LARGE DEVIATIONS PRINCIPLES FOR MARKOV CHAINS AND FOR STRONGLY ADDITIVE ARITHMETIC FUNCTIONS

by

Mehmet Yenisey

Baccalaureus der Mathematik, Albert-Ludwigs-Universität Freiburg, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics
Boğaziçi University
2015

ABSTRACT

LARGE DEVIATIONS PRINCIPLES FOR MARKOV CHAINS AND FOR STRONGLY ADDITIVE ARITHMETIC FUNCTIONS

Large deviations is a branch of *probability theory* with far-reaching applications in other fields of science and can be described to the layperson as the study of "very rare" events. We provide a modest introduction to the subject with an emphasis on large deviations behavior of *continuous-time Markov chains* and we present a recent result on large deviations related to the Erdős-Kac theorem on the number of distinct prime factors of a uniformly chosen random natural number less than or equal to n.

ÖZET

MARKOV ZİNCİRLERİ VE BAZI TOPLAMSAL ARİTMETİK FONKSİYONLAR İÇİN BÜYÜK SAPMALARIN PRENSİPİ

Büyük sapmaların teorisi, matematiksel olasılık teorisinin ender olayları inceleyen bir branşıdır. Bu tezde, temel kavram ve teknikleri açıkladıktan sonra sürekli Markov zincirlerinin ve sayılar teorisinde Erdős-Kac teoremi olarak bilinen teorem ile ilgili bir sürecin, büyük sapmaların teorisi çerçevesinde incelenmelerine yer vereceğiz.

TABLE OF CONTENTS

ABSTRACT iii
ÖZET iv
LIST OF SYMBOLS vii
LIST OF ACRONYMS/ABBREVIATIONS viii
1. INTRODUCTION 1
2. PRELIMINARIES 4
   2.1. Stochastic Processes 4
   2.2. Large Deviations Principle (LDP) 5
   2.3. Two Propositions 8
   2.4. Empirical, Pair Empirical, and Occupation Time Measures 9
   2.5. Markov Chains 12
3. BASIC CONCEPTS AND TECHNIQUES 14
   3.1. Weak LDP and Exponential Tightness 14
   3.2. Exponential Approximations 16
   3.3. The Contraction Principle 17
   3.4. Varadhan's Lemma and Tilting an LDP 18
   3.5. The Gärtner-Ellis Theorem 23
4. LARGE DEVIATIONS FOR SEQUENCES OF I.I.D. RANDOM VARIABLES 29
   4.1. Sanov's Theorem and LDP for the Empirical Measures of Sequences of I.I.D. Random Variables 29
   4.2. LDP for the Pair Empirical Measures of Sequences of I.I.D. Random Variables 32
5. LARGE DEVIATIONS FOR MARKOV CHAINS 35
   5.1. LDPs for the Pair Empirical Measures, the Empirical Measures and the Empirical Means of DTMCs 35
   5.2. LDP for the Occupation Time Measures of CTMCs 39
   5.3. Another Treatment of the LDP for the Occupation Time Measures of CTMCs 44
6. LARGE DEVIATIONS RELATED TO THE ERDŐS-KAC THEOREM . . 52

6.1. The Erdős-Kac Theorem 52
6.2. Large Deviations for Strongly Additive Arithmetic Functions 53
REFERENCES 65

LIST OF SYMBOLS

$\mathcal{A}^{\otimes I}$	The σ-algebra product of a σ-algebra $\mathcal{A}$ indexed by I
$\mathcal{B}(\mathcal{X})$	The Borel σ-algebra on a topological space $\mathcal{X}$
f^*	The Legendre transform of a function $f : \mathbb{R} \to [-\infty, \infty]$
$\mathfrak{M}(\Gamma)$	The set of all probability measures on a finite set Γ
$\mathbb{N}$	The set of all natural numbers excluding the zero
$\mathbb{N}_0$	The set of all natural numbers including the zero
$\mathbb{P}^X$	The law of a stochastic process $(X_t)_{t \in I}$
$\mathbf{P}$	The set of all prime numbers
$\mathcal{P}(M)$	The power set of a set M
$\mathbb{R}$	The set of all real numbers
$\mathbb{R}_{>0}$	The set of all real numbers r with $r > 0$
$\mathbb{R}_{\geq 0}$	The set of all real numbers r with $r \geq 0$
O, o	Bachmann-Landau symbols

