satisfies a LDP with the good rate function

$$\Lambda^*(x) = \frac{x^2}{2\sigma^2}.$$

Proof. Applying Crámer's Theorem (Theorem 2.3.2) we know that $\{\mu_n\}$ satisfies a LDP with the rate function Λ^*, where

$$\Lambda(\lambda) = \log E(e^{\lambda X_1}) = \log e^{\frac{\sigma^2\lambda^2}{2}} = \frac{\sigma^2\lambda^2}{2}.$$

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \frac{\lambda^2 \sigma^2}{2} \right\} = \frac{x^2}{2\sigma^2}.$$

Since $\mathcal{D}_\Lambda = \mathbb{R}$, by part 3 in Lemma 2.2.3, Λ^* is a good rate function. Moreover, note that $\Lambda^*(0) = 0$ and $\Lambda^*(x) > 0$ for $x \neq 0$. In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\mu_n(\Gamma)\}$ for $\Gamma \in \mathcal{B}(\mathbb{R})$ whenever $0 \notin \overline{\Gamma}$.

In particular, when $\sigma^2 = 1$ we recover Equation (1.1.4) of the Example in Chapter 1. Fix $\delta > 0$ and choose $\Gamma = (-\infty, -\delta] \cup [\delta, \infty) \in \mathcal{B}(\mathbb{R})$. Then,

$$-\frac{\delta^2}{2} = -\inf_{x \in \Gamma^\circ} \Lambda^*(x) \leq \liminf_{n \to \infty} \frac{1}{n} \mu_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \mu_n(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} \Lambda^*(x) = -\frac{\delta^2}{2}.$$

So,

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) = \lim_{n \to \infty} \frac{1}{n} \log \mu_n((-\infty, -\delta] \cup [\delta, \infty)) = -\frac{\delta^2}{2}.$$

■

3 General Principles

The objective of this chapter is to study general results about the large deviation principle for families of probability measures on arbitrary topological spaces: their existence and uniqueness, the contraction principle and the concept of exponential approximations.

Since the large deviation principle can be studied in very different settings, it is interesting to consider the problem in an abstract framework and have some useful results that can be applied in the concrete cases. In fact, we are going to prove some techniques that will be used to extend the results of Chapter 4 to the ones in Chapter 5.

3.1 Topological preliminaries

Let $\mathcal{X}$ be any non-empty set with the trivial topology $\{\varnothing, \mathcal{X}\}$. Then, a family of probability measures $\{\mu_\varepsilon\}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfies a LDP with a rate function I if and only if

$$\inf_{x \in \mathcal{X}} I(x) = 0.$$

Note that there are a lot of rate functions with this property. Since we want to avoid such simple cases we are going to consider topological spaces that are Hausdorff. Moreover, in some cases we are going to work with regular spaces.

Definition 3.1.1. A Hausdorff topological space $\mathcal{X}$ is regular if, for any closed set $F \subset \mathcal{X}$ and any point $x \notin F$, there exist disjoint open subsets G_1 and G_2 such that $F \subset G_1$ and $x \in G_2$.

All cases in which the LDP is studied in this project the underlying topological space is, in fact, a metric space. Since every metric space is a regular topological space, such assumption is quite reasonable. We state some properties about regular spaces and rate functions that will be used throughout this chapter.

Lemma 3.1.2. Suppose that $\mathcal{X}$ is a regular topological space. Then

1. For any neighborhood G of $x \in \mathcal{X}$, there exists a neighborhood A of x such that $\overline{A} \subset G$.

2. Let f be a lower semicontinuous function. Then, for any $x \in \mathcal{X}$ and $\delta > 0$ there exists a neighborhood A of x such that

$$\inf_{y \in \overline{A}} f(y) \geq (f(x) - \delta) \wedge \frac{1}{\delta}.$$

Proof. For a reference, see [1], part (a) in page 116 and part(c) in page 117. ■

Lemma 3.1.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be topological spaces and $f : \mathcal{X} \to \mathcal{Y}$. If $\mathcal{X}$ is compact, $\mathcal{Y}$ is Hausdorff and f is a continuous bijection, then f is an homeomorphism between X and Y.

Proof. For a reference, see [2], Theorem 8 in page 141. ■

3.2 The existence of the LDP

The following theorem gives a sufficient condition for the existence of the weak LDP for a family of probability measures on an arbitrary topological space.

Theorem 3.2.1. *Let $\mathcal{X}$ be a topological space, $\{\mu_\varepsilon\}$ a family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $\mathcal{A}$ a base of the topology of $\mathcal{X}$. For every $A \in \mathcal{A}$, define*

$$\mathcal{L}_A := -\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \qquad \text{and} \qquad I(x) := \sup_{\{A\in\mathcal{A}:\ x\in A\}} \mathcal{L}_A. \tag{3.2.1}$$

Then, I is a rate function. If, in addition, for all $x \in \mathcal{X}$,

$$I(x) = \sup_{\{A\in\mathcal{A}:\ x\in A\}} \left[-\limsup_{\varepsilon\to 0} \log \mu_\varepsilon(A) \right], \tag{3.2.2}$$

then, the family $\{\mu_\varepsilon\}$ satisfies a weak LDP with the rate function I.

Proof. Note that I is a nonnegative function. If I is identically equal to 0, it is clear that it is a rate function. So, assume that there exists $\alpha \geq 0$ and $x \in \mathcal{X}$ such that $I(x) > \alpha$. Then, by definition of I there exists $A_x \in \mathcal{A}$ with $x \in A_x$ such that $\mathcal{L}_{A_x} > \alpha$. Moreover, for every $y \in A_x$ we have

$$I(y) = \sup_{\{A\in\mathcal{A}:\ y\in A\}} \mathcal{L}_A \geq \mathcal{L}_{A_x} > \alpha.$$

So, $x \in A_x \subset \{x \in \mathcal{X} : I(x) > \alpha\}$. This proves that the level sets of I are closed, and therefore, I is a rate function.

Let $G \subset \mathcal{X}$ be an open set and $x \in G$. Since $\mathcal{A}$ is a base of the topology of $\mathcal{X}$, there exists $A \in \mathcal{A}$ such that $x \in A \subset G$. Then,

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) = -\mathcal{L}_A \geq -I(x),$$

and taking the supremum over $x \in G$ we get

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq -\inf_{x\in G} I(x).$$

Suppose now that equation (3.2.2) holds and let $F \subset \mathcal{X}$ be a compact set, $x \in F$ and $\delta > 0$. Let I^δ be the δ-rate function, that is,

$$I^\delta(x) := \min\left\{ I(x) - \delta, \frac{1}{\delta} \right\}.$$

By equation (3.2.2) there exists $A_x \in \mathcal{A}$, which also depends on δ, such that $x \in A_x$ and

$$-\limsup_{\varepsilon\to 0} \log \mu_\varepsilon(A_x) \geq I(x) - \delta \geq \min\left\{I(x) - \delta, \frac{1}{\delta}\right\} = I^\delta(x). \tag{3.2.3}$$

Since F is compact we can extract from the open cover $\bigcup\limits_{x\in F} A_x$ of F a finite cover of F by sets $A_{x_1}, ..., A_{x_m}$. So,

$$F \subset \bigcup_{i=1}^m A_{x_i} \Rightarrow \mu_\varepsilon(F) \leq \sum_{i=1}^m \mu_\varepsilon(A_{x_i}).$$

Moreover, applying (1.4.2) of Lemma 1.4.9 and the inequality in (3.2.3) we get

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) &\leq \limsup_{\varepsilon\to 0} \varepsilon \log\left(\sum_{i=1}^m \mu_\varepsilon(A_{x_i})\right) = \max_{1\leq i\leq m} \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon\left(A_{x_i}\right) \\ &\leq \max_{1\leq i\leq m}[-I^\delta(x_i)] = -\min_{i=1}^{m} I^\delta(x_i) \leq -\inf_{x\in F} I^\delta(x).\end{aligned}$$

Finally, taking limit as $\delta \to 0$ and applying Lemma 1.3.5 we obtain

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\inf_{x\in F} I(x).$$

We conclude that the family $\{\mu_\varepsilon\}$ satisfies a weak LDP with the rate function I. ■

Observation 3.2.2. By the definitions in (3.2.1), the condition in (3.2.2) is equivalent to

$$\sup_{\{A\in\mathcal{A}:\ x\in A\}} \left[-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)\right] = \sup_{\{A\in\mathcal{A}:\ x\in A\}} \left[-\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)\right].$$

Hence, if $\lim\limits_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)$ exists for all $A \in \mathcal{A}$ (with $-\infty$ as a possible value), condition (3.2.2) is satisfied.

We already know that condition (3.2.2) implies the existence of the weak LDP. It is interesting to study if the converse is true, that is, if the weak LDP is satisfied, then the rate function is of the form of (3.2.1). The next theorem shows that if the topological space is regular and the full LDP is satisfied, then the converse is true.

Theorem 3.2.3. *Let $\mathcal{X}$ be a regular topological space, $\{\mu_\varepsilon\}$ a family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ that satisfies a LDP with a rate function I. Then, for any base $\mathcal{A}$ of the topology in $\mathcal{X}$, and for any $x \in \mathcal{X}$*

$$I(x) = \sup_{\{A\in\mathcal{A}:\ x\in A\}} \left[-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)\right] = \sup_{\{A\in\mathcal{A}:\ x\in A\}} \left[-\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)\right]. \tag{3.2.4}$$

Proof. Let $x \in \mathcal{X}$ and define

$$J(x) := \sup_{\{A\in\mathcal{A}:\ x\in A\}} \inf_{y\in\overline{A}} I(y).$$

Suppose that $I(x) > J(x)$. In particular $J(x) < \infty$ and $x \in \psi_I(\alpha)^{\mathrm{c}}$ for some $\alpha > J(x)$. Since $\mathcal{A}$ is a base of the topology in $\mathcal{X}$, which is regular, and $\psi_I(\alpha)^{\mathrm{c}}$ is an open set, there exists, by part 1 in Lemma 3.1.2, $A \in \mathcal{A}$ such that $x \in A$ and $\overline{A} \subset \psi_I(\alpha)^{\mathrm{c}}$.

Therefore, $\inf\limits_{y\in\overline{A}} I(y) \geq \alpha$ which implies that $J(x) \geq \alpha$. So, we have obtained a contradiction. We conclude that $J(x) \geq I(x)$.

Since, $\{\mu_\varepsilon\}$ satisfies a LDP with rate function I we have for every $A \in \mathcal{A}$ with $x \in A$:

$$I(x) \geq \inf_{x\in A} I(x) \geq -\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \Rightarrow I(x) \geq \sup_{\{A\in\mathcal{A}:\ x\in A\}} [-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)]. \quad (3.2.5)$$

In addition, for every $A \in \mathcal{A}$ with $x \in A$:

$$-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \geq -\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A) \geq -\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\overline{A}) \geq \inf_{y\in\overline{A}} I(y).$$

Taking supremum over $A \in \mathcal{A}$ with $x \in A$ in the previous inequalities we obtain

$$\sup_{\{A\in\mathcal{A}:\ x\in A\}} [-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)] \geq \sup_{\{A\in\mathcal{A}:\ x\in A\}} [-\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)] \geq J(x). \quad (3.2.6)$$

Finally, since $J(x) \geq I(x)$ we have using (3.2.5) and (3.2.6)

$$J(x) \geq I(x) \geq \sup_{\{A\in\mathcal{A}:\ x\in A\}} [-\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)] \geq \sup_{\{A\in\mathcal{A}:\ x\in A\}} [-\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(A)] \geq J(x).$$

This shows that $I(x)$ satisfies equation (3.2.4). ■

3.3 The uniqueness of the LDP

Another natural question is if a family of probability measures can satisfy a LDP with two different rate functions. In the next proposition we prove the uniqueness of the LDP.

Proposition 3.3.1. *Let $\mathcal{X}$ be a regular topological space and $\{\mu_\varepsilon\}$ a family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then, there exist at most one rate function associated with a possible LDP for $\{\mu_\varepsilon\}$.*

Proof. Suppose that the family $\{\mu_\varepsilon\}$ satisfies a LDP with two different rate functions I_1 and I_2. Without loss of generality, assume that for some $x_0 \in \mathcal{X}$, $I_1(x_0) > I_2(x_0)$.

Let $\delta > 0$. By part 2 in Lemma 3.1.2, there exists a neighborhood A of x_0 such that

$$\inf_{y \in \overline{A}} I_1(y) \geq (I_1(x_0) - \delta) \wedge \frac{1}{\delta}. \tag{3.3.1}$$

In addition, since the family $\{\mu_\varepsilon\}$ satisfies a LDP with the rate functions I_1 and I_2, we have

$$-\inf_{y \in A} I_2(y) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(A) \leq -\inf_{y \in \overline{A}} I_1(y). \tag{3.3.2}$$

Finally, using the inequalities in (3.3.1) and (3.3.2) we obtain

$$I_2(x_0) \geq \inf_{y \in A} I_2(y) \geq \inf_{y \in \overline{A}} I_1(y) \geq (I_1(x_0) - \delta) \wedge \frac{1}{\delta}.$$

Since the last inequality is true for all $\delta > 0$, we deduce that $I_2(x_0) = I_1(x_0)$ (this include the case when $I_1(x_0) = \infty$), which is a contradiction. We conclude that there exists at most one rate function associated with a possible LDP for the family $\{\mu_\varepsilon\}$.

■

3.4 Transformations of LDPs

In this section we study when a transformation preserves the LDP, namely, when the LDP for a family of probability measures $\{\widetilde{\mu}_\varepsilon\}$ can be deduced from the LDP of another family of probability measures $\{\mu_\varepsilon\}$. In particular, we consider the contraction principles and exponential approximations.

3.4.1 Contraction principles

The contraction principle states that the LDP is preserved by continuous maps.

Theorem 3.4.1. ***Contraction principle.*** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces, $f : \mathcal{X} \to \mathcal{Y}$ a continuous function and $I : \mathcal{X} \to [0, \infty]$ a good rate function.*

1. *For each $y \in \mathcal{Y}$, define*

$$J(y) := \inf_{\{x \in \mathcal{X}:\ y = f(x)\}} I(x) = \inf_{x \in f^{-1}(\{y\})} I(x).$$

Then, J is a good rate function on $\mathcal{Y}$.

2. *If a family of probability measures $\{\mu_\varepsilon\}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfies a LDP with the good rate function I, then the family of probability measures $\{\mu_\varepsilon \circ f^{-1}\}$ on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ satisfies a LDP with the good rate function J.*

Proof.
(1) It is clear that J is nonnegative. We are going to prove that for any $\alpha \in [0, \infty)$

$$\psi_J(\alpha) = f(\psi_I(\alpha)).$$

($\subset$) Let $y \in \mathcal{Y}$ such that $J(y) \leq \alpha$. Since Y is Hausdorff and f is continuous, $f^{-1}(\{y\}) \neq \varnothing$ is closed in $\mathcal{X}$. By Lemma 1.3.6, since I is a good rate function the infimum in the definition of J is achieved on some point $x \in \mathcal{X}$. Note that $f(x) = y$ and $I(x) = J(y) \leq \alpha$.

($\supset$) Let $f(x) \in \mathcal{Y}$ with $x \in \mathcal{X}$ and $I(x) \leq \alpha$. Then, $y = f(x) \in \mathcal{Y}$ satisfy $J(y) \leq I(x) \leq \alpha$.

Finally, since $\psi_I(\alpha)$ are compact and f is continuous, $\psi_J(\alpha)$ are also compact.

(2) First, note that for all $A \subset \mathcal{Y}$ we have

$$\inf_{y \in A} J(y) = \inf_{y \in A} \inf_{x \in f^{-1}(\{y\})} I(x) = \inf_{x \in f^{-1}(A)} I(x).$$

Suppose that $A \subset \mathcal{Y}$ is open. Since f is continuous, $f^{-1}(A) \subset \mathcal{X}$ is open and using that $\{\mu_\varepsilon\}$ satisfies a LDP with the good rate function I:

$$-\inf_{y \in A} J(y) = -\inf_{x \in f^{-1}(A)} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(f^{-1}(A)) = \liminf_{\varepsilon \to 0} \varepsilon \log [\mu_\varepsilon \circ f^{-1}](A).$$

The upper bound for the LDP of $\{\mu_\varepsilon \circ f^{-1}\}$ is proved analogously. ■

A reasonable question is whether the reverse of the contraction principle holds, that is, if $\{\mu_\varepsilon \circ f^{-1}\}$ satisfies a LDP, then $\{\mu_\varepsilon\}$ also satisfies a LDP whenever f is a continuous function. The inverse contraction principle shows that in presence of exponential tightness of $\{\mu_\varepsilon\}$ and bijectivity of f we have such result.

Theorem 3.4.2. ***Inverse contraction principle***. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces, $f : \mathcal{X} \to \mathcal{Y}$ a continuous bijection function and $\{\mu_\varepsilon\}$ an exponentially tight family of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*

If the family of probability measures $\{\mu_\varepsilon \circ f^{-1}\}$ on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ satisfies a LDP with a rate function $I : \mathcal{Y} \to [0, \infty]$, then $\{\mu_\varepsilon\}$ satisfies a LDP with the good rate function $J := I \circ f$.

Proof. It is clear that J is nonnegative. We are going to prove that for any $\alpha \in [0, \infty)$

$$\psi_J(\alpha) = f^{-1}(\psi_I(\alpha)).$$

($\subset$) Let $x \in \mathcal{X}$ such that $J(x) = I(f(x)) \leq \alpha$. Then, $y = f(x) \in \mathcal{Y}$ satisfy $I(y) \leq \alpha$.

($\supset$) Let $y \in \mathcal{Y}$ such that $I(y) \leq \alpha$. Since f is a bijection there exists $x \in \mathcal{X}$ with $f(x) = y$. Then, $J(x) = I(f(x)) = I(y) \leq \alpha$.

Since f is continuous we conclude that J is a rate function.

Since $\{\mu_\varepsilon\}$ is an exponentially tight family it is enough by Lemma 1.4.10 to prove a weak LDP with rate function J to conclude that $\{\mu_\varepsilon\}$ satisfies a LDP with the good rate function J.

