ΑΙΚΑΤΕΡΙΝΗ ΠΑΠΑΓΙΑΝΝΟΥΛΗ

Θεωρία των μεγάλων αποκλίσεων και εφαρμογές

Μεταπτυχιακή εργασία για το Δ.Π.Μ.Σ. Εφαρμοσμένες Μαθηματικές επιστήμες

Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών
Εθνικό Μετσόβειο Πολυτεχνείο
Αθήνα, 2014

AIKATERINI PAPAGIANNOULI

Large Deviations and Applications in Option Pricing and Importance Sampling

Thesis for the Interdepartmental Postgraduate Course Programme
Applied Mathematics

FACULTY OF APPLIED MATHEMATICAL AND PHYSICAL SCIENCES
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
ATHENS, 2014

AIKATERINI PAPAGIANNOULI

Large Deviations and Applications in Option Pricing and Importance Sampling

Thesis for the Interdepartmental Postgraduate Course Programme
Applied Mathematics

It has been approved by the advisory committee
Prof. Dr. Michail Loulakis
Prof. Dr. Antonis Papapantoleon
Prof. Dr. Ioannis Spiliotis

FACULTY OF APPLIED MATHEMATICAL AND PHYSICAL SCIENCES
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
ATHENS, 2014

To my teachers

Acknowledgements

In the next few lines I would like to thank people who devoted part of their time so that this thesis be accomplished. First of all, I would like to express my deepest gratitude to my advisor Prof. Dr. Michail Loulakis. He was the teacher who introduced me in Stochastic Analysis and Large Deviations. He was a real teacher, supervising, helping, motivating and inspiring in a very natural and subtle way. Indeed, his serene and intuitive comments made clear any obscurity that arised. Moreover, the freedom that he provided me in exploring my own ideas enhanced my confidence and increased the pleasure of engaging the whole endeavor. I would like also to thank Prof. Dr. I. Spiliotis. He was the teacher who transformed the most vague and unsettling points of Probability into a charming endeavor. Furthermore, I would like to thank Prof. Dr. Antonis Papapantoleon for introducing me in the area of computing option prices using sophisticated methods like Monte-Carlo methods.
Finally, I would like to thank my friends, my family, and especially my father, whose constant support and patience enabled me to achieve my targets.

Contents

Περίληψη

Η θεωρία των μεγάλων αποκλίσεων συνίσταται στην εκτίμηση σπανίων ενδεχομένων- ενδεχομένων δηλαδή που αποκλίνουν απο την τυπική συμπεριφορά στο θεωρούμενο χώρο πιθανότητας. Ο ισχυρός νόμος των μεγάλων αριθμών μας δίνει ένα παράδειγμα τυπικής συμπεριφοράς αφού αν η ακολουθία X_i ικανοποιεί τις υποθέσεις του Ι.Ν.Μ.Α. τότε η τυπική συμπεριφορά της $\frac{S_n}{n}$ είναι (για μεγάλα n) να είναι κοντά στο μέσο της $\bar{x}$. Από τον ασθενή νόμο των μεγάλων αριθμών, το ενδεχομενό ο μέσος όρος να βρίσκεται έξω από μια περιοχή του $\bar{x}$ γίνεται σπάνιο καθώς το n γίνεται μεγάλο. Η ιδέα να εκτιμήσουμε σπάνια ενεχόμενα βρίσκεται στην αλλάγη του μέτρου πιθανότητας κάτω από το οποίο έχουμε μια μη τυπική συμπεριφορά.

Φυσικά το να μελετά κανείς την πιθανότητα σπανίων ενδεχομένων θέτει τα πράγματα σε πολύ γενικό πλαίσιο. Έτσι από τη μια μεν η θεωρία των μεγάλων αποκλίσεων βρίσκει εφαρμογή σε περιοχές τόσο διαφορετικές μεταξύ τους όσο η Στατιστική, Χρηματοοικονομικά, Τηλεπικοινωνίες, Στατιστική Μηχανική, από την άλλη όμως συχνά χρειάζεται να καταφύγει κάποιος σε ειδικές για το εκάστοτε πρόβλημα μεθόδους.

Στην παρούσα εργασία θα προσπαθήσουμε να δώσουμε παραδείγματα ώστε να αναδείξουμε τις κεντρικές ιδέες και τεχνικές της θεωρίας των μεγάλων αποκλίσεων. Τέλος θα δούμε μια εφαρμογή θεωρητική, το πρόβλημα εξόδου ενός στοχαστικού συστήματος από ένα φραγμένο υποσύνολο τού $\mathbb{R}^d$ αλλά και μια εφαρμογή αυτού του προβλήματος στα Χρηματοοικονομικά.

Abstract

The subject of Large Deviations is going back to the early 1930s. It in fact started in Scandinavia, with actuaries working for the insurance industry. The pioneer who started that subject was named Esscher. He was interested in a situation where too many claims could be made against the insurance company. He wanted to calculate the probability of the total claim amount exceeding the reserve fund set aside for paying these claims. So large deviations provides us a formula to estimate tail probabilities. Central Limit Theorem states that the distribution of sums of independent random variables has a Gaussian approximation. However, the error is measured in terms of difference. Both these numbers are very small, therefore the difference between them is small. But, we are interested in how small it is; we are interested in the ratio of these two things, not just the difference of these small numbers. The idea is: how one can shift ones focus so that we can look at the ratio rather than just at the difference. Esscher came up with this idea that is called Esscher"s tilt. It is a way of changing the measure. And from this point of view what was originally considered as a tail event now becomes a central event. Tail events or rare events are events with very small probability, but we would like to have some idea of how small it is. We would like to measure it in logarithmic scale. The main target of the current thesis is to study Large Deviations theory. The thesis is orginised as follows.

In the first part Large Deviations for i.i.d. sequences is presented. More precisely, in Chapter 1, we are studying some basic formualtion and definitions in order to formulate Large Deviations Principle(LDP). The rest sections can be viewed as a sequence of examples, of increasing difficulty, to which the principle developed previously can be applied.

In Chapter 2, we are studying some classical results in Large Deviations area. The most classical result is Cramér theorem. Indeed, we consider the case of real i.i.d. random variables in $\mathbb{R}$. Then Cramér theorem in $\mathbb{R}^d$ is proven through weak LDP and exponential tightness. In Chapter 3 some further general principles are introduced for carrying out Large Deviations results: Varadhan's lemma, contraction principle, relative entropy and Varadhan-Donsker formula. In Chapter 4, Large Deviations for abstract measures in Banach spaces are proven. Indeed, we are studying the example of level-2 LDP.

In many problems the interest is in rare events that depend on random process, and the correspond asymptotics probabilities, usually called sample pathe large deviations. In the second part, we are studying Large Deviation principle for stochastic processes. In chapter 5 we prove Schilder's theorem for a rescaled Browinian motion. Hence, we continue in the next Chapter with two applications of Schilder's theorem. The deriviation of Strassen renowned Law of Iterated Logarithm and the behavior of diffusions with small parameter, proving Freidlin-Wentzel theorem using Euler approximations and superexponential estimates.

In Chapter 7 we deal with the problem of diffusions exit from a bounded domain (the well-known Exit problem). Firstly, we prove this problem as an immediate result of Freidlin-Wentzel theory. Then Exit problem is considered as a parabolic problem. To this end, we introduce an approach which connects large deviations asymptotics of the

corresponding family of measures with an optimal control problem and Hamilton-Bellman-Jacobi equation. This approach is developed within viscosity solutions.

The above problem occurs naturally in Finance. In Chapter 8 we used importance sampling to reduce the variance of a Monte-Carlo computed price of deeply out of the money options. The basic principle of importance sampling is to reduce variance by changing probability measure from which paths are generated. The idea is to change the distribution of the price process and to derive the process to the region of high distribution to the required expectation. We focus on importance sampling for diffusions models and then we show how to obtain an optimal change of measure by large deviation approximations of the required expectation.

Part I

Large Deviations for I.I.D. sequences

Chapter 1

Introduction

1.1 Rare events and Large Deviations

The area of Large Deviations is a set of asymptotic results on rare events probabilities and a set of methods to derive such results. Large deviations is a very active area in applied probabily, and find important applications in finance where questions related to extremal events play an increasing role. In its basic form, the theory of Large Deviations considers the normalizations of $\log \mathbb{P}(A_n)$ for a sequence of events with asymptotically vanishing probabilities. Intuitively, the scope of Large Deviations is the study of deviations far from typical behaviors.

We begin our journey on familiar territory. Let $X_1, X_2, \ldots, X_n,\ n \in \mathbb{N}$ be independent and indentically distributed (briefly i.i.d.) random variables on a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$. Write $\mathbb{E}$ to denote the expectation under $\mathbb{P}$, let

$$\mu = \mathbb{E}X_1 \in \mathbb{R}$$
$$\sigma^2 = VarX_1 \in (0, \infty),$$

Denote the partial sum by S_n, i.e. $S_n = X_1 + X_2 + \cdots + X_n$. Two fundamendal theorems in Probability theory dealing with such sequences give examples of typical behaviors.

Strong Law of Large Numbers(SLLN)

$$\frac{1}{n} S_n \overset{n \to \infty}{\longrightarrow} \mu \quad \mathbb{P}\text{-a.s.}$$

Central Limit theory(CLT)

$$\frac{1}{\sigma\sqrt{n}}(S_n - \mu n) \overset{n \to \infty}{\longrightarrow} Z \quad \text{in law w.r.t } \mathbb{P},$$

where Z is a standard normal random variable.

While the SLLN asserts that the empirical average $\frac{1}{n}S_n$ converges to μ as $n \to \infty$, the CLT quantifies the probability that S_n differs from μn by an amount of order $\sqrt{n}$. Deviations of this sizes are called 'normal'. The SLLN implies that

$$\mathbb{P}\left(\frac{S_n}{n} \notin (\bar{x} - \delta, \bar{x} + \delta)\right) \overset{n \to \infty}{\longrightarrow} 0,$$

for each $\delta > 0$. Events like $\mathbb{R} \setminus (\bar{x} - \delta, \bar{x} + \delta)$, $\delta > 0$ or $[y, \infty)$, $y > \bar{x}$ are considered as rare events for $\frac{S_n}{n}$ as $n \to \infty$.

It is our task to quantify the rate at which probabilities of rare events tends to zero. In, general, a detailed answer to this task is seldom available. However, if one restricts one's attention to events which are 'very deviant' in the sense that the probability of their occurence decays exponentially fast to zero and if one only asks about the exponential rate, then one has a much better chance to find a solution. Now, we give a definition of rare events using the distribution μ_n of Y_n, $n \in \mathbb{N}$:

Definition 1.1. *Suppose that $\{\mu_n : n > 0\}$ is a family of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the property that $\mu_n \Rightarrow \delta_p$ as $n \to \infty$ for some $p \in \mathbb{R}$(i.e. μ_n tends weakly to the point mass δ_p). Then for each open set $U \in p$ we have that $\mu_n(U^c) \to 0$. So we can reasonably say that as $n \to \infty$, the measures μ_n see 'p as being typical'. Equivalently,one can say that events $A \subseteq \mathcal{X}$ lying outside a neibhorhood of p describe increasingly 'deviant' behavior, These events are called rare events.*

Let us begin with one of the most basic computations one can carry out in order to find out how fast the probability of a rare event converges to zero.

Example 1.1. *Consider coin tosses. Let $\{X_i\}$ i.i.d. sequence of Bernoulli random variables with success probability p. The distribution of X_i is $B_p := p\delta_1 + (1-p)\delta_0$ for each $i \in \mathbb{N}$. We want to estimate the rate of convergence of*

$$\mathbb{P}\left(\frac{S_n}{n} \geq q\right) \to 0$$

for $q > p = \mathbb{E}X_i$ and the rate of convergence of

$$\mathbb{P}\left(\frac{S_n}{n} \leq q\right) \to 0$$

for $q < p$.

Proof. Using Chebyshev inequality, one has that

$$\mathbb{P}(S_n \geq nq) = \mathbb{P}\left(\frac{S_n - np}{n} \geq q - p\right) \leq \frac{1}{(q-p)^2 n^2}\mathbb{E}(S_n - np)^2$$

$$= \frac{1}{(q-p)^2 n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{(q-p)^2 n^2} np(1-p) = \frac{p(1-p)}{(q-p)^2}\frac{1}{n}.$$

Thus, the rate of convergence is at least linear. However, it also holds from Chebyshev that

$$\mathbb{P}\left(\frac{S_n}{n} \geq q\right) \leq \frac{1}{(q-p)^4 n^4}\mathbb{E}(S_n - np)^4$$

$$= \frac{1}{n^4(q-p)^4}\left(n\mathbb{E}(X_1 - p)^4 + 6n(n-1)\left[\mathbb{E}(X_1 - p)^2\right]^2\right) \leq \frac{C}{n^2}$$

for $C \leq 0$ constant. Thus, the rate of convergence is at least quadratic. Using existence of higher moments of X_i we can obtain faster rates of decay for $\mathbb{P}\left[\frac{S_n}{n} \geq q\right]$, $q > p$. Indeed, since $\mathbb{E}e^{\lambda X_1} \leq \infty$ for every $\lambda \in \mathbb{R}$, then we can show that the rate of convergence is at least exponential. Using Chebyshev inequality,we have that

$$\mathbb{P}\left(\frac{S_n}{n} \geq q\right) = \mathbb{P}(S_n \geq nq) \leq e^{-nq\lambda}\mathbb{E}e^{\lambda S_n} = e^{-n\lambda q}\left(\mathbb{E}^{\lambda X_1}\right)^n$$

$$= e^{-n\lambda q}\left(pe^{\lambda}+1-p\right)^{n} = e^{-n\lambda q}e^{n\log\left(pe^{\lambda}+1-p\right)}$$
$$= e^{-n\left[q\lambda-\log\left(pe^{\lambda}+1-p\right)\right]}.$$

Therefore,

$$\mathbb{P}\left[\frac{S_n}{n}\geq q\right]\leq e^{-n\sup_{\lambda\geq 0}\left[\lambda q-\log\left(pe^{\lambda}+1-p\right)\right]}.$$

Now for $q>p$, from Jensen inequality we have that

$$\sup_{\lambda>0}[\lambda q-\log\left(pe^{\lambda}+1-p\right)] = \sup_{\lambda\in\mathbb{R}}[\lambda q-\log\left(pe^{\lambda}+1-p\right)],$$

Thus, if we define

$$I\left(q\right)=\sup_{\lambda\in\mathbb{R}}\left[\lambda q-\log\left(pe^{\lambda}+1-p\right)\right]$$

we obtain

$$\mathbb{P}\left[\frac{S_n}{n}\geq q\right]\leq e^{-nI(q)}$$

If we prove that $I\left(q\right)>0$, then we have show that the rate of convergence $\mathbb{P}\left(\frac{S_n}{n}\geq q\right)\to 0$ is exponential with rate function $I\left(q\right)$. Differentiating $I\left(q\right)$, we have that

$$\frac{d}{d\lambda}\left(\lambda q-\log\left(pe^{\lambda}+1-p\right)\right)=q-1+\frac{1-p}{pe^{\lambda}+1-p}$$

If $q>1$ then $q-1+\frac{1-p}{pe^{\lambda}+1-p}>0$ and

$$I\left(q\right)=\lim_{\lambda\to\infty}\left(\lambda q-\log\left(pe^{\lambda}+1-p\right)\right)=\begin{cases}\log\dfrac{1}{p} & q=1\\ \infty & q>1\end{cases}.$$

If, now, $p<q<1$ then $I\left(q\right)$ has unique critical point $\lambda_*:=\log\frac{q(1-p)}{p(1-q)}>0$, which is global maximum. So that,

$$I\left(q\right)=q\lambda_*-\log\left(pe^{\lambda_*}+1-p\right)=q\log\frac{q\left(1-p\right)}{p\left(1-q\right)}-\log\left(\frac{q\left(1-p\right)}{1-q}+1-p\right)$$

$$=q\log\frac{q}{p}+q\log\frac{1-p}{1-q}-\log\frac{1-p}{1-q}=q\log\frac{q}{p}-\left(1-q\right)\log\frac{1-p}{1-q}\Rightarrow$$

$$I(q)=q\log\frac{q}{p}+\left(1-q\right)\log\frac{1-q}{1-p} \tag{1.1}$$

Therefore, for each $q\in(p,1)$ we have that

$$I'\left(q\right)=\log\frac{q}{p}+q\frac{1/p}{q/q}-\log\frac{1-q}{1-p}-\left(1-q\right)\frac{1/\left(1-p\right)}{\left(1-q\right)/\left(1-p\right)}=\log\frac{q\left(1-p\right)}{p\left(1-p\right)}>0$$

and I is monotonically increasing function in $(p,1)$. But, $\lim_{q\to p}I\left(q\right)=0$, thus $I\left(q\right)>0$ for every $q\in(p,1)$. The same holds for every $q<p$. We have that

$$\mathbb{P}\left(\frac{S_n}{n}\leq q\right)\leq e^{nI(q)}\quad\forall n\in\mathbb{N}$$

Similarly, if now $0 < q < p$ thus $I(q)$ is given by 1.1 and if $q \leq 0$ then, then

$$I(q) = \lim_{n\to\infty} \left(\lambda q - \log\left(pe^{\lambda} + 1 - p\right)\right) = \begin{cases} \log \dfrac{1}{1-p} & q = 0 \\ \infty & q < 0 \end{cases}$$

Therefore, taking every case into consideration one has that

$$I(q) = \begin{cases} q \log \dfrac{q}{p} = +(1-q)\log \dfrac{1-q}{1-p} & q \in [0,1] \\ \infty & q \notin [0,1] \end{cases}$$

with the convention that $0 \cdot \infty = 0$.
The function $I(q)$ is called *rate function.* It is infinite outside $[0,1]$, finite and strictly convex inside $[0,1]$ and has a unique zero at $q = p$.
Next, we show that $I(q)$, $p < q < 1$ is the optimal exponential rate of convergence. By this we mean that if there are $n_0 \in \mathbb{N}$ and $\theta > 0$ such that

$$\forall n_0 \in \mathbb{N} \Rightarrow \mathbb{P}(S_n \geq nq) \leq e^{-n\theta}, \tag{1.2}$$

then $\theta \leq I(q)$. It suffices to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nq) \geq -I(q), \tag{1.3}$$

since if 1.2 holds for some $n_0 \in \mathbb{N}$, $\theta > 0$ then

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nq) \leq -\theta,$$

Then, from 1.3 we have that

$$-I(q) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nq) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq nq) \leq -\theta.$$

Thus, $\theta \leq I(q)$. Firstly, since S_n is a the sum of i.i.d. Bernoulli random variables with parameters (n,p) and $p \in (0,1)$, then S_n follows Binomial distribution with parameter p. Therefore,

$$\mathbb{P}(S_n \geq nq) = \sum_{k\geq nq} \mathbb{P}(S_n = k) = \sum_{k\geq nq} \binom{n}{k} p^k (1-p)^{n-k}.$$

Next we estimate the above sum for $q = p$ since its computation is rather difficult its computation. Then we estimate the limit for every q. More presicely,

$$\sum_{k\geq np} \binom{n}{k} p^k (1-p)^{n-k} = \mathbb{P}(S_n \geq np) = \mathbb{P}\left(\frac{S_n - np}{p(1-p)\sqrt{n}} \geq 0\right),$$

Then, using the CLT, we have that

$$\lim_{n\to\infty} \sum_{k\geq nq} \binom{n}{k} p^k (1-p)^{n-k} = \int_0^{\infty} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} = \frac{1}{2}$$

For $q > p$,

$$\mathbb{P}(S_n \geq nq) = \sum_{k\geq nq} \binom{n}{k} q^k (1-q)^{n-k} \left(\frac{p}{q}\right)^k \left(\frac{1-p}{1-q}\right)^{n-k}$$

$$= \left(\frac{1-p}{1-q}\right)^n \sum_{k\geq nq} \binom{n}{k} q^k (1-q)^{n-k} \left(\frac{p(1-q)}{q(1-p)}\right)^k .$$

For any $p \in (0,1)$

$$\sum_{k\geq nq} \binom{n}{k} p^k (1-p)^{n-k} = \mathbb{P}(S_n \geq (p+\epsilon) n) = \mathbb{P}\left(\frac{S_n - np}{\sigma\sqrt{n}} \geq \frac{\epsilon\sqrt{n}}{\sigma}\right) \to 0 \quad \text{as} \quad n \to \infty,$$

for every $\epsilon > 0$ and $\sigma = Var(X_i) = p(1-p)$. Since $\frac{p(1-q)}{q(1-p)} < 1$ then

$$\mathbb{P}(S_n \geq nq) \geq \left(\frac{1-p}{1-q}\right)^n \sum_{nq\leq k\leq n(q+\epsilon)} \binom{n}{k} q^k (1-q)^{n-k} \left(\frac{p(1-q)}{q(1-p)}\right)^{n(q+\epsilon)} .$$

Thus,

$$\frac{1}{n}\log \mathbb{P}(S_n \geq nq) \geq \log\frac{1-p}{1-q} + (q+\epsilon)\log\left(\frac{p(1-q)}{q(1-p)}\right) + \frac{1}{n}\log\left(\sum_{nq\leq k\leq n(q+\epsilon)} \binom{n}{k} q^k (1-q)^{n-k}\right).$$

However, the sum in the last inequality can go as much close to 1 as we want by CLT, therefore

$$\lim_{n\to\infty} \frac{1}{n}\log\left(\sum_{nq\leq k\leq n(q+\epsilon)} \binom{n}{k} q^k (1-q)^{n-k}\right) = 0.$$

Thus,

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq nq) \geq \log\frac{1-p}{1-q} + (q+\epsilon)\log\left(\frac{p(1-q)}{q(1-p)}\right),$$

since this occurs for every $\epsilon > 0$ we have that

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq nq) \geq \log\frac{1-p}{1-q} + q\log\left(\frac{p(1-q)}{q(1-p)}\right) = -I(q),$$

This completes the proof. $\square$

1.2 The Large Deviation Principle

Having seen an example for which it is possible to carry out a succesful analysis of the large deviations, we will now formulate general principles. Firstly, we begin our program by introducing some useful definitions.

Definition 1.2. *Let $\mathcal{X}$ be a Polish space with distance $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$. $f : \mathcal{X} \to [-\infty, +\infty]$ is* ***lower semi-continuous*** *if it satisfies any of the following equivalent properties:*

(i) $\liminf_{n\to\infty} f(x_n) \geq f(x)$ *for all x_n such that $x_n \to x \in \mathcal{X}$.*

(ii) f has closed level sets, i.e., $f^{-1}([-\infty, c]) = \{x \in \mathcal{X} : f(x) \leq c\}$ is closed for all $c \in \mathbb{R}$.

(iii) $\lim_{\epsilon\to 0} \inf_{y\in B_\epsilon(x)} f(y) = f(x)$ *with $B_\epsilon(x) = \{y \in \mathcal{X} : d(x, y) < \epsilon\}$*

Remark 1.1. *A lower semi-continuous function attains a minimum on every non-empty compact set.*

Definition 1.3. *The moment generating function of a distribution $\mu \in \mathbb{R}$ is a function which is given by the following formula:*

$$M_\mu(\lambda) = \int e^{\lambda x} d\mu(x).$$

The set $\mathcal{D}_{M_\mu} := \{\lambda \in \mathbb{R} : M_\mu(\lambda) < \infty\}$ is called essential range of M_μ.

Definition 1.4. *The* ***logarithmic moment generating function*** *is a function in $\mathbb{R}$ such that:*

$$\Lambda_\mu(\lambda) = \log\left(\int_{\mathbb{R}} e^{\lambda x} d\mu(x)\right), \quad \lambda \in \mathbb{R}$$

Remark 1.2. *Note that $\Lambda_\mu(\lambda) \in \mathbb{R}$ is a lower semi-continuous convex function. Indeed, by truncation, it is easy to write Λ_μ as the non-decreasing limit of smooth functions, and the convexity follows from Hölder's inequality.*

So we can define $\Lambda_\mu^*(x)$ be the **Legendre transform** of Λ_μ:

$$\Lambda_\mu^*(x) = \sup\{\lambda x - \Lambda_\mu(\lambda) : \lambda \in \mathbb{R}\}, \quad x \in \mathbb{R}.$$

Note that, by its definition as the point-wise supremum of linear functions, Λ_μ^* is necessarily lower semi-cintinuous and convex.

Definition 1.5. *Suppose that μ is a distribution in $\mathbb{R}$ with exponential moments. We call* ***rate function*** *of μ the function $I_\mu : \mathbb{R} \to [0, \infty]$ such that*

$$I_\mu(x) = \sup_{\lambda\in\mathbb{R}} (\lambda x - \log M_\mu(\lambda)) = \sup_{\lambda\in\mathcal{D}_{M_\mu}} (\lambda x - \log M_\mu(\lambda)).$$

Remark 1.3. *Obviously, $I_\mu \geq 0$ since for every $x \in \mathcal{X}$ zero is included in the domain of the set of which we take the supremum in the definition of $I_\mu(x)$. Of course, the above definition is applicable in a Polish space, with the appropriate modifications. This definition is carried out in following sections.*

Definition 1.6. *The function $I : \mathcal{X} \to [0, \infty)$ is called **good rate function** if:*

(i) $I \neq \infty$

(ii) *I is lower semi-continuous*

(iii) *I has compact level sets.*

Definition 1.7. *A sequence of probability measures $\{\mathbb{P}_n\}_n$ on $\mathcal{X}$ is said to satisfy the **large deviation principle** (LDP) with rate a_n and rate function I if:*

$$\limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(C) \leq -\inf_{x\in C} I(x), \quad \textit{for each closed} \quad C \subseteq \mathcal{X}$$

and

$$\liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(O) \geq -\inf_{x\in O} I(x), \quad \textit{for each open} \quad O \subseteq \mathcal{X}$$

Finally, I is a good rate function.

The following remarks help to explain the basic concept of LDP.

Remark 1.4. *In definition 1.7 it is crucial to distinguish between the asymptotics estimates for open and closed sets. Namely, one might try to replace (i), (ii) by the stronger requirement that*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(S) = -I(S) \quad \forall S \in \mathcal{B}(\mathcal{X}) : \textit{Borel field in} \quad \mathcal{X}. \tag{1.4}$$

However, this would be far too restrictive. Many examples that satisfy (i), (ii) do not satisfy 1.4. For instance, $\mathbb{P}_n$ might be non-atomic for all n. In that case, $\mathbb{P}_n(\{x\}) = 0$ for all $n \in \mathcal{X}$, so picking $S = \{x\}$ we would find that 1.4 could only be true with $I \equiv \infty$, which is contradicted by (i) of definition 1.6.

Remark 1.5. *The role of liminf in open and limsup in closed sets in the LDP reminds us of weak convergence of probability measures where the same boundary issue arises. $(\mathbb{P}_n)$ is said to converge weakly to $\mathbb{P}$ if*

(I) $\limsup_{n\to\infty} \mathbb{P}_n(C) \leq \mathbb{P}(C) \quad \forall C \subset \mathcal{X}$ *closed*

(II) $\liminf_{n\to\infty} \mathbb{P}_n(O) \geq \mathbb{P}(O) \quad \forall O \subset \mathcal{X}$ *open*

One can therefore view (i),(ii) in definition 1.7 as analogues of weak convergence on an exponential scale.

Remark 1.6. *Due to Portmanteau theorem, see [14]: th. 3.25, p.53, (I),(II) are equinalent to*

$$\int_{\mathcal{X}} F(x) \mathbb{P}_n(dx) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} F(x) \mathbb{P}(dx) \quad \forall F \in \mathcal{C}_b(\mathcal{X})$$

with $\mathcal{C}_b(\mathcal{X})$ the space of bounded continuous functions on $\mathcal{X}$. It is intuitively clear that the LDP is ideally suited for handling convergence of integrals of exponential functionals. This intuition will be worked out in Section 3.

Remark 1.7. *The LDP implies that*

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

since $\mathbb{P}_n(\mathcal{X}) = 1$ for all n and $\mathcal{X}$ is closed. Moreover, by remark 1.1 and defintion 1.6, there is an $x \in \mathcal{X}$ such that $I(x) = 0$. In many examples this zero is unique as we have seen in the example in the previous section and correponds to an underlying SLLN, but there are cases where it is not unique as we shall see later in section 5.

Remark 1.8. *It is possible to set up definitions 1.6 and 1.7 in the framework of an arbitrary topological space. We will, however, not insist on this degree of generality. Without the structure of a Polish space the theory tends to become more cumbersome and many results in the Polish space setting fail to carry over. Conversly, the more structure is added to $\mathcal{X}$, the stronger results that can be obtained.*

Now, we are going to give some properties of rate function of definition 1.5 before we prove the Cramér theorem.

Proposition 1.1. *Let μ be the distribution on $\mathbb{R}$ with exponential moments and mean $\bar{x}$. Suppose that I, μ, Λ the rate function,the moment generating function and the logarithmic generating function respectively. Then,*

(i) I is convex and lower semi-continuous.

(ii) $I(x) = 0$ if and only if $x = \bar{x}$.

(iii) $I(x) = \sup_{\lambda \geq 0} (\lambda x - \log M(\lambda))$ for every $x > \bar{x}$ and $I(x) = \sup_{\lambda \leq 0} (\lambda x - \log M(\lambda))$ for every $x < \bar{x}$.

(iv) I is decreasing in $(-\infty, \bar{x})$ and increasing in $(\bar{x}, \infty)$.

(v) For every $c \geq 0$ the level sets $I^{-1}[0, c] = \{x \in \mathbb{R} : I(x) \leq c\}$ are compact.

Proof. (i) I is the supremum of functions $\phi_\lambda(x) = \lambda x - \log M_\mu(\lambda)$, $\lambda \in \mathbb{R}$. Obviously, ϕ_λ are continuous functions, therefore I is lower semi-continuous as the supremum of continuous functions. Also, ϕ_λ are linear functions, so they are convex as the supremum of convex functions.
(ii) Since $e^{\lambda \bar{x}} \leq M_\mu(\lambda)$ then $\lambda \bar{x} \leq \log M_\mu(\lambda)$ for every $\lambda \in \mathbb{R}$. Therefore, $I(\bar{x}) \leq 0$. This proves that $I(\bar{x}) = 0$. We will show that $I(x) = 0 \Rightarrow x = \bar{x}$. Let $x \geq \bar{x}$ such that $I(x) = 0$. Then, $\lambda x \leq \log M_\mu(\lambda)$ and for every $\lambda \in \mathbb{R}$ one has that

$$\frac{e^{\lambda \bar{x}} - 1}{\lambda} \leq \frac{e^{\lambda x} - 1}{\lambda} \leq \frac{M(\lambda) - 1}{\lambda},$$

thus, taking $\lambda \to 0$ we have that $\bar{x} \leq x \leq \bar{x}$.
(iii) Let $x \geq \bar{x}$. Then, for every $\lambda \leq 0$ we have that

$$\lambda x - \log M_\mu(\lambda) \leq \lambda \bar{x} - \log M_\mu(\lambda) \leq I(\bar{x}) = 0$$

Since $I(x) \geq 0$ then every λ that are smaller than 0 don't contribute to the supremum of the definition of $I(x)$. By this we mean that

$$I(x) = \max\left\{\sup_{\lambda \leq 0} (\lambda x - \log M(\lambda)), \sup_{\lambda \geq 0} (\lambda x - \log M(\lambda))\right\}$$

$$= \max\left\{0, \sup_{\lambda \geq 0} (\lambda x - \log M(\lambda))\right\} = \sup_{\lambda \geq 0} (\lambda x - \log M(\lambda)).$$

The same argument holds if $x \leq \bar{x}$ then $\sup_{\lambda \leq 0} (\lambda x - \log M_\mu(\lambda))$.
(iv) Let $x < y < \bar{x}$. Then, for each $\lambda \leq 0$ we have that $\lambda y - \log M_\mu(\lambda) \leq \lambda x - \log M_\mu(\lambda)$ Thus, using (iii) we have that:

$$I(y) = \sup_{\lambda \leq 0} (\lambda y - \log M_\mu(\lambda)) \leq \sup_{\lambda \leq 0} (\lambda x - \log M_\mu(\lambda)) = I(x).$$

Then, I is decreasing in $(-\infty, \bar{x})$. Using the same argument, we can prove that I is increasing in $(\bar{x}, +\infty)$.
(v) Since I is lower semi-continuous, the set $I^{-1}[0,c]$ is closed for every $c \geq 0$ from the definition of lower semi-continuity. It suffices to show that $I^{-1}[0,c]$ is bounded. Since μ has exponential moments, there is $\epsilon > 0$ so that $M(-\epsilon) \vee M(\epsilon) < +\infty$. Also, if $I(x) \leq c$ then

$$(\epsilon x - \Lambda(\epsilon)) \vee (-\epsilon x - \Lambda(-\epsilon)) \leq I(x) \leq c$$

Thus,

$$I^{-1}[0,c] \subseteq \left[-\frac{c + \Lambda(-\epsilon)}{\epsilon}, \frac{c + \Lambda(\epsilon)}{\epsilon}\right] \subseteq \mathbb{R},$$

which means that I is bounded. $\square$

Remark 1.9. *Analogues properties hold, if we are on $\mathbb{R}^d$ or in a Polish space as we will see later.*

Chapter 2

LDP for finite dimensional spaces

2.1 Cramér theorem in $\mathbb{R}$

Assume that $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with common distribution an arbitrary distribution μ and $I:\mathbb{R}\to[0,\infty]$ the rate function of μ. Let μ^n on $\mathbb{R}$ denote the n-fold tensor product of μ with itself. Next, let μ_n on $\mathbb{R}$ denote the distribution of $\frac{1}{n}\sum_{i=1}^n x_i$ under μ^n. We denote $S_n := \frac{1}{n}\sum_{i=1}^n X_i$. Assuming that $\int_{\mathbb{R}} |x|\mu d(x) < \infty$ the Weak Law of Large Numbers implies that $\mu_n \Rightarrow \delta_{\bar{x}}$.

