JACOPO CATTANEO

LARGE DEVIATION PRINCIPLE FOR SUM OF SUBEXPONENTIAL VARIABLES: THE 1D MATCHING PROBLEM

UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

LAUREA MAGISTRALE IN FISICA

LARGE DEVIATION PRINCIPLE FOR SUM OF SUBEXPONENTIAL VARIABLES: THE 1D MATCHING PROBLEM

JACOPO CATTANEO

THESIS ADVISOR

DR. SERGIO CARACCIOLO

THESIS CO-ADVISORS

MAURO PASTORE, ANDREA DI GIOACCHINO

April 4th, 2019

Jacopo Cattaneo: *Large deviation principle for sum of subexponential variables: the 1D matching problem*, , © April 4th, 2019

To the loving memory of my grandparents.

ABSTRACT

Large deviation theory connects different areas of physical interest, being able to justify the average properties of statistical ensembles while characterizing rare events and probabilities of extreme values.

In this work we focused on its application to the cost function of the random euclidean matching problem (REMP) on a compact interval. This is a well known toy-model belonging to the family of disordered spin-glass systems and optimization problems. In our case the cost function is given by the distance of two matched points (*spacing*) raised to some power $p > 1$.

In here we present the main results obtained in the case of independent, identically distributed, spacings, providing a precise asymptotic expression for the probability distribution of the cost function. In particular, we found two threshold sequences depending on the number of random variables n and the power p, which define different regions where different behaviors can be observed. On one hand, we show that in a region close to the expected value, deviations from the mean are exponentially suppressed in the number of spacings, which is typical. On the other hand, in a broader region, probability of rare events is less than exponentially dumped, making extreme values not so unlikely.

Moreover, even when correlations between spacings induced by the distance are taken into account, we show that the distribution of the average optimal cost in the REMP is asymptotically normally distributed in the region predicted by the central limit theorem. We show this region can be extended on a broader interval, defining a *moderate* large deviation principle in a range limited, as before, by a suitable threshold sequence.

Both results are supported by numerical simulations. These were obtained from direct sampling the distribution of the spacings and thus evaluating the cost function for different choices of p.

ACKNOWLEDGMENTS

I would like to thank everybody who, directly and indirectly, participated to this work.

First of all I would like to thank my thesis advisor, professor Sergio Caracciolo, a source of inspirational motivation. A big thanks goes to my co-advisors, Andrea Di Gioacchino and especially Mauro Pastore, that helped and supported me during the whole project.

A big thank goes to my parents, my family and my loved ones.
To all of my friends of Nuova Zelanda, to the Tombino gang, to the old and new acquaintances, to all of you a huge hug for having endured me. I know it was not that easy.

To everybody who would have preferred to see his name cited here, I'm happy to disappoint them. I'm joking, I just had no time.

Cheers!

CONTENTS

Part I

LARGE DEVIATION PRINCIPLE AND THE ONE DIMENSIONAL EUCLIDEAN MATCHING PROBLEM

1

INTRODUCTION TO THE LARGE DEVIATION THEORY

In this chapter we want to give some fundamental notions of large deviation theory and its applications in physics. We want to give some sense of what a large deviation is by a simple introduction to probability theory and collecting the fundamental results in this field. A simple application to statistical mechanics is given at the end of the chapter which tries to justify the physical interest of the topic. A good introduction to this topic can be found in the book of R. Ellis [11] and in the Lecture Notes in Physics book series by Springer [40]

1.1 WHY A LARGE DEVIATION THEORY

Describing the physical properties of macroscopic bodies via the computation of (ensemble) averages was the main focus of statistical mechanics at its early stage. As a matter of fact, as macroscopic bodies are made of a huge number of particles, fluctuations were expected to be too small to be actually observable. Broadly speaking, we can say that the theoretical basis of statistical descriptions was guaranteed by the law of large numbers.
On the other hand, whenever physicists calculate an entropy function or a free energy function, large deviation theory is at play. Indeed, large deviation theory is almost always involved when one studies the properties of many-particle systems, be they equilibrium or non equilibrium systems. It explains, for example, why the entropy and free energy functions are mutually connected by a Legendre transform, and so provides an explanation of the appearance of this transform in thermodynamics. Large deviation theory also explains why equilibrium states can be calculated via the extremum principles that are the (canonical) minimum free energy principle and the (microcanonical) maximum entropy principle.

The earliest origins of large deviation theory lie in the work of Boltzmann on entropy in the 1870ies [34] and Cramér's Theorem from 1938 [7, 8]. A unifying mathematical formalism was only developed starting with Varadhan's definition of a *large deviation principle* (LDP) in 1966 [37, 38].

Basically, large deviation theory centers around the observation that suitable functions f of large numbers of random variables $(X_1, \dots, X_n)$ often have the property that, for $n \gg 1$,

$$\Pr\left(f(X_1, \dots, X_n) \in dx\right) \sim e^{-a_n I(x)}\, dx\,, \tag{1.1}$$

where a_n is a suitable sequence such that $\lim_{n\to\infty} a_n = \infty$ (in most cases simply $a_n = n$). In other words, LDP states that the probability that $f(X_1, \ldots, X_n)$ takes values near a point x decays exponentially fast, with speed a_n, and rate function I.

Large deviation theory has two different aspects. On the one hand, there is the question of how to formalize the intuitive formula (1.1). This leads to the already mentioned definition of large deviation principles and involves quite a bit of measure theory and real analysis.

On the other hand, there is a much richer and much more important side of large deviation theory, which tries to identify rate functions I for various functions f and study their properties. This part of the theory is as rich as the branch of probability theory that tries to prove limit theorems for functions of large numbers of random variables, and has many relations to the latter.

1.2 BASIC ELEMENTS OF PROBABILITY THEORY

We start our study of large deviation theory by considering f as a sum of real random variables (RV for short) having the form

$$S_n = \frac{1}{n} \sum_{i=0}^{n} X_i \,. \tag{1.2}$$

Such a sum is often referred to in mathematics or statistics as a *sample mean*. Three basic questions naturally arise when n is very large:

- The behavior of the sample mean S_n , the possible convergence to an asymptotic value and its dependence on the sequence;
- The statistics of small fluctuations of S_n around $\langle S_n \rangle$, i.e. of $\delta S_n = S_n - \langle S_n \rangle$, when $|\delta S_n|$ is *small*;
- The statistical properties of rare events when such fluctuations are *large*.

The answer to the first question, in the simple case of sequences $(X_1, \ldots, X_n)$ of independent and identically distributed (IID for short) random variables with expected value μ and with finite variance, comes form the *law of large numbers* (LLN) which states the empirical average gets closer and closer to the expected value $\mu = \langle X_i \rangle$, when n is large:

$$\lim_{n\to\infty} \Pr\left(|\delta S_n| < \epsilon\right) = 1 \,. \tag{1.3}$$

The second issue is addressed by the *central limit theorem* (CLT). For instance, in the case of IID RV with expected value μ and finite variance σ^2, the CLT describes the statistics of small fluctuations, $|\delta S_n| \lesssim$

$O(\sigma/\sqrt{n})$, around the mean value when n is very large. Roughly speaking, the CLT proves that, in the limit $n \gg 1$, the quantity

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=0}^{n} (X_i - \mu)$$

is normally distributed, meaning that

$$p_{Z_n}(z) \sim \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{1.4}$$

independently of the distribution of the random variables. Under suitable hypothesis the theorem can be extended to dependent (weakly correlated) variables.

Finally, the last point is the subject of large deviation theory which, roughly, states that in the limit $n \gg 1$

$$p_{S_n}(s) \sim e^{-a_n I(s)}, \tag{1.5}$$

or, equivalently,

$$\lim_{n\to\infty} \frac{1}{a_n} \log p_{S_n}(s) = -I(s). \tag{1.6}$$

Unlike the central limit theorem result with the *universal* limit probability density (1.4), the detailed functional dependence of $I(s)$ – the *Cramér's* or *rate function* – and the speed a_n depends on the probability distribution of $p(X_1, \ldots, X_n)$. However, $I(s)$ possesses some general properties: it is zero for $s = \langle S_n \rangle$ and positive otherwise, moreover – when the variables are independent (or weakly correlated) – it is a convex function. A motivation of this last statement is provided in Appendix A. For a complete study of the properties of the rate function we refer to [36, 39]. For sake of simplicity in the following we will assume $a_n = n$, which is common in most simple cases.

1.3 FROM SMALL TO LARGE DEVIATIONS

An LDP for a random variable, say S_n again, gives us a lot of information about its distribution. The fact that the behavior of $p_{S_n}(s)$ is dominated for large n by a decaying exponential means that the exact distribution of S_n can be written as

$$p_{S_n}(s) = e^{-nI(s)+o(n)}$$

with $o(n)$ a sublinear correction in n. By taking the limit

$$\lim_{n\to\infty} \frac{1}{n} \log p_{S_n}(s) = I(s) + \lim_{n\to\infty} \frac{o(n)}{n} = I(s)$$

which retains only the dominant exponential term of the limiting distribution and neglects the others. For this reason, large deviation

theory is often said to be concerned with estimates of probabilities on the logarithmic scale.

Thus it should be clear that probability is exponentially suppressed in n anywhere the rate function $I(s)$ is non-vanishing. Moreover, we know that $p_{S_n}(s)$ concentrates on certain points which are the typical values of the sample mean S_n in the large n limit. These points correspond to the zeroes of the rate function $I(s)$ and it can be shown to be related mathematically to the LLN. Indeed, an LDP always implies some form of LLN.

Often it is not enough to know that S_n converges in probability to some values and we may also want to determine the likelihood that S_n takes a value away but close to its typical value. Let s_0 be one of this values and assume that $I(s)$ admits a Taylor series around s_0, then

$$I(s) = I(s_0) + I'(s_0)(s - s_0) + \frac{I''(s_0)}{2}(s - s_0)^2 + o\left((s - s_0)^2\right) . \quad (1.7)$$

Since s_0 is a zero of $I(s)$ the first two terms vanish and we are left with a law of *small deviations* of S_n around its typical value

$$p_{S_n}(s) \sim e^{-n\frac{I''(s_0)}{2}(s-s_0)^2} .$$

and in this sense, large deviation theory contains the CLT. At the same time, large deviation theory can be seen as an extension of the latter because it gives information not only about the small deviations but also about *large deviations* far away from its typical value(s).

1.3.1 *A combinatorial example*

A natural way to introduce the large deviation theory and show its deep relation with the concept of entropy is to perform a combinatorial computation. We will consider the simple example of a sequence of independent unfair-coin tosses. Denoting with X_i the result of the i-th toss, the possible outcomes are head $(+1)$ with probability π or tail (-1) with probability $1 - \pi$. Let S_n be the sample mean of such variables, we are now interested in formulating an LDP for the RV S_n.

Actually, the probability of the sample mean taking a specific value can be explicitly computed: in the sequence of n tosses of the coin the ways k heads can occur is given by the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} .$$

Thus, we have for the probability of the sample mean

$$\Pr\left(S_n = \frac{2k}{n} - 1\right) = \frac{n!}{k!\,(n-k)!}\pi^k(1-\pi)^{n-k} , \quad (1.8)$$

that is the binomial distribution.

Since we are interested in the asymptotic expression in the large n limit, both the number of heads and the number of tails will diverge. Thus, by setting

$$k = pn\,,$$
$$n - k = (1-p)n\,,$$

and using Stirling approximation of the factorial, the probability in (1.8) will take the form

$$\Pr\left(S_n = 2p-1\right) \sim e^{-n I_\pi(p)}\,, \tag{1.9}$$

where

$$I_\pi(p) = p\log\frac{p}{\pi} + (1-p)\log\frac{1-p}{1-\pi} \tag{1.10}$$

is the rate function of the probability distribution. We can notice $I_\pi(\pi) = 0$, meaning the only value for which the probability is non vanishing is the expected value of the sample mean. Actually, the CLT can be recovered by a Taylor expansion at $p = \pi$, that is

$$I_\pi(p) = \frac{1}{2}\frac{(p-\pi)^2}{\pi(1-\pi)} + o\left((p-\pi)^2\right)\,. \tag{1.11}$$

From the above computation we understand that it is possible to go beyond the CLT, and to estimate the statistical features of extreme (or tail) events, as the number of observations n grows without bounds.

1.4 SOME USEFUL RESULTS IN LARGE DEVIATION THEORY

The first approximation we can have on the distribution comes from the **Markov's inequality**, that gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant. It relates probabilities to expectations, and provide bounds for the cumulative distribution function of a random variable. It is obtained by noting that, for any non-negative RV X with probability density function (PDF for short) $p(x)$, it holds

$$\begin{aligned}\langle X\rangle &= \int_0^\infty dx\, x p(x) = \int_0^a dx\, x p(x) + \int_a^\infty dx\, x p(x)\\ &\geqslant \int_a^\infty dx\, x p(x) \geqslant a\int_a^\infty dx\, p(x) = a\Pr(X \geqslant a)\,,\end{aligned}$$

for any positive a. Therefore we have

$$\Pr(X \geqslant a) \leqslant \frac{\langle X\rangle}{a} \tag{1.12}$$

A direct consequence of Markov's inequality is the **Chernoff bound** [6], that gives exponentially decreasing bounds on tail distributions

of sums of independent random variables. This is simply obtained by applying (1.12) to the RV e^{tX},

$$\Pr(X \geqslant a) = \Pr\left(e^{tX} \geqslant e^{ta}\right) \leqslant \frac{\left\langle e^{tX}\right\rangle}{e^{ta}},$$

that holds for any $t \geqslant 0$. Now, let $X = S_n$ be the sample mean of IID RVs: by recalling that

$$p^{(n)}(X_1, \ldots, X_n) = \prod_{i=1}^{n} p^{(1)}(X_i), \tag{1.13}$$

then it holds

$$\begin{aligned}\Pr(S_n \geqslant s) &= \Pr\left(\sum_{i=0}^{n} X_i \geqslant ns\right) = \Pr\left(e^{t\sum_{i=0}^{n} X_i} \geqslant e^{tns}\right) \\ &\leqslant \left(\frac{\left\langle e^{tX}\right\rangle}{e^{ts}}\right)^n \leqslant \left(\min_{t\geqslant 0}\left[e^{-ts}\left\langle e^{tX}\right\rangle\right]\right)^n.\end{aligned}$$

Thus, in terms of large deviation theory,

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr(S_n \geqslant s) \leqslant -\sup_{t\geqslant 0}\left[ts - \log M(t)\right] \tag{1.14}$$

where $M(t) \equiv \left\langle e^{tX}\right\rangle$ is the *moment generating function* (MGF) of the variable X.
A similar Chernoff bound can be obtained for the complementary probability: namely, for any $t \leqslant 0$, we have

$$\Pr(S_n \leqslant s) = \Pr\left(e^{t\sum_{i=0}^{n} X_i} \geqslant e^{tns}\right) \leqslant \left(\min_{t\leqslant 0}\left[e^{-ts}\left\langle e^{tX}\right\rangle\right]\right)^n,$$

and thus

$$\lim_{n\to\infty}\frac{1}{n}\log\Pr(S_n \leqslant s) \leqslant -\sup_{t\leqslant 0}\left[ts - \log M(t)\right]. \tag{1.15}$$

The main result of large deviation theory that is widely used to obtain LDPs is called the **Gärtner–Ellis Theorem** (GE Theorem) [12, 14]. Let

$$K(\lambda) = \lim_{n\to\infty}\frac{1}{n}\log\left\langle e^{n\lambda f_n}\right\rangle_{f_n} \tag{1.16}$$

be the *scaled cumulant generating function* (SCGF) for some real RV f_n parametrized by n, where

$$\left\langle e^{n\lambda f_n}\right\rangle_{f_n} = \int \Pr\left(f_n \in dx\right) e^{n\lambda x}. \tag{1.17}$$

In the following we will omit the subscript on which probability measure the expectation value is taken unless it is essential for understanding. The GE Theorem states that if $K(\lambda)$ exists $\forall\lambda \in \mathbb{R}$ and is differentiable then f_n satisfies a LDP

$$\Pr\left(f_n \in dx\right) \sim e^{-nI(x)}\, dx, \tag{1.18}$$

with *rate function* $I(x)$ given by

$$I(x) = \sup_{\lambda} \left[\lambda x - K(\lambda)\right] . \tag{1.19}$$

The transform defined by the supremum is an extension of the Legendre transform referred to as the *Legendre–Fenchel transform* (LFT for short). We refer to Appendix A for a more in-depth discussion.

The GE Theorem thus states in words that, when the SCGF $K(\lambda)$ is differentiable, then f_n obeys a large deviation principle with a rate function $I(x)$ given by the LFT of $K(\lambda)$.
The SCGF has some interesting properties: since the probability measure is normalized $K(0) = 0$ and, for $m \in \mathbb{N}$, it holds

$$\begin{aligned} K^{(m)}(0) &= \frac{\partial^m}{\partial \lambda^m} K(\lambda)\bigg|_{\lambda=0} \\ &= \lim_{n\to\infty} \frac{1}{n} \frac{\partial^m}{\partial \lambda^m} \log \left\langle e^{n\lambda f_n} \right\rangle \bigg|_{\lambda=0} \\ &= \kappa_m \end{aligned}$$

where κ_m is the m-th cumulant of f_n. In particular, the first two cumulants correspond to

$$\begin{aligned} \kappa_1 &= \lim_{n\to\infty} \langle f_n \rangle_{f_n} = \mu \\ \kappa_2 &= \lim_{n\to\infty} n \left(\langle f_n^2 \rangle - \langle f_n \rangle^2 \right) = \lim_{n\to\infty} n \mathrm{var}(f_n) . \end{aligned}$$

The importance of the GE Theorem is to be able to calculate the SCGF without knowing the exact form of $p_{f_n}(x)$. Conversely, the asymptotic expression of the distribution $p_{f_n}(x)$ can be retrieved exactly with this theorem, by the evaluation of the rate function itself.