Consider $K \subset \mathcal{X}$ a compact set. Then, $f(K) \subset \mathcal{Y}$ is compact and since $\mathcal{X}$ Hausdorff, $f(K) \subset \mathcal{Y}$ is closed. Using that $\{\mu_\varepsilon \circ f^{-1}\}$ satisfies a LDP with rate function I and that f is a bijection we get

$$\begin{aligned}\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K) &= \limsup_{\varepsilon\to 0} \varepsilon \log[\mu_\varepsilon \circ f^{-1}](f(K)) \\ &\leq -\inf_{y\in f(K)} I(y) = -\inf_{x\in K} I(f(x)) = -\inf_{x\in K} J(x).\end{aligned}$$

Let $G \subset \mathcal{X}$ be an open set and $x \in G$ with $J(x) = I(f(x)) = \alpha < \infty$. Note that we can assume the fact that $\alpha < \infty$ because otherwise the lower bound for the LDP would be immediate.

Since $\{\mu_\varepsilon\}$ is exponentially tight, there exist a compact set $K_\alpha \subset \mathcal{X}$ such that

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_\alpha^{\mathrm{c}}) < -\alpha. \tag{3.4.1}$$

The set $f(K_\alpha) \subset \mathcal{Y}$ is compact and therefore, $f(K_\alpha)^{\mathrm{c}} = f(K_\alpha^{\mathrm{c}}) \subset \mathcal{Y}$ is open. Using that $\{\mu_\varepsilon \circ f^{-1}\}$ satisfies a LDP with rate function I and that f is a bijection we obtain

$$-\inf_{y\in f(K_\alpha^{\mathrm{c}})} I(y) \leq \liminf_{\varepsilon\to 0} \varepsilon \log[\mu_\varepsilon \circ f^{-1}](f(K_\alpha^{\mathrm{c}})) = \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(K_\alpha^{\mathrm{c}}) < -\alpha.$$

Since $J(x) = I(f(x)) = \alpha$, the previous inequality implies that $x \in K_\alpha$. Note that, by Lemma 3.1.3, f is a homeomorphism between K_α and $f(K_\alpha)$ because K_α is compact, $f(K_\alpha)$ is Hausdorff and the restriction of f to K_α is also a continuous bijection. Observe that

$G \cap K_\alpha$ neighborhood of x in the induced topology on $K_\alpha \subset \mathcal{X}$.

$\Rightarrow f(G \cap K_\alpha)$ neighborhood of $f(x)$ in the induced topology on $f(K_\alpha) \subset \mathcal{Y}$.
$\Rightarrow$ There exist $U \subset \mathcal{Y}$ neighborhood of $f(x)$ in $\mathcal{Y}$ such that $f(G \cap K_\alpha) = U \cap f(K_\alpha)$.

Using that f is a bijection

$$U = (U \cap f(K_\alpha)) \cup (U \cap f(K_\alpha^c)) \subset f(G \cap K_\alpha) \cup f(K_\alpha^c) = f(G \cup K_\alpha^c).$$

Then, for every $\varepsilon > 0$,

$$[\mu_\varepsilon \circ f^{-1}](U) \leq [\mu_\varepsilon \circ f^{-1}](f(G \cup K_\alpha^c)) = \mu_\varepsilon(G \cup K_\alpha^c) \leq \mu_\varepsilon(G) + \mu_\varepsilon(K_\alpha^c). \tag{3.4.2}$$

Finally by the inequalities in (3.4.2) and (3.4.1) and (1.4.3) of Lemma 1.4.9,

$$\begin{aligned} -\alpha = -J(x) = -I(f(x)) &\leq -\inf_{y \in U} I(y) \leq \liminf_{\varepsilon \to 0} \varepsilon \log[\mu_\varepsilon \circ f^{-1}](U) \\ &\leq \liminf_{\varepsilon \to 0} \varepsilon \log[\mu_\varepsilon(G) + \mu_\varepsilon(K_\alpha^c)] \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(G) \vee \limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(K_\alpha^c) \\ &< \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(G) \vee (-\alpha) \Rightarrow -J(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(G). \end{aligned}$$

Taking supremum over $x \in G$ we obtain the lower bound of the LDP for $\{\mu_\varepsilon\}$. ■

The following corollary shows how useful is the inverse contraction principle because it can be used for strengthening the LDP from a coarse topology to a finer one. So, when proving an LDP, in presence of exponential tightness, it is equivalent to study it with a coarser topology as long as it is Hausdorff.

Corollary 3.4.3. Let $\mathcal{X}$ be a set, τ_1 and τ_2 two topologies on $\mathcal{X}$ such that τ_1 is Hausdorff and $\tau_1 \subset \tau_2$, and let $\{\mu_\varepsilon\}$ be an exponentially tight family of probability measures on $(\mathcal{X}, \mathcal{B}(\tau_2))$, where $\mathcal{B}(\tau_i)$ is the σ-field generated by τ_i.

If $\{\mu_\varepsilon\}$ satisfies a LDP in $(\mathcal{X}, \mathcal{B}(\tau_1))$, then $\{\mu_\varepsilon\}$ also satisfies the same LDP in $(\mathcal{X}, \mathcal{B}(\tau_2))$.

Proof. Note that (X, τ_1) and (X, τ_2) are Hausdorff topological spaces,

$$Id : (X, \tau_2) \to (X, \tau_1)$$

is a continuous bijection because $\tau_1 \subset \tau_2$. Since $\{\mu_\varepsilon\}$ is an exponentially tight family of probability measures on $(\mathcal{X}, \mathcal{B}(\tau_2))$ and it satisfies a LDP in $(\mathcal{X}, \mathcal{B}(\tau_1))$, we conclude by Theorem 3.4.2 that $\{\mu_\varepsilon\}$ satisfies the same LDP in $(\mathcal{X}, \mathcal{B}(\tau_2))$. ■

3.4.2 Exponential approximations

It is intuitive that if a family of probability measures $\{\mu_\varepsilon\}$ satisfies a LDP, then for a close enough family of probability measures $\{\widetilde{\mu}_\varepsilon\}$ a LDP is also satisfied. The proper notion of closeness in this situation is the concept of exponential approximation.

Definition 3.4.4. Let $(\mathcal{Y}, d)$ be a separable metric space. Two families of $\mathcal{Y}$-valued random variables $\{Z_\varepsilon\}$ and $\{\widetilde{Z}_\varepsilon\}$ are called exponentially equivalent if, for every $\delta > 0$,

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_\varepsilon) > \delta) = -\infty.$$

We are imposing that the probability that $\widetilde{Z}_\varepsilon$ and Z_ε differ from more than $\delta > 0$ to behave, for example, like $e^{-\frac{1}{\varepsilon^2}}$. So, such sequence not only goes to zero but it does with this exponentially fast ratio.

It can be proved that $\{\omega \in \Omega : \ d(\widetilde{Z}_\varepsilon(\omega), Z_\varepsilon(\omega)) > \delta\} \in \mathcal{F}$ using that $\mathcal{Y}$ is separable.

The next step is to define the concept of exponentially good approximation. The idea is that for a fixed family of random variables $\{\widetilde{Z}_\varepsilon\}$, we consider a sequence of families of random variables $\{Z_{\varepsilon,m}\}$ that asymptotically behaves like it was exponentially equivalent.

Definition 3.4.5. Let $(\mathcal{Y}, d)$ be a separable metric space. A sequence of families of $\mathcal{Y}$-valued random variables $\{Z_{\varepsilon,m}\}$ are called exponentially good approximations of a family of $\mathcal{Y}$-valued random variables $\{\widetilde{Z}_\varepsilon\}$ if, for every $\delta > 0$,

$$\lim_{m \to \infty} \limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta) = -\infty \tag{3.4.3}$$

Note that in the previous definition, if $\{Z_{\varepsilon,m}\}$ does not depend on m we recover Definition 3.4.4. The following important theorem justify all the previous definitions. The main idea is that if $\{Z_{\varepsilon,m}\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$, we can infer the LDP for the laws of $\{\widetilde{Z}_\varepsilon\}$ from the LDP of the laws of $\{Z_{\varepsilon,m}\}$ for fixed m.

Theorem 3.4.6. *Let $(\mathcal{Y}, d)$ be a separable metric space, $\{Z_{\varepsilon,m}\}$ a sequence of families of $\mathcal{Y}$-valued random variables and $\{\widetilde{Z}_\varepsilon\}$ a family of $\mathcal{Y}$-valued random variables with laws $\{\mu_{\varepsilon,m}\}$ and $\{\widetilde{\mu}_\varepsilon\}$, respectively. Suppose that for every $m \geq 1$, the family $\{\mu_{\varepsilon,m}\}$ satisfies a LDP with a rate function I_m and that $\{Z_{\varepsilon,m}\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$. Then,*

1. *$\{\widetilde{\mu}_\varepsilon\}$ satisfies a weak LDP with the rate function*

$$I(y) := \sup_{\delta > 0} \liminf_{m \to \infty} \inf_{z \in B_\delta(y)} I_m(z). \tag{3.4.4}$$

2. If I is a good rate function and for every closed set $F \subset Y$,

$$\inf_{y\in F} I(y) \le \limsup_{m\to\infty} \inf_{y\in F} I_m(y), \tag{3.4.5}$$

then the full LDP holds for $\{\widetilde{\mu}_\varepsilon\}$ with rate function I.

Proof.
(1) The objective is to apply Theorem 3.2.1 in order to show that $\{\widetilde{\mu}_\varepsilon\}$ satisfies a weak LDP. Note that $\mathcal{A} = \{B_\delta(y),\ y \in \mathcal{Y},\ \delta > 0\}$ is a base of $\mathcal{Y}$ and that for any function $L : \mathcal{A} \to [-\infty, \infty]$

$$\sup_{\{A\in\mathcal{A}:\ y\in A\}} L(A) = \sup_{\delta>0} L\left(B_\delta(y)\right).$$

Summarising, we have to check that the function defined in equation (3.4.4) is equal to the rate function defined in Theorem 3.2.1 and that condition (3.2.2) is satisfied. By the previous equation this is equivalent to prove the two following equalities

$$I(y) = \sup_{\delta>0}\left(-\liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon\left(B_\delta(y)\right)\right) = \sup_{\delta>0}\left(-\limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon\left(B_\delta(y)\right)\right). \tag{3.4.6}$$

We can rewrite equation (3.4.6) as

$$I(y) = -\inf_{\delta>0} \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon\left(B_\delta(y)\right) = -\inf_{\delta>0} \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon\left(B_\delta(y)\right). \tag{3.4.7}$$

Now, fix $\delta > 0$ and $y \in \mathcal{Y}$. Then, for every $m \in \mathbb{N}$ and every $\varepsilon > 0$ we have

$$\{Z_{\varepsilon,m} \in B_\delta(y)\} \subset \{\widetilde{Z}_\varepsilon \in B_{2\delta}(y)\} \cup \{d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta\}.$$

Hence,

$$\mu_{\varepsilon,m}(B_\delta(y)) \le \widetilde{\mu}_\varepsilon(B_{2\delta}(y)) + \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta).$$

Since for every $m \ge 1$ the family $\{\mu_{\varepsilon,m}\}$ satisfies a LDP with rate function I_m, applying (1.4.3) of Lemma 1.4.9 we obtain

$$\begin{aligned} -\inf_{z\in B_\delta(y)} I_m(z) &\le \liminf_{\varepsilon\to 0} \varepsilon \log \mu_{\varepsilon,m}(B_\delta(y)) \le \liminf_{\varepsilon\to 0} \varepsilon \log \left[\widetilde{\mu}_\varepsilon(B_{2\delta}(y)) + \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta)\right] \\ &\le \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_{2\delta}(y)) \ \vee\ \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta) \end{aligned}$$

Taking $\limsup$ as $m \to \infty$ and using that $\{Z_{\varepsilon,m}\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$ we obtain

$$\limsup_{m\to\infty}[-\inf_{z\in B_\delta(y)} I_m(z)] \le \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_{2\delta}(y)). \tag{3.4.8}$$

Similarly, using that

$$\{\widetilde{Z}_\varepsilon \in B_\delta(y)\} \subset \{Z_{\varepsilon,m} \in B_{2\delta}(y)\} \cup \{d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta\},$$

and that for every $m \geq 1$ the family $\{\mu_{\varepsilon,m}\}$ satisfies a LDP with rate function I_m we obtain

$$-\inf_{z\in\overline{B}_{2\delta}(y)} I_m(z) \geq \limsup_{\varepsilon\to 0} \varepsilon \log \mu_{\varepsilon,m}(B_{2\delta}(y)) \geq \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)).$$

Taking $\limsup$ as $m \to \infty$,

$$\limsup_{m\to\infty}[-\inf_{z\in\overline{B}_{2\delta}(y)} I_m(z)] \geq \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)). \tag{3.4.9}$$

Taking infimum over $\delta > 0$ in inequality (3.4.8) we obtain

$$\begin{aligned} &\inf_{\delta>0} \limsup_{m\to\infty}[-\inf_{z\in B_\delta(y)} I_m(z)] = -\sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_\delta(y)} I_m(z) = -I(y) \\ &\leq \inf_{\delta>0} \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_{2\delta}(y)) = \inf_{\delta>0} \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)). \end{aligned} \tag{3.4.10}$$

Taking infimum over $\delta > 0$ in inequality (3.4.9) we obtain

$$\begin{aligned} &\inf_{\delta>0} \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)) \leq \inf_{\delta>0} \limsup_{m\to\infty}[-\inf_{z\in\overline{B}_{2\delta}(y)} I_m(z)] \\ &= -\sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in\overline{B}_{2\delta}(y)} I_m(z) \leq -\sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_{3\delta}(y)} I_m(z) \\ &= -\sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_\delta(y)} I_m(z) = -I(y). \end{aligned} \tag{3.4.11}$$

Finally, combining the inequalities in (3.4.10) and (3.4.11) we have

$$-I(y) \leq \inf_{\delta>0} \liminf_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)) \leq \inf_{\delta>0} \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(B_\delta(y)) \leq -I(y),$$

and condition (3.4.7) is satisfied. We conclude that I is a rate function and that $\{\widetilde{\mu}_\varepsilon\}$ satisfies a weak LDP with the rate function I.

(2) We only have to check that $\{\widetilde{\mu}_\varepsilon\}$ satisfies the LDP upper bound for closed sets. Fix $\delta > 0$ and let $F \subset \mathcal{Y}$ be a closed set. Note that for every $m \geq 1$ and every $\varepsilon > 0$

$$\{\widetilde{Z}_\varepsilon \in F\} \subset \{Z_{\varepsilon,m} \in F^\delta\} \cup \{d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta\},$$

where $F^\delta = \{y \in \mathcal{Y} : d(y, F) \leq \delta\}$ is the closed blowup of F. Then, applying (1.4.2) of Lemma 1.4.9, the fact that $F^\delta \subset \mathcal{Y}$ is closed and that $\{\mu_{\varepsilon,m}\}$ satisfies a LDP with the rate function I_m we obtain

$$\limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(F) \leq \limsup_{\varepsilon\to 0} \varepsilon \log \left[\mu_{\varepsilon,m}(F^\delta) + \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta)\right]$$

$$
\begin{aligned}
&= \limsup_{\varepsilon\to 0} \varepsilon \log \mu_{\varepsilon,m}(F^\delta) \vee \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta) \\
&\le [-\inf_{y\in F^\delta} I_m(y)] \vee \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(d(\widetilde{Z}_\varepsilon, Z_{\varepsilon,m}) > \delta).
\end{aligned}
$$

Taking liminf as $m \to \infty$, using that $\{Z_{\varepsilon,m}\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$ and the hypothesis in (3.4.5) for the closed set F^δ we obtain

$$
\begin{aligned}
\limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(F) &\le \liminf_{m\to\infty} [-\inf_{y\in F^\delta} I_m(y)] \\
&= -\limsup_{m\to\infty} \inf_{y\in F^\delta} I_m(y) = -\inf_{y\in F^\delta} I(y).
\end{aligned}
$$

Finally taking limit as $\delta \to 0$ and using part 2 in Lemma 1.3.7 we obtain

$$
\limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(F) \le \lim_{\delta\to 0} [-\inf_{y\in F^\delta} I(y)] = -\inf_{y\in F} I(y).
$$

So, the upper bound for the LDP of $\{\widetilde{\mu}_\varepsilon\}$ is proved. ■

As a consequence of the previous theorem we can prove the following result: if two families of random variables are exponentially equivalent, then the LDP of the laws of one family implies the LDP of the other.

Theorem 3.4.7. *Let $(\mathcal{Y}, d)$ be a separable metric space and, $\{Z_\varepsilon\}$ and $\{\widetilde{Z}_\varepsilon\}$ two exponentially equivalent families of $\mathcal{Y}$-valued random variables with laws $\{\mu_\varepsilon\}$ and $\{\widetilde{\mu}_\varepsilon\}$, respectively. Suppose that the family $\{\mu_\varepsilon\}$ satisfies a LDP with a good rate function J.*

Then, the family $\{\widetilde{\mu}_\varepsilon\}$ also satisfies a LDP with the same good rate function J.

Proof. Define the sequence of families of $\mathcal{Y}$-valued random variables $\{Z_{\varepsilon,m}\}$ with $Z_{\varepsilon,m} := Z_\varepsilon$. Since $\{Z_\varepsilon\}$ are exponentially equivalent to $\{\widetilde{Z}_\varepsilon\}$, $\{Z_{\varepsilon,m}\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$.

Note that we are in the hypothesis to apply Theorem 3.4.6 with $I_m = J$. So, $\{\widetilde{\mu}_\varepsilon\}$ satisfies a weak LDP with the rate function

$$
I(y) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_\delta(y)} J(z) = \sup_{\delta>0} \inf_{z\in B_\delta(y)} J(z) = J(y) \Rightarrow I = J.
$$

In the third equality we have used that J is a lower semicontinuous function and Proposition 1.3.3.

In addition, we are also in the hypothesis to apply part 2 of Theorem 3.4.6 because $I_m = J = I$ for all $m \in \mathbb{N}$. We conclude that $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP with the good rate function J. ■

We finish this section with a very important result that will be used in Chapter 5 when proving the the LDP for stochastic differential equations. In some sense, is an extension of Theorem 3.4.6 and the contraction principle to maps that are not continuous, but that can be approximated well enough by continuous maps.

Theorem 3.4.8. *Let $\mathcal{X}$ be a Hausdorff topological space, $(\mathcal{Y}, d)$ a separable metric space, $\{Z_\varepsilon\}$ a family of $\mathcal{X}$-valued random variables with laws $\{\mu_\varepsilon\}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ that satisfies a LDP with a good rate function I and $\{\widetilde{Z}_\varepsilon\}$ a family of $\mathcal{Y}$-valued random variables with laws $\{\widetilde{\mu}_\varepsilon\}$ on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$. For every $m \geq 1$, let $F_m : \mathcal{X} \to \mathcal{Y}$ be continuous functions and $F : \mathcal{X} \to \mathcal{Y}$ a measurable map such that for every $\alpha < \infty$,*

$$\limsup_{m\to\infty} \sup_{\{x\in\mathcal{X}:\ I(x)\leq\alpha\}} d(F_m(x), F(x)) = 0. \tag{3.4.12}$$

Then, for every $\alpha < \infty$, F is continuous on the level set $\psi_I(\alpha) = \{x \in \mathcal{X} :\ I(x) \leq \alpha\}$.

