INTRODUCTION TO LARGE DEVIATIONS

JULIEN POISAT

Contents

These notes are intended for the students of the MASTER 2 Applied and Theoretical Mathematics (MATH) at Université Paris-Dauphine PSL. They introduce the topic of *Large Deviations*: that is (in short) the branch of Probability Theory which aims to compute or at least estimate the probability of events that *deviate* from the typical law-of-large-number behavior. In many cases, the probability of such rare events is shown to decay exponentially fast in the system (or sample) size and the rate of decay may be expressed as the solution of a certain *variational problem*. For the most part, the structure is borrowed from [2]. In the first two sections we will treat (without much theory) two unavoidable results in this field, that is Cramér's theorem (large deviations for the sum of independent and identically distributed real random variables) and Sanov's theorem (large deviations for the empirical measure of independent and identically distributed random variables). Sanov's theorem makes use of the important notion of *entropy*, which also appears in other fields like information theory and statistical mechanics. General principles will be presented in Section 3, where topological considerations come into play. Those principles will be applied to the so-called Curie-Weiss model (a mean-field spin system) as an illustration of how large deviations and statistical mechanics are connected. Finally, Large Deviations for finite Markov chains will be covered in Section 4. The last section offers a selection of more advanced topics in order to arouse the reader's curiosity (or possibly engage discussions towards a research internship). Technical lemmas and reminders will be gathered in an appendix.

Date: November 17, 2025.

Much of the material (including some proofs) has been condensed in the form of exercises, which will be corrected in class. It is strongly advised to read the relevant sections beforehand (see the class journal on the author's webpage) in order to fully benefit from the lectures.

Notation

- $\mathbb{N} = \{1, 2, \ldots\}$ is the set of positive integers, while $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ is the set of non-negative integers.
- Unless stated otherwise, random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$.
- $\mathcal{N}(m, \sigma^2)$ is the Gaussian probability distribution on $\mathbb{R}$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 \in [0, \infty)$. When $\sigma^2 = 0$, this is a Dirac mass at m.
- We denote by δ_x the Dirac measure at x, i.e. $\delta_x(A) = 1_A(x)$, that is one if $x \in A$ and zero if $x \notin A$.
- By *replicas* of a random variable Z, we mean a collection of independent random variables that all are distributed as Z.
- For every $i, n \in \mathbb{N}$, $i[n]$ is the unique integer in $\{1, \ldots, n\}$ such that $i - i[n] \in n\mathbb{N}_0$.
- For every $\alpha, \beta \in [-\infty, \infty]$, $\alpha \vee \beta = \max(\alpha, \beta)$ and $\alpha \wedge \beta = \min(\alpha, \beta)$.

These notes start with an elementary but fundamental exercise that the reader should keep in mind throughout the course:

Exercise 1 (Warm-up)**.** Let (a_n) and (b_n) be two sequences valued in $[0, \infty]$. We use the convention $\log 0 = -\infty$ and $\log \infty = \infty$.

(1) Assume that $\liminf \frac{1}{n} \log a_n \geq \alpha$ and $\liminf \frac{1}{n} \log b_n \geq \beta$. Prove that

$$\liminf_{n\to\infty} \frac{1}{n} \log(a_n + b_n) \geq \alpha \vee \beta. \tag{0.1}$$

(2) Assume that $\limsup \frac{1}{n} \log a_n \leq \alpha$ and $\limsup \frac{1}{n} \log b_n \leq \beta$. Prove that

$$\limsup_{n\to\infty} \frac{1}{n} \log(a_n + b_n) \leq \alpha \vee \beta. \tag{0.2}$$

(3) Assume that $\lim \frac{1}{n} \log a_n = \alpha$ and $\lim \frac{1}{n} \log b_n = \beta$. Prove that

$$\lim_{n\to\infty} \frac{1}{n} \log(a_n + b_n) = \alpha \vee \beta. \tag{0.3}$$

1. Cramér's theorem for real-valued random variables

Let $X = (X_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) real random variables (r.v.). We assume that the r.v.'s are *integrable*, i.e. $\mathrm{E}(|X_1|) < +\infty$. The (typical) asymptotic behaviour of the *empirical mean* associated to this random sequence is given by the (strong) Law of Large Numbers:

$$\frac{1}{n} \sum_{1\leq i\leq n} X_i \overset{n\to\infty}{\longrightarrow} \mathrm{E}(X_1) < +\infty, \qquad \text{P-a.s. and in } L_1. \tag{1.1}$$

If the r.v.'s are also *square-integrable*, i.e. $\mathrm{E}(X_1^2) < +\infty$, the fluctuations (also called standard deviations) around the mean are given by the Central Limit Theorem:

$$\sqrt{n}\Big(\frac{1}{n} \sum_{1\leq i\leq n} X_i - \mathrm{E}(X_1)\Big) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \sigma^2) \qquad \text{weakly,} \tag{1.2}$$

where $\sigma^2 := \mathrm{V}(X_1) = \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2$. When the sum in (1.2) deviates from its mean by a quantity *of the same order than the mean* (also called a large deviation *), the question of the asymptotic behaviour is answered by the following theorem, known as *Cramér's theorem*, provided that the underlying r.v.'s satisfy a *finite exponential moment condition*.

Theorem 1.1 (Cramér's theorem on the real line)**.** *Suppose that, in addition to the previous assumptions made on X, we have*

$$\forall \lambda \in \mathbb{R}, \qquad M(\lambda) := \mathrm{E}(e^{\lambda X_1}) < +\infty. \tag{1.3}$$

Then, for all $a > \mathrm{E}(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\Big(\sum_{1\le i\le n} X_i \ge an \Big) = -I(a), \tag{1.4}$$

where

$$I(a) := \sup_{\lambda\in\mathbb{R}} \{a\lambda - \log M(\lambda)\} \in [0,\infty]. \tag{1.5}$$

The function M in (1.3) is called the *moment generating function*, while $\log M$ is often referred to as the *logarithmic moment generating function*. The function I is called the *rate function*† and the formula in (1.5) shows that it coincides with the so-called *Fenchel-Legendre transform* (or conjugate) of $\log M$. The rate function quantifies the *unlikelihood* or *cost* of a large deviation. We will see later that for upper deviations ($a > \mathrm{E}(X_1)$) the supremum in (1.5) may be restricted to $\lambda \ge 0$.

Exercise 2. Show that (1.3) implies that $\mathrm{E}(|X_1|^k) < +\infty$ for every $k \in \mathbb{N}$.

Exercise 3. For the most common probability distributions, determine whether (1.3) is satisfied or not.

Proof of Theorem 1.1. We first prove the upper bound corresponding to (1.4).

(1) Prove that l.h.s. (1.4) $\le e^{-\lambda a n} M(\lambda)^n$ for all $\lambda \ge 0$. Hint: use the Markov inequality with an ad-hoc non-negative random variable.

(2) Conclude.