Theorem 2.1. *(Cramér) Assume that $M(\lambda) < \infty$ for every $\lambda\in\mathbb{R}$. Then for every measurable $A\subset\mathbb{R}$ one has that:*

$$-I(x) \leq \liminf_{n\to\infty}\frac{1}{n}\log\mu_n(A) \leq \limsup_{n\to\infty}\frac{1}{n}\log\mu_n(A) \leq -I(x)$$

where $I(x) = \sup_\lambda[\lambda x - \log M(\lambda)]$.

Proof. Upper Bound

Let $x\geq\bar{x}$. Applying Chebyshev inequality, one has that $\forall\lambda\geq 0$:

$$\mu_n([x,\infty)) = \mathbb{P}[S_n\geq nx] \leq e^{-\lambda nx}\mathbb{E}\left(e^{\lambda S_n}\right) = e^{-\lambda nx}M(\lambda)^n = e^{-\lambda nx}e^{\log M(\lambda)^n}$$
$$= e^{-n(\lambda x-\log M(\lambda))} = e^{-n\sup_{\lambda\geq 0}(\lambda x-\log M(\lambda))}$$

optimizing over λ we get the last equality. Thus, taking logarithm in both sides and dividing by n, we have that $\forall\lambda\in\mathbb{R}$:

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}[S_n\geq nx] \leq -(\lambda x-\log M(\lambda))$$
$$\leq \inf_{\lambda\in\mathbb{R}}(-(\lambda x-\log M(\lambda))) = -\sup_{\lambda\in\mathbb{R}}(\lambda x-\log M(\lambda)).$$

However,

$$-\sup_{\lambda\geq 0}(\lambda x-\log M(\lambda)) = -\sup_{\lambda}(\lambda x-\log M(\lambda))$$

Indeed, if $\lambda > 0$, applying Jensen's inequality :

$$\log\int e^{\lambda x}d\mu(x) \geq \int\log e^{\lambda x}d\mu(x) = \int\lambda x d\mu(x) = \lambda\bar{x} \geq \lambda x$$

That is, if $\lambda \leq 0$ then $\log M(\lambda) \geq \lambda x$. Because 0 is always a trivial lower bound replacing $\sup_{\lambda \geq 0}(\lambda x - \log M(\lambda))$ by $\sup_{\lambda}(\lambda x - \log M(\lambda))$ does not increase its value.
Then,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \geq nx] \leq -I(x)$$

Lower Bound

For every $\delta \geq 0$ and $x \geq \bar{x}$, we will show that:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n([x, x+\delta)) = \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}([nx \leq S_n < nx + n\delta)) \geq -I(x)$$

Let $A = [x, x+\delta)$. We consider the case the supremum on I is attained at a finite λ. That is, we suppose there exists λ^* finite such that,

$$I(x) = \sup_{\lambda}(\lambda x - \log M(\lambda)) = \lambda^* x - \log M(\lambda^*).$$

Since $M(\lambda^*) < \infty$,

$$x = \frac{M'(\lambda^*)}{M(\lambda^*)} = \int y \frac{e^{\lambda^* y}}{M(\lambda^*)} d\mu(y)$$

If we now define a new probability distribution $\tilde{\mu}$ by the relation:
$d\tilde{\mu}(y) = \frac{e^{\lambda^*(y)}}{M(\lambda^*)} d\mu(y)$, then $\tilde{\mu}$ has x as its expected value.

$$\begin{aligned}
\mathbb{P}\left[x \leq \frac{S_n}{n} < x+\delta\right] &= \mu^{\otimes n}\left(S_n^{-1}[nx, nx+n\delta]\right) = \int \dots \int_{nx \leq \sum x_i \leq n(x+\delta)} d\mu(x_1)\, d\mu(x_2) \dots d\mu(x_n) \\
&\geq \int \dots \int_{nx \leq S_n \leq nx+n\epsilon} d\mu(x_1)\, d\mu(x_2) \dots d\mu(x_n) \; \forall \epsilon > 0, \quad \delta > \epsilon > 0 \\
&\geq \int \dots \int_{nx \leq S_n \leq nx+n\epsilon} e^{\lambda^* S_n(y) - n\lambda^* x - n|\lambda^*|\epsilon} d\tilde{\mu}(x_1) \dots d\tilde{\mu}(x_n) \\
&= e^{-n\lambda^* x - n|\lambda^*|\epsilon} \int \dots \int_{nx \leq S_n \leq nx+n\epsilon} e^{\lambda^* S_n(y)} d\tilde{\mu}(x_1) \dots d\tilde{\mu}(x_n) \\
&= e^{-n\lambda^* x - n|\lambda^*|\epsilon} \int \dots \int_{nx \leq S_n \leq nx+n\epsilon} e^{\lambda^*(x_1 + \dots + x_n)} d\tilde{\mu}(x_1) \dots d\tilde{\mu}(x_n) \\
&= e^{-n\lambda^* x - n|\lambda^*|\epsilon} \int \dots \int e^{\lambda^*(x_1 + \dots + x_n)} \mathbb{1}_{[nx \leq S_n \leq nx+n\epsilon]} d\tilde{\mu}(x_1) \dots d\tilde{\mu}(x_n) \\
&= e^{-n\lambda^* x - n|\lambda^*|\epsilon} M^n(\lambda^*)\, \tilde{\mathbb{P}}\left[x \leq \frac{S_n}{n} \leq x+\epsilon\right] = e^{-n\lambda^* x - n|\lambda^*|\epsilon} M^n(\lambda^*)\, \tilde{\mathbb{P}}\left[0 \leq \frac{S_n}{n} - x \leq \epsilon\right].
\end{aligned}$$

Hence,

$$\frac{1}{n} \log \mathbb{P}[nx \leq S_n \leq nx + n\delta] \geq -\lambda^* x - |\lambda^*|\epsilon + \log M(\lambda^*) + \frac{1}{n} \log \tilde{\mathbb{P}}\left[0 \leq \frac{S_n}{n} - x \leq \epsilon\right]$$

where $\tilde{\mathbb{P}} := \tilde{\mu}^{\otimes\infty}$ in the measurable product space $\left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}}^{\mathbb{N}}, \tilde{\mathbb{P}}\right)$.
Then, by Central Limit Theorem:

$$\tilde{\mathbb{P}}\left(0 \leq \frac{S_n}{n} - x \leq \epsilon\right) = \tilde{\mathbb{P}}\left(0 \leq \frac{1}{\tilde{\sigma}\sqrt{n}}(S_n - x) \leq \frac{\epsilon\sqrt{n}}{\tilde{\sigma}}\right) = \int_0^\infty e^{-\frac{x^2}{2} \frac{dx}{\sqrt{2\pi}} = \frac{1}{2}}.$$

Therefore since ϵ was arbitrarily chosen,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}[nx \leq S_n \leq nx + n\delta] \geq -\lambda^* x - |\lambda^*|\epsilon + \log M(\lambda^*) = -\sup_{\lambda}[\lambda x - \log M(\lambda^*)]$$

for every $\epsilon > 0$.
Note that the assumption $\mathcal{D}_{M_\mu}$ is not necessary, but we make it for shake of simplicity. An analogous argument can be used in the case when $x \leq \bar{x}$. We refer the interested reader in [4] Theorem 2.2.3 p.27
We also refer in [4] the same chapter, in case that $\Lambda(\lambda)$ doesn't attain at a finite λ^*.

□

Remark 2.1. *We should take note of the structure of the preceding lines of reasoning. Namely, the upper bound comes from optimizing over a family of Chebyshev inequalities, while the lower bound comes from introducing a RADON-NICODYM measure in order to make what was originally "deviant" behavior look like typical behavior. This pattern of proof is recurrent in the theory of Large Deviations. In particular, it will be used extendedly in the following sections.*

2.2 Cramér theorem in $\mathbb{R}^d$

Now, we will extend the Cramér theorem on $\mathbb{R}^d$. Assume that $\{X_i\}_{i\in\mathbb{N}}$ i.i.d. random vectors with common distribution μ and μ_n the distribution of their arithmetic mean $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n X_i$. We denote $\langle\cdot,\cdot\rangle$ the Euclidean inner product in $\mathbb{R}^d$. The same properties for the rate function and the logarithmic generating function holds as in Cramér theorem on $\mathbb{R}$. We give the basic definitions.

Definition 2.1. *Suppose that μ is a distribution on $\mathbb{R}^d$. The moment generating function of μ is the function $M_\mu : \mathbb{R}^d \to [0,\infty]$*

$$M_\mu(x) = \int e^{\langle\lambda,x\rangle} d\mu(x).$$

The set $\mathcal{D}_{M_\mu} = \left\{x \in \mathbb{R}^d : M_\mu(x) < \infty\right\}$ is called essential range of M_μ.

Remark 2.2. *We denote by $\Lambda_\mu = \log M_\mu$ the logarithmic generating function and by I_μ the function that:*

$$I_\mu(x) = \sup_{\lambda\in\mathbb{R}^d} \left(\langle\lambda,x\rangle - \Lambda(\lambda)\right)$$

*That, I_μ is the **Legendre transform** of Λ_μ, and we call I_μ the rate function of μ.*

Next, we state the basic properties of rate functions.

Proposition 2.1. *Let μ be distribution on $\mathbb{R}^d$ with exponential moments and mean $\bar{x}$. Suppose that I, M, Λ are the rate function, the moment generating function and the logarithmic generating function of μ, respectively. Then,*

(i) I is convex, lower semi-continuous and non-negative.

(ii) $I(x) = 0$ if and only if $x = \bar{x}$

(iii) For each $c \geq 0$ the level sets $I^{-1}[0,c] = \left\{x \in \mathbb{R}^d : I(x) \leq c\right\}$ are compact.

Proof. The argument that are used to prove the above properties are the same as in the case for $d = 1$, so we omit the proof. $\square$

Next, we prove the Cramér theorem in the case that the moment generating function of the common distribution μ of $\{X_i\}_{i\in\mathbb{N}}$ i.i.d. has essential range $\mathcal{D}_{M_\mu} = \mathbb{R}^d$.

Theorem 2.2. *(Cramér Theorem on $\mathbb{R}^d$)*
For every closed $F \subseteq \mathbb{R}^d$,

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

and for every open $U \subseteq \mathbb{R}^d$,

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

Before proceeding with the proof of Theorem 2.2 we will introduce two definitions that are useful in order to prove the upper bound.

Definition 2.2. *(Weak Large Deviation Principle)*
We say that the family $\{\mu_n : n \in \mathbb{N}\}$ satisfies the WLDP. If I is a rate function and $\{\mu_n : n \in \mathbb{N}\}$ satisfies:
For all open sets $U \subset \mathbb{R}^d$

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

and for all compact sets K in $\mathbb{R}^d$

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

The passage from a Weak to a full Large Deviation Principle is often accomplished by an application of the following observation:

Definition 2.3. *(exponentially tightness)*
Let $\{\mu_n : n \in \mathbb{N}\}$ a family of measures and assume that, for each $L \geq 0$ there exists a compact set K_L with the property that:

$$\mu_n[K_L^c] \leq e^{-nL} \Leftrightarrow \limsup_{n\to\infty} \frac{1}{n} \log \mu_n[K_L^c] \leq L$$

We say that the family $\{\mu_n : n \in \mathbb{N}\}$ is exponentially tight.

Lemma 2.1. *If $\{\mu_n : n \in \mathbb{N}\}$ a family of measures satisfies weak LDP and exponentially tightness property then it satisfies the full LDP.*

Proof. In order to prove that $\{\mu_n\}_{n\in\mathbb{N}}$ satisfies full LDP we have to prove that for any closed set $K \subset \mathbb{R}^d$

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

First of all we observe that for every closed $F \subset \mathbb{R}^d$ and every compact $K \subset \mathbb{R}^d$ we have that $F = (F \cap K) \cup (F \setminus K) \subset (F \cap K) \cup K^c$. So,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n (F) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mu_n (F \cap K) \vee \limsup_{n\to\infty} \frac{1}{n} \log \mu_n (K^c) \qquad (2.1)$$

$$\leq \left(-\inf_{x\in F\cap K} I\left(x\right)\right) \vee \left(-L\right) \leq -\inf_{x\in F} I\left(x\right) \vee \left(-L\right).$$

So, if $\inf_{x\in F} I\left(x\right) = 0 \Rightarrow \log \mu_n\left(F\right) \leq 0$ for all $n \in \mathbb{N}$ and if $\inf_{x\in F} I\left(x\right) = \infty$ from the exponentially tightness, choosing an appropriate compact set $K \subset \mathbb{R}^d$ we make the right side of 2.1 as small as we want and $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n\left(F\right) = -\infty = -\inf_{x\in F} I\left(x\right)$. Finally if $0 < \inf_{x\in F} I\left(x\right) < \infty$ then by the exponentially tightness we can choose a compact set $K \subset \mathbb{R}^d$ such that $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n\left(F\right) < -\inf_{x\in F} I\left(x\right)$ and for this K we get a propper upper bound for F. The proof is completed.

□

Remark 2.3. *Exponential tightness is the LD analogue of tightness in weak convergence.*

Proof. Upper bound

Firstly,we will prove a proper upper bound for open balls. Then, we will cover compact sets with a finite number of open balls. This will give us an upper bound for compact sets. Finally, we will use the above definition to pass from compact to closed sets.

More precisely,

First Step:

we will prove that for each$x \in \mathbb{R}^d$ and for each $\epsilon > 0$ there exists $\delta = \delta_{x,\epsilon} > 0$ such that:

$$\frac{1}{n} \log \mu_n[B\left(x,\delta\right)] \leq -I^\epsilon\left(x\right) = \left(I\left(x\right) - \epsilon\right) \wedge \frac{1}{\epsilon}$$

We included the term $\frac{1}{\epsilon}$ to simultaneously treat the cases $I\left(x\right) < \infty$ and $I\left(x\right) = \infty$.

Suppose, now, that $x \in \mathbb{R}^d$, $\epsilon > 0$ and observe that for each $\delta > 0$ and $\lambda \in \mathbb{R}^d$, if $\frac{S_n}{n} \in B\left(x,\delta\right)$ then

$$\langle \lambda, S_n \rangle \geq n \inf_{y\in B(x,\delta)} \langle \lambda, y \rangle$$

Thus,

$$\mu_n[B\left(x,\delta\right)] = \mathbb{E}[\mathbb{1}_{\frac{S_n}{n}\in B(x,\delta)}] \leq \mathbb{E}[e^{\langle \lambda, S_n\rangle - n\inf_{y\in B(x,\delta)}\langle \lambda, y\rangle}] \leq e^{-n\inf_{y\in B(x,\delta)}\langle \lambda, y\rangle}\mathbb{E}[e^{\langle \lambda, S_n\rangle}]$$
$$= e^{-n\inf_{y\in B(x,\delta)}\langle \lambda, y\rangle} M^n\left(\lambda\right) = e^{-n\inf_{y\in B(x,\delta)}\langle \lambda, y\rangle} e^{n\log M(\lambda)}$$
$$= e^{-n\left(\inf_{y\in B(x,\delta)}\langle \lambda, y\rangle - \Lambda(\lambda)\right)}$$

Notice that $y \in B\left(x,\delta\right) \Rightarrow |\langle \lambda, y - x\rangle| \leq |\lambda|\delta \Rightarrow \langle \lambda, y\rangle \geq \langle \lambda, x\rangle - |\lambda|\delta$.

Therefore,

$$\frac{1}{n} \log \mu_n[B\left(x,\delta\right)] \leq -\inf_{y\in B(x,\delta)} \left(\langle \lambda, y\rangle - \Lambda\left(\lambda\right)\right) \leq -\left(\langle \lambda, x\rangle + |\lambda|\delta - \Lambda\left(\lambda\right)\right)$$

From the definition of I there exists $\lambda_{x,\epsilon} > 0$ such that:

$$\langle \lambda_{x,\epsilon}, x\rangle - \Lambda\left(\lambda_{x,\epsilon}\right) > \left(I\left(x\right) - \frac{\epsilon}{2} \wedge \frac{1}{\epsilon}\right)$$

Then choosing $\delta_{x,\epsilon} > 0$ such that $|\lambda_{x,\epsilon}|\delta_{x,\epsilon} < \frac{\epsilon}{2}$

$$\mu_n[B\left(x,\delta_{x,\epsilon}\right)] \leq e^{n|\lambda_{x,\epsilon}|\delta_{x,\epsilon}} e^{-n|I(x)-\frac{\epsilon}{2}|\wedge\frac{1}{\epsilon}} \leq e^{-nI(x)\wedge\frac{1}{\epsilon}}$$

Hence, for every $x \in \mathbb{R}^d$, we have found a $\delta_{x,\epsilon} > 0$ such that

$$\frac{1}{n} \log \mu_n[B\left(x,\delta_{x,\epsilon}\right) \leq -\left(I\left(x\right) - \epsilon\right) \wedge \frac{1}{\epsilon} = -I^\epsilon\left(x\right),$$

for each $n \in \mathbb{N}$
second step
Suppose that $K \subseteq \mathbb{R}^d$ is compact set. Then, there exists $N = N(K, \epsilon) \in \mathbb{N}$ and $x_1 \dots x_N \in K$ such that: $K \subseteq \bigcup_{i=1}^{N} B(x_i, \delta_{x_i,\epsilon})$.

$$\mu_n(K) \le \mu_n\left(\bigcup_{i=1}^{N} B(x_i, \delta_{x_i,\epsilon})\right) \le \sum_{i=1}^{N} \mu_n(B(x_i, \delta_{x_i,\epsilon})) \le N \max_{1\le i\le n} \mu_n(B(x_i, \delta_{x_i,\epsilon})).$$

So,

$$\frac{1}{n}\log \mu_n(K) \le \frac{1}{n}\log\left(N \max_{1\le i\le n} \mu_n\left(B\left(x_i, \delta_{x_{i,\epsilon}}\right)\right)\right) = \frac{1}{n}\log \max_{1\le i\le n} N + \frac{1}{n}\log \mu_n(B(x_i, \delta_{x_i,\epsilon}))$$

and taking $n \to \infty$ we get

$$\limsup_{n\to\infty} \frac{1}{n} \max_{1\le i\le n} \log \mu_n(K) \le \limsup_{n\to\infty} \frac{1}{n} \max_{1\le i\le n} \log \mu_n(B(x_i, \delta_{x_i,\epsilon})) \le$$

$$\max_{1\le i\le n} (-I^\epsilon(x_i)) \le \sup_{x\in K} (-I^\epsilon(x)) = -\inf_{x\in K} I^\epsilon(x),$$

for each $\epsilon > 0$.
Then, we make the following observation:

$$\begin{aligned}
\lim_{\epsilon\to 0} \inf_{x\in K} I(x) &= \lim_{\epsilon\to 0} \min\left\{\inf_{x\in K\cap D_I} I^\epsilon(x), \frac{1}{\epsilon}\right\} \\
&= \lim_{\epsilon\to 0} \inf_{x\in K\cap D_I} I^\epsilon(x) \\
&= \inf_{x\in K\cap D_I} I(x) \\
&= \inf_{x\in K} I(x)
\end{aligned}$$

Because I^ϵconverges uniformly to I on D_I when $\epsilon \to 0$.
Thus,

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

for every compact set $K \subseteq \mathbb{R}^d$.
Third step

Now, due to lemma 2.1 we will pass from weak to full LDP proving exponentially tightness.

Firstly, we observe that for every closed set $F \subset \mathbb{R}^d$ and compact $K \subset \mathbb{R}^d$:

$$F = (F \cap K) \cup (F \setminus K) \subseteq (F \cap K) \cup K^c$$

Then,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(F) &\le \limsup_{n\to\infty} \frac{1}{n}\log \mu_n(F\cap K) \vee \limsup_{n\to\infty} \frac{1}{n}\log \mu_n(K^c) \\
&\le -\inf_{x\in K\cap F} I(x) \vee \limsup_{n\to\infty} \frac{1}{n}\log \mu_n(K^c)
\end{aligned}$$

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

Thus, it suffices to show that $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight.
Suppose that $M > 0$ and define:

$$K_t := [-t,t]^d \Rightarrow K_t^c = \left([-t,t]^d\right)^c = \cup_{j=1}^d \left\{x : |x^j| > t\right\}$$

where $x^j : \mathbb{R}^d \to \mathbb{R}$ are the projections of x. Then, the union of events bound yields:

$$\mu_n(K_t^c) = \mu_n\left(\cup_{j=1}^d \left\{|x^j| > t\right\}\right) \leq \sum_{j=1}^d \mu_n^j([t,\infty)) + \sum_{j=1}^d \mu_n^j((-\infty,-t])$$

where μ_n^j , $j = 1,\dots,n$ are the laws of the coordinates of the random vector S_n and $S_j^n = \frac{1}{n}\sum_{i=1}^n X_i^j \left\{X_i^j\right\}_{i\in\mathbb{N}}$ for $j = 1,\dots,n$ i.i.d random vectors.
Thus, by Cramér Theorem on $\mathbb{R}$:
$\mu_n^j([t,\infty)) \leq e^{-n\Lambda_j^*(t)}$ and $\mu_n^j((-\infty,-t]) \leq e^{-n\Lambda_j^*(-t)}$
where Λ_j^* is the Legendre transform of $\log \mathbb{E}[e^{\lambda X_i^j}]$ for $j = 1,\dots,d$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_t^c) \leq \max_{j=1,\dots,d} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n^j([-t,t]^c)$$

$$\leq \max_{j=1,\dots,d} (-I_j(t) \wedge -I_j(-t))$$

$$\leq -\min_{j=1,\dots,d} (I_j(t) \wedge I_j(-t))$$

As $|t| \to \infty$ then $\lim_{|t|\to\infty} I_j(t) = \lim_{|t|\to\infty} (\sup_{t\in\mathbb{R}} \{\lambda t - \Lambda(\lambda)\}) = \infty$
Then, there exists t_0 for each $t > t_0$ such that:

$$\min(I_j(t) \wedge I_j(-t)) > M \Rightarrow \limsup_{n\to\infty} \frac{1}{n} \log \mu_n^j\left(K_{t_0}^c\right) \leq -M$$

Consequently, $\{\mu_n\}$ is a an exponentially tight sequence of probability measures, since the hypercubes K_t are compact.

Lower bound
We take into account two cases, in the first case we suppose that the supremum in the definition of I is attained at a finite λ, whereas in the second case we suppose that the supremum is not achieved in a finite λ.

First case
Suppose that $x \in \mathbb{R}^d$ thus there exists $\rho > 0$ such that $B(x,\rho) \subset U$, where $B(x,\rho)$ is a ball of radius ρ. Then it suffices to show that:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(B(x,\rho)) \geq -\inf I(x)$$

where $I(x) := \sup_{\lambda\in\mathbb{R}^d} \{\langle\lambda, x\rangle - \log M(\lambda)\}$.
It follows from our assumption that there exists a solution λ^* to the saddle-point equation $\Lambda(\lambda)$:

$$x = \nabla\Lambda(\lambda^*) \Leftrightarrow x = \frac{\nabla M(\lambda^*)}{M(\lambda^*)} \Leftrightarrow x = \int y \frac{e^{\langle\lambda^*, y\rangle}}{M(\lambda^*)} d\mu(y) \Leftrightarrow x = \int y d\tilde{\mu}(y)$$

where $d\tilde{\mu}(y) = \frac{e^{\langle \lambda^*, y\rangle}}{M(\lambda^*)} d\mu(y)$.
Define $S_n : \mathbb{R}^d \times \mathbb{R}^d \times \ldots \times \mathbb{R}^d \to \mathbb{R}^d$ such that $S_n(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$ where $x_i, i = 1, 2, \ldots n$ random vectors.
and the product measure: $d\tilde{\mu}^{\otimes n}(y) = \frac{e^{\langle \lambda^*, S_n(x)\rangle}}{M(\lambda^*)} d\mu^{\otimes n}$.
Then observe that for $\frac{S_n}{n} \in B(x, \epsilon)$:

$$\langle \lambda^*, S_n(x)\rangle = n\left\langle \lambda^*, \frac{S_n}{n} + x - x\right\rangle = n\langle \lambda^*, x\rangle + n\left\langle \lambda^*, \frac{S_n}{n} - x\right\rangle \leq n\langle \lambda^*, x\rangle + n|\lambda^*|\epsilon,$$

so

$$e^{-\langle \lambda^*, S_n(x)\rangle} \geq e^{-(n\langle \lambda^*, x\rangle + n|\lambda^*|\epsilon)}$$

Suppose, now that $0 < \epsilon < \rho$. Therefore,

$$\begin{aligned}
\mu_n[B(x,\rho)] \geq \mu_n[B(x,\epsilon)] = \mathbb{P}[\frac{S_n}{n} \in B(x,\epsilon)] &= \mu^{\otimes n}\left\{y \in \mathbb{R}^d \times \ldots \mathbb{R}^d, \frac{S_n(y)}{n} \in B(x,\epsilon)\right\} \\
&= \int \cdots \int_{\left\{\frac{S_n}{n} \in B(x,\epsilon)\right\}} d\mu(y) \ldots d\mu(y) \\
&\geq \int \cdots \int_{\left\{\frac{S_n}{n} \in B(x,\epsilon)\right\}} e^{-\langle \lambda^*, S_n(x)\rangle} M^n(\lambda^*)\, d\tilde{\mu}(y) \ldots d\tilde{\mu}(y) \\
&= M^n(\lambda^*) \int \cdots \int_{\left\{\frac{S_n}{n} \in B(x,\epsilon)\right\}} e^{-\langle \lambda^*, S_n(x)\rangle} d\tilde{\mu}(y) \ldots d\tilde{\mu}(y) \\
&\geq M^n(\lambda^*) \int \cdots \int_{\left\{\frac{S_n}{n} \in B(x,\epsilon)\right\}} e^{-n(\langle \lambda^*, x\rangle + n|\lambda^*|\epsilon)} d\tilde{\mu}(y) \ldots d\tilde{\mu}(y) \\
&= M^n(\lambda^*)\, e^{-n|\lambda^*|\epsilon - n\langle \lambda^*, x\rangle} \int \cdots \int_{\left\{\frac{S_n}{n} \in B(x,\epsilon)\right\}} d\tilde{\mu}(y) \ldots d\tilde{\mu}(y).
\end{aligned}$$

Taking logarithms and dividing by n we have

$$\begin{aligned}
\frac{1}{n}\log \mu_n[B(x,\rho)] &\geq \log M(\lambda^*) - |\lambda^*|\epsilon - \langle \lambda^*, x\rangle + \frac{1}{n}\log \tilde{\mu}^{\otimes n}\left[\frac{S_n}{n} \in B(x,\epsilon)\right] \\
&= -I(x) - |\lambda^*|\epsilon + \frac{1}{n}\log \tilde{\mu}^{\otimes n}[\frac{S_n}{n} \in B(x,\epsilon)]
\end{aligned}$$

We are in the product space: $\left(\mathbb{R}^d \times \ldots \times \mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right) \otimes \ldots \otimes \mathcal{B}\left(\mathbb{R}^d\right), \mathbb{P}^*\right)$,
where: $\mathbb{P}^* := \tilde{\mu}^{\otimes\infty}$. We define $p^i : \mathbb{R}^d \times \ldots \mathbb{R}^d \to \mathbb{R}^d$ i.i.d. random vectors with common distribution $\tilde{\mu}$. Thus, $\tilde{S_n} : \Omega \to \mathbb{R}^d$ such that $\tilde{S_n} = S_n \circ \left(p^1, \ldots, p^n\right) = \sum_{i=1}^n p^i$, for every $n \in \mathbb{N}$ and

$$\mathbb{P}^*\left(\frac{\tilde{S_n}}{n} \in B(x,\epsilon)\right) = \tilde{\mu}^{\otimes n}\left(\frac{S_n}{n} \in B(x,\epsilon)\right).$$

Hence, by the Weak Law of Large Numbers: $\frac{\tilde{S_n}}{n} \to x \in \mathbb{R}^d$ $\mathbb{P}$-a.s.
This means that,

$$\liminf_{n\to\infty} \frac{1}{n}\log \tilde{\mu}^{\otimes n}\left[\frac{S_n}{n} \in B(x,\epsilon)\right] = \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}^*\left[\frac{\tilde{S_n}}{n} \in B(x,\epsilon)\right]$$

therefore,

$$\liminf_{n\to\infty} \frac{1}{n}\log \mu_n[B(x,\epsilon)] \geq -I(x) - |\lambda^*|\epsilon = I^\epsilon(x).$$

Now, if let $\epsilon \to 0$ then $I^{\epsilon}(x) \to I(x)$.
Second case
Now, we consider the case the supremum of $I(x)$ is not attained at any finite λ. More precisely for $x \in D_I \nabla \Lambda\left(\mathbb{R}^d\right)$.
We regularize the common distribution μ of $\{X_i\}$ with an appropriate family of Gaussian measures, $\mu_\epsilon := \mu * N\left(0, \epsilon^2 I\right)$ in order to apply the previous case.

$$X_i^{\epsilon} = X_i + \epsilon Y_i \Rightarrow S_n^{\epsilon} = \sum_{i=1}^{n} X_i + \epsilon \sum_{i=1}^{n} Y_i$$

where Y_i is an i.i.d. sequence of normal distributed vectors independent of X_i. We shall prove that the supremum of $I_{\mu_\epsilon}(x)$ is achieved in $\lambda_\epsilon \in \mathbb{R}^d$.
Suppose that there exists $\{\lambda_n\} \subseteq \mathbb{R}^d$ such that:

$$I_{\mu_\epsilon}(x) = \lim_{n\to\infty} \left(\langle \lambda_n, x\rangle - \Lambda_{\mu_\epsilon}(\lambda_n)\right)$$

When $X \sim N\left(0, \epsilon^2 I\right) \Rightarrow \mathbb{E}\left(e^{\lambda X}\right) = e^{\frac{|\lambda|^2}{2}} \Rightarrow \log M(\lambda) = \frac{|\lambda|^2}{2}$, and $\Lambda_{\mu_\epsilon}(\lambda) = \Lambda_\mu(\lambda) + \frac{\epsilon^2 |\lambda_n|^2}{2}$
Then,

$$I_{\mu_\epsilon}(x) = \lim_{n\to\infty} \left(\langle \lambda_n, x\rangle - \Lambda_\mu(\lambda_n) - \frac{\epsilon^2 |\lambda_n|^2}{2}\right)$$

and $\{\lambda_n\}$ is bounded.
Indeed, if $\{\lambda_n\}$ were not bounded then there would exist $\{\lambda_{k_n}\}$ such that $|\lambda_{k_n}| \to \infty$. Since $\Lambda_\mu(\lambda) > \langle \lambda, \bar{x}\rangle$ for every $\lambda \in \mathbb{R}^d$ we would have

$$0 \le I_{\mu_\epsilon}(x) = I_{\mu_\epsilon}(x) = \lim_{n\to\infty} \left(\langle \lambda_{k_n}, x\rangle - \Lambda_\mu(\lambda_{k_n}) - \frac{\epsilon^2 |\lambda_{k_n}|^2}{2}\right) \le \limsup_{n\to\infty} \left(\langle \lambda_{k_n}, x - \bar{x}\rangle - \frac{\epsilon^2 |\lambda_{k_n}|^2}{2}\right)$$

$$\le \limsup_{n\to\infty} \left(|\lambda_{k_n}||x - \bar{x} - \frac{\epsilon^2 |\lambda_{k_n}|^2}{2}\right) = -\infty$$

which is a contradiction. Then, $\{\lambda_n\}$ is bounded and there exists a subsequence which converges to λ_ϵ and $I_{\mu_\epsilon}(x) = \langle \lambda_\epsilon, x\rangle - \Lambda_{\mu_\epsilon}(\lambda_\epsilon)$,obviously $x \in \nabla \Lambda_{\mu_\epsilon}\left(\mathbb{R}^d\right)$. As a consequence, from the previous case,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n^{\epsilon}}{n} \in B(x, \rho)\right] \ge -I_{\mu}\epsilon(x) \ge -I(x),$$

for every $\rho > 0$.
We set $T_n := \sum_{i=1}^{n} Y_i$, then: $T_n \sim N(0, nI) \Rightarrow \frac{T_n}{\sqrt{n}} \sim N(0,1)$. We make the following observation that:

$$\left\{\frac{S_n^{\epsilon}}{n} \in B(x, \rho)\right\} \subset \left\{\frac{S_n}{n} \in B(x, 2\rho)\right\} \cup \left\{\epsilon |\frac{T_n}{n}| > \rho\right\}$$

More precisely,

$\left|\frac{S_n}{n} - x\right| \le \left|\frac{S_n^{\epsilon}}{n} - x\right| + \epsilon \left|\frac{T_n}{n}\right| \Rightarrow \left|\frac{S_n}{n} - x\right| \le \rho + \epsilon \left|\frac{T_n}{n}\right|$, since $\frac{S_n^{\epsilon}}{n} \in B(x, \rho)$.
If $\frac{S_n}{n} \notin B(x, 2\rho)$ then $2\rho < \left|\frac{S_n}{n} - x\right| \le \rho + \epsilon \left|\frac{T_n}{n}\right| \Rightarrow \left|\frac{T_n}{n}\right| \ge \frac{\rho}{\epsilon}$.
Hence,

$$\mathbb{P}\left[\frac{S_n^{\epsilon}}{n} \in B(x, \rho)\right] \le \mathbb{P}\left[\frac{S_n}{n} \in B(x, \rho)\right] + \mathbb{P}\left[\epsilon \left|\frac{T_n}{n}\right| > \rho\right].$$

We want to show that $\mathbb{P}\left[\epsilon|\frac{T_n}{n}| > \rho\right]$ is quite small so as not to contribute to our computation. By this we mean that:

$$-I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n^\epsilon}{n} \in B(x,\rho)\right] \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\frac{S_n}{n} \in B(x,2\rho)\right] \vee \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\epsilon\left|\frac{T_n}{n}\right| > \rho\right]$$

$$\leq \liminf_{n\to\infty} \frac{1}{n} \log\left(\mu_n[B(x,2\rho)]\right) \vee \mathbb{P}\left[\epsilon\left|\frac{T_n}{n}\right| > \rho\right] = \liminf_{n\to\infty} \frac{1}{n} \log \mu_n\left[B(x,2\rho)\right],$$

since we will prove that:

$$\mathbb{P}\left[\epsilon\left|\frac{T_n}{n}\right| > \rho\right] < \liminf_{n\to\infty} \frac{1}{n} \log \mu_n\left[B(x,2\rho)\right].$$

More precisely,

$$\mathbb{P}\left[\epsilon\left|\frac{T_n}{n}\right| > \rho\right] = \mathbb{P}\left[\left|\frac{T_n}{n}\right| > \frac{\rho}{\epsilon}\right] = \mathbb{P}\left[\left|\frac{T_n}{\sqrt{n}}\right| > \frac{\rho\sqrt{n}}{\epsilon}\right] = \int_{\mathbb{R}^d\setminus B\left(0,\frac{\rho\sqrt{n}}{\epsilon}\right)}^{\infty} e^{-\frac{|x|^2}{2}} dx$$

So, we have to compute:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left[\epsilon\left|\frac{T_n}{n}\right| > \rho\right] = \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}^d\setminus B\left(0,\frac{\rho\sqrt{n}}{\epsilon}\right)}^{+\infty} e^{-\frac{|x|^2}{2}} dx = -\frac{\rho^2}{2\epsilon} \tag{2.2}$$

By this we mean that if one of the above integrals exists then the other exists too and they are equal.

