An Introduction to Large Deviations

Vitalii Konarovskyi*
konarovskiy@gmail.com

September, 2019

Abstract

The mini-course is oriented on master and PhD students as the first look at the theory of large deviations.

Course webpage: http://www.math.uni-leipzig.de/%7Ekonarovskyi/teaching/2019/LDP/LDP_2019.html

1 Lecture 1 – Introduction and some examples

1.1 Introduction

We start from the considering of a coin-tossing experiment. Let us assume that we toss a fair coin. The law of large numbers says us that the frequency of occurrence of "heads" becomes close to $\frac{1}{2}$ as the number of trials increases to infinity. In other words, if $X_1, X_2, \dots$ are independent random variables taking values 0 and 1 with probabilities $\frac{1}{2}$, i.e. $\mathbb{P}\{X_k = 0\} = \mathbb{P}\{X_k = 1\} = \frac{1}{2}$, then we know that for the empirical mean $\frac{1}{n}S_n = \frac{X_1+\dots+X_n}{n}$

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \frac{1}{2}\right| > \varepsilon\right\} \to 0^{1}, \quad n \to \infty,$$

or more strongly

$$\frac{1}{n}S_n \to \frac{1}{2} \ \text{ a.s.}^{2}, \quad n \to \infty.$$

We are going stop more precisely on the probabilities $\mathbb{P}\left\{\left|\frac{1}{n}S_n - \frac{1}{2}\right| > \varepsilon\right\}$. We see that this events becomes more unlikely for large n and their probabilities decay to 0. During the course, we will work with such kind of unlike events and will try to understand the rate of their decay to zero. The knowledge of decay of probabilities of such unlike events has many applications in insurance, information theory, statistical mechanics etc. The aim of the course is to give an introduction to one of the key technique of the modern probability which is called the **large deviation theory**.

Before to investigate the rate of decay of the probabilities $\mathbb{P}\left\{\left|\frac{1}{n}S_n - \frac{1}{2}\right| > \varepsilon\right\}$, we consider an example of other random variable where computations are much more simpler.

*It is a grate pleasure for me to thank Prof. Huilai Li for his kind invitation to Jilin University and for the opportunity to teach this course. I am also grateful Prof. Andrey Dorogovtsev for his useful comments which allow to improve the course. I would finally thank my friends Suli Liu and Jia-nan Zhu for their continuous help during my stay in China.

[1]according to the *weak law of large numbers*

[2]according to the *strong law of large numbers*

Example 1.1. Let $\xi_1, \xi_2, \dots$ be independent normal distributed random variables with mean $\mu = 0$ and variance $\sigma = 1$ (shortly $\xi_k \sim N(0,1)$). Then the random variable S_n has the normal distribution with mean 0 and variance n. This implies $\frac{1}{\sqrt{n}} S_n \sim N(0,1)$.

Now we can consider for $x > 0$

$$\mathbb{P}\left\{\frac{1}{n} S_n \geq x\right\} = \mathbb{P}\left\{\frac{1}{\sqrt{n}} S_n \geq x\sqrt{n}\right\} = \frac{1}{\sqrt{2\pi}} \int_{x\sqrt{n}}^{+\infty} e^{-\frac{y^2}{2}} dy \sim \frac{1}{\sqrt{2\pi} x \sqrt{n}} e^{-\frac{nx^2}{2}}, \quad n \to +\infty,$$

by Exercise 1.1 below. Thus, we have for $x > 0$

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n} S_n \geq x\right\} &= \lim_{n\to\infty} \frac{1}{n} \ln \frac{1}{\sqrt{2\pi} x \sqrt{n}} e^{-\frac{nx^2}{2}} \\ &= -\lim_{n\to\infty} \frac{1}{n} \ln \sqrt{2\pi} x \sqrt{n} - \lim_{n\to\infty} \frac{x^2}{2} = -\frac{x^2}{2} \end{aligned}$$

due to Exercise 1.2 3).

Remark 1.1. By symmetry, one can show that

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n} S_n \leq x\right\} = -\frac{x^2}{2}$$

for all $x < 0$. Indeed,

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n} S_n \leq x\right\} = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{-\frac{1}{n} S_n \geq -x\right\} = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n} S_n \geq -x\right\} = -\frac{(-x)^2}{2}$$

because $-\xi_k \sim N(0,1)$ for $k \geq 1$.

Exercise 1.1. Show that

$$\int_x^{+\infty} e^{-\frac{y^2}{2}} dy \sim \frac{1}{x} e^{-\frac{x^2}{2}}, \quad x \to +\infty.$$

Exercise 1.2. Let $(a_n)_{n\geq1}$ and $(b_n)_{n\geq1}$ be two sequences of positive real numbers. We say that they are **logarithmically equivalent** and write $a_n \simeq b_n$ if

$$\lim_{n\to\infty} \frac{1}{n} (\ln a_n - \ln b_n) = 0.$$

1. Show that $a_n \simeq b_n$ iff $b_n = a_n e^{o(n)}$.
2. Show that $a_n \sim b_n$ implies $a_n \simeq b_n$ and that the inverse implication is not correct.
3. Show that $a_n + b_n \simeq \max\{a_n, b_n\}$.

Exercise 1.3. Let $\xi_1, \xi_2, \dots$ be independent normal distributed random variables with mean μ and variance σ^2. Let also $S_n = \xi_1 + \dots + \xi_n$. Compute $\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n} S_n \geq x\right\}$ for $x > \mu$.

1.2 Coin-tossing

In this section, we come back to the coin-tossing experiment and compute the decay of the probability $\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\}$. Let, us before, $X_1, X_2, \ldots$ be independent random variables taking values 0 and 1 with probabilities $\frac{1}{2}$ and $S_n = X_1 + \cdots + X_n$ denote their partial sum. We will show that

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = -I(x) \tag{1}$$

for all $x \geq \frac{1}{2}$, where I is some function of x.

We first note that for $x > 1$

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = -\infty.$$ [3]

Next, for $x \in \left[\frac{1}{2}, 1\right]$ we observe that

$$\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = \mathbb{P}\left\{S_n \geq xn\right\} = \sum_{k \geq xn} \mathbb{P}\left\{S_n = k\right\} = \frac{1}{2^n} \sum_{k \geq xn} C_n^k,$$

where $C_n^k = \frac{n!}{k!(n-k)!}$. Then we can estimate

$$\frac{1}{2^n} \max_{k \geq xn} C_n^k \leq \mathbb{P}\left\{S_n \geq xn\right\} \leq \frac{n+1}{2^n} \max_{k \geq xn} C_n^k. \tag{2}$$

Note that the maximum is attained at $k = \lfloor xn \rfloor$, the smallest integer $\geq xn$, because $x \geq \frac{1}{2}$. We denote $l := \lfloor xn \rfloor$. Using Stirling's formula

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + O\left(\frac{1}{n}\right)\right),$$

we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \ln \max_{k \geq xn} C_n^k &= \lim_{n\to\infty} \frac{1}{n} \ln C_n^l = \lim_{n\to\infty} \frac{1}{n} \left(\ln n! - \ln l! - \ln(n-l)!\right) \\
&= \lim_{n\to\infty} \left(\ln n - 1 - \frac{l}{n}\ln l + \frac{l}{n} - \frac{n-l}{n} \ln(n-l) + \frac{n-l}{n}\right) \\
&= \lim_{n\to\infty} \left(\frac{l}{n}\ln n + \frac{n-l}{n}\ln n - \frac{l}{n}\ln l - \frac{n-l}{n}\ln(n-l)\right) \\
&= \lim_{n\to\infty} \left(-\frac{l}{n}\ln\frac{l}{n} - \frac{n-l}{n}\ln\frac{n-l}{n}\right) = -x\ln x - (1-x)\ln(1-x),
\end{aligned}$$

because $\frac{l}{n} = \frac{\lfloor xn \rfloor}{n} \to x$ as $n \to +\infty$. This together with estimate (2) implies

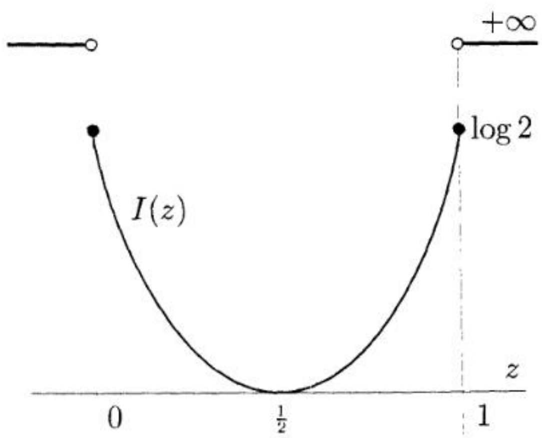
$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = -\ln 2 - x\ln x - (1-x)\ln(1-x)$$

for all $x \in \left[\frac{1}{2}, 1\right]$.

So, we can take

$$I(x) = \begin{cases} \ln 2 + x\ln x + (1-x)\ln(1-x) & \text{if } x \in [0,1], \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

[3] we always assume that $\ln 0 = -\infty$



[4]

Remark 1.2. Using the symmetry, we have that

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \leq x\right\} = -I(x) \tag{4}$$

for all $x \leq \frac{1}{2}$. Indeed,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \leq x\right\} &= \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{n}{n} - \frac{1}{n}S_n \geq 1 - x\right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{(1-X_1)+\cdots+(1-X_n)}{n} \geq 1 - x\right\} = -I(1-x) = -I(x)\end{aligned}$$

because X_k and $1 - X_k$ have the same distribution.

Theorem 1.1. *Let $\xi_1, \xi_2, \ldots$ be independent Bernoulli distributed random variables with parameter p for some $p \in (0,1)$, that is, $\mathbb{P}\{\xi_k = 1\} = p$ and $\mathbb{P}\{\xi_k = 0\} = 1 - p$ for all $k \geq 1$. Let also $S_n = \xi_1 + \cdots + \xi_n$. Then for all $x \geq p$*

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = -I(x),$$

where

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

Exercise 1.4. Prove Theorem 1.1.

Exercise 1.5. Using (1) and (4) show that

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \varepsilon\right\} < \infty,$$

for all $\varepsilon > 0$. Conclude that $\frac{S_n}{n} \to \frac{1}{2}$ a.s. as $n \to \infty$ *(strong low of large numbers).*

(*Hint:* Use the Borel-Cantelly lemma to show the convergence with probability 1)

[4] The picture was taken from [dH00]

2 Lecture 2 – Cramer's theorem

2.1 Comulant generating function

The aim of this lecture is to obtain an analog of Theorem 1.1 for any sequeness of independent identically distributed random variables. In order to understand the form of the rate function I, we will make the following computations, trying to obtain the upper bound for $\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\}$.

Let $\xi_1, \xi_2, \dots$ be independent identically distributed random variables with mean $\mu \in \mathbb{R}$. Let also $S_n = \xi_1 + \dots + \xi_n$. We fix $x > \mu$ and $\lambda \geq 0$ and use Chebyshev's inequality in order to estimate the following probability

$$\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = \mathbb{P}\left\{S_n \geq xn\right\} = \mathbb{P}\left\{e^{\lambda S_n} \geq e^{\lambda xn}\right\} \leq \frac{1}{e^{\lambda xn}}\mathbb{E}\, e^{\lambda S_n} = \frac{1}{e^{\lambda xn}}\prod_{k=1}^{n}\mathbb{E}\, e^{\lambda \xi_k} = \frac{1}{e^{\lambda xn}}\left(\mathbb{E}\, e^{\lambda\xi_1}\right)^n.$$

Thus, we have

$$\varlimsup_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left\{S_n \geq xn\right\} \leq \varlimsup_{n\to\infty}\frac{1}{n}\ln e^{-\lambda xn} + \varlimsup_{n\to\infty}\frac{1}{n}\ln\left(\mathbb{E}\, e^{\lambda\xi_1}\right)^n = -\lambda x + \varphi(\lambda),$$

where $\varphi(\lambda) := \ln\mathbb{E}\, e^{\lambda\xi_1}$. Therefore, taking infimum over all $\lambda \geq 0$, we obtain

$$\varlimsup_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left\{S_n \geq xn\right\} \leq \inf_{\lambda\geq 0}\left\{-\lambda x + \varphi(x)\right\} = -\sup_{\lambda\geq 0}\left\{\lambda x - \varphi(\lambda)\right\} \overset{5}{=} -\sup_{\lambda\in\mathbb{R}}\left\{\lambda x - \varphi(\lambda)\right\}.$$

Later we will see that the function $\sup_{\lambda\in\mathbb{R}}\left\{\lambda x - \varphi(\lambda)\right\}$ plays an important role, namely, it is exactly the rate function I.

Definition 2.1. Let ξ be a random variable on $\mathbb{R}$. The function

$$\varphi(\lambda) := \ln\mathbb{E}\, e^{\lambda\xi}, \quad \lambda \in \mathbb{R},$$

where the infinite values are allowed, is called the **logarithmic moment generating function** or **comulant generating function** associated with ξ.

Example 2.1. We compute the comulant generating function associated with Bernoulli distributed random variables ξ with parameter $p = \frac{1}{2}$. So, since $\mathbb{P}\left\{\xi = 1\right\} = \mathbb{P}\left\{\xi = 0\right\} = \frac{1}{2}$, we obtain

$$\varphi(\lambda) = \ln\mathbb{E}\, e^{\lambda\xi} = \ln\left(e^{\lambda\cdot 1}\frac{1}{2} + e^{\lambda\cdot 0}\frac{1}{2}\right) = -\ln 2 + \ln\left(e^{\lambda} + 1\right), \quad \lambda \in \mathbb{R}.$$

Example 2.2. In this example, we will compute the comulant generating function associated with exponentially distributed random variable ξ with rate γ. We recall that the density of ξ is given by the following formula

$$p_\xi(x) = \begin{cases}\gamma e^{-\gamma x} & \text{if } x \geq 0,\\ 0 & \text{if } x < 0.\end{cases}$$

[5] This equality immediately follows from the fact that the expression in the curly bracket is negative for $\lambda < 0$ (see Exercise 2.2).

So,

$$\varphi(\lambda) = \ln \int_0^\infty e^{\lambda x} \gamma e^{\gamma x} dx = \ln \int_0^\infty \gamma e^{-(\gamma-\lambda)x} dx = \ln \left(-\frac{\gamma}{\gamma-\lambda} e^{-(\gamma-\lambda)x} \Big|_0^\infty \right) = \ln \frac{\gamma}{\gamma-\lambda}$$

if $\lambda < \gamma$. For $\lambda \geq \gamma$ trivially $\varphi(\lambda) = +\infty$. Thus,

$$\varphi(\lambda) = \begin{cases} \ln\gamma - \ln(\gamma-\lambda) & \text{if } \lambda < \gamma, \\ +\infty & \text{if } \lambda \geq \gamma. \end{cases}$$

Exercise 2.1. Show that the function φ is convex[6].