We continue with the lower bound. First, we introduce for every $\lambda \in \mathbb{R}$ a new probability distribution P_λ under which the X_i's remain i.i.d. and such that for every Borel set $A \subseteq \mathbb{R}$,

$$\mathrm{P}_\lambda(X_1 \in A) = \mathrm{E}\Big(\frac{e^{\lambda X_1}}{M(\lambda)} 1_{\{X_1 \in A\}} \Big). \tag{1.6}$$

(1) Show that M and $\log M$ are twice differentiable on $\mathbb{R}$ and that for every $\lambda \in \mathbb{R}$, $\mathrm{E}_\lambda(X_1) = (\log M)'(\lambda)$ and $\mathrm{V}_\lambda(X_1) = (\log M)''(\lambda)$.

(2) Prove that for every $\lambda \ge 0$ and $b > 0$,

$$\text{l.h.s. (1.4)} \ge e^{-\lambda[an + b\sqrt{n}]} M(\lambda)^n \mathrm{P}_\lambda\Big(an \le \sum_{1\le i\le n} X_i \le an + b\sqrt{n} \Big). \tag{1.7}$$

(3) Conclude. Hint: make an appropriate choice of λ then use the Central Limit Theorem. It might be necessary to restrict to values of a that lie strictly below the supremum value in the support of $\mathrm{P}(X_1 \in \cdot)$.

---

*A more rigorous and general definition will follow in Section 3.

†Here we use the term *rate function* quite loosely: a formal definition will be given in Section 3.1.

■

We have just met two important techniques for estimating large deviations:

- the *Chernoff bound* (for the upper bound);
- the *change of measure argument* (for the lower bound) a.k.a. *tilting*.

The law P_λ introduced in (1.6) is called a *tilted* or *biased* law (with respect to P). It may seem at first quite fortunate that this particular (linear) change of measure gives a sharp estimate for the large deviations. We will see in Exercise 10 that there is a deeper reason, apart from the convenient fact the normalizing constant is explicit in terms of the moment generating function.

Exercise 4 (Lower deviations have feelings too)**.** Show that the statement of Theorem 1.1 may be extended to *lower* deviations, i.e. for the probability that $\sum_{1\le i\le n} X_1 \le an$, when $a < \mathrm{E}(X_1)$.

Exercise 5. Compute the rate function I for all well-known probability distributions which satisfy the finite exponential moment condition in (1.3).

Exercise 6 (Regularity and convexity of $\log M$ and I)**.** (1) Show that M and $\log M$ are smooth (i.e. infinitely differentiable).

(2) Show that $\log M$ is convex. When is it strictly convex?

(3) Is the rate function I always finite? When it is finite, is it smooth? Is it (strictly) convex?

(4) On a schematic representation of $\log M$, indicates how to find the quantity $I(a)$ using the idea used for proving the lower bound in Theorem 1.1.

Exercise 7 (Tilting and stochastic domination)**.** (1) Let $f, g\colon \mathbb{R} \to \mathbb{R}$ be two non-increasing functions. Show that for every real-valued r.v. Z,

$$\mathrm{E}(f(Z)g(Z)) \ge \mathrm{E}(f(Z))\mathrm{E}(g(Z)), \tag{1.8}$$

provided that the three expectations are well-defined. We say that $f(Z)$ and $g(Z)$ are *positively correlated*. Hint: consider the random variable $[f(Z_1) - f(Z_2)][g(Z_1) - g(Z_2)]$, where Z_1 and Z_2 are two independent copies (a.k.a. *replicas*) of Z.

(2) (Same notation as in the proof of Theorem 1.1). Show that for every $\lambda > 0$ and $x \in \mathbb{R}$, $P_\lambda(X_1 \le x) \le P(X_1 \le x)$. We say that P_λ stochastically dominates P. What if $\lambda < 0$?

2. Sanov's theorem on a finite state space and relative entropy

In this section we denote by $X = (X_i)_{i\in\mathbb{N}}$ a sequence of i.i.d. r.v.'s on a *finite* state space $\Sigma = \{x_1, \dots, x_r\}$ ($r \in \mathbb{N}$). Using Cramér's theorem (Theorem 1.1) one may compute the large deviation probabilities for the empirical frequency $(1/n)\mathrm{card}\{1 \le i \le n\colon X_i = x\}$, for any $x \in \Sigma$ (try and do it as an exercise). In this section, we will derive (via elementary methods) the large deviation probabilities for all empirical frequencies *simultaneously*, i.e. for the (empirical) *occupation measure*

$$L_n := \frac{1}{n} \sum_{1\le i\le n} \delta_{X_i}, \tag{2.1}$$

that is, with a slight abuse of notation, a random element of

$$\mathcal{M}_1(\Sigma) := \{\nu := (\nu_1, \dots, \nu_r) \in [0,1]^r \colon \nu_1 + \dots + \nu_r = 1\} \tag{2.2}$$

This is the space of probability measures on Σ, that we see as a compact subset of $\mathbb{R}^r$, which we endow with the following metric (also called the *total variation distance*):

$$d(\nu,\mu) = \frac{1}{2}\sum_{x\in\Sigma} |\mu_x - \nu_x|, \qquad \mu,\nu \in \mathcal{M}_1(\Sigma). \tag{2.3}$$

Open balls shall be denoted by $B(\nu, r)$, with ν the center and $r > 0$ the radius. Before considering deviations of L_n, let us notice that its mean is given by the law of X_1, as for every $x \in \Sigma$ (with a slight abuse of notation)

$$\mathrm{E}[L_n(x)] = \frac{1}{n}\sum_{1\le i\le n} \mathrm{P}(X_i = x) = \mathrm{P}(X_1 = x). \tag{2.4}$$

Theorem 2.1 (Sanov's theorem on a finite state space)**.** *Let $\mu \in \mathcal{M}_1(\Sigma)$ be the law of X_1. For every $\nu \in \mathcal{M}_1(\Sigma)$ and $r > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}(L_n \in B(\nu,r)^c) = -\inf_{B(\nu,r)^c} h(\,\cdot\,|\mu), \tag{2.5}$$

where

$$h(\nu|\mu) := \sum_{x\in\Sigma} \nu_x \log\left(\frac{\nu_x}{\mu_x}\right) \ge 0, \qquad \nu \in \mathcal{M}_1(\Sigma), \tag{2.6}$$

denotes the relative entropy *of ν with respect to μ.*

Remark 2.2. *In order for (2.6) to be well-defined, we may assume w.l.o.g. that $\mu_x > 0$ for every $x \in \Sigma$. In general, the relative entropy of ν w.r.t. to μ is defined as in (2.6) when $\nu \ll \mu$ (with the convention $0 \log 0 = 0$) and as $+\infty$ otherwise. The non-negativity of the relative entropy is a consequence of the convexity of $x \mapsto x \log x$ on $[0,\infty)$.*

Relative entropy is a ubiquitous notion that can be found in probability theory, analysis, information theory and statistical physics. Although it does not define a metric (the reader may check that it is not symmetric) it quantifies how much a measure is close to another one. The infimum in (2.5) indicates that *among all unlikely scenarii*, the large deviation event is achieved by *the least unfavorable one*, an idea that is reminiscent of the *least action principle* in physics or the *Laplace principle* for evaluating the asymptotic behaviour of certain integrals.