For $d = 1$ the statement holds. Since

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}\setminus\left(-\frac{\rho\sqrt{n}}{\epsilon},\frac{\rho\sqrt{n}}{\epsilon}\right)} e^{-\frac{|x|^2}{2}} dx = \lim_{n\to\infty} \frac{1}{n} \log \int_{\frac{\rho\sqrt{n}}{\epsilon}}^{+\infty} e^{-\frac{|x|^2}{2}} dx$$

Using the following inequalities

$$\frac{x}{1+x^2} e^{\frac{-x^2}{2}} \leq \int_x^{+\infty} e^{-\frac{|y|^2}{2}} dy \leq \frac{1}{x} e^{-\frac{x^2}{2}}, \quad x > 0 \tag{2.3}$$

and the $\lim_{n\to\infty} \frac{\log x}{x} = 0$ then the 2.2 holds.

For $d = 2$, we use polar coordinates thus

$$\int_{\mathbb{R}^2\setminus B(0,r_0)} e^{-\frac{|x|^2}{2}} dx = 2\pi \int_{r_0}^{\infty} re^{-\frac{r^2}{2}} dr = 2\pi e^{-\frac{r_0^2}{2}}.$$

Therefore

$$\frac{1}{n} \log \int_{\mathbb{R}^2\setminus B\left(0,\frac{\rho\sqrt{n}}{\epsilon}\right)} e^{-\frac{|x|^2}{2}} dx = \frac{\log 2\pi}{n} - \frac{\rho^2}{2\epsilon^2} \to -\frac{\rho^2}{2\epsilon^2} \quad \text{as} \quad n \to \infty.$$

Now, we pass in the d-dimension using once again polar coordinates. Indeed,

$$\int \cdots \int_{\mathbb{R}^d\setminus B\left(0,\frac{\rho\sqrt{n}}{\epsilon}\right)} e^{-\frac{|x|^2}{2}} ds = \int_{S^{d-1}} d\sigma \int_{\frac{\rho\sqrt{n}}{\epsilon}}^{\infty} dre^{-\frac{r^2}{2}} r^{d-1}$$

$$= |S^{d-1}| \int_{\frac{\rho\sqrt{n}}{\epsilon}}^{\infty} dre^{-\frac{\rho^2}{2}}.$$

where $|S^{d-1}|$ is the d- dimension surface area of unit radius sphere,
$S^{d-1} = \left\{(x_1, x_2, \ldots, x_d) : x_1^2 + x_2^2 + \ldots + x_d^2 = 1\right\}$.
Then, we can easy estimate the last integral by 2.3 and we get

$$\frac{1}{n} \log \left(\int \ldots \int_{\mathbb{R}^d \setminus B\left(0, \frac{\rho\sqrt{n}}{\epsilon}\right)} e^{-\frac{|x|^2}{2}} ds \right) \to -\frac{\rho^2}{2\epsilon^2} \quad \text{as } n \to \infty$$

Since we have proved the equality 2.2 for d- dimensions then it holds for every $d \in \mathbb{N}$. This completes the proof.

$\square$

Chapter 3

General Principles

3.1 Varadhan's Lemma and Contraction Principle

In this section, we present an approach to large deviations theory based on Laplace principle, which relies on the evaluation of asymptotics of certain integrals.

We first state a theorem that enables us to generate one LDP from another through contraction. In weak convergence, the continuous mapping theorem plays a key role. The analogous theorem in large deviation theory is the contraction principle. This theorem yields that LDP is preserved under continuous mappings. The contraction principle will turn out to be very useful later on.

Theorem 3.1. *(Contraction Principle)*
If $\{\mathbb{P}_n\}_{n\in\mathbb{N}}$ satisfies LDP on $\mathcal{X}$ with good rate function $I(\cdot)$, and F is a continuous mapping from the Polish space $\mathcal{X}$ to another Polish space $\mathcal{Y}$, then the family $\mathbb{Q}_n = \mathbb{P}_n \circ F^{-1}$ satisfies L.D.P.on $\mathcal{Y}$ with good rate function $J(\cdot)$ given by :

$$J(y) = \inf_{x:F(x)=y} I(x)$$

with the convention that $\inf \emptyset = \infty$.

Proof. Since F is continuous, F^{-1} maps open sets to open sets and closed sets to closed sets. Pick $C \subset \mathcal{Y}$ closed.
Then,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{Q}_n(C) = \limsup_{n\to\infty} \log \mathbb{P}_n \circ F^{-1}(C) \le - \inf_{x\in F^{-1}(C)} I(x) =$$
$$= - \inf_{y\in C} \inf_{x\in F^{-1}(\{y\})} I(x) = - \inf_{y\in C} \inf_{F(x)=y} I(x) = - \inf_{y\in C} J(y).$$

A similar argument works for $O \subset \mathcal{Y}$ open.

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{Q}_n(O) = \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\circ F^{-1}(O) \ge - \inf\left(F^{-1}(O)\right) \ge - \inf_{x\in F^{-1}(O)} I(x) = - \inf_{y\in O} J(y).$$

Hence, it remains to prove that J is a good rate function.

Clearly, $\mathcal{D}_I = \{x \in \mathcal{X} : I(x)\} \neq \emptyset$ implies $\mathcal{D}_J = \{y \in \mathcal{Y} : J(y)\} \neq \emptyset$. In fact, for every $c \in \mathbb{R}$ we have

$$\{y : \ J(y) \le c\} = F\big(\{x : \ I(x) \le c\}\big). \tag{3.1}$$

Indeed, if $J(y) \leq c$, then there exists a sequence $\{x_n\}$ such that $F(x_n) = y$ and $I(x_n) \to c$. Since $\{x_n\}$ is eventually contained in the compact set $\{x : \ I(x) \leq c+1\}$ we may take a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \to x_*$. By lower semicontinuity of I we have $I(x_*) \leq \liminf I(x_{n_k}) = c$, and by continuity of F we have $F(x_*) = y$. Thus, $y \in F\big(\{x : \ I(x) \leq c\}\big)$ and the one inclusion is proved. The inverse inclusion is elementary. In view of (3.1) the level set $\{y : \ J(y) \leq c\}$ is compact as the continuous image of the compact level set $\{x : \ I(x) \leq c\}$. J is lower semi-continuous as the infimum of lower semi-continuous functions. Hence, both of the proceeding arguments prove that J is a good rate function and the proof is complete.

□

We are now ready to formulate the first important general theorem of large deviations which is due to Varadhan. In the evaluation of integrals, large values accomplished in a "small" part of the space may play a key role. Varadhan's integral extends the well known method of Laplace for studying the asymptotics of certain integrals on $\mathbb{R}$: given a continuous function f from $[0,1]$ into $\mathbb{R}$, Laplace's principle states that:

$$\lim_{\epsilon \to 0} \epsilon \log \int_0^1 e^{\frac{f(x)}{\epsilon}} dx = \sup_{x \in [0,1]} f(x)$$

Varadhan's integral generalizes the previous result in the case of integrals not computed under the same measure but under a family of measures that satisfies the LDP.

Lemma 3.1. *(Varadhan's lemma) Suppose that $\{\mu_\epsilon\}_\epsilon$ satisfies LDP on $\mathcal{X}$ with good rate function I. Then for any bounded continuous function $\phi : \mathcal{X} \to \mathbb{R}$ we have that:*

$$\lim_{\epsilon \to 0} \epsilon \log \int e^{\frac{\phi(x)}{\epsilon}} d\mu_\epsilon (x) = \sup_{x \in \mathcal{X}} \big(\phi(x) - I(x)\big).$$

Proof. We break the proof into two parts.
Lower Bound
Since ϕ is continuous, then for every $x \in \mathcal{X}$ and $\delta \geq 0$ there exists a neighborhood U_x of x such that:

$$U_x = \{y \in U_x : \phi(y) > \phi(x) - \delta\},$$

for every $y \in U_x$.
Note that since $x \in U_x$ we have that:

$$\inf \{I(y) : y \in U_x\} \leq I(x)$$

Thus,

$$\int e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon (y) \geq \int_{U_x} e^{\frac{\phi(x)-\delta}{\epsilon}} d\mu_\epsilon (y) \geq e^{\frac{\phi(x)-\delta}{\epsilon}} \int_{U_x} d\mu_\epsilon (y) = e^{\frac{\phi(x)-\delta}{\epsilon}} \mu_\epsilon (U_x).$$

By the LDP lower bound we have that:

$$\liminf_{\epsilon \to 0} \epsilon \log \int e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon (y) \geq \phi(x) - \delta - \inf_{y \in U_x} I(y) \geq \phi(x) - I(x) - \delta.$$

Take $\delta \to 0$ and the supremum in the right side of the inequality, then we have the lower bound.

Upper bound
For each $x \in \mathcal{X}$ there exist a neighborhood U_x of x such that:
$\phi(y) < \phi(x) + \delta$ and $I(y) > I(x) - \delta$, for each $y \in \bar{U}_x$.
We know that for each $M > 0$ $\Psi_I(M) = \{x : I(x) \leq M\}$ is compact set.
Then, U_x balls cover the compact set Ψ_I,

$$\Psi_I(M) \subset \bigcup_{i=1}^N U_{x_i} \text{ and } F = \left(\bigcup_{i=1}^N U_{x_i}\right)^c$$

$$\int e^{\frac{\phi(y)\epsilon}{d}\mu_\epsilon(y)} = \int_{\bigcup_{i=1}^N U_{x_i}} e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) + \int_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) \leq \sum_{i=1}^N \int_{U_{x_i}} e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) + \int_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y)$$

$$\leq \sum_{i=1}^N \int_{U_{x_i}} e^{\frac{\phi(x_i)+\delta}{\epsilon}} d\mu_\epsilon(y) + \int_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) \leq \sum_{i=1}^N \mu_\epsilon(U_{x_i}) e^{\frac{\phi(x_i)+\delta}{\epsilon}} + \int_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y)$$

But, ϕ is a bounded function and μ_ϵ is exponentially tight we have that,

$$\int_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) = \int \mathbb{1}_F e^{\frac{\phi(y)}{\epsilon}} d\mu_\epsilon(y) \leq \int \mathbb{1}_F e^{\frac{\sup \phi}{\epsilon}} d\mu_\epsilon(y) = \mu_\epsilon(F) e^{\frac{\sup \phi}{\epsilon}} \leq e^{\frac{-L}{\epsilon}} e^{\frac{\sup \phi}{\epsilon}}$$

Collecting things together, we have by the LDP upper bound that,

$$\limsup_{\epsilon \to 0} \epsilon \log \int e^{\frac{\phi(y)}{\epsilon} d\mu_\epsilon(y)} \leq \max\left\{\max_i \left(\phi(x_i) + \delta - \inf_{\bar{U}_{x_i}} I(x_i)\right), \sup \phi - L\right\}$$

$$\leq \max\left\{\max_i (\phi(x_i) - I(x_i) + 2\delta), \sup \phi - L\right\}$$

$$\leq \max\left\{\max_i (\phi(x_i) - I(x_i)), \sup \phi - L\right\} + 2\delta$$

Let $\delta \to 0$ and $L \to \infty$ and the proof is complete.

□

Remark 3.1. *Varadhan's integral has the following interpretation. By writing formally the LDP for $\{X^\epsilon\}_\epsilon$ with rate function I as $\mathbb{P}[X_\epsilon \in dx] \cong e^{-\frac{I(x)}{\epsilon}}$, we can write:*

$$\mathbb{E}\left[e^{\frac{\phi(X_\epsilon)}{\epsilon}}\right] = \int e^{\frac{\phi(x)}{\epsilon}} \mathbb{P}[X_\epsilon \in dx] \cong \int e^{\frac{\phi(x)-I(x)}{\epsilon}}$$

As in Laplace's method, Varadhan's formula states that to exponential order, the main contribution to the integral is due to the largest value of the exponent.

Hence, the previous lemma means that the large deviation principle implies the Laplace principle. The next result proves the converse. The Laplace principle implies the large deviation principle with the same good rate function.

Proposition 3.1. *(Bryc's theorem) Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on a metric space $\mathcal{X}$. Assume that $\{\mu_n\}_{n\in\mathbb{N}}$ is exponentially tight. Suppose that the limit:*

$$\Lambda(f) = \lim_{n \to n} \frac{1}{n} \log \int e^{nf(x)} d\mu_n(x)$$

exists for all bounded functions f. Then, LDP holds with good rate function:

$$I(x) = \sup_{f \in \mathcal{C}_b(\mathcal{X})} \{f(x) - \Lambda(f)\}$$

Remark 3.2. *The above theorem reminds us again of weak convergence of probability measures. Varadhan's Lemma implies that* $\Lambda(x) = \sup_{f \in \mathcal{C}_b(\mathcal{X})} \{f(x) - I(f)\}$*, given the LDP. Note, however that the relation between* Λ *and* I *is not the convex duality as we have seen in the first chapter, where the functions* f *are linear. Even though, until this point we have seen convex rate functions, the above rate function doesn't need to be. Also, proving that the limit* $\Lambda(f)$ *exists for all bounded functions may be too hard to achieve. However, one really needs the limit to exist for a rich enough class of functions for the LDP to hold. If* $\mathcal{X}$ *is a metric vector space , then a rich enough class of functions, that ensure the LDP through Bryc's theorem is the class of Lipschitz functions.*

Proof. Firstly, since $\Lambda_0 = 0$, we have that $I \geq 0$. Function I is lower semicontinuous, since it is supremun of continuous functions. Since exponential tightness and the weak LDP together imply the full LDP, we need only prove that weak LDP is satisfied. We start with the upper bound.

Upper bound

As in Cramér theorem on $\mathbb{R}^d$ it suffices to show the upper bound for open balls on $\mathcal{X}$ and then we will cover any compact $K \subset \mathcal{X}$ with a finite number of such open balls.
Pick $\epsilon > 0$, from the definition of I we have that, for each $x \in \mathcal{X}$ there exists $f \in \mathcal{C}_b(\mathcal{X})$ such that:

$$f(x) - \Lambda(f) > I^\epsilon(x) = (I - \epsilon) \wedge \frac{1}{\epsilon}$$

Since, f is continuous, there exists $\delta_x = \delta_{x,\epsilon} > 0$ such that:

$$l_x = \inf_{y \in B(x,\delta_x)} \{f(y) - f(x)\} \geq -\epsilon,$$

Since for every $x \in \mathcal{X}$ we have that $f(x) - f(y) - l_x \geq 0$, then for each $y \in B(x, \delta_x)$

$$\mu_n(B(x,\delta_x)) = \int_{B(x,\delta_x)} d\mu_n \leq \int_{B(x,\delta_x)} e^{n[f(y)-f(x)-l_x]} d\mu(y) = e^{-nl_x} \int_{B(x,\delta_x)} e^{n[f(y)-f(x)]} d\mu(y)$$

$$\leq e^{n[\epsilon - f(x)]} \int e^{nf(y)} d\mu_n(y)$$

for each $x \in \mathcal{X}$, $n \in \mathbb{N}$.

Thus,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(B(x,\delta)) \leq \epsilon - (f(x) - \Lambda(f)) < -I^\epsilon(x) + \epsilon$$

for each $x \in \mathcal{X}$.
Given, now, of any compact $K \subset \mathcal{X}$ there exists, $x_1, \ldots, x_m \in K$, $m \in \mathbb{N}$, such that $K \subset \bigcup_{i=1}^m B(x_i, \delta_{x_i})$, therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K) \leq \max_{i=1,\ldots,m} \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(B(x_i, \delta_{x_i}))$$

$$\epsilon - \min_{i=1,\cdots,m} I^\epsilon(x_i) \leq \epsilon - \inf_{x\in K} I^\epsilon(x).$$

Since ϵ is picked randomly, this result proves weak upper bound of the L.D.P.

Lower bound

Suppose that $x \in \mathcal{X}$ and $U \subseteq \mathcal{X}$ a neighborhood of x. Since $\mathcal{X}$ is a metric space, there exists continuous function such that $f : \mathcal{X} \to [0,1]$ such that $f(x) = 1$ and f vanishes outside U. Take $m > 0$ and define $f_m = m(f-1)$ for each $m \in \mathbb{N}$. Then, by our assumption there exists the limit:

$$\Lambda(f_m) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{nf_m(y)} d\mu_n(y)$$

Thus,

$$\begin{aligned}\int e^{nf_m(y)} d\mu_n(y) &= \int_U e^{nf_m(y)} d\mu_n(y) + \int_{U^c} e^{nf_m(y)} d\mu_n(y) \\ &= \int_U e^{nf_m(y)} d\mu_n(y) + e^{-nm}\mu_n(U^c) \\ &\leq \mu_n(U) + e^{-nm}.\end{aligned}$$

Thus,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(U) \vee (-m) &\geq \liminf_{n\to\infty} \frac{1}{n} \log \int e^{nf_m(y)} d\mu_m(y) = \Lambda(f_m) \\ &= -(f_m(x) - \Lambda(f_m)) \geq -\sup_{f\in\mathcal{C}_b(\mathcal{X})} (f(x) - \Lambda(f)) \\ &= -I(x)\end{aligned}$$

Since the above result is valid for every $m > 0$, the proof is completed if we let $m \to \infty$.

$\square$

If we combine the results of Varadhan and Bryc, then we have the following theorem for exponentially tight families of measures.

Theorem 3.2. *Suppose that, $\{\mu_n\}$ is exponential tight sequence. Then, $\{\mu_n\}$ satisfies LDP if and only if the limit*

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{nf(x)} d\mu_n(x)$$

exists for each $f \in \mathcal{C}_b(\mathcal{X})$ and in that case the rate function I is given by

$$I(x) = \sup_{f\in\mathcal{C}_b(\mathcal{X})} \{f(x) - \Lambda(f)\}$$

and

$$\Lambda(f) = \sup_{x\in\mathcal{X}} \{f(x) - I(x)\}$$

for each $f \in \mathcal{C}_b(\mathcal{X})$.

3.2 Relative entropy and Varadhan-Donsker formula

We next show how one can evaluate expectations arising in Laplace principles, which then can be used to derive the large deviation principle associated with the empirical measures of i.i.d. random variables.

The relative entropy plays a key role in the determination of the rate function. We are given a topological space $\mathcal{X}$, equipped with its Borel σ field. Let $\mathcal{M}(\mathcal{X})$ be the space of finite signed measures on $(\mathcal{X},\mathcal{B})$, $\mathcal{M}_1(\mathcal{X})$ be the space of probability measures on $\mathcal{M}(\mathcal{X})$ and $\mathcal{C}_b(\mathcal{X})$ be the space of bounded continuous functions and $\mathcal{B}_b(\mathcal{X})$ be the space of bounded measurable functions. There is a natural duality between $\mathcal{M}(\mathcal{X})$ and $\mathcal{B}_b(\mathcal{X})$:

$$\langle \nu, g\rangle = \int_{\mathcal{X}} g d\nu$$

for $g \in \mathcal{B}_b(\mathcal{X})$ and $\nu \in \mathcal{M}(\mathcal{X})$.

Definition 3.1. *For $\mu \in \mathcal{M}_1(\mathcal{X})$, the relative entropy $R(\cdot|\mu)$ is a mapping from $\mathcal{P}(\mathcal{X})$ into $\bar{\mathbb{R}}$, defined by*

$$R(\nu|\mu) = \begin{cases} \int_{\mathcal{X}} \left(\log \frac{d\nu}{d\mu}\right) d\nu = \int_{\mathcal{X}} \frac{d\nu}{d\mu}\left(\log \frac{d\nu}{d\mu}\right) d\mu & \text{if } \nu \ll \mu \\ \infty & \text{otherwise} \end{cases}$$

Note that $x\log x \geq -\frac{1}{e}$ so the above integral is well defined.

Remark 3.3. *By observing that $x \log x \geq x-1$ with equality if and only of $x=1$, we see that $R(\nu|\mu) \geq 0$, and $R(\nu|\mu)=0$ if and only if $\nu=\mu$.*

Remark 3.4. *Let us fix $\mu \in \mathcal{M}_1(\mathcal{X})$. Let $p : \mathcal{B}_b(\mathcal{X}) \to \mathbb{R}$ be defined by $p(\phi) = \log \int e^{\phi} d\mu(x)$. Then p and R are convex conjugate of one another and have the following variational representation.*

Proposition 3.2. *Let ϕ be a bounded measurable function on $\mathcal{X}$ and μ a probability measure on $\mathcal{X}$. Then,*

$$\log \int_{\mathcal{X}} e^{\phi} d\mu = \sup_{\nu \in \mathcal{M}_1(\mathcal{X})} \left[\int_{\mathcal{X}} \phi d\nu - R(\nu|\mu)\right],$$

and the supremum is attained uniquely by the probability measure ν_0 defined by

$$\frac{d\nu_0}{d\mu} = \frac{e^{\phi}}{\int_{\mathcal{X}} e^{\phi} d\mu}$$

Proof. In the supremum in the above variational formula, we may restrict to $\nu \in \mathcal{P}(\mathcal{X})$ with finite relative entropy: $R(\nu|\mu) < \infty$. If $R(\nu|\mu) < \infty$, then ν is absolutely continuous with respect to μ, and since μ is equivalent to ν_0, ν is also absolutely continuous with respect to ν_0. Thus,

$$\int_{\mathcal{X}} \phi d\nu - R(\nu|\mu) = \int_{\mathcal{X}} \phi d\nu - \int_{\mathcal{X}} \left(\log \frac{d\nu}{d\mu}\right) d\nu$$
$$= \int_{\mathcal{X}} \phi d\nu - \int_{\mathcal{X}} \log\left(\frac{d\nu}{d\nu_0}\right) d\nu - \int_{\mathcal{X}} \log\left(\frac{d\nu_0}{d\mu}\right) d\nu$$

$$= \int_{\mathcal{X}} \left(\phi - \log \frac{d\nu_0}{d\mu} \right) d\nu - R(\nu|\nu_0)$$
$$= \int_{\mathcal{X}} \log \left(e^{\phi} \frac{d\mu}{d\nu_0} \right) d\nu - R(\nu|\nu_0)$$
$$= \int_{\mathcal{X}} \log \left(\int_{\mathcal{X}} e^{\phi} d\mu \right) d\nu - R(\nu|\nu_0)$$
$$= \log \int_{\mathcal{X}} e^{\phi} d\mu - R(\nu|\nu_0).$$

We conclude by observing that $R(\nu|\nu_0) \geq 0$ and $R(\nu|\nu_0) = 0$ if and only if $\nu = \nu_0$.

□

Proposition 3.3. *(Varadhan-Donsker variational formula) Let $\mathcal{X}$ be a Polish space. For all $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$, we have*

$$R(\nu|\mu) = \sup_{f \in \mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \right\} = \sup_{g \in \mathcal{C}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} g d\nu - \log \int_{\mathcal{X}} e^g d\mu \right\}$$

The dual formula to the above variational formula is known as the Donsker-Varadhan variational formula.

Proof. First we show that

$$R(\nu|\mu) = \sup_{f \in \mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \right\}$$

and later we prove that

$$\sup_{g \in \mathcal{C}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} g d\nu - \log \int_{\mathcal{X}} e^g d\mu \right\} = \sup_{f \in \mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \right\}. \quad (3.2)$$

We denote by $H(\nu, \mu)$ the right side of (3.2). By taking the zero function on $\mathcal{X}$ we observe that $H(\nu, \mu) \geq 0$. From Proposition 3.3, we have for any $f \in \mathcal{B}_b(\mathcal{X})$

$$R(\nu|\mu) \geq \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu$$

and taking the supremum over $f \in \mathcal{B}_b(\mathcal{X})$, we obtain that

$$R(\nu|\mu) \geq \sup_{f \in \mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \right\} = H(\nu, \mu)$$

To prove the inverse inequality, we may assume that $H(\nu, \mu) < \infty$. We first show that under this condition ν is absolutely continuous with respect to μ. Let A be a Borel set for which $\mu(A) = 0$ and take $k > 0$. Since for any $f \in \mathcal{B}_b(\mathcal{X})$

$$\int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \leq H(\nu, \mu) < \infty$$

We obtain, upon substituting $f = k\mathbb{1}_A$

$$\int_{\mathcal{X}} k\mathbb{1}_A d\nu - \log \int_{\mathcal{X}} e^{k\mathbb{1}_A} d\mu \leq H(\nu, \mu)$$

$$k\int_A d\nu \leq H(\nu,\mu) \Rightarrow k\nu(A) \leq H(\nu,\mu)$$

Taking $k \to \infty$ gives $\nu(A)$ thus $\nu \ll \mu$. Since $\nu \ll \mu$ we can define the Radon-Nikodym derivative $h = \frac{d\nu}{d\mu}$. If h is uniformly positive and bounded then $f = \log h$ is bounded and measurable and substituting this function into variational formula 3.3 yields

$$R(\nu|\mu) = \int_{\mathcal{X}} \log h d\nu \leq H(\nu,\mu)$$

If h is uniformly positive and not bounded, we truncate the function and set $h_n = h \wedge n$ and substitute $f = \log h_n$ into the variational formula 3.3. Using the Monotone Convergence theorem we again obtain,

$$R(\nu|\mu) = \int_{\mathcal{X}} \log h d\nu = \lim_{n\to\infty} \int_{\mathcal{X}} \log h_n d\nu \leq H(\nu,\mu) + \lim_{n\to\infty} \log \int_{\mathcal{X}} h_n d\mu = H(\nu,\mu).$$

We now treat the general case where f is neither uniformly positive nor bounded. For $t \in [0,1]$ we define

$$\nu_t = t\mu + (1-t)\nu \text{ and } h_t = \tfrac{d\nu_t}{d\mu} = t + (1-t)h$$

for each $t \in (0,1]$ h_t is uniformly positive and so by the preceding calculation:

$$R(\nu_t|\mu) \leq H(\nu_t,\mu).$$

We now want to prove that

$$\lim_{t\to 0} R(\nu_t|\mu) = R(\nu|\mu) \text{ and } \lim_{t\to 0} H(\nu_t,\mu) = H(\nu,\mu).$$

Since $x\log x$ is convex on $[0,\infty)$

$$R(\nu_t|\mu) = \int_{\mathcal{X}} h_t \log h_t d\mu \leq (1-t)\int_{\mathcal{X}} h\log h d\mu$$

Furthermore, $\log x$ is concave on $(0,\infty)$ then,

$$\log h_t \geq \log t \vee (1-t)\log h.$$

This means that

$$\begin{aligned} R(\nu_t|\mu) &= \int_{\mathcal{X}} t\log h_t d\mu + \int_{\mathcal{X}} (1-t)\, h\log h_t d\mu \\ &\geq t\log t + (1-t)^2 \int h\log h d\mu \\ &= t\log t + (1-t)^2 R(\nu|\mu). \end{aligned}$$

If we let $t \to 0$ and combine the above inequalities then

$$\lim_{t\to 0} R(\nu_t|\mu) = R(\nu|\mu)$$

For the second limit, we observe that $H(\mu,\mu) = 0$ and $t \mapsto H(\nu_t,\mu)$ is lower semicontinuous and convex, as the supremum of linear functions in t. Thus, for every $t \in [0,1]$

$$0 \leq H(\nu_t,\mu) \leq tH(\mu,\mu) + (1-t)H(\nu,\mu) = (1-t)H(\nu,\mu) \leq H(\nu,\mu) < \infty$$

and so $H(\nu_t, \mu)$ is also bounded. Also, we know that a function f that is convex and lower semicontinuous on $\mathbb{R}$ and is finite on the closed bounded interval $[a, b]$ then f is continuous on $[a, b]$. As a result, $H(\nu_t, \mu)$ is continuous and therefore

$$\lim_{t\to 0} H(\nu_t, \mu) = H(\nu_0, \mu) = H(\nu, \mu).$$

We showed that

$$R(\nu|\mu) = \sup_{g\in\mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} g d\nu - \log \int_{\mathcal{X}} e^g d\mu \right\} = H(\nu, \mu). \tag{3.3}$$

Now, we will use Lusin's theorem to approximate $f = \frac{d\nu}{d\mu}$ by bounded continuous functions with respect to both μ and ν and pass to the limit. In fact, we will find a rather small set that will work with respect to both of the measures. Then we will have that

$$R(\nu|\mu) = \sup_{g\in\mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} g d\nu - \log \int_{\mathcal{X}} e^g d\mu \right\} = \sup_{f\in\mathcal{C}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \log \int_{\mathcal{X}} e^f d\mu \right\}.$$

In order to prove 3.2 we need the following lemma

Lemma 3.2. *For each $f \in \mathcal{B}_b(\mathcal{X})$ and $\mu, \nu \in \mathcal{M}(\mathcal{X})$ there exists a sequence $\{f_n\} \subseteq \mathcal{C}_b(\mathcal{X})$ such that*

$$f_n \overset{n\nearrow\infty}{\longrightarrow} f \quad \text{in } L^1(\mu) \text{ and } L^1(\nu)$$

$$f_n \overset{n\nearrow\infty}{\longrightarrow} f \quad \mu\text{-a.s. and } \nu\text{-a.s.},$$

$$\|f_n\|_u \le \|f\|_u - \frac{1}{n}, \forall n \in \mathbb{N}.$$

Proof. Since each measurable function is uniformly approximated by simple functions, it suffices to show that in the case that f is the characteristic function $\mathbb{1}_E$ of a Borel set $E \subseteq \mathcal{X}$. Suppose that $E \subseteq \mathcal{X}$ Borel set and $\epsilon > 0$. We are looking for a $g \in \mathcal{C}_b(\mathcal{X})$ such that

$$\int_{\mathcal{X}} |g - \mathbb{1}_E| d\mu \vee \int_{\mathcal{X}} |g - \mathbb{1}_E| d\nu < \epsilon.$$

Since μ, ν are finite Borel measure in a Polish space, they are regular measures and there exist compact sets $K_\mu, K_\nu \subseteq E$ and open set $A_\mu, A_\nu \supseteq E$ such that

$$K_\rho \subseteq E \subseteq A_\rho, \quad \rho(A_\rho \setminus K_\rho) < \epsilon, \quad \rho = \mu, \nu.$$

If we set $K := K_\mu \cup K_\nu$, $A := A_\mu \cap A_\nu$ then K is compact and A is open and

$$\rho(A \setminus K) \le \rho(A_\rho \setminus K) \le \rho(A_\rho \setminus K_\rho) \le \epsilon, \quad \rho = \mu, \nu$$

Pick a function $g \in \mathcal{C}_b(\mathcal{X})$ such that $\mathbb{1}_K \le g \le \mathbb{1}_A$, then

$$|g - \mathbb{1}_E| \le \mathbb{1}_{A\setminus K}$$

and

$$\int_{\mathcal{X}} |g - \mathbb{1}_E| d\rho \le \rho(A \setminus K) < \epsilon, \quad \rho = \mu, \nu.$$

Therefore we prove that for each $f \in \mathcal{B}_b(\mathcal{X})$ there exists $\{g_n\} \subseteq \mathcal{C}_b(\mathcal{X})$ such that $g_n \to f$ in $L^1(\mu)$ and $L^1(\nu)$. Now if we pass in the subsequence of $\{g_n\}$ we can assume that $\{g_n\}$ converges pointwise in f μ-a.s. and ν-a.s.. If we define

$$f_n := \left(-\|f\|_u + \frac{1}{n}\right) \vee \left[g_n \wedge \left(\|f\|_u - \frac{1}{n}\right)\right], \quad \forall n \in \mathbb{N},$$

then we obtain that $f_n \to f$ pointwise μ-a.s. and ν-a.s.. Since $\|g_n\|_u \leq \|f_n\|_u$ for every $n \in \mathbb{N}$, we have that $f_n \to f$ in $L^1(\mu)$ and $L^1(\nu)$ by bounded convergence theorem. Then f_n satisfies all the assumptions of the Lemma 3.2. □

We have proved that

$$R(\nu|\mu) = \sup_{f \in \mathcal{B}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \int_{\mathcal{X}} h(f) \, d\mu \right\}$$

where $h(f) = h \log f$ and $f = \frac{d\mu}{d\nu}$. Also we know that

$$R(\nu|\mu) = H(\nu, \mu) \geq H'(\nu, \mu) = \sup_{f \in \mathcal{C}_b(\mathcal{X})} \left\{ \int_{\mathcal{X}} f d\nu - \int_{\mathcal{X}} h(f) \, d\mu \right\} \geq 0.$$

Since $f \in \mathcal{B}_b(\mathcal{X})$, $h(f)$ is lower bounded and the integral $\int_{\mathcal{X}} f(f)$ is defined for every $f \in \mathcal{B}_b(\mathcal{X})$.
Assume that $\mu, \nu \in \mathcal{M}(\mathcal{X})$, we need to show that for any function $f \in \mathcal{B}_b(\mathcal{X})$ such that $f \in L^1(\mu)$ there is $\{f_n\} \subseteq \mathcal{C}_b(\mathcal{X})$, such that

$$\int_{\mathcal{X}} f_n d\nu - \int_{\mathcal{X}} h(f_n) \, d\mu \overset{n \to \infty}{\longrightarrow} \int_{\mathcal{X}} f d\nu - \int_{\mathcal{X}} h(f) \, d\mu.$$

Suppose, now, that there is $\{f_n\}$ that satisfies the assumptions of Lemma 3.2. Then

$$\int_{\mathcal{X}} f_n d\mu \to \int_{\mathcal{X}} f d\nu \quad \text{as } n \to \infty$$

So it remains to show that

$$\lim_{n \to \infty} \int h(f_n) \, d\mu \to \int_{\mathcal{X}} h(f) \, d\mu \quad \text{as } n \to \infty \tag{3.4}$$

Indeed, h is continuous as a convex function. Then

$$h(f_n) \to h(f) \qquad \text{pointwise } \mu\text{-a.s. as } n \to \infty.$$

Also, h is continuous and bounded in any closed interval and $|f_n(x)| \leq \|f\|_u$ for each $n \in \mathbb{N}$ and for each $x \in \mathcal{X}$. Then we have that

$$\|h(f_n)\|_u \leq \sup_{-\|f\|_u \leq t \leq \|f\|_u} h(t) \leq \infty \quad \text{for each } n \in \mathbb{N}.$$

By bounded convergence theorem we obtain 3.4 and so the proof is completed.