Solution. In order to show the convexity of φ, we will use Hölder's inequality.[7] We take $\lambda_1, \lambda_2 \in \mathbb{R}$ and $t \in (0,1)$. Then for $p = \frac{1}{t}$ and $q = \frac{1}{1-t}$

$$\begin{aligned} \varphi(t\lambda_1 + (1-t)\lambda_2) &= \ln \mathbb{E} \left[e^{t\lambda_1 \xi} e^{(1-t)\lambda_2 \xi} \right] \leq \ln \left(\left[\mathbb{E}\, e^{\lambda_1 \xi} \right]^t \left[\mathbb{E}\, e^{\lambda_2 \xi} \right]^{1-t} \right) \\ &= t \ln \mathbb{E}\, e^{\lambda_1 \xi} + (1-t) \ln \mathbb{E}\, e^{\lambda_2 \xi} = t\varphi(\lambda_1) + (1-t)\varphi(\lambda_2). \end{aligned}$$

Exercise 2.2. Assume that a random variable ξ has a finite first moment $\mathbb{E}\,\xi = \mu$ and let φ be the comulant generating function associated with ξ. Show that for every $x > \mu$ and all $\lambda < 0$

$$\lambda x - \varphi(\lambda) \leq 0.$$

(*Hint:* Use Jensen's inequality.[8])

Exercise 2.3. Let φ be a comulant generating function associated with ξ. Show that the function φ is differentiable in the interior of the domain $\mathcal{D}_\varphi := \{x \in \mathbb{R} : \ \varphi(x) < \infty\}$. In particular, show that $\varphi'(0) = \mathbb{E}\,\xi$ if $0 \in \mathcal{D}_\varphi^\circ$.

(*Hint:* To show the differentiability of φ, it is enough to show that $\mathbb{E}\, e^{\lambda\xi}$, $\lambda \in \mathbb{R}$, is differentiable. For the differentiability of the latter function, use the definition of the limit, the dominated convergence theorem[9] and the fact that the function $\frac{e^{\varepsilon a}-1}{\varepsilon} = \int_0^a e^{\varepsilon x} dx$ increases in $\varepsilon > 0$ for each $a \geq 0$.)

2.2 Fenchel-Legendre transform and Cramer's theorem

In this section, we discuss the Fenchel-Legendre transform of a convex function that apeared in the previous secton. Let $f : \mathbb{R} \to (-\infty, +\infty]$ be a convex function.

Definition 2.2. The function

$$f^*(y) := \sup_{x \in \mathbb{R}} \{yx - f(x)\}$$

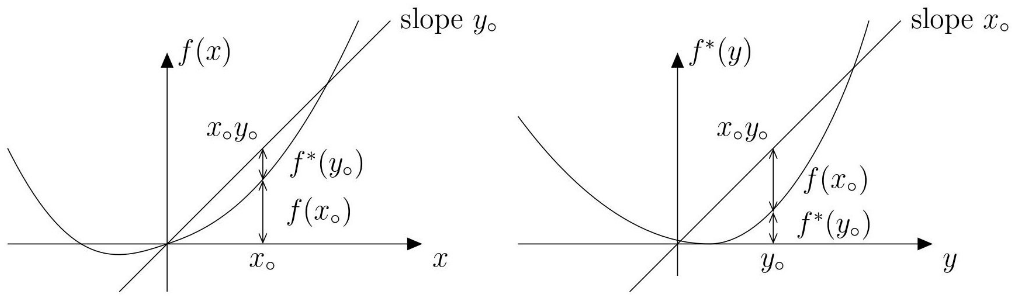
is called the **Fenchel-Legendre transform** of f.

[6] A function $f : \mathbb{R} \to (-\infty, +\infty]$ is convex if for all $x_1, x_2 \in \mathbb{R}$ and $t \in (0,1)$, $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$

[7] Let $p, q \in (1, +\infty)$, $\frac{1}{p} + \frac{1}{q} = 1$ and ξ, η be random variables. Then $\mathbb{E}\,(\xi\eta) \leq (\mathbb{E}\,\xi^p)^{\frac{1}{p}} \, (E\eta^q)^{\frac{1}{q}}$.

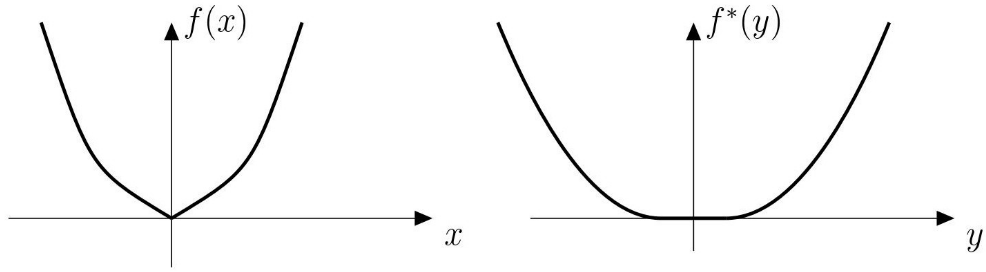
[8] For any random variable ξ with a finite first moment and a convex function $f : \mathbb{R} \to \mathbb{R}$, one has $f(\mathbb{E}\,\xi) \leq \mathbb{E}\, f(\xi)$.

[9] For the dominated convergence theorem see [Kal02, Theorem 1.21]



[10]

Fenchel-Legendre transformation: definition



[11]

Fenchel-Legendre transformation of a function f

Exercise 2.4. Show that the Fenchel-Legendre transform of a convex function f is also convex.

(*Hint:* Show first that the supremum of convex functions is a convex function. Then note that the function $\lambda x - \varphi(\lambda)$ is convex in the variable x)

Exercise 2.5. Compute the Fenchel-Legendre transform of the comulant generating function associated with the Bernoulli distribution with $p = \frac{1}{2}$.

Solution. Let ξ be a Bernoulli distributed random variable with parameter $p = \frac{1}{2}$, i.e. $\mathbb{P}\{\xi = 1\} = \mathbb{P}\{\xi = 0\} = \frac{1}{2}$. We first write its comulant generating function:

$$\varphi(\lambda) = -\ln 2 + \ln\left(1 + e^{\lambda}\right)$$

(see Example 2.1). In order to compute the Fenchel-Legendre transform of φ, we have to find the supremum of the function

$$g(\lambda) := \lambda x - \varphi(\lambda) = \lambda x + \ln 2 - \ln\left(1 + e^{\lambda}\right), \quad \lambda \in \mathbb{R},$$

for every $x \in \mathbb{R}$. So, we fix $x \in \mathbb{R}$ and find

$$g'(\lambda) = x - \frac{e^{\lambda}}{1 + e^{\lambda}} = 0.$$

[10] It turns out that the Fenchel-Legendre transform of f^* coincides with f (see e.g. [Swa12, Proposition 2.3]). The picture was taken from [Swa12].

[11] The picture was taken from [Swa12].

Hence
$$\lambda = \ln \frac{x}{1-x} \quad \text{if} \quad x \in (0,1)$$
is a local maximum. Due to the convexity of φ, this point is also the global maximum. Consequently,
$$\begin{aligned}\varphi^*(x) &= \sup_{\lambda \in \mathbb{R}} \{\lambda x - \varphi(\lambda)\} = x \ln \frac{x}{1-x} + \ln 2 - \ln\left(1 + \frac{x}{1-x}\right) \\ &= x\ln x - x\ln(1-x) + \ln 2 + \ln(1-x) = \ln 2 + x\ln x + (1-x)\ln(1-x), \quad x \in (0,1).\end{aligned}$$
If $x < 0$ or $x > 1$ then $\varphi^*(x) = +\infty$. For $x = 0$ and $x = 1$ one can check that $\varphi^*(x) = \ln 2$.

Exercise 2.6. Show that the function φ^* from the previous exercise equals $+\infty$ for $x \in (-\infty, 0) \cup (1, +\infty)$ and $\ln 2$ for $x \in \{0, 1\}$.

Compering the Fenchel-Legendre transformation φ^* of the comulant generating function associated with the Bernoulli distribution ξ and the rate function I given by (3), we can see that thay coinside.

Exercise 2.7. a) Show that the Fenchel-Legendre transform of the comulant generating function associated with $N(0,1)$ coincides with $\frac{x^2}{2}$.

b) Show that the Fenchel-Legendre transform of the comulant generating function associated with Bernoulli distribution with paramiter $p \in (0,1)$ coincides with the function I from Theorem 1.1.

c) Find the Fenchel-Legendre transform of the comulant generating function associated with exponential distribution.

Exercise 2.8. Suppose that φ^* is the Fenchel-Legendre transform of the cumulant generating function of a random variable ξ with $\mathbb{E}\,\xi = \mu$. Show that

(i) $\varphi^*(x) \geq 0$ for all $x \in \mathbb{R}$. (*Hint:* Use the fact that $\varphi(0) = 0$)

(ii) $\varphi^*(\mu) = 0$. (*Hint:* Use (i) and Jensen's inequality to show that $\varphi^*(\mu) \leq 0$)

(iii) φ^* increases on $[\mu, \infty)$ and decreases on $(-\infty, \mu]$. (*Hint:* Use the convexity of φ^* (see Exercise 2.4) and (ii))

(iv) $\varphi^*(x) > 0$ for all $x \neq \mu$. (*Hint:* Use the fact that φ is convex and differentiable at 0 with $\varphi'(0) = \mu$ (see Exercise 2.3))

Now we are ready to formulate a statement on the decay of the probability $\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\}$ in the general case.

Theorem 2.1 (Cramer's theorem). *Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables with mean $\mu \in \mathbb{R}$ and comulant generating function φ. Let also $S_n = \xi_1 + \cdots + \xi_n$. Then, for every $x \geq \mu$*
$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\} = -\varphi^*(x),$$
and for every $x \leq \mu$
$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \leq x\right\} = -\varphi^*(x),$$
where φ^* is the Fenchel-Legendre transform of the comulant generating function φ associated with ξ_1.

The general proof of Cramer's theorem can be found in [Kal02, Theorem 27.3] or [DZ98, Section 2.1.2]. Proofs in some particular cases are given in [dH00, Section I.3], [Swa12, Section 2.2], [MÖ08, Section 1]

3 Lecture 3 – Large deviation principle for Gaussian vectors

3.1 Large deviations for Gaussian vectors in $\mathbb{R}^d$

We recall that, in the last lecture, we have investigated the decay of the probability $\mathbb{P}\left\{\frac{1}{n}S_n \geq x\right\}$, where $S_n = \xi_1 + \dots + \xi_n$ and ξ_k, $k \in \mathbb{N}$, ware independent identically distributed random variables in $\mathbb{R}$. We start this lecture from some example of random variables in higher dimension and investigate the decay of similar probabilities. This will lead us to the general concept of large deviation principle in the next section. We note that the case $\xi_k \sim N(0,1)$ was very easy for computations (see Example 1.1). So similarly, we take independent $\mathbb{R}^d$-valued random element (or random vector) $\eta_k = (\eta_k^{(1)}, \dots, \eta_k^{(d)})$, $k \geq 1$, with standard Gaussian distributions[12]. We will study the decay of the probability

$$\mathbb{P}\left\{\frac{1}{n}S_n \in A\right\} = \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} = \mathbb{P}\left\{\eta \in \sqrt{n}A\right\},$$

where A is a subset of $\mathbb{R}^d$, $S_n = \eta_1 + \dots + \eta_n$, and η has a standard Gaussian distribution.

The upper bound. We recall that the density of $\frac{1}{\sqrt{n}}\eta$ is given by the formula

$$p_{\frac{\eta}{\sqrt{n}}}(x) = \frac{(\sqrt{n})^d}{(\sqrt{2\pi})^d} e^{-\frac{n\|x\|^2}{2}}, \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where $\|x\|^2 = x_1^2 + \dots + x_d^2$. Now, we can estimate

$$\begin{aligned}
\varlimsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} &= \varlimsup_{n\to\infty} \frac{1}{n} \ln \int_A p_{\frac{\eta}{\sqrt{n}}}(x)dx = \varlimsup_{n\to\infty} \frac{1}{n} \ln \frac{(\sqrt{n})^d}{(\sqrt{2\pi})^d} \\
&+ \varlimsup_{n\to\infty} \frac{1}{n} \ln \int_A e^{-\frac{n\|x\|^2}{2}} dx \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln \int_A \left(\sup_{y\in A} e^{-\frac{n\|y\|^2}{2}}\right) dx \\
&= \varlimsup_{n\to\infty} \frac{1}{n} \ln \left(e^{-n \inf_{x\in A} \frac{\|x\|^2}{2}} |A|\right) \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln |A| + \varlimsup_{n\to\infty} \frac{1}{n} \ln e^{-n \inf_{x\in A} I(x)} \\
&= -\inf_{x\in A} I(x) \leq -\inf_{x\in \bar{A}} I(x),
\end{aligned}$$

where $I(x) := \frac{\|x\|^2}{2}$, $\bar{A}$ is the closure of A and $|A|$ denotes the Lebesgue measure of A.

The lower bound. In order to obtain the lower bound, we assume that the interior A° of A is non-empty and fix $x_0 \in A^\circ$. Let $B_r(x_0) = \left\{x \in \mathbb{R}^d : \|x - x_0\| < r\right\}$ denotes the ball in $\mathbb{R}^d$ with center x_0 and radius r. We estimate

$$\begin{aligned}
\varliminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} &= \varliminf_{n\to\infty} \frac{1}{n} \ln \int_A p_{\frac{\eta}{\sqrt{n}}}(x)dx \geq \varliminf_{n\to\infty} \frac{1}{n} \ln \frac{(\sqrt{n})^d}{(\sqrt{2\pi n})^d} \\
&+ \varliminf_{n\to\infty} \frac{1}{n} \ln \int_{B_r(x_0)} e^{-\frac{n\|x\|^2}{2}} dx \geq \varliminf_{n\to\infty} \frac{1}{n} \ln \int_{B_r(x_0)} e^{-\frac{n(\|x_0\| + \|x - x_0\|)^2}{2}} dx \\
&\geq \varliminf_{n\to\infty} \frac{1}{n} \ln \int_{B_r(x_0)} e^{-\frac{n(\|x_0\|+r)^2}{2}} dx = \varliminf_{n\to\infty} \frac{1}{n} \ln e^{-\frac{n(\|x_0\|+r)^2}{2}} |B_r(x_0)| \\
&= -\frac{(\|x_0\| + r)^2}{2}.
\end{aligned}$$

[12] $\eta_k^{(i)} \sim N(0,1)$, $i = 1, \dots, d$, and are independent

Making $r \to 0+$, we have

$$\varliminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} \geq -\frac{\|x_0\|^2}{2}.$$

Now, maximizing the right hand side over all points x_0 from the interior A°, we obtain

$$\varliminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} \geq \sup_{x_0\in A^\circ}\left(-\frac{\|x_0\|^2}{2}\right) = -\inf_{x\in A^\circ} I(x).$$

Thus, combining the lower and upper bounds, we have prove that for any Borel measurable set A

$$-\inf_{x\in A^\circ} I(x) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{\sqrt{n}}\eta \in A\right\} \leq -\inf_{x\in \bar{A}} I(x). \tag{5}$$

3.2 Definition of large deviation principle

Let $(\xi_\varepsilon)_{\varepsilon>0}$ be a family of random elements on a metric space E and I be a function from E to $[0,\infty]$.