The rigorous proof of the GE Theorem is too technical to be presented here. However, there is a way to justify this result by deriving in a heuristic way another result known as Varadhan's Theorem. The latter theorem is concerned with the evaluation of a functional expectation of the form

$$W_n[g] = \left\langle e^{n g(f_n)} \right\rangle = \int dx\, p_{f_n}(x) e^{n g(x)} \tag{1.20}$$

with g some function of the real RV f_n. Assuming f_n satisfies a LDP with rate function $I(x)$, then we can write

$$\begin{aligned} W_n[g] &\sim \int dx\, e^{n[g(x) - I(x)]} \\ &\sim e^{n \sup_x [g(x) - I(x)]} \end{aligned}$$

where the last equality follows from the *saddle-point approximation* or *Laplace approximation*, and it is justified in the context of large deviation theory because the corrections to this approximation are

subexponential in n, as are those of the LDP. By defining the following functional

$$K[g] = \lim_{n\to\infty} \frac{1}{n} \log W_n[g] \,,$$

using a limit similar to the limit defining the LDP, we then obtain

$$K[g] = \sup_x [g(x) - I(x)] \,. \tag{1.21}$$

The result above is what is referred to as **Varadhan's Theorem** [37], which proved this result for a wide class of RV.

To connect Varadhan's Theorem with the GE Theorem we consider the special case $g(s) = \lambda s$ with $\lambda \in \mathbb{R}$. Then equation in (1.21) becomes

$$K(\lambda) = \sup_x [\lambda x - I(x)] \,, \tag{1.22}$$

which is the same function defined in (1.16).
Thus it is clear that if S_n satisfies a LDP with rate function $I(s)$, then the SCGF $K(\lambda)$ is the LFT of $I(s)$.

This heuristic derivation illustrates two important points about large deviation theory. The first is that LFTs appear into this theory as a natural consequence of the saddle point approximation. The second is that the Gärtner–Ellis Theorem is essentially a consequence of the large deviation principle combined with Laplace's approximation.

To complete this presentation of known results it is useful to have a look at the application of GE Theorem in the special case of the sample mean of a set of IID RV $(X_1, \ldots, X_n)$. Namely, by taking $f_n = S_n$ as in (1.2) and by recalling that

$$p^{(n)}(X_1, \ldots, X_n) = \prod_{i=1}^{n} p^{(1)}(X_i) \,,$$

we can write for the SCGF

$$\begin{aligned} K_{S_n}(\lambda) &= \lim_{n\to\infty} \frac{1}{n} \log \left\langle e^{n\lambda S_n} \right\rangle_n \\ &= \lim_{n\to\infty} \frac{1}{n} \log \left[\prod_{i=1}^{n} \left\langle e^{\lambda X_i} \right\rangle_1 \right] \\ &= \log \left\langle e^{\lambda X} \right\rangle_1 \,. \end{aligned} \tag{1.23}$$

As a result the SCGF K_{S_n} is simply derived by the *cumulant generating function* of the single RV X and the LDP for S_n is retrieved by taking the LFT just as in (1.19). This result goes under the name of **Cramér's Theorem** and plays a central role in determining a large deviation principle for the sample mean of different RV.
Note that the differentiability condition of the GE Theorem need not be checked for IID sample means because the *moment generating function*

(MGF) or *Laplace transform* $\left\langle e^{\lambda X}\right\rangle_1$ of a RV is always real analytic when it exists $\forall\lambda\in\mathbb{R}$. On the other hand, the existence of the MGF on the positive real axis (usually referred to as *Cramér's condition*) must not be taken for granted and sometimes more refined tools may be necessary to establish an LDP.

1.5 LARGE DEVIATIONS IN STATISTICAL MECHANICS

As it has been said the mathematical basis for the notion of thermodynamic behavior is the LLN. The idea is that the outcomes of a *macrostate*, say $M_n(\omega)$, involving n particles should concentrate in probability around certain equilibrium values despite the fact that the particles' *microstate* is modeled by a random variable ω taking values in the *phase space* Λ^n. Large deviation theory enters this picture by noting that, in many cases, the outcomes of the macrostate are ruled by a large deviation principle, and that, in these cases, the concentration of probability measure p_{M_n} around these equilibrium values is exponentially fast in the large n limit. Consequently, all that is needed to describe the state of a large many-particle system at the macroscopic level is to know the equilibrium values of M_n which correspond to the global minima of the rate function governing the fluctuations.

Let us consider the *microcanonical ensemble*, that is a closed system with constant energy. We can start by writing the probability that the energy ϵ_n sits in a range $d\epsilon$ in terms of the probability measure of the microstates

$$\Pr(\epsilon_n\in d\epsilon)=\int_{\Lambda^n}\delta\left(\epsilon_n(\omega)-\epsilon\right)\Pr(d\omega)\,.$$

By taking the uniform measure $\Pr(d\omega)=d\omega\,/|\Lambda|^n$ it appears that $\Pr(\epsilon_n\in d\epsilon)$ is proportional to the volume in the phase space occupied by microstates which satisfy $\epsilon_n(\omega)=\epsilon$, namely

$$\Omega_{\epsilon_n}(\epsilon)=\int_{\Lambda^n}\delta\left(\epsilon_n(\omega)-\epsilon\right)d\omega\ .$$

Thus, by admitting ϵ_n satisfies an LDP with rate function $I(\epsilon)$, we have

$$I(\epsilon)=-\lim_{n\to\infty}\frac{1}{n}\log\Pr(\epsilon_n\in d\epsilon)\equiv -s(\epsilon)$$

where

$$s(\epsilon)=\lim_{n\to\infty}\frac{1}{n}\log\frac{\Omega_{\epsilon_n}(\epsilon)}{|\Lambda|^n}$$

is the *microcanonical entropy density*. Moreover, proportionality between $\Pr(\epsilon_n\in d\epsilon)$ and $\Omega_{\epsilon_n}(\epsilon)$ implies for the SCGF

$$K(\lambda)=\lim_{n\to\infty}\frac{1}{n}\log\left\langle e^{n\lambda\epsilon_n}\right\rangle=-\beta\varphi(\beta)|_{\beta=-\lambda}$$

where $\varphi(\beta)$ is the *free energy density*. More explicitly

$$\begin{aligned}\varphi(\beta) &= -\lim_{n\to\infty}\frac{1}{\beta n}\log Z_n(\beta)\\ &= -\lim_{n\to\infty}\frac{1}{\beta n}\log\int_{\Lambda^n} d\omega\, e^{-n\beta\epsilon_n(\omega)}\\ &= -\lim_{n\to\infty}\frac{1}{\beta n}\log\left\langle e^{-n\beta\epsilon_n}\right\rangle .\end{aligned}$$

As a consequence, from Varadhan's and GE Theorem, we can establish the relation between $\varphi(\beta)$ and $s(\epsilon)$

$$\begin{aligned}\beta\varphi(\beta) &= \inf_{\epsilon}[\beta\epsilon + s(\epsilon)],\\ s(\epsilon) &= -\inf_{\beta}[\beta\epsilon - \beta\varphi(\beta)],\end{aligned}$$

that is the entropy and the free energy can be obtained by an LFT.

As it may have become clear, LDT is at play each time we are interested in average properties of statistical ensembles, while accounting for probability of rare events. In particular, we showed that both the SCGF and the rate function itself are related to thermodynamic quantities of great physical interest, that is the free energy and entropy.

2

THE ONE DIMENSIONAL RANDOM EUCLIDEAN MATCHING PROBLEM

2.1 DISORDERED SYSTEMS

Before we start studying a specific spin-glass model, we need to introduce a couple of simple concepts, which we will extensively use all along the dissertation. Each of them would deserve much more space than we can afford: we refer to [5, 24] for a background in statistical mechanics and glassy systems.

2.1.1 *What is disorder*

As already pointed out in the previous chapter, when we study a system involving a large number n of observables, we are mainly interested in finding the expression of some average properties, that is the value of a macroscopic measurable quantity.

In the language of statistical mechanics this is usually achieved by the evaluation of the **partition function**

$$Z(\beta) = \sum_{\omega \in \mathcal{S}} e^{-\beta \epsilon(\omega)}, \tag{2.1}$$

where each microstate ω, in the space of configurations $\mathcal{S}$, is associated with the corresponding energy $\epsilon(\omega)$. This dependence between energy ϵ and microstates ω is expressed by the dependence of the energy on the parameters describing the microstate, that is the Hamiltonian H. From the partition function in (2.1), we can obtain a lot of information on the average, or typical, values that different quantities of physical interest can assume. As an example, the mean energy can be retrieved by a simple derivation of the partition function, that is

$$E = -\frac{\partial}{\partial \beta} \log Z(\beta). \tag{2.2}$$

The simplest way to introduce disorder in any system is considering some of the parameters describing the microstate ω as stochastic variables. There are two main classes of disordered systems: the first we are going to discuss are *quenched* disordered systems. In these systems the disorder is explicitly present in the Hamiltonian, typically under the form of random couplings $\mathbf{J}$ among the degrees of freedom $\boldsymbol{\sigma}$

$$H = H(\mathbf{J}, \boldsymbol{\sigma}).$$

The disorder introduced by the set of RVs **J** is completely specified by their probability distribution $p(J)$ which is assumed to be the same for each different coupling constant in the system. A famous example is the Edwards-Anderson model [10], described by the Hamiltonian

$$H = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j \,,$$

where the spins $\sigma_i = \pm 1$ are the degrees of freedom of the system, and the couplings J_{ij} are Gaussian RVs. This is a finite dimensional model, since the sum is performed over nearest-neighbor spins. In this case we say the disorder is **quenched**, meaning that the set of RVs **J** are constant on the time scale over which the the degrees of freedom **σ** fluctuate. This fact is realized in physical systems where the (microscopic) parameters governing the evolution of the system, that is the variables on which the Hamiltonian depends on, at sufficiently low temperatures can be separated into two different classes, *slow* and *fast* observables. The difference between the two classes arises from noticing that the typical time scale of evolution of an observable (in this case the couplings J_{ij}) is much larger than the time scale on which the spins σ_i interact. These systems are called **spin-glass systems** and an extended literature has been produced around this kind of problems [1, 24, 35].

As a matter of fact disorder creates **frustration**: it becomes impossible to satisfy all the couplings at the same time, as it would be in a ferromagnetic system. Formally, a system is said to be frustrated if there exists a loop on which the product of the couplings is negative. This can be better understood by looking at a frustrated loop: if we fix an initial spin, and starting from it we try to chain-fix the other spins one after the other according to the sign of the couplings, we are bound to return to the initial spin and flip it. As a consequence, the energy of a frustrated loop is not located at its minimum, as it would be if the couplings J_{ij} could explore the whole configurations' space.

On the other hand, if the time scale of the parameters describing the system are of the same magnitude, this implies the time evolution of all the observables must be taken into account simultaneously when computing ensemble averages. Such kind of randomness is usually referred to as **annealed** disorder. This is often the case of spin-glass systems at high temperature, where the frustration induced by the disorder is irrelevant, as the system can visit a lot of different, often high energy, configurations due to the effect of the entropic force.

In the following we will deal only with quenched disordered systems, that is statistical systems where a bunch of degrees of freedom cannot fluctuate freely and the space of configurations is restricted due to the specific realization of the parameters.

2.1.2 *Large deviations in glassy systems*

In the study of disordered systems nearly all predictions concern the most likely behavior, but there is also considerable interest in developing techniques to compute the probability distribution of rare events, i. e. the probability of finding systems that have properties different from the typical ones. Systems with quenched disorder have been studied intensively for the last two decades. Thermodynamic properties in these systems, such as the free energy, fluctuate from sample to sample, but not very much: indeed, they are self-averaging if the disorder does not have long range correlations [19]. This means that typical values of the free energy density (to name but one quantity) deviate arbitrarily little from a fixed value in the large volume limit.

Because of this, little work has considered large deviations, i. e. the probability of finding a rare sample (realization of the disorder). Indeed it is well known that the probability of large deviations is related to the free energy function $\varphi(\beta)$. Thus we are interested in big oscillations of this quantity when we try to formulate an LDP for any observable in a disordered system. As already discussed in Section 1.5, in LDT the analog of the free energy is the SCGF: this will play a central role in determining big oscillations from the typical values for the quantities of interest.

2.2 RANDOM OPTIMIZATION PROBLEMS

Combinatorial optimization is a branch of operational research which deals with the problem of optimizing a cost function over a finite set of configurations. Let $\mathcal{S}$ be the set of all the possible configurations ω a system can explore. Then, given a cost function $\mathcal{E}(\omega) \in \mathbb{R}$, we are interested in finding the optimal configuration $\omega^* \in \mathcal{S}$ that minimizes such cost, namely

$$\mathcal{E}(\omega^*) = \min_{\omega \in \mathcal{S}} \mathcal{E}(\omega) .$$

In terms of statistical systems discussed in the previous section, the cost function $\mathcal{E}(\omega)$ can be interpreted as the energy of a configuration. In this case the problem of finding the optimal cost corresponds to find the average energy in the zero temperature limit. More explicitly, the optimal cost is retrieved from (2.2) by simply taking the limit for $\beta \to \infty$, that is

$$\mathcal{E}(\omega^*) = - \lim_{\beta \to \infty} \frac{\partial}{\partial \beta} \log Z(\beta) .$$

When considering a *random optimization problem* we mean a special kind of optimization problem where the space of configuration $\mathcal{S}$ depends on some random parameters. It is quite clear that in this case $\mathcal{E} = \mathcal{E}(\omega^*)$, i. e. the set $\mathcal{S}$ of possible configurations, will depend on

the particular realization of the set of RVs, which remains fixed for any realization. As a consequence, this kind of systems belong to the bigger family of glassy systems.

We call $\mathcal{S}$ an *instance* of the problem and we are interested in finding the average properties of the optimal solution. In particular, denoting with $\langle\bullet\rangle$ the average over the instances, we could ask what is the *average minimal cost*

$$\langle \mathcal{E} \rangle = \left\langle \min_{\omega \in \mathcal{S}} \mathcal{E}(\omega) \right\rangle = - \lim_{\beta \to \infty} \left\langle \frac{\partial}{\partial \beta} \log Z(\beta) \right\rangle . \tag{2.3}$$

After the seminal works of Kirkpatrick [18] Orland [30], and Mézard and Parisi [21], random optimization problems have been successfully studied using statistical physics techniques. The average appearing in the previous equation can be tackled using the celebrated *replica trick*, which allowed the derivation of fundamental results for many relevant random combinatorial optimization problems, like random matching problems [21–23, 30] or the traveling salesman problem in its random formulation [20, 30].

Very often the set of the configurations $\mathcal{S}$ is described in the modern abstract framework of graph theory. It seems useful to briefly recall some definitions of this field: an (*undirected*) *graph* $\mathcal{G}$ is a couple (V, E) where

- V is the set of *vertices* of the graph;
- $E \subseteq V \times V$ is the set of *edges*, where $e \in E$ is an unordered pair of vertices, namely $e = (u, v) \subset V$, $v \neq u$.

The vertices u and v are said to be the *ends* of the edge $e = (u, v)$ and thus we say the edge e is *incident* onto both v and u. A graph $\mathcal{K} = (V, E)$ is said to be *complete* if each pair of vertices are connected by an edge (i.e. $(u, v) \subset V \Leftrightarrow e = (u, v) \in E$).
With this definitions we are now able to present the main topic of this chapter.

2.2.1 *The matching problem*

The matching problem is a rather simple system which has similarities with spin glasses with finite range interactions [24]. Its applications go from biology [13], to traffic modeling [3, 31], to neural networks [32, 41]. Here we want to define the generic problem in terms of graph theory and give a hint of the different flavors it can assume.

Given a graph $\mathcal{G} = (V, E)$ a *matching* $\mu \subseteq E$ is a set of edges having the property that none of the edges in μ have an end in common. More explicitly, for any $e_1, e_2 \in \mu$, then $e_1 \cap e_2 = \varnothing$. We say that a vertex $v \in V$ is *matched* if there is an edge incident to v in the matching, otherwise the vertex is *unmatched*.

Denoting by $|\mu|$ the cardinality of μ, we define $\nu(\mathcal{G}) \equiv \max_\mu |\mu|$ the *matching number* of $\mathcal{G}$. As a consequence a matching μ is *maximum* if there is no matching of greater cardinality, that is if $|\mu| = \nu(\mathcal{G})$. In particular, a maximum matching is called *perfect* if every vertex of $\mathcal{G}$ is matched. Obviously any perfect matching is maximum and thus maximal and we will denote by M the set of perfect matchings.

Now, we can imagine to assign a cost to each edge in the graph: let $w_e \geqslant 0$ be a weight corresponding to the edge $e \in E$ in the graph $\mathcal{G}$. Then we can associate to each perfect matching $\mu \in M$ the *total cost* function

$$\mathcal{E}(\mu) \equiv \sum_{e \in \mu} w_e \tag{2.4}$$

and the *average cost* per edge

$$\varepsilon(\mu) \equiv \frac{\mathcal{E}(\mu)}{\nu(\mathcal{G})} = \frac{1}{\nu(\mathcal{G})} \sum_{e \in \mu} w_e \,. \tag{2.5}$$

In the weighted matching problem we search for the perfect matching μ such that the total cost in (2.4) is minimized, that is the optimal matching μ^* is such that

$$\mathcal{E}(\mu^*) = \min_{\mu \in M} \mathcal{E}(\mu) \,. \tag{2.6}$$

In the following we will deal with random matching problems, where the costs $\{w_e\}_{e \in E}$ are RVs. In this case, the average properties of the optimal solution are of a certain interest, and in particular the typical optimal cost, $\langle \mathcal{E} \rangle = \langle \min_\mu \mathcal{E}(\mu) \rangle$. The simplest way we can imagine to introduce randomness in the problem is to consider the weights as IID RVs. A number of rigorous result were obtained in this mean field theory, starting from the work of Parisi and Mézard [21].