LIST OF ACRONYMS/ABBREVIATIONS

CTMC	Continuous-time Markov chain
DTMC	Discrete-time Markov chain
i.i.d.	Independent and identically distributed
LDP	Large deviations principle
WLDP	Weak large deviations principle

1. INTRODUCTION

Firstly, a mathematical meaning of rare events should involve the idea of some probability measures at hand assuming small values, and secondly, the natural environment for a study of these events should be in the context of the theory of *stochastic processes*, where numerous interesting limit theorems already take place, with *stochastic processes* being sequences or families of *random variables* modeling a random process evolving over time. *Large deviations* is roughly speaking a study of "logarithms of probabilities" extending on a theorem of H. Cramér from 1938, see [1], which deals with an asymptotic estimation of a certain sequence of probabilities in the i.i.d. setting, referred to as "large deviations". (We shall present this theorem in the next chapter.) A new scaling favoring the "rare events", achieved by considering the logarithms of the probabilities yields way to an abstract theory with analytic, probabilistic and topological components. The results complement the various limit theorems, and they have important applications in physics, biology, computer science, statistics, operations research, and finance among other fields.

The theory is centered around the definition of a *large deviations principle* (LDP) for a family of probability measures on a topological space. The definition is due S. R. S. Varadhan, see [2], and is actually tailored to suit the need of computing certain exponential expectations with respect to these measures over "time". A correct estimation of these type of integrals play an important role in applications of probability theory. Under *large deviations theory* we understand the collection of definitions, theorems, techniques particularly relating to this definition, in its abstract setting.

Namely, a family $(\mu_t)_{t>0}$ of probability measures on $\mathcal{B}$, with (E, d) a metric space and $\mathcal{B}$ the Borel σ-algebra on E, is said to satisfy a *large deviations principle* with rate $I : E \to [0, \infty]$, if

$$\sup_{x \in A^o}\{-I(x)\} \leq \liminf_{t\to\infty} \frac{1}{t} \log \mu_t(A^o) \leq \limsup_{t\to\infty} \frac{1}{t} \log \mu_t(\bar{A}) \leq \sup_{x\in\bar{A}}\{-I(x)\},$$

for all $A \in \mathcal{B}$, and if some further technical condition on the function I holds, which is useful for the implications. We also sometimes say that a family of random variables satisfy an LDP, if their distributions satisfy the definition.

After a chapter on preliminaries, we present the basics of the theory in Chapter 3, where we exhibit the techniques for showing LDPs. Here we point out to the *Gärtner-Ellis theorem* as a nice tool for this purpose. The rate function turns out to be a *Legendre transform* in this case. This chapter is also where we present *Varadhan's lemma*, which enables the estimation of the exponential expectations we have mentioned above.

Chapter 4 deals with the LDP for the *empirical* and the *pair empirical measures* of i.i.d. random variables. These are *random measures* constructed by considering the relative frequencies of states, respectively, the relative frequencies of pairs of states occurring in succession, in given realizations of the process up to a certain time. The LDPs are followed by using a combinatorial argumentation for the estimations without using strong tools, which are not needed in this case, and thus some insight can be gained as to how the corresponding rate functions actually arise.

Then the results of Chapter 4 are used to deduce the LDP for the pair empirical measures for finite state Markov chains in discrete-time, by introducing a *tilt* argumentation. The LDP for the empirical measures follows via *contraction principle.* Afterwards an approximation argumentation is applied to extend the results to the continuous-time setting. We conclude the chapter by showing an alternative approach which uses the Gärtner-Ellis theorem together with the *Perron-Frobenius theory* of positive matrices directly in the continuous-time setting.