□

We next prove other three important properties of relative entropy: convexity, lower semicontinuity, and compactness of level sets. In this section we develop only the properties of relative entropy that will be useful later in this section. All of the results in the following lemma are formulated for arbitrary Polish spaces.

Lemma 3.3. *Let $\mathcal{X}$ be a Polish space and $\mu \in \mathcal{M}_1(\mathcal{X})$. The relative entropy has the following properties:*

(i) (convexity) $R(\cdot|\mu)$ is strictly convex on the set $\{\nu \in \mathcal{M}_1(\mathcal{X}) : R(\nu|\mu) < \infty\}$.

(ii) (lower semicontinuity) $R(\cdot|\mu)$ is lower semicontinuous.

(iii) (compactness of level sets) $R(\cdot|\mu)$ has compact level sets. That is, for each $M < \infty$ the set $\{\nu \in \mathcal{M}_1(\mathcal{X}) : R(\nu|\mu) \leq M\}$ is a compact subset of $\mathcal{M}_1(\mathcal{X})$.

Proof. (i) Convexity is immediate by Proposition 3.3, since $\nu \mapsto R(\nu|\mu)$ is the supremum of linear functions in ν. To prove strict convexity recall that

$$R(\nu|\mu) = \int_{\mathcal{X}} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu$$

for any $\nu \in \mathcal{M}_1(\mathcal{X})$ satisfying $R(\nu|\mu) < \infty$. We know that $h(x) = x \log x$ is strictly convex for $x \in [0, \infty)$.

Suppose that $\nu_0, \nu_1 \in \mathcal{M}_1(\mathcal{X})$ and $\nu_t = (1-t)\nu_0 + t\nu_1 \in \mathcal{M}_1(\mathcal{X})$, $t \in (0,1)$. Then, we have that

$$h\left(\frac{d\nu_t}{d\mu}\right) = h\left((1-t)\frac{d\nu_0}{d\mu} + t\frac{d\nu_1}{d\mu}\right) \leq (1-t)\,h\left(\frac{d\nu_0}{d\mu}\right) + th\left(\frac{d\nu_1}{d\mu}\right)$$

with equality holding if and only if $\frac{d\nu_0}{d\mu} = \frac{d\nu_1}{d\mu}$. Thus,

$$\begin{aligned} R(\nu_t|\mu) = \int_{\mathcal{X}} h\left(\frac{d\nu_t}{d\mu}\right) d\mu \leq (1-t)\int_{\mathcal{X}} h\left(\frac{d\nu_0}{d\mu}\right) d\mu + t\int_{\mathcal{X}} h\left(\frac{d\nu_1}{d\mu}\right) d\mu \\ = (1-t)\,R(\nu_0|\mu) + tR(\nu_1|\mu)\,, \end{aligned}$$

with equality holding if and only if $\frac{d\nu_0}{d\mu} = \frac{d\nu_1}{d\mu}$ μ-a.s., that is if and only if $\nu_1 = \nu_2$, proving that $R(\cdot|\mu)$ is strictly convex on $\{\nu \in \mathcal{M}_1(\mathcal{X}) : \ R(\nu|\mu) < +\infty\}$.

(ii) For any $f \in \mathcal{C}_b(\mathcal{X})$ the map $\nu \mapsto \int f d\nu$ is continuous. In view of Proposition 3.3 $R(\cdot|\mu)$ is lower semicontinuous as the supremum of continuous functions.

(iii) Let $\{\nu_n, n \in \mathbb{N}\}$ be any sequence in $\{\nu \in \mathcal{M}_1(\mathcal{X}) : \ R(\nu|\mu) \leq M\}$. According to the variational formula 3.3 for any bounded measurable function $f \in \mathcal{B}_b(\mathcal{X})$ mapping $\mathcal{X}$ into $\mathbb{R}$ we have that

$$\int_{\mathcal{X}} f d\nu_n - \log \int_{\mathcal{X}} e^f d\mu \leq M$$

Pick any $\delta \geq 0$ and $\epsilon > 0$. The tightness of μ guarantees that there exists a compact set K such that $\mu(K^c) \leq \epsilon$. Substituting into the last display the function f that equals to 0 on K and $\log\left(1 + \frac{1}{\epsilon}\right)$ on K^c, we have that for each $n \in \mathbb{N}$

$$\int_{K^c} \log\left(1 + \frac{1}{\epsilon}\right) d\nu_n - \log\left(\int_K d\mu + \int_{K^c} 1 + \frac{1}{\epsilon} d\mu\right) \leq M \Leftrightarrow$$

$$\log\left(1 + \frac{1}{\epsilon}\right) \nu_n(K^c) - \log\left(1 + \frac{1}{\epsilon}\mu(K^c)\right) \leq M \Leftrightarrow$$

$$\nu_n(K^c) \leq \frac{1}{\log\left(1 + \frac{1}{\epsilon}\right)} \left(M + \log\left(1 + \frac{1}{\epsilon}\mu(K^c)\right)\right) \Rightarrow$$

$$\nu_n(K^c) \leq \frac{1}{\log\left(1 + \frac{1}{\epsilon}\right)} (M + \log 2)\,.$$

Since ϵ is picked arbitrary we can choose ϵ such that $\frac{1}{\log\left(1+\frac{1}{\epsilon}\right)}\left(M+\log 2\right) \leq \delta$. This implies that $\{\nu_n\}$ is tight. Applying Prohorov's theorem there exists a subsequence $\{\nu_{n_k}\}$ weakly converging to some $\nu \in \mathcal{M}_1\left(\mathcal{X}\right)$. Lower semicontinuity yields that

$$R\left(\nu|\mu\right) \leq \liminf_{k\to\infty} R\left(\nu_{n_k}|\mu\right) \leq M$$

This implies that $\{\nu \in \mathcal{M}_1\left(\mathcal{X}\right) : R\left(\nu|\mu\right) \leq M\}$ is compact, and the proof is complete. $\square$

Properties (ii) and (iii) above state in particular that the function $\nu \mapsto R\left(\nu|\mu\right)$ is a good rate function defined on $\mathcal{M}_1\left(\mathcal{X}\right)$.

Chapter 4

LDP for abstract measures

4.1 Sanov's theorem

We are now ready to prove large deviations for empirical measures of a sequence of i.i.d. random variables on a Polish space. The topology that determines the open and closed set in $\mathcal{M}(\mathcal{X})$ is the weak topology generated by $\mathcal{C}_b(\mathcal{X})$. Let $\{X_n\}$ be a sequence of i.i.d. random variables valued on a Polish space $\mathcal{X}$ and with common distribution μ. Then the sample distribution of $\{X_i\}$ is a sequence of random measures (empirical measures):

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ_x is the Dirac measure in $x \in \mathcal{X}$ and maps $\mathcal{X}^n \to \mathcal{M}_1(\mathcal{X})$ and the product measure $\mathbb{P}^n$ will generate a measure $\mathbb{P}_n$ on $\mathcal{M}_1(\mathcal{X})$ which is the distribution of the empirical distribution. The weak law of Large Numbers essentially implies that :

$$\mathbb{P}_n \Rightarrow \delta_\mu$$

This means that the empirical distribution μ_n approaches the true distribution μ. Close here is in the sense of weak convergence. To this end, we have that

$$L_n(x_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_j\}}(X_j)$$

is a random variable. Therefore, the sequence $\{Y_{x,j}\}$ which is defined by $Y_{j,x} = \mathbb{1}_{\{x\}}X_j$ is i.i.d. sequence with mean value $\mathbb{E}[Y_{j,x}] = \mathbb{P}[x_1 = \mu] = \mu(x)$. Thus, for every $x \in \mathcal{X}$, by the weak law of large numbers:

$$L_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_{X,j} \overset{n\to\infty}{\longrightarrow} \mu(x) \quad \mathbb{P}\text{-a.s.}$$

Thus, it is reasonable to inquire about the large deviations of $\{\mathbb{P}_n : n \geq 1\}$. In fact, we will see that the large deviations of $\{\mathbb{P}_n : n \geq 1\}$ are governed by the rate function:

$$I(\nu) = R(\nu|\mu)$$

and that $R(\cdot|\mu)$ is a good rate function as we proved previously.

Theorem 4.1. *The sequence $\{\mathbb{P}_n\}$ satisfies LDP on $\mathcal{M}_1(\mathcal{X})$ with the good rate function the relative entropy $R(\nu|\mu)$.*

Proof. Upper bound

We are going to follow the techniques that we used to prove Cramér theorem on $\mathbb{R}^d$. For any given ν and any $\epsilon > 0$, there is a small neighborhood U_ν around ν such that:

$$U_\nu := \{\rho : |\langle f, \nu\rangle - \langle f, \rho\rangle| < \epsilon\}.$$

Thus,

$$\begin{aligned}\mathbb{P}_n[U_\nu] &= \mathbb{P}_n[\rho : |\langle f, \nu\rangle - \langle f, \rho\rangle| < \epsilon] \\ = \mathbb{P}^n\left[|\frac{1}{n}\sum_{i=1}^n f(X_i) - \langle f, \nu\rangle| < \epsilon\right] &\leq e^{-n\langle f,\nu\rangle + n\epsilon}\mathbb{E}^n\left[e^{\sum_{i=1}^n f(X_i)}\right] \\ &= e^{-n\langle f,\nu\rangle + n\epsilon}\left(\int e^{f(x)}d\mu(x)\right)^n\end{aligned}$$

Then,

$$\begin{aligned}\frac{1}{n}\log\mathbb{P}_n[U_\nu] &\leq \frac{1}{n}\log e^{-n\langle f,\nu\rangle} + \frac{1}{n}\log e^{n\epsilon} + \frac{1}{n}\log\left(\int_{\mathcal{X}} e^{f(x)}d\mu(x)\right)^n \\ &= -\langle f,\nu\rangle + \epsilon + \frac{1}{n}\log\left(\int_{\mathcal{X}} e^{f(x)}d\mu(x)\right)^n \\ &= \langle f,\nu\rangle + \epsilon + \log\int_{\mathcal{X}} e^f(x)\,d\mu(x).\end{aligned}$$

We may now choose $f \in \mathcal{C}_b(\mathcal{X})$ so that

$$\langle f,\nu\rangle - \log\int e^f d\mu \geq R(\nu|\mu) - \epsilon.$$

Therefore,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[U_\nu] \leq -I^\epsilon(\nu) + \epsilon$$

If D is any compact subset of $\mathcal{M}_1(\mathcal{X})$, then D can be covered by a finite number of U_ν. Hence, $C \subseteq \bigcup_{i=1}^N U_{\nu_i}$.

$$\mathbb{P}_n[D] \leq \mathbb{P}_n[\bigcup_{i=1}^N U_{\nu_i}] \leq \sum_{i=1}^n \mathbb{P}_n[U_{\nu_i}] \leq N\max_{i\leq N}\mathbb{P}[U_{\nu_i}]$$

Then,

$$\frac{1}{n}\log\mathbb{P}_n[D] \leq \frac{1}{n}\max_{i\leq N}\log N + \frac{1}{n}\log\mathbb{P}[U_{\nu_i}] \Rightarrow$$

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[D] &\leq \max_{i=1,\cdots,N}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[U_{\nu_i}] \\ &\leq \max_{i=1,\cdots,n}(-I^\epsilon(\nu_i)) \leq \sup_{\nu\in D}(-I^\epsilon(\nu)) = -\inf_{\nu\in D}I^\epsilon(\nu)\end{aligned}$$

Since ϵ is arbitrary we actually have that

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[D] \leq -\inf_{\nu\in D}I(\nu)$$

for every compact $D \subseteq \mathcal{M}_1(\mathcal{X})$.

In order to get an upper bound for general closed set, it is enough to establish exponential tightness of $\{\mathbb{P}_n\}$. We want to show that for every $L > 0$, there exists a compact set K_L such that $\mathbb{P}_n[K_l^c] \leq e^{-nL}$ for all $n \in N$. Using Ulam's theorem, for any probability measure $\mu \in \mathcal{M}_1(\mathcal{X})$ and for any $\epsilon > 0$, there exists a compact set $K \subset \mathcal{X}$ such that $\mu(K^c) \leq \epsilon$. So, we can pick a compact set A_L such that $\mu(A_L^c) < e^{-L^2}$. Then, $D_L = \left\{\nu : \nu(A_L) \geq 1 - \frac{1}{L}\right\}$ is closed in the weak topology $\sigma(\mathcal{M}(\mathcal{X}), \mathcal{C}_b(\mathcal{X}))$ because the Portmanteau theorem implies that for every closed set A_L

$$\limsup_{j\to\infty} \nu_j(A_L) \leq \nu(A_L)$$

if ν_j weakly converges to ν.
The set $K_L = \bigcap_{L\geq l} D_L$ is also closed. Since K_L is also tight, it is compact by Prohorov's theorem. But, now, we have that

$$\begin{aligned}
\mathbb{P}_n[D_L^c] = \mathbb{P}\left[L_n[A_L^c] > \frac{1}{L}\right] &= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{A_L^c}(X_i) > \frac{1}{L}\right] \\
&= \mathbb{P}\left[\sum_{i=1}^n \mathbb{1}_{A_L^c}(X_i) > \frac{n}{L}\right] \\
&\leq e^{-\frac{nL^2}{L}}\mathbb{E}^{\mathbb{P}}\left[e^{L^2\sum_{i=1}^n \mathbb{1}_{A_L^c}}(X_i)\right] \\
&= e^{-nL}\mathbb{E}^{\mu}\left[e^{L^2\mathbb{1}_{A_L^c}(X_i)}\right]^n \\
&= e^{-nL}\left(e^{L^2}e^{-L^2}+1\right)^n \leq e^{-nL}2^n
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{P}_n[K_L^c] &= \mathbb{P}_n[(\cap_{L\geq l} D_L)^c] = \mathbb{P}_n[\cup_{L\leq l} D_L^c] \\
\leq \sum_{L\geq l}\mathbb{P}_n[D_L^c] &\leq \sum_{L\geq l} 2^n e^{-nL} \leq \frac{2^n e^{-nL}}{1-e^{-n}} \leq 2e{-}n(l-1)
\end{aligned}$$

We use the same argument as in the Cramér theorem on $\mathbb{R}^d$. $F = (F\cap K_L)\cup(F\setminus K_L)$.

Hence,

$$\mathbb{P}_n[F] \leq \mathbb{P}_n[F\cap K_L] + \mathbb{P}_n[K_L^c] \Rightarrow$$

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[F] \leq \max\left\{-\inf_{\nu\in F} I(\nu), -l\right\}$$

and by letting $l \to \infty$ we get the upper bound

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n[F] \leq -\inf_{\nu\in F} I(\nu)$$

for every closed $F \subset \mathcal{X}$.

Lower bound
To prove the lower bound, we tilt the measure from $\mathbb{P}_n$ to $\mathbb{Q}_n$ through Radon-Nikodym derivative based on i.i.d. random vectors with μ for each component. This way relative

entropy enters the calculation. Let U_ν be a neighborhood around ν. We assume that $R(\nu|\mu) < \infty$. We want to show that $\forall \nu \in \mathcal{M}_1(\mathcal{X})$ and $n \in U_\nu$ open

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n[U_\nu] \geq -R(\nu|\mu).$$

This implies that $\nu \ll \mu$. Let $b = \frac{d\nu}{d\mu}$ and let $\mathbb{Q} = \nu^{\otimes n}$ and $\mathbb{P} = \mu^{\otimes n}$ be the law of the i.i.d. with marginal μ.

Then,

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(x_1, \cdots, x_n) = \prod_{i=1}^{n} b(x_i) = b_n(\mathbf{x})$$

Here, we used the notation $\mathbf{x} = (x_i)_{i\geq 1}$.
Now, we write

$$\begin{aligned}
\mathbb{P}[L_n \in U_\nu] &\geq \int_{L_n \in U_\nu} \mathbb{1}_{\{b_n > 0\}} d\mathbb{P} \\
&= \int_{L_n \in U_\nu} b_n^{-1}(\mathbf{x})\, d\mathbb{Q} \\
&= \int_{L_n \in U_\mu} (b(X_1) \cdots b(X_n))^{-1}\, d\mathbb{Q} \\
&= \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu\}} b_n^{-1}(\mathbf{x})] \\
&= \frac{\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}]}{\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}]} \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu\}} b_n^{-1}(X)]
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{1}{n}\log \mathbb{P}[L_n \in U_\nu] &\geq \frac{1}{n}\log\left[\frac{1}{\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}]}\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu\}} b_n^{-1}(X)]\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}]\right] \\
&= \frac{1}{n}\log\left[\frac{1}{\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}]}\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu\}} b_n^{-1}(X)]\right] + \frac{1}{n}\log \mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}] \\
&\geq -\frac{1}{n\mathbb{Q}[\mathbb{1}_{\{L_n \in U_\mu\}}]}\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\mu\}} \log b_n] + \frac{1}{n}\log \mathbb{Q}[\mathbb{1}_{\{L_n \in U_\mu\}}].
\end{aligned}$$

In the third line, in the first term of the second part of the inequality , we used Jensen's inequality in the convex function $-\log x$. Now, we will use the fact that $x \log x \geq -\frac{1}{e}$ to write,

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu\}} \log b_n] &= E^{\mathbb{Q}}[\log b_n] - \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{L_n \in U_\nu^c\}} \log b_n] \\
&= n\mathbb{E}^{\nu}[\log b] - \mathbb{E}^{\mathbb{P}}[\mathbb{1}_{\{L_n \in U_\nu^c\}} b_n \log b_n] \\
&\leq nR(\nu|\mu) + \frac{1}{e}.
\end{aligned}$$

Therefore,

$$\frac{1}{n}\log \mathbb{P}[L_n \in U_\nu] \geq \frac{1}{\mathbb{Q}[L_n \in U_\nu]}\left(-R(\nu|\mu) - \frac{1}{ne}\right) + \frac{1}{n}\log \mathbb{Q}[\mathbb{1}_{\{L_n \in U_\nu\}}].$$

By the Law of Large Numbers, since $\mathbb{Q} \circ X_i^{-1} = \nu$ that is the distribution of X_i's under $\mathbb{Q}$. $\mathbb{Q}[L_n \in U_\nu]$ converges to 1. We, thus, have

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}[U_\nu] \geq -R(\nu|\mu).$$

The proof is now complete, since, we have already showed that $R(\nu|\mu) = I(\nu)$ is a good rate function.

$\square$

Part II

Large Deviations for Stochastic Processes

Chapter 5

Sample path Large Deviations

5.1 Schilder's theorem

In many problems, the interest is in rare events that depended on random process, and the corresponding asymptotics probabilities, usually called sample path large deviations, were developed by Freidlin-Wentzell and Donsker-Varadhan.
The first example is known as Schilder's theorem, and concerns large deviations for the process $\{W_\epsilon\}_\epsilon = \sqrt{\epsilon}W$, as ϵ goes to zero, (the family of rescaled Brownian Motion) where $W = \{W_t\}_{t\in[0,T]}$ is a Brownian motion in $\mathbb{R}^d$. Denote by $C_0[0,1]$ the space of continuous functions on $[0,1]$. The family of paths is defined on an infinite dimensional space: $(\Theta, \|\cdot\|_\infty)$, where

$$\Theta = \left\{\theta \in \left(C_0[0,1];\mathbb{R}^d\right) : \theta(0) = 0\right\}$$

Θ is a separable Banach space with respect to the uniform norm. Now, we identify the dual space of Θ. The dual space of Θ is the set of all vector signed measures on $[0,1]$ and with finite variation. The duality relation is given by:

$$\int_0^1 \sum_{i=1}^d \theta_i(s)\,\lambda(ds) = \langle \lambda, \theta\rangle.$$

We consider $d = 1$ for sake of simplicity.

Having identified C_0' the dual space of C_0 we follow the scheme of Ellis-Gärtner theorem in order to compute for any $\lambda \in C_0'$ the limit

$$\Lambda(\lambda) = \lim_{\epsilon\to 0} \epsilon \log \tilde{\mathbb{Q}}_\epsilon\left(\frac{\lambda}{\epsilon}\right)$$

for $\tilde{\mathbb{Q}}_\epsilon \in C_0'$ and the convex conjugate Λ^* of Λ, that is for $\psi \in C_0$

$$\Lambda^*(\psi) = \sup_{\lambda\in C_0'} \left(\langle \lambda, \psi\rangle - \Lambda(\lambda)\right)$$

Lemma 5.1. *Given, $\lambda \in C_0'$ and $\psi \in \Theta$ then for all $\lambda \in \Theta^*$*

$$\Lambda(\lambda) = \frac{1}{2}\int_0^1 |\lambda[s,1]|^2 ds$$

Proof. First of all, we observe that

$$\omega \to \langle \lambda, W(\omega) \rangle = \int_0^1 W_s(\omega)\,\lambda(ds) \tag{5.1}$$

is a centered Gaussian random variable and it is well known that the mean of the exponential of a centered random variable with variance σ^2 is equal to $e^{\frac{1}{2}\sigma^2}$. So we have to compute σ^2 of $\langle \lambda, W \rangle$.

$$\begin{aligned}
\sigma^2 = \mathbb{E}[\langle \lambda, W \rangle^2] &= \mathbb{E}\left[\left(\int_0^1 W(s)\,\lambda(ds)\right)^2\right] \\
&= \mathbb{E}\left[\int_0^1 W(s)\,\lambda(ds)\int_0^1 W(t)\,\lambda(dt)\right] \\
&= \mathbb{E}\left[\int_0^1\int_0^1 W(s)\,W(t)\,\lambda(dt)\,\lambda(ds)\right] \\
&= \int_0^1 d\lambda(s)\int_0^1 \mathbb{E}[W(s)\,W(t)]\lambda(dt) \\
&= \int_0^1 \lambda(ds)\int_0^1 t\wedge s\lambda(dt)\,.
\end{aligned}$$

However,

$$t\wedge s = \int_0^1 \mathbb{1}_{[0,s]}(u)\,du$$

$$\begin{aligned}
= \int_0^1 t\wedge s\lambda(dt) &= \int_0^1\int_0^t \mathbb{1}_{[0,s]}(u)\,du\lambda(dt) \\
&= \int_0^1 du\int_0^t \mathbb{1}_{[0,s]}(u)\,\lambda(dt) \\
&= \int_0^s \lambda([u,1])\,du.
\end{aligned}$$

Therefore,

$$\mathbb{E}[\langle \lambda, W \rangle^2] = \int_0^1 \lambda(ds)\int_0^s \lambda[u,1]du = \int_0^1 \lambda_[u,1]du\int_u^1 \lambda(ds) = \int_0^1 \lambda^2([u,1])\,du.$$

so,

$$\Lambda(\lambda) = \lim_{\epsilon\to 0}\epsilon\log\tilde{\mathbb{Q}}_\epsilon\left(\frac{\lambda}{\epsilon}\right) = \frac{1}{2}\int_0^1 |\lambda([s,1])|^2 ds.$$

Remark 5.1. *Using stochastic integration, the variance of the random variable defined in 5.1 can be derived more easily by considering that, by integration by parts,*

$$\int_0^1 W(\omega)\lambda(ds) = \int_0^1 \lambda([s,1])\,dW(s)\,.$$

Next define $\mathcal{H}$ be the space of $\psi \in \Theta$, consisting of absolutely continuous functions with the property that

$$\psi(t) = \int_0^1 \dot{\psi}(s)\,ds, \quad \dot{\psi} \in L^2\left([0,1];\mathbb{R}^d\right).$$

$\mathcal{H}$ is called Cameron-Martin space and is Hilbert space with respect to scalar product

$$\langle f, g \rangle = \int_0^1 \dot{f}_s\dot{g}_s ds.$$

Remark 5.2. *It is important to notice that by Hölder's inequality the paths in $\mathcal{H}$ are Hölder continuous of index $\frac{1}{2}$ for $0 < s < 1$*

$$|\psi_s - \psi_t| = \left|\int_s^t \dot{\psi}_u du\right| \leq |t-s|^{1/2}\left(\int_s^t |\dot{\psi}_s|^2 ds\right)^{1/2} \leq \|\psi\|_{\mathcal{H}}|t-s|^{1/2}. \quad (5.2)$$

This implies that bounded sets of paths in $\mathcal{H}$ are uniformly bounded and uniformly equicontinuous. Thus, by ths Ascoli-Arzela theorem, bounded sets of $\mathcal{H}$ are relative compact in $C_0[0,1]$. We will use this argument later in order to construct relative compact set in $\mathcal{H}$ and prove the upper bound.

Lemma 5.2. *For any $\psi \in C_0$ the Legendre transform of $\Lambda(\lambda)$ is*

$$\Lambda^*(\psi) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds & \psi \in \mathcal{H} \\ \infty & \textit{otherwise} \end{cases}$$

Proof. The Legendre transform of $\Lambda(\lambda)$ will be

$$\Lambda^*(\psi) = \sup_{\lambda \in C_0'} \{\langle \lambda, \psi\rangle - \Lambda(\lambda)\}$$

$$= \sup_{\lambda \in C_0'} \left\{\int_0^1 \psi(s)\, d\lambda_s - \frac{1}{2}\int_0^1 |\lambda([s,1])|^2 ds\right\}$$

$$= \sup_{\lambda \in C_0'} \left\{\int_0^1 \lambda([s,1])\, \dot{\psi}_s ds - \frac{1}{2}\int_0^1 |\lambda([s,1])|^2 ds\right\}$$

$$= \sup_{\lambda \in C_0'} \left\{-\frac{1}{2}\int_0^1 |\lambda([s,1]) - \dot{\psi}_s|^2 ds + \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds\right\}$$

$$\leq \frac{1}{2}\int_0^1 |\dot{\psi}(s)|^2 ds. \quad (5.3)$$

In order to prove that the equality holds we need to consider the following lemma.

Lemma 5.3. *$h \in \mathcal{H}$ if and only if*

$$\lim_{n\to\infty} 2^n \sum_{i=1}^{2^n} \left|h\left(\frac{i}{2^n}\right) - h\left(\frac{i-1}{2^n}\right)\right|^2 < \infty \quad (5.4)$$

and is equal to $\|h\|_{\mathcal{H}}^2 = \int_0^1 \left|\dot{h}_s\right|^2 ds$, where $\mathcal{H}$ is Cameron-Martin space.

Proof. Suppose that $h \in \mathcal{H}$ then we devide $[0,1]$ in 2^n equal parts. Then there is $s \in \left(\frac{i-1}{2^n}, \frac{i}{2^n}\right) \subset [0,1]$. We define

$$g_n(s) = 2^n \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} \dot{h}_s ds = 2^n \left(h\left(\frac{i}{2^n}\right) - h\left(\frac{i-1}{2^n}\right)\right).$$

Then

$$g_n(t) = \int_0^1 \dot{g}_n(s)\, ds \quad (5.5)$$

Thus this function gives us h for each point $\frac{i}{2^n}$, $i = 0, 1, 2, \ldots, 2^n$. But this means that g is linear between two such points, since it is linear interpolation of h at these points $\frac{i}{2^n}$. Then the norm of the interpolating function is

$$|g_n|_{\mathcal{H}} = 2^n \sum_{i=1}^{2^n} \left| h\left(\frac{1}{2^n}\right) - h\left(\frac{i-1}{2^n}\right) \right|^2 < \infty.$$

We need to prove that $g_n \to h$ in $\mathcal{H}$. We consider the σ-algebra which is generetad by these intervals. More precisely, we define

$$\mathcal{F}_n = \left\{ \left[\frac{i-1}{2^n}, \frac{i}{2^n}\right] : 1 \leq i \leq 2^n \right\}.$$

Therefore,

$$\dot{g}_n = \mathbb{E}[\dot{h}|\mathcal{F}_n]$$

which means that $\dot{g}_n$ is $\mathcal{F}_n$-martingale and bounded in $\mathcal{H}$ and $\dot{g}_n$ converges to $\dot{h}$.
Now, suppose that 5.4 holds. We want to prove that $h \in \mathcal{H}$. It suffices to show that

$$\mathbb{E}[\dot{g}_{n+1}|\mathcal{F}_n] = \dot{g}_n$$

which means that $\dot{g}_n$ is $\mathcal{F}_n$ martingale. Since $\mathcal{F}_n$ is generated by a partition of $[0, 1]$ we know that

$$\mathbb{E}[\dot{g}_{n+1}|\mathcal{F}_n] = \frac{1}{\frac{1}{2^n}} \int_{\frac{i-1}{2^n}}^{\frac{i}{2^n}} \dot{h}_s ds = \dot{g}_n < \infty$$

Then $\dot{g}_n$ is a martingale with respect to $\mathcal{F}_n$ which is bounded by 5.4 in $\mathcal{L}^2$. Thus there exists $\dot{g}$ such that $\dot{g}_n \to \dot{g}$ in $\mathcal{L}^2$, which means that

$$\dot{g}_n = \mathbb{E}[\dot{g}|\mathcal{F}_n].$$

Since

$$\int_0^t \dot{g}_n(s)\, ds \to \int_0^t \dot{g}(s)\, ds$$

then necessarily $h(s) = \int_0^t \dot{g}(s)$ and $h \in \mathcal{H}$.

□

Now, we construct a sequence of measures using the above lemma in order to prove that the equality holds in 5.3.

$$\lambda_n = 2^n \sum_{i=1}^{2^n} \left| \psi\left(\frac{i}{2^n}\right) - \psi\left(\frac{i-1}{2^n}\right) \right| \left(\delta_{\frac{i}{2^n}} - \delta_{\frac{i-1}{2^n}} \right).$$

We set $b_{n=}\delta_{\frac{i}{2^n}} - \delta_{\frac{i-1}{2^n}}$, therefore

$$b_n([u, 1]) = \begin{cases} 1 & \dfrac{i-1}{2^n} \leq u \leq \dfrac{i}{2^n} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\lambda_n([u, 1]) = 2^n \left| \psi\left(\frac{i}{2^n}\right) - \psi\left(\frac{i-1}{2^n}\right) \right| \qquad \text{for } \tfrac{i-1}{2^n} \leq u \leq \tfrac{i}{2^n}$$

and

$$\Lambda\left(\lambda_n\right)=\frac{1}{2}2^n\sum_{i=1}^{2^n}\left|\psi\left(\frac{i}{2^n}\right)-\psi\left(\frac{i-1}{2^n}\right)\right|^2.$$

Finally, we get

$$\begin{aligned}\langle\lambda_n,\psi\rangle&=2^n\sum_{i=1}^{2^n}\left|\psi\left(\frac{i}{2^n}\right)-\psi\left(\frac{i-1}{2^n}\right)\right|^2-\frac{1}{2}2^n\sum_{i=1}^{2^n}\left|\psi\left(\frac{i}{2^n}\right)-\psi\left(\frac{i-1}{2^n}\right)\right|^2\\&=\frac{1}{2}2^n\sum_{i=1}^{2^n}\left|\psi\left(\frac{1}{2^n}\right)-\psi\left(\frac{i-1}{2^n}\right)\right|^2\to\frac{1}{2}\int_0^1\left|\dot{\psi}_s\right|^2ds\qquad\text{as } n\to\infty.\end{aligned}$$

if $\psi\in\mathcal{H}$. $\square$

Now, we are ready to prove Schilder's theorem. Schilder's theorem gives us an estimate for the probability that a rescaled B.M. will stray far from the mean path which is constant with value 0.