In this work, we will focus on the *random Euclidean matching problem* (REMP), where the graph $\mathcal{G}$ is supposed to be embedded in a d-dimensional Euclidean domain $\Lambda \subseteq \mathbb{R}^d$ through an embedding function Φ, in such a way that each vertex $v \in V$ of the graph is associated to a random Euclidean point $v \mapsto \Phi(v) \in \Lambda$. In this case the random weights w_e associated to each edge $e = (u, v) \in E$ will be some function of the distance of the corresponding points in Λ, namely

$$w_e = f\left(\|\Phi(u) - \Phi(v)\|\right) \,.$$

REMPs are usually more difficult to investigate than the purely random case of $\{w_e\}$ IID RVs, due to the presence of Euclidean correlations among the weights. This induces a non-trivial dependence between the RVs, that result in the problem being more complex to investigate.

2.3 THE RANDOM EUCLIDEAN MATCHING PROBLEM ON AN INTERVAL

In the following we will restrict ourselves to a specific toy model in one dimension, that is the random Euclidean matching problem. Here we refer to the work of Caracciolo, D'Achille, Sicuro [4], that obtained a number of results for this system.

We will focus only on the case in which $\mathcal{G} = \mathcal{K}_{2n}$ is a complete graphs with $2n$ vertices associated with a set of points $\Xi_{2n} \equiv \{x_i\}_{i=1,\dots,2n}$ independently and uniformly generated on the compact interval $\Lambda = [0,1]$. In this case a perfect matching μ of $2n$ points corresponds to any partition of the set Ξ_{2n} made up of two elements only, its cardinality being n.

As weight function of the edge $e = (x_i, x_j)$ we will consider

$$w_e \equiv w_{i,j} = |x_i - x_j|^p \,, \quad p > 1 \,.$$

Hence the average cost function in (2.5) will take the form

$$\varepsilon_n^{(p)}(\mu) = \frac{1}{n} \sum_{(i,j)\in\mu} |x_i - x_j|^p \,. \tag{2.7}$$

The reason for this specific choice of the weight function is that, with this definition, it is a monotonically increasing, convex function of the Euclidean distance. Among the numerous consequences, it can be shown [4] the cost function is self averaging quantity, meaning that the probability measure concentrates around the typical values while vanishing elsewhere. Moreover, in this particular case, the optimal configuration has a simple structure. Let the elements in Ξ_{2n} be indexed such that $0 \leqslant x_1 \leqslant x_2 \cdots \leqslant x_{2n} \leqslant 1$, then the pair (x_i, x_j) belongs to the optimal matching if and only if i is odd and $j = i + 1$. This follows from the direct investigation of the simplest non-trivial case for $n = 2$: the possible outcomes are shown in Figure 2.1 where the arcs represent the edges in each matching.
The solution in Figure 2.1b is non-optimal since, for any $p > 1$,

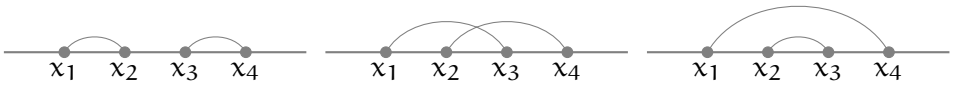
$$|x_3 - x_1|^p = |(x_3 - x_2) + (x_2 - x_1)|^p \geqslant |x_2 - x_1|^p$$

and similarly $|x_4 - x_2|^p \geqslant |x_4 - x_3|^p$. Moreover, for the case in Figure 2.1c, we have

$$\begin{aligned} |x_4 - x_1|^p &= |(x_4 - x_3) + (x_3 - x_2) + (x_2 - x_1)|^p \\ &\geqslant |x_4 - x_3|^p + |x_3 - x_2|^p + |x_2 - x_1|^p \\ &\geqslant |x_4 - x_3|^p + |x_2 - x_1|^p \end{aligned}$$

the last line being the cost of the optimal matching in Figure 2.1a.

The study of the properties of the optimal matching is reduced therefore to the study of spacings between successive random points



(a) (b) (c)

Figure 2.1: Representation of the three possible matchings in the simple case of $n = 2$ with cost function $w_{i,j} = |x_i - x_j|^p$, $p > 1$. Both solutions in (b) and (c) are non-optimal, their cost being greater by direct inspection. As a consequence, the optimal solution with generic n has always the structure (a) with edges $e_i = (x_{2i-1}, x_{2i})$, $i = 1, \dots, n$.

on Λ. By defining $\varphi_i \equiv |x_i - x_{i-1}|$ the i-th *spacing*, the optimal cost in (2.7) takes the form

$$\min_{\mu \in M} \varepsilon_n^{(p)}(\mu) = \frac{1}{n} \sum_{i=1}^{n} \varphi_{2i-1}^p \,. \tag{2.8}$$

Having defined the structure of the optimal solution we can now study its average properties.

2.3.1 *Probability distribution of random uniform spacings on an interval*

We are interested in finding the explicit expression for the probability distribution of the set of RVs $\boldsymbol{\varphi} = (\varphi_0, \dots, \varphi_{2n})$. Let us firstly observe that the distribution of the ordered set $\mathbf{X} = (X_1, \dots, X_{2n})$ of random points on Λ is given by

$$p_n(\mathbf{x}) = (2n)! \prod_{i=0}^{2n} \theta(x_{i+1} - x_i) \tag{2.9}$$

being $x_0 \equiv 0$ and $x_{2n+1} \equiv 1$. It follows that for the set $\boldsymbol{\varphi}$

$$\rho_n^{(2n+1)}(\boldsymbol{\varphi}) = (2n)! \left[\prod_{i=0}^{2n} \theta(\varphi_i) \right] \delta \left(\sum_{j=0}^{2n} \varphi_j - 1 \right) . \tag{2.10}$$

This multivariate distribution is known as *Dirichlet distribution*, often denoted $\mathrm{Dir}(\boldsymbol{\alpha})$, with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{2n+1})$ a vector of positive real parameters. The generic PDF for a set of non-negative RVs $\mathbf{Y} = (Y_1, \dots, Y_K) \sim \mathrm{Dir}(\boldsymbol{\alpha})$ is given by

$$p_D(\mathbf{y}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} y_i^{\alpha_i - 1} \,. \quad \text{with} \quad \sum_{i=1}^{K} y_i = 1 \,, \tag{2.11}$$

The normalizing constant in front of the distribution is defined as

$$B(\boldsymbol{\alpha}) = \frac{\prod_j \Gamma(\alpha_j)}{\Gamma\left(\sum_j \alpha_j\right)}$$

which denotes the multivariate *beta function*. The Dirichlet distribution, because of the constraint on the sum of y_is being 1, obviously induces correlations among the RVs. Actually, the joint probability distribution is degenerated, meaning it can be expressed in terms of $K-1$ variables, while the K-th one is uniquely determined by the constraint, namely

$$y_K = 1 - \sum_{i=1}^{K-1} y_i \,. \tag{2.12}$$

As it can be easily seen from (2.11), the PDF is completely symmetric for any exchange in the couples (y_i, α_i). This implies that the explicit dependence in (2.12) of one variable in terms of the $K-1$ remaining ones can be carried out for any of the K RVs, while the distribution remaining consistent. This property, that is the complete exchangeability of the RVs under any permutation of the indices, is a consequence of the probability distribution (2.11) having support on the $(K-1)$-dimensional simplex. This is a generalization of a triangle embedded in the next-higher dimension. For example, with $K = 3$, the support is an equilateral 2-dimesional triangle embedded in a downward-angle fashion in 3-dimensional space, with vertices at $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, i. e. touching each of the coordinate axes at a point 1 unit away from the origin.
The set of positive parameters $\boldsymbol{\alpha}$ can be interpreted as the weights of the RVs. Their meaning can be better understood by taking the symmetric case where $\alpha_i = \alpha$, for all i. For values of α larger than 1 the resulting distribution favors evenly distributed RVs, meaning Y_i are close to each other. On the other hand, when α is smaller than 1, the distribution selects sparse samples, meaning that most of the contribution to their sum comes from few RVs taking a large value. In our case the distribution in (2.10) is retrieved simply by setting $\alpha_i = 1$, $i = 1, \dots, 2n+1$, that corresponds to the uniform distribution on the $(2n+1)$-dimesional simplex.

It is useful to evaluate the expected value of a generic product of Dirichlet distributed RVs

$$\begin{aligned}\left\langle \prod_{i=1}^{K} y_i^{p_i} \right\rangle &= \frac{1}{B(\boldsymbol{\alpha})} \left[\prod_{i=1}^{K} \int_0^\infty dy_i \, y_i^{p_i} \right] \delta\left(1 - \sum_{j=0}^{2n} y_j \right) \\ &= \frac{\Gamma\left(\sum_j \alpha_j\right)}{\prod_j \Gamma(\alpha_j)} \int_0^\infty dy_1 \dots \\ &\quad \dots \int_0^\infty dy_{K-1} \, y_1^{p_1} \dots \left(1 - \sum_{j=0}^{2n-1} y_j \right)^{p_K} ,\end{aligned}$$

which reduces to

$$\left\langle \prod_{i=0}^{K} y_i^{p_i} \right\rangle = \frac{\Gamma\left(\sum_j \alpha_j\right)}{\Gamma\left(\sum_j \alpha_j + p_j\right)} \prod_{i=0}^{K} \frac{\Gamma(\alpha_i + p_i)}{\Gamma(\alpha_i)} \tag{2.13}$$

$$= \frac{B(\boldsymbol{\alpha} + \mathbf{p})}{B(\boldsymbol{\alpha})} \,. \tag{2.14}$$

Thus, specializing to our case,

$$\left\langle \prod_{i=0}^{2n} \varphi_i^{p_i} \right\rangle = \frac{\Gamma(2n+1)}{\Gamma\left(2n+1+\sum_j p_j\right)} \prod_{i=0}^{2n} \Gamma(p_i+1) \,. \tag{2.15}$$

This formula can be used to evaluate any moment and correlation function, which is finite for any choice of $n \in \mathbb{N}$ and $p > 1$.

The distribution in (2.10) can now be marginalized to obtain the PDF for the single variable by subsequent integrations. Performing this operation over $2n$ variables we have

$$\begin{aligned}
\rho_n^{(1)}(\varphi_i) &= (2n)! \left[\prod_{r\neq i} \int_0^\infty d\varphi_r\right] \delta\left(\sum_{s=0}^{2n} \varphi_s - 1\right) \\
&= (2n)! \lim_{\epsilon\to 0^+} \int_{\mathbb{R}} \frac{d\xi}{2\pi} e^{(-i\xi+\epsilon)(1-\varphi_i)} \\
&\qquad \times \left[\prod_{r\neq i} \int_0^\infty d\varphi_r \; e^{(i\xi-\epsilon)\varphi_r}\right] \\
&= (2n)!\, i^{2n} \lim_{\epsilon\to 0^+} e^{\epsilon(1-\varphi_i)} \int_{\mathbb{R}} \frac{d\xi}{2\pi} \frac{e^{-i\xi(1-\varphi_i)}}{(\xi+i\epsilon)^{2n}}
\end{aligned}$$

where we used the Fourier representation of the Dirac delta function. Thus, from a simple contour integral around the pole $\xi = -i\epsilon$ of $2n$-th order, it follows

$$\rho_n^{(1)}(\varphi) = 2n(1-\varphi)^{2n-1}\theta(\varphi)\theta(1-\varphi) \,, \tag{2.16}$$

which is known as the *beta distribution* $\mathcal{B}(\alpha, \beta)$, the parameters being in this case $\alpha = 1$, $\beta = 2n$. It is important to stress out that no dependence on i appears in $\rho_n^{(1)}$ because of the symmetry from Dirichlet distributed RVs. From a similar calculation we can also evaluate the joint probability density for the set of RVs $\boldsymbol{\varphi}_{\text{odd}} = (\varphi_1, \ldots, \varphi_{2n-1})$, i. e. the set of spacings appearing in the cost function (2.8) of the optimal matching. It follows

$$\rho_n^{(n)}(\boldsymbol{\varphi}_{\text{odd}}) = \frac{(2n)!}{n!} \left(1 - \sum_{i=1}^{n} \varphi_{2i-1}\right)^{2n-1} \,, \tag{2.17}$$

with $\varphi_i \geqslant 0$ and $\sum_i \varphi_{2i-1} \leqslant 1$.

From the expression in (2.16), we can now retrieve the average cost per edge, that is

$$\left\langle \varepsilon_n^{(p)} \right\rangle \equiv \langle \varphi^p \rangle = \frac{\Gamma(2n+1)\Gamma(p+1)}{\Gamma(2n+p+1)} \tag{2.18}$$

$$= \frac{\Gamma(p+1)}{(2n)^p} \left[1 + \frac{p(p+1)}{4n} + o\left(\frac{1}{n}\right) \right] . \tag{2.19}$$

Moreover, successive moments can be evaluated in the same way:

$$\left\langle \left(\varepsilon_n^{(p)}\right)^2 \right\rangle = \frac{1}{n^2} \sum_{i,j=1}^{n} \left\langle \varphi_{2i-1}^p \varphi_{2j-1}^p \right\rangle$$

$$= \frac{1}{n^2} \left(n \left\langle \varphi^{2p} \right\rangle + n(n-1) \left\langle \varphi_i^p \varphi_j^p \right\rangle \right)$$

and by using the expression in (2.15) we have

$$\left\langle \varphi^{2p} \right\rangle = \frac{\Gamma(2n+1)\Gamma(2p+1)}{\Gamma(2n+2p+1)} , \tag{2.20}$$

$$\left\langle \varphi_i^p \varphi_j^p \right\rangle = \frac{\Gamma(2n+1)\Gamma^2(p+1)}{\Gamma(2n+2p+1)} , \tag{2.21}$$

which, once again, do not depend on the specific i and j. As a consequence, the variance of the average cost results

$$\operatorname{var}\left[\varepsilon_n^{(p)}\right] \equiv \left\langle \left(\varepsilon_n^{(p)}\right)^2 \right\rangle - \left\langle \varepsilon_n^{(p)} \right\rangle^2$$

$$= \frac{2\Gamma(2p+1) - (2+p^2)\Gamma^2(p+1)}{(2n)^{2p+1}} + o\left(\frac{1}{n^{2p+1}}\right) . \tag{2.22}$$

It is useful to stress out the dependence of the set of RVs $\boldsymbol{\varphi}$, i. e. the two point function brings non-zero contribution, arises from the sum of the spacings being constrained. Moreover, it is explicitly related to the weight function, that is (2.21) depends on the exponent p in $\varepsilon_n^{(p)}$.

Since at leading order $\langle \varphi^p \rangle = O(n^{-2p})$ this suggests, in the large n limit, the substitution $\varphi_i = \phi_i/(2n)$. By plugging in (2.16) we have

$$\hat{\rho}_n^{(1)}(\phi) = \left(1 - \frac{\phi}{2n}\right)^{2n-1} \theta(\phi)\theta(2n-\phi) \tag{2.23}$$

$$= e^{-\phi} \left[1 - \frac{\phi^2 - 2\phi}{4n} + o\left(\frac{1}{n}\right) \right] , \tag{2.24}$$

that is, in the large n limit, the spacings ϕ_i are exponentially distributed with $\langle \phi^p \rangle = \Gamma(p+1) = O(1)$. In this frame we can now explicit the *weak dependence* of the RVs. Using the result in (2.19) and taking the series expansion for $n \gg 1$ in (2.21), after the proper rescaling in n, we have

$$\langle \phi \rangle = 1 - \frac{1}{2n} + o\left(\frac{1}{n}\right) , \tag{2.25}$$

$$\langle \phi_i \phi_j \rangle = 1 - \frac{3}{2n} + o\left(\frac{1}{n}\right) , \tag{2.26}$$

that results in

$$\operatorname{cov}[\phi_i \phi_j] \equiv \langle \phi_i \phi_j \rangle - \langle \phi \rangle^2 \tag{2.27}$$

$$= -\frac{3}{2n} + o\left(\frac{1}{n}\right) . \tag{2.28}$$

As we could expect form a set of RVs with a constrained sum, the covariance is negative. This can be interpreted as the consequence of the fact that, given any instance of the problem, namely given any $2n$ points on the interval Λ, if any spacing is stretched to become larger, the others must shrink as a compensation for the sum being fixed.

Moreover, even looking at the leading order of covariance in (2.28), it vanishes when n approaches infinity. This suggests the RVs ϕ_i act more and more like IID when the number of spacings increases. Driven by this observation we can now try to establish an LDP for the RV $\varepsilon_n^{(p)}$.

3

A LARGE DEVIATION PRINCIPLE FOR THE INDEPENDENT CASE

In the previous chapter we have obtained the average properties of the one dimensional REMP and looked at the asymptotic behavior when the number of variables n is large. Since the RVs are *weakly dependent*, i. e. $\mathrm{cov}[\phi_i \phi_j] = O(n^{-1})$, here we want to formulate an LDP for the sum of powers of IID RV.

In the following we will take $\mathbf{X} = (X_1, \ldots, X_n)$ a set of n IID RVs and we will focus on giving an asymptotic expression for the tail probability

$$\overline{F}_n(x) = 1 - F_n(x) \equiv \Pr(S_n > x) \tag{3.1}$$

where

$$S_n = S_n(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i \,. \tag{3.2}$$

Firstly, it is useful to introduce ourselves in providing an LDP for a simple case, that is the sample mean of exponential RVs. Since a slight complication of this problem will be treated in the following sections, we will show how a small modification can result in a catastrophic outcome, with the loss of validity of Cramér's Theorem.