In the last chapter it is shown that a large deviations principle is also satisfied in the context of the Erdős-Kac theorem. The Erdős-Kac theorem is in a sense a *central limit theorem*, for the number of distinct prime factors of a uniformly chosen random natural number from $\{1, 2, ..., n\}$, saying that, loosely speaking, the distributions after subtraction of the "means" and division by the "variances" follow the standard normal distribution. The results are presented in a general form for arbitrary *strongly additive arithmetic functions* in the place of the *prime factor count function*. Here again the Gärtner-Ellis theorem is employed together with a series of *exponential approximations* reducing the problem to a simpler model. The results here are due B. Mehrdad and L. Zhu [3].

2. PRELIMINARIES

2.1. Stochastic Processes

Here are some basic definitions that we shall use.

Definition 2.1. (Stochastic process). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ a measurable space, and let I serve as an indexing set. A* stochastic process *with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(E, \mathcal{E})$ is a family $(X_t)_{t\in I}$ of $(\mathcal{F}, \mathcal{E})$-measurable functions*

$$X_t : \Omega \to E,\ \ t \in I;$$

and for fixed $\omega_0 \in \Omega$, the function

$$X_\bullet(\omega_0) : I \to E,\ \ s \mapsto X_s(\omega_0),$$

is called a realization *of the process.*

Definition 2.2. (The law of a stochastic process). *Let $(X_t)_{t\in I}$ be a discrete-time stochastic process with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(E, \mathcal{E})$. Let further $\mathcal{E}^{\otimes I}$ denote the the smallest σ-algebra on E^I, which contains all* finite rectangles*, i.e., all subsets of the form*

$$\{f \in E^I; f(i_1) \in E_1, ..., f(i_k) \in E_k\},\ \ k \in \mathbb{N},\ \ i_1, ..., i_k \in I,\ \ E_1, ..., E_k \in \mathcal{E}.$$

The function

$$X : \Omega \to E^I,\ \ \omega \mapsto X_\bullet(\omega),$$

is $(\mathcal{F}, \mathcal{E}^{\otimes I})$-measurable, *and*

$$\mathbb{P}^X : \mathcal{E}^{\otimes I} \to \mathbb{R},\ B \mapsto \mathbb{P}^X(B) := \mathbb{P}(X^{-1}(B)),$$

is a probability measure on $(E^I, \mathcal{E}^{\otimes I})$*. It is called the* law of the process $(X_t)_{t\in I}$.

2.2. Large Deviations Principle (LDP)

The definition of an LDP is set on a *Polish space*, see [4], for the sake of generality.

Definition 2.3. (Polish space)*. A topological space* $\mathcal{X}$ *is said to be* Polish*, if there exists a complete metric defining its topology and if this topology has a countable basis.*

Definition 2.4. (Lower semi-continuity)*. Let* $\mathcal{X}$ *be a Polish space with the metric* $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$*, and* $f : \mathcal{X} \to [-\infty, \infty]$ *a function.* f *is called* lower semi-continuous*, if it satisfies the following equivalent conditions:*

(i) For all $x \in \mathcal{X}$ *and all sequences* $(x_n)_{n\in\mathbb{N}}$ *in* $\mathcal{X}$ *with* $\lim_{n\to\infty} x_n = x$*, the inequality*

$$\liminf_{n\to\infty} f(x_n) \geq f(x)$$

holds.

(ii) For all $x \in \mathcal{X}$,

$$\lim_{\epsilon\to 0^+} \inf_{y\in B_\epsilon(x)} \{f(y)\} = f(x),$$

where $B_\epsilon(x) := \{y \in \mathcal{X}; d(x, y) < \epsilon\}$.

(iii) f *has closed level sets, i.e.,*

$$f^{-1}([-\infty, c]) = \{x \in \mathcal{X}; f(x) \leq c\}$$

is closed for all $c \in \mathbb{R}$.

Definition 2.5. (Rate function). *Let $\mathcal{X}$ be a Polish space, and $f : \mathcal{X} \to [0,\infty]$ a function. f is called a* rate function, *if the following conditions hold:*

(i) $f \not\equiv \infty$.