Theorem 5.1. *Let $\mathbb{Q}$ be the Wiener measure on $C_0[0,1]$ and for $\epsilon>0$ let $\mathbb{Q}_\epsilon$ its image through the mapping $W\to\sqrt{\epsilon}W$. Then the family of measures $\{\mathbb{Q}_\epsilon\}_\epsilon$ satisfies LDP with good rate function given by $\Lambda^*\left(\psi\right)$ for $\psi\in C_0$.*

Proof. Upper bound

First step: We shall prove the upper bound for a small neighborhood of ψ, $B_\delta\left(\psi\right)$ open balls of radius δ,

$$B_\delta\left(\psi\right)=\left\{f\in C_0[0,1]:\sup\left|f\left(t\right)-\psi\left(t\right)\right|<\delta\right\}.$$

We need to show that

$$\limsup_{\epsilon\to0}\epsilon\log\mathbb{Q}_\epsilon[B_\delta\left(\psi\right)]\le-\inf_{\psi}\Lambda^*_{\mathbb{Q}}\left(\psi\right).\tag{5.6}$$

$$\begin{aligned}\mathbb{Q}_\epsilon[B_\delta\left(\psi\right)]=\mathbb{P}[\sqrt{\epsilon}W\in B_\delta\left(\psi\right)]&=\mathbb{P}[\sup_{0\le t\le1}|\sqrt{\epsilon}W_t-\psi_t|<\delta]=\mathbb{P}\left[\sup|\sqrt{\epsilon}\left(W_t-\frac{\psi_t}{\sqrt{\epsilon}}\right)<\delta\right]\\&=\mathbb{P}\left[\sup_{0\le t\le1}|W_t-\frac{\psi_t}{\sqrt{\epsilon}}<\frac{\delta}{\sqrt{\epsilon}}\right]=\mathbb{Q}\left[B_{\frac{\delta}{\sqrt{\epsilon}}}\left(\frac{\psi}{\sqrt{\epsilon}}\right)\right]\\&=\mathbb{E}^{\mathbb{Q}}\left[\mathbb{1}_{B_{\frac{\delta}{\sqrt{\epsilon}}}\left(\frac{\psi}{\sqrt{\epsilon}}\right)}\right]\\&\le\int_{B_{\frac{\delta}{\sqrt{\epsilon}}}\left(\frac{\psi}{\sqrt{\epsilon}}\right)}\exp[\langle\lambda,W\rangle-\inf_{\phi\in B_\delta(\psi)}\langle\lambda,\phi\rangle]d\mathbb{Q}\\&\le e^{-\inf_{\phi\in B_\delta(\psi)}\langle\lambda,\phi\rangle}\int_{B_{\frac{\delta}{\sqrt{\epsilon}}}\left(\frac{\psi}{\sqrt{\epsilon}}\right)}e^{\langle\lambda,W\rangle}d\mathbb{Q}\\&\le e^{-\langle\lambda,\psi\rangle}\int_{B_{\frac{\delta}{\sqrt{\epsilon}}}(0)}e^{\langle\lambda,W-\frac{\psi}{\sqrt{\epsilon}}\rangle}d\mathbb{Q}.\end{aligned}$$

However by integration by parts we get

$$\langle \lambda, W \rangle = \int_0^1 W_s d\lambda_s = -\int_0^1 \lambda\left([s,1]\right) dW_s \leq \frac{1}{2}\int_0^1 \lambda^2\left([s,1]\right) ds.$$

Then we have that

$$\begin{aligned}
\mathbb{Q}_\epsilon[B_\delta\left(\psi\right)] &\leq e^{\langle \lambda, \psi \rangle} \int_{B_{\frac{\delta}{\sqrt{\epsilon}}}(0)} e^{\langle \lambda, W\rangle - \langle \lambda, \frac{\psi}{\sqrt{\epsilon}}\rangle} d\mathbb{P} \\
&< e^{-\left(\langle \lambda, \psi\rangle + \langle \lambda, \frac{\psi}{\sqrt{\epsilon}}\rangle\right)} \int_{B_{\frac{\delta}{\sqrt{\epsilon}}}(0)} e^{\frac{1}{2}\int_0^1 \lambda^2([s,1])ds} d\mathbb{P} \\
&< e^{-\frac{1}{\epsilon}\left(\langle \lambda, \psi\rangle - \frac{1}{2}\int_0^1 \lambda^2([s,1])ds\right)} \int_{B_{\frac{\delta}{\sqrt{\epsilon}}}(0)} d\mathbb{P} \\
&= e^{\frac{\Lambda^*(\psi)}{\epsilon}} \mathbb{P}[B_{\frac{\delta}{\sqrt{\epsilon}}}(0)].
\end{aligned}$$

Thus, if we take the logarithm and let $\epsilon \to 0$ we obtain the proper upper bound for open balls

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[B_\delta\left(\psi\right)] \leq -\inf \Lambda^*\left(\psi\right). \tag{5.7}$$

Second step: Now, we will show the upper bound for $K \subset \Theta$ compact set. If K is compact then $K \subset \bigcup_{i=1}^n B_{\delta_i}\left(\psi_i\right)$. The technique is similar to that of Cramér theorem in $\mathbb{R}^d$ and by 5.7:

$$\begin{aligned}
\limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[K] &\leq \max_i \limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[B_{\delta_i}\left(\psi_i\right)] \\
&\leq \max_i \left(-\inf \Lambda^*\left(\psi_i\right)\right) \leq -\inf \Lambda^*\left(\psi\right).
\end{aligned}$$

Third step: Finally, we prove the upper bound for $F \subset \Theta$ closed sets. Firstly, we make the following observation

$$F = (F \cap K) \cup (F \setminus K) \subset (F \cap K) \cup K^c.$$

Therefore

$$\begin{aligned}
\limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[F] &\leq \limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[F \cap K] \vee \limsup_{\epsilon \to 0} \log \mathbb{Q}_\epsilon[K^c] \\
&\leq -\inf_{\psi \in F \cap K} \Lambda^*\left(\psi\right) \vee \limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[K^c].
\end{aligned}$$

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[F] \leq -\inf_{\psi \in F} \Lambda^*\left(\psi\right) \vee \limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[K^c]. \tag{5.8}$$

The problem is to remove compactness restriction from the above inequality. The idea is to construct a compact set K_L, $L > 0$ of paths with the property that

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbb{Q}_\epsilon[K_L^c] \leq -L. \tag{5.9}$$

What this mean is that as $L \to \infty$ the events become so deviant that they cannot even been seen on the scale at which we are looking. Therefore, they cannot contribute to our calculation.

Recall that 5.2, because of Ascoli-Arzelá theorem, all sets which are uniformly bounded and equicontinuous are relative compact. For $0 < \alpha < \frac{1}{2}$ let us consider the Hölder norm on $C_0[0,1]$

$$|\psi|_\alpha = \sup_{0 \leq s \leq t \leq 1} \frac{|\psi\left(t\right) - \psi\left(s\right)|}{|t-s|^\alpha}$$

But it is well-known that the paths of the Brownian motion are α-Hölder continuous for every $\alpha < \frac{1}{2}$, so that $|\cdot|_\alpha$ is finite $\mathbb{Q}$-almost everywhere. Indeed, $K_L = \left\{|\psi|_\alpha \leq \sqrt{L}/c\right\}$ are relative compact because of 5.2, so we have that

$$\mathbb{Q}_\epsilon[K_L^c] = \mathbb{P}\left[|\sqrt{\epsilon}W|_\alpha > \frac{\sqrt{L}}{C}\right] = \mathbb{P}\left[e^{C|W|_\alpha^2} > e^{\frac{L}{\epsilon}}\right] \leq \mathbb{E}\left[e^{C|W|_\alpha^2}\right] e^{\frac{-L}{\epsilon}} \tag{5.10}$$

the last inequality is due to Markov inequality. We need to show that

$$\mathbb{E}\left[e^{C|W|_\alpha^2}\right] = c_1 < \infty \tag{5.11}$$

There are several ways one can establish the property 5.11. The method which will adopt here will be to construct a function $\Phi : \Theta \to [0, \infty]$ such that:

(1) Φ is sub-additive

(2) $\Phi(a\theta) = |a|\Phi(\theta)$ for all $a \in \mathbb{R}$ and $\theta \in \Theta$.

(3) $\mathbb{P}[\{\theta : \Phi(\theta) < \infty\}] = 1$

In order to construct such a Φ and to pass from the fact that it exists to 5.9, we will make use of the following beatiful and powerful estimate due to X. Fernique.

Theorem 5.2. *(Fernique). Let $\mathcal{X}$ be a real seperable Fréchét space and $\Phi : \mathcal{X} \to [0, \infty]$ a measurable function sub-additive function with the property that $\Phi(ax) = |a|\Phi(x)$ for all $a \in \mathbb{R}$ and $x \in \mathcal{X}$. Next define μ a probability measure on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ with the property that μ^2 on $\left(\mathcal{X}^2, \mathcal{B}_{\mathcal{X}^2}\right)$ is invariant under the transformation:*

$$F(x, y) = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right) \in \mathcal{X}^2.$$

If $\mu[\{x : \Phi(x) < \infty\}] = 1$ then there exists an $a > 0$ for which

$$\int_\mathcal{X} e^{a\Phi(x)^2} d\mu(x) < \infty.$$

Remark 5.3. *Let us see how this result allows us to derive 5.11. Indeed, it is enough to apply it to the semi-norm $\Phi = |\cdot|_\alpha$, $0 < \alpha < 1/2$ and for $\mathcal{X} = C_0[0, 1]$, μ is the Wiener measure. Since we know that the paths of the Brownian motion are α-Hölder continuous for every $\alpha < 1/2$ it holds that $\Phi < \infty$ μ-a.s.. The invariance property of $\mu \otimes \mu$ under F also comes easily from the fact that $\mu \otimes \mu$ is the law on $C \times C$ of a bi-dimensional Brownian motion $W = (W_1, W_2)$ and Wiener measure is invariant under the $\frac{\pi}{4}$ rotation. To prove the invariance property is equivalent to show that if*

$$V_1(t) = \frac{1}{\sqrt{2}}(W_1(t) + W_2(t)), \quad V_2(t) = \frac{1}{\sqrt{2}}(W_1(t) - W_2(t))$$

then $V = (V_1, V_2)$ is still a bi-dimensional Brownian motion, which is immediate as V has the same finite dimensional distributions of W.

Proof. Given that $0 < s < t$ and $A = \left\{(x, y) \in \mathcal{X}^2 : \phi(x) \leq s, \phi(\psi) \geq t\right\}$ we have

$$\mu \otimes \mu(A) = \mu \otimes \mu\left(\left\{(x, y); \phi\left(\frac{x-y}{\sqrt{2}}\right) \leq s, \phi\left(\frac{x+y}{\sqrt{2}}\right) \geq t\right\}\right)$$

$$= \mu \otimes \mu \left(\left\{ (x,y) ; \phi (x-y) \leq \sqrt{2}s, \phi (x+y) \geq \sqrt{2}t \right\} \right)$$

$$\leq \mu \otimes \mu \left(\left\{ |\phi (x) - \phi (y)| \leq \sqrt{2}s, \phi (x) + \phi (y) \geq \sqrt{2}t \right\} \right)$$

$$\leq \mu \otimes \mu \left(\left\{ (x,y) : -\max (\phi (x), \phi (y)) + \min (\phi (x), \phi (y)) \geq -\sqrt{2}s, \right. \right.$$

$$, \max (\phi (x), \phi (y)) + \min (\phi (x) + \phi (y)) \geq \sqrt{2}t$$

$$\leq \mu \otimes \mu \left(\left\{ (x,y), \min (\phi (x), \phi (y)) \geq \frac{1}{\sqrt{2}} (t-s) \right) \right.$$

$$= \left(\mu \left(x : \phi (x) \geq \frac{1}{\sqrt{2}} (t-s) \right) \right)^2 .$$

In the third inequality, we used the argument that

$$|\phi (x) - \phi (y)| = \max (\phi (x), \phi (y)) - \min (\phi (x), \phi (y)) .$$

Thus we have,

$$\mu (\{x : \phi (x) \leq s\}) \cdot \mu (\{y : \phi (y) \geq t\}) \leq \left(\mu \left(\left\{ x : \phi (x) \geq \frac{1}{\sqrt{2}} (t-s) \right\} \right) \right)^2 . \tag{5.12}$$

Since ϕ is finite μ-a.s. there exists $s \geq 0$ such that

$$\mu (\{x : \phi (x) \leq s\}) > \frac{1}{2}.$$

Thus we define by recurrence a sequence $(t_n)_n$ by

$$t_0 = s \quad t_n = \sqrt{2} t_{n-1} + s$$

$$t_n = s \left(1 + \sqrt{2} + \left(\sqrt{2} \right)^2 + \cdots + \left(\sqrt{2} \right)^n \right) = s \frac{\left(\sqrt{2} \right)^{n+1} - 1}{\sqrt{2} - 1} \leq s(\underbrace{\frac{\sqrt{2}}{\sqrt{2}-1}}_{=\sqrt{b}}) \cdot 2^{n/2}.$$

By 5.12 we derive that

$$\mu (\phi \leq s) \mu (\phi \geq t_n) \leq (\mu (\phi \geq t_{n-1}))^2 .$$

Iterating this inequality we obtain:

$$\frac{\mu (\phi \geq t_n)}{\mu (\phi \leq s)} \leq \left(\underbrace{\frac{\mu (\phi \geq t_{n-1})}{\mu (\phi \leq s)}}_{\leq 1} \right)^2 = e^{-2^n c}$$

where $c = -\log \left[\frac{\mu(\phi(x) \geq s)}{\mu(\phi(x) \leq s)} \right]$
and therefore

$$\int_{\mathcal{X}} e^{a\phi^2(x)} \mu dx \leq \underbrace{\int_{\{\phi \leq \sqrt{b}\}} e^{a\phi^2(x)} \mu dx}_{\leq e^{ab}} + \int_{\{\phi \geq \sqrt{b}\}} e^{a\Phi^2(x)} \mu dx.$$

$$\int_{\{\phi\geq\sqrt{b}\}} e^{a\phi^2(x)}\mu(dx) \leq \sum_{n=1}^{\infty}\int_{\{b2^n\leq\phi^2\leq b2^{n+1}\}} e^{a\phi^2(x)}\mu(dx)$$

$$\leq \sum_{n=1}^{\infty} e^{ab2^{n+1}}\mu\left(\{\phi^2\geq b2^n\}\right).$$

Since $t_n^2 \leq b2^n$ 5.12 gives

$$\int_{\{\phi\leq\sqrt{b}\}} e^{a\phi^2(x)}\mu(dx) \leq \sum_{n=1}^{\infty} e^{ab2^{n+1}}\mu(\phi\geq t_n)$$

$$\leq \sum_{n=1}^{\infty} e^{ab2^{n+1}}\mu(\phi\geq s)\, e^{-2^n c}$$

which for $2ab < c$ gives convergent series and this concluded the proof. □

Using the above theorem, the statement 5.10 holds and so the family of measures $\{\mathbb{Q}_\epsilon\}$ is exponential tight and the statement 5.9 holds too. Therefore, by 5.8 we have the proper upper bound for closed sets.

Lower bound

Let $\psi \in H^1$. We want to prove that

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbb{Q}_\epsilon[B_\delta(\psi)] \geq -\inf_{\psi\in H^1}\Lambda^*(\psi)$$

$B_\delta(\psi)$ being the open ball radius δ centered at ψ in the uniform norm. The idea is, as always for a lower bound, based in a change of measure.

$$\mathbb{Q}_\epsilon[B_\delta(\psi)] = \mathbb{P}[\sqrt{\epsilon}W \in B_\delta(\psi)] = \mathbb{P}[\sqrt{\epsilon}W \in B_\delta(\psi)]$$

$$= \mathbb{P}[\sqrt{\epsilon}\left(W - \frac{\psi}{\sqrt{\epsilon}}\right) \in B_\delta(0)]$$

$$\mathbb{P}[W - \frac{\psi}{\sqrt{\epsilon}} \in D_{\frac{\delta}{\sqrt{\epsilon}}}(0)]$$

By Gisranov theorem the law of $(W_t - \psi\sqrt{\epsilon}) = \tilde{W}$ has a density with respect to Wiener measure $\mathbb{Q}$ which is given by:

$$\frac{d\mathbb{Q}_{\epsilon,\psi}}{d\mathbb{Q}_\epsilon} = \exp\left(\frac{1}{\sqrt{\epsilon}}\int_0^1 \dot{\psi}_s dW_s - \underbrace{\frac{1}{2\epsilon}\int_0^1 |\dot{\psi}_s|^2 ds}_{=\frac{\Lambda^*(\psi)}{\epsilon}}\right).$$

Thus,

$$\mathbb{P}[\sqrt{\epsilon}\left(W - \frac{\psi}{\sqrt{\epsilon}}\right) \in B_\delta(0)] = \mathbb{Q}_{\epsilon,\psi}[B_\delta(0)] = \int_{B_\delta(0)} \frac{d\mathbb{Q}_{\epsilon,\psi}}{d\mathbb{Q}_\epsilon} d\mathbb{Q}_\epsilon$$

$$= \int_{B_\delta(0)} \exp\left(\frac{1}{\sqrt{\epsilon}}\int_0^1 \dot{\psi}_s dw_s - \frac{\Lambda^*(\psi)}{\epsilon}\right) d\mathbb{Q}_\epsilon$$

$$= e^{-\frac{\Lambda^*(\psi)}{\epsilon}} \int_{B_\delta(0)} \exp\left(\frac{1}{\sqrt{\epsilon}}\int_0^1 \dot{\psi}_s dw_s\right) d\mathbb{Q}_\epsilon$$

$$= e^{-\frac{\Lambda^*(\psi)}{\epsilon}} \mathbb{Q}_\epsilon[B_\delta(0)] \frac{1}{\mathbb{Q}_\epsilon[B_\delta(0)]} \int_{B_\delta(0)} \exp\left(\frac{1}{\sqrt{\epsilon}} \int_0^1 \dot{\psi}_s dw_s\right) d\mathbb{Q}_\epsilon$$

$$\geq e^{-\frac{\Lambda^*(\psi)}{\epsilon}} \mathbb{Q}_\epsilon[B_\delta(0)] \exp \frac{1}{\mathbb{Q}_\epsilon[B_\delta(0)]} \int_{B_\delta(0)} \left(\frac{1}{\sqrt{\epsilon}} \int_0^1 \dot{\psi}_s dw_s\right) d\mathbb{Q}_\epsilon.$$

In the last display the inequality is due to Jensen's inequality. Furthermore,

$$\int_{B_\delta(0)} \left(\frac{1}{\sqrt{\epsilon}} \int_0^1 \dot{\psi}_s dw_s\right) d\mathbb{Q}_\epsilon = 0$$

since of Brownian motion symmetry $W \sim -W$ and $\mathbb{Q}_\epsilon[B_\delta(0)] \to 1$ as $\epsilon \to 0$ since

$$\mathbb{Q}_\epsilon[B_\delta(0)] = \mathbb{P}[\sqrt{\epsilon}W \in B_\delta(0)] = \mathbb{P}[\sup_{0\leq t\leq 1} |\sqrt{\epsilon}W_t| < \delta]$$

$$= \mathbb{P}[\sup_{0\leq t\leq 1} |W_t| < \frac{\delta}{\sqrt{\epsilon}}] = 1 - \mathbb{P}[\sup_{0\leq t\geq 1} |W_t| > \frac{\delta}{\sqrt{\epsilon}}].$$

But

$$\mathbb{P}[\sup_{0\leq t\leq 1} |W_t| > \frac{\delta}{\sqrt{\epsilon}}] = \mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] + \mathbb{P}[\inf_{0\leq t\leq 1} W_t < -\frac{\delta}{\sqrt{\epsilon}}]$$

$$= \mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] + \mathbb{P}[-\sup_{0\leq t\leq 1} -W_t < -\frac{\delta}{\sqrt{\epsilon}}]$$

$$= 2\mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] \qquad \text{since } W \sim -W$$

We consider the exponential martingale for $\lambda > 0$: $e^{\lambda W_t - \frac{\lambda^2}{2}t} = M_t$ then

$$\mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] \leq \mathbb{P}[\sup_{0\leq t\leq 1} M_t > e^{\lambda\frac{\delta}{\sqrt{\epsilon}} + \frac{\lambda^2}{2}t}]$$

$$= \mathbb{P}[e^{-\lambda\frac{\delta}{\sqrt{\epsilon}} + \frac{\lambda^2}{2}t} \sup_{0\leq t\leq 1} M_t \geq 1] \leq e^{-\lambda\frac{\delta}{\sqrt{\epsilon}} + \frac{\lambda^2}{2}} \mathbb{E}[M_t]$$

$$= e^{-\lambda\frac{\delta}{\sqrt{\epsilon}} + \frac{\lambda^2}{2}} \mathbb{E}[M_0] \leq e^{-\lambda\frac{\delta}{\sqrt{\epsilon}} + \frac{\lambda^2}{2}}.$$

By minimizing the latter expression in $\lambda > 0$ we have that

$$\mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] \leq e^{-\frac{1}{2}\frac{\delta^2}{\epsilon}}$$

which leads to

$$\mathbb{P}[\sup_{0\leq t\leq 1} |W_t| > \frac{\delta}{\sqrt{\epsilon}}] \leq 2e^{-\frac{1}{2}\frac{\delta^2}{\sqrt{\epsilon}}}.(5.13)$$

Otherwise, by reflection principle we have that

$$\mathbb{P}[\sup_{0\leq t\leq 1} W_t > \frac{\delta}{\sqrt{\epsilon}}] = 2\mathbb{P}[W_1 > \frac{}{\sqrt{\epsilon}}] = 2\int_{\frac{\delta}{\sqrt{\epsilon}}}^{\infty} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}.$$

Therefore,

$$\mathbb{Q}_\epsilon[B_\delta(\psi)] \geq e^{-\frac{\Lambda^*(\psi)}{\epsilon}}$$

and

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbb{Q}_\epsilon[B_\delta(\psi)] \geq -\inf_{\psi\in\mathcal{H}} \Lambda^*(\psi).$$

□

We now give some examples of Schilder theorem.
First example
Suppose that W is 1-dimensional Brownian motion. We estimate the rate of decay of the probability

$$\mathbb{P}[\sup_{0\le t\le 1} W_t \ge M]$$

as $M \to \infty$

Proof.

$$\mathbb{P}[\sup_{0\le t\le 1} W_t \ge M] = \mathbb{P}[\frac{1}{M}\sup_{0\le t\le 1} W_t \ge 1] = \mathbb{P}[\sup_{0\le t\le 1}\frac{1}{M} W_t > 1]$$

We will use Schilder's theorem to prove that

$$\mathbb{P}[\frac{1}{M}W \in A] \le e^{-\inf_{\psi\in A}\Lambda^*(\psi)}$$

where

$$A = \{\psi : \psi \in C_0[0,1], \psi(t) \ge 1, \quad \text{for some} \quad t > 1\}$$

So, we have to estimate the infimumm over A of $\Lambda^*(\psi) = \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds$.
First case if $\psi(t) = 1$ then by Cauchy-Schwartz inequality

$$1 = \int_0^t \dot{\psi}_s ds \le \sqrt{t}\left(\int_0^t |\dot{\psi}_s|^2 ds\right)^{\frac{1}{2}} \le \sqrt{t}|\psi|_{\mathcal{H}}$$

then —$\psi|_{\mathcal{H}} > \frac{1}{\sqrt{t}}.So,\Lambda^*(\psi) = \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds = \frac{1}{2}|\psi|_{\mathcal{H}} \le \frac{1}{2}$ for $t = 1$.
Second case if $\psi(t) = t \Rightarrow \dot{\psi}(t) = 1$. This means that $\Lambda^*(\psi) = \frac{1}{2}\int_0^1 ds = \frac{1}{2}$ then $\inf_{\psi\in A}\Lambda^*(\psi) = \frac{1}{2}$. But A is a closed set of paths then from the upper bound of Schilder theorem we have that

$$\lim_{M\to\infty}\sup\frac{1}{M^2}\log\mathbb{P}[\frac{1}{M}W \in A] \le -\inf_{\psi\in A}\Lambda^*(\psi) \le -\frac{1}{2}.$$

Since the infimum is attained into A we have that

$$\inf_{\psi\in A}\Lambda^*(\psi) = \inf_{\psi\in A^o}\Lambda^*(\psi).$$

Then we get that

$$\limsup_{M\to\infty}\log\mathbb{P}[\sup_{t\le 1} W_t \ge M] = -\frac{1}{2}$$

However, by the reflection principle we get that

$$\mathbb{P}\left[\sup_{t\le 1} W_t \le M\right] = 2\mathbb{P}[W_1 \le 1] = 2\frac{1}{\sqrt{2\pi}}\int_M^\infty e^{-\frac{x^2}{2}}dx \sim \frac{2}{M\sqrt{2\pi}}e^{-\frac{M^2}{2}}.$$

□

Second example
Suppose that W is a d-dimensional Brownian motion, D is an open set containing the origin. We define

$$\tau = \inf\{t : W_t \notin D\}.$$

We estimate the exponential rate of decay towards 0 of the probability $\mathbb{P}[\tau \le t]$ as $t \to 0$.

Proof. First, we observe that

$$\{\tau \leq t\} = \{w, W_s(w) \notin D \quad \text{for some} \quad s \leq t\}$$

and $\frac{W_{st}}{\sqrt{t}} \overset{d}{=} W_s$. Thus,

$$\begin{aligned}\{\tau \leq t\} &= \{w, W_s(w) \notin D \quad \text{for some} \quad s \leq t\} \\ &= \{w, W_{st}(w) \notin D \quad \text{for some} \quad s \leq 1\} \\ &= \left\{\sqrt{t}\left(\frac{W_{st}}{\sqrt{t}} \notin D \quad \text{for some} \quad s \leq 1\right\}\right. \\ &= \left\{\sqrt{t}W \in A_D\right\}\end{aligned}$$

where $A_D = \{\psi, \psi(s) \notin D \quad \text{for some} \quad s \leq 1\}$. Then we can apply Schilder's theorem and we have to compute $\inf_{\psi \in A_D} \Lambda^*(\psi)$. Let $x \in \partial D$ be the point inthe boundary of D which minimizes the distance from the origin. If $\psi(t) = tx$ is the line segment joining the origin to x then $\Lambda^*(\psi) = \frac{1}{2}\int_0^1 |\dot{\psi}_s^2 ds = \frac{|x|^2}{2}$.
If $\psi \in \mathcal{H}$ is any path such that $\psi(t) = z \in \partial D$ then

$$z = \left|\int_0^t \dot{\psi}_s ds\right| \leq \sqrt{t}\left(\int_0^s |\dot{\psi}_s|^2 ds\right)^{\frac{1}{2}} \leq \sqrt{s}|\psi|_{\mathcal{H}}.$$

So,

$$\Lambda^*(\psi) = \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds = \frac{1}{2}|\psi|^2_{\mathcal{H}} \geq \frac{|z|^2}{2}$$

and $\inf_{\psi \in A_D} \Lambda^*(\psi) = \frac{d^2}{2}$ where $d = dist(x, \partial D)$. Therefore,

$$\liminf_{t \to 0} t \log \mathbb{P}[\tau \leq t] = \limsup_{t \to 0} t \log \mathbb{P}[\tau \leq t] = -\frac{d^2}{2}.$$

$\square$

Chapter 6

Two applications of Schilder's theorem

6.1 Strassen theorem

Let $\mathbb{P}$ be the Wiener measure on the space $\Omega = C\left(([0,1)\right)$ of continuous functions on $[0,1]$ that starts from the point 0 at time 0. For $n \geq 3$ we define the rescaled process:

$$\xi_n\left(t, W\right) = \frac{W\left(nt\right)}{\sqrt{n \log \log n}} = \frac{W_{n\cdot}}{\sqrt{n \log \log n}} \tag{6.1}$$

As $n \to \infty$, $\xi_n\left(t, W\right) \to 0$ in probability with respect to $\mathbb{P}$. But, the convergence will not be almost sure. Our first goal in this section will be to show how Schilder's theorem provides the key estimates in the proof of Strassen's law of the iterated logarithm.

Remark 6.1. *The original proof of Strassen made use of some special feature of Brownian motion, the exact knowledge of its distribution and the reflection principle.*

Theorem 6.1. *(Strassen)For $n \geq 3$ define:*

$$\xi_n\left(t, W\right) = \frac{W\left(nt\right)}{\sqrt{n \log \log n}}, \quad \left(t, W\right) \in [0,1] \times \Theta.$$

and set $K = \left\{\psi \in H : \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds \leq 1\right\}$ then for $\mathbb{P}$-almost sure the sequence $\left\{\xi_n\left(\cdot\right)\right\}_3^\infty$ has the following property:

The family $\left\{\xi_n\left(\cdot\right) : n \geq 3\right\}$ is relative compact and K is its limit set.

Corollary 6.1. *For every $\psi \in K$ there is a subsequence of $\left\{\xi_n\left(\cdot\right)\right\}_3^\infty$ which converges to ψ in Θ.*

Remark 6.2. *Notice that $\left\{\xi_n\left(\cdot\right)\right\}_{n\geq 3}$ contains in a compressed form the whole sample path $W\left(t\right)$.*

Proof. The proof consists in three applications of the Borel-Cantelli lemma, the large deviations estimates being necessary in order to prove that the series converge or diverge. (i) First we shall prove that all limit points of $\left\{\xi_n\left(\cdot\right)\right\}_{n\geq 3}$ lie in K as $n \to \infty$. If K_δ is a neighborhood of size δ around K then for almost all w $\xi_n\left(\cdot\right) \in \mathrm{K}_\delta$ for sufficiently large n. We need to show that for every $\delta > 0$ there exists $n \geq n_0\left(w\right)$ such that

$$d\left(\xi_n\left(w\right), K\right) < \delta \quad \forall n \geq n_0$$

This is proved in two steps.
Firstly, we sample $\xi_n(\cdot)$ along a discrete sequence $n = \rho^m$ for some $\rho > 1$ and show that almost surely, for any such ρ, $\xi_{\rho^m}(\cdot) \in K_\delta$ for sufficiently large n. $K_\delta = \{\psi \in C_0 : d(\psi, K) < \delta\}$, is an open set thus K_δ^c is a closed set.
Moreover, $\inf(\Lambda^*(\psi); \psi \notin K_\delta) > 1$. Indeed, this is obvious since the good rate function Λ^* always attains its infimum of the closed sets. Thus, there exists $\psi \in K_\delta^c$ such that

$$\inf(\Lambda^*(\psi), \psi \notin K_\delta) = \Lambda^*(\psi_0) = \lambda > 1.$$

The key remark here is that $\frac{W_{\cdot t}}{\sqrt{t}} \overset{d}{=} W_\cdot$.
Thus, for every $\delta > 0$ and from Schilder's theorem for any closed set

$$\begin{aligned}
\mathbb{P}[\xi_{\rho^m} \notin K_\delta] &= \mathbb{P}[\xi_{\rho^m} \in K_\delta^c] \\
&= \mathbb{P}\left[\frac{W_\cdot}{\sqrt{\log(\log \rho^m)}} \in K_\delta^c\right] \\
&\leq e^{\Lambda^*(\psi)\log(\log \rho^m)} \\
&\leq e^{\lambda \log(m \log \rho)} \\
&= \frac{\log \rho}{m^\lambda}
\end{aligned}$$

where $\mathbb{P}$ is the Wiener measure scaled by

$$\frac{1}{\sqrt{\log(\log \rho^m)}} \sim \frac{1}{\sqrt{\log m}}.$$

Therefore we obtain that

$$\sum_{m=1}^{\infty} \mathbb{P}[\xi_{\rho^m} \in K_\delta^c] < \infty$$

since

$$\sum_{m=1}^{\infty} \frac{\log \rho}{m^\lambda} < \infty \quad \text{for all} \quad \lambda > 1.$$

This requires just the Borel-Cantelli lemma. We have just showed that for almost $w \in C_0$ there exists $m_0 = m_0(w)$ such that:

$$\sum_{m=1}^{\infty} \mathbb{P}[\xi_{\rho^m}(w) \in K_\delta^c] < \infty \tag{6.2}$$

then

$$\mathbb{P}[\limsup_{m\to\infty} \xi_{\rho^m}(w) \in K_\delta^c] = 0.$$

The second step is to show that the price for sampling is not too much. To this end, we need to control the behavior of (ξ_n) for n between two numbers of the form ρ^m. More precisely, we need to prove the following lemma.

Lemma 6.1. *Let $\delta > 0$ be fixed. There exists $\rho > 1$ such that for almost every w there exists $m_1 = m_1(w)$ such that $\forall m \geq m_1$*

$$Y_n = \sup_{\rho^m < n < \rho^{m+1}} \|\xi_n(\cdot) - \xi_{\rho^m}(\cdot)\|_\infty \leq \delta.$$

Remark 6.3. *This lemma concludes the proof of the first part of the proof since obviously for $m \geq \min\{m_0, m_1\}$ one has that*

$$d(\xi_n(\cdot), K) \leq d(\xi_{\rho^m}, K) + \delta \leq 2\delta$$

where $\rho^m < n < \rho^{m+1}$.