In addition, assume that the sequence of families of $\mathcal{Y}$-valued random variables $\{F_m \circ Z_\varepsilon\}$, which have law $\{\mu_\varepsilon \circ F_m^{-1}\}$ on $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$.

Then, the family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $\mathcal{Y}$ with the good rate function.

$$J(y) := \inf_{\{x\in\mathcal{X}:\ y=F(x)\}} I(x) = \inf_{x\in F^{-1}(\{y\})} I(x).$$

Proof. Let $\alpha < \infty$ and $x_1, x_2 \in \psi_I(\alpha)$, then

$$d(F(x_1), F(x_2)) \leq d(F(x_1), F_m(x_1)) + d(F_m(x_1), F_m(x_2)) + d(F_m(x_2), F(x_2))$$

The continuity of F on $\psi_I(\alpha)$ follows from condition (3.4.12) and the fact that F_m are continuous.

For the second part of the Theorem, we know by the contraction principle (Theorem 3.4.1) that for every $m \geq 1$, the family $\{\mu_\varepsilon \circ F_m^{-1}\}$ satisfies a LDP in $\mathcal{Y}$ with the good rate function

$$I_m(y) = \inf_{\{x\in\mathcal{X}:\ y=F_m(x)\}} I(x) = \inf_{x\in F_m^{-1}(\{y\})} I(x).$$

which has level sets $\psi_{I_m}(\alpha) = F_m(\psi_I(\alpha))$.

Then, since $\{F_m \circ Z_\varepsilon\}$ are exponentially good approximations of $\{\widetilde{Z}_\varepsilon\}$, by part 1 in Theorem 3.4.6 we have that the family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a weak LDP in $\mathcal{Y}$ with rate function

$$\overline{J}(y) := \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_\delta(y)} I_m(z). \tag{3.4.13}$$

So, our objective is to prove that $J = \overline{J}$ and that we are under the hypothesis of part 2 in Theorem 3.4.6.

We start proving that J is a good rate function. For this, we check that for any $\alpha < \infty$, $\psi_J(\alpha) = F(\psi_I(\alpha))$.

($\subset$) Let $y \in \mathcal{Y}$ such that $J(y) \leq \alpha$. By the definition of infimum, there exists $x \in \mathcal{X}$ with $y = F(x)$ and $J(y) \leq I(x) \leq J(y) + 1 \leq \alpha + 1$. So, $x \in \psi_I(\alpha + 1)$ and $y \in F(\psi_I(\alpha + 1))$. Since $\{y\} \subset \mathcal{Y}$ is closed and F is continuous on $\psi_I(\alpha + 1)$, we have that $F^{-1}(\{y\}) \neq \varnothing$ is closed in $\mathcal{X}$. By Lemma 1.3.6, since I is a good rate function the infimum in the definition of J is achieved on some point $x^{'} \in \mathcal{X}$. Note that $y = F(x^{'})$ and $I(x^{'}) = J(y) \leq \alpha$.

($\supset$) Let $F(x) \in \mathcal{Y}$ with $I(x) \leq \alpha$. Then, $y = F(x) \in \mathcal{Y}$ and $J(y) \leq I(x) \leq \alpha$.

Since for any $\alpha < \infty$, F is continuous on $\psi_I(\alpha)$, I is a good rate function and $\psi_J(\alpha) = F(\psi_I(\alpha))$, we conclude that J is a good rate function.

We now prove that condition (3.4.5) is satisfied for J. Consider $C \subset \mathcal{Y}$ a closed subset and for any $m \geq 1$ define

$$\gamma_m := \inf_{y \in C} I_m(y) = \inf_{y \in C} \inf_{x \in F_m^{-1}(\{y\})} I(x) = \inf_{x \in F_m^{-1}(C)} I(x).$$

Suppose that $\gamma := \liminf_{m \to \infty} \gamma_m < \infty$ and consider a subsequence $\{\gamma_{m_k}\}$ with $\lim_{k \to \infty} \gamma_{m_k} = \gamma$. Let $\alpha < \infty$ be such that $\sup_{k \geq 1} \gamma_{m_k} = \alpha$.

Note that $F_{m_k}^{-1}(C) \subset \mathcal{X}$ is closed and non-empty because $\gamma_{m_k} \leq \alpha$. Since I is a good rate function, by Lemma 1.3.6 there exists $x_k \in \mathcal{X}$ such that $F_{m_k}(x_k) \in C$ and

$$I(x_k) = \inf_{x \in F_{m_k}^{-1}(C)} I(x) = \gamma_{m_k} \leq \alpha.$$

Therefore, for any $\delta > 0$, there exists k_δ such that for all $k \geq k_\delta$ we have

$$F(x_k) \in C^\delta = \left\{ y \in \mathcal{Y} : \ d(y, C) := \inf_{c \in C} d(y, c) \leq \delta \right\}$$

because by condition (3.4.12)

$$d(F(x_k), F_{m_k}(x_k)) \leq \sup_{\{x \in \mathcal{X} : \ I(x) \leq \alpha\}} d(F(x), F_{m_k}(x)) \xrightarrow[k \to \infty]{} 0.$$

Then, for $k \geq k_\delta$,

$$\inf_{y \in C^\delta} J(y) \leq J(F(x_k)) \leq I(x_k) = \gamma_{m_k},$$

Taking limit as $k \to \infty$ in the previous inequalities, we obtain

$$\inf_{y \in C^\delta} J(y) \leq \lim_{k\to\infty} \gamma_{m_k} = \gamma = \liminf_{m\to\infty} \gamma_m = \liminf_{m\to\infty} \inf_{y \in C} I_m(y).$$

Note that the previous inequality trivially holds if $\gamma = \infty$. Taking limit as $\delta \to 0$ and using part 2 in Lemma 1.3.7 we have

$$\inf_{y \in C} J(y) = \lim_{\delta \to 0} \inf_{y \in C^\delta} J(y) \leq \liminf_{m\to\infty} \inf_{y\in C} I_m(y). \tag{3.4.14}$$

We conclude that J satisfies condition (3.4.5) of Theorem 3.4.6. To finish the proof of this Theorem it only remains to check that $J = \overline{J}$, where $\overline{J}$ is defined in (3.4.13). Consider $y \in \mathcal{Y}$, then,

$$\begin{aligned} J(y) &= \sup_{\delta>0} \inf_{z\in B_\delta(y)} J(z) = \sup_{\delta>0} \inf_{z \in \overline{B}_\delta(y)} J(z) \\ &\leq \sup_{\delta>0} \liminf_{m\to\infty} \inf_{y\in\overline{B}_\delta(y)} I_m(z) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{y \in B_\delta(y)} I_m(z) = \overline{J}(y), \end{aligned}$$

where in the first equality we have used Proposition 1.3.3 and in the inequality the result in (3.4.14) with $C = \overline{B}_\delta(y)$. So $J(y) \leq \overline{J}(y)$.

For the converse inequality, we can assume without loss of generality that $J(y) = \alpha < \infty$ because we already know that $J(y) \leq \overline{J}(y)$. Then, $y \in \psi_J(\alpha) = F(\psi_I(\alpha))$. So, there exists $x \in \mathcal{X}$ such that $y = F(x)$ and $I(x) \leq \alpha$.

Note that $y_m := F_m(x) \in F_m(\psi_I(\alpha)) = \psi_{I_m}(\alpha)$. Hence, $I_m(y_m) \leq \alpha$ and by condition (3.4.12) we have

$$d(y, y_m) = d(F(x), F_m(x)) \leq \sup_{\{x \in \mathcal{X}:\ I(x) \leq \alpha\}} d(F(x), F_m(x)) \xrightarrow[m\to\infty]{} 0$$

In summary, for every $\delta > 0$, there exists $m_\delta \geq 1$ such that for all $m \geq m_\delta$, $y_m \in B_\delta(y)$. This implies that $\inf_{z\in B_\delta(y)} I_m(z) \leq \alpha$ and

$$\overline{J}(y) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_\delta(y)} I_m(z) \leq \alpha = J(y).$$

This finishes the proof of the Theorem. ■

4 LDP for the Brownian motion: Schilder's Theorem

In this chapter we are going to study the large deviation principle for the sample paths of Brownian motion. The idea is to perturb a standard Brownian motion in such a way that its sample paths converge to a deterministic function and study the exponential velocity of this convergence.

An important difference between this chapter and Chapter 2 is that in this case we are going to study the LDP problem in an infinite dimensional space. This implies that we will need some powerful techniques, like Girsanov Theorem.

We fix some notation that will be used through this chapter.

Notation 4.0.1. For $x \in \mathbb{R}^d$, we will write $x = (x^{(1)}, ..., x^{(d)})$ and

$$|x| := \sqrt{(x^{(1)})^2 + ... + (x^{(d)})^2},$$

for the Euclidean norm on $\mathbb{R}^d$.

For this chapter, unless otherwise specified, $B = \{B(t),\ t \in [0,1]\}$ will denote a standard Brownian motion in $\mathbb{R}^d$. That is, $B(t) = (B^{(1)}(t), ..., B^{(d)}(t))$ where $B^{(i)}$ are independent standard Brownian motions in $\mathbb{R}$, namely, for all $1 \leq i \leq d$,

1. $B^{(i)}(0) = 0$ a.s.
2. For any $0 \leq s \leq t \leq 1$, $B^{(i)}(t) - B^{(i)}(s)$ is independent of $\sigma\left(B^{(i)}(r),\ 0 \leq r \leq s\right)$.
3. For any $0 \leq s \leq t \leq 1$, $B^{(i)}(t) - B^{(i)}(s) \stackrel{\mathrm{d}}{=} N(0, t-s)$.

4.1 Preliminaries of Brownian Motion

We start with some basic but useful results on Brownian motion.

Lemma 4.1.1. Let $X \stackrel{\mathrm{d}}{=} N(0,1)$. Then, for any $\delta > 0$,

$$\mathbb{P}(X \geq \delta) \leq \frac{1}{\sqrt{2\pi}\delta} \exp\left(\frac{-\delta^2}{2}\right).$$

Proof.

$$\mathbb{P}(X \geq \delta) = \frac{1}{\sqrt{2\pi}} \int_\delta^\infty \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \leq \frac{1}{\sqrt{2\pi}} \int_\delta^\infty \frac{x}{\delta} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}t = \frac{1}{\sqrt{2\pi}\delta} \exp\left(\frac{-\delta^2}{2}\right).$$

■

Lemma 4.1.2. Désiré Andre's reflection principle. Let $B = \{B(t),\ t \in [0,1]\}$ be a standard Brownian motion in $\mathbb{R}$. Then, for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{0\le t\le 1} B(t) \ge \delta\right) \le 2\ \mathbb{P}\left(B(1) \ge \delta\right).$$

Proof. For a reference, see [3], Theorem 2.14 in page 74. ∎

Thanks to the two previous results, we can prove the following important result which will be used several times in this chapter.

Lemma 4.1.3. For any $\delta > 0$ and $0 < \tau \le 1$ we have

$$\mathbb{P}\left(\sup_{0\le t\le \tau} |B(t)| \ge \delta\right) \le \frac{4\sqrt{d^3\tau}}{\sqrt{2\pi}\delta} \exp\left(\frac{-\delta^2}{2d\tau}\right). \tag{4.1.1}$$

Proof.

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le \tau} |B(t)| \ge \delta\right) = \mathbb{P}\left(\sup_{0\le t\le \tau} |B(t)|^2 \ge \delta^2\right) \\
&\le \mathbb{P}\left(\sum_{i=1}^{d} \sup_{0\le t\le \tau} (B^{(i)}(t))^2 \ge \delta^2\right) \le \mathbb{P}\left(\bigcup_{i=1}^{d} \sup_{0\le t\le \tau} (B^{(i)}(t))^2 \ge \frac{\delta^2}{d}\right) \\
&\le \sum_{i=1}^{d} \mathbb{P}\left(\sup_{0\le t\le \tau} (B^{(i)}(t))^2 \ge \frac{\delta^2}{d}\right) = d\ \mathbb{P}\left(\sup_{0\le t\le \tau} (B^{(1)}(t))^2 \ge \frac{\delta^2}{d}\right) \\
&\overset{(1)}{=} d\ \mathbb{P}\left(\sup_{0\le t\le \tau} \sqrt{\tau}|B^{(1)}(t/\tau)| \ge \frac{\delta}{\sqrt{d}}\right) = d\ \mathbb{P}\left(\sup_{0\le t\le 1} |B^{(1)}(t)| \ge \frac{\delta}{\sqrt{d\tau}}\right) \\
&= d\ \mathbb{P}\left[\left(\sup_{0\le t\le 1} B^{(1)}(t) \ge \frac{\delta}{\sqrt{d\tau}}\right) \bigcup \left(\sup_{0\le t\le 1} -B^{(1)}(t) \ge \frac{\delta}{\sqrt{d\tau}}\right)\right] \\
&\le d\ \mathbb{P}\left(\sup_{0\le t\le 1} B^{(1)}(t) \ge \frac{\delta}{\sqrt{d\tau}}\right) + d\ \mathbb{P}\left(\sup_{0\le t\le 1} -B^{(1)}(t) \ge \frac{\delta}{\sqrt{d\tau}}\right) \\
&\overset{(2)}{=} 2d\ \mathbb{P}\left(\sup_{0\le t\le 1} B^{(1)}(t) \ge \frac{\delta}{\sqrt{d\tau}}\right) \overset{(3)}{\le} 4d\ \mathbb{P}\left(B^{(1)}(1) \ge \frac{\delta}{\sqrt{d\tau}}\right) \\
&\overset{(4)}{\le} \frac{4\sqrt{d^3\tau}}{\sqrt{2\pi}\delta} \exp\left(\frac{-\delta^2}{2d\tau}\right),
\end{aligned}$$

where we have used

1. $\{B^{(1)}(t),\ 0 \le t \le \tau\}$ and $\{\sqrt{\tau}B^{(1)}(t/\tau),\ 0 \le t \le \tau\}$ have the same law.
2. $\{B^{(1)}(t),\ 0 \le t \le 1\}$ and $\{-B^{(1)}(t),\ 0 \le t \le 1\}$ have the same law.

3. Lemma 4.1.2 with $B^{(1)}$ as a standard Brownian motion in $\mathbb{R}$,
4. Lemma 4.1.1.

■

In fact, for the proof of Shchilder's Theorem it is enough to use a particular case of Girsanov Theorem: the Cameron-Martin formula. In order to state this theorem, we first set some notation.

Notation 4.1.4. Let $\psi \in L^2([0,1])$ where

$$L^2([0,1]) := \left\{ \phi : [0,1] \to \mathbb{R}^d \text{ measurable and such that } \int_0^1 |\psi(s)|^2 \mathrm{d}s < \infty \right\}$$

Then, we write for $0 \leq t \leq 1$,

$$\int_0^t \psi(s) \mathrm{d}B(s) := \sum_{i=1}^d \int_0^t \psi^{(i)}(s) \mathrm{d}B^{(i)}(s),$$

and

$$\int_0^t \psi(s) \mathrm{d}s := \left(\int_0^t \psi^{(1)}(s) \mathrm{d}s \;, \dots \;, \int_0^t \psi^{(d)}(s) \mathrm{d}s \right).$$

Note that under such hypothesis, every integral is well defined.

Theorem 4.1.5. Cameron-Martin (1944), Girsanov (1966). *Consider $\psi \in L^2([0,1])$. Then,*

1. *$\widetilde{\mathbb{P}}$ defined by*

$$\widetilde{\mathbb{P}}(A) := \int_A Z_\psi \, d\mathbb{P}, \qquad A \in \mathcal{F},$$

is a probability in $(\Omega, \mathcal{F})$ where

$$Z_\psi := \exp\left(\int_0^1 \psi(s) \, dB(s) - \frac{1}{2} \int_0^1 |\psi(s)|^2 \, ds \right).$$

2. *The process $\widetilde{B} := \{\widetilde{B}(t),\ t \in [0,1]\}$ defined by*

$$\widetilde{B}(t) := B(t) - \int_0^t \psi(s) \, ds, \qquad 0 \leq t \leq 1,$$

is a standard Brownian motion in $\mathbb{R}^d$ with respect to $\widetilde{\mathbb{P}}$.

Proof. For a reference, see [3], Theorem 2.23 in page 86. ■

The two following results are natural extensions of the properties of the Itô integral and Brownian motion in $\mathbb{R}$ to $\mathbb{R}^d$, which will be used in the proof of Schilder's Theorem.