Definition 3.1. We say that the family $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the **large deviation principle (LDP) in** E **with rate finction** I if for any Borel set $A \subset E$ we have

$$-\inf_{x\in A^\circ} I(x) \leq \varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A\right\} \leq \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A\right\} \leq -\inf_{x\in \bar{A}} I(x). \tag{6}$$

We remark that in the case of a countable family of random elements $(\xi_n)_{n\geq 1}$, the large deviation principle corresponds to the statement

$$-\inf_{x\in A^\circ} I(x) \leq \varliminf_{n\to\infty} a_n \ln \mathbb{P}\left\{\xi_n \in A\right\} \leq \varlimsup_{n\to\infty} a_n \ln \mathbb{P}\left\{\xi_n \in A\right\} \leq -\inf_{x\in \bar{A}} I(x)$$

for some sequence $a_n \to 0$. In fact, we have proved in the previous section that the family $\left(\frac{1}{n}S_n\right)_{n\geq 1}$ or $\left(\frac{\eta}{\sqrt{n}}\right)_{n\geq 1}$ satisfies the large deviation principle in $\mathbb{R}^d$ with rate function $I(x) = \frac{\|x\|^2}{2}$, $x \in \mathbb{R}^d$ and $a_n = \frac{1}{n}$ (see inequality (5)).

Lemma 3.1. *A family $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the large deviation principle in E with rate function I iff*

$$\varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F\right\} \leq -\inf_{x\in F} I(x) \tag{7}$$

for every closed set $F \subset E$, and

$$\varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in G\right\} \geq -\inf_{x\in G} I(x) \tag{8}$$

for every open set $G \subset E$.

Proof. We first remark that inequalities (7) and (8) immediately follow from the definition of LDP and the fact that $F = \bar{F}$ and $G = G^\circ$.

To prove (6), we fix a Borel measurable set $A \subset E$ and estimate

$$\begin{aligned}-\inf_{x\in A^\circ} I(x) &\overset{(8)}{\leq} \varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A^\circ\right\} \leq \varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A\right\} \\ &\leq \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A\right\} \leq \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in \bar{A}\right\} \overset{(7)}{\leq} -\inf_{x\in \bar{A}} I(x).\end{aligned}$$

□

Remark 3.1. We note that repeating the proof from the previous section, one can show that the family $(\sqrt{\varepsilon}\xi)_{\varepsilon>0}$ satisfies the LDP in $\mathbb{R}^d$ with rate function $I(x) = \frac{1}{2}\|x\|^2$, where ξ is a standard Gaussian random vector in $\mathbb{R}^d$.

Exercise 3.1. Let there exist a subset E_0 of the metric space E such that

1) for each $x \in E_0$
$$\lim_{r\to 0} \varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in B_r(x)\right\} \geq -I(x),$$
where $B_r(x)$ denotes the ball with center x and radius r;

2) for each x satisfying $I(x) < \infty$, there exists a sequence $x_n \in E_0$, $n \geq 1$, such that $x_n \to x$, $n \to \infty$, and $I(x_n) \to I(x)$, $n \to \infty$.

Show that lower bound (8) holds for any open set G.

Solution. First we note that it is enough to prove the lower bound for all open $G \subseteq E$ satisfying $\inf_{x\in G} I(x) < \infty$.

Let δ be an arbitrary positive number. Then there exists $x_0 \in G$ such that
$$I(x_0) < \inf_{x\in G} I(x) + \delta.$$
Hence, by 2) and the openness of G we can find $\tilde{x} \in G \cap E_0$ that satisfies
$$I(\tilde{x}) < I(x_0) + \delta.$$
Next, using 1) and the openness of G, there exists $r > 0$ such that $B_r(\tilde{x}) \subseteq G$ and
$$\varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\{\xi_\varepsilon \in B_r(\tilde{x})\} \geq -I(\tilde{x}) - \delta.$$
Consequently, we can now estimate
$$\begin{aligned}\varliminf_{\varepsilon\to 0} \varepsilon \ln\mathbb{P}\{\xi_\varepsilon \in G\} &\geq \varliminf_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\{\xi_\varepsilon \in B_r(\tilde{x})\} \\ &\geq -I(\tilde{x}) - \delta > -I(x_0) - 2\delta > -\inf_{\varphi\in G} I(x) - 3\delta.\end{aligned}$$
Making $\delta \to 0$, we obtain the lower bound (8).

Exercise (3.1) shows the local nature of the lower bound.

Exercise 3.2. Let $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in E with rate function I. Show that

a) if A is such that $\inf_{x\in A^\circ} I(x) = \inf_{x\in \bar{A}} I(x)$, then
$$\lim_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in A\right\} = -\inf_{x\in A} I(x);$$

b) $\inf_{x\in E} I(x) = 0$.

Exercise 3.3. Let $E = \mathbb{R}$ and $\xi \sim N(0,1)$. Show that the family $(\varepsilon\xi)_{\varepsilon>0}$ satisfies the LDP with rate function

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

Compare this claim with the result of Remark 3.1.

Exercise 3.4. Let a_n, b_n, $n \geq 1$, be positive real numbers. Show that

$$\varlimsup_{n\to\infty} \frac{1}{n} \ln(a_n + b_n) = \varlimsup_{n\to\infty} \frac{1}{n} \ln a_n \vee \varlimsup_{n\to\infty} \frac{1}{n} \ln b_n,$$

where $a \vee b$ denotes the maximum of the set $\{a, b\}$.

(*Hint:* Use the inequality $\ln(a_n + b_n) - \ln a_n \vee \ln b_n \leq \ln 2$)

Exercise 3.5. Let $\eta_1, \eta_2 \sim N(0,1)$. Let also for every $\varepsilon > 0$ a random variable ξ_ε have the distribution defined as follows

$$\mathbb{P}\left\{\xi_\varepsilon \in A\right\} = \frac{1}{2}\mathbb{P}\left\{-1 + \sqrt{\varepsilon}\eta_1 \in A\right\} + \frac{1}{2}\mathbb{P}\left\{1 + \sqrt{\varepsilon}\eta_2 \in A\right\}$$

for all Borel sets A. Show that the family $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the LDP with rate function $I(x) = \frac{1}{2}\min\left\{(x-1)^2, (x+1)^2\right\}$, $x \in \mathbb{R}$.

(*Hint:* Show first that both families $(\sqrt{\varepsilon}\eta_1)_{\varepsilon>0}$ and $(\sqrt{\varepsilon}\eta_2)_{\varepsilon>0}$ satisfy LDP and find the corresponding rate functions. Then use Exercise 3.4)

4 Lecture 4 – Lower semi-continuity and goodness of rate functions

4.1 Lower semi-continuity of rate functions

Let $(\xi_\varepsilon)_{\varepsilon>0}$ satisfy the LDP in a metric space with rate function $I : E \to [0, \infty]$. In this section, we are going to answer the question when the rate function I is unique.

Example 4.1. Let $\xi \sim N(0,1)$. We know that the family $(\xi_\varepsilon := \sqrt{\varepsilon}\xi)_{\varepsilon>0}$ satisfies the LDP in $\mathbb{R}$ with the good rate function $I(x) = \frac{x^2}{2}$.[13] We take another function

$$\tilde{I}(x) = \begin{cases} \frac{x^2}{2} & \text{if } x \neq 0, \\ +\infty & \text{if } x = 0, \end{cases}$$

and show that the family $(\xi_\varepsilon)_{\varepsilon>0}$ also satisfies the LDP with rate function $\tilde{I}$.

Indeed, if G is an open set in $\mathbb{R}$, then trivially

$$\inf_{x\in G} I(x) = \inf_{x\in G} \tilde{I}(x).$$

For a closed F which does not contain 0 or 0 is its limit point[14], we have

$$\inf_{x\in F} I(x) = \inf_{x\in F} \tilde{I}(x).$$

We have only to check the upper bound for the case if 0 is an isolated point of F. Then $F \setminus \{0\}$ is closed. So,

$$\varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F\right\} = \varlimsup_{\varepsilon\to 0} \varepsilon \ln\left(\mathbb{P}\left\{\sqrt{\varepsilon}\xi \in F \setminus \{0\}\right\}\right) \leq -\inf_{x\in F\setminus\{0\}} I(x) = -\inf_{x\in F} \tilde{I}(x).$$

This example shows that the same family of random variables can satisfy the LDP with different rate functions. In the rest of the section, we will impose some additional conditions on rate functions to provide the uniqueness.

Definition 4.1. A function $f : E \to [-\infty, +\infty]$ is called lower semi-continuous if

$$\varliminf_{n\to\infty} f(x_n) \geq f(x) \quad \text{whenever} \quad x_n \to x.$$

Lemma 4.1. *A function $f : E \to [-\infty, +\infty]$ is lower semi-continuous iff for each $\alpha \in [-\infty, +\infty]$ the* **level set** *$\{x \in E : \quad f(x) \leq \alpha\}$ is a closed subset of E.*

Exercise 4.1. Prove Lemma 4.1.

(*Hint:* To prove that every level set of f is closed, assume that it is not so. Then there exists a subsequence in a level set whose limit is outside this set. Conclude that it contradicts the inequality from Definition 4.1.

In order to prove the sufficiency, assume that the inequality of the definition is not correct. Then take α such that $\varliminf_{n\to\infty} f(x_n) < \alpha \leq f(x)$ and show that the corresponding level set is not closed)

Remark 4.1. Note that the function $\tilde{I}$ from Example 4.1 is not lower semi-continuous. Indeed, the inequality from Definition 4.1 does not hold e.g. for $x_n = \frac{1}{n}$.

[13] see Remark 3.1

[14] x is a limit point of F if there is a sequence $(x_n)_{n\geq 1}$ of elements from F such that $x_n \to x$, $n \to \infty$

Next we are going to show that one can always replace a rate function by a lower semi-continuous rate function. Moreover, it turns out that a lower semi-continuous rate function is unique. For this, we introduce a transformation produces a lower semi-continuous function f_{lsc} from an arbitrary function $f : E \to [-\infty, +\infty]$ (for more details see [RAS15, Section 2.2]).

The **lower semi-continuous regularization** of f is defined by

$$f_{\mathrm{lsc}}(x) = \sup \left\{ \inf_{y \in G} f(y) : \ G \ni x \text{ and } G \text{ is open} \right\}. \tag{9}$$

Exercise 4.2. Show that the function $\tilde{I}_{\mathrm{lsc}}$ coincides with $I(x) = \frac{x^2}{2}$, $x \in \mathbb{R}$, where $\tilde{I}$ was defined in Example 4.1.

Exercise 4.3. Let $f(x) = \mathbb{I}_{\mathbb{Q}}(x)$, $x \in \mathbb{R}$, where $\mathbb{Q}$ denotes the set of rational numbers. Find the function f_{lsc}.

Lemma 4.2. *The function f_{lsc} is lower semi-continuous and $f_{lsc}(x) \leq f(x)$ for all $x \in E$. If g is lower semi-continuous and satisfies $g(x) \leq f(x)$ for all x, then $g(x) \leq f_{lsc}(x)$ for all x. In particular, if f is lower semi-continuous, then $f = f_{lsc}$.*

The Lemma 4.2 says that the lower semi-continuous regularization f_{lsc} of f is the maximal lower semi-continuous function less or equal that f.

Proof of Lemma 4.2. The inequality $f_{\mathrm{lsc}} \leq f$ is clear. To show that f_{lsc} is lower semi-continuous we use Lemma 4.1. Let $x \in \{f_{\mathrm{lsc}} > \alpha\}$. Then there is an open set G containing x such that $\inf_G f > \alpha$. Hence by the supremum in the definition of f_{lsc}, $f_{\mathrm{lsc}}(y) \geq \inf_G f > \alpha$ for all $y \in G$. Thus G is an open neighbourhood of x contained in $\{f_{\mathrm{lsc}} > \alpha\}$. So $\{f_{\mathrm{lsc}} > \alpha\}$ is open.

To show that $g \leq f_{\mathrm{lsc}}$ one just needs to show that $g_{\mathrm{lsc}} = g$. Then

$$\begin{aligned} g(x) = g_{\mathrm{lsc}}(x) &= \sup \left\{ \inf_G g : \ x \in G \text{ and } G \text{ is open} \right\} \\ &\leq \sup \left\{ \inf_G f : \ x \in G \text{ and } G \text{ is open} \right\} = f_{\mathrm{lsc}}(x). \end{aligned}$$

We already know that $g_{\mathrm{lsc}} \leq g$. To show the other direction let α be such that $g(x) > \alpha$. Then, $G = \{g > c\}$ is an open set containing x and $\inf_G g \geq \alpha$. Thus, $g_{\mathrm{lsc}}(x) \geq \alpha$. Now increasing α to $g(x)$, we obtain the needed inequality. $\square$

Exercise 4.4. 1) Show that if $x_n \to x$, then $f_{\mathrm{lsc}}(x) \leq \underline{\lim}_{n\to\infty} f(x_n)$.

(*Hint:* Use Lemma 4.2, namely that the function f_{lsc} is lower semi-continuous and $f_{\mathrm{lsc}} \leq f$)

2) Show that for each the supremum in (9) can only be taken over all ball with center x, namely

$$f_{\mathrm{lsc}}(x) = \sup_{r>0} \inf_{y \in B_r(x)} f(y) \tag{10}$$

(*Hint:* Use the fact that any open set G containing x also contains a ball $B_r(x)$ for some $r > 0$. It will allow to prove the inequality $f_{\mathrm{lsc}}(x) \leq \sup_{r>0} \inf_{y \in B_r(x)} f(y)$. The inverse inequality just follows from the observation that supremum in the right hand side of (10) is taken over smaller family of open sets)

3) Prove that for each $x \in E$ there is a sequence $x_n \to x$ such that $f(x_n) \to f_{\mathrm{lsc}}(x)$ (the constant sequence $x_n = x$ is allowed here). This gives the alternate definition

$$f_{\mathrm{lsc}}(x) = \min\left\{ f(x), \varliminf_{y \to x} f(y) \right\}.$$

(*Hint:* Use part 2) of the exercise to construct the corresponding sequence x_n, $n \geq 1$)

Proposition 4.1. *Let $(\xi_\varepsilon)_{\varepsilon>0}$ satisfy the LDP in a metric space E with rate function I. Then it satisfies the LDP in E with the rate function I_{lsc}. Moreover, there exists only unique lower semi-continuous associated rate function.*

Proof. We first show that $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in E with the lower semi-continuous function I_{lsc}. For this we check the inequalities of Lemma 3.1. We note that the upper bound immediately follows from the inequality $I_{\mathrm{lsc}} \leq I$. For the lower bound we observe that $\inf\limits_G I_{\mathrm{lsc}} = \inf\limits_G I$ when G is open. Indeed, the inequality $\inf\limits_G I_{\mathrm{lsc}} \leq \inf\limits_G I$ follows from Lemma 4.2. The inverse inequality follows from the definition of lower semicontinuous regularization.