Proof of Theorem 2.1. This is taken from [2]. For all $n \in \mathbb{N}$, L_n takes values in

$$K_n := \mathcal{M}_1(\Sigma) \cap (\tfrac{\mathbb{N}_0}{n})^r. \tag{2.7}$$

(1) Using Stirling's formula, show that for all $\nu \in K_n$,

$$\mathrm{P}(L_n = \nu) = n! \prod_{1\le s\le r} \frac{\mu_s^{n\nu_s}}{(n\nu_s)!} = \exp\Big(-nh(\nu|\mu) + O(\log n)\Big), \tag{2.8}$$

where the term in $O(\log n)$ is uniform on K_n.

(2) Clearly, for all $A \subseteq \mathcal{M}_1(\Sigma)$,

$$\max_{\nu\in K_n\cap A} \mathrm{P}(L_n = \nu) \le \mathrm{P}(L_n \in A) \le |K_n| \max_{\nu\in K_n\cap A} \mathrm{P}(L_n = \nu). \tag{2.9}$$

Evaluate $|K_n|$ and conclude.

∎

The chain of inequalities in (2.9) is typical of Large Deviation bounds: the lower bound is obtained by picking the most likely scenario, while for the upper bound one needs to consider all other cases. In this simple example, it is enough to compute the total number of possible outcomes. In other cases, more elaborate *entropy reduction techniques* may be necessary.

Unless stated otherwise, the notation used in the exercises below is the same as the one used in the beginning of this section.

Exercise 8. Show that the estimate in (2.8) can be made more precise, namely, for every $n \in \mathbb{N}$ and $\nu \in K_n$,

$$(n+1)^{-|\Sigma|} e^{-nh(\nu|\mu)} \le \mathrm{P}(L_n = \nu) \le e^{-nh(\nu|\mu)}. \tag{2.10}$$

See [1, Lemma 2.1.9].

Exercise 9. Recalling that $\mu = \mathrm{P}(X_1 \in \cdot)$, prove that

$$\mathrm{P}(L_n = \mu) \ge \mathrm{P}(L_n = \nu), \qquad \nu \in K_n. \tag{2.11}$$

Exercise 10 (Important exercise: towards the contraction principle)**.** Let $(X_i)_{i\in\mathbb{N}}$ be as above, with Σ finite. Let $f\colon \Sigma \to \mathbb{R}$ and $Y_i = f(X_i)$ for every $i \in \mathbb{N}$. The goal of this exercise is to estimate the large deviations for the empirical mean of the Y_i's via Sanov's theorem and compare it with Cramér's theorem (as well as enlighten its proof).

(1) Assume for simplicity that $\mathrm{E}(Y_1) = 0$ and let $\varepsilon > 0$. Prove that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathrm{P}\Big(\sum_{1\le i\le n} Y_i \ge \varepsilon n \Big) = -\inf \Big\{ h(\nu|\mu) \colon \nu \in \mathcal{M}_1(\Sigma),\ \langle f, \nu\rangle \ge \varepsilon \Big\}, \tag{2.12}$$

where $\langle f, \nu \rangle := \int_\Sigma f \mathrm{d}\nu$.

(2) Show that the infimum in (2.12) is attained for a unique probability measure ν which corresponds to a tilted measure similar as the one that was introduced in the proof of Theorem 1.1.

(3) Finally, show that the r.h.s. in (2.12) coincides with what we would get by applying Cramér's theorem.

We have been able to derive the large deviation asymptotic behaviour of $\langle f, L_n\rangle$ from that of L_n. This is a general idea known as *contraction principle* that will be formalized in Section 3.3.

Exercise 11 (A general lower bound using relative entropy)**.** Prove that, for $\nu(A) > 0$,

$$\log\Big(\frac{\mu(A)}{\nu(A)}\Big) \ge -\frac{1}{\nu(A)}\Big(h(\nu|\mu) + \frac{1}{e}\Big). \tag{2.13}$$

Hint: use the convexity of $x \mapsto x\log x$ on $[0,\infty)$.

Exercise 12. For every $\mu \in \mathcal{M}_1(\Sigma)$ we let $\mu^{\otimes n}$ denote the law of $(X_1, \dots, X_n)$ on Σ^n, where the X_i's are replicas of $X_1 \sim \mu$.

(1) Show that for every $n \in \mathbb{N}$, $h(\nu^{\otimes n}|\mu^{\otimes n}) = nh(\nu|\mu)$.

(2) Give an alternative proof of the lower bound in Sanov's theorem using (1) and Exercise 11.

Exercise 13 (A challenging exercise: large deviations for pair occupation measures)**.** From [2, Section II.2]. Let $\mu = \mathrm{P}(X_1 \in \cdot) \in \mathcal{M}_1(\Sigma)$ with Σ finite. The goal of this exercise is to provide an analogue of Theorem 2.1 for the *pair* occupation measure (with periodic boundary conditions) defined as

$$L_n^{(2)} = \frac{1}{n} \sum_{1 \le i \le n} \delta_{(X_{i[n]}, X_{i+1[n]})}, \qquad n \in \mathbb{N}. \tag{2.14}$$

Observe that the pair occupation measure belongs to the set

$$\widetilde{\mathcal{M}}_1(\Sigma \times \Sigma) := \Big\{ \nu \in \mathcal{M}_1(\Sigma \times \Sigma) \colon \sum_{y \in \Sigma} \nu_{xy} = \sum_{y \in \Sigma} \nu_{yx}, \ \forall x \in \Sigma \Big\}. \tag{2.15}$$

For every $\nu \in \widetilde{\mathcal{M}}_1(\Sigma \times \Sigma)$, we shall denote by $\bar{\nu}$ the (common) marginal distribution on Σ. Let us now define

$$K_n^{(2)} := \Big\{ k = (k_{xy}) \in \mathbb{N}_0^{\Sigma \times \Sigma} \colon \sum_{x,y \in \Sigma} k_{xy} = n, \ \sum_{y \in \Sigma} k_{xy} = \sum_{y \in \Sigma} k_{yx} \Big\}. \tag{2.16}$$

For every $k \in K_n^{(2)}$, denote by $G(k)$ the oriented graph with Σ as its set of vertices and k_{xy} distinct edges from $x \in \Sigma$ to $y \in \Sigma$. An *Euler circuit* of $G(k)$ is a looped path respecting the arrows and making use of each arrow precisely once. Two looped paths that coincide up to a cyclic time shift count as the same Euler circuit.