(ii) *f has compact level sets, i.e., $f^{-1}((-\infty, c])$ is compact for all $c \in [0,\infty)$.*

Remark 2.1. Notice that a rate function on a Polish space is thus semi-continuous. Also notice that a lower semi-continuous function $f : \mathcal{X} \to [-\infty,\infty]$ on a Polish space $\mathcal{X}$ attains a minimum on every non-empty compact subset $K \subset \mathcal{X}$.

We now present the definition of a large deviations principle. As we have mentioned in the Introduction, this definition is particularly set with the estimation of exponential integrals in mind, and has established itself as a central definition in the field of large deviations. For similar principles and approaches to the subject see [5].

Definition 2.6. (Large deviations principle). *Let $\mathcal{X}$ be a Polish space, $\mathcal{B}(\mathcal{X})$ the Borel σ-algebra on $\mathcal{X}$, $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\rho_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$, and $I : \mathcal{X} \to [0,\infty]$ a function. The sequence $(P_n)_{n\in\mathbb{N}}$ is said to satisfy an* LDP *with speed ρ_n and rate I, if:*

(i) I is a rate function.

(ii) For all open sets $G \subset \mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{\rho_n} \log P_n(G) \geq -I(G);$$

(iii) For all closed sets $F \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(F) \leq -I(F),$$

where $I(S) := \inf\{I(x); x \in S\}$ for $S \subset \mathcal{X}$, with the convention $\inf \emptyset = \infty$. (ii) and (iii) are then referred to as the LDP lower bounds, *and the* LDP upper bounds *respectively.*

Remark 2.2. Clearly, we may extend this definition to define the concept of an LDP for a family of measures indexed by real numbers in the obvious way by altering the parameters of the limits appearing above. Also we point out here that the theorems we shall prove in the next chapter for families of measures satisfying LDPs are usually stated for the ones satisfying the LDPs with speed $\rho_n = n$ and with the discrete definition, for the sake of brevity in the statements and in the argumentation. However it is a simple task to check whether the theorems hold in general with arbitrary speeds ρ_n, ρ_t with $\lim_{n\to\infty} \rho_n = \infty$, $\lim_{t\to\infty} \rho_t = \infty$ in the corresponding settings.

At this point we state Cramér's theorem mentioned in the Introduction. Its resemblance with the definition of an LDP is apparent.

Theorem 2.7. (Cramér's theorem). *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of $\mathbb{R}$-valued i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\phi(t) := \mathbb{E}[\mathrm{e}^{tX_1}] < \infty$ for all $t \in \mathbb{R}$. Let further $\mathbb{P}^X : \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}} \to [0,1]$ be the law of $(X_i)_{i\in\mathbb{N}}$, and $S_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Then, for all $a > \mathbb{E}[X_1]$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^X\left(\frac{S_n}{n} \geq a\right) = -I(a), \tag{2.1}$$

where $I(a) := \sup_{t\in\mathbb{R}}\{t \cdot a - \log\phi(t)\}$ for $a \in \mathbb{R}, a > \mathbb{E}[X_1]$.

Proof. We defer a proof until Section 3.5. For now we just point out that the Equation 2.1 is equivalent to

$$\mathbb{P}\left(\frac{S_n}{n} \geq a\right) = \mathrm{e}^{-nI(a)+o(n)},$$

for $I(a) < \infty$. This is clearly a statement about the decay rate of the probabilities $\mathbb{P}(\frac{S_n}{n} \geq a)$, which should converge to 0 by the *weak law of large numbers*, since $a > \mathbb{E}[X_1]$. These probabilities are sometimes referred to as *large deviations away from the mean*, and apparently the corresponding events are, loosely speaking, rare events. □

The function $I : \mathbb{R} \to [-\infty, \infty]$, $x \mapsto I(x) := \sup_{t\in\mathbb{R}}\{tx - \log\phi(t)\}$, appearing in the above theorem is called the *Legendre transform* of the function $\log\phi$ there:

Definition 2.8. (The Legendre transform, a.k.a. the convex conjugate). *Let $f:\mathbb{R}\to[-\infty,\infty]$ be a function. Then, the function $f^*:\mathbb{R}\to[-\infty,\infty]$ defined by*

$$f^*(x) := \sup_{t\in\mathbb{R}}\{tx - f(t)\},\ \textit{for } x\in\mathbb{R},$$

is called the Legendre transform *of the function f.*

2.3. Two Propositions

The following proposition is well-known and needs no introduction.