Proof. Let us define for simplicity $\phi(t) = \sqrt{t \log\log t}$. Then one has that

$$\xi_n(t) - \xi_{\rho^m}(t) = \frac{W_{nt}}{\phi(n)} - \frac{W_{\rho^m t}}{\phi(\rho^m)}$$

$$= \frac{W_{nt}}{\phi(n)} + \frac{W_{\rho^m t}}{\phi(n)} - \frac{W_{\rho^m t}}{\phi(n)} - \frac{W_{\rho^m t}}{\phi(\rho^m)}$$

$$= \frac{W_{nt} - W_{\rho^m t}}{\phi(n)} + \frac{W_{\rho^m t}}{\phi(\rho^m)} \left(\frac{\phi(\rho^m)}{\phi(n)} - 1 \right) := S_t^{n,m} + R_t^{n,m} \tag{6.3}$$

Let us start we the second term of the right side.

$$\|R_t^{n,m}\|_\infty \leq \|\xi_{\rho^m}\|_\infty \left| 1 - \frac{\phi(\rho^m)}{\phi(n)} \right|.$$

From the first part we know that $\xi_{\rho^m} \in K$ so that there exists $M \geq 0$ such that $\|\xi_{\rho^m}\|_\infty \leq M$ since K is compact and thus bounded. Also an elemantary computation gives us

$$\lim_{m\to\infty} \left(1 - \frac{\phi(n)}{\phi\rho^m} \right) = 1 - \frac{1}{\sqrt{\rho}}.$$

Thus for m large we have that $\|R^{n,m}\|_\infty \leq M\left(1 - \frac{1}{\sqrt{\rho}}\right) \leq \delta$.As for the term $S^{n,m}$ we have that

$$|S_t^{n,m}| = \left| \frac{W_{nt} - W_{\rho^m t}}{\phi(n)} \right| \leq \left| \frac{W_{nt} - W_{\rho^m t}}{\phi(\rho^m)} \right| = \frac{\phi(\rho^{m+1})}{\phi(\rho^m)} \left| \xi^{\rho^{m+1}}_{t\frac{n}{\rho^{m+1}}} - \xi^{\rho^{m+1}}_{\frac{t}{\rho}} \right|.$$

Since $\rho^m < n < \rho^{m+1}$ then the quotient between the two time instants $t\frac{n}{\rho^{m+1}}$ and $\frac{t}{\rho}$ is comprised between $\frac{1}{\rho}$ and 1. Thus

$$\|S^{n,m}\|_\infty = \sup_{0\leq t\leq 1} |S_t^{n,m}| \leq \frac{\phi(\rho^{m+1})}{\phi(\rho^m)} \sup_{0\leq t\leq 1, \frac{t}{\rho}\leq s\leq t} \left| \xi_t^{\rho^{m+1}} - \xi_s^{\rho^{m+1}} \right|.$$

We already know that

$$\lim_{m\to\infty} \frac{\phi(\rho^{m+1})}{\phi(\rho^m)} = \sqrt{\rho}.$$

Now we will show that $\mathbb{P}[\|S^{n,m}\| + \overline{\infty > \delta] = 0}$, using Borel-Cantelli lemma. So,

$$\mathbb{P}\left[\|S^{n,m}\|_\infty > \delta] \leq \mathbb{P}[\sup_{0\leq t\leq 1, \frac{t}{\rho}\leq s\leq t} \left| \xi_t^{\rho^{m+1}} - \xi_s^{\rho^{m+1}} \right| > \delta \right] = \mathbb{P}[\xi_{\rho^{m+1}} \in A]$$

where A is the set of paths

$$A = \left\{ \psi \in C_0, \sup_{0\leq t\leq 1, \frac{t}{\rho}\leq s\leq t} \left| \xi_t^{\rho^{m+1}} - \xi_s^{\rho^{m+1}} \right| \leq \frac{\delta}{2} \right\}.$$

So by Schilder's theorem in order to estimate the behavior of $\mathbb{P}[\xi_{\rho^{m+1}} \in A]$ we need to compute $\Lambda^*(\psi)$ over A. Since

$$\mathbb{P}[\xi_{\rho^{m+1}} \in A] \leq e^{-\phi\left(\rho^{m+1}\right) \inf_{\psi \in A} \Lambda^*(\psi)}.$$

Recall that $\lambda^*(\psi) = \frac{1}{2}\int_0^1 |\dot{\psi}_s|^2 ds$ and $0 \leq t \leq 1$, $\frac{t}{\rho} \leq s \leq t$. Therefore,

$$\frac{\delta}{2} \leq |\psi_t - \psi_s| = \left|\int_0^t \dot{\psi}_u du\right| \leq \sqrt{|t-s|}\left(\int_s^t |\dot{\psi}_u|^2 du\right)^{\frac{1}{2}}$$

$$\leq 2\sqrt{|t - \frac{t}{\rho}|\Lambda^*(\psi)} = 2\sqrt{t\left(1 - \frac{1}{\rho}\right)\Lambda^*(\psi)} \leq 2\sqrt{\frac{\rho-1}{\rho}\Lambda^*(\psi)}.$$

Then

$$\Lambda^*(\psi) \geq \frac{\delta\sqrt{\rho}}{4\sqrt{\rho-1}}$$

which means that for every $\delta > 0$ the infimum over A can be made as large as we want, provided that ρ is close to 1. In particular, if ρ is small enough then the infimum of Λ^* over A is larger than 1, so that the series

$$\sum_0^\infty e^{-\phi\left(\rho^{m+1}\right) \inf_{\psi \in A} \Lambda^*(\psi)} < \infty.$$

By Borel-Cantelli lemma this implies that for every $\delta > 0$ there exists $m_0 = m_0(w)$ such that $|S^{n,m}(w)| > \delta$ for all $m \geq m_0$. This combined with 6.3 and 6.1 concludes the proof.

$\square$

One of the consequences of the result proved in the first part is that for any continuous functional $F : \Omega \to \mathbb{R}$ almost surely,

$$\limsup_{m\to\infty} F(\xi_{\rho^m}(\cdot)) \leq \sup_{\psi \in K} F(\psi)$$

(ii) We need to prove that every $\psi \in K$ is a limit point of $\{\xi_n(\cdot)\}_{n\geq 3}$. More precisely , we want to show that for every $\psi \in K$ such that $\frac{1}{2}\int_0^1 |\dot{\psi}_s| ds < 1$ there exists $(n_k)_k$ such that $\|\xi_{n_k} - \psi\|_\infty \to \infty$ as $k \to \infty$:

$$\mathbb{P}[d(\xi_{n_k}, \psi) < \delta \quad \text{for infinite indices} \quad n] = 1.$$

It is sufficient to show that

$$\sum_{m=1}^\infty \mathbb{P}[d(\xi_{\rho^m} - \psi) < \delta] = \infty.$$

Thus if the events $\{d(\xi_{\rho^m} - \psi) < \delta\}$ were independent the result will be immediate by second lemma of Borel-Cantelli. Unfortunately, these events are not independent, so we need to construct independent sequences. To this end, let us define $a_m = \rho^m - \rho^{m-1}$, $0 < t < 1$

$$y_m(t) = \frac{1}{\sqrt{a_m \log\log a_m}}[W\left(\rho^{m-1}t + a_m t\right) - W\left(\rho^{m-1}\right)]$$

$W\left(\rho^{m-1}t + a_m t\right) - W\left(\rho^{m-1}\right)$: increments, which consisting, as m varies, of families of independent variables and $y_m(t)$ is a Brownian motion scaled by $\frac{1}{\sqrt{a_m \log\log a_m}}$.
Therefore

$$\mathbb{P}[d(y_m(t), \psi) < \delta] = \mathbb{P}[\|y_m(t) - \psi\| < \delta]$$

$$= \mathbb{P}[\sup |y_m(t) - \psi| < \delta]$$
$$= \mathbb{P}[y_m \in A]$$

such that $A = \{f \in C([0,1]) : |f - \psi| < \delta\}$ is open set, then by Schilder's theorem for any open set we obtain that

$$\mathbb{P}[y_m(\cdot) \in A] \le e^{-\Lambda^*(\psi_0)\log(\log a_m)}$$
$$\le e^{-\lambda \log(\log a_m)} = \frac{1}{(\log a_m)^\lambda}.$$

Then

$$\sum_{m=1}^{\infty} \mathbb{P}[y_m(\cdot) \in A] = \infty$$

since

$$\sum_{m=1}^{\infty} \frac{1}{(\log a_m)} = \infty \quad \text{when} \quad \lambda < 1.$$

Applying Borel-Cantelli lemma $y_n(\cdot)$ returns infinitely often to the δ neighborhood of ψ. The last part of the proof is to show that almost surely

$$\limsup_{m\to\infty} \|\xi_{\rho^m}(\cdot) - y_m(\cdot)\| \le \epsilon(\rho).$$

$$|\xi_{\rho^m}(t) - y_m(t)| = \left| \frac{W(\rho^m t)}{\sqrt{\rho^m \log\log \rho^m}} - \frac{\left[W\left(\rho^{m-1}t - a_m t\right) - W\left(\rho^{m-1}\right)\right]}{\sqrt{a_m \log\log a_m}} \right|$$

$$= \left| \frac{W(\rho^m t)}{\sqrt{\rho^m \log\log \rho^m}} + \frac{W(\rho^m t)}{\sqrt{a_m \log\log a_m}} - \frac{W(\rho^m t)}{\sqrt{a_m \log\log a_m}} - \frac{\left[W\left(\rho^{m-1} + a_m t\right) - \theta\left(\rho^{m-1}\right)\right]}{\sqrt{a_m \log\log a_m}} \right|$$

$$\le \left| \frac{W(\rho^m t)}{\sqrt{\rho^m \log\log \rho^m}} - \frac{W(\rho^m t)}{\sqrt{a_m \log\log a_m}} \right| + \frac{1}{\sqrt{a_m \log\log a_m}} \left|W(\rho^m t) - W\left(\rho^{m-1}t + a_m t\right) + W\left(\rho^{m-1}\right)\right|$$

$$\le |W(\rho^m t)| \left| \frac{1}{\sqrt{\rho^m \log\log \rho^m}} - \frac{1}{\sqrt{a_m} \log\log a_m} \right| + \frac{1}{\sqrt{a_m \log\log a_m}} |W(\rho^m t) - W\left(\rho^{m-1} + a_m t\right)|$$
$$+ \frac{1}{\sqrt{a_m \log\log a_m}} |W(\rho^m)|$$

$$\le \left| \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} - 1 \right| |\xi_{\rho^m}(t)| + \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} \left| \xi_{\rho^m}(t) - \frac{W\left(\rho^{m-1} + a_m t\right)}{\sqrt{\rho^m \log\log \rho^m}} \right|$$
$$+ \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} \left| \frac{W(\rho^m)}{\sqrt{\rho^m \log\log \rho^m}} \right|$$

$$\le \left| \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} - 1 \right| |\xi_{\rho^m}(t)| + \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} \left| \xi_{\rho^m}(t) - \xi_{\rho^m}\left(\frac{1}{\rho} + \left(1 - \frac{1}{\rho}\right)t\right) \right|$$
$$+ \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} \left| W_{\rho^m}\left(\frac{1}{\rho}\right) \right|.$$

Taking the supremum over $0 \le t \le 1$

$$\|\xi_{\rho^m}(t) - y_m(t)\| \le \left| \sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}} - 1 \right| \|\xi_{\rho_m}(t)\|$$

$$+\sqrt{\frac{\rho^m \log\log \rho^m}{a_m \log\log a_m}}\left[\sup_{|t-s|\leq\frac{1}{\rho}} |\xi_{\rho^m}(t)-\xi_{\rho^m}(s)|+\xi_{\rho^m}\left(\frac{1}{\rho}\right)\right]$$

$$\leq \left|\sqrt{\frac{\rho}{\rho-1}}-1\right| \|\xi_{\rho^m}(t)\| + \sqrt{\frac{\rho}{\rho-1}}[\sup_{|t-s|\leq\frac{1}{\rho}} |\xi_{\rho^m}(t)-\xi_{\rho^m}(s)|+\xi_{\rho^m}\left(\frac{1}{\rho}\right)]$$

$$\leq \left|\sqrt{\frac{\rho}{\rho-1}}-1\right| \sup_{\psi\in K}\|\psi\| + \sqrt{\frac{\rho}{\rho-1}}\left[\sup_{\psi\in K}\psi\left(\frac{1}{\rho}\right)+\sup_{|t-s|\leq\frac{1}{\rho}}|\psi(t)-\psi(s)|\right]$$

$$=\epsilon(\rho).$$

Then $\epsilon(\rho)\to 0$ as $\rho\to\infty$. This completes the proof. □

6.2 The Freidlin-Wentzel theory

Behavior of diffusions with a small noise parameter:
The results of Section 5 are extended here to the case of strong solutions of stochastic differential equations. Let W_t be a d-dimensional Brownian motion and let $\mathcal{N}_t$ be the σ-algebra generated by the random variables W_t for $s \leq t$. We consider the stochastic differential equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t \quad X_0 = x \tag{6.4}$$

in R^d. Here $b(X_t) = \left(b^1(x), \dots, b^d(x)\right)$ is a vector in $\mathbb{R}^d$ and $\sigma(x) = \sigma^i_j(x)$ is an invertible matrix having d columns and d rows.
We assume that the coefficients $b(x)$, $\sigma(x)$ satisfy the following conditions:

(i) $\sum_i |b^i(x) - b^j(y)| + \sum_{i,j} |\sigma^j_i(x) - \sigma^j_i(y)| \leq K|x-y|$,

(ii) $\sum_i |b^i(x)| + \sum_{i,j} |\sigma^j_i(x)| \leq K(1+|x|)$,

under these conditions it can be proved that the above SDE 6.4 has a unique solution which $X^x_t(\omega)$, $t\geq 0$ which is continuous with probability one, the random variable $X^x_t(\omega)$ is measurable with respect to the σ-algebra $\mathcal{N}_t$ for every $t \geq 0$, and $\int_a^b \mathbb{E}|X^x_t|^2 dt < \infty$ for any $0 < a < b$.
The corresponding generator to 6.4 is

$$L = \frac{1}{2}\sum_{i,j} a^{ij}(x)\frac{\partial^2}{\partial x_1 \partial x_j} + \sum_i b^i(x)\frac{\partial}{\partial x_i},$$

where $a^{ij}(x) = \sigma(x)\,\sigma^T(x)$ is a symmetric matrix valued.
Moreover, we assume that the diffusions coefficients are uniformly elliptic.
Now we are interested in the situation where L depends on a small parameter ϵ that is small.
In particular, the diffusion X^ϵ is a random perturbation of the deterministic system:

$$\frac{dX(t)}{dt} = b(X(t)), \quad X(0) = x. \tag{6.5}$$

We consider the following perturbation of the above deterministic system.

$$dX^\epsilon_t = b_\epsilon(X^\epsilon_t)\,dt + \sqrt{\epsilon}\sigma_\epsilon(X^\epsilon_t)\,dW_t \tag{6.6}$$

if $\epsilon \to 0$ then the above SDE transformed to a deterministic function.
Let $\tilde{\mathbb{Q}}_\epsilon$ be the measure induced by $X_\epsilon(\cdot)$ on the space of $\mathbb{R}^d$-valued continuous functions on some arbitrary but finite interval.
As $\epsilon \to 0$, $\tilde{\mathbb{Q}}_\epsilon$ converges weakly to the degenerate measure concentrated at a single trajectory $x_0(t)$ which solves uniquely the O.D.E 6.5: $\dot{X}_0(t) = b(X_0(t))$. Then we have that $\tilde{\mathbb{Q}}_\epsilon \Rightarrow \delta_{x_0}$ as $\epsilon \to 0$. This means that

$$\lim_{\epsilon} \mathbb{P}\left\{\sup_{0\le s\le 1} |X_s^\epsilon - X_s^0| > \delta\right\} = 0$$

Once again, we are dealing with a family for which it is reasonable to ask if it satisfies large deviation principle. First we consider the relatively simple situation. Let $\{X_t^\epsilon\}_{\epsilon>0}$ be the diffusion process that is the unique strong solution of the SDE

$$dX_t^\epsilon = b(X_t^\epsilon)\,dt + \sqrt{\epsilon}dW_t \tag{6.7}$$

and the coefficient b satisfies the following conditions

(i) $|b(x) - b(y)| \le K|x-y|$

(ii) $|b(x)| \le C$

for K, C are constants. Let $\tilde{\mathbb{Q}}_\epsilon$ denote the measure induced by the strong solution $\{X_t^\epsilon\}$ of 6.7 on $C_0[0,1]$ then $\tilde{\mathbb{Q}}_\epsilon = \mathbb{Q}_\epsilon \circ F^{-1}$ where $\mathbb{Q}_\epsilon$ is the measure induced by $\{\sqrt{\epsilon}W\}$ and the deterministic map $F : C_0[0,1] \to C_0[0,1]$ is defined by $f = F(g)$ where f is the unique continuous solution of

$$f(t) = \int_0^t b(f(s))\,ds + g(t), \quad t \in [0,1].$$

Then the LDP associated with $\{X_t^\epsilon\}$ is a direct application of the contraction principle with respect to the map F and Schilder's theorem 5.1.

Theorem 6.2. *$\{X_t^\epsilon\}_\epsilon$ satisfies the LDP in $C_0[0,1]$ with the good rate function*

$$I(f) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 |\dot{f}(t) - b(f(t))|^2 dt & f \in \mathcal{H} \\ \infty & \textit{otherwise} \end{cases}.$$

Proof. We shall prove that

$$F : \sqrt{\epsilon}W_t \to X_t^\epsilon \quad \text{is continuous.}$$

Indeed

$$\begin{aligned} F(g) = f(t) &= x + \int_0^t b(f(s))\,ds + \sqrt{\epsilon}\int_0^t dW_s \\ &= x + \int_0^t b(f(s))\,ds + \int_0^t d\left(\sqrt{\epsilon}W_s\right) \\ &= x + \int_0^t b(f(s))\,ds + g(t). \end{aligned}$$

We assume that $x = 0$, then

$$f(t) = F(g(t)) = \int_0^t b(f(s))\,ds + g(t) \qquad t \in [0,1].$$

Therefore, for every $f_1, f_2 \in C_0\left([0,1];\mathbb{R}^d\right)$ we define

$$f_1(t) = F(g_1(t)), \qquad f_2(t) = F(g_2(t)).$$

Hence, we have that

$$f_1(t) - f_2(t) = \int_0^t \left(b(f_1(s)) - b(f_2(s))\right) ds + g_1(t) - g_2(t).$$

Now consider for $g_1, g_2 \in C_0[0,1]$ and Lipshitz continuity of b, we have

$$|f_1(t) - f_2(t)| \le \int_0^t |b(f_1(s)) - b(f_2(s))|ds + |g_1(t) - g_2(t)| \le K\int_0^t |f_1(s) - f_2(s)|ds + |g_1(t) - g_2(t)|$$

$$\le K\int_0^t |f_1(s) - f_2(s)|ds + \delta \le \delta e^{Kt}$$

The last inequality is due to Gronwall inequality. So

$$\|f_1 - f_2\| \le \delta e^K$$

and the continuity of F is now established.
Now, we combine Schilder's theorem and contraction principle to obtain our result. Firstly we obtain a lower bound for every G open set in $C_0[0,1]$.

$$\liminf_{\epsilon\to 0} \log \epsilon \log \tilde{\mathbb{Q}}_\epsilon(G) = \liminf_{\epsilon\to 0} \epsilon \log \mathbb{Q}_\epsilon \circ F^{-1}(G)$$

$$\ge -\inf_{f\in G} I(f)$$

where

$$I(f) = \inf_{g\in\mathcal{H}: f = F(g)} \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 ds.$$

We observe that F is an injection and $g \in \mathcal{H}$ which implies that $f = F(g)$ is differentiable almost everywhere with

$$\dot{f}(t) = b(f(t)) + \dot{g}(t), \qquad f(0) = 0.$$

$$|\dot{f}(t) - \dot{f}(0)| \le K\int_0^t |\dot{f}(s)|ds + \dot{g}(t) \le \delta e^{Kt}$$

since $g \in \mathcal{H}$ we have that $f \in \mathcal{H}$. Similarly, if F is a closed set of $C_0\left([0,1];\mathbb{R}^d\right)$ we obtain an upper bound

$$\limsup_{\epsilon\to 0} \epsilon \log \tilde{\mathbb{Q}}_\epsilon(F) \le -\inf_{f\in F} I(f).$$

□

In other words, Schilder's theorem leads directly to a large deviation result for $\left\{\tilde{\mathbb{Q}}_\epsilon : \epsilon > 0\right\}$. The precceding example of Freidlin-Wentzel's theory is as sipmle as because F is continuous and its inverse is easy to compute. In general, the maps involved are not only more complicated but are not even continuous. To be more precise, the following case describes how to solve the problem of a function that is not continuous but measurable.
Let $\{X_t^\epsilon\}$ be the difussion process that is the unique solution of the SDE

$$dX_t^\epsilon = b(X_t^\epsilon)\,dt + \sqrt{\epsilon}\sigma(X_t^\epsilon)\,dW_t, \quad 0 \le t \le 1, \quad X_0^\epsilon = x. \tag{6.9}$$

Theorem 6.3. *If b, σ are uniformly bounded and Lipschitz continuous functions then $\{X_t^\epsilon\}$ the unique strong solution of 6.9 satisfies LDP on $C[0,1]$ with good rate function*

$$I_x(f) = \begin{cases} \inf\limits_{\left\{g\in\mathcal{H}: f(t)=x+\int_0^t b(f(s))ds+\int_0^t \sigma(f(s))\dot{g}(s)ds\right\}} \dfrac{1}{2}\displaystyle\int_0^1 |\dot{g}(s)|^2 ds & f\in\mathcal{H} \\ \infty & \textit{otherwise} \end{cases}$$

Remark 6.4. *It suffices to proof the theorem for $x_0^\epsilon = x = 0$ since x may always be moved to the origin by a translation of coordinates. Then the measure $\tilde{\mathbb{Q}}_\epsilon$ of X^ϵ is supported on $C_0([0,1])$.*

Proof. The proof here is based on approximating the process X_t^ϵ by $X_t^{n,\epsilon}$ in the following way. For each $s > 0$ we construct an approximate solution of 6.9 as follows

$$X_t^{n,\epsilon} = x + \int_0^t b\left(X^{\epsilon,n}\left(\pi_n(s)\right)\right) ds + \sqrt{\epsilon}\int_0^t \sigma\left(X^{\epsilon,n}\left(\pi_n(s)\right)\right) dW_s \tag{6.10}$$

where $\pi_n(s) = \frac{[ns]}{n}$ is Euler approximations and $X_0^{\epsilon,n} = 0$, $t\in[0,1]$. In fact, we amounts to freezing the coefficients over the time interval $[\frac{k}{n}, \frac{k+1}{m}]$ and then updating them every $\frac{1}{n}$ unit of time. The advantage in using $X_{n,\epsilon}(t)$ is that in fact

$$F^n : \sqrt{\epsilon} W_t \to X_t^{\epsilon,n} \quad \text{continuous.}$$

More precisely, let the map F^n defined by $F^n(g) = h$ where

$$h(t) = h\left(\frac{k}{n}\right) + b\left(h\left(\frac{k}{n}\right)\right)\left(t - \frac{k}{n}\right) + \sigma\left(h\left(\frac{k}{n}\right)\right)\left(g(t) - g\left(\frac{k}{n}\right)\right)$$

and $t\in[\frac{k}{n}, \frac{k+1}{n}]$, $k = 0, 1, \ldots, n-1$, $h(0) = 0$. We observe that $F^n : C_0[0,1] \to C_0[0,1]$ and $X^{\epsilon,n} = F^n(\sqrt{\epsilon}W)$. By assumptions of $b(\cdot), \sigma(\cdot)$ and that $g_1, g_2 \in C_0[0,1]$ we have that F^n is continuous. Indeed,

$$\begin{aligned} |F^n(g_1) - F^n(g_2)| = |h_1(t) - h_2(t)| &\le b\left(h_1\left(\frac{k}{n}\right)\right) - b\left(h_2\left(\frac{k}{n}\right)\right)\left|t - \frac{k}{n}\right| + \\ &+\sigma\left(h_1\left(\frac{k}{n}\right)\right) - \sigma\left(h_2\left(\frac{k}{n}\right)\right)||g_1(t) - g_2(t)| \\ &\le C\left[\left|h_1\left(\frac{k}{n}\right) - h_2\left(\frac{k}{n}\right)\right| + |g_1(t) - g_2(t)|\right] \end{aligned}$$

then

$$\sup_{t\in[\frac{k}{n}, \frac{k+1}{n}]} |h_1(t) - h_2(t)| \le C\left[\left|h_1\left(\frac{k}{n}\right) - h_2\left(\frac{k}{n}\right)\right| + \|g_1 - g_2\|\right]$$

So the continuity of F^n with respect to the supremum norm is established by iterating this bound over $k = 0, 1, \ldots, n-1$.
Now let F defined by $f = F(g)$ as the unique solution of the integral

$$f(t) = \int_0^t b(f(s))\, ds + \int_0^t \sigma(f(s))\,\dot{g}(s)\, ds, \qquad 0 \le t \le 1.$$

We shall show that for every $c < \infty$

$$\lim_{n\to\infty} \sup_{\{g:\|g\|_{\mathcal{H}} < c\}} \|F^n(g) - F(g)\| = 0 \tag{6.11}$$

To this end, we fix $c < \infty$, $g \in \mathcal{H}$ and $\|g\|_{\mathcal{H}} \leq c$ then

$$h(t) - h\left(\frac{[nt]}{n}\right) = \int_0^t b\left(h\left(\frac{[ns]}{n}\right)\right) ds + \int_0^t \sigma\left(h\left(\frac{[ns]}{n}\right)\right) \dot{g}(s)\, ds$$

by boundedness of b, σ and the Cauchy-Schwartz inequality in the second part of the right side we have that

$$|h(t) - h\left(\frac{[nt]}{n}\right)| \leq \delta_n t + \delta_n c$$

and

$$\sup_{0\leq t\leq 1}\left|h(t) - h\left(\frac{[nt]}{n}\right)\right| \leq c(\delta_n + 1)$$

where δ_n is independent of g. Therefore

$$|f(t) - h(t)| = \left|\int_0^t b(f(s))\, ds + \int_0^t \sigma(f(s))\,\dot{g}(s)\, ds - \left(\int_0^t b\left(h\left(\frac{[ns]}{n}\right)\right) ds + \int_0^t \sigma\left(h\left(\frac{[ns]}{n}\right)\right)\dot{g}(s)\, ds\right)\right|$$

$$= \left|\int_0^t b(f(s)) - b\left(h\left(\frac{[ns]}{n}\right)\right) ds\right| + \left|\int_0^t \left(\sigma\left(f(s) - \sigma\left(h\left(\frac{[ns]}{n}\right)\right)\right)\right)\dot{g}(s)\, ds\right|$$

$$\leq \int_0^t \left|b(f(s)) - b\left(h\left(\frac{[ns]}{n}\right)\right)\right| ds + \int_0^t \left|\sigma(f(s)) - \sigma\left(h\left(\frac{[ns]}{n}\right)\right)\right|\dot{g}(s)\, ds$$

$$\leq C\left(\int_0^t \left|f(s) - h\left(\frac{[ns]}{n}\right)\right|^2 ds\right)^{\frac{1}{2}} + C\left(\int_0^t \left(\left|f(s) - h\left(\frac{[ns]}{n}\right)\right|\dot{g}(s)\right)^2 ds\right)^{\frac{1}{2}}$$

$$\leq C\int_0^t \left|f(s) - h\left(\frac{[ns]}{n}\right)\right| ds\,(1+c)\,.$$

Thus

$$|f(t) - h(t)|^2 \leq C\int_0^t \left|f(s) - h\left(\frac{[ns]}{n}\right)\right|^2 ds + C\delta_n^2 \leq C\delta_n^2 e^{Ct}$$

the last inequality is due to Cronwall inequality. So,

$$\sup_{\{g:\|g\|_{\mathcal{H}}<c\}} \|F^n(g) - F(g)\| \leq \sqrt{C}\delta_n e^C$$

and the 6.11 is established. The proof of theorem is completed by proving that for any $\delta > 0$, $X^{\epsilon,n}$ are exponentially good approximations of X^ϵ.

$$\lim_{n\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}[\sup_{0\leq t\leq 1}|X^{n,\epsilon}(t) - X^\epsilon(t)\| \geq \delta] = -\infty \tag{6.12}$$

These estimates are called superexponential estimates. □

We prove the following lemma in which the proof of 6.12 is based.

Lemma 6.2. *Let $z(t)$ be a process satisfying*

$$dz(t) = b_t dt + \sqrt{\epsilon}\sigma_t dW_t$$

where z_0 is deterministic. Let $\tau_0 \in [0,1]$ be a stopping time with respect to the filtration of $\{W_t\}$, $\tau_0 \in [0,1]$. Suppose that the coefficients of the diffusion are uniformly bounded

$$|\sigma_t| \leq M\left(\rho^2 + |z_t|^2\right)^{\frac{1}{2}}$$

$$|b_t| \leq B\left(\rho^2+|z_t|^2\right)^{\frac{1}{2}}$$

where B, ρ, M are constants and $t \in [0,\tau_1]$. Then for any $\delta > 0$ and $\epsilon \leq 1$

$$\mathbb{P}[\sup_{t\in[0,\tau_1]} |z_t| > \delta] \leq \frac{\rho^2+|z_0|^2}{\rho^2+\delta^2}e^K$$

where $K = 2B + M^2(2+d)$.

Proof. Let $u_t = (z_t)$ and $\phi(y) = \left(\rho^2+|y|^2\right)^{\frac{1}{\epsilon}}$. Applying Ito's formula in u_t we have that

$$du_t = \nabla\phi(z_t)\,dz_t + \frac{\epsilon}{2}Tr[\sigma_t\sigma_t' D^2(z_t)]dt$$

$$= \left(\nabla\phi(z_t)\,b_t + \frac{\epsilon}{2}Tr[\sigma_t\sigma_t' D^2\phi(z_t)]\right)dt + \sqrt{\epsilon}\sigma_t\nabla\phi(z_t)\,dW_t$$

$$= g(t)\,dt + h(t)\,dW_t.$$

Note that

$$\nabla\phi(y) = \frac{1}{\epsilon}\frac{2\phi(y)}{\rho^2+|y|^2}y.$$

$$|\nabla\phi(y)\,b_t| \leq \left|\frac{1}{\epsilon}\frac{2\phi(z_t)}{\rho^2+|z_t^2|}|z_t|\right| B\left(\rho^2+|z_t|^2\right)^{1/2}$$

$$\leq \frac{2B}{\epsilon}u_t.$$

$$|\frac{\epsilon}{2}Tr[\sigma_t\sigma_t' D^2(z_t)]| \leq \frac{\epsilon}{2}M^2\left(\rho^2+|z_t|^2\right)\left(\frac{2}{\epsilon}\left(\frac{\phi(z_t)}{\rho^2+|z_t|^2}\right)' y + \frac{2}{\epsilon}\frac{\phi(z_t)}{\rho^2+|z_t|^2}\right)$$

$$\leq \frac{M}{\epsilon}(2+d)\,u_t.$$

Then

$$|g(t)| \leq \frac{2B+M^2(2+d)}{\epsilon}u_t = \frac{K}{\epsilon}u_t, \qquad t\in[0,\tau_1]$$

where $K = 2b + M^2(d+2)$ constant and

$$|h(t)| \leq \frac{2M}{\sqrt{\epsilon}}u_t.$$

Now, fix $\delta > 0$ we define the stopping time $\tau_2 := \inf\{t : |z_t| \geq \delta\} \wedge \tau_1$. Then g_t, h_t are uniformly bounded on $[0,\tau_2]$, $h_t dW_t$ is a martingale and $u_t - \int_0^t g_s ds$ is a continuous martingale on $[0,\tau_2]$. So we can apply Doob's theorem

$$\mathbb{E}[u_{t\wedge\tau_2} - \int_0^t g_s ds] = \mathbb{E}[u_0]$$

$$\mathbb{E}[u_{t\wedge\tau_2}] = u_0 + \mathbb{E}\left[\int_0^{t\wedge\tau_2} g_s ds\right]$$

$$_0 + \mathbb{E}\left[\int_0^{t\wedge\tau_2}\frac{K}{\epsilon}u_s ds\right]$$

$$= u_0 + \frac{K}{\epsilon}\mathbb{E}\left[\int_0^{t\wedge\tau_2} u_{s\wedge\tau_2} ds\right]$$

$$\leq u_0 + \frac{K}{\epsilon}\mathbb{E}\left[\int_0^t u_{s\wedge\tau_2} ds\right]$$
$$= u_0 + \frac{K}{\epsilon}\int_0^t \mathbb{E}[u_{s\wedge\tau_2}]ds$$
$$\leq u_0 e^{\frac{K}{\epsilon}t}$$
$$\mathbb{E}[u_{\tau_2}] = \mathbb{E}[u_{1\wedge\tau_2}] \leq u_0 e^{\frac{K}{\epsilon}}.$$

Therefore, since ϕ is monotone increasing in $|y|$

$$\mathbb{P}[|z_{\tau_2}| \leq \delta] = \mathbb{P}[\phi(z_{\tau_2}) \leq \phi(\delta)] \leq \frac{\mathbb{E}[\phi(z_{\tau_2})]}{\phi(\delta)} = \frac{\mathbb{E}[u_{\tau_2}]}{\phi(\delta)}$$

the last inequality is due to Chebyshev inequality. Finally,

$$\mathbb{P}[|z_{\tau_2}| \leq \delta] = \mathbb{P}[\sup_{t\in[0,\tau_1]} |z_t| \geq \delta] \leq \frac{u_0}{\phi(\delta)}e^{\frac{K}{\epsilon}} = \left(\frac{\rho^2+|z_0|^2}{\rho^2+\delta^2}\right)^{\frac{1}{\epsilon}} e^{\frac{K}{\epsilon}}.$$

□

Now we proceed in the proof of 6.12.