Lemma 4.1.6. Isometry Property. Let $\phi \in L^2([0,1])$. Then,

$$E\left(\left(\int_0^1 \phi(s)\mathrm{d}B(s)\right)^2\right) = \int_0^1 |\phi(s)|^2\mathrm{d}s.$$

Proof.

$$\begin{aligned}
E\left(\left(\int_0^1 \phi(s)\mathrm{d}B(s)\right)^2\right) &= E\left(\left(\sum_{i=1}^d \int_0^1 \phi^{(i)}(s)\mathrm{d}B^i(s)\right)^2\right) \\
&= \sum_{i=1}^d E\left(\left(\int_0^1 \phi^{(i)}(s)\mathrm{d}B^{(i)}(s)\right)^2\right) + \sum_{i\neq j} E\left(\int_0^1 \phi^{(i)}(s)\mathrm{d}B^{(i)}(s)\int_0^1 \phi^{(j)}(s)\mathrm{d}B^{(j)}(s)\right) \\
&\overset{(1)}{=} \sum_{i=1}^d E\left(\int_0^1 \left(\phi^{(i)}(s)\right)^2 \mathrm{d}s\right) = \int_0^1 \sum_{i=1}^d \left(\phi^{(i)}(s)\right)^2 \mathrm{d}s = \int_0^1 |\phi(s)|^2\mathrm{d}s,
\end{aligned}$$

where in (1) we have used the isometry property of the Itô integral, the fact that

$$\int_0^1 \phi^{(i)}(s)\mathrm{d}B^{(i)}(s) \quad \text{and} \quad \int_0^1 \phi^{(j)}(s)\mathrm{d}B^{(j)}(s)$$

are independent for $i \neq j$ and that the Itô integral is centered. ■

Lemma 4.1.7. Stationary increments. Let $0 \le s \le t \le 1$, then,

$$|B(t) - B(s)| \overset{\mathrm{d}}{=} |B(t-s)|.$$

Proof. Since $|B(t) - B(s)|$ and $|B(t-s)|$ are positive random variables, it is enough to prove that

$$|B(t) - B(s)|^2 \overset{\mathrm{d}}{=} |B(t-s)|^2.$$

Since each Brownian motion $B^{(i)}$ has stationary increments,

$$\left(B^{(i)}(t) - B^{(i)}(s)\right)^2 \overset{\mathrm{d}}{=} \left(B^{(i)}(t-s)\right)^2, \qquad i = 1, ..., d.$$

Then, writing φ_X for the characteristic function of a random variable X and using the fact that the Brownian motions $B^{(i)}$ are independent we have

$$\varphi_{|B(t)-B(s)|^2} = \varphi_{\sum_{i=1}^d \left(B^{(i)}(t)-B^{(i)}(s)\right)^2} = \prod_{i=1}^d \varphi_{\left(B^{(i)}(t)-B^{(i)}(s)\right)^2}$$
$$= \prod_{i=1}^d \varphi_{\left(B^{(i)}(t-s)\right)^2} = \varphi_{\sum_{i=1}^d \left(B^{(i)}(t-s)\right)^2} = \varphi_{|B(t-s)|^2}.$$

By injectivity of the characteristic function we deduce that

$$|B(t)-B(s)|^2 \overset{\mathrm{d}}{=} |B(t-s)|^2.$$

■

4.2 Preliminaries of Schilder's Theorem

We modify the Brownian motion in the following way

$$B_\varepsilon := \{B_\varepsilon(t) := \sqrt{\varepsilon} B(t),\ t \in [0,1]\}, \qquad \varepsilon > 0.$$

Our objective is to study how fast are the sample paths of B modified when $\varepsilon \to 0$. Since such sample paths are continuous and vanishing at the origin it is useful to introduce the following space of functions

$$C_0([0,1]) := \{\phi : [0,1] \to \mathbb{R}^d \text{ continuous and } \phi(0) = 0\}.$$

with the supremum norm $\|\phi\| := \sup_{t\in[0,1]} |\phi(t)|$. So, $(C_0([0,1]), \|\cdot\|)$ is a normed space.

Then, the sample paths of B_ε can be seen as the map

$$\begin{aligned} B_\epsilon(\cdot)\colon \Omega &\longrightarrow C_0([0,1]) \\ \omega &\longmapsto B_\varepsilon(\cdot)(\omega)\colon [0,1] \longrightarrow \mathbb{R}^d \\ &\qquad\qquad\qquad\qquad t \longmapsto B_\varepsilon(t)(\omega). \end{aligned}$$

A natural question is whether this map is measurable. As a consequence of the Weierstrass Approximation Theorem we have the following result.

Lemma 4.2.1. $C_0([0,1])$ is a separable metric space.

Proof. For a reference, see [4], Theorem 7.26 in page 159. ■

Proposition 4.2.2. $B_\varepsilon(\cdot) : (\Omega, \mathcal{F}) \to (C_0([0,1]), \mathcal{B}(C_0([0,1])))$ *is measurable.*

Proof. Let $G \subset C_0([0,1])$ be an open subset. In order to prove that $B_\varepsilon(\cdot)$ is measurable it is enough to prove that $B_\varepsilon(\cdot)^{-1}(G) \in \mathcal{F}$.

By Lemma 4.2.1, G is also separable and let $G_1 \subset G$ be a dense countable subset. Note that we have

$$\{\omega \in \Omega :\ B_\varepsilon(\cdot)(\omega) \in G\} = \bigcap_{n\geq 1} \bigcup_{\phi_1 \in G_1} \left\{\omega \in \Omega :\ \|B_\varepsilon(\cdot)(\omega) - \phi_1\| < \frac{1}{n}\right\}. \tag{4.2.1}$$

Moreover, since $B_\varepsilon(\omega)(\cdot) - \phi_1$ are continuous functions we also have

$$\left\{\omega \in \Omega :\ \|B_\varepsilon(\cdot)(\omega) - \phi_1\| < \frac{1}{n}\right\} = \left\{\omega \in \Omega :\ \sup_{t\in[0,1]\cap\mathbb{Q}} |B_\varepsilon(t)(\omega) - \phi_1(t)| < \frac{1}{n}\right\},$$

which is a measurable set because the supremum is taken over a countable set. Finally, $B_\varepsilon(\cdot)^{-1}(G) \in \mathcal{F}$ because by (4.2.1) is the countable intersection of the countable union of measurable sets. ■

As a consequence of the previous proposition, it makes senses to consider the law of $B_\varepsilon(\cdot)$ in $C_0([0,1])$, which we will write as μ_ε. That is, for $G \in \mathcal{B}(C_0([0,1]))$,

$$\mu_\varepsilon(G) = \mathbb{P}(B_\varepsilon(\cdot) \in G) = \mathbb{P}(\{\omega \in \Omega :\ B_\varepsilon(\cdot)(\omega) \in G\}).$$

The next proposition shows, as one expects, that the sample paths of B_ε converge (in probability) to the constant function equal to 0. In fact, we prove a stronger result: such convergence is exponentially fast in $1/\varepsilon$ as a consequence of Lemma 4.1.3. Hence, the next proposition can be seen as a partial result of the LDP problem that we are studying.

Proposition 4.2.3. *Let* $\delta > 0$. *Then,*

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\|B_\varepsilon(\cdot)\| \geq \delta) = \frac{-\delta^2}{2d}.$$

In particular,

$$\lim_{\varepsilon\to 0} \mathbb{P}(\|B_\varepsilon(\cdot)\| \geq \delta) = 0,$$

that is, B_ε *converges in probability in* $C_0([0,1])$ *to the constant function equal to* 0.

Proof. Let $\delta > 0$. Using Lemma 4.1.3 we have

$$\mathbb{P}\left(\|B_\varepsilon(\cdot)\| \geq \delta\right) = \mathbb{P}\left(\sup_{0\leq t\leq 1} |B_\varepsilon(t)| \geq \delta\right) = \mathbb{P}\left(\sup_{0\leq t\leq 1} |B(t)| \geq \frac{\delta}{\sqrt{\varepsilon}}\right) \leq \frac{4\sqrt{d^3\varepsilon}}{\sqrt{2\pi}\delta} \exp\left(\frac{-\delta^2}{2d\varepsilon}\right).$$

Then,

$$\varepsilon \log \mathbb{P}\left(\|B_\varepsilon(\cdot)\| \geq \delta\right) = \varepsilon \left[\log\left(\frac{4\sqrt{d^3}}{\sqrt{2\pi}\delta}\right) + \log\left(\sqrt{\varepsilon}\right) - \frac{\delta^2}{2d\varepsilon}\right].$$

Using that $\varepsilon \log(\sqrt{\varepsilon}) \to 0$ as $\varepsilon \to 0$ we obtain

$$\lim_{\varepsilon\to 0} \varepsilon \log \mathbb{P}\left(\|B_\varepsilon(\cdot)\| \geq \delta\right) = \frac{-\delta^2}{2d}.$$

■

4.3 Schilder's Theorem

First of all, we introduce some new notation that will be used.

Notation 4.3.1. We will write $\mathcal{P}$ for the set of all partitions of the interval $[0, 1]$. That is,

$$\mathcal{P} := \{\{0 = t_1 < ... < t_N = 1\} : \ N \geq 1\}.$$

We also introduce a subspace of $C_0([0, 1])$ that will appear in the incoming results:

$$H_1 := \left\{f \in C_0([0, 1]) : \ f(t) = \int_0^t g(s)\mathrm{d}s, \ g \in L^2([0, 1])\right\}$$

Note that H_1 is the space of all absolutely continuous functions with square integrable derivative.

Let $\phi \in H_1$, then we will write

$$\dot{\phi} := \left(\frac{\mathrm{d}\phi^{(1)}}{\mathrm{d}t}, ..., \frac{\mathrm{d}\phi^{(d)}}{\mathrm{d}t}\right).$$

In order to prove that the rate function that appears in Schilder's Theorem is, in fact, a good rate function we will need the two following results of analysis. Lemma 4.3.2 is due to Riesz and Lemma 4.3.3 is a version of the Arzelà-Ascoli Theorem.

Lemma 4.3.2. Let $\phi \in C_0([0,1])$. Then

$$\phi \in H_1 \Leftrightarrow \sup_{P\in\mathcal{P}} \sum_{i=1}^{\#P-1} \frac{|\phi(t_{i+1})-\phi(t_i)|^2}{t_{i+1}-t_i} < \infty.$$

Moreover, in that case,

$$\sup_{P\in\mathcal{P}} \sum_{i=1}^{\#P-1} \frac{|\phi(t_{i+1})-\phi(t_i)|^2}{t_{i+1}-t_i} = \int_0^1 |\dot{\phi}(t)|^2 \mathrm{d}t.$$

Proof. For a reference, see [5], Lemma 18 in page 75. ■

Lemma 4.3.3. Arzelà-Ascoli. Let $\mathcal{F} \subset C_0([0,1])$. Then,

$$\mathcal{F} \text{ is compact } \Leftrightarrow \mathcal{F} \text{ is closed, uniformly bounded and equicontinuous}$$

Proof. For a reference, see [4], Theorem 7.25 in page 158. ■

Proposition 4.3.4. *The function $I : C_0([0,1]) \to [0,+\infty]$ defined by*

$$I(\phi) := \begin{cases} \frac{1}{2}\int_0^1 |\dot{\phi}(t)|^2 dt & \text{if } \phi \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

is a good rate function.

Proof. Let $\phi \in C_0([0,1])$ and $\{\phi_n\} \subset C_0([0,1])$ such that $\lim_{n\to\infty} \phi_n = \phi$. By Proposition 1.3.4, in order to prove that I is a rate function, we have to check that

$$\liminf_{n\to\infty} I(\phi_n) \geq I(\phi).$$

Applying Lemma 4.3.2 to ϕ we have

$$\begin{aligned} I(\phi) &= \frac{1}{2} \sup_{P\in\mathcal{P}} \sum_{i=1}^{\#P-1} \frac{|\phi(t_{i+1})-\phi(t_i)|^2}{t_{i+1}-t_i} \\ &= \frac{1}{2} \sup_{P\in\mathcal{P}} \lim_{n\to\infty} \sum_{i=1}^{\#P-1} \frac{|\phi_n(t_{i+1})-\phi_n(t_i)|^2}{t_{i+1}-t_i} \\ &= \frac{1}{2} \sup_{P\in\mathcal{P}} \liminf_{n\to\infty} \sum_{i=1}^{\#P-1} \frac{|\phi_n(t_{i+1})-\phi_n(t_i)|^2}{t_{i+1}-t_i} \end{aligned}$$

$$\leq \frac{1}{2} \liminf_{n\to\infty} \sup_{P\in\mathcal{P}} \sum_{i=1}^{\#P-1} \frac{|\phi_n(t_{i+1}) - \phi_n(t_i)|^2}{t_{i+1} - t_i}$$
$$= \liminf_{n\to\infty} I(\phi_n).$$

So, we conclude that I is a rate function.

In order to prove that I is a good rate function, we are going to apply Lemma 4.3.3. Let $\alpha \in [0, \infty)$. We already know that the level set

$$\psi_I(\alpha) = \{\phi \in C_0([0,1]) : I(\phi) \leq \alpha\} = \{\phi \in H_1 : I(\phi) \leq \alpha\}$$

is closed. Let $\phi \in \psi_I(\alpha)$ and $0 \leq s \leq t \leq 1$, applying Hölder inequality

$$|\phi(t) - \phi(s)|^2 = \left| \int_s^t \dot{\phi}(u) \mathrm{d}u \right|^2 = \sum_{i=1}^{d} \left(\int_t^s \frac{\mathrm{d}\phi^{(i)}}{\mathrm{d}t}(u)\ \mathrm{d}u \right)^2$$
$$\leq \sum_{i=1}^{d} (t-s) \int_s^t \left(\frac{\mathrm{d}\phi^{(i)}}{\mathrm{d}t}(u) \right)^2\ \mathrm{d}u = (t-s) \int_s^t |\dot{\phi}(u)|^2 \mathrm{d}u$$
$$\leq 2(t-s) I(\phi) \leq 2(t-s)\alpha,$$

which implies that $\psi_I(\alpha)$ is equicontinuous. Moreover, using the previous inequality with $s = 0$ we deduce that for any $\phi \in \psi_I(\alpha)$

$$\sup_{t\in[0,1]} |\phi(t)| \leq \sqrt{2\alpha},$$

which implies that $\psi_I(\alpha)$ is uniformly bounded. Finally, by Lemma 4.3.3 we conclude that $\psi_I(\alpha)$ is a compact subset of $C_0([0,1])$. So, I is a good rate function. ■

Finally, we have all the tools to prove Schilder's Theorem. Recall that μ_ε is the law of $B_\varepsilon(\cdot)$ in $C_0([0,1])$.

The strategy to prove the lower bound of the LDP is to make a translation of the original Brownian motion with a deterministic function in such a way that our computations will be more tractable. The key point is to use Girsanov Theorem to determine a new probability such that this translation is a new Brownian motion.

On the other hand, in order to prove the upper bound of the LDP we will take a linear approximation of the Brownian motion and bound its error in such a way that we will be able to transfer the properties of the approximation to the original process.

Theorem 4.3.5. Schilder. *The family $\{\mu_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function*

$$I(\phi) := \begin{cases} \frac{1}{2}\int_0^1 |\dot{\phi}(t)|^2\, dt & \text{if } \phi \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

That is,

1. *For any open set $G \subset C_0([0,1])$,*

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(G) \geq -\inf_{\phi\in G} I(\phi). \tag{4.3.1}$$

2. *For any closed set $F \subset C_0([0,1])$,*

$$\limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(F) \leq -\inf_{\phi\in F} I(\phi). \tag{4.3.2}$$

Proof.
(1) Let $G \subset C_0([0,1])$ be a non-empty open set. If $I(\phi) = \infty$ for all $\phi \in G$ the inequality in (4.3.1) is clear. So, suppose that there exist $\phi \in G$ with $I(\phi) < \infty$ and define for $\varepsilon > 0$

$$\psi_\varepsilon := -\frac{1}{\sqrt{\varepsilon}}\dot{\phi}.$$

Note that $\psi_\varepsilon : [0,1] \to \mathbb{R}^d$ is measurable and satisfies

$$\int_0^1 |\psi_\varepsilon(s)|^2 \mathrm{d}s = \frac{1}{\varepsilon}\int_0^1 |\dot{\phi}(s)|^2 \mathrm{d}s = \frac{2}{\varepsilon} I(\phi) < \infty.$$

Hence, $\psi_\varepsilon \in L^2([0,1])$ and by Girsanov Theorem (Theorem 4.1.5), $\widetilde{B}_\varepsilon := \{\widetilde{B}_\varepsilon(t),\ t \in [0,1]\}$ defined by

$$\widetilde{B}_\varepsilon(t) = B(t) - \int_0^t \psi_\varepsilon(s)\mathrm{d}s = B(t) + \frac{1}{\sqrt{\varepsilon}}\int_0^t \dot{\phi}(s)\mathrm{d}s = B(t) + \frac{1}{\sqrt{\varepsilon}}\phi(t),$$

is a standard Brownian motion with respect to the probability $\mathrm{d}\widetilde{\mathbb{P}}_\varepsilon = Z_{\psi_\varepsilon}\mathrm{d}\mathbb{P}$, where

$$\begin{aligned} Z_{\psi_\varepsilon} &= \exp\left(\int_0^1 \psi_\varepsilon(s)\mathrm{d}B(s) - \frac{1}{2}\int_0^1 |\psi_\varepsilon(s)|^2\mathrm{d}s\right) \\ &= \exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)\right)\exp\left(-\frac{1}{\varepsilon}I(\phi)\right), \end{aligned}$$

Since G is open and $\phi \in G$, there exists $\delta > 0$ such that $B_\delta(\phi) \subset G$, where $B_\delta(\phi)$ denotes the ball of radius δ and center ϕ. Then,

$$
\begin{aligned}
\mu_\varepsilon(G) &= \mathbb{P}\left(B_\varepsilon(\cdot) \in G\right) = \mathbb{P}\left(\sqrt{\varepsilon}B(\cdot) \in G\right) = \widetilde{\mathbb{P}}_\varepsilon\left(\sqrt{\varepsilon}\widetilde{B}_\varepsilon(\cdot) \in G\right) \\
&= \widetilde{\mathbb{P}}_\varepsilon\left(\sqrt{\varepsilon}B(\cdot) + \phi \in G\right) \geq \widetilde{\mathbb{P}}_\varepsilon\left(\sqrt{\varepsilon}B(\cdot) + \phi \in B_\delta(\phi)\right) = \widetilde{\mathbb{P}}_\varepsilon\left(\left\|\sqrt{\varepsilon}B(\cdot)\right\| < \delta\right) \\
&= \widetilde{\mathbb{P}}_\varepsilon\left(\|B(\cdot)\| < \frac{\delta}{\sqrt{\varepsilon}}\right) = \widetilde{\mathbb{P}}_\varepsilon\left(\sup_{0\leq t\leq 1}|B(t)| < \frac{\delta}{\sqrt{\varepsilon}}\right) = \int_{\{\sup\limits_{0\leq t\leq 1}\sqrt{\varepsilon}\ |B(t)|<\delta\}} Z_{\psi_\varepsilon}(1)\mathrm{d}\mathbb{P} \\
&= \exp\left(-\frac{1}{\varepsilon}I(\phi)\right)\int_{\{\sup\limits_{0\leq t\leq 1}\sqrt{\varepsilon}\ |B(t)|<\delta\}} \exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)\right)\mathrm{d}\mathbb{P}.
\end{aligned}
\tag{4.3.3}
$$

On one hand, by Proposition 4.2.3 we know that

$$
\lim_{\varepsilon\to 0}\mathbb{P}\left(\sup_{0\leq t\leq 1}\sqrt{\varepsilon}\ |B(t)| < \delta\right) = 1.
$$

In particular, there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$

$$
\mathbb{P}\left(\sup_{0\leq t\leq 1}\sqrt{\varepsilon}\ |B(t)| < \delta\right) \geq \frac{2}{3}. \tag{4.3.4}
$$