3.1 A SIMPLE CASE: THE SAMPLE MEAN OF EXPONENTIAL RANDOM VARIABLES

Let $\mathbf{X}$, be a set of IID RVs with common probability density function

$$\rho(x) = e^{-x}, \quad x \geqslant 0 \,. \tag{3.3}$$

It is quite easy to characterize the distribution of S_n: in fact we can check Cramér's condition is satisfied, that is the MGF is finite, provided that $\lambda < 1$. By explicitly computing the MGF

$$\begin{aligned} M(\lambda) &\equiv \left\langle e^{\lambda X} \right\rangle \\ &= \int_0^{+\infty} dx\, e^{-(1-\lambda)x} = \frac{1}{1-\lambda}, \end{aligned}$$

thus from (1.23) the SCGF results

$$K(\lambda) = \log \frac{1}{1-\lambda} \,. \tag{3.4}$$

We can now plug this result in (1.19)

$$\begin{aligned} I(x) &\equiv \sup_{\lambda<1}[\lambda x - K(\lambda)] \\ &= \lambda x - K(\lambda)|_{\frac{\partial}{\partial\lambda}K(\lambda)=x} \\ &= \left(1-\frac{1}{x}\right)x + \log\left[1+\left(1-\frac{1}{x}\right)\right]. \end{aligned}$$

which leads to the rate function

$$I(x) = x - 1 - \log x\,. \tag{3.5}$$

Thus, we can state a large deviation principle holds for S_n, with

$$\overline{F}_n(x) \sim x^n e^{-n(x-1)}\,. \tag{3.6}$$

As a result, large deviations from the expected value $\langle X\rangle = 1$ are exponentially damped in the size n of the sample, making a large value of $(S_n - \langle X\rangle)$ extremely unlikely.

3.2 THE CASE OF REMP

In the REMP studied in Chapter 2 we had to deal with the sample mean of powers of the RVs, namely we are interested in formulating an LDP for the quantity

$$S_n^{(p)}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^p\,, \quad p > 1\,, \tag{3.7}$$

by assuming the the set $\mathbf{X}$ are IID RVs, with common probability distribution $\rho_m : [0,1] \to \mathbb{R}^+$, $m \in \mathbb{N}$, given by

$$\rho_m(x) = 2m(1-x)^{2m-1}\,, \tag{3.8}$$

that is the distribution found in (2.16). By simply applying Cramér's Theorem in (1.23) to $S_n^{(p)}$ and taking the expansion of the exponential we have for the MGF

$$\begin{aligned} M_{p,m}(\lambda) \equiv \left\langle e^{\lambda x^p}\right\rangle_m &= \int_0^1 dx\,\rho_m(x)e^{\lambda x^p} \\ &= \sum_{k=0}^{\infty}\frac{\langle x^{kp}\rangle}{\Gamma(k+1)}\lambda^k \\ &= \sum_{k=0}^{\infty}\frac{\Gamma(2m+1)\Gamma(kp+1)}{\Gamma(k+1)\Gamma(2m+kp+1)}\lambda^k \end{aligned}$$

which is a regular finite function $\forall\lambda \in \mathbb{R}$, for $\rho_m(x)e^{\lambda x^p}$ being regular on the compact interval $[0,1]$. As an example, for $p = 2$, we have

$$M_{2,m}(\lambda) = {}_2F_2\left(\frac{1}{2}, 1; \frac{1}{2}+m, 1+m; \lambda\right)$$

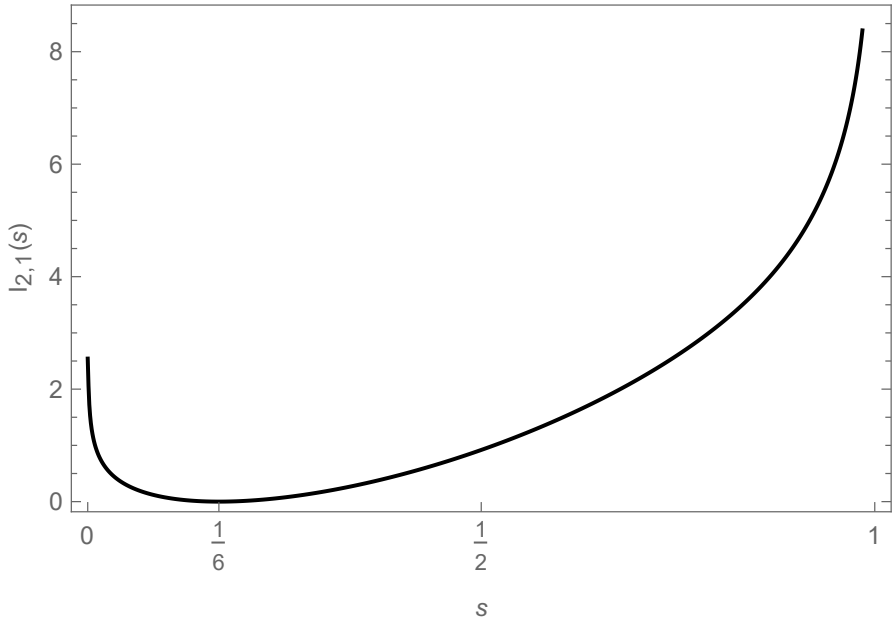


Figure 3.1: Rate function $I_{2,1}(s)$ of the sample mean $S_n^{(2)} = \sum_i X_i^2/n$ over the probability distribution $\rho_1(x) = 2(1-x)$. The function is obtained via numeric evaluation of the Legendre–Fenchel transform of the scaled cumulant generating function $K_{2,1}(\lambda)$. The function is defined on the open interval $(0,1)$ taking its minimum at $\langle X^2 \rangle = \frac{1}{6}$.

where ${}_2F_2$ denotes the *generalized hypergeometric function*, that is an entire function of the variable λ, defined by

$$ {}_2F_2\left(\frac{1}{2}, 1; \frac{1}{2}+m, 1+m; \lambda\right) \equiv \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_k (1)_k}{\left(\frac{1}{2}+m\right)_k (1+m)_k} \lambda^k $$

and $(a)_k$ is the *Pochhammer symbol*

$$ (a)_k \equiv a(a+1)\dots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)}, \quad k \geqslant 0\,. $$

The rate function $I_{p,m}(x)$ for the RV $S_n^{(p)}$ can be obtained as in (1.19) by a Legendre–Fenchel transform of the SCGF $K_{p,m}(\lambda) = \log M_{p,m}(\lambda)$

$$ I_{p,m}(s) = \sup_{\lambda} [\lambda s - K_{p,m}(\lambda)]\,. \tag{3.9} $$

Although an analytic expression is unfeasible, the inversion can be carried out numerically as in Figure 3.1.

Since an LDP always involve taking the limit for large number of RVs, we want to take the asymptotic expression of $\rho_m(x)$ for $m \gg 1$. By setting

$$ y = 2mx\,, \quad \hat{\rho}_m(y) = e^{-y} + O(m^{-1}) $$

as in (2.24) and taking the limit, we have that y is exponentially distributed, with finite moments given by

$$\mu_p \equiv \langle Z \rangle_p = \Gamma(p+1)\,, \tag{3.10}$$

$$\sigma_p^2 \equiv \langle Z^2 \rangle_p - \langle Z \rangle_p^2 = \Gamma(2p+1) - \Gamma(p+1)^2\,. \tag{3.11}$$

Thus, we have for the MGF referring to the RV $\widehat{S}_n^{(p)}(\mathbf{Y}) \equiv S_n^{(p)}(2m\mathbf{X}) = (2m)^p S_n^{(p)}(\mathbf{X})$

$$\widehat{M}_p(\lambda) \equiv \lim_{m\to\infty} M_{p,m}\left((2m)^p\lambda\right) = \int_0^{+\infty} dy\, e^{-y+\lambda y^p} \tag{3.12}$$

which clearly is a well defined function only for $\lambda \leqslant 0$. In this case Cramér's condition (i. e. the existence of the MGF for some positive λ) is violated and Cramér's Theorem does not apply. Anyway, by expanding the exponential $e^{\lambda y^p}$, we can express the MGF by the power series

$$\widehat{M}_p(\lambda) = \sum_{k=0}^{\infty} \frac{\Gamma(kp+1)}{\Gamma(k+1)}\lambda^k \tag{3.13}$$

that returns the right moments

$$\left\langle (Y^p)^k \right\rangle = \frac{\partial^k}{\partial\lambda^k}\widehat{M}_p(\lambda)\bigg|_{\lambda=0} = \Gamma(kp+1) < \infty\,, \quad \forall k \geqslant 0.$$

For example, this specializes in the $p = 2$ case to

$$\widehat{M}_2(\lambda) = \int_0^{+\infty} dy\, e^{-y+\lambda y^2} = e^{-\frac{1}{4\lambda}}\sqrt{-\frac{\pi}{4\lambda}}\,\overline{\Phi}\left(\frac{1}{\sqrt{-4\lambda}}\right) \tag{3.14}$$

where

$$\overline{\Phi}(x) = 1 - \Phi(x) \equiv \frac{2}{\sqrt{\pi}}\int_x^{+\infty} dz\, e^{-z^2}$$

denotes the *complementary error function*. A sketch of the function can be found in Figure 3.2. Here we want to stress out that the existence of the MGF on the positive real axis is a fundamental requirement for the proof of Cramér's Theorem. In particular, for the rate function to be properly defined, we need for the SCGF, and thus for the MGF, to be defined in a neighborhood of the origin. This is no surprise, since all the properties of the distribution, that is all the moments of the RV, can be retrived by subsequent derivations of $\widehat{M}_p(\lambda)$ at $\lambda = 0$.

Anyway, we can still extract a bound for the probability

$$F_n^{(p)}(s) \equiv \Pr\left(\widehat{S}_n^{(p)} \leqslant s\right)\,.$$

Recalling the Chernoff bound in (1.15), we have

$$\lim_{n\to\infty}\frac{1}{n}F_n^{(p)}(s) \leqslant -\sup_{t\leqslant 0}\left[-ts + \log\widehat{M}_p(-t)\right] \equiv -\widehat{I}_p^-(s)\,. \tag{3.15}$$

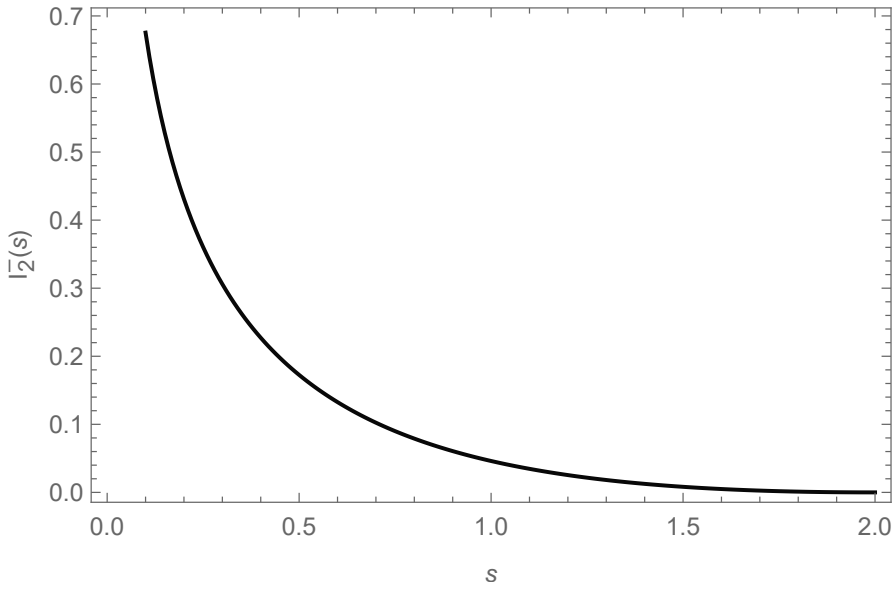


Figure 3.2: Plot of the rate function $\widehat{I}_2^-(s)$ of the sample mean $S_n^{(2)} = \sum_i Y_i^2/n$ over the probability distribution $\hat{\rho}(y) = e^{-y}$. The function is obtained via numeric evaluation of the Legendre–Fenchel transform of the scaled cumulant generating function $K_2(\lambda)$. The function is positive only in the region $s < \langle Y^2 \rangle$, while vanishing elsewhere, taking its minimum at $\langle Y^2 \rangle = \Gamma(3) = 2$.

Here $\widehat{I}_p^-(s)$ is very similar in definition to the rate function apart from taking the LFT only in the region of non-positive t. From convexity of $\log \widehat{M}_p(t)$ we have that $\widehat{I}_p^-(s) > 0$ in the region

$$s < \left.\frac{\widehat{M}_p'(t)}{\widehat{M}_p(t)}\right|_{t=0} = \langle Y^p \rangle = \Gamma(p+1)\,, \tag{3.16}$$

while $\widehat{I}_p^-(s) = 0$ otherwise. This implies oscillations of $\widehat{S}_n^{(p)}$ from the left-hand side of the average $\left\langle \widehat{S}_n^{(p)} \right\rangle = \langle Y^p \rangle$ are suppressed at least exponentially in the size n of the sample.

We are left now with the problem of formulating an LDP for big oscillations on the right-hand side (that is the *tail* of the distribution) of the RV $\widehat{S}_n^{(p)}$.

3.3 SUBEXPONENTIAL DISTRIBUTIONS

As it can be easily seen from (3.12), the existence of the MGF depends heavily on the nature of the tails of the PDF. Even in the simple example of a sample mean S_n of a set on IID RVs $\mathbf{X}$ we have

$$M(\lambda) = \left\langle e^{\lambda X} \right\rangle = \int dx\, \rho(x) e^{\lambda x} \tag{3.17}$$

that stays finite for some $\lambda > 0$ only if $\rho(x)$ decays at least exponentially for $x \gg 1$. A slower than exponential decay for $\rho(x)$ in the large x

limit prevents $M(\lambda)$ to be defined on the positive real axis. Such distributions are known as *subexponential* or *heavy-tailed distributions*, provided that $X > 0$. They are defined by the limit

$$\lim_{x\to\infty} \left| \frac{\bar{F}_n(x)}{n\bar{F}_1(nx)} - 1 \right| = 0\,, \quad n \in \mathbb{N}. \tag{3.18}$$

Let $X_n^* = \max_{i\leqslant n}[X_1, \ldots, X_n]$, then it can be easily seen that

$$\lim_{x\to\infty} \frac{\Pr(X_n^* > x)}{n\bar{F}_1(x)} = 1, \tag{3.19}$$

which allows for the interpretation that the large deviations of sums of independent heavy-tailed random variables are typically realized by just one of these variables taking a very large value. This is well known since the classical works of Heyde [17] and Nagaev [28, 29]. As (3.20) suggests, we can explicit the asymptotic expression for large values of x

$$\sup_{x\geqslant d_n} \left| \frac{\bar{F}_n(x)}{n\bar{F}_1(nx)} - 1 \right| = o(1)\,, \tag{3.20}$$

for a suitable sequence d_n.

Thus we are led to ask what kind of approximation to the tail probabilities $\bar{F}_n(x)$ can be expected in the finite x region. A natural bound comes from the CLT which implies, given $\langle X\rangle = \mu$, $\left\langle (X-\mu)^2 \right\rangle = \sigma^2 < \infty$,

$$\sup_x \left| \bar{F}_n(x) - \overline{\Phi}\left(\frac{\sqrt{n}(x-\mu)}{\sigma} \right) \right| = o(1)\,, \tag{3.21}$$

where, once again, Φ denotes the standard *error function*. This is formally identical to the formulation we gave in Section 1.2. The last relation can be rewritten as

$$\sup_{(x-\mu)\in[a_n,b_n]} \left| \frac{\bar{F}_n(x)}{\overline{\Phi}\left(\frac{\sqrt{n}(x-\mu)}{\sigma} \right)} - 1 \right| = o(1), \tag{3.22}$$

for

$$a_n = \frac{a}{\sqrt{n}}\,, \quad b_n = \frac{b}{\sqrt{n}}\,, \quad a, b \in \mathbb{R}\,. \tag{3.23}$$

We can expect that an asymptotic expression of this type may hold even for large deviations, namely that exists a sequence $c_n \geqslant b_n$, with $c_n\sqrt{n} \to \infty$ sufficiently slowly and such that (3.22), with $a_n = O(n^{-1/2})$, holds. In his famous work, Cramér [7, 8] proved that, given the existence of the moment generating function of X in a neighborhood of the origin, (3.22) holds with $a_n = O(n^{-1/2})$and $c_n = o(n^{-1/3})$, while (3.22) fails in general for $c_n = O(n^{-1/3})$.

From the previous discussion it seems typical for $\overline{F}_n(x)$ (and it actually is, see [25]) that there exist two threshold sequences $c_n \leqslant d_n$ such that

$$\overline{F}_n(x) \sim \begin{cases} \overline{\Phi}\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right) & x-\mu \ll c_n \\ n\overline{F}_1(nx) & x-\mu \gg d_n \end{cases} . \tag{3.24}$$

The rigorous treatment of where these sequences arise from would require a number of technical tools; we refer to [25] for the details. Despite this, an heuristic argument can be provided: from the previous discussion it may have become clear that there exist different types of large deviation results on different intervals, where either the CLT applies or the extremes in the sample dominate $\overline{F}_n(x)$. A separating sequence c_n can be expected at the border, where both the CLT and the extremal behavior overlap, i. e. where

$$\overline{F}_n(x) \sim \overline{\Phi}\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right) \sim \Pr(X_n^* > x) \sim n\overline{F}_1(nx) . \tag{3.25}$$

Thus a natural definition of b_n comes from the relation

$$\overline{\Phi}\left(\sqrt{n}c_n\right) \sim n\overline{F}_1(nc_n) ,$$

which implies

$$\overline{\Phi}\left(\sqrt{n}(x-\mu)\right) = o\left(n\overline{F}_1(nx)\right) , \quad x \gg c_n .$$

On the other hand, the estimate of d_n is rather more complicated. It can be retrieved both from *extreme value theory* arguments or more simply by looking at the distribution of $S_n - X_n^*/n$, conditional upon $S_n = x$, that is

$$\Pr\left(S_n - \frac{X_n^*}{n} = u \,\middle|\, S_n = x\right) = n\frac{p_{n-1}(u)}{p_n(x)}p_1(nx-nu)$$

In this case the threshold sequence d_n arise from $p_{n-1}(u)$, i. e. the quantity $S_n - X_n^*/n$, being asymptotically negligible. This implies that the large deviation of S_n occurs only on account of X_n^*.