Lemma 6.3. *For any $\delta > 0$, the solutions $X_{n,\epsilon}(\cdot)$ and $X_\epsilon(\cdot)$ of*

$$X_t^{\epsilon,n} = x + \int_0^t b(X^{\epsilon,n}(\pi_n(s)))\,ds + \sqrt{\epsilon}\int_0^t \sigma(X^{\epsilon,n}(\pi_n(s)))\,dW_s$$

and

$$X_t^{\epsilon} = x + \int_0^t b(X_s^{\epsilon})\,ds + \sqrt{\epsilon}\int_0^t \sigma(X_s^{\epsilon})\,dW_s$$

respectively, satisfy:

$$\lim_{n\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}[\sup_{0\leq t\leq 1}\|X^{\epsilon,n}(t) - X^{\epsilon}(t)\| \geq \delta] = -\infty.$$

Proof. We define $z_t = X_t^{\epsilon,n} - X_t^{\epsilon}$ and for any $\rho > 0$ the stopping time $\tau_1 = \inf\left\{t : |X_t^{\epsilon,n} - X_{\frac{[nt]}{n}}^{\epsilon,n}| > \rho\right\} \wedge 1$. For the process z_t we have that

$$z_t = X_t^{\epsilon,n} - X_{\frac{[nt]}{n}}^{\epsilon,n}$$
$$= \int_{\frac{k}{n}}^{\frac{k+1}{n}} \left(b(X_t^{\epsilon,n}) - b\left(X_{\frac{[nt]}{n}}^{\epsilon,n}\right)\right)dt + \sqrt{\epsilon}\int_{\frac{k}{n}}^{\frac{k+1}{n}}\left(\sigma(X_t^{\epsilon,n}) - \sigma\left(X_{\frac{[nt]}{n}}^{\epsilon,n}\right)\right)dW_t$$

where $z_0 = 0$. Then z_t satisfies the conditions of the lemma 6.2 and it follows that

$$\mathbb{P}[\sup_{0\leq t\leq\tau_1}|X_t^{\epsilon,n} - X_{\frac{[nt]}{n}}^{\epsilon,n}| \geq \delta] \leq e^{\frac{K}{\epsilon}}\left(\frac{\phi(z_0)}{\phi(\delta)}\right)^{\frac{1}{\epsilon}}$$

with K constant and

$$\limsup_{\rho\to 0}\sup_{n\geq 1}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}[\sup_{0\leq t\leq\tau_1}|X_t^{\epsilon,n} - X_{\frac{[nt]}{n}}^{\epsilon,n}| \geq \delta] = -\infty.$$

We want to show that

$$\lim_{n\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}[\sup_{0\le t\le 1}|X_t^{\epsilon,n}-X_t^{\epsilon}|\ge\delta]=-\infty.$$

Now since

$$\{\|X^{,n}-X^{\epsilon}\|\ge\delta\}\subseteq\{\tau_1<1\}\cup\left\{\sup_{0\le t\le\tau_1}|X_t^{\epsilon,n}-X_t^{\epsilon}|\ge\delta\right\}$$

the proof is completed as soon as we show that

$$\lim_{n\to\infty}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}[\sup_{0\le t\le 1}|X_t^{\epsilon,n}-X^{\epsilon}_{\frac{[nt]}{n}}|\ge\delta]=-\infty.$$

By boundedness of $|b(\cdot)|,|\sigma(\cdot)|$ we have

$$|X_t^{\epsilon,n}-X^{\epsilon,n}_{\frac{[nt]}{n}}|\le C\left(\frac{1}{n}+\sqrt{\epsilon}\max_{k=0,\dots,n-1}\sup_{0\le s\le\frac{1}{n}}|W_{s+\frac{k}{n}}-W_{\frac{k}{n}}|\right)$$

Hence,

$$\begin{aligned}\mathbb{P}[\sup_{0\le t\le 1}|X_t^{\epsilon,n}-X^{\epsilon,n}_{\frac{[nt]}{n}}|\le\rho]&\le m\mathbb{P}[\left(\frac{C}{n}+C\sqrt{\epsilon}\sup_{0\le s\le\frac{1}{n}}|W_{s+\frac{K}{n}}-W_{\frac{K}{n}}|\right)\le\rho]\\&=n\mathbb{P}[\sup_{0\le s\le\frac{1}{n}}|W_{s+\frac{K}{n}}-W_{\frac{K}{n}}|\le\frac{\rho-\frac{C}{n}}{C\sqrt{\epsilon}}]\\&=n\mathbb{P}[\sup_{0\le s\le\frac{1}{n}}|W_s|>\frac{\rho-\frac{C}{n}}{C\sqrt{\epsilon}}]\\&\le 2ne^{-\frac{1}{2}\frac{\left(\rho-\frac{C}{n}\right)^2 n}{C^2\epsilon}}\end{aligned}$$

the last inequality is due to 5.13 and the proof is completed. □

The following theorem strengthens Theorem 6.3 by allowing for ϵ dependent initial conditions.

Theorem 6.4. *Assume the conditions of Theorem 6.3. Let $\{X_t^{\epsilon,y}\}$ denote the solution of*

$$dX_t^{\epsilon}=b(X_t^{\epsilon})\,dt+\sqrt{\epsilon}\sigma(X_t^{\epsilon})\,dW_t$$

for the initial condition $X_0=y$. Then for any compact $K\subset\mathbb{R}^d$ and any closed $F\subset C([0,1])$

$$\limsup_{\epsilon\to 0}\epsilon\log\sup_{y\in K}\mathbb{P}[X^{\epsilon,y}\in F]\le-\inf_{f\in F}I_y(f). \tag{6.13}$$

$$\liminf_{\epsilon\to 0}\epsilon\log\inf_{y\in K}\mathbb{P}[X^{\epsilon,y}\in G]\ge-\sup_{y\in K}\inf_{f\in G}I_y(f). \tag{6.14}$$

Proof. Let $-I_K=-\inf_{f\in F}I_y(f)$. We fix $\delta>0$ and $I_K^{\delta}=(I_K-\delta)\wedge\frac{1}{\delta}$. Then for any $x\in K$ there is ϵ_x such that for every $\epsilon<\epsilon_x$

$$\epsilon\log\sup_{y\in B_{\epsilon_x}(x)}\mathbb{P}[X^{\epsilon,y}\in F]\le-I_K^{\delta}.$$

But $x_1,x_2,\dots,x_k\in K$ and K is compact so $K\subseteq\bigcup_{i=1}^k B_{\epsilon_{x_i}}(x_i)$. Then we choose $\epsilon\le\min_{i=1,2,,k}\epsilon_{x_i}$

$$\epsilon\log\sup_{y\in K}\mathbb{P}[X^{\epsilon,y}\in F]\le-I_K.$$

By first considering $\epsilon\to 0$ and the $\delta\to 0$ we obtain 6.13. The same arguments works in order to prove tha lower bound 6.14. □

Chapter 7

Exit problem

7.1 A solution through Freidlin-Wentzel theory

Remark 7.1. *"The basic idea is that among a bunch of very unlikely things the least unlikely thing is the most likely to occur first."*

Let us consider the problem of exit from a domain. We consider the system

$$dX_t^\epsilon = b\left(X_t^\epsilon\right) dt + \sqrt{\epsilon}\sigma\left(X_t^\epsilon\right) dW_t, \qquad X_t^\epsilon \in \mathbb{R}^d \quad X_0^\epsilon = x(7.1)$$

in the open, bounded $G \subseteq \mathbb{R}^d$ and let ∂G be its boundary, which we assume to be smooth for the sake of simplicity, $b\left(\cdot\right), \sigma\left(\cdot\right)$ are uniformly Lipschitz continuous functions of d-dimensions and W is d-dimensional Brownian motion. In this section we shall assume that $b\left(x\right) \cdot \eta\left(x\right) < 0$ for $x \in \partial G$, where $\eta\left(x\right)$ is the exterior normal to the boundary of G, so that the curves $x_x\left(t\right)$ cannot leave G for $x \in G$. The trajectories of the O.D.E. system 6.5 vanishes within G only at one point, the equilibrium point. More precisely, we assume that there is a globally stable equilibrium point 0 in G such that for every $x \in G$ the solution $x\left(t\right)$ of 6.5 lies in G for $t > 0$ and $x\left(t\right) \to 0$ as $t \to \infty$. As $\epsilon \to 0$ the trajectories of the diffusion process are close to the deterministic trajectories with a very high probability. In the limit the deterministic trajectory doesn't exit at all from the set G so the exit time and the exit place are not defined. We need a new formulation to calculate the limit of the hitting distribution on ∂G as $\epsilon \to 0$. If we define the stopping time

$$\tau^\epsilon = \inf\left\{t : X_t^\epsilon \notin G\right\}$$

then events like this $\left\{\tau^\epsilon < T\right\}$ are rare events, indeed

$$\mathbb{P}[\tau^\epsilon < T] \to 0 \quad \text{as } \epsilon \to 0$$

for any $T < \infty$, so we are dealing with a family that is reasonable to ask about large deviation principle. Motivated by Theorem 6.3, we define the cost function

$$V\left(y, z, t\right) \triangleq \inf_{f \in C([0,1]) : f_t = z} I_{y,t}\left(f\right) \tag{7.2}$$

$$= \inf_{g \in L_2([0,t]) : f_t = z, f_s = y + \int_0^s b(f_u)du + \int_0^s \sigma(f_u)\dot{g}du} \frac{1}{2} \int_0^t |\dot{g}_s|^2 ds.$$

where $I_{y,t}$ is the good rate function of 6.3 which controls the LDP associated with 7.1. Heuristicaly $V\left(y, z, t\right)$ is the cost of forcing the system 7.1to be at the point z at time t when starting at y. We define

$$V\left(y, z\right) \triangleq \inf_{t > 0} V\left(y, z, t\right)$$

The function $V(0,z)$ is called the quasi-potential. The picture that emerges is that a typical path will go quickly near the equilibrium point, wander around it for exponentially long time making periodic futile short lived attempts to get out which are determined by $\inf_{z\in\partial G} V(0,z)$. Finally a successful excursion takes place. The rationale here is that any excursion off the stable point $x=0$ has an overwhelmingly high probability of being pulled back there. What is matter is to find the path for a direct, fast exit due to a rare segment in the Brownian motion's path.
The following assumptions prevail throghout this section:
A-1 The unique equilibrium point in G of the d-dimensional ordinary differential equation

$$\dot{f}_t = b(f_t) \tag{7.3}$$

is at $0 \in G$, and

$$f_0 \in G \Rightarrow \forall t > 0, f_t \in G \text{ and } \lim_{t\to\infty} f_t = 0$$

A-2 All the trajectories of the deterministic system 7.3 starting at $f_0 \in \partial G$ converge to 0 as $t \to \infty$.
A-3 $\bar{V} \triangleq \inf_{z\in\partial G} V(0,z) < \infty$
A-4 There exists $M < \infty$ such that for all $\rho > 0$ small enough and all x, y with $|x-y| \le |x-z| + |y-z| \le \rho$ for some $z \in \partial G \cup \{0\}$ there is a function $g \in L_2$ such that $\|g\| < M$ where

$$f_t = x + \int_0^t b(f_s)\, ds + \int_0^t \sigma(f_s)\, \dot{g} ds.$$

Remark 7.2. *Assumption A-3 is natural otherwise all points on ∂G are equally unilikely on the large deviation scale. Assumption A-4 is related to the controlability of system 7.1 where a smooth controls replaces the Brownian motion. Also this assumption implies the following useful continuity property.*

Lemma 7.1. *Assume the condition of A-4. For any $\delta > 0$, there exists $\rho > 0$ small enough such that*

$$\sup_{x,y\in B_\rho} \inf_{t\in[0,1]} V(x,y,t) < \delta \tag{7.4}$$

$$\sup_{x,y:\inf_{z\in\partial G}\{|y-z|+|x-z|<\rho\}} \inf_{t\in[0,1]} V(x,y,t) < \delta. \tag{7.5}$$

Throughout this section we also denote as $B_\rho = \{x : |x| \le \rho\}$ and $S_\rho = \{x : |x| = \rho\}$. The first lemma gives a uniform lower bound on the probability of exit from G.

Lemma 7.2. *For any $\eta > 0$ and $\rho > 0$ small enough, there is $T < \infty$ such that*

$$\liminf_{\epsilon\to 0} \epsilon \log \inf_{x\in B_\rho} \mathbb{P}_x[\tau^\epsilon < T] > -\left(\bar{V} + \eta\right)$$

Proof. We fix $\eta > 0$ and $\rho > 0$ then from lemma 7.1

$$\sup_{x\in B_\rho} \inf_{t\in[0,1]} V(x,0,t).$$

Then there exists a path ψ^x of length $t_x < 1$ such that

$$I(\psi^x) < \frac{\eta}{3}, \text{where } \psi_0^x = x \text{ and } \psi_{t_x}^x = 0.$$

From assumption A-3 there exists $z \notin \bar{G}$ for $t < \infty$ and the distance Δ of z from $\bar{G}$ is positive. Then there exsists a path $\phi \in C([0,T])$ such that

$$I_{0,T}(\phi) \le \bar{V} + \frac{\eta}{3} \text{where } \phi_0 = 0 \text{ and } \phi_T = z$$

Now we construct the following path ϕ^x by concatenating ψ^x and ϕ, in that order, and extending the resulting function to be of length $T_0 = T + 1$ by following the trajectory of 7.3 after reaching z. Then it follows that

$$I_{x,T_0}(\phi^x) < \bar{V} + \frac{2\eta}{3}.$$

Consider the set

$$\Psi \triangleq \bigcup_{x \in B_\rho} \left\{ \psi \in C([0, T_0]) : \|\psi - \phi^x\| < \frac{\Delta}{2} \right\}.$$

We observe that Ψ is an open subset of $C([0, T_0])$ that contains the functions $\{\phi^x\}_{x \in B_\rho}$. Therefore by Theorem 6.3

$$\limsup_{\epsilon \to 0} \epsilon \log \inf_{x \in B_\rho} \mathbb{P}_x[X^\epsilon \in \Psi] \geq - \sup_{x \in B_\rho} \inf_{\psi \in \Psi} I_{x,T_0}(\psi)$$

$$\geq_{x \in B_\rho} I_{x,T_0}(\phi^x) > -\left(\bar{V} + \eta\right).$$

But $\{\tau^\epsilon \leq T_0\} \supseteq \{X^\epsilon \in \Psi\}$. Since if $\psi \in \Psi$ then $\psi_t \notin \bar{G}$ for some $t \in [0, T_0]$ and the proof is complete. □

Next notice that the probability the diffusion 7.1 wanders in G for an arbitrary long time without hitting a small meighborhood of 0 is exponential negligible. More precisely,let

$$\sigma_\rho \triangleq \inf\{t : X_t^\epsilon \in B_\rho \cup G\} \tag{7.6}$$

where $B_\rho \subset G$. Then

$$\lim_{t \to} \limsup_{\epsilon \to 0} \log \sup_{x \in G} \mathbb{P}_x[\sigma_\rho > t] = -\infty. \tag{7.7}$$

Now we give an upper bound relates the quasi-potential with the probability the excursion started from a small sphere of 0 hits a given subset of ∂G before hitting an even smaller sphere.

Lemma 7.3. *For any closed set $N \subset \partial G$*

$$\limsup_{\epsilon \to 0} \epsilon \log \sup_{x \in S_{2\rho}} \mathbb{P}_x[X_{\sigma_\rho}^\epsilon > t] \leq - \inf_{z \in N} V(0, z)$$

where σ_ρ is difined by 7.6.

Proof. We fix a closed set $N \subseteq \partial G$ and define $V_N \triangleq (\inf_{z \in N} V(0, z) - \delta) \wedge \frac{1}{\delta}$. By lemma 7.1

$$\inf_{y \in S_{2\rho}, z \in N} V(y, z) \geq \inf_{z \in N} V(0, z) - \sup_{y \in S_{2\rho}} V(0, y) \geq V_N$$

Moreover, by 7.7 there exists $T < \infty$ large enough for

$$\limsup_{\epsilon \to 0} \epsilon \log \sup_{\psi \in S_{2\rho}} \mathbb{P}_y[\sigma_\rho > T] \leq -V_N.$$

We consider the set

$$\Phi \triangleq \{\phi \in C([0, T]) : \exists t \in [0, T] \text{such that } \phi_t \in N\}$$

Notice that

$$\inf_{y \in S_{2\rho}, \phi \in \Phi} I_{y,T}(\phi) \geq \inf_{y \in S_{2\rho}, z \in N} V(y, z) \geq -V_N$$

then by Theorem 6.3

$$\limsup_{\epsilon\to 0}\epsilon\log\sup_{y\in S_{2\rho}}\mathbb{P}_y[X^\epsilon\in\Phi]\le -\inf_{y\in S_{2\rho},\phi\in\Phi}I_{y,T}(\phi)\le -V_N$$

Since

$$\left\{X^\epsilon_{\sigma_\rho}\in N\right\}\subseteq\{\sigma_\rho>T\}\cup\{X^\epsilon\in\Phi\}$$

we obtain

$$\limsup_{\epsilon\to 0}\epsilon\log\sup_{y\in S_{2\rho}}\mathbb{P}_y[X^\epsilon_{\sigma_\rho}\in N]\le -V_N.$$

If we let $\delta\to 0$ the proof of the lemma is completed. □

In order to extend the upper bound to hold for every $X^\epsilon_0\in G$ we observe,that as $\epsilon\to 0$ with high probability X^ϵ attracted to a a small neighborhood of 0 without hitting ∂G on its way.

Lemma 7.4. *For every $\rho>0$ such that $B_\rho\subset G$ and all $x\in G$*

$$\lim_{\epsilon\to 0}\mathbb{P}_x[X^\epsilon_{\sigma_\rho}\in B_\rho]=1$$

Fix $x\in G\setminus B_\rho$, let f denote the trajectory of 7.3 with initial condition $f_0=x$ and let $T=\inf\{t: f_t\in S_\rho\}<\infty$. Since f is a continuous path that does not hit the compact set ∂G,then $d=dist\,(f,\partial G)>0$ for $t\le T$. Suppose that this distance is smaller than ρ and let X^ϵ_t be the solution of 6.4 with $X^\epsilon_0=x$. Then

$$\sup_{t\in[0,T]}|X^\epsilon_t-f_t|\le\frac{d}{2}\Rightarrow X^\epsilon_{\sigma_\rho}\in B_\rho.$$

By uniform Lipscitz continuity of $b(\cdot)$ we have that

$$\begin{aligned}|X^\epsilon_t-f_t|&=\int_0^t b(X^\epsilon_s)-b(f_s)\,ds+\sqrt{\epsilon}\int_0^t\sigma(X^\epsilon_s)\,dW_s\\&\le B\int_0^t|x^\epsilon_s-f_s|ds+\sqrt{\epsilon}\left|\int_0^s\sigma(X^\epsilon_s)\,dW_s\right|.\end{aligned}$$

Hence, By Cronwall's lemma we have that

$$\sup_{t\in[0,T]}|X^\epsilon_t-f_t|\le\sqrt{\epsilon}e^{BT}\sup_{t\in[0,t]}\left|\int_0^t\sigma(X^\epsilon_s)\,dW_s\right|$$

Thus

$$\begin{aligned}\mathbb{P}_x[X^\epsilon_{\sigma_\rho}\in\partial G]&\le\mathbb{P}[\sup_{t\in[0,T]}|X^\epsilon_t-f_t|>d/2]\\&=\mathbb{P}\left[\frac{e^{-BT}}{\sqrt{\epsilon}}\sup_{t\in[0,T]}|X^\epsilon_t-f_t|>\frac{de^{-BT}}{\sqrt{\epsilon}}\right]\\&\le\mathbb{P}\left[\sup_{t\in[0,T]}\left|\int_0^t\sigma(X^\epsilon_s)\,dW_s\right|>\frac{de^{-BT}}{\sqrt{\epsilon}}\right].\end{aligned}$$

But $M_t=\int_0^t\sigma(X^\epsilon_s)\,dW_s$ is a martingale, then $\langle M\rangle_t=\int_0^T trace\sigma(X^\epsilon_s)\,\sigma(X^\epsilon_s)'\,ds$. Therefore, by maximal inequality we have that

$$\mathbb{P}\left[\sup_{t\in[0,T]}\left|\int_0^t\sigma(X^\epsilon_s)\,dW_s\right|>\frac{de^{-BT}}{\sqrt{\epsilon}}\right]\le k\epsilon\mathbb{E}_x\left[\int_0^T trace\sigma(X^\epsilon_s)\,\sigma(X^\epsilon_s)'\,ds\right]\to 0$$

as $\epsilon\to 0$, where k is independent of ϵ and $k<\infty$.

Remark 7.3. *In the language of the theory of differential equations this theorem can be formulated in the following equivalent form.*
Let $g(x)$ be a continuous function defined on the boundary ∂G of a domain G. Let us consider the Dirichlet problem in G

$$\frac{\epsilon}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2 u^\epsilon(x)}{\partial x_i \partial x_j} + \sum_i b_i(x)\frac{\partial u^\epsilon(x)}{\partial x_i} = 0, \quad x \in G$$

$$u^\epsilon(x) = g(x), \quad x \in \partial G$$

If the assumption of the theorem are satisfied , then there exist a unique $y_0 \in \partial G$ such that

$$\lim_{\epsilon\to 0} u^\epsilon(x) = g(y_0).$$

In the next section we formulate the above result in the language of the theory of PDEs, in particular as a result of viscosity solution of a parabolic problem. The interest reader should refer in Chapter 10 of [6] for a more detailed introduction in viscosity solutions of a Hamilton-Jacobi-Bellman equation.

7.2 Viscosity solution approach

Recall that $\{X_s^\epsilon, t \le s \le \tau\}$ be the diffusion with small parameter noise and τ_ϵ the exit time from an open set G. Then

$$\lim_{\epsilon\to 0} \ln \mathbb{P}[\tau^\epsilon \le T] = -\inf\{I(\psi) : \psi \in \mathcal{H}([0,T]) : \psi(t) = x \quad \tau(\psi) \le T\} = -V_0(t,x).$$

We set

$$\Phi^\epsilon(t,x) = \mathbb{P}[\tau^\epsilon < t_1] \tag{7.8}$$

where $t_1 < \infty$ with boundary data

$$\Phi^\epsilon(t,x) = 1 \qquad (t,x) \in [0,t_1) \times \partial G$$

$$\Phi^\epsilon(t_1,x) = 0 \qquad x \in \bar{G}$$

Then 7.8 satisfies the backward stochastic differential equation

$$-\frac{\partial \Phi^\epsilon(t,x)}{\partial t} - b(t,x) D_x \Phi^\epsilon(t,x) - \frac{\epsilon}{2}\sum_j \sum_i a_{ij}(t,x)\frac{\partial^2 \Phi^\epsilon(t,x)}{\partial x_i \partial x_j} = 0. \tag{7.9}$$

We, now, make the logarithmic transformation:

$$V^\epsilon = -\epsilon \log \Phi^\epsilon(t,x) \tag{7.10}$$

then 7.9 become

$$-\frac{\partial V^\epsilon(t,x)}{\partial t} - b(t,x) D_x V^\epsilon(t,x) - \frac{\epsilon}{2}\sum_i \sum_j a_{ij}(t,x)\frac{\partial^2 V^\epsilon(t,x)}{\partial x_i \partial x_j}$$

$$+\frac{1}{2}\sum_i \sum_j a_{ij}(t,x) D_x V^\epsilon (D_x V^\epsilon)' = 0 \tag{7.11}$$

and the boundary data become

$$V^\epsilon(t,x) = 0, \quad (t,x) \in (0,t_1] \times \partial G$$

$$\lim_{t \to t_1} V^{\epsilon}(t,x) = \infty, \qquad x \in G.$$

The later boundary condition means that $x(s)$ must reach the boundary before t_1. Because of the nonstandard form of the terminal data, the stability results for viscosity solutions are not directly applicable to V^{ϵ}. However, we could truncate V^{ϵ} to a level as follow $V^{\epsilon} = -\epsilon \log \Phi^{\epsilon}(t,x) \wedge \frac{1}{\epsilon}$ or two estimates of V^{ϵ} could control the convergence of V^{ϵ}. As we can see later. As $\epsilon \to 0$ the PDE 7.11 becomes a first order PDE

$$-\frac{\partial V^0(t,x)}{\partial t} - b(t,x) D_x V^0(t,x) + \frac{1}{2}\sum_i \sum_j a_{ij}(t,x) \left(D_x V^0\right)' \left(D_x V^0\right) = 0 \tag{7.12}$$

By PDEs methods and viscosity solutions we can prove that $V^{\epsilon} \to V^0$ solution to the above PDE. Moreover V^0 has a representation in terms of control theory. Next, we consider the Hamiltonian function

$$H(t,x,p) = -b(t,x)\,p + \frac{1}{2}p'\sigma\sigma'(t,x)\,p.$$

so that

$$-\frac{\partial V^0}{\partial t}(t,x) + H(t,x,p) = 0$$

Since the Hamiltonian is quadratic and particular convex in p, we can use the Legendre transform and may rewrite

$$H(t,x,p) = \sup_{u \in \mathbb{R}^d} \{-up - L(t,x,u)\}$$

$$= -\inf_{u \in \mathbb{R}^d} \{up + L(t,x,u)\}$$

where

$$L(t,x,u) = \sup_{p \in \mathbb{R}^d} \{-up - H(t,x,u)\}$$

$$= \frac{1}{2}(u - b(t,x)) \left(\sigma\sigma'(t,x)\right)^{-1} (u - b(t,x))'$$

and $(t,x,u) \in [0,t_1] \times G \times \mathbb{R}^d$. Hence, 7.12 is rewritten as:

$$-\frac{\partial V^0(t,x)}{\partial t} + \inf_{u \in \mathbb{R}^d} \{up + L(t,x,u)\} = 0 \tag{7.13}$$

which together with the boundary data is associated to the value function for the following calculus of variation problem.

$$V^0 = \inf_{x(\cdot) \in \mathcal{H}} \int_0^{t_1} L(s,x,u)\,ds$$

$$= \inf \int_0^{t_1} \frac{1}{2}(\dot{x}(s) - b(x,s)) \left(\sigma\sigma^T\right)^{-1}(t,x)\,(\dot{x}(s) - b(x(s)))'\,ds.$$

Then from control theory the solution to the Hamilton-Jacobi-Bellman equation is represented by a unique viscosity solution V^0, where $\mathcal{H}$ is the Cameron-Martin space that we defined in the previous section. Therefore, the large deviation results stated as

$$\lim_{\epsilon \to 0} \epsilon \log \Phi^{\epsilon}(t,x) = -V^0(t,x)$$

where $V^0(t,x)$ is the rate function. Now, we have to prove that V^0 is the unique viscosity solution of 7.13 which is the value function of an optimal control problem that we will define now: Let $\mathcal{U}(t,x) = \{u(\cdot) \in U^0(t) : \tau < t_1\}$, where $\mathcal{U}$ denote the set of all controls, and U^0 is the space of all bounded, Lebesgue measurable, $\mathbb{R}^n$ valued functions on $[0, t_1]$. Then $\dot{x}(s) = u(s) \Rightarrow u(s) = \frac{1}{t_1}\int_0^{t_1} \dot{x}(s)\,ds$.

$$V(t,x) = \inf_{x(\cdot)\in\mathcal{H}(t,x)} J(t,x,u)$$

where

$$J(t,x,u) = \int_t^{\tau} L(s, x(s), \dot{x}(s))\,ds.$$

and τ is the first exit time of $(s, x(s))$ from G.
We make the following assumptions: σ is invertible, σ and b are bounded and Lipschitz continuous on $[t_0, t_1] \times \mathbb{R}^d$ and L has the following properties:

(i) $L \geq 0$, $L_{uu} \geq 0$

(ii) There exists k such that $k|u|^2 \leq L(t,x,u)$ when $u \geq R$

(iii) $|L_x(t,x,u)| + |L_u(t,x,u)| \leq K\left(1 + |u|^2\right)$.

Next, we continue with two estimates of V^ϵ which can be derived by probabilistic methods as we have seen in the previous section.

Lemma 7.5. *Suppose that ∂G is smooth. Then there exists $K > 0$ satisfying*

$$V^\epsilon(t,x) \leq \frac{K dist(x, \partial G)}{t_1 - t}, \quad (t,x) \in \bar{G}, \epsilon \in (0,1]. \tag{7.14}$$

Proof. Suppose x_1 is the first component of the vector x, there is a constant μ satisfying $x_1 + \mu > 0$ for all $x \in G$. For $\lambda, \gamma > 0$, where γ is a constant, we define an auxiliary function

$$g^\epsilon(t,x) = \exp\left(-\frac{\lambda(x_1+\mu)}{\epsilon(t_\gamma - t)}\right), \tag{7.15}$$

where $(t,x) \in \bar{G}_\gamma$, $t_\gamma = t_1 - \gamma$, $G_\gamma = [0, t_\gamma) \times D$.
Thus, $g^\epsilon(t,x)$ is a subsolution of 7.9. More precisely,

$$\begin{aligned}
&\frac{\partial}{\partial t} g^\epsilon(t,x) - \frac{\epsilon}{2} tra(t,x) D_x^2 g^\epsilon(t,x) - b(t,x) D_x g^\epsilon(t,x) \\
&= \frac{1}{\epsilon} g^\epsilon(t,x) \frac{\lambda(x_1+\mu)}{(t_\gamma - t)^2} + \frac{1}{\epsilon} g^\epsilon(t,x) \frac{\lambda b_1(t,x)}{(t_\gamma - t)} - \frac{1}{\epsilon} a_{11}(t,x) g^\epsilon(t,x) \frac{\lambda^2}{(t_\gamma - t)^2} \\
&= -\frac{1}{\epsilon} \frac{g^\epsilon(t,x)}{(t_\gamma - t)^2} \left(\frac{1}{2} a_{11}(t,x)\lambda^2 - \lambda(x_1+\mu) - (t_\gamma - t)\lambda b_1(t,x)\right) < 0
\end{aligned}$$

where a_{11} is the first element of the matrix $a(t,x)$. Since $b(t,x)$, $a^{-1}(t,x)$ are bounded, $a_{11}(t,x)$ is uniformly bounded away from zero on $\bar{G}$. So the above expression is non positive for large enough $\lambda = \lambda^*$. Therefore, $g^\epsilon(t,x)$ is a subsolution of 7.9 in G. Moreover,

$$g^\epsilon(t_\gamma, x) = 0 \leq \Phi^\epsilon(t_\gamma, x).$$

Using the maximum principle for parabolic equations we have that

$$\Phi^\epsilon(t,x) \geq g^\epsilon(t,x), \qquad (t,x) \in \bar{G}.$$

Otherwise we could use Ito's formula to prove the above inequality. Since, the boundary of G is smooth enough there exists $\delta > 0$ such that $d(x) = dist(x, \partial G)$ and

$$G_\delta = \{x \in G : d(x) < \delta\}.$$

Now, we set

$$\tilde{g}^\epsilon(t,x) = \exp\left(-\frac{1}{\epsilon}\frac{Kd(x)}{(t_\gamma - t)}\right), \qquad (t,x) \in G_\gamma,$$

therefore,

$$\tilde{g}^\epsilon(t,x) \le g^\epsilon(t,x) \le \Phi^\epsilon(t,x)$$

since $K > 0$ satisfy

$$K \ge \frac{1}{\delta}\sup\left\{\lambda^*(x_1 + \mu) : x \in \bar{G}\right\}$$

and $d(x) > \delta$. This means that,

$$\tilde{g}^\epsilon(t,x) \le \Phi^\epsilon(t,x), \forall (t,x) \in [t_0, t_\gamma) \times D_\delta^c. \tag{7.16}$$

Then we observe that $\tilde{g}^\epsilon(t,x)$ is a subsolution on $[0, t_\gamma) \times G_\delta$. In particular,

$$\frac{\partial}{\partial t}\tilde{g}^\epsilon(t,x) - \frac{\epsilon}{2}tra(t,x)D_x^2\tilde{g}^\epsilon(t,x) - b(t,x)\tilde{g}^\epsilon(t,x)$$

$$= \frac{1}{\epsilon}\frac{Kd(x)}{(t_\gamma - t)^2}\tilde{g}^\epsilon(t,x) - \frac{1}{2\epsilon}\frac{K^2Dd(x)Dd(x)}{(t_\gamma - t)^2}\tilde{g}^\epsilon(t,x) + \frac{1}{2}\frac{KD^2d(x)}{(t_\gamma - t)}\tilde{g}^\epsilon(t,x) + \frac{1}{\epsilon}\frac{Kd(x)}{(t_\gamma - t)}\tilde{g}^\epsilon(t,x)$$

$$= -\frac{K\tilde{g}^\epsilon(t,x)}{\epsilon(t_\gamma - t)^2}\left(\frac{K}{2}a(t,x)Dd(x)Dd(t,x) - \epsilon\frac{(t_\gamma - t)}{2}tra(t,x)D^2d(x) - (t_\gamma - t)b(t,x)Dd(x) - d(x)\right)$$

$$\le -\frac{K\tilde{g}^\epsilon(t,x)}{\epsilon(t_\gamma - t)^2}\left(Ka_0 - \frac{\epsilon}{2}(t_\gamma - t)|a(t,x)||D^2d(x)| + (t_\gamma - t)|b(t,x)| - d(x)\right) < 0.$$

Since, a^{-1} is bounded and uniformly elliptic, there is a constant $a_0 > 0$ such that

$$a(t,x)\xi\xi \ge a_0|\xi|^2, \forall (t,x) \in \bar{D}, \xi \in \mathbb{R}^n$$

and $|Dd(x)| = 1$ on G_δ, so we obtain

$$a(t,x)Dd(x)Dd(x) \ge a_0|Dd(x)|^2 = a_0, \forall (t,x) \in \bar{G}$$

Using theses facts the above expression is negative for sufficiently large K. So we have prove that $\tilde{g}^\epsilon(t,x)$ is a subsolution of 7.9 on $(0, t_\gamma) \times G_\delta$. Also on $[t_0, t_\gamma] \times \partial G_\delta \cup \{t_\gamma\} \times G_\delta$ $\tilde{g}^\epsilon(t,x) = \Phi^\epsilon(t,x) = 1$. Thus, the maximum principle for parabolic equations yields

$$\tilde{g}^\epsilon(t,x) \le \Phi^\epsilon(t,x), \quad \text{on } [t_0, t_1] \times G_\delta.$$

Moreover, $\Phi^\epsilon(t,x) = \exp\left(-\frac{1}{\epsilon}V^\epsilon(t,x)\right)$, then this combined with 7.16 and the above inequality imply 7.14. ☐

Lemma 7.6. *For any $M > 0$ and $d(x) = dist(x, \partial G)$ in $C^2(\bar{G})$ with $d(x) = 0$ for all $x \in \partial G$, there exists $K_M > 0$ such that*

$$V^\epsilon(t,x) \ge Md(x) - K_M(t_1 - t), (t,x) \in [t_0, t_1] \times \bar{D}, \epsilon \in (0,1). \tag{7.17}$$

Proof. We define the auxiliary function

$$\bar{g}^{\epsilon}(t,x) = \exp\left(-\frac{1}{\epsilon}Md(x) - \frac{1}{\epsilon}K_M(t_1-t)\right)$$

and we show that $\bar{g}^{\epsilon}(t,x)$ is a supersolution of 7.9. Particular,

$$-\frac{\partial}{\partial t}\bar{g}^{\epsilon}(t,x) - \frac{\epsilon}{2}tra(t,x)D_x^2\bar{g}^{\epsilon}(t,x) - b(t,x)D_x\bar{g}^{\epsilon}(t,x)$$

$$= \bar{g}^{\epsilon}(t,x)\frac{K_M}{\epsilon} - \frac{1}{2}tra(t,x)MDd(x)\bar{g}^{\epsilon}(t,x) + \frac{1}{2\epsilon}M^2tra(t,x)Dd(x)\bar{g}^{\epsilon}(t,x)Dd(x) + b(t,x)\frac{1}{\epsilon}MDd(x)\bar{g}^{\epsilon}(t,x)$$

$$= \frac{1}{\epsilon}\bar{g}^{\epsilon}(t,x)\left(K_M - \frac{\epsilon}{2}tra(t,x)MD^2d(x) + \frac{1}{2}M^2tra(t,x)Dd(x)Dd(x) + b(t,x)Md(x)\right) > 0$$

since we set

$$K_M = \sup_{(t,x)\in[0,t_1]\times\bar{G}}\left\{-\frac{\epsilon}{2}Mtra(t,x)D^2d(x) + F(t,x,MDd(x))\right\}.$$

Consequently, $\bar{g}^{\epsilon}(t,x)$ is a supersolution of 7.9. Besides the fact that Φ^{ϵ} is not continuous at $\{t_1\}\times\partial G$, $\bar{g}^{\epsilon}(t,x)$ is continuous since,

$$\bar{g}^{\epsilon}(t_1,x) = 1, \quad x\in\partial G.$$

Using maximum principle for parabolic equations once again we obtain 7.17. □

Remark 7.4. *Using this subsolution and supersolution of $\Phi^{\epsilon}(t,x)$, we manage to find a way to control the convergence of $V^{\epsilon}(t,x)$. Also, $V^{\epsilon}\geq 0$ since $\Phi^{\epsilon}\leq 1$ and by 7.14 $V^{\epsilon}(t,x)$ is uniformly bounded for $\epsilon\in(0,1]$, $(t,x)\in[0,T]\times\bar{D}$ with any $T<t_1$. We use the Barles and Perthame procedure. We give a brief outline of this procedure, the interesting reader should refer in Chapter VII paragraph $3-4-5-6$ of [8] for the detailed steps.*
For $(t,x)\in(t_0,t_1)\times\bar{G}$, define

$$V^*(t,x) = \limsup_{(s,y)\to(t,x)} V^{\epsilon}(s,y), \tag{7.18}$$

$$V_*(t,x) = \liminf_{(s,y)\to(t,x)} V^{\epsilon}(s,y) \tag{7.19}$$

for $(s,y)\in[t_0,t_1]\times\bar{D}$. These functions however are not necessarily continuous. In fact we may only infer that they are semi-continuous. Therefore, we conclude that V^*, V_* are respectively viscosity subsolution and supersolution of 7.13 in $[t_0,T]\times D$ for every $T<t_1$. Then, using the equation 7.9 and its boundary data yields that any viscosity subsolution of this problem is dominated by any viscosity supersolution, $V_*\geq V^*$. However, by construction, $V_*\leq V^*$. Although, the terminal data of the problem is infinite, the stability result still holds. Hence, 7.17 implies that $V^*(t,x)$ and $V_*(t,x)$ converges to ∞ as $t\to t_1$ uniformly on compact subsets of D. However, this convergence is controlled by 7.14. The above properties of V^*, V_* will be used later to show the convergence of V^{ϵ} to $V^0 = V^* = V_*$ which is the unique viscosity solution of HJB equation defined previously 7.13.