(1) Prove that

$$\mathrm{P}\Big(L_n^{(2)}(x,y) = \frac{k_{xy}}{n}, \ \forall x, y \in \Sigma \Big) = \sum_{C \in \mathcal{E}(k)} \frac{m(C)}{\prod_{x,y \in \Sigma} k_{xy}!} \prod_{x \in \Sigma} \mu_x^{\bar{k}_x}, \tag{2.17}$$

where $\mathcal{E}(k)$ is the set of Euler circuits on $G(k)$, $m(C)$ is the number of cyclic shifts of C giving distinct paths $(X_1, \dots, X_n)$, and $\bar{k}_x = \sum_y k_{xy}$.

(2) Prove that for every $k \in K_n^{(2)}$,

$$\prod_{x \in \Sigma \colon \bar{k}_x > 0} (\bar{k}_x - 1)! \le \mathrm{card}(\mathcal{E}(k)) \le \prod_{x \in \Sigma} \bar{k}_x! \tag{2.18}$$

(3) Conclude, using Stirling's formula, that for every $\nu_0 \in \widetilde{\mathcal{M}}_1(\Sigma \times \Sigma)$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathrm{P}(L_n^{(2)} \in B(\nu_0, r)^c) = - \inf_{\nu \in B(\nu_0, r)^c} h(\nu | \bar{\nu} \otimes \mu), \tag{2.19}$$

where $B(\nu_0, r)$ is the ball on $\widetilde{\mathcal{M}}_1(\Sigma \times \Sigma)$ with center ν_0 and radius $r > 0$, for the total variation distance.

As an application of (2.19) we will give an alternative proof of the large deviation principles for finite Markov chains in Exercise 21.

3. Large Deviations: general principles

We already gave a taste of what large deviation principles are with Cramér and Sanov's theorems in Sections 1 and 2. It is now time to turn to general principles.

3.1. **Large Deviation principles.** Let $\mathcal{X}$ be a metric space. A function $I\colon \mathcal{X} \to [0,\infty]$ is said to be a *rate function* if it is *lower semi-continuous*, i.e. if for every $x_n \to x \in \mathcal{X}$,

$$\liminf_{n\to\infty} I(x_n) \geq I(x). \tag{3.1}$$

The following exercise gives another useful characterization of lower semi-continuity.

Exercise 14. Show that $f\colon \mathcal{X} \to [-\infty,\infty]$ is lower semi-continuous if and only if f has *closed level sets*, i.e. $\{f \leq \alpha\} := \{x \in \mathcal{X}\colon f(x) \leq \alpha\}$ is closed for every $\alpha \in \mathbb{R}$.

A function $I\colon \mathcal{X} \to [0,\infty]$ is said to be a *good rate function* if it has *compact* level sets.

Exercise 15. Show that a lower semi-continuous function achieves its infimum on any compact set.

Good rate functions will play a role in Section 3.3. ‡

Let us equip $\mathcal{X}$ with its Borel σ-algebra $\mathcal{B}_{\mathcal{X}}$.

Definition 3.1. *We say that a sequence of probability measures (μ_n) on $(\mathcal{X},\mathcal{B}_{\mathcal{X}})$ satisfies a Large Deviation Principle (LDP) with speed (a_n) and rate function $I\colon \mathcal{X} \to [0,\infty]$ if*

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{a_n} \log \mu_n(O) \geq -\inf_{x\in O} I(x), && \textit{for all open sets } O,\\ &\limsup_{n\to\infty} \frac{1}{a_n} \log \mu_n(F) \leq -\inf_{x\in F} I(x). && \textit{for all closed sets } F. \end{aligned} \tag{3.2}$$

The infimum over empty sets is understood as $+\infty$.

Exercise 16 (Uniqueness of the rate function)**.** (1) Prove that a function $f\colon \mathcal{X} \to [-\infty,\infty]$ is lower semi-continuous if and only if for every $x \in \mathcal{X}$,

$$f(x) = \lim_{r\to 0} \inf\{f(y)\colon d(x,y) < r\}. \tag{3.3}$$

(2) Prove that the rate function of an LDP is unique.

Definition 3.1 may also be applied to *random variables* instead of probability measures: a sequence of random variables is said to satisfy an LDP if the corresponding sequence of probability distributions (or laws) does. We remark an analogy with weak convergence (see the Portmanteau theorem) as one asks for a *lower bound* for *open sets* and an *upper bound* for *closed sets*. In general, the limit for an arbitrary set may or may not exist. Theorems 1.1 and 2.1 may now be reformulated in terms of a proper LDP:

Theorem 3.2 (Cramér's theorem: LDP reformulation)**.** *Let $X = (X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. real r.v.'s with finite exponential moments, i.e. $M(\lambda) := \mathrm{E}(e^{\lambda X_1})$ is finite for every $\lambda \in \mathbb{R}$. Then, the sequence $(\frac{1}{n}\sum_{i=1}^n X_i)_{n\in\mathbb{N}}$ satisfies an LDP on $\mathbb{R}$ (equipped with the usual topology) with speed n and rate function*

$$I(a) := \sup_{\lambda\in\mathbb{R}}\{a\lambda - \log M(\lambda)\} \in [0,\infty], \qquad a \in \mathbb{R}. \tag{3.4}$$

Theorem 3.3 (Sanov's theorem: LDP reformulation)**.** *Let $X = (X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. r.v.'s valued in a finite set Σ, with common marginal distribution $\mu \in \mathcal{M}_1(\Sigma)$. Then, the sequence of occupation measures $(L_n)_{n\in\mathbb{N}}$ as defined in (2.1) satisfies an LDP on $\mathcal{M}_1(\Sigma)$ (equipped with the total variation distance) with speed n and rate function $h(\cdot|\mu)$, that is the relative entropy functional defined in (2.6).*

‡Be careful that in some references, including [2], good rate functions are actually called rate functions, and rate functions are called weak rate functions.

We will prove Theorem 3.2 and leave the proof of Theorem 3.3 to the reader. Beforehand, we need the following definition:

Definition 3.4 (Weak LDP)**.** *A sequence of probability measures is said to satisfy a* weak *LDP when the lower bound holds for open sets and the upper bound holds for* compact *sets instead of closed sets.*

In contrast with the weak version, the LDP from Definition 3.1 is sometimes referred to as *strong* or *full* LDP.

Proof of Theorem 3.2 (weak version). To simplify, we assume w.l.o.g. that $\mathrm{E}(X_1) = 0$. Prove the lower bound for open intervals first, then for any open set (hint: use an approximate minimizer). As for the upper bound, we prove it for compact sets first (use that from any covering of a compact set by open balls one may extract a finite covering). ∎

We will see in the following section how to obtain the (strong/full) LDP from the weak one using the notion of *exponential tightness*.

3.2. **Exponential tightness.** The notion of *exponential tightness* is an "exponential version" of the notion of tightness used for weak convergence of measures.