Proposition 2.9. *We have, as $n\to\infty$:*

(i) $n! = n^n e^{-n}\sqrt{2\pi n}(1+O(\frac{1}{n}))$. *"Stirling's Formula"*
(ii) $\log n! = n\log n - n + \frac{1}{2}\log n + \log\sqrt{2\pi} + O(\frac{1}{n})$.

Proof. For (i) see e.g., [6]. (ii) follows from (i) with ease using

$$\log(1+O(\frac{1}{n})) = O(\frac{1}{n}),$$

where the $O(\frac{1}{n})$ in the argument of the logarithm function on the left-hand side denotes an unspecified function f such that $f = O(\frac{1}{n})$. □

The following proposition sheds some light on the definition of a large deviations principle.

Proposition 2.10. (Exponential rate of growth of sums). *Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be two sequences in $\mathbb{R}_{>0}$, and $(\rho_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty}\rho_n = \infty$.*

Then,

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log(\alpha_n + \beta_n) = \limsup_{n\to\infty} \frac{1}{\rho_n} \log(\max\{\alpha_n, \beta_n\})$$
$$= \max\{\limsup_{n\to\infty} \frac{1}{\rho_n} \log \alpha_n, \limsup_{n\to\infty} \frac{1}{n} \log \beta_n\}.$$

Proof. Let's call these quantities A, B and C respectively, the claim being thus $A = B = C$. We have

$$A \leq \limsup_{n\to\infty} \frac{1}{\rho_n} \log\left(2\max\{\alpha_n, \beta_n\}\right) = \limsup_{n\to\infty} \frac{1}{\rho_n} \left(\log 2 + \log\left(\max\{\alpha_n, \beta_n\}\right)\right) = B.$$

Also $B \leq A$ is clear, and hence $A = B$. Next, observe that $C \leq B$ is clear too. We show $C \leq B$ to complete the proof. For this we use the fact that B must be the greatest *limit point* of the sequence

$$\left(\frac{1}{\rho_n} \log\left(\max\{\alpha_n, \beta_n\}\right)\right)_{n\in\mathbb{N}}.$$

Since the "maximum" is either the α_n for infinitely many of the n's or it is the β_n for infinitely many of the n's, we deduce that B is also a limit point of at least one of the sequences $(\frac{1}{\rho_n} \log \alpha_n)_{n\in\mathbb{N}}$, $(\frac{1}{\rho_n} \log \beta_n)_{n\in\mathbb{N}}$. And so, indeed $B \leq C$ and the result is established. $\square$

2.4. Empirical, Pair Empirical, and Occupation Time Measures

We need rigorous definitions for the measures, for which we shall present large deviations principles. We establish a part of this task here.

Definition 2.11. ($\mathfrak{M}(\Gamma)$). *Let $r \in \mathbb{N}$ and Γ be a set with $|\Gamma| = r$.*

$$\mathfrak{M}(\Gamma) := \{\nu = (\nu_1, ..., \nu_r) \in [0,1]^r; \sum_{s=1}^{r} \nu_s = 1\} \subset [0,1]^r.$$

$\mathfrak{M}(\Gamma)$ can be interpreted as the set of all probability measures on $(\Gamma, \mathcal{P}(\Gamma))$. It is a

Polish space together with the topology endowed to it by the total variation distance

$$d(\mu,\nu)=\frac{1}{2}\sum_{s=1}^{r}|\mu_s-\nu_s| \text{ for } \mu,\nu\in\mathfrak{M}(\Gamma).$$