These heuristic arguments explain that the maximum of the sample begins to have influence on the large deviations of S_n for $x \gg c_n$, and that it dominates the large deviations when $x \gg d_n$. In the region (c_n, d_n), the partial sums and the extremes have influence. Therefore, in the latter region, explicit asymptotic expressions for $\overline{F}_n(x)$ are quite difficult to obtain. The choice for the two threshold sequences heavily depend on the nature of the tail of the distribution. A complete case study has been collected by Nagaev and Mikosh [25] where a number of known results are presented.

3.3.1 *Large deviations for sample mean of stretched exponential random variables*

Let $\mathbf{X}$ be a set of n IID RVs distributed according to the exponential distribution $\rho(x)$ as in (3.3) and $S_n^{(p)}(\mathbf{X})$ the sample mean of X_i^p as in (3.7).
After a proper rescaling $X_i^p = Z_i$, this problem is formally identical to the one of considering the sample mean $S_N(\mathbf{Z})$ of the set of IID RVs $\mathbf{Z}$, sorted according to the probability density

$$\tilde{\rho}^{(p)}(z) = \frac{z^{(1-p)/p}}{p} e^{-z^{1/p}} \qquad z \geqslant 0\,. \tag{3.26}$$

Distributions of the type in (3.26) are called *stretched exponential distributions* due to a slower than exponential decay. This kind of distributions clearly exhibit a subexponential tail in the large z region. As a check we can evaluate

$$\begin{aligned}
\Pr(Z_n^* > z) &= 1 - \Pr(Z_n^* < z) \\
&= 1 - \Pr(Z_1 < z, Z_2 < z, \ldots, Z_n < z) \\
&= 1 - \prod_{i=1}^{n} \Pr(Z_i < z)
\end{aligned}$$

which leads to

$$\Pr(Z_n^* > z) = 1 - \left[1 - e^{-z^{1/p}}\right]^n . \tag{3.27}$$

Thus, by replacing in (3.19), we have

$$\begin{aligned}
\lim_{z\to\infty} \frac{\Pr(Z_n^* > z)}{n\bar{F}_1(z)} &= \lim_{z\to\infty} \frac{1 - \left[1 - e^{-z^{1/p}}\right]^n}{n e^{-z^{1/p}}} \\
&= \lim_{z\to\infty} \frac{1 - e^{-n e^{-z^{1/p}}}}{n e^{-z^{1/p}}} \\
&= 1\,.
\end{aligned}$$

As a consequence we can state the tail probability $\bar{F}_n^{(p)}(x)$ for the sum $S_n(\mathbf{Z}) = S_n^{(p)}(\mathbf{X})$ takes the asymptotic form

$$\bar{F}_n^{(p)}(x) \sim \bar{\Phi}\left(\frac{\sqrt{n}(x - \mu_p)}{\sigma_p}\right), \quad x - \mu_p \ll c_n^{(p)}$$

with μ_p, σ_p given in (3.10) and (3.11), while

$$\bar{F}_n^{(p)}(x) \sim n\bar{F}_1^{(p)}(nx) = n e^{-(nx)^{1/p}}, \quad x - \mu_p \gg d_n^{(p)}\,.$$

This implies large deviations are less than exponentially dumped in the size n of the sample.

It can be shown the threshold sequences are given by

$$c_n^{(p)} = \begin{cases} c \cdot n^{-\frac{1}{3}} & p \in (1,2) \\ c \cdot n^{\frac{1-p}{2p-1}} & p \geqslant 2 \end{cases}, \tag{3.28}$$

$$d_n^{(p)} = d \cdot n^{\frac{2-p}{2p-2}}, \tag{3.29}$$

for any $c, d \in \mathbb{R}^+$.
As it can be easily seen, $c_n^{(p)}$ always vanishes in the large n limit, being $c_n^{(p)} \leqslant O(n^{-1/3})$, due to the concentration of probability measure, that is the CLT. On the other hand, the behavior of the threshold sequence $d_n^{(p)}$ strongly depends on the parameter p that determines the behavior of the heavy-tail. In fact, by taking the exponent in (3.29)

$$\frac{2-p}{2p-2} > 0 \quad \Longleftrightarrow \quad p \in (1,2). \tag{3.30}$$

Hence, provided that $p \in (1,2)$, the stretched exponential behavior of $\bar{F}_n^{(p)}(x)$ is recovered only for extremely large deviations, namely $(x - \mu_p) \gg n^\delta$, $\delta > 0$.

As a check, as $p \to 1$, that is as $S_n^{(p)}$ approaches the sample mean S_n of exponentially sorted random variables, $d_n^{(p)} \to \infty$ and no stretched exponential decay occurs. This is strongly related to the fact that Cramér's condition still holds for distributions with proper exponential decay like in (3.3). Thus the moment generating function is not ill-defined and a rate function of the type in (3.5) can be recovered.

3.4 THE LARGE DEVIATION PRINCIPLE FOR SUM OF POWERS OF INDEPENDENT EXPONENTIAL VARIABLES

Here we want to collect the results we found in the independent case. We started from the simple case of the sample mean S_n of IID RVs sorted according to the exponential distribution $\rho(x) = e^{-x}$. This case present no difficulties, for the MGF being finite up to a certain value on the positive real axis, i. e. Cramér's condition holds. By applying Cramér's Theorem we found the moment generating function $I(x)$ in (3.5) that assures large deviations from the mean value are suppressed exponentially in the number n of RVs in S_n.

We found the same procedure applies to the distribution $\rho_m(x)$ in (3.8) (i. e. the marginal probability in the REMP) for the RV $S_n^{(p)}$ as defined in (3.7), for finite values of m. Even in this case Cramér's Theorem applies and the MGF $M_{p,m}(\lambda)$ can be expressed as a power series of λ. Despite the fact that $I_{m,p}(x)$ has no simple analytic expression, it can be computed by numerically evaluating the LFT of the SCGF, as defined in (3.9), which leads for example to the RF in Figure 3.1.

As a result, once again we have an exponential decay in n for the tail probability, that is

$$\lim_{n\to\infty} \frac{1}{n} \log \overline{F}^{(p)}_{n,m}(s) = -I_{m,p}(s) \tag{3.31}$$

Since the parameter m in the distribution $\rho_m(x)$ controls the number of intervals in the REMP, that is the number of the RVs in the sum $S^{(p)}_n$, we took the limit for large m. By rescaling the RVs $\mathbf{Y} = 2m\mathbf{X}$ we found

$$\lim_{m\to\infty} \rho_m\left(y/(2m)\right) \equiv \hat{\rho}(y) = e^{-y}\,. \tag{3.32}$$

Thus we focused on finding an asymptotic expression for the probability $\overline{F}^{(p)}_n(s) \equiv \Pr\left(\widehat{S}^{(p)}_n(\mathbf{Y}) \geqslant s\right)$, with the set of RVs $\mathbf{Y}$ sorted according to $\hat{\rho}(y)$, as in (3.3). This time we found that Cramér's condition is violated, that is the MGF $\widehat{M}_p(\lambda)$ is defined only in the region $\lambda \leqslant 0$. Despite this, by using the Chernoff bound for the probability $F^{(p)}_n(s) \equiv 1 - \overline{F}^{(p)}_n(s)$, we had

$$\lim_{n\to\infty} \frac{1}{n} F^{(p)}_n(s) \leqslant -\widehat{I}^{-}_p(s)\,,$$

where $\widehat{I}^{-}_p(s)$ is a regular positive function in the region $s < \langle Y^p\rangle \equiv \mu_p = \Gamma(p+1)$, while it vanishes otherwise. In this case we have an LDP only on the left-hand side from the expected value of $\widehat{S}^{(p)}_n$, where the probability of deviations exhibits an exponential dumping in n, while the right-hand side, that is for $s > \langle Y^p\rangle$, has no proper bound. This suggests a different regime from the exponential one for the probability $\overline{F}^{(p)}_n(s)$ must be taken into account.

From the analysis made in Section 3.3 it should have become clear that the problem of formulating an LDP for the RV $\widehat{S}^{(p)}_n(\mathbf{Y})$ is identical to the one of finding the asymptotic expression for the probability $\Pr(S_n(\mathbf{Z}) > s)$, where the set of RVs $\mathbf{Z}$ is distributed according to the PDF $\tilde{\rho}^{(p)}(x) \sim e^{-x^{1/p}}$ as in (3.26). Here we found two threshold sequences $c^{(p)}_n$, $d^{(p)}_n$ such that, for $(s-\mu_p) \ll c^{(p)}_n$ the Gaussian regime is still valid, even beyond the usual $(s-\mu_p) = O(n^{-1/2})$, that is the well known CLT. Moreover, for $(s-\mu_p) \gg d_n$, we observed the distribution of $\widehat{S}^{(p)}_n(\mathbf{Y})$ is influenced by the maximum of the RVs taking a very large value, that is the tail probability $\overline{F}^{(p)}_n(s) \sim n\overline{F}^{(p)}_1(ns)$.

In conclusion we can state that, given a set of RVs $\mathbf{Y}$ distributed according to the exponential distribution

$$\hat{\rho}(y) = e^{-y}$$

an LDP holds for the RV $\widehat{S}^{(p)}_n(\mathbf{Y})$ with different bound in different regions. In particular

$$F^{(p)}_n(s) \leqslant e^{-n\widehat{I}^{-}_p(s)}\,, \quad s-\mu_p \leqslant 0\,, \tag{3.33}$$

with $I_p^-(s)$ as given in (3.15), while

$$\overline{F}_n^{(p)}(s) \sim \begin{cases} \overline{\Phi}\left(\frac{\sqrt{n}(s-\mu_p)}{\sigma_p}\right) & 0 < s - \mu_p \ll c_n^{(p)} \\ ne^{-(ns)^{1/p}} & s - \mu_p \gg d_n^{(p)} \end{cases}, \tag{3.34}$$

with μ_p, σ_p given in (3.10) and (3.11), and with the threshold sequences $c_n^{(p)}$, $d_n^{(p)}$ as given in (3.28) and (3.29).

Thus the distribution of $\widehat{S}_n^{(p)}(\mathbf{Y})$ exhibits a different scaling with n on the left and on the right-hand side of the average value μ_p. In particular, while in the region $s \ll c_n^{(p)}$ deviations are always suppressed exponentially in n, we have for $s \gg d_n^{(p)}$ that the probability assumes a stretched exponential behavior, with scaling speed $n^{1/p}$, making extreme values more likely.

3.5 NUMERICAL SIMULATIONS

In this section we want to discuss the numerical methods to have an estimate of the shape of the rate function. This problem is quite difficult to approach, mainly because of the nature of large deviations. From the previous discussion it should have become clear that LDT has to deal with extreme value theory and the probability of rare events, while the number of RVs is very large. Since from CLT we have that probability concentrates around the typical values of the distribution as the number of RVs increases, sampling rare events, that is extracting a precise asymptotic trend for the tail probability, can result in an unfeasible task. In other words, given the sample mean S_n of RVs, regardless the speed of the scaling, large deviation probabilities are always exponentially suppressed in n, while, to obtain an accurate approximation for the rate function $I(s)$, we would need n to be large. This comes from the asymptotic expression of the PDF for large values of n retaining only the dominant term in the exponential, while neglecting subleading orders.

There are different techniques to treat this problem: here we refer to one of the most simple tools we can imagine to extract a rate function from numerical simulations, that is the *direct sampling*.

3.5.1 *Direct sampling method*

The problem addressed here is to obtain a numerical estimate of the PDF $p_{S_n}(s)$ for the real RV S_n satisfying an LDP, and to extract from this an estimate of the rate function $I(s)$. To be general, we will consider $S_n = S_n(\mathbf{X})$ to be a function of the set $\mathbf{X}$ of n RVs, which, at this point, are not necessarily IID.

Numerically, we cannot of course obtain $p_{S_n}(s)$ or, equivalently, $I(s)$ for all $s \in \mathbb{R}$, but only for a finite number of values s, which we take

for simplicity to be equally spaced with a small step Δs. Thus, we can estimate the coarse-grained PDF

$$p_{S_n}(s) = \frac{\Pr\left(S_n \in [s, s+\Delta s]\right)}{\Delta s} = \frac{\Pr(S_n \in \Delta_s)}{\Delta s}, \tag{3.35}$$

where $\Delta_s \equiv [s, s+\Delta s]$ denotes the interval of amplitude Δ anchored to the value s.

To construct this estimate, we follow the *statistical sampling* or *Monte Carlo method*, which we can broke down into the following steps

- generate the sample $\{\mathbf{X}^{(j)}\}_{j=1}^{L}$ of L copies of the sequence $\mathbf{X}$ from its PDF;
- obtain from this sample, the set $\{s^{(j)}\}_{j=1}^{L}$ of the realizations of S_n, that is

  $$s^{(j)} = S_n(\mathbf{X}^{(j)});$$

- estimate the probability $\Pr(S_n \in \Delta_s)$ by evaluating the sample mean

  $$P_L(\Delta_s) \equiv \frac{1}{L}\sum_{j=1}^{L} \chi_{\Delta_s}\left(s^{(j)}\right),$$

  where $\chi_A(x)$ denotes the indicator function of the set A, that is

  $$\chi_A(x) \equiv \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases};$$

- use the sample mean $P_L(\Delta_s)$ to estimate the probability distribution of S_n:

  $$p_L(s) \equiv \frac{P_L(\Delta_s)}{\Delta_s} = \frac{1}{\Delta_s L}\sum_{j=1}^{L} \chi_{\Delta_s}\left(s^{(j)}\right).$$

Note that $p_L(s)$ above is nothing but an empirical vector for S_n or, equivalently, a histogram normalized over the total counts of the sample $\{s^{(j)}\}_{j=1}^{L}$. The reason for choosing $p_L(s)$ as our estimator of $p_{S_n}(s)$ is that it is an unbiased estimator, in the sense that

$$\langle p_L(s)\rangle = p_{S_n}(s)$$

for any L. Moreover, we know from the LLN that $p_L(s)$ converges in probability to its mean $p_{S_n}(s)$ as L increases. Therefore, the larger our sample, the closer we should get to a valid estimation of $p_{S_n}(s)$.

The rate function can be easily computed by recalling the symptotic expression of the PDF, that leads to

$$I_n^{(L)}(s) = \frac{1}{n}\log p_L(s). \tag{3.36}$$

We can repeat the whole process for larger and larger integer values of n and L to improve the accuracy of the rate function.

This method presents a severe limitation: a basic rule in statistical sampling, suggested by the LLN, is that an event with probability P will appear in a sample of size L roughly LP times. Thus to get at least one instance of that event in the sample, we must have $L > P^{-1}$, as an approximate lower bound for the size of our sample. In terms of LDT we see that if a RV S_n satisfies an LDP with rate function $I(s)$ and speed n then we would need $L > e^{nI(s)}$ to see just one event. As a consequence, increasing the size n of the sample mean to maximize accuracy, determines the number of instances L to get exponentially large.

3.5.2 *Simulation results*

Here we collect the results of the numerical computations in the independent case. The simulations presented here are obtained via the direct sampling method discussed in the previous paragraph. As a consequence of the limitations of this procedure, the plots should be considered as a qualitative check, without claiming to confirm nor reject the rigorous results obtained in Section 3.4.

As a first check, we extrapolated the rate function $I_p^-(s)$ for the RV $S_n^{(p)}$ we found in (3.15) from the direct sampling of the exponential distribution. As an example, in Figure 3.3 the plot of the rate function $I_2^-(s)$ is shown in gray with the estimated rate functions, as in (3.36), for fixed $L = 10^7$ instances of the sample mean $S_n^{(2)}$ and for different values of n. As we can see, the larger the value of n the closer the points are to the expected rate function. On the other hand, it is important to notice that the range of the simulations, that is the interval on which the estimated rate function is defined, decreases as the number of RVs n accounted for in $S_n^{(2)}$ becomes larger. This directly follows from the deterioration of the statistics far away from the expected value due to the direct sampling method: in fact, as n increases we are probing only the (small) Gaussian deviations, that is the CLT.

Moreover, we can check the exponential scaling of the tail probability $\bar{F}_n^{(p)}(s)$ is $n^{1/p}$, as we found in (3.34). An estimation of $\bar{F}_n^{(p)}(s) = \Pr\left(S_n^{(2)} > s\right)$ can be obtained in the same fashion of the previous paragraph, by counting the number of instances having a realization $s^{(j)} > s$ and normalizing over the total number of instances L.
By recalling that

$$\bar{F}_n^{(p)}(s) \sim e^{-n^{1/p}(x-\mu_p)^{1/p}},$$

we can extract the exponent of n from the relation

$$\log\left[-\log \bar{F}_n^{(p)}(s)\right] \equiv \log Q_n^{(p)}(s) = \frac{1}{p}\log n + \log(x-\mu_p)^{1/p}. \quad (3.37)$$

Thus, from a linear regression in $\log n$, we can obtain the exponent as the slope of the fitting line. An example is shown in Figure 3.4 for the case $p = 5$, which we expect to show a more pronounced subexponential tail, with $L = 10^7$ number of instances and different values of $\log n$.