To prove the uniqueness, assume that (6) holds for two lower semi-continuous functions I and J, and let $I(x) < J(x)$ for some $x \in E$. By the lower semi-continuity of J, we may choose a neighbourhood G of x such that $\inf\limits_{\bar{G}} J > I(x)$, taking e.g. G as an open neighbourhood of x such that $\bar{G} \subset \left\{ y : \ J(y) > I(x) + \frac{J(x)-I(x)}{2} \right\}$. Then applying (6) to both I and J yields the contradiction

$$-I(x) \leq -\inf_G I \leq \varliminf_{\varepsilon \to 0} \varepsilon \ln \mathbb{P}\left\{ \xi_\varepsilon \in G \right\} \leq -\inf_{\bar{G}} J < -I(x).$$

□

Exercise 4.5. Assume φ^* that is the Fenchel-Legendre transform of the comulant generating function. Show that φ^* is lower semi-continuous.

(*Hint:* Show that supremum of a family of continuous functions is lower semi-continuous)

4.2 Goodness of rate functions

We remark that in many cases the rate function satisfies better properties than lower semi-continuity.

Definition 4.2. We say that a rate function $I : E \to [0, +\infty]$ is **good** if the level sets $\{x \in E : \ I(x) \leq \alpha\}$ are compact (rather than just closed) for all $\alpha \geq 0$.

Example 4.2. Show that the rate function $I(x) = \frac{\|x\|^2}{2}$, $x \in \mathbb{R}^d$, from Exercise 1.3 is good.

Remark 4.2. The rate functions from all previous examples are also good.

Now we consider another example of a good rate function which is the rate function for LDP for Brownian motion. We obtain the LDP for Brownian motion later and here we just show that the associated rate function is good.

Let $\mathrm{C}_0[0,T]$ denote a normal space of continuous functions from $[0,T]$ satisfying $f(0)=0$ endowed with the uniform norm.[15] Let $H_0^2[0,T]$ be the set of all absolutely continuous[16] functions $f \in \mathrm{C}_0[0,T]$ with $\dot{f} \in L_2[0,T]$.

Exercise 4.6. Let $f \in \mathrm{C}_0^1[0,T]$.[17] Show that f is absolutely continuous and $\dot{f}$ coincides with the classical derivative f' of f. Conclude that $f \in H_0^2[0,T]$.

Exercise 4.7. Show that the function $f(x) = 1 - |x-1|$, $x \in [0,2]$, belongs to $H_0^2[0,2]$ but is not continuously differentiable.

(*Hint:* Show that $\dot{f}(x) = \begin{cases} 1 & \text{if } x \in [0,1]\,, \\ -1 & \text{if } x \in (1,2]\,. \end{cases}$)

We consider a function from $\mathrm{C}_0[0,T]$ to $[0,+\infty]$ defined as follows

$$I(f) = \begin{cases} \frac{1}{2}\int_0^T \dot{f}^2(x)dx & \text{if } f \in H_0^2[0,T], \\ +\infty & \text{otherwise}\ . \end{cases} \tag{12}$$

Exercise 4.8. Let $I : \mathrm{C}_0[0,T] \to [0,+\infty]$ be defined by (12). Show that the set $\{f \in \mathrm{C}_0[0,T] :\ I(f) \le \alpha\}$ is equicontinuous[18] and bounded in $\mathrm{C}_0[0,T]$ for all $\alpha \ge 0$.

(*Hint:* Using Hölder's inequality, show that $|f(t)-f(s)|^2 \le |t-s| \int_0^T \dot{f}^2(x)dx$ for all $t,s \in [0,T]$ and each $f \in H_0^2[0,T]$)

Exercise 4.9. Let $I : \mathrm{C}_0[0,T] \to [0,+\infty]$ be defined by (12). Show that I is good.

(*Hint:* Show that level sets are closed and then use Example 4.8 and the Arzela-Ascoli theorem[19]. To show that level sets of I are closed in $\mathrm{C}_0[0,T]$, use Alaoglu's theorem[20])

[15] The uniform norm on $\mathrm{C}_0[0,T]$ is defined as $\|f\|_C = \max_{x\in[0,T]} |f(x)|$. The space $\mathrm{C}_0[0,T]$ endowed with this norm is a separable Banach space

[16] A continuous function $f \in \mathrm{C}[0,T]$ (not necessarily $f(0)=0$) is said to be **absolutely continuous** if there exists a function $h \in L_1[0,T]$ such that

$$f(t) = f(0) + \int_0^t h(s)ds, \quad t \in [0,T]. \tag{11}$$

Such a function h is denoted by $\dot{f}$ and is called the derivative of f.

[17] $\mathrm{C}_0^m[0,T]$ consists of all functions from $\mathrm{C}_0[0,T]$ which are m times continuously differentiable on $(0,T)$

[18] see Definition VI.3.7 [Con90]

[19] see Corollary VI.3.10 [Con90]

[20] see Theorem V.3.1 [Con90]

5 Lecture 5 – Weak large deviation principle and exponential tightness

5.1 Weak large deviation principle

Example 3.1 shows that lower bound inequality (8) is enough to show only for open balls. Unfortunately, it is not enough for upper bound (7). Later, in Proposition 5.1, we will show that upper bound (7) for closed (or open) balls will only imply the upper bound for compact sets F. To have the upper bound for any closed set we need one extra condition, called exponential tightness, which we will discuss in the next section. Let us consider the following example taken from [DZ98, P. 7] which demonstrates that upper bound for all compact sets does not imply inequality (7) for any closed set.

Example 5.1. We consider random variables $\xi_\varepsilon := \frac{1}{\varepsilon}$, $\varepsilon > 0$, in $\mathbb{R}$ and set $I(x) := +\infty$, $x \in \mathbb{R}$. Then for any compact set F in $\mathbb{R}$ (which is also bounded) we have

$$\overline{\lim_{\varepsilon\to 0}} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F\right\} = -\infty = -\inf_{x\in F} I(x)$$

because there exists $\varepsilon_0 > 0$ such that $\mathbb{P}\left\{\xi_\varepsilon \in F\right\} = 0$ for all $\varepsilon \in (0, \varepsilon_0)$. But it is easily seen that this inequality is not preserved for the closed set $F = \mathbb{R}$. Indeed,

$$\overline{\lim_{\varepsilon\to 0}} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in \mathbb{R}\right\} = \overline{\lim_{\varepsilon\to 0}} \varepsilon \ln 1 = 0 \not\leq -\infty = \inf_{x\in\mathbb{R}} I(x).$$

We also remark here that the family $(\xi_\varepsilon)_{\varepsilon>0}$ and the function I satisfy lower bound (8).

Consequently, it makes sense to introduce a relaxation of the *full* LDP, where we will require the upper bound only for compact sets.

Definition 5.1. We say that the family $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the **weak large deviation principle (weak LDP) in E with rate function I** if

$$\overline{\lim_{\varepsilon\to 0}} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F\right\} \leq -\inf_{x\in F} I(x) \tag{13}$$

for every *compact* set $F \subset E$, and

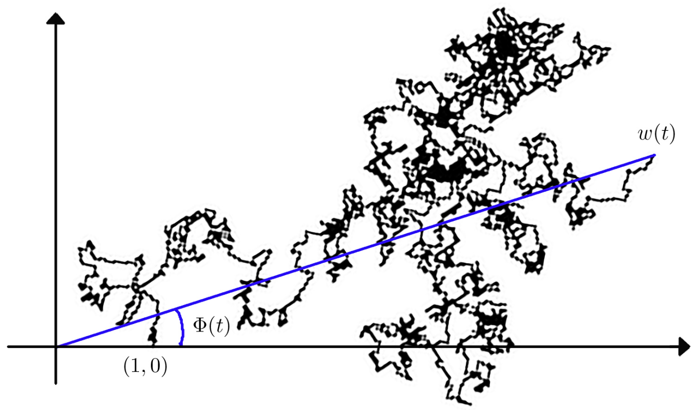
$$\underline{\lim_{\varepsilon\to 0}} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in G\right\} \geq -\inf_{x\in G} I(x) \tag{14}$$

for every open set $G \subset E$.

We remark that Example 5.1 shows that the family $\left(\xi_\varepsilon = \frac{1}{\varepsilon}\right)_{\varepsilon>0}$ satisfies the weak LDP in $\mathbb{R}$ with good rate function $I(x) = +\infty$, $x \in \mathbb{R}$, but it does not satisfy the full LDP.

Let us consider another interesting example of a family of random elements in $\mathrm{C}[0,T]$ which satisfies the weak LDP but it does not satisfies the full LDP for any rate function. This is a recent result obtained by V. Kuznetsov in [Kuz15].

Example 5.2 (Winding angle of Brownian trajectory around the origin)**.** Let $w(t) = (w_1(t), w_2(t))$, $t \in [0,T]$, be a two dimensional Brownian motion started from the point $(1,0)$. We denote for every $t \in [0,T]$ the angle between the vector $w(t)$ and the x-axis (the vector $(1,0)$) by $\Phi(t)$ and set $\Phi_\varepsilon(t) = \Phi(\varepsilon t)$, $t \in [0,T]$. It turns out that the family $(\Phi_\varepsilon)_{\varepsilon>0}$ satisfies only the weak LDP in the space of continuous functions $\mathrm{C}[0,T]$.



Winding angle of Brownian motion

In the next section, we will consider conditions on a family $(\xi_\varepsilon)_{\varepsilon>0}$ which will guarantee the implication. Now we will give a useful statement which allows to check the upper bound for the weak LDP.

Proposition 5.1. *Let $(\xi_\varepsilon)_{\varepsilon>0}$ be a family of random variables in a metric space E and let I be a function from E to $[0,+\infty]$. Then upper bound (13) follows from*

$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in B_r(x)\right\}\leq -I(x)$$

for all $x\in E$.

Proof. Let F be a compact set. We set $\alpha:=\inf\limits_{x\in F}I(x)$ and assume that $\alpha<\infty$. Remark that for every $x\in F$ $I(x)\geq\inf\limits_{x\in F}I(x)=\alpha$. Hence, for any fixed $\delta>0$

$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in B_r(x)\right\}\leq -I(x)\leq-\alpha<-\alpha+\delta$$

for all $x\in F$. Consequently, by the definition of limit, for every $x\in F$ there exists $r_x>0$ such that

$$\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in B_{r_x}(x)\right\}\leq -I(x)<-\alpha+\delta.$$

Since the family of balls $B_{r_x}(x)$, $x\in F$, is an open cover[21] of F. Be the compactness of F, there exists a finite subcover of F, i.e. there exist $x_1,\dots,x_m\in F$ such that $F\subset\bigcup_{k=1}^m B_{r_{x_k}(x_k)}$. Now we can estimate

$$\begin{aligned}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in F\right\}&\leq\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in\bigcup_{k=1}^m B_{r_{x_k}}(x_k)\right\}\leq\varlimsup_{\varepsilon\to 0}\varepsilon\ln\left(\sum_{k=1}^m\mathbb{P}\left\{\xi_\varepsilon\in B_{r_{x_k}}(x_k)\right\}\right)\\&\overset{\text{Exercise 3.4}}{=}\max_{k=1,\dots,m}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in B_{r_{x_k}}(x_k)\right\}<-\alpha+\delta=-\inf_{x\in F}I(x)+\delta.\end{aligned}$$

Making $\delta\to 0$, we obtain

$$\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{\xi_\varepsilon\in F\right\}\leq-\inf_{x\in F}I(x).$$

Similarly, one can show inequality (13) in the case $\alpha=+\infty$ replacing $-\alpha+\delta$ by $-\frac{1}{\delta}$. $\square$

Exercise 5.1. Finish the proof of Proposition 5.1 in the case $\inf\limits_{x\in F}I(x)=+\infty$.

[21]Each set $B_{r_x}(x)$ is open and $F\subset\bigcup_{x\in F}B_{r_x}(x)$

5.2 Exponential tightness

We start from the definition of exponential tightness.

Definition 5.2. A family of random elements $(\xi_\varepsilon)_{\varepsilon>0}$ is said to be **exponentially tight in** E if for any number $\beta > 0$ there exists a compact set $K \subset E$ such that

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \notin K\right\} \le -\beta. \tag{15}$$

We remark that in the case of a countable family of random elements $(\xi_n)_{n\ge1}$, the exponential tightness corresponds to the statement

$$\overline{\lim_{n\to\infty}}\, a_n \ln \mathbb{P}\left\{\xi_n \notin K\right\} \le -\beta \tag{16}$$

for some $a_n \to 0$.

Exercise 5.2. Prove that a family $(\xi_\varepsilon)_{\varepsilon>0}$ is exponentially tight in E if and only if for any $b > 0$ there exists a compact $K \subset E$ and $\varepsilon_0 > 0$ such that

$$\mathbb{P}\left\{\xi_\varepsilon \notin K\right\} \le e^{-\frac{1}{\varepsilon}b}, \quad \varepsilon \in (0, \varepsilon_0).$$

Exercise 5.2 shows that the exponential tightness is much more stronger than the tightness[22].

Exercise 5.3. Let E be a complete and separable metric space.

a) Show that exponential tightness implies tightness for a countable family of random variables.

(*Hint:* Prove a similar inequality to one in Exercise 5.2 and then use the fact that any random element on a complete and separable metric space is tight (see Lemma 3.2.1 [EK86]))

b) Show that tightness does not imply exponential tightness.

Example 5.3. Let ξ be a standard Gaussian vector in $\mathbb{R}^d$. We consider as before $\xi_\varepsilon = \sqrt{\varepsilon}\xi$ and check the family $(\xi_\varepsilon)_{\varepsilon>0}$ is exponentially tight in $\mathbb{R}^d$. For this we will use the fact that this family satisfies the LDP.

So, we fix $\beta > 0$ and take a compact set $K_a := [-a, a]^d$ such that $\inf\limits_{\mathbb{R}^d\setminus(-a,a)^d} I \ge \beta$, where $I(x) = \frac{\|x\|^2}{2}$, $x \in \mathbb{R}^d$, is the rate function for the family $(\xi_\varepsilon)_{\varepsilon>0}$[23]. Since the family $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the LDP with rate function I, we have

$$\overline{\lim_{\varepsilon\to 0}}\, \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \notin K_a\right\} \le \overline{\lim_{\varepsilon\to 0}}\, \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in \mathbb{R}^d \setminus (-a,a)^d\right\} \le -\inf_{\mathbb{R}^d\setminus(-a,a)^d} I \le -\beta.$$

It turns out, that the full LDP implies exponential tightness, but only for a countable family $(\xi_\varepsilon)_{n\ge1}$.