On the other hand, for any $\lambda > 0$ using Markov's inequality and the isometry property of the Itô integral (Lemma 4.1.6) we have

$$
\begin{aligned}
&\mathbb{P}\left(\int_0^1 \dot{\phi}(s)\mathrm{d}B(s) \geq \lambda\sqrt{I(\phi)}\right) \leq \mathbb{P}\left(\left(\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)\right)^2 \geq \lambda^2 I(\phi)\right) \\
&\leq \frac{1}{\lambda^2 I(\phi)}E\left(\left(\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)\right)^2\right) = \frac{1}{\lambda^2 I(\phi)}\int_0^1 |\dot{\phi}(s)|^2\mathrm{d}s = \frac{2}{\lambda^2}.
\end{aligned}
\tag{4.3.5}
$$

Then, using that

$$
\mathbb{P}(A\cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cup B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1,
$$

for $A, B \in \mathcal{F}$, and the inequalities (4.3.4) and (4.3.5) with $\lambda = \sqrt{6}$ we obtain that for all $0 < \varepsilon \leq \varepsilon_0$

$$
\mathbb{P}\left(\sup_{0\leq t\leq 1}\sqrt{\varepsilon}\ |B(t)| < \delta, \frac{1}{\sqrt{\varepsilon}}\int_0^1 \dot{\phi}(s)\mathrm{d}B(s) < \frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\right) \geq \frac{2}{3} + \left(1 - \frac{2}{6}\right) - 1 = \frac{1}{3}\ . \tag{4.3.6}
$$

Going back to equation (4.3.3), we can write using the inequality in (4.3.6)

$$\int_{\{\sup_{0\le t\le 1}\sqrt{\varepsilon}\;|B(t)|<\delta\}} \exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^t \dot{\phi}(s)\mathrm{d}B(s)\right)\mathrm{d}\mathbb{P}$$
$$\ge \int_{\{\sup_{0\le t\le 1}\sqrt{\varepsilon}\;|B(t)|<\delta, \frac{1}{\sqrt{\varepsilon}}\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)<\frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\}} \exp\left(-\frac{1}{\sqrt{\varepsilon}}\int_0^t \dot{\phi}(s)\mathrm{d}B(s)\right)\mathrm{d}\mathbb{P}$$
$$> \mathbb{P}\left(\sup_{0\le t\le 1}\sqrt{\varepsilon}\;|B(t)|<\delta, \frac{1}{\sqrt{\varepsilon}}\int_0^1 \dot{\phi}(s)\mathrm{d}B(s)<\frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\right)\exp\left(-\frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\right)$$
$$\ge \frac{1}{3}\exp\left(-\frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\right).$$

Finally, by equation (4.3.3) and the previous inequality we obtain

$$\mu_\varepsilon(G) \ge \frac{1}{3}\exp\left(-\frac{1}{\varepsilon}I(\phi)\right)\exp\left(-\frac{1}{\sqrt{\varepsilon}}\sqrt{6I(\phi)}\right).$$

Therefore,

$$\varepsilon\log\mu_\varepsilon(G) \ge \varepsilon\log\frac{1}{3} - I(\phi) - \frac{\varepsilon}{\sqrt{\varepsilon}}\sqrt{6I(\phi)} \Rightarrow \liminf_{\varepsilon\to 0}\varepsilon\log\mu_\varepsilon(G) \ge -I(\phi).$$

Taking supremum over G we obtain the lower bound for the LDP of $\{\mu_\varepsilon\}$.

(2) Let $F \subset C_0([0,1])$ be a non-empty closed set.

Let $n \ge 2$ and consider the linear interpolation of step size $1/n$. That is,

$$\begin{aligned}\pi_n\colon C_0([0,1]) &\longrightarrow C_0([0,1])\\ \psi &\longmapsto \pi_n(\psi)\end{aligned}$$

defined by

$$\begin{aligned}\pi_n(\psi)\colon [0,1] &\longrightarrow \mathbb{R}^d\\ t &\longmapsto n\left[\psi\left(\frac{j+1}{n}\right)-\psi\left(\frac{j}{n}\right)\right]\left(t-\frac{j}{n}\right)+\psi\left(\frac{j}{n}\right),\ \text{if } t\in\left[\frac{j}{n},\frac{j+1}{n}\right],\end{aligned}$$

where $j = 0, ..., n-1$.

Let $\delta > 0$ and consider the closed blowup of F

$$F^\delta = \left\{\phi \in C_0([0,1]) : d(\phi, F) := \inf_{\psi\in F}\|\phi-\psi\| \le \delta\right\}.$$

Then,

$$\begin{aligned} F &= \{\phi \in F : \pi_n(\phi) \in F^\delta\} \cup \{\phi \in F : \pi_n(\phi) \notin F^\delta\} \\ &\subset \{\phi \in F : \pi_n(\phi) \in F^\delta\} \cup \{\phi \in F : \|\pi_n(\phi) - \phi\| > \delta\}. \end{aligned} \tag{4.3.7}$$

(2.1) Define $I_\delta := \inf_{\phi \in F^\delta} I(\phi)$. Then,

$$\begin{aligned} \mu_\varepsilon\left(\{\phi \in F : \pi_n(\phi) \in F^\delta\}\right) &\leq \mu_\varepsilon\left(\{\phi \in C_0([0,1]) : I(\pi_n(\phi)) \geq I_\delta\}\right) \\ &= \mathbb{P}\left(I(\pi_n(B_\varepsilon(\cdot))) \geq I_\delta\right). \end{aligned} \tag{4.3.8}$$

We can write

$$\begin{aligned} I(\pi_n(B_\varepsilon(\cdot))) &= \frac{1}{2}\int_0^1 |\dot{\pi}_n(B_\varepsilon(s))|^2 \mathrm{d}s = \frac{n}{2}\sum_{j=0}^{n-1}\left|B_\varepsilon\left(\frac{j+1}{n}\right) - B_\varepsilon\left(\frac{j}{n}\right)\right|^2 \\ &= \frac{n\varepsilon}{2}\sum_{j=0}^{n-1}\sum_{i=1}^{d}\left(B^{(i)}\left(\frac{j+1}{n}\right) - B^{(i)}\left(\frac{j}{n}\right)\right)^2 \\ &= \frac{\varepsilon}{2}\sum_{j=0}^{n-1}\sum_{i=1}^{d}\left[\sqrt{n}\left(B^{(i)}\left(\frac{j+1}{n}\right) - B^{(i)}\left(\frac{j}{n}\right)\right)\right]^2. \end{aligned}$$

Note that, the random variables

$$\sqrt{n}\left(B^{(i)}\left(\frac{j+1}{n}\right) - B^{(i)}\left(\frac{j}{n}\right)\right), \qquad 1 \leq i \leq d,\ 0 \leq j \leq n-1,$$

are independent and $N(0,1)$ distributed. Therefore,

$$\frac{2}{\varepsilon}\ I(\pi_n(B_\varepsilon(\cdot))) \overset{\mathrm{d}}{=} \chi^2_{nd}.$$

Suppose that $I_\delta < \infty$. Then, since the density of a χ^2_{nd} random variable is

$$f(x) = \frac{1}{2^{\frac{nd}{2}}\Gamma\left(\frac{nd}{2}\right)} x^{\frac{nd}{2}-1} e^{-\frac{x}{2}},$$

we have,

$$\begin{aligned} &\mathbb{P}\left(I(\pi_n(B_\varepsilon(\cdot))) \geq I_\delta\right) = \mathbb{P}\left(\frac{2}{\varepsilon} I(\pi_n(B_\varepsilon(\cdot))) \geq \frac{2}{\varepsilon} I_\delta\right) \\ &= \int_{\frac{2}{\varepsilon}I_\delta}^\infty \frac{1}{2^{\frac{nd}{2}}\Gamma\left(\frac{nd}{2}\right)} x^{\frac{nd}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x = \frac{e^{-\frac{I_\delta}{\varepsilon}}}{\Gamma\left(\frac{nd}{2}\right)} \int_0^\infty e^{-y}\left(y + \frac{I_\delta}{\varepsilon}\right)^{\frac{nd}{2}-1} \mathrm{d}y. \end{aligned} \tag{4.3.9}$$

where in the last integral we have done the change of variables $y = \frac{x}{2} - \frac{I_\delta}{\varepsilon}$. Using that $\frac{nd}{2} - 1 \geq 0$, defining $p := \left[\frac{nd}{2} - 1\right] + 2 \geq 2$ (where $[\cdot]$ is the floor function) and q such that $\frac{1}{p} + \frac{1}{q} = 1$, we have using Hölder's inequality

$$\left(y + \frac{I_\delta}{\varepsilon}\right)^{\frac{nd}{2}-1} \leq \left(y + \frac{I_\delta}{\varepsilon} + 1\right)^{\frac{nd}{2}-1} \leq \left(y + \frac{I_\delta}{\varepsilon} + 1\right)^{p} \leq 2^{\frac{p}{q}} \left[(y+1)^p + \left(\frac{I_\delta}{\varepsilon}\right)^p\right].$$

Then, applying the previous inequality and making the change of variables $z = y+1$ we get

$$\begin{aligned}&\int_0^\infty e^{-y} \left(y + \frac{I_\delta}{\varepsilon}\right)^{\frac{nd}{2}-1} \mathrm{d}y \leq 2^{\frac{p}{q}} \left[\int_0^\infty e^{-y}(y+1)^p \mathrm{d}y + \left(\frac{I_\delta}{\varepsilon}\right)^p \int_0^\infty e^{-y} \mathrm{d}y\right] \\ &\leq 2^{\frac{p}{q}} \left[e \int_1^\infty e^{-z} z^p \mathrm{d}z + \left(\frac{I_\delta}{\varepsilon}\right)^p\right] \leq 2^{\frac{p}{q}} \left[e\Gamma(p+1) + \left(\frac{I_\delta}{\varepsilon}\right)^p\right].\end{aligned}$$

This implies that

$$\limsup_{\varepsilon \to 0} \varepsilon \log \int_0^\infty e^{-y} \left(y + \frac{I_\delta}{\varepsilon}\right)^{\frac{nd}{2}-1} \mathrm{d}y \leq \lim_{\varepsilon \to 0} \varepsilon \log \left(2^{\frac{p}{q}} \left[e\Gamma(p+1) + \left(\frac{I_\delta}{\varepsilon}\right)^p\right]\right) = 0 \tag{4.3.10}$$

We conclude by (4.3.8), (4.3.9) and (4.3.10) that

$$\begin{aligned}&\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon \left(\{\phi \in F : \pi_n(\phi) \in F^\delta\}\right) \\ &\leq -I_\delta + \limsup_{\varepsilon \to 0} \varepsilon \log \frac{1}{\Gamma\left(\frac{nd}{2}\right)} + \limsup_{\varepsilon \to 0} \varepsilon \log \int_0^\infty e^{-y} \left(y + \frac{I_\delta}{\varepsilon}\right)^{\frac{nd}{2}-1} \mathrm{d}y \leq -I_\delta. \end{aligned} \tag{4.3.11}$$

Observe that the previous inequality is also true when $I_\delta = \infty$ because in this case by (4.3.8)

$$\mu_\varepsilon \left(\{\phi \in F : \pi_n(\phi) \in F^\delta\}\right) = 0.$$

(2.2) On the other hand,

$$\begin{aligned}&\mu_\varepsilon \left(\{\phi \in F : \|\pi_n(\phi) - \phi\| > \delta\}\right) \leq \mu_\varepsilon \left(\{\phi \in C_0([0,1]) : \|\pi_n(\phi) - \phi\| > \delta\}\right) \\ &= \mathbb{P}\left(\|\pi_n(B_\varepsilon(\cdot)) - B_\varepsilon(\cdot)\| \geq \delta\right) = \mathbb{P}\left(\sup_{0 \leq t \leq 1} |\pi_n(B_\varepsilon(t)) - B_\varepsilon(t)| > \delta\right) \\ &\leq \sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} |\pi_n(B_\varepsilon(t)) - B_\varepsilon(t)| > \delta\right).\end{aligned} \tag{4.3.12}$$

Note that if $\frac{j}{n} \leq t \leq \frac{j+1}{n}$, we have

$$|\pi_n(B_\varepsilon(t)) - B_\varepsilon(t)| = \left| n \left[B_\varepsilon\left(\frac{j+1}{n}\right) - B_\varepsilon\left(\frac{j}{n}\right)\right] \left(t - \frac{j}{n}\right) + B_\varepsilon\left(\frac{j}{n}\right) - B_\varepsilon(t) \right|.$$

Therefore,

$$\begin{aligned}
&\sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} |\pi_n(B_\varepsilon(t)) - B_\varepsilon(t)| \\
&\leq \sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} \left| B_\varepsilon(t) - B_\varepsilon\left(\frac{j}{n}\right)\right| + \sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} \left| n \left[B_\varepsilon\left(\frac{j+1}{n}\right) - B_\varepsilon\left(\frac{j}{n}\right)\right]\left(t - \frac{j}{n}\right)\right| \\
&\leq \sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} \left| B_\varepsilon(t) - B_\varepsilon\left(\frac{j}{n}\right)\right| + \left| B_\varepsilon\left(\frac{j+1}{n}\right) - B_\varepsilon\left(\frac{j}{n}\right)\right| \\
&\leq 2 \sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} \left| B_\varepsilon(t) - B_\varepsilon\left(\frac{j}{n}\right)\right|.
\end{aligned}$$

Going back to (4.3.12), by stationarity of the increments of Brownian motion (Lemma 4.1.7) and Lemma 4.1.3

$$\begin{aligned}
&\sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} |\pi_n(B_\varepsilon(t)) - B_\varepsilon(t)| > \delta \right) \leq \sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{\frac{j}{n} \leq t \leq \frac{j+1}{n}} \left| B_\varepsilon(t) - B_\varepsilon\left(\frac{j}{n}\right)\right| > \frac{\delta}{2}\right) \\
&= n\, \mathbb{P}\left(\sup_{0 \leq t \leq \frac{1}{n}} |B(t)| \geq \frac{\delta}{2\sqrt{\varepsilon}} \right) \leq \frac{8\sqrt{d^3 n \varepsilon}}{\sqrt{2\pi}\delta} \exp\left(-\frac{\delta^2 n}{8\varepsilon d}\right).
\end{aligned}$$

Combining (4.3.12) and the last inequality we obtain

$$\begin{aligned}
&\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon\left(\{\phi \in F : \|\pi_n(\phi) - \phi\| > \delta\}\right) \\
&\leq \limsup_{\varepsilon \to 0} \varepsilon \left[\log\left(\frac{8\sqrt{d^3 n}}{\sqrt{2\pi}\delta}\right) + \log\left(\sqrt{\varepsilon}\right) - \frac{\delta^2 n}{8\varepsilon d}\right] = -\frac{\delta^2 n}{8d}.
\end{aligned} \tag{4.3.13}$$

Finally, by (4.3.7), Lemma 1.4.2 and inequalities (4.3.11) and (4.3.13) we obtain

$$\begin{aligned}
&\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(F) \\
&\leq \limsup_{\varepsilon \to 0} \varepsilon \log \left[\mu_\varepsilon\left(\{\phi \in F : \pi_n(\phi) \in F^\delta\}\right) + \mu_\varepsilon\left(\{\phi \in F : \|\pi_n(\phi) - \phi\| > \delta\}\right)\right] \\
&\leq \max\left\{-I_\delta, -\frac{\delta^2 n}{8d}\right\} = -\min\left\{I_\delta, \frac{\delta^2 n}{8d}\right\}.
\end{aligned}$$

Let $n \to \infty$, then,

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(F) \leq -I_\delta.$$

Let $\delta \to 0$, using the fact that F is closed and Lemma 1.3.7 we obtain that

$$\limsup_{\varepsilon \to 0} \varepsilon \log \mu_\varepsilon(F) \leq \lim_{\delta \to 0} -I_\delta = -\lim_{\delta \to 0} \inf_{\phi \in F^\delta} I(\phi) = -\inf_{\phi \in F} I(\phi).$$

This finishes the proof of the upper bound of the LDP for μ_ε.

Moreover, note that $I(0(\cdot)) = 0$ and $I(\phi) > 0$ for $\phi \neq 0(\cdot)$, where $0(\cdot)$ is the constant function equal to 0. In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\mu_\varepsilon(\Gamma)\}$ for $\Gamma \in \mathcal{B}(C_0([0,1]))$ whenever $0(\cdot) \notin \overline{\Gamma}$. ■

5 LDP for SDE: The Freidlin-Wentzell Theory

In this chapter we are going to study the large deviation principle for the sample paths of strong solutions of stochastic differential equations. The idea is to perturb a SDE in such a way that its solution converges to the solution of a deterministic differential equation and study the exponential velocity of this convergence.

An interesting point is that the tools that we are going to use are Schilder's Theorem, the contraction principle and the use of exponential approximations. So, in some sense, the LDP problems that we are going to consider are extensions of the ones in Chapter 4.

First of all, we state basic results about solutions of SDE that will be used throughout this chapter and then we will consider different types of SDE, starting with a simple one and finishing with a more general case.

5.1 The setting

Like in the previous chapter, $B = \{B(t),\ t \in [0,1]\}$ will denote a standard Brownian motion in $\mathbb{R}^d$ and $\{\mathcal{F}_t,\ t \in [0,1]\}$ will be the filtration generated by B.

For the rest of this chapter, b and σ will denote Lipschitz continuous functions

$$\begin{aligned} b\colon \mathbb{R}^d &\longrightarrow \mathbb{R}^d & \qquad \sigma\colon \mathbb{R}^d &\longrightarrow \mathbb{R}^{d\times d} \\ x &\longmapsto b(x) = (b^{(i)}(x))_{1\le i\le d} & x &\longmapsto \sigma(x) = (\sigma^{(i,j)}(x))_{1\le i,j\le d} \end{aligned}$$

with linear growth. This means, that there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^d$

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \le L|x - y| \quad \text{and} \quad |b(x)| + |\sigma(x)| \le L(1 + |x|), \tag{5.1.1}$$

where $|\cdot|$ denotes both the Euclidean norm on $\mathbb{R}^d$ and on $\mathbb{R}^{d\times d}$. In fact, the linear growth of b and σ is a consequence of their Lipschitz continuity.

We have to introduce the $L^k_{a,1}$ spaces of stochastic processes, which will be used in the definition of a strong solution of a SDE. Then, we define the notions of path-wise uniqueness and strong solutions, and we state the theorem of existence and uniqueness, which will be constantly used during this chapter. Finally, we write down the very useful Gronwall's inequality.