Definition 3.5. *We say that a sequence of probability measures (μ_n) on a metric space $\mathcal{X}$ is exponentially tight (on a scale a_n) if for every $A > 0$, there exists a compact set $K_A \subseteq \mathcal{X}$ such that:*

$$\limsup_{n\to\infty} \frac{1}{a_n} \log \mu_n(K_A^c) \leq -A. \tag{3.5}$$

Proposition 3.6. *If a sequence of probability measures satisfies a weak LDP and is exponentially tight (on a common scale) then it also satisfies the strong/full LDP.*

Proof of Proposition 3.6. Hint: Let $F \subseteq \mathcal{X}$ be a non-empty closed set and $x \in F$ an arbitrary element. Let $A := I(x)$ and decompose $F = (F \cap K_A) \cup (F \cap K_A^c)$ where K_A is the compact set provided by the definition of exponential tightness. ∎

We may now complete the proof of Theorem 3.2.

Exercise 17. Prove that the sequence $(\frac{1}{n}\sum_{i=1}^n X_i)_{n\in\mathbb{N}}$ from Theorem 3.2 is exponentially tight (on the scale n) and conclude the proof of Theorem 3.2 (strong/full LDP).

3.3. **The contraction principle.** We have already met the idea behind the contraction principle in Exercise 10. Here is a general statement.

Theorem 3.7 (Contraction principle)**.** *If (μ_n) satisfies an LDP with* good *rate function I on a metric space $\mathcal{X}$ and if T is a* continuous *map from $\mathcal{X}$ to another metric space $\mathcal{Y}$, then $(\mu_n \circ T^{-1})$ satisfies an LDP with* good *rate function J given by*

$$J(y) = \inf_{x\in\mathcal{X}:\, T(x)=y} I(x), \qquad (\inf\emptyset = +\infty). \tag{3.6}$$

Proof of Theorem 3.7. First, prove the lower and upper large deviation bounds using the fact that T^{-1} maps open sets into open sets and closed sets into closed sets. Then, prove that for every $\alpha \in \mathbb{R}$, $\{J \leq \alpha\} = T(\{I \leq \alpha\})$ and conclude that J is a good rate function. ∎

The next exercise shows that Theorem 3.7 (partly) fails without assuming goodness of the rate function.

Exercise 18. Let $(U_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that U_n is uniformly distributed on $[-n, n]$.

(1) Show that (U_n) satisfies a LDP with (trivial) rate function $I(x) = 0$ for $x \in \mathbb{R}$.

(2) Show that I is not a good rate function.

(3) Let $Z_n := \exp(U_n)$. Show that (Z_n) satisfies lower and upper large deviation bounds but that the function corresponding to (3.6) with $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and $T = \exp$ is *not* a rate function.

3.4. **Varadhan's lemma.** The next lemma allows to obtain the asymptotic behaviour of certain exponential moments when the underlying random variables satisfy an LDP. This is extremely useful for instance in statistical mechanics, see Exercise 19 below.

Theorem 3.8 (Varadhan's lemma)**.** *Let (μ_n) be a sequence of probability measures on $\mathcal{X}$ that satisfies a LDP with speed n and rate function I. Let $F\colon \mathcal{X} \mapsto \mathbb{R}$ be a* continuous *function that is* bounded from above. *Then,*

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)} \mu_n(\mathrm{d}x) = \sup_{x\in\mathcal{X}} \{F(x) - I(x)\}. \tag{3.7}$$

Proof of Theorem 3.8. Let us start with the lower bound.

(1) Let $x_0 \in \mathcal{X}$. Using the open set $O_{x_0,\varepsilon} := F^{-1}((F(x_0) - \varepsilon, \infty))$, prove that

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} e^{nF(x)} \mu_n(\mathrm{d}x) \geq F(x_0) - I(x_0) \tag{3.8}$$

and conclude.

We continue with the upper bound. Let $b := \sup F$ and $a := \sup[F - I]$.

(2) Show that $-\infty < a \leq b < \infty$.

(3) Let $C := F^{-1}([a, b])$ and for $N \in \mathbb{N}$, $1 \leq i \leq N$, define

$$C_i^N := F^{-1}([c_{i-1}^N, c_i^N]), \qquad \text{where} \quad c_i^N := a + \tfrac{i}{N}(b - a). \tag{3.9}$$

Prove that

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_C e^{nF(x)} \mu_n(\mathrm{d}x) \leq \max_{1\leq i\leq N} [c_i^N - I(C_i^N)]. \tag{3.10}$$

(4) Show that

$$\inf_{C_i^N} F - \inf_{C_i^N} I \leq \sup_{C_i^N} [F - I]. \tag{3.11}$$

(5) Deduce thereof that

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_C e^{nF(x)} \mu_n(\mathrm{d}x) \leq \sup_C [F - I] + \frac{b - a}{N} \tag{3.12}$$

and conclude.

■

Exercise 19 (The Curie-Weiss model)**.** Let $\sigma = (\sigma_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. r.v.'s uniformly distributed on $\{-1, 1\}$ and β a positive parameter. The *partition function* of the Curie-Weiss model for a system of size $n \in \mathbb{N}$ and *inverse temperature* β is

$$Z_{n,\beta} = \mathrm{E}[\exp(\beta H_n(\sigma))], \qquad \text{where} \quad H_n(\sigma) := \frac{1}{n} \sum_{1\leq i<j\leq n} \sigma_i \sigma_j. \tag{3.13}$$

(1) Show that $H_n(\sigma) = \frac{1}{2}(nM_n(\sigma)^2 - 1)$, where $M_n(\sigma) := \frac{1}{n}\sum_{i=1}^n \sigma_i$ is the *mean magnetization.*

(2) Using Varadhan's lemma, prove that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log Z_{n,\beta} = f(\beta) := \sup_{m\in[-1,1]} \left\{ \tfrac{1}{2}\beta m^2 - I(m) \right\}, \\ &\text{where} \quad I(m) := -\frac{1+m}{2} \log\left(\frac{1+m}{2}\right) - \frac{1-m}{2} \log\left(\frac{1-m}{2}\right). \end{aligned} \tag{3.14}$$

(3) Investigate the (existence and values of) maximizers in the variational formula above. What do you notice? (This is an example of a *phase transition* in statistical mechanics).

4. Large Deviations for finite Markov chains

We show in this section that Sanov's theorem for the occupation measure of i.i.d. random variables (Theorem 2.1) may be extended to finite Markov chains. Let $X = (X_i)_{i\in\mathbb{N}_0}$ be an irreducible Markov chain on a *finite* state space Σ, with transition matrix $Q = (Q_{xy})_{x,y\in\Sigma}$ and unique stationary probability distribution $\pi = (\pi_x)_{x\in\Sigma}$. The initial distribution is arbitrary. We assume for simplicity that $Q_{xy} > 0$ for all $x, y \in \Sigma$. Recall the definition of L_n in (2.1).