Proposition 5.2. *Let E be a complete and separable metric space. If $(\xi_n)_{n\ge1}$ satisfies the full LDP with a good rate function I. Then it is exponentially tight.*

[22] A family of random variables (ξ_ε) is **tight** if for any $\delta > 0$ there exists a compact set $K \subset E$ such that $\mathbb{P}\left\{\xi_\varepsilon \notin K\right\} \le \delta$ for all ε

[23] see Exercise 3.3

For the proof see e.g. Lemma 1.11 [Swa12].

Exercise 5.4. Let $(\xi_\varepsilon)_{\varepsilon>0}$ be a family of random variables in $\mathbb{R}$ such that there exist $\lambda > 0$ and $\kappa > 0$ such that $\mathbb{E}\, e^{\frac{\lambda}{\varepsilon}|\xi_\varepsilon|} \leq \kappa^{\frac{1}{\varepsilon}}$ for all $\varepsilon > 0$. Show that this family is exponentially tight.

(*Hint:* Use Chebyshev's inequality)

Proposition 5.3. *If a family $(\xi_\varepsilon)_{\varepsilon>0}$ is exponentially tight and satisfies a weak LDP in E with rate function I, then it satisfies a full LDP.*

Proof. In order to prove a full LDP, we need to stay only upper bound (7) for each closed set $F \subset E$. So, let F be a given closed set and K and $\beta > 0$ be the corresponding set and constant from the definition of exponential tightness (see (15)). Then, using properties of probability, we have

$$\mathbb{P}\left\{\xi_\varepsilon \in F\right\} = \mathbb{P}\left\{\xi_\varepsilon \in F \cap K\right\} + \mathbb{P}\left\{\xi_\varepsilon \in K^c\right\},$$

where $K = E \setminus K$ is the complement of K. Consequently, using Exercise 3.4 and the fact that the set $K \cap F$ is compact[24], one can estimate

$$\begin{aligned}\varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F\right\} &= \varlimsup_{\varepsilon\to 0} \varepsilon \ln\left(\mathbb{P}\left\{\xi_\varepsilon \in F \cap K\right\} + \mathbb{P}\left\{\xi_\varepsilon \in K^c\right\}\right) \\ &= \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in F \cap K\right\} \vee \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi \not\in K\right\} \\ &\leq \left(-\inf_{x\in F\cap K} I(x)\right) \vee (-\beta) \leq \left(-\inf_{x\in F} I(x)\right) \vee (-\beta).\end{aligned}$$

Letting $\beta \to +\infty$, we get upper bound (7). □

[24] since $K \cap F$ is a closed subset of the compact set K

6 Lecture 6 – Large deviation principle for Brownian motion

6.1 Multidimensional Cramer's theorem

This lecture is devoted to the LDP for a Brownian motion. We first consider the LDP for finite dimensional distributions of a Brownian motion, which will explain better the form of the rate function. Then we study the admissible shifts of Brownian motion, which allows as in the next lecture rigorously prove the LDP. We start from the LDP for empirical mean of random vectors. This result generalises Theorem 2.1.

Similarly to the one-dimensional case we introduce the **comulant generating function** associated with a random vector ξ in $\mathbb{R}^d$ as follows

$$\varphi_\xi(\lambda) = \ln \mathbb{E}\, e^{\lambda\cdot\xi}, \quad \lambda \in \mathbb{R}^d,$$

where $a \cdot b = a_1 b_1 + \dots + a_d b_d$ for $a = (a_1, \dots, a_d)$ and $b = (b_1, \dots, b_d)$ from $\mathbb{R}^d$. As in one-dimensional case[25], one can show that the function φ is convex. So, we can introduce the Fenchel-Legendre transform

$$\varphi_\xi^*(x) = \sup_{\lambda\in\mathbb{R}^d} \{\lambda \cdot x - \varphi(\lambda)\}, \quad x \in \mathbb{R}^d,$$

of a function φ.

Exercise 6.1. For any random vector $\xi \in \mathbb{R}^d$ and non-singular $d \times d$ matrix A, show that $\varphi_{A\xi}(\lambda) = \varphi_\xi(\lambda A)$ and $\varphi^*_{A\xi}(x) = \varphi^*_\xi(A^{-1}x)$.

Exercise 6.2. For any pair of independent random vectors ξ and η show that $\varphi_{\xi,\eta}(\lambda, \mu) = \varphi_\xi(\lambda) + \varphi_\eta(\mu)$ and $\varphi^*_{\xi,\eta}(x, y) = \varphi^*_\xi(x) + \varphi^*_\eta(y)$.

(*Hint:* To prove the second equality, use the equality $\sup\limits_{\lambda,\mu} f(\lambda, \mu) = \sup\limits_{\lambda} \sup\limits_{\mu} f(\lambda, \mu)$)

The following theorem is a multidimensional Cramer's theorem.

Theorem 6.1. *Let $\xi_1, \xi_2, \dots$ be a sequence of independent identically distributed random vectors in $\mathbb{R}^d$ with comulant generating function φ and let $S_n = \xi_1 + \dots + \xi_n$. If φ is finite in a neighbourhood of 0 then the family $\left(\frac{1}{n}S_n\right)_{n\geq 1}$ satisfies the large deviation principle with good rate function φ^*, that is, for every Borel set $A \subset \mathbb{R}^d$*

$$-\inf_{x\in A^\circ} \varphi^*(x) \leq \varliminf_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \in A\right\} \leq \varlimsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{1}{n}S_n \in A\right\} \leq -\inf_{x\in \bar{A}} \varphi^*(x).$$

For proof of Theorem 6.1 see e.g. [RAS15, P.61] (for simpler proof in the case $\varphi(\lambda) < \infty$, $\lambda \in \mathbb{R}^d$, see e.g. [Var84, Theorem 3.1], [Kal02, Theorem 27.5] or [DZ98, Theorem 2.2.30].

Exercise 6.3. Let $\xi_1, \xi_2, \dots$ be independent random vectors in $\mathbb{R}^d$ whose coordinates are independent exponentially distributed random variables with rate γ.[26] Show that the empirical means $\left(\frac{1}{n}S_n\right)_{n\geq 1}$ satisfies the LDP in $\mathbb{R}^d$ and find the corresponding rate function I.

(*Hint:* Use Proposition 6.1. For computation of the rate function use exercises 2.7 and 6.2)

[25] see Exercise 2.1
[26] see also Example 2.2

6.2 Schilder's theorem

We start this section with computation of the rate function for finite dimensional distributions of a Brownian motion. So, let $w(t)$, $t \in [0,T]$, denote a standard Brownian motion on $\mathbb{R}$.[27] We take a partition $0 = t_0 < t_1 < \dots < t_d = T$ and consider the random vector $\xi = (w(t_1), \dots, w(t_d))$ in $\mathbb{R}^d$. Let $\xi_1, \xi_2, \dots$ be independent copies of ξ. Then the distribution of

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} \xi_k$$

coincides with the distribution of $\frac{1}{\sqrt{n}}(w(t_1), \dots, w(t_d))$. Consequently, one can use Theorem 6.1 to conclude that the family $\left(\frac{1}{\sqrt{n}}(w(t_t), \dots, w(t_d))\right)_{n \geq 1}$ satisfies the LDP with good rate function φ^*_ξ. Next we compute φ^*_ξ to see the precise form of the rate function. We remark that the random vector

$$\eta = \left(\frac{w(t_1)}{\sqrt{t_1}}, \frac{w(t_2) - w(t_1)}{\sqrt{t_2 - t_1}}, \dots, \frac{w(t_d) - w(t_{d-1})}{\sqrt{t_d - t_{d-1}}} \right)$$

is a standard Gaussian vector in $\mathbb{R}^d$. According to exercises 6.2 and 2.7,

$$\varphi^*_\eta(x) = \frac{\|x\|^2}{2}, \quad x \in \mathbb{R}^d.$$

We observe that

$$\eta = A \begin{pmatrix} w(t_1) \\ \dots \\ w(t_d) \end{pmatrix},$$

where A is some non-singular $d \times d$-matrix. Thus, by Exercise 6.1,

$$\varphi^*_\xi(x) = \varphi^*_{A^{-1}\eta}(x) = \varphi^*_\eta(Ax) = \frac{\|Ax\|^2}{2} = \frac{1}{2} \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{t_k - t_{k-1}},$$

where $x_0 = 0$.

Let us denote

$$w_\varepsilon(t) = \sqrt{\varepsilon} w(t), \quad t \in [0,T].$$

Then taking a function $f \in \mathrm{C}[0,T]$, we should expect that the family $(w_\varepsilon)_{\varepsilon>0}$ will satisfy the LDP in $\mathrm{C}[0,T]$ with rate function

$$I(f) = \frac{1}{2} \int_0^T f'^2(t) dt.$$

Now we give a rigorous statement about the LDP for a Brownian motion. So, let $H^2_0[0,T]$ be a set of all absolutely continuous functions $h \in \mathrm{C}_0[0,T]$ with $\dot{h} \in L_2[0,T]$ (see also Section 4.2 for more details).

Theorem 6.2 (Schilder's theorem). *The family $(w_\varepsilon)_{\varepsilon>0}$ satisfies the large deviation principle in $\mathrm{C}_0[0,T]$ with good rate function*

$$I(f) = \begin{cases} \frac{1}{2} \int_0^T \dot{f}^2(t) dt & \text{if } f \in H^2_0[0,T], \\ +\infty & \text{otherwise }. \end{cases}$$

[27] $w(t)$, $t \in [0,T]$ is a Brownian motion with $w(0) = 0$ and $\operatorname{Var} w(t) = t$

In order to prove Schilder's theorem, we are going estimate probabilities $\mathbb{P}\{w_\varepsilon \in B_r(f)\}$, where $B_r(f) = \{g \in \mathrm{C}[0,T] : \|g - f\|_C < r\}$ is the ball in $\mathrm{C}[0,T]$ with center f and radius r. This will be enough to prove the weak LDP according to Proposition 5.1 and Exercise 3.1. Then we will prove the exponential tightness of $(w_\varepsilon)_{\varepsilon>0}$ that will guarantee the full LDP by Proposition 5.3. The fact that the rate function I is good was considered as an exercise (see Exercise 4.9 above).

6.3 Cameron-Martin formula

In order to estimate the probability $\mathbb{P}\{w_\varepsilon \in B_r(f)\} = \mathbb{P}\{\|w_\varepsilon - f\|_C < r\}$, we need to work with distribution of the process $w_\varepsilon(t) - f(t)$, $t \in [0,T]$. We start the section with a simple observation. Let a random variable $\eta \sim N(0,1)$ be given on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $a \in \mathbb{R}$. It turns out that one can change the probability measure $\mathbb{P}$ in such a way that the random variable $\eta - a$ has a standard normal distributed. We note that

$$e^{-\frac{(x-a)^2}{2}} = e^{ax - \frac{a^2}{2}} e^{-\frac{x^2}{2}}.$$

Considering the new probability measure on Ω defined as

$$\mathbb{P}^a\{A\} = \mathbb{E}\,\mathbb{I}_A e^{a\eta - \frac{a^2}{2}}, \quad A \in \mathcal{F},$$

we claim that the random variable $\eta - a$ has a standard normal distribution on $(\Omega, \mathcal{F}, \mathbb{P}^a)$.

Exercise 6.4. Show that $\mathbb{P}^a$ is a probability measure on Ω, i.e. $\mathbb{P}^a\{\Omega\} = 1$.

Indeed, for any $z \in \mathbb{R}$ we have

$$\begin{aligned}
\mathbb{P}^a\{\eta - a \leq z\} &= \mathbb{E}\,\mathbb{I}_{\{\eta - a \leq z\}} e^{a\eta - \frac{a^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z+a} e^{ax - \frac{a^2}{2}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z+a} e^{-\frac{(x-a)^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx.
\end{aligned}$$

It turns out that for a Brownian motion we can do the same. So, let $w_{\sigma^2}(t)$, $t \in [0,T]$, be a Brownian motion with diffusion rate[28] σ^2 defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We introduce a new probability measure on Ω defined as follows

$$\mathbb{P}^h\{A\} = \mathbb{E}\,\mathbb{I}_A e^{\int_0^T h(t) dw_{\sigma^2}(t) - \frac{\sigma^2}{2} \int_0^T h^2 dt}, \quad A \in \mathcal{F},$$

where h is a fixed function from $L_2[0,T]$.

Proposition 6.1. *The process*

$$w_{\sigma^2}(t) - \sigma^2 \int_0^t h(s) ds, \quad t \in [0,T],$$

is a Brownian motion with diffusion rate σ^2 on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^h)$.

We remark that the statement of Proposition 6.1 is a consequence of more general Cameron-Martin theorem about admissible shifts of Brownian motion (see [Kal02, Theorems 18.22]).

[28] $w(t)$, $t \in [0,T]$, is a Brownian motion with $\operatorname{Var} w(t) = \sigma^2 t$

Exercise 6.5. Let $w(t)$, $t \in [0,T]$, be a Brownian motion with diffusion rate σ^2 defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $f \in H_0^2[0,T]$. Find a probability measure $\tilde{\mathbb{P}}$ such that $w(t) - f(t)$, $t \in [0,T]$, is a Brownian motion on $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$.

(*Hint:* Use Proposition 6.1 and definition of absolutely continuous functions)

Exercise 6.6. Show that for every $a \in \mathbb{R}$ and $\delta > 0$

$$\mathbb{P}\left\{w_{\sigma^2}(t) + at < \delta, \quad t \in [0,T]\right\} > 0.$$

(*Hint:* Use Proposition 6.1 and the fact that $\sup\limits_{t\in[0,T]} w_{\sigma^2}(t)$ and $|w_{\sigma^2}(T)|$ have the same distribution[29])

[29] see Proposition 13.13 [Kal02]

7 Lecture 7 – Proof of Schilder's theorem

7.1 Proof of weak LDP for Brownian motion

The goal of this lecture is to prove the LDP for a family $(w_\varepsilon)_{\varepsilon>0}$ of Brownian motions, where $w_\varepsilon(t) = \sqrt{\varepsilon}w(t)$, $t \in [0,T]$. The rigorous statement was formulated in the previous lecture (see Theorem 6.2). In this section, we will prove the weak LDP.

For the proof of **the lower bound** we use Exercise 3.1. So, we need to show that

1) for every $f \in \mathrm{C}_0^2[0,T]$
$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon \ln \mathbb{P}\left\{w_\varepsilon \in B_r(f)\right\} \geq -I(f);$$
where
$$I(f) = \begin{cases} \frac{1}{2}\int_0^T \dot{f}^2(t)dt & \text{if } f \in H_0^2[0,T], \\ +\infty & \text{otherwise.}\end{cases}$$

2) For every $f \in H_0^2[0,T]$ there exists a sequence f_n, $n \geq 1$, from $\mathrm{C}_0^2[0,T]$ such that $f_n \to f$ in $\mathrm{C}_0[0,T]$ and $I(f_n) \to I(f)$, $n \to \infty$.