We shall mean this topological space when we speak of $\mathfrak{M}(\Gamma)$*.*

Definition 2.12. ($\widetilde{\mathfrak{M}}(\Gamma\times\Gamma)$)*. Let $r\in\mathbb{N}$ and Γ be a set with $|\Gamma|=r$. Consider the set $\mathfrak{M}(\Gamma\times\Gamma)$ using the previous definition.*

$$\widetilde{\mathfrak{M}}(\Gamma\times\Gamma):=\{\nu=(\nu_{st})_{(s,t)\in\Gamma\times\Gamma}\in\mathfrak{M}(\Gamma\times\Gamma);\forall s\in\Gamma:\sum_{t\in\Gamma}\nu_{st}=\sum_{t\in\Gamma}\nu_{ts}\}\subset[0,1]^{r\times r}.$$

$\widetilde{\mathfrak{M}}(\Gamma\times\Gamma)$ *has an interpretation as being the set of all probability measures on the space $(\Gamma\times\Gamma,\mathcal{P}(\Gamma\times\Gamma))$ each of whose marginals coincide. It is a Polish space together with the topology endowed to it by the total variation distance*

$$d(\mu,\nu)=\frac{1}{2}\sum_{(s,t)\in\Gamma\times\Gamma}|\mu_{st}-\nu_{st}| \text{ for } \mu,\nu\in\widetilde{\mathfrak{M}}(\Gamma\times\Gamma).$$

We shall mean this topological space when we speak of $\widetilde{\mathfrak{M}}(\Gamma\times\Gamma)$*.*

Definition 2.13. (The functions $L_n:\Gamma^{\mathbb{N}}\to\mathfrak{M}(\Gamma)$)*. Let $r\in\mathbb{N}$ and $\Gamma=\{\gamma_1,...,\gamma_r\}$ be a set with $|\Gamma|=r$. For $n\in\mathbb{N}$, we define*

$$L_n:\Gamma^{\mathbb{N}}\to\mathfrak{M}(\Gamma),\ x=(x_k)_{k\in\mathbb{N}}\mapsto L_n(x):=(l_{n,x}(\gamma_1),...,l_{n,x}(\gamma_r)),$$

where for $n\in\mathbb{N},\ x=(x_k)_{k\in\mathbb{N}}\in\Gamma^{\mathbb{N}}$ the functions $l_{n,x}$ are defined as

$$l_{n,x}:\Gamma\to[0,1],\ \gamma\mapsto l_{n,x}(\gamma):=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{x_i\}}(\gamma).$$

L_n is $\big(\mathcal{P}(\Gamma)^{\otimes\mathbb{N}},\mathcal{B}(\mathfrak{M}(\Gamma))\big)$-measurable for all $n\in\mathbb{N}$.

Remark 2.3. In the above definition $l_{n,x}(\gamma)\in[0,1]$ is the *relative frequency* of the value $\gamma\in\Gamma$ among the first n entries of $x=(x_k)_{k\in\mathbb{N}}\in\Gamma^{\mathbb{N}}$. And, the measurability bit is also clear, since L_n relates only to the first n entries of $x=(x_k)_{k\in\mathbb{N}}$.

Definition 2.14. (Empirical measures of a stochastic process). *Let* $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, $L_n : \Gamma^{\mathbb{N}} \to \mathfrak{M}(\Gamma)$ *for* $n \in \mathbb{N}$ *the* $\big(\mathcal{P}(\Gamma)^{\otimes\mathbb{N}}, \mathcal{B}(\mathfrak{M}(\Gamma))\big)$*-measurable functions defined in Definition 2.13, and let* $(X_n)_{n\in\mathbb{N}}$ *be a stochastic process with probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, state space* $(\Gamma, \mathcal{P}(\Gamma))$ *and law* $\mathbb{P}^X : \mathcal{P}(\Gamma)^{\otimes\mathbb{N}} \to [0, 1]$*. This constellation gives rise to a sequence of probability measures on* $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$ *given by* $B \mapsto \mathbb{P}^X(L_n^{-1}(B))$*, called the* distributions of the empirical measures *of the process, and denoted by* $\mathbb{P}^X(L_n \in \bullet)$*.*

Remark 2.4. Notice here that we have not defined empirical measures, but their distributions. Empirical measures are the *random measures*, i.e., random variables $(\Omega, \mathcal{F}) \to (\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$, defined in a similar manner, whose distributions are the objects we have defined. We do not stick to the perhaps more deliberate construction of the above definition throughout the work.