Definition 5.1.1. Let $X = \{X(t),\ t \in [0,1]\}$ be a stochastic process taking values in $\mathbb{R}$. We write $X \in L^k_{a,1}$ if

1. X is jointly measurable in (t, ω) with respect to the product σ-field $\mathcal{B}([0,1]) \otimes \mathcal{F}$.
2. X is adapted to $\{\mathcal{F}_t,\ t \in [0,1]\}$.
3. $$E\left(\int_0^1 |X(t)|^k \mathrm{d}t\right) < \infty.$$

Definition 5.1.2. A stochastic process $X = \{X(t),\ t \in [0,1]\}$ taking values in $\mathbb{R}^d$, jointly measurable and adapted to $\{\mathcal{F}_t,\ t \in [0,1]\}$ is a strong solution to the SDE

$$\begin{aligned} \mathrm{d}X(t) &= b(X(t))\mathrm{d}t + \sigma(X(t))\mathrm{d}B(t), \\ X(0) &= x_0 \in \mathbb{R}^d, \end{aligned} \tag{5.1.2}$$

if

1. The processes $\{\sigma^{(i,j)}(X(t)),\ t \in [0,1]\}$ belong to $L^2_{a,1}$ for any $1 \leq i,j \leq d$.
2. The processes $\{b^{(i)}(X(t)),\ t \in [0,1]\}$ belong to $L^1_{a,1}$ for any $1 \leq i \leq d$.
3. For any $0 \leq t \leq 1$,
$$X_t = x_0 + \int_0^t b(X(s))\mathrm{d}s + \int_0^t \sigma(X(s))\mathrm{d}B(s), \quad \text{a.s,}$$
or coordinate-wise, for any $0 \leq t \leq 1$ and $1 \leq i \leq d$,
$$X_t^{(i)} = x_0^{(i)} + \int_0^t b^{(i)}(X(s))\mathrm{d}s + \sum_{j=1}^d \int_0^t \sigma^{(i,j)}(X(s))\mathrm{d}B^{(j)}(s), \quad \text{a.s.}$$

Definition 5.1.3. The SDE in (5.1.2) has a path-wise unique solution if any two strong solutions X_1 and X_2 of (5.1.2) are indistinguishable, that is,

$$\mathbb{P}\left(X_1(t) = X_2(t),\ \text{for any } t \in [0,1]\right) = 1.$$

Theorem 5.1.4. *Under the assumptions on b and σ given above, there exists a path-wise unique strong solution to the SDE in* (5.1.2). *The sample paths of this path-wise unique strong solution are continuous.*

Proof. For a reference, see [6], Theorem 2.5 in page 287. ■

Lemma 5.1.5. Gronwall. Let $g : [0,1] \to \mathbb{R}$ be a nonnegative, continuous function satisfying

$$g(t) \leq \alpha(t) + \beta \int_0^t g(s)\mathrm{d}s, \qquad 0 \leq t \leq 1,$$

where $\alpha : [0,1] \to \mathbb{R}$ is measurable, $\int_0^1 |\alpha(t)|\mathrm{d}t < \infty$ and $\beta \geq 0$. Then,

$$g(t) \leq \alpha(t) + \beta \int_0^t \alpha(s) e^{\beta(t-s)}\mathrm{d}s, \qquad 0 \leq t \leq 1.$$

Proof. For a reference, see [1], Lemma E.6 in page 360. ■

5.2 Case 1

In this first case we are going to consider that the diffusion matrix σ is the identity matrix and that the initial condition is the origin. The strategy will be to use Schilder's Theorem together with the contraction principle and will be a motivation for the strategy that we will use in the next case.

Fix $\varepsilon > 0$ and let X_ε be the path-wise unique strong solution of the following SDE

$$\begin{aligned} &\mathrm{d}X_\varepsilon(t) = b(X_\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}B(t), \\ &X_\varepsilon(0) = 0. \end{aligned} \tag{5.2.1}$$

Recall the space of functions introduced in Chapter 4,

$$C_0([0,1]) := \{f : [0,1] \to \mathbb{R}^d \text{ continuous and } f(0) = 0\}$$

endowed with the norm $\|f\| := \sup\limits_{0 \leq t \leq 1} |f(t)|$ and the subspace

$$H_1 := \left\{ f \in C_0([0,1]) : \ f(t) = \int_0^t g(s)\mathrm{d}s, \ g \in L^2([0,1]) \right\}$$

endowed with the norm $\|f\|_{H_1}^2 := \int_0^1 |\dot{f}(t)|^2\mathrm{d}t$.

Since the sample paths of X_ε are continuous, it makes sense to consider the map

$$\begin{aligned} X_\epsilon(\cdot)\colon \Omega &\longrightarrow C_0([0,1]) \\ \omega &\longmapsto X_\epsilon(\cdot)(\omega)\colon [0,1] \longrightarrow \mathbb{R}^d \\ &\qquad\qquad\qquad\qquad t \longmapsto X_\epsilon(t)(\omega). \end{aligned}$$

Then, following the same proof of Proposition 4.2.2 one can check that the previous map $X_\varepsilon(\cdot)$ is measurable. Hence, we can consider the law of $X_\varepsilon(\cdot)$ in $C_0([0,1])$, which we will write as $\widetilde{\mu}_\varepsilon$.

One expects that as $\varepsilon \to 0$, $X_\varepsilon(\cdot)$ converges in probability to the deterministic function X that solves the following deterministic differential equation (see Proposition 5.4.2 for the proof in a more general case)

$$\begin{aligned} \mathrm{d}X(t) &= b(X(t))\mathrm{d}t, \\ X(0) &= 0. \end{aligned} \tag{5.2.2}$$

This motivates the definition of the function F in Lemma 5.2.1, which will satisfy $\widetilde{\mu}_\varepsilon = \mu_\varepsilon \circ F^{-1}$, where μ_ε is the law of

$$B_\varepsilon := \{B_\varepsilon(t) := \sqrt{\varepsilon}B(t),\ t \in [0,1]\},$$

in $C_0([0,1])$. Then, the proof of the LDP for $\{\widetilde{\mu}_\varepsilon\}$ will be an application of Schilder's Theorem and the contraction principle.

Lemma 5.2.1. Let $F : C_0([0,1]) \to C_0([0,1])$ be defined in the following way: for $g \in C_0([0,1])$, $f = F(g)$ is the unique solution of

$$f(t) = \int_0^t b(f(s))\mathrm{d}s + g(t), \qquad 0 \le t \le 1. \tag{5.2.3}$$

Then, F is well defined and continuous in the supremum norm.

Proof. The existence and uniqueness of the solution of (5.2.3) is standard and it is a consequence of the Lipschitz continuity of b. For a reference, see [7], Theorem 3.1.2 in page 105.

Let $g_1, g_2 \in C_0([0,1])$ and write $f_1 = F(g_1)$, $f_2 = F(g_2)$ and $e(t) := |f_1(t) - f_2(t)|$. Then,

$$e(t) \le \int_0^t |b(f_1(s)) - b(f_2(s))|\mathrm{d}s + |g_1(t) - g_2(t)| \le L\int_0^t e(s)\mathrm{d}s + \|g_1 - g_2\|$$

Applying Gronwall's Lemma (Lemma 5.1.5) we obtain that

$$e(t) \le \|g_1 - g_2\| + L\int_0^t \|g_1 - g_2\|\, e^{L(t-s)}\mathrm{d}s = \|g_1 - g_2\|\, e^{Lt}$$

This implies that

$$\|F(g_1) - F(g_2)\| = \|f_1 - f_2\| = \sup_{0\le t\le 1} e(t) \le \|g_1 - g_2\|\, e^{L},$$

which implies that F is continuous. ■

Lemma 5.2.2. $\widetilde{\mu}_\varepsilon = \mu_\varepsilon \circ F^{-1}$, where $\widetilde{\mu}_\varepsilon$ is the law of $X_\varepsilon(\cdot)$, μ_ε is the law of $B_\varepsilon(\cdot)$ and F is the function defined in Lemma 5.2.1.

Proof. Observe that

$$\begin{aligned} F(B_\varepsilon)(t) &= \int_0^t b(F(B_\varepsilon)(s))\mathrm{d}s + B_\varepsilon(t), \\ F(B_\varepsilon)(0) &= 0, \end{aligned}$$

so, both X_ε and $F(B_\varepsilon)$ are strong solutions of the SDE in (5.2.1). We conclude that X_ε and $F(B_\varepsilon)$ are indistinguishable. Then, for $G \in \mathcal{B}(C_0([0,1]))$ we have

$$\widetilde{\mu}_\varepsilon(G) = \mathbb{P}(X_\varepsilon(\cdot) \in G) = \mathbb{P}(F(B_\varepsilon(\cdot)) \in G) = \mathbb{P}(B_\varepsilon(\cdot) \in F^{-1}(G)) = \left(\mu_\varepsilon \circ F^{-1}\right)(G).$$

■

Theorem 5.2.3. *The family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function*

$$I(f) := \begin{cases} \frac{1}{2}\int_0^1 |\dot{f}(t) - b(f(t))|^2 dt & \text{if } f \in H_1 \\ \infty & \text{otherwise.} \end{cases} \tag{5.2.4}$$

Proof. By Schilder Theorem (Theorem 4.3.5), the contraction principle (Theorem 3.4.1) and Lemmas 5.2.1 and 5.2.2 we conclude that the family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function

$$I(f) = \inf_{\{g \in H_1 :\ f = F(g)\}} \frac{1}{2}\int_0^1 |\dot{g}(t)|^2 \mathrm{d}t = \inf_{\{g \in H_1 :\ f(t) = \int_0^t b(f(s))\mathrm{d}s + g(t)\}} \frac{1}{2}\int_0^1 |\dot{g}(t)|^2 \mathrm{d}t.$$

In order to identify I with the function in (5.2.4) note that if there exists $g \in H_1$ satisfying

$$f(t) = \int_0^t b(f(s))\mathrm{d}s + g(t), \tag{5.2.5}$$

then $f \in H_1$ because

$$\dot{f}(t) = b(f(t)) + \dot{g}(t) \Rightarrow |\dot{f}(t)| \le |b(f(t))| + |\dot{g}(t)| \le L(1 + |f(t)|) + |\dot{g}(t)|,$$

Therefore, if $f \notin H_1$, there is no $g \in H_1$ satisfying (5.2.5) and we conclude that $I(f) = \infty$.

On the other hand, if $f \in H_1$ then there exists a unique $g \in H_1$ satisfying (5.2.5). This is because g is defined by

$$g(t) = f(t) - \int_0^t b(f(s))\mathrm{d}s$$

which implies that $g \in H_1$ because

$$\dot{g}(t) = \dot{f}(t) - b(f(t)) \Rightarrow |\dot{g}(t)| \leq |\dot{f}(t)| + |b(f(t))| \leq |\dot{f}(t)| + L(1 + |f(t)|).$$

This finishes the proof of the Theorem. Moreover, note that $I(X) = 0$ and $I(f) > 0$ for $f \neq X$ where X is the deterministic function that solves (5.2.2). In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\widetilde{\mu}_\varepsilon(\Gamma)\}$ for $\Gamma \in \mathcal{B}(C_0([0,1]))$ whenever $X \notin \overline{\Gamma}$. ■

5.3 Case 2

Now we are going to consider a general diffusion matrix σ with the initial condition still being the origin. In addition to the conditions of b and σ described in (5.1.1) we are going to assume that they are bounded. That is,

$$\sup_{x \in \mathbb{R}^d} (|b(x)| + |\sigma(x)|) \leq L,$$

where $L > 0$ is the same constant as in the Lipschitz and linear growth condition.

Fix $\varepsilon > 0$ and let X_ε be the path-wise unique strong solution of the following SDE

$$\begin{aligned} \mathrm{d}X_\varepsilon(t) &= b(X_\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_\varepsilon(t))\mathrm{d}B(t), \\ X_\varepsilon(0) &= 0. \end{aligned} \tag{5.3.1}$$

As in the previous case, we will denote by $\widetilde{\mu}_\varepsilon$ the law of $X_\varepsilon(\cdot)$ in $C_0([0,1])$. By Proposition 5.4.2 below, $X_\varepsilon(\cdot)$ converges in probability to the deterministic function X that solves the deterministic differential equation

$$\begin{aligned} \mathrm{d}X(t) &= b(X(t))\mathrm{d}t, \\ X(0) &= 0. \end{aligned} \tag{5.3.2}$$

In order to prove the LDP in this more general case, we could try to follow the same procedure as before, that is, to construct a continuous map $F : C_0([0,1]) \to C_0([0,1])$ such that $\widetilde{\mu}_\varepsilon = \mu_\varepsilon \circ F^{-1}$ and then apply the contraction principle. However, in this case, it is known that such F need not be continuous due to the called "Wong-Zakai correction" term. For a reference, see [9].

The new strategy will be to construct exponentially good approximations of $X_\varepsilon(\cdot)$, say $X_{\varepsilon,m}(\cdot)$ with law $\mu_{\varepsilon,m}$ in $C_0([0,1])$, and continuous maps $F_m : C_0([0,1]) \to C_0([0,1])$ such that $\mu_{\varepsilon,m} = \mu_\varepsilon \circ F_m^{-1}$ and such that they approximate well enough a map F to be defined. Then the LDP will be an application of Theorem 3.4.8.

Before constructing such exponentially good approximations we state some preliminaries results about stopping times and SDE that we will need.

5.3.1 Preliminaries for Case 2

Lemma 5.3.1. Let $Y = \{Y(t), t \in [0,1]\}$ be a stochastic process taking values in $\mathbb{R}^d$, adapted to $\{\mathcal{F}_t,\ t \in [0,1]\}$ with continuous sample paths and $C \subset \mathbb{R}^d$ a closed subset. Then the hitting time

$$\tau := \inf\{t \in [0,1] :\ Y(t) \in C\} \wedge 1,$$

is a stopping time.

Proof. For a reference, see [6], Problem 2.7 in page 7. ■

Definition 5.3.2. Let $Y = \{Y(t),\ t \in [0,1]\}$ be a stochastic process taking values in a measurable space $(S, \mathcal{A})$. The process Y is said to be progressively measurable if, for every $t \in [0,1]$, the map

$$\begin{aligned} [0,t] \times \Omega &\longrightarrow S \\ (s,\omega) &\longmapsto Y(s)(\omega) \end{aligned}$$

is jointly measurable in (s,ω) with respect to the product σ-field $\mathcal{B}([0,t]) \otimes \mathcal{F}$.

Lemma 5.3.3. Let $Y = \{Y(t),\ t \in [0,1]\}$ be an adapted process to $\{\mathcal{F}_t,\ t \in [0,1]\}$ taking values in $\mathbb{R}^d$ and with continuous sample paths. Then, Y is progressively measurable.

Proof. For a reference, see [6], Proposition 1.13 in page 5. ■

Lemma 5.3.4. Let $Z = \{Z(t),\ t \in [0,1]\}$ be a stochastic process taking values in $\mathbb{R}^d$ such that

$$\begin{aligned} \mathrm{d}Z(t) &= \Theta(t)\mathrm{d}t + \sqrt{\varepsilon}\Sigma(t)\mathrm{d}B(t), \\ Z(0) &= z_0 \in \mathbb{R}^d, \end{aligned}$$

where $\Theta = \{\Theta(t),\ t \in [0,1]\}$ and $\Sigma = \{\Sigma(t),\ t \in [0,1]\}$ are progressively measurable processes taking values in $\mathbb{R}^d$ and $\mathbb{R}^{d\times d}$, respectively, and let τ_1 be a stopping time.

Suppose that the for some positive constants N_Σ, M_Σ, M_Θ, ρ we have

$$\sup_{0\le t\le 1} |\Sigma(t)| \le N_\Sigma$$

and for any $t \in [0, \tau_1]$,

$$|\Sigma(t)| \le M_\Sigma(\rho^2 + |Z(t)|^2)^{1/2} \qquad \text{and} \qquad |\Theta(t)| \le M_\Theta(\rho^2 + |Z(t)|^2)^{1/2}.$$

Then, for any $\delta > 0$ and any $\varepsilon \leq 1$,

$$\varepsilon \log \mathbb{P}\left(\sup_{t\in[0,\tau_1]} |Z(t)| \geq \delta\right) \leq K + \log\left(\frac{\rho^2 + |Z(0)|^2}{\rho^2 + \delta^2}\right),$$

where $K = 2M_\Theta + M_\Sigma^2(2+d)$.

Proof. For a reference, see [1], Lemma 5.6.18 in page 217. ■

5.3.2 Proof of Case 2

For $m \geq 1$, define $\pi_m(t) := \frac{[mt]}{m}$ and let $X_{\varepsilon,m}$ be the path-wise unique strong solution of

$$\begin{aligned} \mathrm{d}X_{\varepsilon,m}(t) &= b\left(X_{\varepsilon,m}\left(\pi_m(t)\right)\right)\mathrm{d}t + \sqrt{\varepsilon}\sigma\left(X_{\varepsilon,m}\left(\pi_m(t)\right)\right)\mathrm{d}B(t), \\ X_{\varepsilon,m}(0) &= 0. \end{aligned} \tag{5.3.3}$$

Since $t \in \left[\frac{k}{m}, \frac{k+1}{m}\right)$ implies that $\pi_m(t) = \frac{k}{m}$, the previous SDE is similar to the SDE in (5.3.1) but freezing b and σ over the time intervals $[\frac{k}{m}, \frac{k+1}{m})$. In the next Lemma we prove that $\{X_{\varepsilon,m}(\cdot)\}$ are exponentially good approximations of $\{X_\varepsilon(\cdot)\}$.