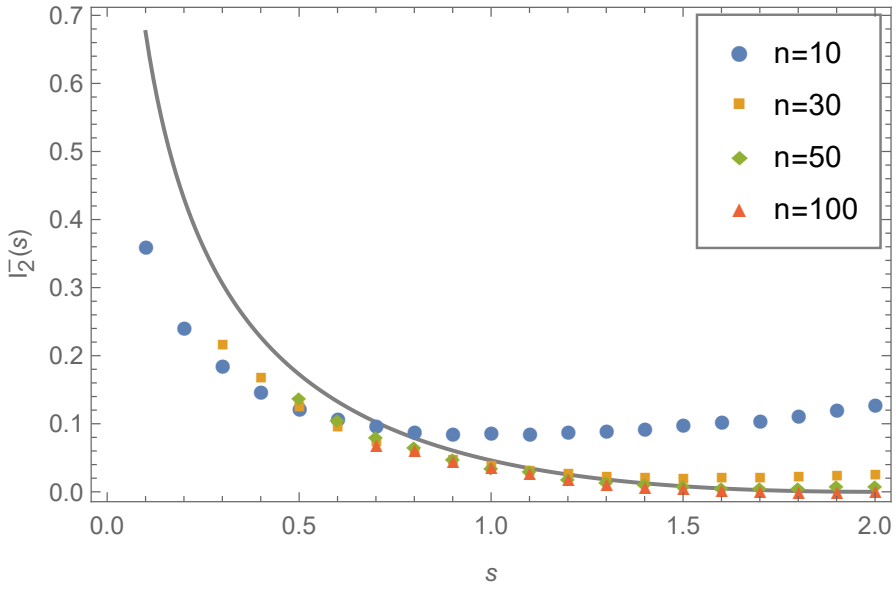


Figure 3.3: Plot of the rate function $\widehat{I}_2^-(s)$ as in (3.15) of the sample mean $S_n^{(2)} = \sum_i Y_i^2/n$ over the exponential PDF $\hat{\rho}(y) = e^{-y}$ and the estimated rate functions from *direct sampling method* for different values of n. As the number of RVs n increases the estimated rate functions (i. e. the PDF of S_n) get closer to the asymptotic expression $\widehat{I}_2^-(s)$, while the range of definition shrinks due to concentration of measure, that is the central limit theorem.

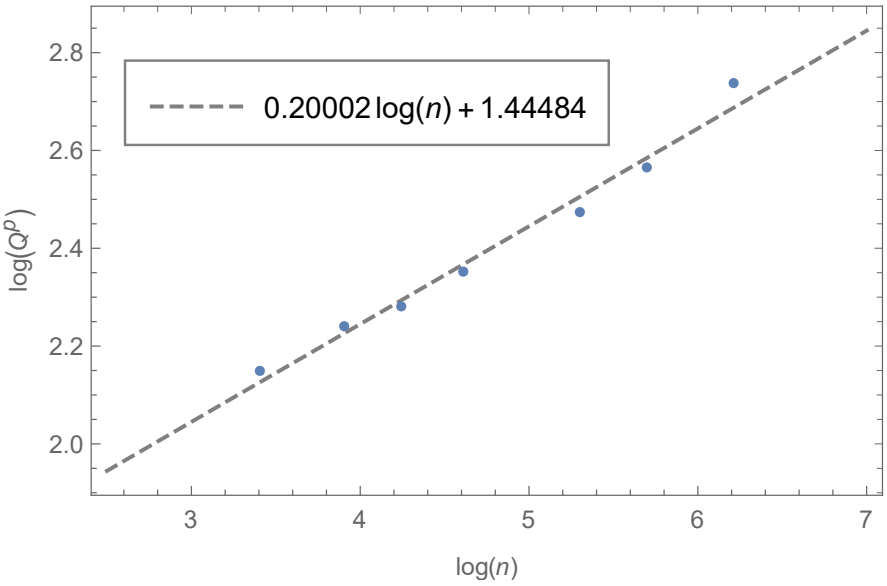


Figure 3.4: Linear regression as defined in (3.37) to check the scaling speed for the tail probability $\overline{F}_n^{(5)} = \Pr\left(S_n^{(5)} > s\right)$ in the subexponential region ($p = 5$). The slope of the fitting line coincide with the exponent of n in the asymptotic expression of $\overline{F}_n^{(5)} \sim \exp\{-n^{1/5}(x - \Gamma(6))^{1/5}\}$ as in (3.34). The result matches with the expected value $p^{-1} = 5^{-1} = 0.2$.

4

A LARGE DEVIATION PRINCIPLE FOR THE AVERAGE COST OF ONE DIMENSIONAL REMP

In this section we collect the main results we obtained for the distribution of the average cost of the one dimensional matching problem discussed in Chapter 2 in the large n limit. We start by proving the total cost in REMP is asymptotically normally distributed, that is the probability measure concentrates around the typical value, as stated by the central limit theorem. We show that actually the Gaussian regime can be extended to *moderate* large deviations, for a suitable threshold sequence $c_n^{(p)}$ as in the case of independent identically distributed random variables. A short comment on numerical simulations is provided at the end of this chapter, where some interesting features of the distribution can be noticed.

4.1 THE CENTRAL LIMIT THEOREM FOR DEPENDENT VARIABLES

In the following we want to explicit that the limiting distribution for $n \to \infty$ of the total optimal cost $\mathcal{E}(\mu^*)$ for the REMP, as gven in (2.6), is Gaussian. Although this is not a result of LDT, we want to stress that the CLT does not apply in its standard formulation to dependent variables, as the one we are considering in our problem. Here we refer to the work of Darling [9] which proved this limit holds for the generalized sum

$$W_n^{(h)} = \sum_{i=0}^{2n} h(\varphi_i) \tag{4.1}$$

for a wide range of real-valued functions h.

4.1.1 *The fundamental formula*

Let us start by noting that, given $\boldsymbol{\varphi} = (\varphi_0, \dots, \varphi_{2n})$ the set of subintervals in which the unit interval $\Lambda = [0, 1]$ is divided by $2n$ random points, for $\mathbf{f} = (f_0, \dots, f_{2n})$ a set of real-valued function, it holds

$$\left\langle \prod_{j=0}^{2n} f_j(\varphi_j) \right\rangle = \frac{(2n)!}{2\pi i} \int_{c-i\infty}^{c+i\infty} dz\, e^z \left[\prod_{j=0}^{2n} \int_0^\infty dr_j\, e^{-r_j z} f_j(r_j) \right] . \tag{4.2}$$

Here c is a constant larger than all the abscissas of convergence of the corresponding Laplace transforms of the f_i, the path of integration being on the complex plane $\mathrm{Re}(z) = c$. The formula in (4.2) can be

obtained by observing that the expectation value of the product of $f_i(\varphi_i) = f_i(x_{i+1} - x_i)$, $0 = x_0 \leqslant x_1 \leqslant x_2 \cdots \leqslant x_{2n} \leqslant x_{2n+1} = 1$, is given by

$$\left\langle \prod_{j=0}^{2n} f_j(\varphi_j) \right\rangle = (2n)! \int_0^1 dx_{2n} \int_0^{x_{2n}} dx_{2n-1} \cdots \int_0^{x_2} dx_1 \, f_0(x_1) f_1(x_2 - x_1) \dots f_{2n}(1 - x_{2n}) \, .$$

This is simply the convolution $f_0 * f_1 * \cdots * f_{2n}(1)$, where

$$g * h(x) \equiv \int_0^x dt \, g(x-t) h(t) \, .$$

Hence, by recalling that the Laplace transforms multiply under convolution, we obtain

$$\int_0^\infty dx \, f_0 * f_1 * \cdots * f_{2n}(x) e^{-zx} = \prod_{j=0}^{2n} \int_0^\infty dr_j \, f_j(r_j) e^{-z r_j} \, .$$

Now, by applying the complex inversion for the Laplace transform and setting $x = 1$, we obtain (4.2).

The formula we just found is completely general and allows to evaluate numerous expectation values over the probability distribution of uniform spacings. By simply setting $f_i(\varphi_i) = e^{i\xi h(\varphi_i)}$ we have

$$\begin{aligned} \left\langle e^{i\xi W_n^{(h)}} \right\rangle &\equiv \left\langle \prod_{i=0}^{2n} e^{i\xi h(\varphi_i)} \right\rangle \\ &= \frac{(2n)!}{2\pi i} \int_{c-i\infty}^{c+i\infty} dW \, e^W \left[\int_0^\infty dr \, e^{-rW + i\xi h(r)} \right]^{2n+1} , \end{aligned} \tag{4.3}$$

that is the *characteristic function*, i. e. the Fourier transform of the distribution, for the generalized sum $W_n^{(h)}$.

4.1.2 *The Gaussian regime*

The proof of asymptotic normality of the distribution of the total cost in REMP follows directly from the evaluation of the characteristic function. It is useful to recall that in the REMP we have a set of $(2n+1)$ RVs, namely the subintervals $\boldsymbol{\varphi}$, while the total optimal cost $\mathcal{E}_n^{(p)} = n\varepsilon_n^{(p)}$ is evaluated by summing only on the set $\boldsymbol{\varphi}_{\text{odd}} = (\varphi_1, \dots, \varphi_{2n-1})$. Thus, from (4.3) we have

$$\begin{aligned} \mathcal{F}_n^{(p)}(\xi) &\equiv \left\langle e^{i\xi \mathcal{E}_n^{(p)}} \right\rangle \\ &= \frac{(2n)!}{2\pi i} \int_{c-i\infty}^{c+i\infty} d\mathcal{E} \, e^{\mathcal{E}} \left[\int_0^\infty dr \, e^{-r\mathcal{E} + i\xi r^p} \right]^n \left[\int_0^\infty dr \, e^{-r\mathcal{E}} \right]^{n+1} . \end{aligned} \tag{4.4}$$

Now, letting

$$\xi = (2n+1)^{p-1/2}t\,, \quad \mathcal{E} = (2n+1)z\,,$$

and shifting the contour parallel to itself we have

$$\mathcal{F}_n^{(p)}\left((2n+1)^{p-1/2}t\right) = \frac{(2n+1)!}{(2n+1)^{2n+1}}\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} dz\,\frac{e^{(2n+1)z}}{z^{2n+1}}\left[B_n(z,t)\right]^n\,, \quad (4.5)$$

where

$$B_n(z,t) \equiv (2n+1)z\int_0^\infty dr\, e^{-r(2n+1)z+it(2n+1)^{p-1/2}r^p}\,. \quad (4.6)$$

As it will turn out, $[B_n(z,t)]^n$ is actually a bounded function with a well defined limit for $n \to \infty$ for t bounded and $\forall z$. Thus the integral in (4.5) can be evaluated by just taking the dominant term of $e^{(2n+1)z}z^{-(2n+1)} = e^{(2n+1)f(z)}$ by a saddle point approximation, with $f(z) = z - \log z$. Now, solving for $f'(z) = 0$, we have $z = 1$, while $f''(1) = 1$. Hence in (4.5) we simply take $c = 1$ to get the contour of steepest descent. By setting

$$z = 1 + \frac{iy}{\sqrt{2n+1}} \quad (4.7)$$

with

$$y \in I_n = \left[-(2n+1)^\delta, (2n+1)^\delta\right] \qquad \delta \in \left(0, \frac{1}{2}\right)\,, \quad (4.8)$$

the integral has its essential contribution in this range. It is useful to stress out the Gaussian regime results as a consequence of this approximation, for the CLT being valid in an interval around the average value of order $\sqrt{n}$. Plugging the expression of $z = z(y)$ as in (4.7) in (4.5) we have for the constants preceding the integral

$$\frac{(2n+1)!\,e^{2n+1}}{(2n+1)^{2n+\frac{3}{2}}}\frac{1}{2\pi} = \frac{1}{\sqrt{2\pi}} + o(1)$$

which follows from Stirling's approximation of the factorial $(2n+1)! = \Gamma(2n+2)$. Thus, for the characteristic function in (4.5), we have

$$\mathcal{F}_n^{(p)}\left((2n+1)^{p-1/2}t\right) = \int_{I_n}\frac{dy}{\sqrt{2\pi}}e^{-y^2/2}\left[B_n(z,t)\right]^n\left[1+o(1)\right]\,, \quad (4.9)$$

Now we take the asymptotic expansion in (4.6) that gives

$$\begin{aligned} B_n(z,t) &= (2n+1)z\int_0^\infty dr\, e^{-r(2n+1)z} \\ &\quad\times\left(1 + it(2n+1)^{p-1/2}r^p + \frac{(it)^2}{2}(2n+1)^{2p-1}r^{2p} + \dots\right) \\ &= 1 + (it)\frac{\Gamma(p+1)}{z^p\sqrt{2n+1}} + \frac{(it)^2}{2}\frac{\Gamma(2p+1)}{z^{2p}(2n+1)} + o(n^{-1})\,, \end{aligned}$$

and letting $z = z(y)$ we get

$$\begin{aligned} n \log B_n(z,t) &= n \log \left[1 + (it) \frac{\Gamma(p+1)}{z^p \sqrt{2n+1}} + \frac{(it)^2}{2} \frac{\Gamma(2p+1)}{z^{2p}(2n+1)} + o(n^{-1}) \right] \\ &= (it) \frac{\Gamma(p+1)}{\sqrt{2}} \sqrt{n} + ty \frac{p\Gamma(p+1)}{2} \\ &\quad + \frac{(it)^2}{2} \frac{\Gamma(2p+1) - \Gamma(p+1)^2}{2} + o(1) . \end{aligned}$$

Plugging this estimate in (4.9) we obtain

$$\begin{aligned} &\mathcal{F}_n^{(p)} \left((2n+1)^{p-1/2} t \right) e^{-it \frac{\Gamma(p+1)}{\sqrt{2}} \sqrt{n}} \\ &\quad = e^{-\frac{t^2}{2} \frac{\Gamma(2p+1) - \Gamma(p+1)^2}{2}} \int_{I_n} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2 + ytp\Gamma(p+1)/2} \left(1 + o(1)\right) , \end{aligned}$$

and by taking the limit of large n we get

$$\begin{aligned} &\lim_{n\to\infty} \left\langle \exp\left\{ it \left((2n+1)^{p-\frac{1}{2}} \mathcal{E}_n^{(p)} - \frac{\Gamma(p+1)}{\sqrt{2}} \sqrt{n} \right) \right\} \right\rangle \\ &\quad = \lim_{n\to\infty} \left\langle \exp\left\{ it(2n)^{p-1/2} \left[\mathcal{E}_n^{(p)} - \frac{\Gamma(p+1)}{2^p n^{p-1}} \right] \right\} \right\rangle \\ &\quad = \exp\left\{ -\frac{t^2}{2} \frac{2\Gamma(2p+1) - 2\Gamma(p+1)^2}{4} \right\} \\ &\qquad \times \int_{-\infty}^{\infty} \frac{dy}{\sqrt{2\pi}} e^{-y^2/2 + ytp\Gamma(p+1)/2} \\ &\quad = \exp\left\{ -\frac{t^2}{2} \frac{2\Gamma(2p+1) - (2+p^2)\Gamma(p+1)^2}{4} \right\} . \end{aligned}$$

Since the Fourier transform of a normal distribution is normal distributed in the conjugate variable, we have established the assumption. As a consequence, we have also an independent derivation for the asymptotic moments given in (2.19) and (2.22).

4.2 THE EXTENSION TO LARGE DEVIATIONS

Here we collect the results of large deviations which we obtained for the average cost for the REMP. The main achievement is the extension of Gaussian regime we just extracted to a broader interval.

4.2.1 *The Gaussian moderate deviations*

In this section we will try to show the asymptotic normality of the distribution of the average cost in the REMP can be extended in a larger region than the typical one of CLT. As in the previous chapter, to properly formulate an LDP for the average cost we would need a number of technical tools of probability theory. In here we refer to the work of Mirakhmedov [26, 27], that found a number of precise large

deviations for the generalized sum W_n, as in (4.1), under suitable conditions of the function h.

The topic seems to be of great interest for asymptotic efficiency of goodness-of-fit tests [15, 33]. These are aimed to confirm or reject the hypothesis that a certain data set is distributed according to some *a priori* PDF, which in this case is the uniform distribution on the unit interval Λ. Typically, this is achieved by looking at the discrepancy between the expected and empirical value of a generalized sum $W_n^{(h)}$, for some suitable function h of the spacings between successive points. Although a number of possible choices is given for the actual function h, it seems common to choose $h(x) = x^2$, called *Greenwood statistics*, named after Major Greenwood [16] who firstly introduced this test to estimate the distribution of genes in the chromosomes of living organisms. The asymptotic expression of the tail probability for this kind of statistics has a specialized result, as it will be shown subsequently.

After the discussion in Chapter 3, it should have become clear that the right observable to be taken into account to formulate an LDP is the rescaled RV $\boldsymbol{\phi} = 2n\boldsymbol{\varphi}$, which has a well defined distribution in the large n limit. In particular, we will refer to the average cost per edge of the REMP, namely to the quantity $\hat{\varepsilon}_n^{(p)} = (2n)^p \varepsilon_n^{(p)}$ whose limit moments, given in (3.10) and (3.11), are positive and finite.