Definition 2.15. (The functions $L_n^{(2)}$, defined with periodic boundaries). *Let* $r \in \mathbb{N}$ *and* Γ *be a set with* $|\Gamma| = r$*. For* $n \in \mathbb{N}$*, we define*

$$L_n^{(2)} : \Gamma^{\mathbb{N}} \to \widetilde{\mathfrak{M}}(\Gamma \times \Gamma),\ x = (x_k)_{k\in\mathbb{N}} \mapsto L_n^{(2)}(x) := \big(l_{n,x}^{(2)}(\alpha, \beta)\big)_{(\alpha,\beta)\in\Gamma\times\Gamma},$$

where for $n \in \mathbb{N},\ x = (x_k)_{k\in\mathbb{N}} \in \Gamma^{\mathbb{N}}$,

$$l_{n,x}^{(2)} : \Gamma \times \Gamma \to [0, 1],\ (\alpha, \beta) \mapsto\ l_{n,x}^{(2)}(\alpha, \beta) := \frac{1}{n}\sum_{i=1}^{n-1} \mathbf{1}_{\{(x_i,x_{i+1})\}}(\alpha, \beta) + \frac{1}{n}\mathbf{1}_{\{(x_n,x_1)\}}(\alpha, \beta).$$

$L_n^{(2)}$ *is* $\Big(\mathcal{P}(\Gamma)^{\otimes\mathbb{N}}, \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma))\Big)$*-measurable for all* $n \in \mathbb{N}$*.*

Remark 2.5. In the above definition $l_{n,x}^{(2)}(\alpha, \beta) \in [0, 1]$ is the relative frequency of the pair $(\alpha, \beta) \in \Gamma \times \Gamma$ among the n pairs $(x_1, x_2), (x_2, x_3), ..., (x_{n-1}, x_n), (x_n, x_1)$, with $x = (x_k)_{k\in\mathbb{N}} \in \Gamma^{\mathbb{N}}$. And, the measurability bit is also clear, since $L_n^{(2)}$ relates only to the first n entries of $x = (x_k)_{k\in\mathbb{N}}$.

Definition 2.16. (Pair empirical measures with periodic boundary conditions). *Let* $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, $L_n^{(2)} : \Gamma^{\mathbb{N}} \to \widetilde{\mathfrak{M}}(\Gamma \times \Gamma)$ *for* $n \in \mathbb{N}$ *the* $\Big(\mathcal{P}(\Gamma)^{\otimes\mathbb{N}}, \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma))\Big)$*-measurable functions defined in Definition 2.15, and let* $(X_n)_{n\in\mathbb{N}}$ *be a stochastic process with probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, state space* $(\Gamma, \mathcal{P}(\Gamma))$ *and law* $\mathbb{P}^X : \mathcal{P}(\Gamma)^{\otimes\mathbb{N}} \to [0, 1]$*.*

These give rise to a sequence of probability measures on $\left(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma), \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma))\right)$ *given by* $B \mapsto \mathbb{P}^X((L_n^{(2)})^{-1}(B))$, *called the* distributions of the pair empirical measures *of the process, and denoted by* $\mathbb{P}^X(L_n^{(2)} \in \bullet)$.

Remark 2.6. Notice also here that we have not actually defined the pair empirical measures, but their distributions. Pair empirical measures are the random measures $(\Omega, \mathcal{F}) \to \left(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma), \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma))\right)$ defined in a similar manner, whose distributions are the objects we have defined. We do not stick to the construction of the above definition throughout the work, and use further constructions for the measures we shall be dealing with when we believe they are more suitable. These will be rigorously defined at the appropriate places. The same remarks hold for the next two definitions.