Theorem 4.1 (LDP for the occupation measure of finite Markov chains)**.** *The sequence $(L_n)_{n\in\mathbb{N}}$ satisfies a LDP on $\mathcal{M}_1(\Sigma)$ with speed n and rate function:*

$$I(\nu) = \sup_{u>0} - \sum_{x\in\Sigma} \nu_x \log\left[\frac{(Qu)_x}{u_x}\right], \tag{4.1}$$

where $u > 0$ denotes the set of positive *functions $u\colon \Sigma \to (0,\infty)$.*

Exercise 20. Check that Theorem 4.1 is consistent with Theorem 2.1.

The arguments for the proof are borrowed from [3]. Other lines of proof are suggested in the exercises at the end of this section.

Proof of Theorem 4.1. We start by proving that I is a rate function.

(1) Show that I is a non-negative function.

(2) Show that I is lower semi-continous. Hint: it writes as the supremum of a collection of continuous functions.

Let us now establish the upper bound. Let u be a positive function on Σ.

(3) Show that the sequence $(M_n)_{n\in\mathbb{N}_0}$ defined by

$$\begin{aligned} &M_n = u(X_n) \exp\Big(- \sum_{0\le i<n} W(X_i) \Big), \\ &\text{where} \quad W(x) := \log \frac{(Qu)_x}{u_x}, \qquad x\in\Sigma, \end{aligned} \tag{4.2}$$

(empty sum is zero) is a martingale (for the natural filtration).

(4) Show that there exists a constant $C > 0$ such that for all $x \in \Sigma$

$$\mathrm{E}_x\Big[\exp\Big(-n\int_\Sigma W(y)L_n(\mathrm{d}y)\Big)\Big] \leq C, \tag{4.3}$$

where P_x denotes the law of Markov chain started at x.

(5) Deduce thereof that for every Borel set $A \subseteq \mathcal{M}_1(\Sigma)$,

$$\mathrm{P}_x(L_n \in A) \leq C\exp\Big(n\sup_{\nu\in A}\sum_{y\in\Sigma}\nu_y W(y)\Big), \tag{4.4}$$

and finally

$$\mathrm{P}_x(L_n \in A) \leq C\exp\Big(-n\sup_{u>0}\inf_{\nu\in A} -\sum_{y\in\Sigma}\nu_y\log\Big[\frac{(Qu)_y}{u_y}\Big]\Big). \tag{4.5}$$

It remains to interchange the supremum and the infimum when A is closed, which we henceforth assume. To this end, we adapt a compactness argument from [4]. For convenience, let us rewrite (4.5) as

$$\mathrm{P}_x(L_n \in A) \leq C\exp\Big(-n\sup_{u>0}\inf_{\nu\in A}\langle\nu, W(u)\rangle\Big). \tag{4.6}$$

where $\langle\cdot\rangle$ is the Euclidian scalar product on $\mathbb{R}^d$ and, slightly changing notation,

$$W(u)_y := -\log\Big[\frac{(Qu)_y}{u_y}\Big], \qquad \forall y \in \Sigma. \tag{4.7}$$

Let $\ell \geq 0$ be the infimum of I over A, and $\varepsilon > 0$.

(6) Show that for every $\nu \in A$, there exists $u_\nu > 0$ such that $\langle\nu, W(u_\nu)\rangle \geq \ell - \varepsilon$.

(7) Show that for every $\nu \in A$, there exists an open neighborhood of ν in $\mathcal{M}_1(\Sigma)$, denoted by $N(\nu)$, such that

$$\forall\mu \in N(\nu), \qquad \langle\mu, W(u_\nu)\rangle \geq \ell - 2\varepsilon. \tag{4.8}$$

(8) Show that there exists a finite collection $\nu_1, \ldots, \nu_k$ in $\mathcal{M}_1(\Sigma)$ such that the union of the $N(\nu_i)$'s contains A. Hint: A is a closed set of the compact space $\mathcal{M}_1(\Sigma)$.

(9) Deduce thereof that

$$\limsup_{n\to\infty}\tfrac{1}{n}\log\mathrm{P}_x(L_n \in A) \leq -\min_{1\leq i\leq k}\sup_{u>0}\inf_{\mu\in N(\nu_i)}\langle\mu, W(u)\rangle. \tag{4.9}$$

and conclude. Hint: recall Exercise 1 and (4.8).

This concludes the proof of the upper bound.

We finally turn to the lower bound. Suppose A is open and let $\nu \in A$ be fixed.

(10) Show that there exists $\varepsilon > 0$ such that

$$\mathrm{P}(L_n \in A) \geq \mathrm{P}(L_n \in B(\nu, \varepsilon)). \tag{4.10}$$

(11) Show that the supremum in (4.1) is achieved by some positive function u_0, which we fix in the rest of the proof. Hint: observe that the function $W(u)$ defined in (4.7) is invariant under the maps $u \mapsto \lambda u$ for all $\lambda \in (0, \infty)$.

(12) Show that ν is the invariant probability distribution for the Markov chain with transition matrix $\widetilde{Q}$ defined as

$$\frac{\widetilde{Q}(x,y)}{Q(x,y)} = \frac{u_0(y)}{(Qu_0)(x)}, \qquad x, y \in \Sigma. \tag{4.11}$$

(13) Let $\widetilde{\mathrm{P}}_x$ denote the law of the Markov chain started at $x \in \Sigma$ with transition matrix $\widetilde{Q}$. Prove that there exists a constant $C > 0$ such that

$$\mathrm{P}_x(L_n \in B(\nu,\varepsilon)) \geq C \exp\big[-n\big(\langle \nu, W(u_0)\rangle + \varepsilon \|W(u_0)\|_\infty\big)\big] \widetilde{\mathrm{P}}_x(L_n \in B(\nu,\varepsilon)), \tag{4.12}$$

and conclude using the ergodic theorem for irreducible finite Markov chains.

■

Remark 4.2. *The proof of the upper bound shares similarities with the Chernoff bound that was used to derive the upped bound in Cramér's theorem: here the random variable under consideration is the occupation measure instead of the empirical mean and the relevant linear functional is integration against the functions W defined in (4.2) instead of scalar multiplication. The proof of the lower bound uses a change of measure argument, as we did for proving Cramér's theorem.*

Remark 4.3 (An important step in the proof of the upper bound)**.** *In order to obtain (4.9), it is crucial to use the result from Exercise 1 before applying the intermediate estimate in (4.6) to each $N(\nu_i)$ for $1 \leq i \leq k$. If instead we directly apply (4.6) to the union of the $N(\nu_i)$'s, we would get $\sup_{u>0} \min_{1\leq i\leq k} \inf_{\mu \in N(\nu_i)}(\dots)$ instead of $\min_{1\leq i\leq k} \sup_{u>0} \inf_{\mu\in N(\nu_i)}(\dots)$ in the r.h.s. of (4.9). That would not be enough, since it is not possible to interchange $\sup_{u>0}$ and $\min_{1\leq i\leq k}$ in general.*

Exercise 21 (Alternative proof of the LDP for Markov chains using the pair occupation measure)**.** The goal of the exercise is to derive the LDP for the pair occupation measure of finite Markov chains from that of i.i.d. random variables, see Exercise 13. Let us recall the definition of the pair occupation measure $L_n^{(2)}$ defined in (2.14). To avoid any confusion, we let P_x be the law of an irreducible Markov chain started at $x \in \Sigma$ with transition matrix Q and invariant probability distribution π, and $\mathrm{P}_{\text{i.i.d.}}$ be the law of a sequence of i.i.d. random variables with common distribution π. We have (almost) proved in Exercise 13 that under $\mathrm{P}_{\text{i.i.d.}}$, $L_n^{(2)}$ satisfies an LDP on $\widetilde{\mathcal{M}}_1(\Sigma\times\Sigma)$ with speed n and rate function $h(\nu|\bar{\nu}\otimes\pi)$.