Lemma 5.3.5. Let $\delta > 0$. Then

$$\lim_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\|X_{\varepsilon,m}(\cdot) - X_\varepsilon(\cdot)\| > \delta) = -\infty. \tag{5.3.4}$$

Proof. For $\rho > 0$, define

$$\tau_1 := \inf\{t \in [0,1] : \ |X_{\varepsilon,m}(t) - X_{\varepsilon,m}(\pi_m(t))| \geq \rho\} \wedge 1.$$

By Lemma 5.3.1, τ_1 is a stopping time and we have that

$$\begin{aligned} &\{\|X_{\varepsilon,m}(\cdot) - X_\varepsilon(\cdot)\| > \delta\} = \{\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| > \delta\} \\ &= \left(\{\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| > \delta\} \cap \{\tau_1 = 1\}\right) \\ &\quad \bigcup \left(\{\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| > \delta\} \cap \{\tau_1 < 1\}\right) \\ &\subset \{\sup_{0\leq t\leq \tau_1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| > \delta\} \cup \{\tau_1 < 1\} \\ &\subset \{\sup_{0\leq t\leq \tau_1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| \geq \delta\} \cup \{\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right)| \geq \rho\} \end{aligned}$$

Then, applying (1.4.2) of Lemma 1.4.9 and the previous inclusions we obtain

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}(\|X_{\varepsilon,m}(\cdot) - X_\varepsilon(\cdot)\| > \delta) \\
&\leq \max\left[\limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}\left(\sup_{0\leq t\leq \tau_1} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| \geq \delta\right)\right. \\
&\qquad \left., \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}\left(\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_{\varepsilon,m}(\pi_m(t))| \geq \rho\right)\right] \qquad (5.3.5)
\end{aligned}$$

(1) For the first limit inside the maximum of (5.3.5) we will apply Lemma 5.3.4 with

$$Z := \{Z(t) := X_{\varepsilon,m}(t) - X_\varepsilon(t),\ 0 \leq t \leq 1\}.$$

Note that

$$\begin{aligned}
\mathrm{d}Z(t) &= \Theta(t)\mathrm{d}t + \sqrt{\varepsilon}\Sigma(t)\mathrm{d}B(t), \\
Z(0) &= z_0 = 0,
\end{aligned}$$

with

$$\begin{aligned}
\Theta(t) &:= b\left(X_{\varepsilon,m}\left(\pi_m(t)\right)\right) - b(X_\varepsilon(t)), \\
\Sigma(t) &:= \sigma\left(X_{\varepsilon,m}\left(\pi_m(t)\right)\right) - \sigma(X_\varepsilon(t)).
\end{aligned}$$

By Lipschitz continuity of b and σ, the fact that the processes X_ε and $X_{\varepsilon,m}$ are adapted to $\{\mathcal{F}_t,\ t \in [0,1]\}$ and have continuous sample paths, we deduce that the processes Θ and Σ are progressively measurable applying Lemma 5.3.3.

Note that

$$\sup_{0\leq t\leq 1} |\Sigma(t)| \leq 2L,$$

and for $t \in [0, \tau_1]$,

$$\begin{aligned}
|\Sigma(t)| &\leq L\left|X_{\varepsilon,m}\left(\pi_m(t)\right) - X_\varepsilon(t)\right| \leq L\left(|X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right)| + |Z(t)|\right) \\
&\leq L(\rho + |Z(t)|) \leq \sqrt{2}L\left(\rho^2 + |Z(t)|^2\right)^{1/2}.
\end{aligned}$$

The same argument works for $|\Theta(t)|$. We conclude that we can apply Lemma 5.3.4 with $N_\Sigma = 2L$, $M_\Sigma = \sqrt{2}L$ and $M_\Theta = \sqrt{2}L$. So, for any $\delta > 0$ and any $\varepsilon \leq 1$,

$$\varepsilon \log \mathbb{P}\left(\sup_{t\in[0,\tau_1]} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| \geq \delta\right) \leq K + \log\left(\frac{\rho^2}{\rho^2 + \delta^2}\right),$$

with $K = 2L(\sqrt{2} + L(2+d))$. Then,

$$\lim_{\rho\to 0} \limsup_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}\left(\sup_{t\in[0,\tau_1]} |X_{\varepsilon,m}(t) - X_\varepsilon(t)| \geq \delta \right) = -\infty. \tag{5.3.6}$$

(2) For the second limit inside the maximum of (5.3.5), note that since $X_{\varepsilon,m}$ satisfy the SDE in (5.3.3), we can write

$$X_{\varepsilon,m}(t) = \int_0^t b\left(X_{\varepsilon,m}\left(\pi_m(s)\right)\right) \mathrm{d}s + \int_0^t \sqrt{\varepsilon}\sigma\left(X_{\varepsilon,m}\left(\pi_m(s)\right)\right) \mathrm{d}B(s).$$

Then, if $t \in \left[\frac{k}{m}, \frac{k+1}{m}\right)$, $\pi_m(t) = \frac{k}{m}$ and since b and σ are bounded by L, we have

$$\begin{aligned} X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right) &= \int_{\frac{k}{m}}^t b\left(X_{\varepsilon,m}\left(\frac{k}{m}\right)\right) \mathrm{d}s + \int_{\frac{k}{m}}^t \sqrt{\varepsilon}\sigma\left(X_{\varepsilon,m}\left(\frac{k}{m}\right)\right) \mathrm{d}s \\ &= b\left(X_{\varepsilon,m}\left(\frac{k}{m}\right)\right)\left(t - \frac{k}{m}\right) + \sqrt{\varepsilon}\sigma\left(X_{\varepsilon,m}\left(\frac{k}{m}\right)\right)\left(B(t) - B\left(\frac{k}{m}\right)\right). \end{aligned}$$

Hence, for any $t \in \left[\frac{k}{m}, \frac{k+1}{m}\right)$,

$$|X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right)| \leq L\left[\frac{1}{m} + \sqrt{\varepsilon} \sup_{\frac{k}{m}\leq s\leq \frac{k+1}{m}} \left|B(s) - B\left(\frac{k}{m}\right)\right|\right].$$

Finally, we can write

$$\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right)| \leq L\left[\frac{1}{m} + \sqrt{\varepsilon} \max_{0\leq k\leq m-1} \sup_{0\leq s\leq \frac{1}{m}} \left|B\left(s + \frac{k}{m}\right) - B\left(\frac{k}{m}\right)\right|\right].$$

Then, for $m > \frac{L}{\rho}$,

$$\begin{aligned} &\mathbb{P}\left(\sup_{0\leq t\leq 1} |X_{\varepsilon,m}(t) - X_{\varepsilon,m}\left(\pi_m(t)\right)| > \rho\right) \\ &\leq \mathbb{P}\left(\max_{0\leq k\leq m-1} \sup_{0\leq s\leq \frac{1}{m}} \left|B\left(s + \frac{k}{m}\right) - B\left(\frac{k}{m}\right)\right| \geq \frac{\rho - L/m}{\sqrt{\varepsilon}L}\right) \\ &\overset{(1)}{\leq} \sum_{k=0}^{m-1} \mathbb{P}\left(\sup_{0\leq s\leq \frac{1}{m}} |B(s)| \geq \frac{\rho - L/m}{\sqrt{\varepsilon}L}\right) \\ &\overset{(2)}{\leq} m\frac{4\sqrt{\varepsilon d^3}L}{\sqrt{2\pi m}(\rho - L/m)} \exp\left(\frac{-m(\rho - L/m)^2}{2d\varepsilon L^2}\right). \end{aligned}$$

We have used

1. Stationary increments of B. (Lemma 4.1.7).
2. Lemma 4.1.3.

The previous equation implies that

$$\lim_{m\to\infty}\limsup_{\varepsilon\to 0}\varepsilon\log\mathbb{P}\left(\sup_{0\leq t\leq 1}|X_{\varepsilon,m}(t)-X_{\varepsilon,m}(\pi_m(t))|\geq\rho\right)=-\infty. \tag{5.3.7}$$

Finally, by (5.3.5), (5.3.6) and (5.3.7) we have proved (5.3.4). ■

Having the exponential good approximations of $X_\varepsilon(\cdot)$, we now define continuous maps $F_m : C_0([0,1]) \to C_0([0,1])$ such that $\mu_{\varepsilon,m} = \mu_\varepsilon \circ F_m^{-1}$, where $\mu_{\varepsilon,m}$ is the law of $X_{\varepsilon,m}(\cdot)$ in $C_0([0,1])$, and such that they approximate well enough a map F to be defined. We prove this results in the three following Lemmas.

Lemma 5.3.6. Let $F_m : C_0([0,1]) \to C_0([0,1])$ be defined in the following way: for $g \in C_0([0,1])$, $h = F_m(g)$ where

$$\begin{aligned} h(t) =& h\left(\frac{k}{m}\right)+b\left(h\left(\frac{k}{m}\right)\right)\left(t-\frac{k}{m}\right)+\sigma\left(h\left(\frac{k}{m}\right)\right)\left(g(t)-g\left(\frac{k}{m}\right)\right), \\ & t\in\left[\frac{k}{m},\frac{k+1}{m}\right],\ k=0,...,m-1,\ h(0)=0. \end{aligned} \tag{5.3.8}$$

Then, F_m is well defined and continuous in the supremum norm.

Proof. To see that F_m is well defined observe that for $t \in \left[0, \frac{1}{m}\right]$,

$$h(t) = b(0)t + \sigma(0)g(t).$$

Reasoning inductively on k we see that F_m is well defined and that $F_m(g) \in C_0([0,1])$.

Consider $g_1, g_2 \in C_0([0,1])$ and write $f_1 = F_m(g_1)$, $f_2 = F_m(g_2)$ and $e(t) := |f_1(t) - f_2(t)|$. Then, if $t \in \left[\frac{k}{m}, \frac{k+1}{m}\right]$,

$$\begin{aligned} e(t) \leq & \left|f_1\left(\frac{k}{m}\right)-f_2\left(\frac{k}{m}\right)\right|+\left|b\left(f_1\left(\frac{k}{m}\right)\right)-b\left(f_2\left(\frac{k}{m}\right)\right)\right|\left|t-\frac{k}{m}\right| \\ & +\left|\sigma\left(f_1\left(\frac{k}{m}\right)\right)\left(g_1(t)-g_1\left(\frac{k}{m}\right)\right)-\sigma\left(f_2\left(\frac{k}{m}\right)\right)\left(g_2(t)-g_2\left(\frac{k}{m}\right)\right)\right| \\ \leq & e\left(\frac{k}{m}\right)+\frac{L}{m}e\left(\frac{k}{m}\right)+\left|\left[\sigma\left(f_1\left(\frac{k}{m}\right)\right)-\sigma\left(f_2\left(\frac{k}{m}\right)\right)\right]\left(g_1(t)-g_1\left(\frac{k}{m}\right)\right)\right| \end{aligned}$$

$$+\left|\sigma\left(f_2\left(\frac{k}{m}\right)\right)\left(g_2(t)-g_2\left(\frac{k}{m}\right)-g_1(t)+g_1\left(\frac{k}{m}\right)\right)\right|$$
$$\leq e\left(\frac{k}{m}\right)+\frac{L}{m}e\left(\frac{k}{m}\right)+2Le\left(\frac{k}{m}\right)\|g_1\|+2L\|g_1-g_2\|.$$

We conclude that

$$\sup_{\frac{k}{m}\leq t\leq\frac{k+1}{m}} e(t)\leq\left(1+\frac{L}{m}+2L\|g_1\|\right)e\left(\frac{k}{m}\right)+2L\|g_1-g_2\|$$

Since $e(0)=0$, iterating the previous bound over $k=0,...,m-1$ we obtain that F_m is continuous. ■

Lemma 5.3.7. $F_m\circ B_\varepsilon(\cdot)=X_{\varepsilon,m}(\cdot)$ a.s, where F_m the function defined in Lemma 5.3.6.

Proof. Observe that if $t\in\left[\frac{k}{m},\frac{k+1}{m}\right)$,

$$F_m(B_\varepsilon)(t)-F_m(B_\varepsilon)\left(\frac{k}{m}\right)$$
$$=b\left(F_m(B_\varepsilon)\left(\frac{k}{m}\right)\right)\left(t-\frac{k}{m}\right)+\sigma\left(F_m(B_\varepsilon)\left(\frac{k}{m}\right)\right)\left(B_\varepsilon(t)-B_\varepsilon\left(\frac{k}{m}\right)\right)$$

Since $\frac{k}{m}=\pi_m(t)$ and $F_m(B_\varepsilon(0))=0$, we see that both $X_{\varepsilon,m}$ and $F_m(B_\varepsilon(t))$ are strong solutions of the same SDE in (5.3.3). We conclude that they are indistinguishable and that $F_m\circ B_\varepsilon(\cdot)=X_{\varepsilon,m}(\cdot)$ a.s. ■

Lemma 5.3.8. Let $F:H_1\to C_0([0,1])$ be defined in the following way: for $g\in H_1$, $f=F(g)$ is the unique solution of

$$f(t)=\int_0^t b(f(s))\mathrm{d}s+\int_0^t\sigma(f(s))\dot{g}(s)\mathrm{d}s,\qquad 0\leq t\leq 1. \tag{5.3.9}$$

Then, F is well defined and for every $\alpha<\infty$

$$\lim_{m\to\infty}\sup_{\{g\in H_1:\ \|g\|_{H_1}\leq\alpha\}}\|F_m(g)-F(g)\|=0.$$

In particular, F is continuous in H_1 in the supremum norm.

Proof. The existence and uniqueness of the solution (5.3.9) is standard and it is a consequence of the Lipschitz continuity of b and σ. For a reference, see [7], Theorem 3.1.2 in page 105.

Let $\alpha < \infty$ and consider $g \in H_1$ such that $\|g\|_{H_1} \leq \alpha$. Write $h = F_m(g)$, $f = F(g)$ and $e(t) = |f(t) - h(t)|^2$. By the definition of F_m in (5.3.8) we can write

$$h(t) = \int_0^t b\left(h\left(\pi_m(s)\right)\right) \mathrm{d}s + \int_0^t \sigma\left(h\left(\pi_m(s)\right)\right) \dot{g}(s) \mathrm{d}s.$$

If $t \in \left[0, \frac{1}{m}\right]$, the previous equality is clear. By induction, if $t \in \left[\frac{k}{m}, \frac{k+1}{m}\right]$, then

$$\begin{aligned} h(t) = & \int_0^{\frac{k}{m}} b\left(h\left(\pi_m(s)\right)\right) \mathrm{d}s + \int_0^{\frac{k}{m}} \sigma\left(h\left(\pi_m(s)\right)\right) \dot{g}(s) \mathrm{d}s \\ & + \int_{\frac{k}{m}}^t b\left(h\left(\pi_m(s)\right)\right) \mathrm{d}s + \int_{\frac{k}{m}}^t \sigma\left(h\left(\pi_m(s)\right)\right) \dot{g}(s) \mathrm{d}s \\ = & h\left(\frac{k}{m}\right) + b\left(h\left(\frac{k}{m}\right)\right)\left(t - \frac{k}{m}\right) + \sigma\left(h\left(\frac{k}{m}\right)\right)\left(g(t) - g\left(\frac{k}{m}\right)\right), \end{aligned}$$

and we recover the definition of F_m. Applying Hölder inequality

$$\begin{aligned} & \left|h(t) - h\left(\pi_m(t)\right)\right| \leq \int_{\pi_m(t)}^t \left|b\left(h\left(\pi_m(s)\right)\right)\right| \mathrm{d}s + \int_{\pi_m(t)}^t \left|\sigma\left(h\left(\pi_m(s)\right)\right)\right| \left|\dot{g}(s)\right| \mathrm{d}s \\ & \leq L\left|t - \pi_m(t)\right| + \left(\int_{\pi_m(t)}^t \left|\sigma\left(h\left(\pi_m(s)\right)\right)\right|^2 \mathrm{d}s\right)^{1/2} \left(\int_{\pi_m(t)}^t \left|\dot{g}(s)\right|^2 \mathrm{d}s\right)^{1/2} \\ & \leq \frac{L}{m} + \frac{L}{\sqrt{m}}\alpha \leq (\alpha+1)\frac{L}{\sqrt{m}}. \end{aligned} \tag{5.3.10}$$

Applying again Hölder inequality,

$$\begin{aligned} \sqrt{e(t)} & = |f(t) - h(t)| \\ & \leq \int_0^t \left|b(f(s)) - b\left(h\left(\pi_m(s)\right)\right)\right| \mathrm{d}s + \int_0^t \left|\sigma(f(s)) - \sigma\left(h\left(\pi_m(s)\right)\right)\right| \left|\dot{g}(s)\right| \mathrm{d}s \\ & \leq L\left(\int_0^t \left|f(s) - h\left(\pi_m(s)\right)\right|^2 \mathrm{d}s\right)^{1/2} \left(\int_0^t 1^2 \mathrm{d}s\right)^{1/2} \\ & + L\left(\int_0^t \left|f(s) - h\left(\pi_m(s)\right)\right|^2 \mathrm{d}s\right)^{1/2} \left(\int_0^t \left|\dot{g}(s)\right|^2 \mathrm{d}s\right)^{1/2} \\ & \leq (\alpha+1)L\left(\int_0^t \left|f(s) - h\left(\pi_m(s)\right)\right|^2 \mathrm{d}s\right)^{1/2}. \end{aligned} \tag{5.3.11}$$

And finally by (5.3.11) and (5.3.10) we have,

$$e(t) \leq 2(\alpha+1)^2 L^2 \left[\int_0^t |f(s) - h(s)|^2 \mathrm{d}s + \int_0^t \left|h(s) - h\left(\pi_m(s)\right)\right|^2 \mathrm{d}s\right]$$

$$\leq 2(\alpha+1)^2 L^2 \left[\int_0^t e(s)\mathrm{d}s + (\alpha+1)^2 \frac{L^2}{m}\right] \leq K \int_0^t e(s)\mathrm{d}s + \frac{K}{m},$$

with $K := \max\{2(\alpha+1)^2L^2, 2(\alpha+1)^4L^4\}$. By Gronwall's Lemma (Lemma 5.1.5), we obtain

$$e(t) \leq \frac{K}{m} + K\int_0^t \frac{K}{m} e^{K(t-s)}\mathrm{d}s = \frac{K}{m} + K\frac{K}{m}\frac{e^{Kt}-1}{K} = \frac{K}{m}e^{Kt}.$$

This implies that

$$\|F_m(g) - F(g)\| = \sup_{0\leq t\leq 1} \sqrt{e(t)} \leq \frac{\sqrt{K}}{\sqrt{m}} e^{K/2}.$$

Since K does not depend on m, we conclude that

$$\lim_{m\to\infty} \sup_{\{g\in H_1:\ \|g\|_{H_1}\leq\alpha\}} \|F_m(g) - F(g)\| = 0. \tag{5.3.12}$$

To see that F is continuous consider $g_1, g_2 \in H_1$ and $\alpha < \infty$ such that

$$\max\{\|g_1\|_{H_1}, \|g_2\|_{H_1}\} \leq \alpha.$$

Then,

$$\|F(g_1) - F(g_2)\| \leq \|F(g_1) - F_m(g_1)\| + \|F_m(g_1) - F_m(g_2)\| + \|F_m(g_2) - F(g_2)\|$$

By (5.3.12) and the continuity of F_m (Lemma 5.3.6) we conclude that $F : H_1 \to C_0([0,1])$ is continuous in the supremum norm. ■

We finally have all the tools to prove the LDP for $\{\widetilde{\mu}_\varepsilon\}$.