We start from checking 1). Take a function $f \in \mathrm{C}_0^2[0,T]$ and estimate $\mathbb{P}\left\{\|w_\varepsilon - f\|_C < r\right\}$ from below, using Proposition 6.1. We set $h := f'$ and consider the following transformation of the probability measure $\mathbb{P}$
$$\mathbb{P}^{h,\varepsilon}\{A\} = \mathbb{E}\,\mathbb{I}_A e^{\int_0^T \frac{h(t)}{\varepsilon}dw_\varepsilon(t) - \frac{\varepsilon}{2}\int_0^T \frac{h^2(t)}{\varepsilon^2}dt} = \mathbb{E}\,\mathbb{I}_A e^{\frac{1}{\varepsilon}\left[\int_0^T h(t)dw_\varepsilon(t) - \frac{1}{2}\int_0^T h^2(t)dt\right]}.$$

Then the process
$$w_\varepsilon(t) - \varepsilon\int_0^t \frac{h(s)}{\varepsilon}ds = w_\varepsilon(t) - \int_0^t f'(s)ds = w_\varepsilon(t) - f(t), \quad t \in [0,T],$$
is a Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^{h,\varepsilon})$ with diffusion rate ε, according to Proposition 6.1. Integrating by parts in the first integral,[30] we have
$$\int_0^T h(t)dw_\varepsilon(t) - \frac{1}{2}\int_0^T h^2(t)dt = h(T)w_\varepsilon(T) - \int_0^T h'(t)w_\varepsilon(t)dt - \frac{1}{2}\int_0^T h^2(t)dt =: \Phi(h, w_\varepsilon).$$

Now, we can estimate
$$\begin{aligned}\mathbb{P}\left\{w_\varepsilon \in B_r(f)\right\} &= \mathbb{E}\left[\mathbb{I}_{\{w_\varepsilon \in B_r(f)\}}e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}e^{-\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}\right] \\ &\geq \mathbb{E}\left[\mathbb{I}_{\{w_\varepsilon \in B_r(f)\}}e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}e^{-\frac{1}{\varepsilon}\sup\limits_{g\in B_r(f)}\Phi(h,g)}\right] \\ &= e^{-\frac{1}{\varepsilon}\sup\limits_{g\in B_r(f)}\Phi(h,g)}\mathbb{P}^{\varepsilon,h}\left\{\|w_\varepsilon - f\|_C < r\right\} = e^{-\frac{1}{\varepsilon}\sup\limits_{g\in B_r(f)}\Phi(h,g)}\mathbb{P}\left\{\|w_\varepsilon\|_C < r\right\},\end{aligned} \tag{17}$$

[30] see Exercise 7.3 below

where the latter equality follows from Proposition 6.1. Hence,

$$\begin{aligned}\lim_{r\to 0}\varliminf_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{w_\varepsilon\in B_r(f)\right\} &\geq -\lim_{r\to 0}\sup_{g\in B_r(f)}\Phi(h,g)\\ &= \Phi(h,f)=h(T)f(T)-\int_0^T h'(t)f(t)dt-\frac{1}{2}\int_0^T h^2(t)dt\\ &=\int_0^T h^2(t)dt-\frac{1}{2}\int_0^T h^2(t)dt=I(f)\end{aligned}$$

because $\mathbb{P}\left\{\|w_\varepsilon\|_C<r\right\}\to 1$ as $\varepsilon\to 0$,[31] and the function Φ is continuous on $\mathrm{C}_0[0,T]$ in the second argument. This finishes the proof of 1). The proof of 2) is proposed as an exercise (see Exercise 7.2).

Exercise 7.1. Let $w_\varepsilon(t)$, $t\in[0,T]$, denote a Brownian motion with diffusion rate ε for every $\varepsilon>0$. Show that $\mathbb{P}\left\{\|w_\varepsilon\|_C<r\right\}\to 1$ as $\varepsilon\to 0$, for all $r>0$.

Exercise 7.2. Show that for any $f\in H_0^2[0,T]$ there exists a sequence $(f_n)_{n\geq 1}$ from E_0 such that $f_n\to f$ in $\mathrm{C}_0[0,T]$ and $I(f_n)\to I(f)$ as $n\to\infty$.

(*Hint:* Use first the fact that $\mathrm{C}^1[0,T]$ is dense in $L_2[0,T]$. Then show that if $h_n\to h$ in $L_2[0,T]$, then $\int_0^\cdot h_n(s)ds$ tends to $\int_0^\cdot h(s)ds$ in $\mathrm{C}_0[0,T]$, using Hölder's inequality)

Exercise 7.3. Let $h\in\mathrm{C}^1[0,T]$ and $w(t)$, $t\in[0,T]$, be a Brownian motion. Show that

$$\int_0^T h(t)dw(t)=h(T)w(T)-h(0)w(0)-\int_0^T h'(t)w(t)dt.$$

(*Hint:* Take a partition $0=t_0<t_1<\dots<t_n=T$ and check first that functions $h_n=\sum_{k=1}^n h(t_k)\mathbb{I}_{[t_{k-1},t_k)}$ converge to h in $L_2[0,T]$ as the mesh of partition goes to 0, using e.g. the uniform continuity of h on $[0,T]$. Next show that

$$\sum_{k=1}^n h(t_{k-1})(w(t_k)-w(t_{k-1}))=h(t_{n-1})w(T)-h(0)w(0)-\sum_{k=1}^{n-1}w(t_k)(h(t_k)-h(t_{k-1}))$$

Then prove that the first partial sum converges to the integral $\int_0^T h(t)dw(t)$ in L_2 and the second partial sum converges to $\int_0^T w(t)dh(t)$ a.s. as the mesh of partition goes to 0)

To prove **the upper bound**[32] (13) for any compact set $F\subset\mathrm{C}_0[0,T]$, we will use Proposition 5.1. We are going to show that for any $f\in\mathrm{C}_0[0,T]$

$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{w_\varepsilon\in B_r(f)\right\}\leq -I(f).$$

So we fix any $f\in\mathrm{C}_0[0,T]$ and $h\in\mathrm{C}^1[0,T]$, and estimate

$$\begin{aligned}\mathbb{P}\left\{w_\varepsilon\in B_r(f)\right\}&=\mathbb{E}\left[\mathbb{I}_{\{w_\varepsilon\in B_r(f)\}}e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}e^{-\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}\right]\\ &\leq\mathbb{E}\left[\mathbb{I}_{\{w_\varepsilon\in B_r(f)\}}e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}e^{-\frac{1}{\varepsilon}\inf_{g\in B_r(f)}\Phi(h,g)}\right]\\ &\leq e^{-\frac{1}{\varepsilon}\inf_{g\in B_r(f)}\Phi(h,g)}\,\mathbb{E}\,e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}=e^{-\frac{1}{\varepsilon}\inf_{g\in B_r(f)}\Phi(h,g)},\end{aligned}$$

[31] see Exercise 7.1
[32] The method applied here was taken from [DMRYZ04], see also [KvR19] for an infinite dimensional state space

because $\mathbb{E}\, e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)} = 1$. The last equality follows from the fact that $\mathbb{P}^{h,\varepsilon}$ is a probability measure and $\mathbb{E}\, e^{\frac{1}{\varepsilon}\Phi(h,w_\varepsilon)}$ is the expectation of 1 with respect to $\mathbb{P}^{h,\varepsilon}$. Consequently,

$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{w_\varepsilon \in B_r(f)\right\} \leq -\lim_{r\to 0}\inf_{g\in B_r(f)}\Phi(h,g) = -\Phi(h,f).$$

Now, taking infimum over all $h \in \mathrm{C}^1[0,T]$, we obtain

$$\lim_{r\to 0}\varlimsup_{\varepsilon\to 0}\varepsilon\ln\mathbb{P}\left\{w_\varepsilon \in B_r(f)\right\} \leq \inf_{h\in \mathrm{C}^1[0,T]}(-\Phi(h,f)) = -\sup_{h\in \mathrm{C}^1[0,T]}\Phi(h,f).$$

It remains only to show that

$$\sup_{h\in \mathrm{C}^1[0,T]}\Phi(h,f) = I(f). \tag{18}$$

7.2 A variational problem

We will check equality (18) inly for $f \in \mathrm{C}_0^2[0,T]$. This case is much more simple. The general case is based on the Riesz representation theorem and can be found e.g. in [KvR19, Proposition 4.6]. We observe that for $f \in \mathrm{C}_0^2[0,T]$ and any $h \in \mathrm{C}^1[0,T]$

$$\begin{aligned}\Phi(h,f) &= h(T)f(T) - \int_0^T h'(t)f(t)dt - \frac{1}{2}\int_0^T h^2(t)dt \\ &= \int_0^T h(t)f'(t)dt - \frac{1}{2}\int_0^T h^2(t)dt \leq \frac{1}{2}\int_0^T f'^2(t)dt = I(f),\end{aligned}$$

where we first used the integration by parts and then the trivial inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$, $a, b \in \mathbb{R}$. Moreover, we see that the last inequality becomes an equality if $h = f'$. So, the supremum is attained at the point $h = f'$ and $\Phi(f', f) = I(f)$. This proves (18).

7.3 Exponential tightness of I

To finish the proof of Schilder's theorem, it remains to prove that $(w_\varepsilon)_{\varepsilon>0}$ is exponentially tight. Exercise 5.4 shows that the estimate

$$\mathbb{E}\, e^{\frac{\lambda}{\varepsilon}|\xi_\varepsilon|} \leq \kappa^{\frac{1}{\varepsilon}}$$

for some $\kappa > 0$, $\lambda > 0$ and all $\varepsilon > 0$, is enough to conclude the exponential tightness of $(\xi_\varepsilon)_{\varepsilon>0}$ in $\mathbb{R}$. It turns out, that a similar estimate allows to get exponential tightness in the space of continuous functions. However, one has to control the Hölder norm.

Proposition 7.1. *Let $(\xi_\varepsilon)_{\varepsilon>0}$ be a family of random elements in $\mathrm{C}_0[0,T]$. If there exist positive constants γ, λ and κ such that for all $s, t \in [0,T]$, $s < t$, and $\varepsilon > 0$*

$$\mathbb{E}\, e^{\frac{\lambda}{\varepsilon}\frac{|\xi_\varepsilon(t)-\xi_\varepsilon(s)|}{(t-s)^\gamma}} \leq \kappa^{\frac{1}{\varepsilon}},$$

then the family $(\xi_\varepsilon)_{\varepsilon>0}$ is exponentially tight in $\mathrm{C}_0[0,T]$.

The proof of Proposition 7.1 follows from Corollary 7.1 [Sch97].

Lemma 7.1. *Let $w(t)$, $t \in [0,T]$, be a standard Brownian motion. Then the family $(\sqrt{\varepsilon}w)_{\varepsilon>0}$ is exponentially tight in $\mathrm{C}_0[0,T]$.*

Proof. To check the exponential tightness of $(\sqrt{\varepsilon}w)_{\varepsilon>0}$, we will use Proposition 7.1. We first remark that $\mathbb{E}\, e^{\alpha(w(t)-w(s))-\frac{\alpha^2}{2}(t-s)} = 1$ for all $\alpha \in \mathbb{R}$ and $s,t \in [0,T]$, $s<t$ (see Exercise 7.4 below). So, we can estimate for $\varepsilon > 0$, $s<t$ and $\alpha > 0$

$$\mathbb{E}\, e^{\alpha|\sqrt{\varepsilon}w(t)-\sqrt{\varepsilon}w(s)|} \leq \mathbb{E}\, e^{\alpha(\sqrt{\varepsilon}w(t)-\sqrt{\varepsilon}w(s))} + \mathbb{E}\, e^{-\alpha(\sqrt{\varepsilon}w(t)-\sqrt{\varepsilon}w(s))}$$
$$= 2\mathbb{E}\, e^{\alpha(\sqrt{\varepsilon}w(t)-\sqrt{\varepsilon}w(s))} = 2\mathbb{E}\, e^{\alpha\sqrt{\varepsilon}(w(t)-w(s))-\frac{\alpha^2\varepsilon}{2}(t-s)+\frac{\alpha^2\varepsilon}{2}(t-s)} = 2e^{\frac{\alpha^2\varepsilon}{2}(t-s)}.$$

Taking $\alpha := \frac{\sqrt{2}}{\varepsilon\sqrt{t-s}}$, we obtain

$$\mathbb{E}\, e^{\frac{\sqrt{2}|\sqrt{\varepsilon}w(t)-\sqrt{\varepsilon}w(s)|}{\varepsilon\sqrt{t-s}}} \leq 2\varepsilon^{\frac{1}{2}} \leq (2\varepsilon)^{\frac{1}{\varepsilon}}.$$

This implies the exponential tightness. □

Exercise 7.4. Let $w(t)$, $t \in [0,T]$, be a standard Brownian motion. Show directly that for each $\alpha \in \mathbb{R}$, $s,t \in [0,T]$, $s<t$

$$\mathbb{E}\, e^{\alpha(w(t)-w(s))-\frac{\alpha^2}{2}(t-s)} = 1.$$

(*Hint:* Use Exercise 6.4)

Remark 7.1. Let $w(t)$, $t \in [0,T]$, be a standard Brownian motion on $\mathbb{R}^d$. Then using the same argument, one can prove that the family $(w_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in $\mathrm{C}_0([0,T],\mathbb{R}^d)$ with rate function

$$I(f) = \begin{cases} \frac{1}{2}\int_0^T \|\dot{f}(t)\|^2 dt & \text{if } f \in H_0^2([0,T],\mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mathrm{C}_0([0,T],\mathbb{R}^d) = \{f = (f_1,\dots,f_d) : \ f_i \in \mathrm{C}_0[0,T], \quad i=1,\dots,d\}$, $H_0^2([0,T],\mathbb{R}^d) = \{f = (f_1,\dots,f_d) : f_i \in \mathrm{C}_0[0,T], \quad i=1,\dots,d\}$ and $\dot{f} = (\dot{f}_1,\dots,\dot{f}_d)$.

8 Lecture 8 – Contraction principle and Freidlin-Wentzell theory

8.1 Contraction principle

The goal of this section is the transformation of LDP under a continuous map.

Theorem 8.1 (Contraction principle)**.** *Consider a continuous function f between two metric spaces E and S, and let ξ_ε be random elements in E. If $(\xi_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in E with rate function I, then the images $f(\xi_\varepsilon)$ satisfy the LDP in S with rate function*

$$J(y) = \inf\left\{I(x) : \ f(x) = y\right\} = \inf_{f^{-1}(\{y\})} I, \quad y \in E. \tag{19}$$

Moreover, J is a good rate function on S whenever the function I is good on E.

Proof. We take a closed set $F \subset S$ and denote $f^{-1}(F) = \{x : \ f(x) \in F\} \subset E$. Then

$$\begin{aligned}\varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{f(\xi_\varepsilon) \in F\right\} &= \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\xi_\varepsilon \in f^{-1}(F)\right\} \\ &\leq -\inf_{x\in f^{-1}(F)} I(x) = -\inf_{y\in F}\inf_{f(x)=y} I(x) = -\inf_{y\in F} J(y).\end{aligned}$$

The lower bound can be proved similarly.