(1) Show that for every $A \subseteq \Sigma^n$ and $x \in \Sigma$,

$$\mathrm{P}_x((X_1,\dots,X_n)\in A) = \mathrm{E}_{\text{i.i.d.}}\Big[\exp\Big(O(1) + \sum_{1\leq i\leq n} \log \frac{Q(X_{i-1},X_i)}{\pi(X_i)}\Big) 1_{\{(X_1,\dots,X_n)\in A\}}\Big]. \tag{4.13}$$

(2) Deduce thereof that, under P_x, $L_n^{(2)}$ satisfies an LDP on $\widetilde{\mathcal{M}}_1(\Sigma\times\Sigma)$ with speed n and rate function

$$h(\nu|\bar{\nu}\otimes Q) := \sum_{s,t\in\Sigma} \log\Big(\frac{\nu_{st}}{\bar{\nu}_s Q_{st}}\Big)\nu_{st}. \tag{4.14}$$

Hint: recall the proof of Varadhan's lemma (Theorem 3.8).

(3) Using the contraction principle (Theorem 3.7), check that the rate function in (4.14) is consistent with Theorem 4.1.

Exercise 22 (Alternative proof of the LDP for Markov chains using Exercise 11)**.** Let P_n (resp. $\widetilde{\mathrm{P}}_n$) denote the law of $(X_1, \ldots, X_n)$ when X is an irreducible Markov chain on a finite state space Σ with transition matrix Q (resp. $\widetilde{Q}$) and invariant probability distribution π.

(1) Show that (recall the definition of relative entropy in (2.6)):

$$\lim_{n\to\infty} \frac{1}{n} h(\widetilde{\mathrm{P}}_n|\mathrm{P}_n) = \sum_{x,y\in\Sigma} \log\Big(\frac{\widetilde{Q}_{xy}}{Q_{xy}}\Big)\pi_x Q_{xy}. \tag{4.15}$$

(2) Using the estimate in Exercise 11, retrieve the large deviation lower bound corresponding to Theorem 4.1.

Let us close this section by mentioning another path to the LDP of finite Markov chains, which consists in deriving the asymptotic behaviour of moment generating functions for additive functionals of the Markov chain via the use of Perron-Frobenius eigenvalues (i.e. principal eigenvalues of irreducible matrices with positive entries), see [8] or [1, Section 3].

5. Further topics

5.1. **Sample path large deviations.** Since the first part of the course (on weak convergence) is devoted to Donsker's theorem, that is a functional version of the Central Limit Theorem, it is quite natural to ask about a functional version of Cramér's theorem. As in Section 1, we assume that $X = (X_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with all finite exponential moments and we recall the definition of the rate function I in (1.5). Define

$$S_0 = 0, \qquad S_k = X_1 + \ldots + X_k, \quad k \in \mathbb{N}. \tag{5.1}$$

Let us denote by $L_\infty([0,1])$ the space of real-valued and (essentially) bounded functions on $[0,1]$, $L_1([0,1])$ the space of integrable real-valued functions on $[0,1]$, and $C_{\mathrm{abs}}([0,1])$ the space of *absolutely continuous* real-valued functions on $[0,1]$. We recall that $\phi \in C_{\mathrm{abs}}([0,1])$ if and only if there exists a function in $L_1([0,1])$ denoted by ϕ' (with a slight abuse of notation since ϕ need not be everywhere differentiable) such that $\phi(t) = \int_0^t \phi'(s)\mathrm{d}s$ for every $0 \le t \le 1$.

Theorem 5.1 (Mogulskii)**.** *The sequence of random processes* $\big(\frac{1}{n}S_{\lfloor nt\rfloor}\big)_{0\le t\le 1}$ *satisfies an LDP in* $L_\infty([0,1])$ *(equipped with the supremum norm) with speed* n *and good rate function*

$$\mathfrak{I}(\phi) = \begin{cases} \int_0^1 I(\phi'(t))\mathrm{d}t & \text{if} \quad \phi \in C_{\mathrm{abs}}([0,1]) \quad \text{and} \quad \phi(0) = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{5.2}$$

The proof of Theorem 5.1 is outside the scope of these introductory notes (see [7] for a self-contained proof or [1, Section 5] for a *projective limit* approach). Nevertheless, as an exercise, we will prove an upper bound in order to enlighten the expression of Mogulskii's rate function.

Exercise 23. For convenience, let us note $Z_n(t) = \frac{1}{n}S_{\lfloor nt\rfloor}$ for $0 \le t \le 1$.

(1) Show that for every *fixed* $0 \le t \le 1$, the sequence $(Z_n(t))_{n\in\mathbb{N}}$ satisfies an LDP and identify its rate function.

Let $0 = t_0 < t_1 < \ldots < t_k = 1$ for $k \in \mathbb{N}$ be any subdivision of the unit interval, $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ and $\varepsilon > 0$.

(2) Prove that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathrm{P}\big(|Z_n(t_i) - a_i| \le \varepsilon,\ 1 \le i \le k\big) \le -\sum_{1\le i\le k} (t_i - t_{i-1}) \inf_{x\in C_i} I(x), \tag{5.3}$$

where $C_i := \{x \in \mathbb{R} \colon |(t_i - t_{i-1})x - (a_i - a_{i-1})| \le \varepsilon\}$.

Let $\phi \colon [0, 1] \to \mathbb{R}$ be continuous and differentiable on $[0, 1]$ (that is differentiable on $(0, 1)$ with left (resp. right) derivative at 1 (resp. 0).

(3) Show that

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathrm{P}\big(\|Z_n - \phi\|_{\infty,[0,1]} \le \varepsilon\big) \le -\int_0^1 I(\phi'(t))\mathrm{d}t. \tag{5.4}$$

Let us conclude this section with a special case: if X_1 is the standard Gaussian distribution on the real line then $I(x) = (1/2)x^2$ and the expression of the rate function obtained thereby in Theorem 5.1, that is $\mathfrak{I}(\phi) = (1/2) \int_0^1 \phi'(t)^2 \mathrm{d}t$ actually coincides with the rate function for the LDP of Brownian motion sample paths (Schilder's theorem). Results are available for diffusions as well (Freidlin-Wentzell theory). We refer to [8] and [1, Section 5] for a larger overview of sample path large deviations and their applications.