Theorem 5.3.9. *The family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ an LDP with the good rate function*

$$I(f) = \inf_{\{g\in H_1:\ f(t)=\int_0^t b(f(s))ds + \int_0^t \sigma(f(s))\dot{g}(s)ds\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 ds.$$

Proof. We are going to apply Theorem 3.4.8 with $\mathcal{X} = \mathcal{Y} = C_0([0,1])$, $Z_\varepsilon = B_\varepsilon(\cdot)$, $\widetilde{Z}_\varepsilon = X_\varepsilon(\cdot)$, F_m the functions defined in Lemma 5.3.6 and F the function defined in Lemma 5.3.8.

By Schilder's Theorem (Theorem 4.3.5) we know that the family $\{\mu_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function

$$I_2(f) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{f}(t)|^2\mathrm{d}t = \frac{1}{2}\|f\|_{H_1} & \text{if } f \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

By Lemma 5.3.6 the functions F_m are continuous and by Lemma 5.3.8 $F : H_1 \to C_0([0,1])$ is a continuous function satisfying for every $\alpha < \infty$

$$\lim_{m\to\infty} \sup_{\{g\in C_0([0,1]):\ I_2(g)\le\alpha\}} \|F_m(g) - F(g)\| = \lim_{m\to\infty} \sup_{\{g\in H_1:\ \|g\|_{H_1}\le 2\alpha\}} \|F_m(g) - F(g)\| = 0.$$

Note that we can extend $F : C_0([0,1]) \to C_0([0,1])$ in such a way that is measurable. For example, for $g \in C_0([0,1]) \setminus H_1$ just define $F(g) = 0(\cdot)$ where $0(\cdot)$ is the constant function equal to 0. Then, the measurability of F follows from the fact that H_1 is closed in $C_0([0,1])$, $F|_{H_1}$ is continuous and $F|_{H_1^c}$ is constant.

By Lemma 5.3.5, we have that $\{X_{\varepsilon,m}(\cdot)\}$ are exponentially good approximations of $\{X_\varepsilon(\cdot)\}$. So, by Lemma 5.3.7, $\{F_m \circ B_\varepsilon(\cdot)\} = \{X_{\varepsilon,m}(\cdot)\}$ are exponentially good approximations of $\{X_\varepsilon(\cdot)\}$.

In summary, we have all the conditions of Theorem 3.4.8. We conclude that the family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function

$$I(f) = \inf_{\{g\in C_0([0,1]):\ f=F(g)\}} I_2(g) = \inf_{\{g\in H_1:\ f(t)=\int_0^t b(f(s))\mathrm{d}s+\int_0^t \sigma(f(s))\dot{g}(s)\mathrm{d}s\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 \mathrm{d}s.$$

Moreover, note that $I(X) = 0$, where X is the deterministic function that solves (5.3.2), because then $g = 0(\cdot) \in H_1$ satisfies the condition in the infimum of I. In addition, $I(f) > 0$ for $f \neq X$ because then a function $g \in H_1$ that satisfies the condition in the infimum of I can not be constant equal to 0. In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\widetilde{\mu}_\varepsilon(\Gamma)\}$ for $\Gamma \in \mathcal{B}(C_0([0,1]))$ whenever $X \notin \overline{\Gamma}$. ■

5.4 Case 3

Now we are going to consider that b and σ satisfy the same conditions as in Case 2, but the initial condition need not be the origin. After doing a translation of coordinates we can prove this LDP using the results of the previous one, so the hard work has been done.

Fix $\varepsilon > 0$ and let X_ε be the path-wise unique strong solution of the following SDE

$$\begin{aligned} \mathrm{d}X_\varepsilon(t) &= b(X_\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_\varepsilon(t))\mathrm{d}B(t), \\ X_\varepsilon(0) &= x_0 \in \mathbb{R}^d. \end{aligned} \tag{5.4.1}$$

We will denote by $\widetilde{\mu}_\varepsilon$ the law of $X_\varepsilon(\cdot)$ in $C([0,1])$ where

$$C([0,1]) := \{f : [0,1] \to \mathbb{R}^d \text{ continuous}\}$$

with the standard supremum norm. We prove here the fact that $X_\varepsilon(\cdot)$ converge in probability to the deterministic function X that solves the deterministic differential equation

$$\begin{aligned} \mathrm{d}X(t) &= b(X(t))\mathrm{d}t, \\ X(0) &= x_0 \in \mathbb{R}^d. \end{aligned} \tag{5.4.2}$$

We will require to use Burkholder's inequality.

Lemma 5.4.1. Burkholder. Let $Y = \{Y(t), t \in [0,1]\}$ be a stochastic process taking values in $\mathbb{R}^d$ such that $Y^{(i)} \in L^2_{a,1}$ for any $1 \leq i \leq d$. Then, for any $p > 0$, there exists a constant C_p such that

$$E\left(\sup_{0\leq t\leq 1}\left|\int_0^t Y(s)\mathrm{d}B(s)\right|^p\right) \leq C_p E\left(\left(\int_0^1 |Y(s)|^2\mathrm{d}s\right)^{p/2}\right)$$

Proof. For a reference, see [6], Theorem 3.28 in page 166. ■

Proposition 5.4.2. *Let* $\delta \geq 0$, X_ε *the path-wise unique strong solution of* (5.4.1) *and* X *the deterministic function defined by* (5.4.2). *Then,*

$$\lim_{\varepsilon\to 0}\mathbb{P}(\|X_\varepsilon(\cdot) - X\| \geq \delta) = 0, \tag{5.4.3}$$

that is, $X_\varepsilon(\cdot)$ *converge in probability in* $C([0,1])$ *to the deterministic function* X*.*

Proof. Write $S := \sup\limits_{0\leq t\leq 1}\left|\int_0^t \sigma(X_\varepsilon(s))\mathrm{d}B(s)\right|$. Then, for $0 \leq t \leq 1$

$$\begin{aligned} |X_\varepsilon(t) - X(t)| &\leq \int_0^t |b(X_\varepsilon(s)) - b(X(s))|\,\mathrm{d}s + \sqrt{\varepsilon}\left|\int_0^t \sigma(X_\varepsilon(s))\mathrm{d}B(s)\right| \\ &\leq L\int_0^t |X_\varepsilon(s) - X(s)|\mathrm{d}s + \sqrt{\varepsilon}S. \end{aligned}$$

Applying Gronwall's Lemma (Lemma 5.1.5) we have

$$|X_\varepsilon(t) - X(t)| \leq \sqrt{\varepsilon}S + \sqrt{\varepsilon}LS\int_0^t e^{L(t-s)}\mathrm{d}s \leq \sqrt{\varepsilon}S(1 + Le^L).$$

So,

$$\mathbb{P}(\|X_\varepsilon(\cdot) - X\| \geq \delta) \leq \mathbb{P}(\sqrt{\varepsilon}S(1 + Le^L) \geq \delta) = \mathbb{P}\left(S \geq \frac{\delta}{\sqrt{\varepsilon}(1 + Le^L))}\right) \tag{5.4.4}$$

We check that S has finite expectation applying Burkholder's inequality (Lemma 5.4.1)

$$E(S) = E\left(\sup_{0\le t\le 1}\left|\int_0^t \sigma(X_\varepsilon(s))\mathrm{d}B(s)\right|\right) \le C_1 E\left(\left(\int_0^1 |\sigma(X_\varepsilon(s)|^2\,\mathrm{d}s\right)^{1/2}\right) \le C_1 L,$$

where $C_1 > 0$ is some constant. Finally, applying Chebyshev's inequality in (5.4.4) we have

$$\mathbb{P}(\|X_\varepsilon(\cdot) - X\| \ge \delta) \le \frac{\sqrt{\varepsilon}(1+Le^L))}{\delta}E(S) \le \frac{\sqrt{\varepsilon}(1+Le^L))C_1 L}{\delta}.$$

which implies (5.4.3). ■

Note that the previous proof applies in Case 2 and also in Case 1 because in that last case the diffusion matrix σ is also bounded. We now proceed to prove the LDP.

Theorem 5.4.3. *The family $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C([0,1])$ with the good rate function*

$$I(f) = \inf_{\{g\in H_1:\ f(t)=x_0+\int_0^t b(f(s))ds+\int_0^t \sigma(f(s))\dot{g}(s)ds\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 ds. \tag{5.4.5}$$

Proof. Define $Y_\varepsilon := \{Y_\varepsilon(t) := X_\varepsilon(t) - x_0,\ t \in [0,1]\}$. Then Y_ε is the path-wise unique strong solution of the next SDE

$$\begin{aligned} \mathrm{d}Y_\varepsilon(t) &= b_Y(Y_\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\sigma_Y(Y_\varepsilon(t))\mathrm{d}B(t), \\ Y_\varepsilon(0) &= 0, \end{aligned} \tag{5.4.6}$$

where $b_Y(x) := b(x+x_0)$ and $\sigma_Y(x) := \sigma(x+x_0)$. Note that b_Y and σ_Y satisfy the Lipschitz condition and are bounded by the same constant L of b and σ. So, if $\widetilde{\nu}_\varepsilon$ is the law of $Y_\varepsilon(\cdot)$ in $C_0([0,1])$, by Theorem 5.3.9, the family $\{\widetilde{\nu}_\varepsilon\}$ satisfies a LDP in $C_0([0,1])$ with the good rate function

$$I_Y(f) = \inf_{\{g\in H_1:\ f(t)=\int_0^t b_Y(f(s))\mathrm{d}s+\int_0^t \sigma_Y(f(s))\dot{g}(s)\mathrm{d}s\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2\mathrm{d}s.$$

Let $C \subset C([0,1])$ be a closed subset, then, since $X_\varepsilon(0) = x_0$ we have

$$\widetilde{\mu}_\varepsilon(C) = \widetilde{\mu}_\varepsilon\left(C \cap \{f \in C:\ f(0) = x_0\}\right)$$

Write $\widehat{C} := C \cap \{f \in C:\ f(0) = x_0\}$. Then $\widehat{C} - x_0 \subset C_0([0,1])$ is a closed subset and applying the LDP upper bound of $\{\widetilde{\nu}_\varepsilon\}$ we obtain that

$$\limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(C) = \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\mu}_\varepsilon(\widehat{C}) = \limsup_{\varepsilon\to 0} \varepsilon \log \widetilde{\nu}_\varepsilon(\widehat{C} - x_0)$$

$$\le -\inf_{f\in\widehat{C}-x_0} I_Y(f) = -\inf_{f\in\widehat{C}} I_Y(f-x_0)$$
$$= -\inf_{f\in\widehat{C}} \inf_{\{g\in H_1:\ f(t)-x_0=\int_0^t b_Y(f(s)-x_0)\mathrm{d}s+\int_0^t \sigma_Y(f(s)-x_0)\dot{g}(s)\mathrm{d}s\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2\mathrm{d}s$$
$$= -\inf_{f\in C} \inf_{\{g\in H_1:\ f(t)=x_0+\int_0^t b(f(s))\mathrm{d}s+\int_0^t \sigma(f(s))\dot{g}(s)\mathrm{d}s\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2\mathrm{d}s = -\inf_{f\in C} I(f),$$

where I is the function in (5.4.5). This proves the upper bound of the LDP of $\{\widetilde{\mu}_\varepsilon\}$.

By the same argument, we can prove that lower bound of the LDP of $\{\widetilde{\mu}_\varepsilon\}$.

Since I_Y is a good rate function and $I(f) = I_Y(f-x_0)$, I is also a good rate function because

$$\{f\in C([0,1]):\ I(f)\le\alpha\} = \{f\in C([0,1]):\ I(f)\le\alpha,\ f(0)=x_0\}$$
$$= \{f\in C([0,1]):\ I_Y(f-x_0)\le\alpha,\ f(0)=x_0\} = \{f+x_0\in C_0([0,1]):\ I_Y(f)\le\alpha\}$$

and $\{f+x_0\in C_0([0,1]):\ I_Y(f)\le\alpha\}$ is compact.

Moreover, note that $I(X)=0$, where X is the deterministic function that solves (5.4.2), because then $g=0(\cdot)\in H_1$ satisfies the condition in the infimum of I. In addition, $I(f)>0$ for $f\neq X$ because then a function $g\in H_1$ that satisfies the condition in the infimum of I can not be constant equal to 0. In conclusion, we are under the hypothesis of Proposition 1.4.3 and we have a control of the exponential decay of $\{\widetilde{\mu}_\varepsilon(\Gamma)\}$ for $\Gamma\in\mathcal{B}(C_0([0,1]))$ whenever $X\notin\overline{\Gamma}$. ■

5.5 Case 4

Note that the good rate function in Theorem 5.4.3 seems quite complicated because it involves computing an infimum. We are going to see that with some extra conditions on σ we can obtain a simpler version of I.

We are going to assume that $\sigma(x)$ is invertible for all $x\in\mathbb{R}^d$ and that σ^{-1} is also bounded, that is,

$$\sup_{x\in\mathbb{R}^d} |\sigma^{-1}(x)| \le L,$$

where $L>0$ is the same constant as in the Lipschitz and boundedness condition of b and σ.

Fix $\varepsilon > 0$ and let X_ε be the path-wise unique strong solution of the following SDE

$$\begin{aligned} \mathrm{d}X_\varepsilon(t) &= b(X_\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_\varepsilon(t))\mathrm{d}B(t), \\ X_\varepsilon(0) &= x_0 \in \mathbb{R}^d, \end{aligned}$$

and let $\widetilde{\mu}_\varepsilon$ be the law of $X_\varepsilon(\cdot)$ in $C([0,1])$. Then, with the extra hypothesis on σ we have the following result.

Theorem 5.5.1. *$\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C([0,1])$ with the good rate function*

$$I(f) = \begin{cases} \frac{1}{2}\int_0^1 \left[\dot{f}(t) - b(f(t))\right]' a^{-1}(f(t)) \left[\dot{f}(t) - b(f(t))\right] dt & \text{if } f \in H_1^{x_0} \\ \infty & \text{otherwise.} \end{cases} \tag{5.5.1}$$

where $'$ indicates the transpose of a vector or a matrix, $a(x) := \sigma(x)\sigma'(x)$ and

$$H_1^{x_0} := \left\{ f \in C_0([0,1]) : \ f(t) = x_0 + \int_0^t g(s)\,ds, \ g \in L_2([0,1], \mathbb{R}^d) \right\}.$$

Proof. By Theorem 5.4.3 we know that $\{\widetilde{\mu}_\varepsilon\}$ satisfies a LDP in $C([0,1])$ with the good rate function

$$I(f) = \inf_{\{g \in H_1:\ f(t) = x_0 + \int_0^t b(f(s))\mathrm{d}s + \int_0^t \sigma(f(s))\dot{g}(s)\mathrm{d}s\}} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 \mathrm{d}s.$$

Suppose that there exists $g \in H_1$ such that

$$f(t) = x_0 + \int_0^t b(f(s))\mathrm{d}s + \int_0^t \sigma(f(s))\dot{g}(s)\mathrm{d}s, \tag{5.5.2}$$

then,

$$|\dot{f}(t)| \le |b(f(t))| + |\sigma(f(t))||\dot{g}(t)| \le L + L|\dot{g}(t)|.$$

This implies that $f \in H_1^{x_0}$. Therefore, if $f \notin H_1^{x_0}$, there is no $g \in H_1$ satisfying (5.5.2) and we conclude that $I(f) = \infty$.

On the other hand, if $f \in H_1^{x_0}$, then there exists a unique $g \in H_1$ satisfying (5.5.2). This is because g is defined by $g(0) = 0$ and

$$\dot{g}(t) = \sigma^{-1}(f(t))(\dot{f}(t) - b(f(t)))$$

after taking derivatives in (5.5.2). To see that $g \in H_1$ observe that

$$|\dot{g}(t)| \le L(|\dot{f}(t)| + L).$$

Finally,

$$\begin{aligned}|\dot{g}(t)|^2 &= \left[\sigma^{-1}(f(t))(\dot{f}(t) - b(f(t)))\right]' \left[\sigma^{-1}(f(t))(\dot{f}(t) - b(f(t)))\right] \\ &= \left[\dot{f}(t) - b(f(t))\right]' a^{-1}(f(t)) \left[\dot{f}(t) - b(f(t))\right].\end{aligned}$$

Hence, we have obtained the function in (5.5.1). ■

6 Conclusions

The notion of large deviations has its roots in the study of rare events which can be viewed as opposite of concentration inequalities of random events around the mean. Cramér's Theorem, presented in Chapter 2, was motivated by evaluation of risk in insurance policies. In this work we have seen the evolution of Cramér's ideas from the simple setting of random sequences to more sophisticated objects namely, stochastic processes that are solutions to stochastic differential equations driven by Brownian motion. These are random vectors taking values in the space of continuous functions equipped with the supremum norm.

The toolbox of the memoir is Chapter 3, which provides the fundamental theory to address the study of LDPs in a variety of situations. In particular, we see in Section 3.4 how to transfer LDPs from (elementary) families of random vectors to (not necessary continuous) transformations of these families. With this ideas in mind, we present the LDP for Brownian motion (Schilder's theorem) in Chapter 4 which in Chapter 5 is transferred to the LDP for SDEs.

In conclusion, the memoir provides the foundations of the theory of the LDP with fundamental illustrations, highlighting the crucial role of Brownian motion.

References

[1] Dembo, Amir.; Zeitouni, Ofer. *Large Deviations Tecnhiques and Applications*. Second Edition. United States of America: Springer, 2010. ISBN 0387984062.

[2] Kelley, John L. *General Topology*. First Edition. United States of America: Springer, 1975. ISBN 0387901256.

[3] Olivieri, Enzo.; Eulália Vares, Maria. *Large deviations and metastability*. First edition. New York: Cambridge University Press, 2005. ISBN 0521591635.

[4] Rudin, Walter. *Principles of Mathematical Analysis*. Third edition. United States of America: McGraw-Hill Education, 1976. ISBN 007054235X.

[5] Riesz, Frigyes.; SZ.-Nagy, Béla. *Functional Analysis*. Translated from the Second French edition. United States of America: Dover, 1990. ISBN 0486662896.

[6] Karatzas, Ioannis.; E. Shreve, Steven. *Brownian motion and stochastic calculus*. Second Edition. New York: Springer, 1996. ISBN 3540976558.

[7] Brunner, Hermann. *Volterra Integral Equations*. First edition. United Kingdom: Cambridge University Press, 2017. ISBN 9781107098725.

[8] Varadhan, S. R. S. *Large Deviations and Applications*. First edition. Philadelphia: Society for Industrial and Applied Mathematics, 1984. ISBN 0898711894.

[9] Wong, E.; Zakai, M. *Riemann-Stieltjes approximations of stochastic integrals*. Z. Wahrscheinlichkeitstheorie verw Gebiete 12, 87–97, 1969.