As in Section 3.3, the validity of asymptotic normality for the distribution of the RV $\hat{\varepsilon}_n^{(p)}$ is constrained by a threshold sequence $c_n^{(p)}$. This defines the region where the approximation holds, namely how far from the expected value $\left\langle \hat{\varepsilon}_n^{(p)} \right\rangle$ the deviation can get for the distribution to be Gaussian. In this case the sequence arises from noting that the problem of finding the distribution of the random vector $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{2n})$ can be mapped in the one of $2n+1$ exponential RVs $\mathbf{Z} = (Z_0, \ldots, Z_{2n})$ conditional to their sum being $2n+1$. More explicitly, by denoting with $\mathcal{D}(\mathbf{X})$ the distribution of the random vector $\mathbf{X}$, we have

$$\mathcal{D}(\boldsymbol{\phi}) = \mathcal{D}(\mathbf{Z} | S_{2n+1}(\mathbf{Z}) = 1) \,, \tag{4.10}$$

where, as always, $S_n(\mathbf{X})$ stands for the sample mean of the random vector $\mathbf{X}$. In this case it can be proved that, for any $p > 1$, the asymptotic expression for the tail probability takes the form

$$\Pr\left(\hat{\varepsilon}_n^{(p)} > x\right) \sim \overline{\Phi}\left(\sqrt{\frac{n}{\operatorname{var}\left[\hat{\varepsilon}_n^{(p)}\right]}}(x - \mu_p)\right) \,, \quad 0 < x - \mu_p \ll c_n^{(p)} \,, \tag{4.11}$$

where

$$c_n^{(p)} = c_n = c \cdot \sqrt{\frac{\log n}{n}} \,, \quad c \in \mathbb{R}^+ \,, \tag{4.12}$$

while $\mu_p = \langle Z^p \rangle$ as given in (3.10) and $\operatorname{var}\left[\hat{\varepsilon}_n^{(p)}\right]$ as in (2.22). It is important to notice that the expectation value of $\hat{\varepsilon}_n^{(p)}$ is exactly μ_p only in the large n limit, namely

$$\lim_{n\to\infty} \left\langle \hat{\varepsilon}_n^{(p)} \right\rangle = \mu_p \,.$$

This fact arises exactly from (4.10), the difference of the two distribution being $o(1)$ in the region defined by (4.12).

This result can be specialized to the case $p = 2$, that is the Greenwood statistics: in his work Mirakhmedov shows that the cumulants of the RV $\hat{\varepsilon}_n^{(2)}$ and the ones of $S_n^{(2)}(\mathbf{Z})$ can be both bounded by the same quantity, making the two RVs asymptotically equivalent. This method of extracting large deviation probabilities is known as Statulevičious' condition [2], which allows to verify that $\hat{\varepsilon}_n^{(2)}$ is normally distributed in a region broader than the one defined by (4.12). This results for the tail probability taking the limit expression

$$\Pr\left(\hat{\varepsilon}_n^{(2)} > x\right) \sim \overline{\Phi}\left(\frac{\sqrt{n}(x-\mu_2)}{\sigma_2}\right), \quad 0 < x - \mu \ll c_n^{(2)}, \tag{4.13}$$

with

$$c_n^{(2)} = c \cdot n^{-1/3}, \quad c \in \mathbb{R}^+. \tag{4.14}$$

As a consequence, from both (4.11) and (4.13), we can state the probability distribution is exponentially suppressed in the size of the sample with speed n due to $\hat{\varepsilon}_n^{(p)}$ being normally distributed up to a distance from the average value defined by the threshold sequences given in (4.12) and (4.14).

4.2.2 *Large deviations for the tail probability*

In the previous section we discussed the extension of the Gaussian regime up to a threshold sequence $c_n^{(p)}$. As in the independent case discussed in Section 3.4, we tried to establish an LDP even in the large x region, where we suspect the behavior of the tail probability $\Pr\left(\hat{\varepsilon}_n^{(p)} > x\right)$ is qualitatively identical to the one of the independent case. As already mentioned, this was driven by the observation that, as the number of spacings n increases, the spacings $\boldsymbol{\varphi}$ (that is $\boldsymbol{\phi}$) act more and more like independent and identically distributed. This comes from the RVs being weakly dependent, as we already pointed out in Section 2.3.1. As a matter of fact, no rigorous result could be provided to support this assumption. Nevertheless, a number of numerical simulations were carried out to see if any LDP could be extracted from the directly inspecting the distribution of costs.

4.3 NUMERICAL SIMULATIONS

In this section we present the results of the simulations carried out with the direct sampling method discussed in Section 3.5.1.
Firstly, we checked that the average cost per edge $\hat{\varepsilon}_n^{(p)}$ is normally distributed for large values of n. An example of this fact is illustrated in Figure 4.1 for the $p = 2$ case, where the normalized histogram of the average cost $p_L(x)$, sampled over $L = 10^7$ instances, and the expected Gaussian distribution is given. Actually, the convergence to the normal distribution comes quite rapidly as n increases: the heavy tail behavior for $n = 50$, which is stressed out in the logarithmic plot (Figure 4.1b), gets almost entirely suppressed as $n = 500$ (Figure 4.1d), where no appreciable subexponential trend can be noticed.

It is useful to recall from Section 3.3.1 that, in case of independent RVs, the value $p = 2$ was quite exceptional. For smaller values, that is for $p \in (1, 2)$, the subexponential tail can be recovered only in the region of $x \gg n^\delta$, with $\delta > 0$, that is for extremely large deviations from the expected value. On the other hand, for $p > 2$, the stretched exponential tail dominates a region that gets closer and closer to the expected value μ_p as n increases.

Driven by this observations we could ask if a similar behavior applies to the (correlated) case of REMP. Namely, we could ask if, even for large values of n, a stretched exponential tail can be observed for $p > 2$. In Figure 4.2 we collected the distribution of the average cost in the case $p = 5$ for different values of n with the expected Gaussian distribution as a comparison. It appears clearly that the convergence to the Gaussian regime is slower than the previous case as n increases, mostly due to the variance being larger as a consequence of larger p. Nevertheless, especially in the logarithmic plot (see Figure 4.2b), it appears the distribution exhibit a subexponential tail in the region of large x, even for values of n where the Gaussian regime is dominant (Figure 4.2f). This could suggests a slower than exponential dumping in n may occur in the large x region, as in the case of independent RVs. Although, given the great inefficiency of the sampling method we used for large deviations, extracting a precise speed of scaling for the tail distribution of the average cost is unfeasible. This, as already discussed in Section 3.5.1, has to deal with both the nature of rare events and the limitations of this procedure. As a consequence, no proper bound for the tail probability far away from the expected value can be extracted from the numerical simulations.

4.4 CONCLUSION AND OUTCOMES

As already pointed out, we have proved the distribution of the average cost in the REMP is normally distributed around the typical value in the limit of large n, that is the Gaussian regime applies in a region

predicted by the CLT. We provided some motivations to the fact that this regime can be extended in a broader region than the classical one by using results coming from statistical inference and goodness-of-fit tests.

We would have liked to obtain some rigorous results regarding the behavior of the tail probability of the distribution of cost in the very large x region. This would have allowed to establish a proper large deviation principle for the whole distribution of the average cost. As a consequence, the characterization of probability of rare events and how fast values different from the typical one become unlikely is not completely resolved. This is a challenging task to accomplish, mostly because of non-trivial correlations induced by the Euclidean distance. Numerical simulations that we performed, despite their limitations, seems to suggest that the independent and the correlated case could have a similar asymptotic expression far away from the expected average cost.

Actually, this remains an open problem that needs to be further investigated.

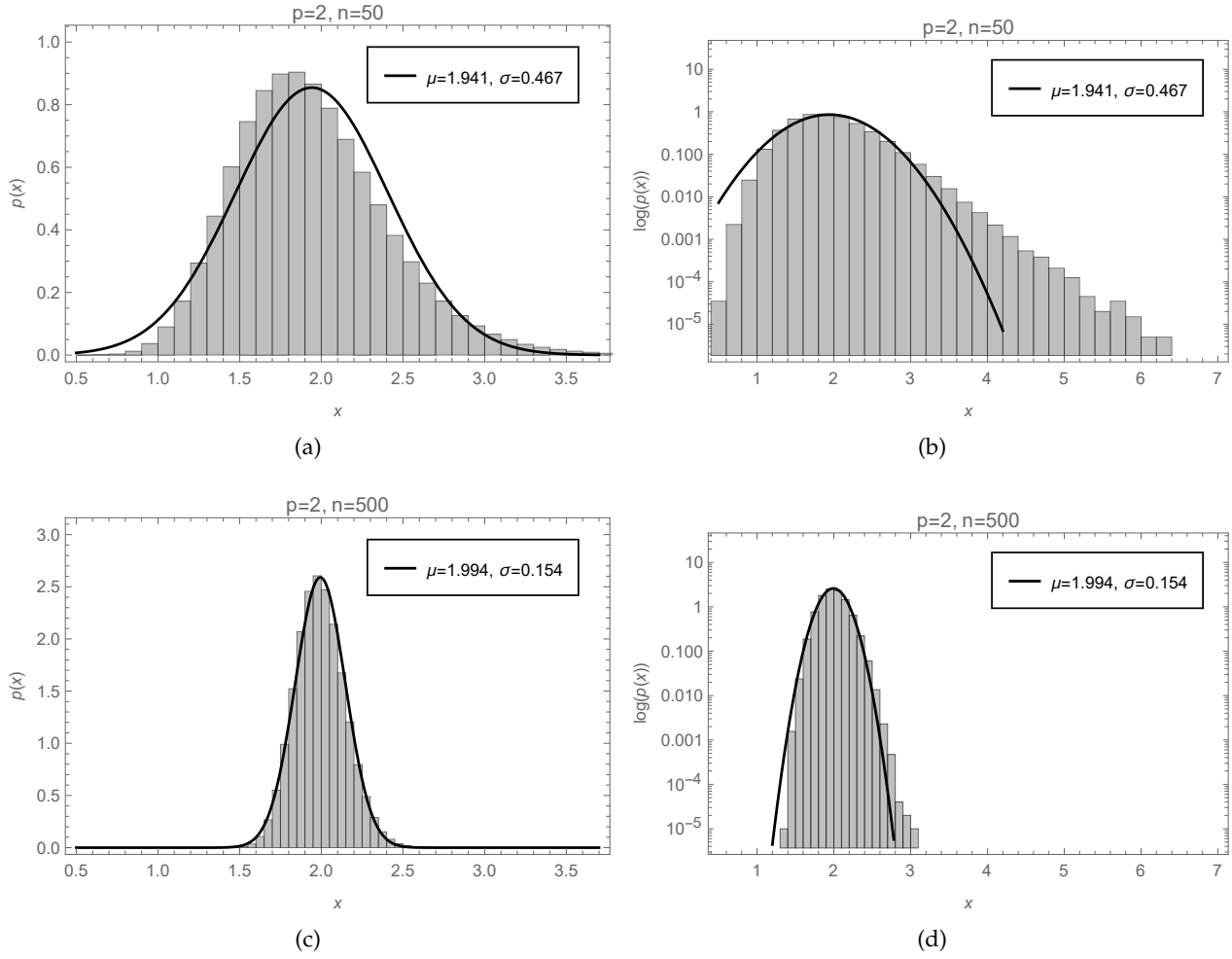


Figure 4.1: Normalized histogram of the average cost $\hat{\varepsilon}_n^{(2)}$ for $p = 2$, from the direct sampling method. The number of instances $L = 10^7$ is fixed and different values of n are collected ($n = 50$ in (a) and (b), $n = 500$ in (c) and (d)). Convergence towards the expected normal distribution (in black, with given mean and variance) is quite fast as n increases: while the probability concentrates around the typical value from (a) to (c), the subexponential tail which can be observed in the logarithmic plot (b) for $n = 50$ is almost entirely suppressed in (d) for $n = 500$.

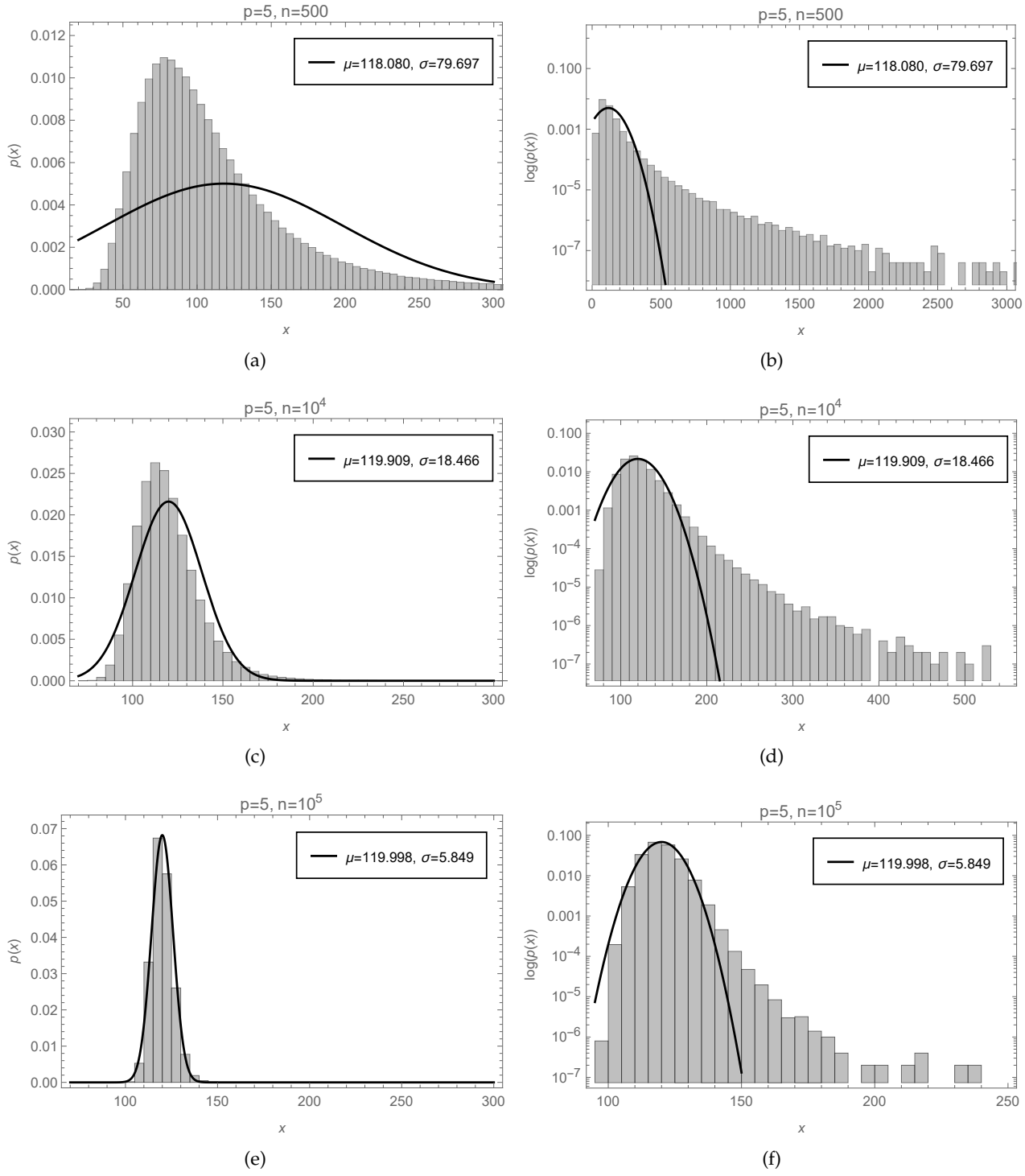


Figure 4.2: Normalized histogram of the average cost $\hat{\varepsilon}_n^{(5)}$ for $p = 5$, from the direct sampling method. The number of instances $L = 10^7$ is fixed and different values of n are collected. Convergence towards the expected normal distribution (in black, with given mean and variance) happens to be slower in n than the $p = 2$ case (see Figure 4.1), although the Gaussian regime is almost totally recovered in (e) for $n = 10^5$. While the probability concentrates around the typical value from (a) to (e), the subexponential tail in the logarithmic plot (b) for $n = 500$ persists even for larger values of n, as in (f) for $n = 10^5$. This may suggest the subexponential regime for large values of x can reflect in a subexponential dumping in n of the tail probability.

Part II

APPENDIX

A

THE LEGENDRE-FENCHEL TRANSFORM

In this appendix we want to give a proper definition and enunciate some useful properties of the Legendre-Fenchel transform, which are extensively used during the dissertation.

A.1 DEFINITION AND FIRST PROPERTIES

Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function: we define the **Legendre-Fenchel transform** (LFT) of $f(x)$ by the variational formula

$$f^*(k) \equiv \sup_{x\in\mathbb{R}} [kx - f(x)] \,. \tag{A.1}$$

In the same way, the LFT of $f^*(k)$ can be defined as

$$f^{**}(x) \equiv (f^*(k))^* = \sup_{k\in\mathbb{R}} [kx - f^*(k)] \,, \tag{A.2}$$

which corresponds also to the double LFT of $f(x)$. It is important to notice LFT is not necessarily *involutive*: that is to say, f^{**} does not need necessarily to be equal to f.

Obviously, the LFT can be defined using an infimum rather than a supremum:

$$g^*(k) = \inf_{x\in\mathbb{R}} [kx - g(x)] \,. \tag{A.3}$$

This can be shown to be equivalent to expression in (A.1) by introducing minus signs at the right place. Explicitly

$$\begin{aligned} -f^*(k) &= -\sup_{x\in\mathbb{R}} [kx - f(x)] \\ &= \inf_{x\in\mathbb{R}} [-kx + f(x)] \end{aligned}$$

so that, by setting $g(x) = -f(x)$, we have $g^*(k) = -f^*(-k)$. That is, defining the LFT by the infimum as in (A.3) simply results in the point reflection of the function f^*.