When I is good, we claim that

$$\{J \leq \alpha\} = f\left(\{I \leq \alpha\}\right) = \left\{f(x) : \ I(x) \leq \alpha\right\}, \quad \alpha \geq 0. \tag{20}$$

To see this, fix any $\alpha \geq 0$, and let $x \in \{I \leq \alpha\}$, i.e. $I(x) \leq \alpha$. Then

$$J(f(x)) = \inf\left\{I(u) : \ f(u) = f(x)\right\} \leq I(x) \leq \alpha,$$

which means that $f(x) \in \{J \leq \alpha\}$. Since I is good and f is continuous, the infimum in (19) is attained at some $x \in E$, and we get $y = f(x)$ with $I(x) \leq \alpha$. Thus, $y \in f\left(\{I \leq \alpha\}\right)$, which completes the proof of (20). Since continuous maps preserve compactness, $\{J \leq \alpha\}$ is compact, by (20). □

Exercise 8.1. Let I be a good rate function on E and f be a continuous function from E to S. Show that the infimum in (19) is attained, that is, there exists $x \in E$ such that $f(x) = y$ and $J(y) = I(x)$.

Exercise 8.2. Let $w(t)$, $t \in [0,T]$, be a Brownian motion on $\mathbb{R}$ with diffusion rate σ^2 and $w(0) = x_0$. Show that $(\sqrt{\varepsilon} w)_{\varepsilon>0}$ satisfies the LDP in $\mathrm{C}[0,T]$ and find the associated rate function.

(*Hint:* Take the continuous map $\Phi(f)(t) = \sigma f(t) + x_0$, and use the contraction principle and Schilder's Theorem 6.2)

Remark 8.1. Let us explain the form of the rate function for Brownian motion using a concept of white noise and contraction principle. We recall that the white nose $\dot{w}(t)$, $t \in [0,T]$, formally can be defined as a Gaussian process with covariance $\mathbb{E}\,\dot{w}(t)\dot{w}(s) = \delta_0(t-s)$, where δ_0 denotes the Dirac delta function. One should interpret the white noise as a family of uncountable numbers of independent identically distributed Gaussian random variables. Similarly, as for Gaussian vectors, where the rate function is given by the formula $\frac{\|x\|^2}{2} = \sum_{k=1}^d \frac{x_k^2}{2}$, the rate function for the family $(\sqrt{\varepsilon}\dot{w})_{\varepsilon>0}$ should be

$$I_{\dot{w}}(x) = \frac{1}{2}\int_0^T x^2(t)dt.$$

We remark that a Brownian motion formally appears as a continuous function of white noise, namely the process $w(t) := \int_0^t \dot{w}(r)dr$, $t \in [0,T]$, defines a standard Brownian motion. Indeed, it is a Gaussian process with covariance

$$\mathbb{E}\left(\int_0^s \dot{w}(r_1)dr_1 \int_0^t \dot{w}(r_2)dr_2\right) = \int_0^s\int_0^t \mathbb{E}\,\dot{w}(r_1)\dot{w}(r_2)dr_1dr_2 = \int_0^s\int_0^t \delta_0(r_1-r_2)dr_1dr_2 = \int_0^s 1dr_1 = s$$

if $s \leq t$. So, $w = \Phi(\dot{w})$, where Φ denotes the integration procedure. By the contraction principle, the rate function of the family $(w_\varepsilon = \Phi(\dot{w}_\varepsilon))_{\varepsilon>0}$ has to be

$$I_w(x) = I_{\dot{w}}(\Phi^{-1}(x)) = \frac{1}{2}\int_0^T x'^2(t)dt.$$

8.2 Freidlin-Wentzell theory

In this section, we prove the LDP for solutions of stochastic differential equations (shortly SDE). Let consider a family $(z_\varepsilon)_{\varepsilon>0}$ of solutions to the following SDEs

$$dz_\varepsilon(t) = a(z_\varepsilon(t))dt + \sqrt{\varepsilon}dw(t), \quad z_\varepsilon(0) = 0, \tag{21}$$

where $a : \mathbb{R} \to \mathbb{R}$ is a bounded Lipschitz continuous[33] function and $w(t)$, $t \in [0,T]$, is a standard Brownian motion. We recall that a continuous process $z_\varepsilon(t)$, $t \in [0,T]$, is a solution to (21) if

$$z_\varepsilon(t) = \int_0^t a(z_\varepsilon(s))ds + \sqrt{\varepsilon}w(t), \quad t \in [0,T].$$

By Theorem 21.3 [Kal02], equation (21) has a unique solution.

Theorem 8.2 (Freidlin-Wentzell theorem)**.** *For any bounded Lipschitz continuous function $a : \mathbb{R} \to \mathbb{R}$ the solutions the family $(z_\varepsilon)_{\varepsilon>0}$ satisfies the large deviation principle in $\mathrm{C}[0,T]$ with good rate function*

$$I(f) = \begin{cases} \frac{1}{2}\int_0^T [\dot{f}(t) - a(f(t))]^2 dt & \text{if } f \in H_0^2[0,T], \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. To prove the theorem, we will use the contraction principle. We first remark that the equation

$$z(t) = \int_0^t a(z(s))ds + g(t), \quad t \in [0,T], \tag{22}$$

has a unique solution for any $g \in \mathrm{C}[0,T]$, since the function a is bounded and Lipschitz continuous. So, there exists a function $\Phi : \mathrm{C}[0,T] \to \mathrm{C}[0,T]$ such that $z = \Phi(g)$. Let us show that Φ is continuous. Take $g_1, g_2 \in \mathrm{C}[0,T]$ and set $z_1 := \Phi(g_1)$, $z_2 := \Phi(g_2)$. Then one can estimate

$$\begin{aligned} |z_1(t) - z_2(t)| = |\Phi(g_2) - \Phi(g_2)| &= \left|\int_0^t a(z_1(s)) - a(z_2(s))ds + g_1(t) - g_2(t)\right| \\ &\leq \int_0^t |a(z_1(s)) - a(z_2(s))|ds + |g_1(t) - g_2(t)| \\ &\leq L\int_0^t |z_1(s) - z_2(s)|ds + \|g_1 - g_2\|_C. \end{aligned}$$

[33] $a : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous function if there exists a constant L such that $|a(x_1) - a(x_2)| \leq L|x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$

Gronwall's Lemma 21.4 [Kal02] yields $|z_1(t) - z_2(t)| \leq \|g_1 - g_2\|_C e^{Lt}$ for all $t \in [0, T]$. Hence,

$$\|\Phi(g_1) - \Phi(g_2)\|_C = \|z_1 - z_2\|_C \leq e^{LT}\|g_1 - g_2\|_C,$$

which shows that Φ is continuous. Using Schilder's theorem 6.2 along with the contraction principle (see Theorem 8.1), we conclude that the family $(z_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in $\mathrm{C}[0, T]$ with the good rate function

$$I(f) = \inf\{I_w(g) : \ \Phi(g) = f\} = \inf\left\{I_w(g) : \ f(t) = \int_0^t a(f(s))ds + g(t)\right\},$$

where I_w is defined in Theorem 6.2. Due to the uniqueness of solutions to differential equation (22), the function Φ is bijective. Moreover, g and $f = \Phi(g)$ lie simultaneously in $H_0^2[0, T]$, in which case $\dot{g} = \dot{f} - a(f)$ almost everywhere.[34] Thus

$$I(f) = \begin{cases} \int_0^T (\dot{f}(t) - a(f(t)))^2 dt & \text{if } f \in H_0^2[0, T], \\ +\infty & \text{otherwise.} \end{cases}$$

$\square$

Exercise 8.3. Let $\Phi : \mathrm{C}[0, T] \to \mathrm{C}[0, T]$ be defined in the proof of Theorem 8.2.

1) Show that the function Φ is bijective.
2) Prove that $g \in H_0^2[0, T]$ if and only if $f = \Phi(g) \in H_0^2[0, T]$.
   (*Hint:* Use equation (22) and the definition of $H_0^2[0, T]$)
3) Show that $\dot{g} = \dot{f} - a(f)$ almost everywhere for every $g \in H_0^2[0, T]$ and $f = \Phi(g)$.

8.3 Contraction principle for some discontinuous functions

In the first section, we showed that LDP is preserved under a continuous transformation. But very often one must work with discontinuous transformations. It turns out that LDP can also be preserved in some cases. Let us consider the following example which was taken from [DO10].

Given a standard Brownian motion $w(t)$, $t \in [0, T]$, in $\mathbb{R}^d$ and a closed set $B \subset \mathbb{R}^d$, we consider the stopping time

$$\tau := \inf\{t : w(t) \in B\} \wedge T,$$

where $a \wedge b = \min\{a, b\}$. Let $y(t) := w(t \wedge \tau)$, $t \in [0, T]$, denote the stopped Brownian motion and $y_\varepsilon(t) := y(\varepsilon t)$, $t \in [0, T]$. We are interesting in the LDP for the family $(y_\varepsilon)_{\varepsilon>0}$. We remark that the process y_ε is obtained as an image of $w_\varepsilon(t) = w(\varepsilon t)$, $t \in [0, T]$. Indeed, let us define for a function $f \in \mathrm{C}_0([0, T], \mathbb{R}^d)$

$$\tau(f) := \inf\{t : \ f(t) \in B\} \wedge T,$$

and

$$\Phi(f)(t) := f(t \wedge \tau(f)), \quad t \in [0, T]. \tag{23}$$

Then, by Exercise 8.4, Φ is a map from $\mathrm{C}_0([0, T], \mathbb{R}^d)$ to $\mathrm{C}_0([0, T], \mathbb{R}^d)$ and $y_\varepsilon = \Phi(w_\varepsilon)$. Unfortunately, we cannot apply the contraction principle here since Φ is discontinuous. But still, one can use the idea of contraction principle to obtain the LDP for $(y_\varepsilon)_{\varepsilon>0}$. We remark also that the set B could be chosen

[34] see Exercise 8.3

by such a way that the set of discontinuous points of the map Φ has a positive Wiener measure[35] (for more details see Example 4.1 [DO10]).

Proposition 8.1. *The family* $(y_\varepsilon)_{\varepsilon>0}$ *satisfies the LDP in* $\mathrm{C}_0([0,T],\mathbb{R}^d)$ *with rate function*

$$I(f) = \begin{cases} \frac{1}{2}\int_0^T \dot{f}^2(t)dt & \text{if } f \in H_0^2([0,T],\mathbb{R}^d) \cap \operatorname{Im}\Phi, \\ +\infty & \text{otherwise,} \end{cases} \tag{24}$$

where $\operatorname{Im}\Phi = \left\{\Phi(f) :\ f \in \mathrm{C}_0([0,T],\mathbb{R}^d)\right\}$.

Proof. A detailed proof of the proposition can be found in [DO10, Section 4]. We present here only the main idea. For the proof of the lower bound we take a closed set $F \subset \mathrm{C}_0([0,T],\mathbb{R}^d)$ and estimate from above the upper limit

$$\begin{aligned} \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{y_\varepsilon \in F\right\} &= \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{\Phi(w_\varepsilon) \in F\right\} = \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{w_\varepsilon \in \Phi^{-1}(F)\right\} \\ &\le \varlimsup_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}\left\{w_\varepsilon \in \overline{\Phi^{-1}(F)}\right\} \le -\inf_{\overline{\Phi^{-1}(F)}} I_w, \end{aligned}$$

where I_w is the rate function defined in Theorem 6.2 and $\Phi^{-1}(F) = \left\{f \in \mathrm{C}_0([0,T],\mathbb{R}^d) :\ \Phi(f) \in F\right\}$. So, in order to obtain the upper bound (7), one needs to prove

$$\inf_{\overline{\Phi^{-1}(F)}} I_w = \inf_{\Phi^{-1}(F)} I_w \quad \left(= \inf_F \tilde{I}\right).$$ [36]

Similarly, for the proof of the lower bound (8), one needs to show that

$$\inf_{\Phi^{-1}(G)^\circ} I_w = \inf_{\Phi^{-1}(G)} I_w \quad \left(= \inf_G \tilde{I}\right)$$

for any open set $G \subset \mathrm{C}_0([0,T],\mathbb{R}^d)$. The prove of those equalities can be found in [DO10, Section 4] □

Exercise 8.4. Let Φ be defined by (23) for $d = 1$ and $B = \{0\}$. Show that Φ maps $\mathrm{C}[0,T]$ to $\mathrm{C}[0,T]$. Prove that it is discontinuous.

[35] $\mathbb{P}\left\{w \text{ is a point of discontinuity of } \Phi\right\} > 0$

[36] We remark that this equality and the equality for open G trivially holds if Φ is continuous, since $\Phi^{-1}(F)$ is closed and $\Phi^{-1}(G)$ is open

9 Lecture 9 – Some applications of large deviations

9.1 Curie-Weiss model of ferromagnetism

This section is taken from [RAS15, Section 3.4].

In this section, we consider an application of LDP in statistical mechanics, using a toy model of ferromagnetism. Let us imagine that a piece of material is magnetized by subjecting it to a magnetic field. Then assume that the field is turned off. We are interesting if the magnetization persist. To answer this question, we introduce a model called the Curie-Weiss ferromagnet and will try to understand this using large deviation theory.

Let us start from the description of the model. Consider n atoms each of them have a ± 1 valued spin ω_i, $i = 1, \ldots, n$. The space of n-spin configurations is $\Omega_n = \{-1, 1\}^n$. The energy of the system is given by the **Hamiltonian**

$$\mathcal{H}_n(\omega) = -\frac{J}{2n} \sum_{i,j=1}^{n} \omega_i \omega_j - h \sum_{j=1}^{n} \omega_j = -\frac{J}{2} \sum_{i=1}^{n} \omega_i \left(\frac{1}{n} \sum_{j=1}^{n} \omega_j \right) - h \sum_{j=1}^{n} \omega_j. \tag{25}$$

A ferromagnet has a positive coupling constant $J > 0$ and $h \in \mathbb{R}$ is the external magnetic field. Since nature prefers low energy, ferromagnet spins tend to align with each other and with the magnetic field h, if $h \neq 0$. The **Gibbs measure** for n spins is

$$\gamma_n(\omega) = \frac{1}{Z_n} e^{-\beta \mathcal{H}_n(\omega)} P_n(\omega), \quad \omega \in \Omega_n.$$

Here $P_n(\omega) = \frac{1}{2^n}$, $\beta > 0$ is the **inverse temperature** and Z_n is the normalization constant.