Lemma 7.7. *For every $T<t_1$, V^0 is a viscosity solution of 7.13 in $[0,T)\times G$ and is Lipschitz continuous on $[0,T]\times\bar{G}$.*

Proof. We give the proof in four steps.
First step
We show that $V^0(t,x)$ is bounded on $[0,T]\times G$. In particular, we show that there exists M such that:

$$0 \le V^0(t,x) \le M dist(x,\partial G), \quad (t,x)\in[0,T]\times G. \tag{7.20}$$

To this end, let $\bar{x}\in\partial G$ nearest to x, $dist(x,\partial G)\le|x-\bar{x}|$. We know that

$$u=\frac{1}{t_1-t}\int_t^{t_1}\dot{x}(s)\,ds$$

then

$$x_0(s)=x+u_0(s-t), \quad u_0=c|\bar{x}-x|^{-1}(\bar{x}-x),$$

where $c=\frac{diamG}{t_1-T}$. Let τ_0 be the exit time from G of $x_0(s)$. Then

$$\tau_0-t=\frac{|\bar{x}-x|}{c}\le\frac{diamG}{c}=t_1-T$$

and $\tau_0<t_1$ when $t\le T$. So, for every $x(\cdot)\in\mathcal{H}(t,x)$

$$0\le J(t,x,u)\le\int_t^{\tau_0}L(s,x_0(s),u_0)\,ds\le C(\tau_0-t)\le C\frac{diamG}{c}\le M dist(x,\partial G)$$

since L is bounded from its definition $L(s,y,u)\le C$ for all (s,y) and $M=\frac{C}{c}$.
Second step
We show that $V^0(t,\cdot)$ is Lipschtz continuous on $\bar{G}$. More precisely, we show that for any $x,y\in D$ there is $M_1>0$ such that

$$|V^0(t,x)-V^0(s,y)|\le M_1|x-y|.$$

For any $\lambda\in(0,1)$, let $y_\lambda(s)=x(s)+\lambda(y-x)$. Let τ_1 be the exit time of $(s,y_1(s))$ from G and $\tau_2=\min(\tau,\tau_1)$. By the dynamic programming principle we have that

$$V^0(t,y)\le\int_t^{\tau_2}L(s,y_1(s),\dot{x}(s))\,ds+V(\tau_2,y_1(\tau_2))$$

$$=\int_t^{\tau_2}L(s,x(s),\dot{x}(s))\,ds+\int_t^{\tau_2}\int_0^1 L_x(s,y(s),\dot{x})\lambda(y-x)\,d\lambda ds+V(\tau_2,y_1(\tau_2)).$$

From the definition of L that $L(s,y,\dot{x})=\int_t^{t_1}\frac{1}{2}\left(\dot{x}(s)-b(x,s)\left(\sigma\sigma^T\right)(s,x)(\dot{x}(s)-b(x,s))\right)'ds$ and the properties of b,σ we have that there exists K such that

$$|L_x(s,x(s),\dot{x}(s))\le K\left(1+|\dot{x}(s)|^2\right)$$

for all $(s,y)\in[0,T]\times\bar{G}$ and $\dot{x}\in\mathbb{R}^n$.
Therefore,

$$V(t,y)\le J(t,x,\dot{x})+K|x-y|\int_t^{\tau_2}\left(1+|\dot{x}(s)|^2\right)ds+V(\tau_2,y_1(\tau_2)).$$

Now, if $\tau_2=t_1\le\tau$, then $y_1(\tau_2)\in\partial G$ and $V(\tau_2,y_1(\tau_2))=0$, and if $\tau_2=\tau<\tau_1$ then $x(\tau_2)\in\partial G$. By 7.20 we obtain

$$V(\tau_2,y_1(\tau_2))\le M|y_1(\tau_2)-x(\tau_2)|<M|x-y|.$$

Thus,

$$V(t,y) \le V(t,x) + k + K|x-y|(t_1 - t_0) + K|x-y|C_1 + M|x-y|$$
$$\le V(t,x) + k + |x-y|\left((t_1 - t_0 + C_1)K + M\right)$$

since $J(t,x,u) < V(t,x) + k$ for any $k \in [0,1]$, $\int_t^\tau |c\dot{x}(s)|^2 ds < C_1$, and we set $M_1 = K(t_1 - t_0 + C_1) + M$. Then k is chosen arbitrarily and we can write

$$|V(t,y) - V(s,y)| \le M_1|x-y|, \tag{7.21}$$

for $0 \le t \le T$ and $x, y \in \bar{G}$.

Third step

Next, we prove that V^0 is a viscosity solution of HBJ equation 7.13.

Let $(t,x) \in [0,T) \times G$, the dynamic programming principle implies that

$$V^0(t,x) = \inf_{\dot{x}(\cdot)\in\mathcal{H}} \left\{ \int_t^{\tilde{\tau}} L(s,x(s),\dot{x}(s))\,ds + V^0(\tilde{\tau}, x(\tau)) \right\}$$

where $\tilde{\tau} = \tau \wedge T$. From stochastic control we verify that the terminal cost function satisfies

$$V^0(\tilde{\tau}, x(\tau)) = \begin{cases} 0 & \text{if} \quad (t,x) \in [0,T] \times \partial G \\ V^0(T,x), & \text{if} \quad x \in G \end{cases}$$

also from the previous step we know that V^0 is Lipschitz continuous. Also, using the assumptions of the integrand L we can apply the Theorem 10.4 of [8] that the value function V^0 is the unique Lipschitz continuous viscosity solution to the HBJ equation 7.9 with the same boundary data. Therefore, V^0 is Lipschitz continuous and a viscosity solution to the HBJ 7.13 on $[t_0,T) \times G$.

Fourth step

Finally, we verify the boundary and terminal data of HBJ equation 7.13. First, $V \ge 0$ since L is positive. Then, we choose $x(\cdot) \in \mathcal{H}(t,x)$ satisfying $\tau = t$, for $(t,x) \in [0,t_1) \times \partial G$ we have that $V^0(t,x) = 0$. To prove $V^0(t,x) = \infty$, let $x \in G$, thus

$$dist(x, \partial G) \le |x - \bar{x}| = \left| \int_t^\tau \dot{x}(s)\,ds \right|$$

since $x(\tau) \in \partial G$. Also, the boundedness of $a(t,x)$ yields

$$L(t,x,\dot{x}) \ge c_0|\dot{x} - b(t,x)|^2$$

for c_0 constant. This means that

$$\int_t^\tau L(s,x(s),\dot{x}(s))\,ds \ge c_0 \int_t^\tau |\dot{x}(s) - b(s,x(s))|^2 ds$$
$$\ge c_0 \int_t^\tau |\dot{x}(s)|^2 ds - K$$
$$\ge \frac{c_0}{\tau - t} \left| \int_t^\tau \dot{x}(s)\,ds \right|^2 - K,$$

where K is a constant. Finally, we obtain

$$V^0(t,x) \ge \frac{c_0\,(dist(x,\partial G))^2}{t_1 - t} - K$$

since $\tau - t \le t_1 - t$ and we have that

$$\lim_{t\to t_1} V^0(t,x) = \infty$$

and this completes the proof. $\square$

Now, we are ready to prove that V^ϵ converges uniformly to V^0 as $\epsilon \to 0$.

Theorem 7.1. *Assume that the properties of b, a, a^{-1} satisfied. Then V^ϵ converges to V^0 uniformly on compact subsets of $[t_0, t_1) \times \bar{D}$ as $\epsilon \to 0$.*

Proof. Recall that $V_*(t,x) = \liminf_{\epsilon\to 0} V^\epsilon(t,x)$ and $V^* = \limsup_{\epsilon\to 0} V^\epsilon(t,x)$. Then,

$$V^*(t,x) - V_*(t,y) \le V^0(t+\delta,x) - V^0(t-\delta,y)\,| \le C(\delta + |x-y|)$$

and

$$V^*(t,x) - V_*(t,y) \ge V_*(t,x) - V^*(t,x) \ge -C(\delta + |x-y|)$$

Therefore

$$|V^*(t,x) - V_*(t,y)\,| \le C(\delta + |x-y|)\,.$$

Now fix x. In order to show the uniformly convergence of V^ϵ to V^0, it suffices to show that

$$V^*(t,x) \le V^0(t+\delta,x) \quad \text{ for } (t,x) \in (t_0, t_1-\delta) \times \bar{G}$$

and

$$V^0(t-\delta,x) \le V_*(t,x) \quad \text{ for } (t,x) \in (t_0+\delta, t_1) \times \bar{G}.$$

First, we show the first inequality,

$$V^*(t,x) \le V^0(t+\delta,x) \quad \text{for} \quad (t,x) \in [0, t_1-\delta) \times \bar{G}. \tag{7.22}$$

We observe that

$$\liminf_{T\to t_1} \sup_{x\in\bar{D}} \left\{V^0(T,x) - V^*(T-\delta,x)\right\} > 0 \tag{7.23}$$

We set $\sup_{x\in\bar{G}} \{V(T,x) - V^*(T-\delta,x)\} = a(T,\delta)$. In particular, choose any sequence $(T_n, x_n) \to (t_1, \bar{x})$.
When $\bar{x} \in \partial G$, then the estimate $V^\epsilon(t,x) \le \frac{K dist(x,\partial G)}{t_1 - t}$ and the positivity of V^0 means that 7.23 satified.
When $\bar{x} \in G$, then the fact that $V^0(t,x) \ge \frac{c_0 (dist(x,\partial G))^2}{t_1-t} - K$
together with $V^\epsilon(t,x) \le \frac{K dist(x,\partial G)}{t_1-t}$ give 7.23.
Now, for any $\delta > 0$ and $T < t_1$ note that $V^0(t,x) + a(T,\delta)$ and $V^*(t,x)$ are a viscosity subsolution and supersolution respectively of 7.9 in $(0, t_1-\delta) \times \bar{G}$. Thus, the boundary and terminal data are the following

$$V^0(t,x) = a(t,\delta)\,, \quad \text{for} \quad (t,x) \in (t_0+\delta, T) \times \partial G$$

$$V^0(t,x) = V(T,x) + a(T,\delta)\,, \quad \text{for} \quad x \in G.$$

Since, V^0 is Lipschitz continuous then a weak comparison result holds giving

$$V^0(t+\delta,x) + a(T,\delta) \ge V^*(t,\delta)\,.$$

Finally, if we let $T \to t_1$ we get the 7.22.
The second inequality

$$V^0(t-\delta,x) \le V_*(t,x) \quad \text{for} \quad (t,x) \in (0+\delta, t_1) \times \bar{G}$$

is a consequence of

$$\liminf_{t\to t_1} \inf_{x\in\bar{G}} \left\{V_*(t,x) - V^0(t-\delta,x) - V^0(t-\delta,x)\right\} > 0$$

and the weak comparison principle that we used before. The above inequality yields due to the fact that

$V^\epsilon(t,x) = 0, \quad$ for $\quad (t,x) \in (\delta, t_1) \times \partial G$ and $V^\epsilon(t,x) \ge M d(x) - K_M(t_1 - t)$. ☐

Remark 7.5. *The above convergence result can be restated as*

$$\Phi^{\epsilon}(t,x) = \exp\left(-\frac{1}{\epsilon}\left(V^{0}(t,x) + h^{\epsilon}(t,x)\right)\right), \tag{7.24}$$

where h^{ϵ} *converges uniformly on compact subsets of* $[0,t_1)\times\bar{G}$ *as* $\epsilon\to 0$*. When* $V^{0}(t,x) > 0$*, we conclude that* $\Phi^{\epsilon}(t,x)\to 0$ *exponentially fast as* $\epsilon\to 0$*. However, if* $V^{0}(t,x)=0$ *the expansion 7.24 does not provide any information. So it is interest the case that* $V^{0}(t,x) > 0$ *on* $[0,t_1]\times\bar{G}$*.*

Part III

Applications in Finance

Chapter 8

Introduction

Large Deviations finds important applications in finance where questions related to external events play an increasingly important role. Large deviations arise in various financial contexts. They occur in risk management for the computation of the probability of large losses of a portfolio subject to market risk or the defaut probabilities of a portfolio under credit risk. Large deviations methods are largely used in rare events simulation and so appear naturally in the approximation of option pricing, in particular for barrier option and out of the money options. More recently, there has been a growing literarure on various asymptotics (small-time, large time, fast mean-reverting, extreme strike) for stochastic volatility mdels, see [1],[15].

We illustrate our purpose with the following toy example. Let, X be a real valued random variable, and consider the problem of computing or estimating $\mathbb{P}[X > l]$, the probability that X exceeds some level l. In finance, we may think of X as the loss of a portfolio subject to credit or market risk, and we are interested in the probability of large loss or default probability. The r.v. X may also correspond to the terminal value of a stock price, and the quantity $\mathbb{P}[X > l]$ appear typically in the computation of a call or barrier option, with small probability of payoff when the option is out of the money or the barrier l is large. To estimate $p = \mathbb{P}[X > l]$, a basic technique is Monte Carlo simulation: generate n independent paths $X_1, X_2, \ldots, X_n$ of X and use the sample mean

$$\bar{S}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \text{with} \quad Y_i = \mathbb{1}_{\{X_i > l\}}$$

The convergence of this estimate (when $n \to \infty$) follows from the law of large numbers, while the standard rate of convergence is given through the central limit theorem, in terms of the variance $v = p\,(1-p)$ of Y_i:

$$\mathbb{P}\left[|\bar{S}_n - p| \geq \frac{a}{\sqrt{n}}\right] \to 2\Phi\left(-\frac{a}{\sqrt{v}}\right),$$

where Φ is the cucmulative distribution function of the standard normal law. Furthermore, the convergence of the estimator $\bar{S}_n$ is precised with the large deviation result, known here as the Cramér theorem.

Let us now turn again to the estimation of $p = \mathbb{P}[X \geq l]$. As mentioned above, the rate of convergence of the estimator $\bar{S}_n$ is determined by:

$$Var\left(\bar{S}_n\right) = \frac{Var\left(\mathbb{1}_{\{X>l\}}\right)}{n} = \frac{p\,(1-p)}{n},$$

and the relative error is

$$\text{relative error} = \frac{\text{standard deviation of} \quad \bar{S}_n}{\text{mean of} \quad \bar{S}_n} = \frac{\sqrt{p\,(1-p)}}{p\sqrt{n}}.$$

Hence, if $p = \mathbb{P}[X > l]$ is small and since $\sqrt{p - q^2}/p \to \infty$ as p goes to zero, we realize that a large sample size is required for the estimator to achieve a reasonable relative error bound. This is a common occurance when estimating rare events. In order to improve the estimate of a tail probability $\mathbb{P}[X > l]$, one is tempted to use importance sampling to reduce variance and hence speed up the computation by requiring fewer samples. This consists basically in changing measures to try to give more weight to important outcomes, (increase the default probability). Since large deviations also deal with rare events we can see a strong link with importance sampling. There are a lot of book and notes in the literature, that the interested reader could find for Monte-Carlo methods and especially for the importance sampling technique for variance reduction see [2] and [11].
To make the idea concrete, consider again the problem of estimating $p = \mathbb{P}[X > l]$, and suppose taht X has the distrubution $\mu(dx)$. Let us took at an alernative sampling distribution $\nu(dx)$ absolutely continuous with respect t $\mu(dx)$, with density $f(x) = \frac{d\nu(x)}{d\mu(x)}$. The tail probability can be rewritten as

$$p = \mathbb{P}[X > l] = \int \mathbb{1}_{\{x>l\}} \phi(x) \nu(dx) = \mathbb{E}^{\nu}[\mathbb{1}_{\{X>l\}} \phi(X)],$$

where $\phi = \frac{1}{f}$, and $\mathbb{E}$ denotes the expectation under the measure ν. By generating i.i.d samples $\tilde{X}_1, \ldots, \tilde{X}_n$ with distribution , we have an alternative unbiased and convergent estiamte of p with

$$\tilde{S}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\tilde{X}_i>l\}} \phi\left(\tilde{X}_i\right),$$

and whose the convergence is determined by

$$Var\left(\tilde{S}_n\right) = \frac{1}{n} \int \left(\mathbb{1}_{\{x>l\}} - pf(x)\right)^2 \phi(x) \nu(dx)$$

The minimization of this quantity over all possible ν leads to a zero variance with the choise of a density $f(x) = \frac{\mathbb{1}_{\{x>l\}}}{p}$. This is of course only a theoritical result since it requires the knowledge of p, the very thing we want ro estimate! However, by noting that in this case $\nu(dx) = f(x)\mu(dx) = \mathbb{1}_{\{x>l\}}\mu(dx) / \mathbb{P}[X > l]$ is nothing else than the conditional distribution of X given $\{X > l\}$, this suggests to use an importance sampling change of measure that makes the rare event more likely to occur. This method of suitable change of measure is also the key step in proving large deviations results. In this chapter we show how to use large deviations approximation through importance sampling for Monte-Carlo computation of expectation arising in option pricing. In the frame of continuous time models, we are interested in the computation of

$$I_g = \mathbb{E}[g(S_t, 0 \le t \le T)]$$

where S is the underlying asset price, and g is the payoff function of the option, eventually path-dependent, i.e. depending on the path process, S_t, $0 \le t \le T$. The Monte-Carlo approximation technique consists in simulating N independent sample paths $\left(S_t^i\right)_{0 \le t \le T}$, $i = 1, 2, 3, \ldots, N$, in the distribution of $(S_t)_{0 \le t \le T}$, and approximating the required expectation by the sample mean estimator:

$$I_g^N = \frac{1}{N} \sum_{i=1}^{N} g\left(S^i\right).$$

The consistency of the estimator is ensured by the Law of Large Numbers, while the error approxiamtion is given by the variance of this estimator from the central limit theorem.

The lower is the variance of $g(S)$, the better is the approxiamtion for a given number N of simulations As already mentioned, the basic idea of importance sampling is to reduce variance by changing probability measure from which paths are generated. Here, the idea is to change the distribution of the price process to be simulated in order to take account the specifities of the payoff function g, and to derive the process to the region of high contribution to the required expectation. We focus in this section in the importance sampling technique within the context of diffusions models, and then show how to obtain an optimal change of measure by a large deviations approximation of the required expectation.

8.1 Importance sampling for diffusions via Girsanov's theorem

In this section, we briefly describe the importance sampling variance reduction technique for diffusions. Let X be a d-dimensional diffusion process gonverned by

$$dX_s = b(X_s)\,ds + \sigma(X_s)\,dW_s, \tag{8.1}$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the borel-measurable functions b, σ satisfy the Lipschitz condition ensuring the existence of a strong solution to the 8.1. We denote by $X_s^{t,x}$ the solution to 8.1 starting from x at time t, and we define the function

$$u(t,x) = \mathbb{E}\left[g\left(X_s^{t,x}, t \leq s \leq T\right)\right] \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

Let, ϕ_t be a $\mathbb{R}^d$-valued adapted process on $[0,T]$ such that the process

$$M_t = \exp\left\{\int_0^t \phi_s dW_s - \frac{1}{2}\int_0^t |\phi_s|^2 ds\right\}, \quad t \in [0,T]$$

is a local martingale by Ito's formula. Since $M_t(\phi)$ is non-negative, Fatou's lemma implies that $M_t(\phi)$ is a supermartingale. Then a supermartingale is a martingale if $\mathbb{E}[M_T(\phi)] = \mathbb{E}[M_0(\phi)] = 1$. But, this is ensured by the Novikov condition that $\mathbb{E}[\exp\left(\frac{1}{2}\int_0^T |\phi_u|^2 du\right)] < \infty$. Therefore, we can have a probability measure $\mathbb{Q}$ absolutely continuous with respect to $\mathbb{P}$ on $(\Omega, \mathcal{F})$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T.$$

It is worth noting that for all $\mathcal{F}$-measurable functions $t < T$, $\frac{d\mathbb{Q}}{d\mathbb{P}} = M_t$ because of the martingale property of M_t. Moreover, by Girsanov's theorem, the process $\tilde{W}_t = W_t - \int_0^t \phi_s ds$, $0 \leq t \leq T$ is a Brownian motion under the new probability measure $\mathbb{Q}$, and the dynamics, now, of X with respect to $\mathbb{Q}$ is given by

$$dX_s = \left(b(X_s) - \sigma(X_s)\phi_s\right)ds + \sigma(X_s)\,d\tilde{W}_s.$$

Thus, from Bayes formula the expectation we want to compute can be rewritten as

$$u(t,x) = \mathbb{E}^{\mathbb{Q}}\left[g\left(X_s^{t,x}, t \leq s \leq T\right)L_T\right], \tag{8.2}$$

where L is the $\mathbb{Q}$ martingale

$$L_t = \frac{1}{M_t} = \exp\left(\int_0^t \phi_s d\tilde{W}_s - \frac{1}{2}\int_0^t |\phi_s|^2 ds\right), \quad 0 \leq t \leq T. \tag{8.3}$$

Due to 8.2 we have an alternative Monte-Carlo estimator for $u(t,x)$ for any choise of ϕ

$$I_{g,\phi}^{N}(t,x)=\frac{1}{N}\sum_{i=1}^{N}g\left(X^{i,t,x}\right)L_{T}^{i},$$

by simulating N independent sample paths $\left(X^{i,t,x}\right)$ and L_T^i of $X^{t,x}$ and L_T under $\mathbb{Q}$ given by 8.2 and 8.3. Hence, the change of probability measure through the choise of ϕ leads to a modification of the drift process in the simulation of X. The variance reduction technique consists in determining a process ϕ, which induces a smaller variance for the corresponding estimator $I_{g,\phi}$ than the initial I_g. In the next section we present an approach leading to the construction of such process ϕ. In this approach, the process ϕ is stochastic and requires an approximation of the expectation of interest. This approach relies on asymptotic results from large daviations techniques.

8.2 Option pricing approximation with a Freidlin-Wentzell large deviation priniciple

In this section, we are looking for a stochastic process ϕ, which allows us to reduce the variance (possibly to zero!) the variance of the corresponding estimator. We give the theoritical approach of the problem which is due to [10]. The heuristics for achieving this goal is based on the following argument. Suppose that the payoff function g depends only on the terminal value X_T. Then by applying Ito's formula to the $\mathbb{Q}$-martingale $u\left(s,X_S^{t,x}\right)L_s$ for $t\leq s\leq T$, we have that

$$g\left(X_T^{t,x}\right)L_T=u(t,x)L_t+\int_t^T L_s\left(D_xu\left(s,X_s^{t,x}\right)'\sigma\left(X_s^{t,x}\right)+u\left(s,X_s^{t,x}\right)\phi_s'\right)d\tilde{W}_s.$$

Hence, the variance of $I_{g,\phi}^N(t,x)$ is given by

$$Var_{\mathbb{Q}}\left(I_{g,\phi}^N(t,x)\right)=\frac{1}{N}\mathbb{E}^{\mathbb{Q}}\left[\int_t^T L_s^2|\left(D_xu\left(s,X_s^{t,x}\right)'\sigma\left(X_s^{t,x}\right)+u\left(s,X_s^{t,x}\right)\phi_s'\right)|^2ds\right].$$

So, if the function u were known, then one could vanish the variance by choosing

$$\phi_s=\phi_s^*=-\frac{1}{u\left(s,X_s^{t,x}\right)}\sigma'\left(X_s^{t,x}\right)D_xu\left(s,X_s^{t,x}\right),\quad t\leq s\leq T. \tag{8.4}$$

Indeed,the function u is unknown, this is precisely what we want to compute, but this suggest to use a process ϕ from the above formula with an approxiamtion of the function u. We may then reasonably hope to reduce the variance and also to use such a method for more general payoff functions, possibly path-dependent. We shall use a large deviation approxiamtion for the function u.
The basic idea for use the large deviations approxiamation to the expectation function u is the following.
Suppose the option of interest, characterized by its payoff function g, has a low probability of excersice, e.g. it is deeply out of the money. Then, a large proportion of simulated paths end up out of the excersice domain, giving no contribution to the Monte-Carlo estimator but increasing the variance. In order to reduce the variance, it is interesting to change of drift in the simulation of price process to make the domain excersice more likely. This is achieved with a large deviation approximation of the process of interest

in the asymptotics of small diffusion term: such a result is known in the literature as Freidlin-Wentzell sample path large deviations principle. Equivalently, by time-scaling, this amounts to large deviation approximation of the process in small time, studied by Varadhan.

To illustrate our purpose, let us cosider an option that pays one unit of numéraire iff the underlying asset reached a given up-barrier K. Within a stochastic volatility model $X = (S, Y)$ as in 8.1 and given by:

$$dS_t = \sigma(Y_t) S_t dW_t^1 \tag{8.5}$$

$$dY_t = \eta(Y_t) dt + \gamma(Y_t) dW_t^2 \tag{8.6}$$

with $d\langle W_1, W_2\rangle_t = \rho dt$, its price is given by

$$u(t,x) = \mathbb{E}\left[\mathbb{1}_{\max_{t\le u\le T} S_u^{t,x} \ge K}\right] = \mathbb{P}[\tau_{t,x} \le T], \quad t \in [0,T], x = (s,y) \in (0,\infty)\times\mathbb{R},$$

where

$$\tau_{t,x} = \inf\left\{u \ge t : X_u^{t,x} \notin D\right\}, \quad D = (0,K)\times\mathbb{R}.$$

Then, the event $\left\{\max_{t\le u\le T} S_u^{t,x} \ge K\right\} = \{\tau_{t,x} \le T\}$ is rare when $x = (s,y) \in D$, i.e. $s < K$ (out of the money option) and the time to maturity $T-t$ is very small. The large deviations asymptotics for the exit probability $u(t,x)$ in small time to maturity $T-t$ is provided by the Freindlin-Wentzell and Varadhan theories. Indeed, we see from the time-homogeneity of the coefficients of the difussion and by time-scaling that we may write $u(t,x) = u_{T-t}(o,x)$, where for $\epsilon > 0$, u_ϵ is the function defined on $[0,1]\times(0,\infty)\times\mathbb{R}$ by

$$u_\epsilon(t,x) = \mathbb{P}_x[\tau^\epsilon \le 1],$$

and $X_{t,x}^\epsilon$ is the solution to

$$dX_s^\epsilon = \epsilon b(X_s^\epsilon) ds + \sqrt{\epsilon}\sigma(X_s^\epsilon) dW_s, \quad X_t^\epsilon = x.$$

and

$$\tau_{t,x}^\epsilon = \inf\left\{s \ge t : X_s^{\epsilon,t,x} \notin D\right\}.$$

From the large deviation result that we have already proven in section 8.2, the problem of exit from a domain using viscosity solution, we have

$$\lim_{t\to T} -(T-t)\ln u(t,x) = V_0(t,x),$$

where

$$V_0(t,x) = \inf_{x(\cdot)\in\mathcal{A}(t,x)} \int_t^1 \frac{1}{2}\dot{x}(u)' a(x(u))\dot{x}(u)\, du \quad (t,x) \in [0,1)\times D,$$

where $\sigma(x)$ is the diffusion matrix of $X = (S,Y)$, $a(x) = (\sigma\sigma'(x))^{-1}$ and

$$\mathcal{A}(t,x) = \{x(\cdot) \in C[(0,1)] : x(t) = x, \quad \text{and} \quad \tau(x) \le 1\}.$$

There is another interpretation of the positive function V_0 in terms of Riemann distance on $\mathbb{R}^d$ associated in the the metric $a(x) = (\sigma\sigma')^{-1}$. One can prove, see [10], that $L_0(x) = \sqrt{2V_0(0,x)}$ is the unique viscosity solution of the eikonal equation

$$(D_x L_0)' \sigma'(x) D_x L_0 = 1, \quad x \in D$$

$$L_0(x) = \inf_{z\in\partial D} L_0(x,z), \quad x \in \partial D$$

and it may be represented as

$$L_0(x) = \inf_{z \in \partial D} L_0(x,z), \quad x \in D$$

where

$$L_0(x,z) = \inf_{x(\cdot) \in A(x,z)} \int_0^1 \sqrt{\dot{x}(u)' a(x(u)) \dot{x}(u)} du,$$

and

$$A(x,z) = \{x(\cdot) \in C[(0,1)] : x(0) = x \quad \text{and} \quad x(1) = z\}.$$

Hence the function L_0 can be computed either by numerical resolution of the eikonal equation or by using the above represantation. $L_0(x)$ is interpreted as the minimal length of the path $x(\cdot)$ allowing to reach the boundary ∂D from x.
From the above large deviations result, the viscosity solution of the eikonal equation and the equation for the optimal theoritical ϕ^*, we use a change of probability measure with

$$\phi(t,x) = \frac{L_0(x)}{T-t} \sigma'(x) D_x L_0(x).$$

Such a process ϕ may also appear interesting to use in more general framework than this model. One can use it for computing any option whose excersice domain looks similar to this one. We also expect the variance reduction is more significant as the excersice probability is low, i.e. for deep out-of-the money options.

Remark 8.1. *One can estimate ϕ with a method due to [12], which in contrast with the above approach, does not require the knowledge of the option price and restricts to deterministics change of drifts. The change of drift is selected through Varadhan-Laplace principle and is shown to be optimal in an asymptotic sense.*

Bibliography

[1] AVELLADA M., BOYER-OLSON D., BUSCA J., FRIZ P., *Application of Large Deviations Methods to the Pricing Index Options in Finance* C.C.Acad. Sci. Paris, Ser. I 336, 263-266, 2003.

[2] BAYER C., *Computational Finance*, Lectures notes, Berlin, 2010.

[3] DEUSCHEL J.D and STROOK D.W., *Large Deviations*, Academic Press, London, 1989.

[4] DEMBO A. and ZEITOUNI O., *Large Deviations Techniques and Applications* (2nd edition), Springer, New York, 1998.

[5] DUDLEY R.M., *Real Analysis and Probability* Cambridge University Press, 1989.

[6] EVANS L.C. *Partial Differential Equations*, American Mathematical society, 1998.

[7] FREIDLIN M.I. and WENTZELL A.D., *Random Pertubations of Dynamical Systems* (2nd edition), Springer, New York, 1998.

[8] FLEMING W. H. and SONER M., *Controlled Markov Processes and Viscosity Solutions*, (2nd edition), Springer, New York, 1992.

[9] FENG J., KURTZ T., *Large Deviations for Stochastic Processes*, 2005.

[10] FOURNIÉ E. , LARSY J.M., TOUZI N., *Monte Carlo Methods for Stochastic Volatility Methods in Finance* Numerical Methods in Finance, L.C.G. Rogers et D.Talay, Cambridge, University press, 1997.

[11] GLASSERMANN P., *Monte-Carlo methods in financial engineering*, Springer, New York, 2003.

[12] GLASSERMAN P, HEIDELBERGER P.,SHAHABUDDIN P., *Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options*, Mathematical Finance, Vol. 9, No.2, 117-152, 1999.

[13] HOLLANDER F., *Large Deviations* American Mathematical Society, 1999.

[14] KALLENBERG O., *Foundations of Modern Probability*, Springer, 1997.

[15] KELLER-RESSEL M., *Moment Explotions and Long term Behavior of Affine Stochastic Volatility Models*, Mathematical Finance, 2009.

[16] PHAM H., *Some Applications of Large Deviations in Finance and Insurance*, Paris-Princeton Lecture Notes on Mathematical Finance, 1919, 191-244, 2007.

[17] REVUZ D., YOR M., *Continuous Martingales and Brownian Motion* (3rd edition), Berlin, Springer, 1999.

[18] STAMATAKIS M., *Large Deviations*, Lectures Notes, University of Crete, 2010

[19] VARADHAN S.R.S, *Large Deviations and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1984.