The LFT is often referred to in physics as *Legendre transform*. Actually the Legendre transform is simply defined by considering

$$f^*(k) = kx_k - f(x_k) \tag{A.4}$$

where x_k is determined as the solution of the equation

$$\frac{d}{dx} f(x) \equiv f'(x) = k \,. \tag{A.5}$$

This form is more limited in scope than the LFT since it applies only to differentiable and convex functions, as we will see later. In this sense, the LFT is a generalization of the Legendre transform, which extends to nonconvex, non differetiable functions.

The definition of the LFT given in (A.1) can trivially be generalized to functions defined on higher-dimensional spaces, namely functions $f(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$, with d a positive integer, by replacing the normal real-number product kx by the scalar product $\mathbf{k} \cdot \mathbf{x}$, where $\mathbf{k}$ and $\mathbf{x}$ are both d-dimensional vectors.

A.2 THEORY OF LEGENDRE-FENCHEL TRANSFORM

The theory of LFT deals almost entirely with two question

- how the choice of the function f affects the expression of f^*;
- when an LFT is involutive, i. e. when it holds $f^{**} = f$.

These two questions are answered by using a fundamental concept of convex analysis known as a **supporting line**. We say that the function $f : \mathbb{R} \to \mathbb{R}$ admits a supporting line at $x \in \mathbb{R}$ if there exists a real constant α such that

$$f(y) \geqslant f(x) + \alpha(y - x), \quad \forall y \in \mathbb{R}. \tag{A.6}$$

The constant α is called the *slope* of the supporting line. We further say that a supporting line is *strictly supporting* at x if

$$f(y) > f(x) + \alpha(y - x), \quad \forall y \neq x. \tag{A.7}$$

From now on we will replace the expression "f admits a supporting line in x" with the one "f is **convex** in x". If, on the other hand, f does not admit a supporting line, we will say f is **nonconvex** at x. It should be clear that, for any differentiable function f, the property of admitting a supporting line in x, i. e. to be convex at x, simply implies $f''(x)$ to be positive, while $\alpha = f'(x)$.

Ideed, it can be shown that the LFT $f^*(k)$ of a real-valued function $f(x)$ is always convex for any $k \in \mathbb{R}$. This requires the double transform $f^{**}(x)$ to be a convex function as well, without being guaranteed $f(x)$ was convex in the first place. We now understand why the equality $f^{**}(x) = f(x)$ does not always hold, namely if f is *nonconvex* in x then we are guaranteed that $f^{**}(x) \neq f(x)$.

We can now answer to the first question by using the result of the following theorem.

Theorem A.2.1 *If f admits a supporting line at x with slope k, then the LFT f^* admits a supporting line at k with slope x.*

This fact specializes to the case of f being a strictly convex function.

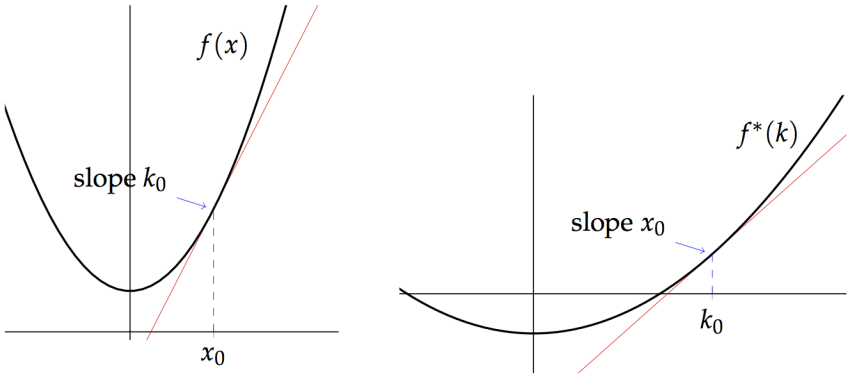


Figure A.1: Illustration of the duality property for supporting lines: points of f are mapped by the Legendre-Fenchel transform into slopes of f^*, and, vice versa, slopes of f are transformed into points of f^*.

Theorem A.2.2 *If f admits a strict supporting line at x with slope k, then f^* is differentiable at k and it admits a tangent supporting line with slope $f^{*\prime}(k) = x$.*

These facts go under the name of *supporting line duality*, i. e. taking the LFT of a convex function f simply means to write the function in terms of the slope k it assumes at the point x. An illustration of this property is given in Figure A.1.

The second answer, namely when the equality $f^{**}{=}f$ holds, is addressed by the following theorems.

Theorem A.2.3 *The function f admits a supporting line at x if and only if $f(x) = f^{**}(x)$.*

Theorem A.2.4 *If f^* is differentiable at k, then $f = f^{**}$ at $x = f^{*\prime}(k)$.*

From the last result we can state that if f^* is everywhere differentiable in its domain, then f is a convex function and it holds $f(x) = f^{**}(x)$, $\forall x \in \mathbb{R}$. This says in words that the LFT is completely involutive if $f^*(k)$ is everywhere differentiable.

All this result can become more clear if we consider again the simple case of a smooth, differentiable real function f: in this case the LFT is simply given by (A.4), where x_k is the solution, for a given k, of the equation

$$\frac{\partial}{\partial x} F(x,k) \equiv \frac{\partial}{\partial x} \left[xk - f(x)\right] = 0\,, \tag{A.8}$$

that is

$$x_k = \arg\sup_{x\in\mathbb{R}} F(x,k)\,. \tag{A.9}$$

In other words, for any fixed k, the supremum returns the maximal distance between the line $y = kx$ with slope k and the underlying

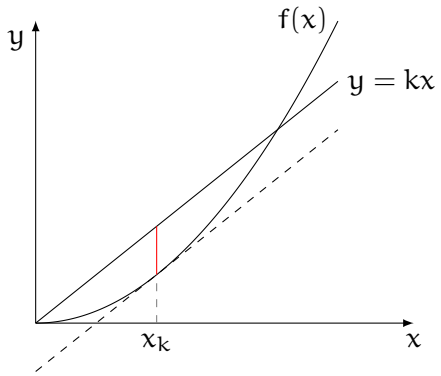


Figure A.2: Geometrical representation of the Legendre transform: maximizing the function $F(x,k) = kx - f(x)$ with respect to x is equivalent to finding the maximal distance between the line $y = kx$ and the underlying function $f(x)$. Here the maximal distance (red) having length $kx_k - f(x_k)$, where x_k is determined by the differential equation $\partial_x F(x,k) = 0$.

function $y = f(x)$. A geometrical interpretation of this fact is given in Figure A.2. Obviously there could be more than one critical points of $F(x,k)$ that would solve the differential equation in (A.8). To make sure that there is actually only one solution for every $k \in \mathbb{R}$, we need to require f' to be monotonically increasing in the variable x, while $f'(x) \to \pm\infty$ for $x \to \pm\infty$. With these two assumptions we are guaranteed that there exists a unique value x_k that satisfies

$$f'(x_k) = k\,,$$

and which maximizes $F(x,k)$. Thus we have proved the LFT reduces to a simple Legendre transform for the case of differentaible and strictly convex functions, the latter descending from monotonicity of f'. As a consequence of this fact, from Theorem A.2.3 we also have $f^{**}(x) = f(x)$, $\forall x \in \mathbb{R}$, meaning the LFT is completely involutive. Thus, from the definition we gave in (A.2) we have

$$f^{**}(x) = f(x) = xk_x - f^*(k_x)\,, \tag{A.10}$$

where, analogously to (A.9), k_x is the unique solution of the equation

$$f^{*\prime}(k) = x\,. \tag{A.11}$$

Now we could ask what would happen if we remove the assumption that $f'(x)$ is monotone, that is if f is a differetiable nonconvex function of x. In this case we have at least one open interval $I = (x_1, x_2)$ in which the function f does not admit a supporting line. This means, by recalling the results in Theorem A.2.1 and Theorem A.2.2, that the points in the interval I have no representation in terms of the LFT f^*. In other words, there is no point k of f^* which admits a supporting line with slope in the range (x_1, x_2).

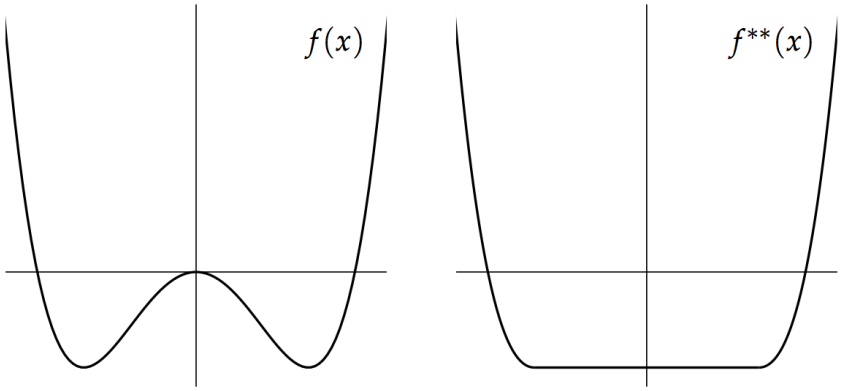


Figure A.3: On the right a nonconvex smooth function $f(x)$. The double Legendre-Fenchel transform $f^{**}(x)$ coincides with the convex envelope of $f(x)$: it is obtained by replacing the nonconvex branch of f with the supporting line connecting the two convex branches.

From the definition we gave in (A.6), it is easy to see that the points x_1 and x_2 must share the same supporting line. Namely, since f is differentiable, the supporting line of f in x_1 coincides with its tangent line

$$y = f(x_1) + f'(x_1)(x - x_1) \leqslant f(x)\,, \quad \forall x \in \mathbb{R}\,,$$

and taking $x = x_2$ it follows that $f'(x_2) = f'(x_1)$. As a consequence The LFT f^* must have a non differentiable point at k_c, with $k_c = f'(x_1) = f'(x_2)$ equal to the slope of the supporting line connecting the two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$. Thus, in a certain way, f^* must have two slopes at k_c that are x_1 and x_2.

We define the **convex hull** or **convex envelope** of f to be the function obtained by replacing the nonconvex branch of $f(x)$, i. e. $f(x)|_I$, by the supporting line connecting the two convex branches of f. Then, both the LFT of f and its convex envelope yield f^*. It should be evident that the convex hull of f is nothing but f^{**}, the double LFT of f. A graphical illustration of this property is shown in Figure A.3. This implies the LFT has the following structure:

$$f \rightharpoonup f^* \leftrightharpoons f^{**}\,. \tag{A.12}$$

This diagram clearly shows that the LFT is not involutive in general. Only in the special case of convex functions, i. e. for functions admitting supporting lines everywhere, it holds $f = f^{**}$, thus the diagram reduces to

$$f \leftrightharpoons f^*\,. \tag{A.13}$$

BIBLIOGRAPHY

[1] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. "Spin-glass models of neural networks." In: *Physical Review A* 32.2 (1985), p. 1007.

[2] NN Amosova. "Necessity of the Cramer, Linnik, and Statulevičius Conditions for Large-Deviation Probabilities." In: *Journal of Mathematical Sciences* 109.6 (2002), pp. 2031–2036.

[3] Alan A Bertossi, Paolo Carraresi, and Giorgio Gallo. "On some matching problems arising in vehicle scheduling models." In: *Networks* 17.3 (1987), pp. 271–281.

[4] Sergio Caracciolo, Matteo D'Achille, and Gabriele Sicuro. "Random Euclidean matching problems in one dimension." In: *Physical Review E* 96.4 (2017), p. 042102.

[5] Tommaso Castellani and Andrea Cavagna. "Spin-glass theory for pedestrians." In: *Journal of Statistical Mechanics: Theory and Experiment* 2005.05 (2005), P05012.

[6] Herman Chernoff et al. "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations." In: *The Annals of Mathematical Statistics* 23.4 (1952), pp. 493–507.

[7] Harald Cramér. "Sur un nouveau théoreme-limite de la théorie des probabilités." In: *Actual. Sci. Ind.* 736 (1938), pp. 5–23.

[8] Harald Cramér and Hugo Touchette. "On a new limit theorem in probability theory (Translation of'Sur un nouveau th\'eor\eme-limite de la th\'eorie des probabilit\'es')." In: *arXiv preprint arXiv:1802.05988* (2018).

[9] DA Darling. "On a class of problems related to the random division of an interval." In: *The Annals of Mathematical Statistics* (1953), pp. 239–253.

[10] Samuel Frederick Edwards and Phil W Anderson. "Theory of spin glasses." In: *Journal of Physics F: Metal Physics* 5.5 (1975), p. 965.

[11] Richard S Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.

[12] Richard S Ellis et al. "Large deviations for a general class of random vectors." In: *The Annals of Probability* 12.1 (1984), pp. 1–12.

[13] Luca Maria Gambardella, É D Taillard, and Marco Dorigo. "Ant colonies for the quadratic assignment problem." In: *Journal of the operational research society* 50.2 (1999), pp. 167–176.

[14] Jürgen Gärtner. "On large deviations from the invariant measure." In: *Theory of Probability & Its Applications* 22.1 (1977), pp. 24–39.

[15] Kaushik Ghosh and SR Jammalamadaka. "Some recent results on inference based on spacings." In: *Asymptotics in Statistics and Probability: Papers in Honor of George Gregory Roussas, VSP Utrecht* (2000), pp. 185–196.

[16] Major Greenwood. "The statistical study of infectious diseases." In: *Journal of the Royal Statistical Society* 109.2 (1946), pp. 85–110.

[17] Chris C Heyde. "On large deviation probabilities in the case of attraction to a non-normal stable law." In: *Sankhyā: The Indian Journal of Statistics, Series A* (1968), pp. 253–258.

[18] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. "Optimization by simulated annealing." In: *science* 220.4598 (1983), pp. 671–680.

[19] M. Lifshitz, S. A. Gredeskul, and L. A. Pastur. *Introduction to the theory of disordered systems*. Wiley, 1988.

[20] Carlo Lucibello, Enrico M Malatesta, Giorgio Parisi, and Gabriele Sicuro. "The random fractional matching problem." In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.5 (2018), p. 053301.

[21] Marc Mézard and Giorgio Parisi. "Replicas and optimization." In: *Journal de Physique Lettres* 46.17 (1985), pp. 771–778.

[22] Marc Mézard and Giorgio Parisi. "On the solution of the random link matching problems." In: *Journal de Physique* 48.9 (1987), pp. 1451–1459.

[23] Marc Mézard and Giorgio Parisi. "The Euclidean matching problem." In: *Journal de Physique* 49.12 (1988), pp. 2019–2025.

[24] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. Vol. 9. World Scientific Publishing Company, 1987.

[25] Thomas Mikosch and Aleksandr V Nagaev. "Large deviations of heavy-tailed sums with applications in insurance." In: *Extremes* 1.1 (1998), pp. 81–110.

[26] Sherzod M Mirakhmedov. "On Greenwood goodness-of-fit test." In: *Journal of Statistical Planning and Inference* 140.11 (2010), pp. 3017–3025.

[27] Sherzod Mira'zam Mirakhmedov et al. "The probability of large deviations for the sum functions of spacings." In: *Int. J. Math. Mathematical Sciences* 2006 (2006), pp. 58738–1.

[28] Aleksandr V Nagaev. "Integral limit theorems taking large deviations into account when Cramér's condition does not hold. I." In: *Theory of Probability & Its Applications* 14.1 (1969), pp. 51–64.

[29] Sergey V Nagaev. "Large deviations of sums of independent random variables." In: *The Annals of Probability* (1979), pp. 745–789.

[30] Henri Orland. "Mean-field theory for optimization problems." In: *Journal de Physique Lettres* 46.17 (1985), pp. 763–770.

[31] Michael Patriksson. *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.

[32] Carsten Peterson and Bo Söderberg. "A new method for mapping optimization problems onto neural networks." In: *International Journal of Neural Systems* 1.01 (1989), pp. 3–22.

[33] Ronald Pyke. "Spacings." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 27.3 (1965), pp. 395–436.

[34] Kim Sharp and Franz Matschinsky. "Translation of Ludwig Boltzmann's Paper "On the Relationship between the Second Fundamental Theorem of the Mechanical Theory of Heat and Probability Calculations Regarding the Conditions for Thermal Equilibrium" Sitzungberichte der Kaiserlichen Akademie der Wissenschaften. Mathematisch-Naturwissen Classe. Abt. II, LXXVI 1877, pp 373-435 (Wien. Ber. 1877, 76: 373-435). Reprinted in Wiss. Abhandlungen, Vol. II, reprint 42, p. 164-223, Barth, Leipzig, 1909." In: *Entropy* 17.4 (2015), pp. 1971–2009.

[35] David Sherrington and Scott Kirkpatrick. "Solvable model of a spin-glass." In: *Physical review letters* 35.26 (1975), p. 1792.

[36] Hugo Touchette. "The large deviation approach to statistical mechanics." In: *Physics Reports* 478.1-3 (2009), pp. 1–69.

[37] SR Srinivasa Varadhan. "Asymptotic probabilities and differential equations." In: *Communications on Pure and Applied Mathematics* 19.3 (1966), pp. 261–286.

[38] SR Srinivasa Varadhan. *Large deviations and applications*. Vol. 46. Siam, 1984.

[39] SR Srinivasa Varadhan. "Large deviations." In: *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*. World Scientific. 2010, pp. 622–639.

[40] Angelo Vulpiani, Fabio Cecconi, Massimo Cencini, Andrea Puglisi, and Davide Vergni. "Large deviations in physics." In: *The Legacy of the Law of Large Numbers (Berlin: Springer)* (2014).

[41] Jun Zhang, Yong Yan, and Martin Lades. "Face recognition: eigenface, elastic matching, and neural nets." In: *Proceedings of the IEEE* 85.9 (1997), pp. 1423–1435.

DECLARATION

All work and no play makes me a dull boy

Milan, April 4th, 2019

Jacopo Cattaneo

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of April 1, 2019 (classicthesis v4.6).