The Gibbs measure captures the competition between the ordering tendency of the energy term $\mathcal{H}(\omega)$ and the randomness represented by P_n. Indeed, let $h = 0$. If the temperature is high (β close to 0), then noise dominates and complete disorder reigns at the limit, $\lim_{\beta \to 0} \gamma_n(\omega) = P_n$. But if temperature goes to zero, then the limit $\lim_{\beta \to \infty} \gamma_n(\omega) = \frac{1}{2}(\delta_{\omega=1} + \delta_{\omega=-1})$ is concentrated on the two **ground states**. The key question is the existence of phase transition: namely, if there is a **critical temperature** β_c^{-1} **(Curie point)** at which the infinite model undergoes a transition that reflects the order/disorder dichotomy of the finite model.

Let a random vector $(\eta_1, \ldots, \eta_n)$ have distribution γ_n. We define magnetization as the expectation $M_n(\beta, h) = \mathbb{E}\, S_n$ of the total spin $S_n = \sum_{i=1}^n \eta_i$. We show that $\frac{1}{n} S_n$ converges and there exists a limit

$$m(\beta, h) = \lim_{n \to \infty} \frac{1}{n} M_n(\beta, h).$$

Then we will see something interesting as $h \to 0$. We remark, that $m(\beta, 0) = 0$, since $\gamma_n(\omega) = \gamma_n(-\omega)$.

Proposition 9.1. *The family* $\left(\frac{1}{n} S_n\right)_{n \geq 1}$ *satisfies the LDP in* $\mathbb{R}$ *with rate function*

$$I(x) = \begin{cases} \frac{1}{2}(1-x)\ln(1-x) + \frac{1}{2}(1+x)\ln(1+x) - \frac{1}{2}J\beta x^2 - \beta h x - c & \text{if } x \in [-1, 1], \\ +\infty & \text{otherwise,} \end{cases}$$

where $c = \inf\limits_{x \in [-1,1]} \left\{ \frac{1}{2}(1-x)\ln(1-x) + \frac{1}{2}(1+x)\ln(1+x) - \frac{1}{2}J\beta x^2 - \beta h x \right\}$.

For the proof of the proposition see [RAS15, Section 3.4].

In order to understand limit of $\frac{1}{n}S_n$, we find the minimizers of the rate function I. Critical points satisfy $I'(x) = 0$ that is equivalent to the equation

$$\frac{1}{2}\ln\frac{1+x}{1-x} = J\beta x + \beta h, \quad x \in [-1,1]. \tag{26}$$

Theorem 9.1. *Let $0 < \beta, J < \infty$ and $h \in \mathbb{R}$.*

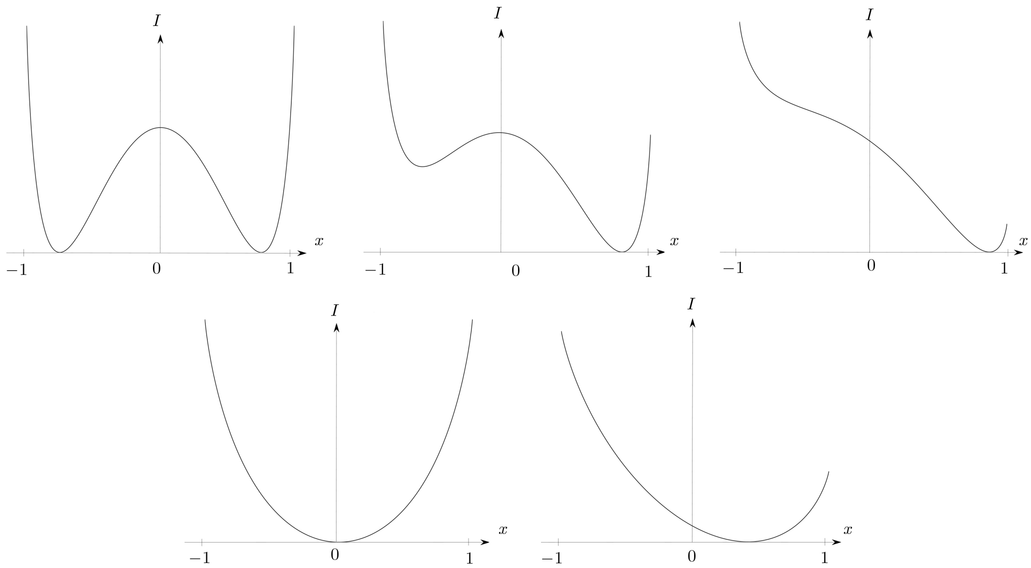
(i) For $h \neq 0$, $m(\beta, h)$ is the unique solution of (26) that has the same sign as h.

(ii) Let $h = 0$ and $\beta < J^{-1}$. Then $m(\beta, 0) = 0$ is the unique solution of (26) and $m(\beta, \tilde{h}) \to 0$ as $\tilde{h} \to 0$.

(iii) Let $h = 0$ and $\beta > J^{-1}$. Then (26) has two nonzero solutions $m(\beta,+) > 0$ and $m(\beta,-) = -m(\beta,+)$. Spontaneous magnetization happens: for $\beta > J^{-1} =: \beta_c$,

$$\lim_{\tilde{h}\to 0+} m(\beta,\tilde{h}) = m(\beta,+) \quad \textit{and} \quad \lim_{\tilde{h}\to 0-} m(\beta,\tilde{h}) = m(\beta,-).$$

We note that statements (ii) and (iii) follows directly from the form of equation (26). Statement (i) is the direct consequence of the further proposition and the dominated convergence theorem.



The graphs of the rate function I. Top plots have $\beta > J^{-1}$ while bottom plots have $\beta \leq J^{-1}$. Top left to right: $h = 0$, $0 < h < h_0(J,\beta)$ and $h > h_0(J,\beta)$. Bottom left to right, $h = 0$ and $h > 0$. The case $h < 0$ is symmetric to that of $h > 0$.

Proposition 9.2. *(i) Suppose that either $h \neq 0$, or $h = 0$ and $\beta \leq J^{-1}$. Then $\frac{1}{n}S_n \to m(\beta,h)$.*

(ii) If $h = 0$ and $\beta > J^{-1}$, then $\frac{1}{n}S_n \to \zeta$ weakly, where $\mathbb{P}\{\zeta = m(\beta,+)\} = \mathbb{P}\{\zeta = m(\beta,-)\} = \frac{1}{2}$.

Proof. We note that the first part of the proposition follows from the fact that the rate function I has a unique minimizer. Indeed,

$$\varlimsup_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\left\{\left|\frac{1}{n}S_n - m(\beta,h)\right| \geq \varepsilon\right\} \leq -\inf_{|x-m(\beta,h)|\geq\varepsilon} I(x) < 0.$$

For part (ii) the large deviation upper bound can be obtained similarly

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}S_n - m(\beta,-)\right| < \varepsilon \quad \text{or} \quad \left|\frac{1}{n}S_n - m(\beta,+)\right| < \varepsilon\right\} = 1.$$

Form $\gamma_n(\omega) = \gamma_n(-\omega)$ it follows that S_n is symmetric and so

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}S_n - m(\beta,-)\right| < \varepsilon\right\} = \lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}S_n - m(\beta,+)\right| < \varepsilon\right\} = \frac{1}{2}.$$

This shows the weak convergence of $\frac{1}{n}S_n$ to ζ. □

9.2 Varadhan formula

The goal of the present section is to show a connection of diffusion processes with the underlying geometry of the state space. This result was obtained by Varadhan in [Var67]. So, we are interesting in deviations of solution $x(t)$ of the SDE in $\mathbb{R}^d$

$$dx(t) = \sigma(x(t))dw(t), \quad x(0) = x_0, \tag{27}$$

from the initial value of x_0 as $t \to 0$, where $w(t)$, $t \in [0,1]$, denotes a standard Brownian motion in $\mathbb{R}^d$ and the $d \times d$-matrix σ is Lipschitz continuous.

We first consider the following family of SDEs

$$dx_\varepsilon(t) = \sigma(x_\varepsilon(t))dw_\varepsilon(t), \quad x(0) = x_0, \tag{28}$$

where $w_\varepsilon(t)$, $t \in [0,1]$, and $\varepsilon > 0$. For every $\varepsilon > 0$ the solution is the diffusion process corresponding to the operator

$$L_\varepsilon(f) = \frac{\varepsilon}{2}\sum_{i,j=1}^d a_{ij}\frac{\partial^2 f}{\partial x_i x_j},$$

with $a = \sigma\sigma^*$. We also assume that the matrix a is bounded and uniformly elliptic, that is, there exists $c > 0$ ans $C > 0$ such that for all $\lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$

$$c\|\lambda\|^2 \le \lambda a \lambda \le C\|\lambda\|^2,$$

where $\lambda a \lambda = \sum_{i,j=1}^d a_{ij}\lambda_i\lambda_j$. We remark that for every $\varepsilon > 0$ SDE (28) has a unique solution x_ε on the space $\mathrm{C}([0,1],\mathbb{R}^d)$.[37] The proof of the following theorem can be found in [Var84, Section 6].

Theorem 9.2. *The family $(x_\varepsilon)_{\varepsilon>0}$ satisfies the LDP in $\mathrm{C}([0,1],\mathbb{R}^d)$ with rate function*

$$I(f) = \begin{cases} \frac{1}{2}\int_0^1 \dot f(t) a^{-1}(f(t))\dot f(t)dt & \text{if } f \in H^2_{x_0}([0,1],\mathbb{R}^d), \\ +\infty & \text{otherwise,} \end{cases}$$

where $H^2_{x_0}([0,T];\mathbb{R}^d)$ is defined similarly as $H^2_0([0,T];\mathbb{R}^d)$, the only difference is $f(0) = x_0$.

[37] see e.g. Theorem 21.3 [Kal02]

Now we are going to obtain the LDP for the family $(x(\varepsilon))_{\varepsilon>0}$ where $x(t)$, $t \in [0,1]$, is a solution to equation (27). It is easily to see that $x(\varepsilon) = x_\varepsilon(1)$. So, we can apply the contraction principle to the LDP for $(x_\varepsilon)_{\varepsilon>0}$. We take the following continuous map on $\mathrm{C}([0,T],\mathbb{R}^d)$

$$\Phi(f) = f(1), \quad f \in \mathrm{C}([0,T],\mathbb{R}^d).$$

Then the family $(x(\varepsilon) = \Phi(x_\varepsilon))_{\varepsilon>0}$ satisfies the LDP with rate function

$$\begin{aligned} I_{x_0}(x_1) &= \inf\left\{ I(f) : \ f \in H_0^2([0,T],\mathbb{R}^d), \quad f(1) = x_1 \right\} \\ &= \frac{1}{2} \inf_{f(0)=x_0,\ f(1)=x_1} \int_0^1 \dot f(t) a^{-1}(f(t)) \dot f(t) dt =: \frac{d^2(x_0,x_1)}{2}, \end{aligned}$$

where the later infimum is taken over all functions $f \in H^2_{x_0}([0,1],\mathbb{R}^d)$ which end at x_1 (and begin at x_0).

Let us define locally the metric on $\mathbb{R}^d$ as

$$ds^2 = \sum_{i,j=1}^d a_{ij} dx_i dx_j.$$

Then the distance

$$d(x_0,x_1) = \left(\inf\left\{ \int_0^1 \dot f(t) a^{-1}(f(t)) \dot f(t) dt : \ f \in H^2_{x_0}([0,1],\mathbb{R}^d),\ f(1) = 1 \right\} \right)^{\frac{1}{2}}, \quad x_0, x_1 \in \mathbb{R}^d,$$

coincides with the global geodesic distance

$$d_{geod}(x_0,x_1) = \inf\left\{ \int_0^1 \sqrt{\dot f(t) a^{-1}(f(t)) \dot f(t)} dt : \ f \in H^2_{x_0}([0,1],\mathbb{R}^d),\ f(1) = 1 \right\}, \quad x_0, x_1 \in \mathbb{R}^d,$$

induced by this metric.

Exercise 9.1. Show that d_{geod} is a distance of $\mathbb{R}^d$.

We remark that the operator L is the Laplace-Beltrami operator on the Riemannian manifold $\mathbb{R}^d$ (with metric ds^2) and the associated process $x(t)$, $t \in [0,1]$, plays a role of Brownian motion on this space.

For further applications of large deviation principle see also [Var08].

References

[Con90] John B. Conway, *A course in functional analysis*, second ed., Graduate Texts in Mathematics, vol. 96, Springer-Verlag, New York, 1990. MR 1070713

[DE97] Paul Dupuis and Richard S. Ellis, *A weak convergence approach to the theory of large deviations*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1997, A Wiley-Interscience Publication. MR 1431744

[dH00] Frank den Hollander, *Large deviations*, Fields Institute Monographs, vol. 14, American Mathematical Society, Providence, RI, 2000. MR 1739680

[DMRYZ04] C. Donati-Martin, A. Rouault, M. Yor, and M. Zani, *Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes*, Probab. Theory Related Fields **129** (2004), no. 2, 261–289. MR 2063378

[DO10] A. A. Dorogovtsev and O. V. Ostapenko, *Large deviations for flows of interacting Brownian motions*, Stoch. Dyn. **10** (2010), no. 3, 315–339. MR 2671379

[DS89] Jean-Dominique Deuschel and Daniel W. Stroock, *Large deviations*, Pure and Applied Mathematics, vol. 137, Academic Press, Inc., Boston, MA, 1989. MR 997938

[DZ98] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, second ed., Applications of Mathematics (New York), vol. 38, Springer-Verlag, New York, 1998. MR 1619036

[EK86] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. MR 838085

[Ell06] Richard S. Ellis, *Entropy, large deviations, and statistical mechanics*, Classics in Mathematics, Springer-Verlag, Berlin, 2006, Reprint of the 1985 original. MR 2189669

[FW84] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 260, Springer-Verlag, New York, 1984, Translated from the Russian by Joseph Szücs. MR 722136

[Kal02] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR 1876169

[Kuz15] V. A. Kuznetsov, *On the large-deviation principle for the winding angle of a Brownian trajectory around the origin*, Theory Stoch. Process. **20** (2015), no. 2, 63–84. MR 3510229

[KvR19] Vitalii Konarovskyi and Max-K. von Renesse, *Modified massive Arratia flow and Wasserstein diffusion*, Comm. Pure Appl. Math. **72** (2019), no. 4, 764–800. MR 3914882

[MÖ8] Peter Mörters, *Large deviation theory and applications*, Lecture notes, 2008.

[RAS15] Firas Rassoul-Agha and Timo Seppäläinen, *A course on large deviations with an introduction to Gibbs measures*, Graduate Studies in Mathematics, vol. 162, American Mathematical Society, Providence, RI, 2015. MR 3309619

[Sch97] Alexander Schied, *Moderate deviations and functional LIL for super-Brownian motion*, Stochastic Process. Appl. **72** (1997), no. 1, 11–25. MR 1483609

[Swa12] Jan Swart, *Large deviation theory*, Lecture notes, 2012.

[Var67] S. R. S. Varadhan, *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Comm. Pure Appl. Math. **20** (1967), 431–455. MR 0208191

[Var84] ______, *Large deviations and applications*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984. MR 758258

[Var08] ______, *Large deviations*, Ann. Probab. **36** (2008), no. 2, 397–419. MR 2393987