5.2. **Beyond independence: the Gärtner-Ellis theorem.** Apart from Section 4, these notes deal with large deviations related to a sample of *independent* and identically distributed random variables. It is however possible to go beyond that setup, as we will see now, using a result due to Gärtner [6] and Ellis [5]. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of real random variables (we stick to $d = 1$ for simplicity) and assume that the following limit exists for every $\lambda \in \mathbb{R}$,

$$M(\lambda) := \lim_{n\to\infty} \mathrm{E}(e^{n\lambda Z_n})^{1/n} \in [0, +\infty]. \tag{5.5}$$

Assume moreover that there exists a neighborhood of the origin where M is finite. If Z_n is the empirical mean of i.i.d. real random variables, then this definition of M is consistent with (1.3). We call $\log M$ the *limiting* logarithmic moment generating function and define, similarly as in (1.5)

$$I(a) := \sup_{\lambda\in\mathbb{R}} \{a\lambda - \log M(\lambda)\} \in [0, \infty], \qquad a \in \mathbb{R}. \tag{5.6}$$

The reader may check that I is a convex good rate function.

Theorem 5.2 (Gärtner-Ellis)**.** *Under the above conditions, we have, as $n \to \infty$,*

(1) $\limsup \frac{1}{n} \log \mathrm{P}(Z_n \in F) \le -\inf\{I(x),\, x \in F\}$*, for every $F \subseteq \mathbb{R}$ closed.*

(2) $\liminf \frac{1}{n} \log \mathrm{P}(Z_n \in O) \ge -\inf\{I(x),\, x \in O \cap E\}$*, for every $O \subseteq \mathbb{R}$ open, where*

$$E := \{x \in \mathbb{R} \colon \exists \tau \in \mathbb{R} \colon M(\tau) < \infty,\ I(y) - I(x) > (y - x)\tau,\ \forall y \ne x\}. \tag{5.7}$$

This way of obtaining a LDP by scaling of the moment generating functions is reminiscent of the Central Limit Theorem (CLT) obtained by scaling of the characteristic functions. The proof is quite close to that of Cramér's theorem, with some modifications needed for the lower bound. It is possible to remove the set E (the *exposed* points) from that lower bound under additional assumptions on the limiting logarithmic moment generating function, see [2, Theorem V.6] and [1, Exercise 2.3.20]. Finally, the LDP for i.i.d. random variables and Markov chains may be recovered from the Gärtner-Ellis theorem, see [2, Section V.4] and a more abstract version in topological vector spaces is available in [2, Section 4.5].

Proof. (1) Let $\delta > 0$. Show that for every $x \in \mathbb{R}$, there exists $\lambda(x) \in \mathbb{R}$ such that

$$\lambda(x)x - \log M(\lambda(x)) \geq \min(I(x) - \delta, \tfrac{1}{\delta}), \tag{5.8}$$

and a neighborhood of x, denoted by A_x, such that

$$\lambda(x)(y - x) \geq -\delta, \qquad \forall y \in A_x. \tag{5.9}$$

Then, show that

$$\mathrm{P}(Z_n \in A_x) \leq e^{n(\delta - \lambda(x)x)} \mathrm{E}(e^{n\lambda(x) Z_n}). \tag{5.10}$$

Deduce theoreof the large deviation upper bound for *compact* sets, and extend it to closed sets using exponential tightness, see Section 3.2.

(2) Let $x \in O \cap E$ and $\varepsilon > 0$ small enough so that $(x-\varepsilon, x+\varepsilon) \subseteq O$. By (5.7), there exists $\tau \in \mathbb{R}$ such that $M(\tau)$ is finite and $I(y) - I(x) > (y - x)\tau$ for every $y \neq x$. The idea is, as in the proof of Cramér's theorem, to obtain a lower bound on $\mathrm{P}(|Z_n - x| < \varepsilon)$ by using a change-of-measure argument. If $\mathrm{P}_n := \mathrm{P}(Z_n \in \cdot)$, then we define $\hat{\mathrm{P}}_n$ by

$$\frac{\mathrm{d}\hat{\mathrm{P}}_n}{\mathrm{d}\mathrm{P}_n}(y) = \frac{e^{n\tau y}}{\mathrm{E}(e^{n\tau Z_n})}, \qquad y \in \mathbb{R}. \tag{5.11}$$

The delicate part is proving that the interval $(x-\varepsilon, x+\varepsilon)$ becomes "typical" under $\hat{\mathrm{P}}_n$ without resorting to the CLT. Compute the limiting logarithmic moment generating function associated to the sequence of probability measures $(\hat{\mathrm{P}}_n)$ and, using the upper bound of Item (1), show that there exists $x_0 \neq x$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \hat{\mathrm{P}}_n((-\infty, x-\varepsilon] \cup [x+\varepsilon, \infty)) \leq -\hat{I}(x_0) < 0, \tag{5.12}$$

where $\hat{I}(y) := I(y) - y\tau + \log M(\tau)$. Conclude.

■

Exercise 24. Let $(X_i)_{i\in\mathbb{Z}}$ be a Gaussian stationary sequence of centered real random variables with *summable correlations*, i.e. $\sum_{n\geq 0} |\operatorname{Cov}(X_0, X_n)| < \infty$. Show that the empirical mean $\frac{1}{n}\sum_{1\leq i\leq n} X_i$ satisfies an LDP and identify the rate function.

Appendix A. Constrained optimisation

Theorem A.1. *Let Ω be a non-empty subset of $\mathbb{R}^n$ and*

$$U = \{x \in \Omega \colon g_i(x) = 0,\ 1 \leq i \leq p\}, \tag{A.1}$$

where $g_1, \dots, g_p$ are functions from Ω to $\mathbb{R}$. Let $f\colon \Omega \to \mathbb{R}$ and $a \in U$. Assume that

(1) *a is a local extremum of f in U;*
(2) *f is differentiable at a;*
(3) *$g_1, \dots, g_p$ are continuously differentiable in a neighborhood of a;*
(4) *$\nabla g_1(a), \dots, \nabla g_p(a)$ are linearly independent.*

Then, there exist p real numbers $\lambda_1, \dots, \lambda_p$, uniquely defined (and called the Lagrange multipliers*), such that*

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \ldots + \lambda_p \nabla g_p(a). \tag{A.2}$$

References

[1] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[2] Frank den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[3] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.*, 28:1–47; ibid. 28 (1975), 279–301, 1975.

[4] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Wiener integrals for large time. In *Functional integration and its applications (Proc. Internat. Conf., London, 1974)*, pages 15–33, 1975.

[5] Richard S. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, 12(1):1–12, 1984.

[6] Jurgen Gertner. On large deviations from an invariant measure. *Teor. Verojatnost. i Primenen.*, 22(1):27–42, 1977.

[7] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.

[8] S. R. S. Varadhan. Large deviations. *Ann. Probab.*, 36(2):397–419, 2008.