Definition 2.17. (The functions $L_t : \Gamma^{\mathbb{R}_{\geq 0}} \to \mathfrak{M}(\Gamma)$). *Let* $r \in \mathbb{N}$, $\Gamma = \{\gamma_1, ..., \gamma_r\}$ *be a set with* $|\Gamma| = r$, *and let* $\mathcal{L}(\Gamma^{\mathbb{R}_{\geq 0}})$ *denote the set of Lebesgue measurable functions* $\mathbb{R}_{\geq 0} \to \Gamma$. *For* $t \in \mathbb{R}_{\geq 0}$, *we define*

$$L_t : \mathcal{L}(\Gamma^{\mathbb{R}_{\geq 0}}) \to \mathfrak{M}(\Gamma),\ x = (x_s)_{s \in \mathbb{R}_{\geq 0}} \mapsto L_t(x) := (l_{t,x}(\gamma_1), ..., l_{t,x}(\gamma_r)),$$

where for fixed $t \in \mathbb{R}_{\geq 0}$, $x = (x_s)_{s \in \mathbb{R}_{\geq 0}} \in \Gamma^{\mathbb{R}_{\geq 0}}$, *and* $\gamma \in \Gamma$, $l_{t,x}(\gamma)$ *is defined by*

$$l_{t,x}(\gamma) := \frac{1}{t} \int_0^t \mathbf{1}_{\{x_s\}}(\gamma) \mathrm{d}s.$$

Definition 2.18. (Occupation time measures associated to a stochastic process). *Let* $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, $L_t : \mathcal{L}(\Gamma^{\mathbb{R}_{\geq 0}}) \to \mathfrak{M}(\Gamma)$ *for* $t \in \mathbb{R}_{\geq 0}$ *the functions defined in Definition 2.17, and let* $(X_t)_{t \in \mathbb{R}_{\geq 0}}$ *be a stochastic process with probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *state space* $(\Gamma, \mathcal{P}(\Gamma))$, *and law* $\mathbb{P}^X$. *These give rise to a family of probability measures on* $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$ *given by* $B \mapsto \mathbb{P}^X(L_t^{-1}(B))$, *denoted by* $\mathbb{P}^X(L_t \in \bullet)$, *and called the* distributions of the occupation time measures *of the process.*

2.5. Markov Chains

Markov chains are in a sense the next level of complexity above the i.i.d. setting. These are stochastic processes considered by A. Markov for the purpose of exhibiting limit theorems replacing the i.i.d. assumption with a moderate dependence assumption, now called the *Markovian property*, which basically demands that the path of the process can be disregarded for the prediction of its future behavior, see [7]. (In this view it is interesting to compare the large deviations results for the i.i.d and the Markovian settings, i.e., the Theorems 4.2 and 5.2 below.) We firstly define *stochastic matrices* and also using the opportunity we define the *irreducibility* of a matrix, then we follow with the definitions of a Markov chain in the discrete- and the continuous-time settings.

Definition 2.19. (Stochastic matrix). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$ and $P = (p_{ij})_{i,j \in \Gamma} \in \mathbb{R}^{r \times r}$ be an $(r \times r)$-matrix with coefficients in $[0, 1]$. P is called a* stochastic matrix*, if $\sum_{j \in \Gamma} p_{ij} = 1$ for all $i \in \Gamma$.*

Definition 2.20. (Irreducible matrix). *Let $n \in \mathbb{N}$ and $A \in \mathbb{R}^{n \times n}$ be an $(n \times n)$-matrix with coefficients in $\mathbb{R}$. A is called* irreducible*, if for every $i, j \in \{1, ..., n\}$ there exists an $m \in \mathbb{N}$ such that $(A^m)_{ij} \neq 0$, where $(A^m)_{ij}$ denotes the coefficient of the matrix A^m corresponding to the index pair (i, j).*

Definition 2.21. (Finite state, discrete-time Markov chain (DTMC)). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $(X_n)_{n \in \mathbb{N}}$ be a stochastic process with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(\Gamma, \mathcal{P}(\Gamma))$. $(X_n)_{n \in \mathbb{N}}$ is called a* DTMC*, if there exists a stochastic matrix $P = (p_{ij})_{i,j \in \Gamma}$, such that for all $n \in \mathbb{N}$ and all $i_1, i_2, ..., i_{n+1} \in \Gamma$ with $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, ..., X_1 = i_1) > 0$,*

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, ..., X_1 = i_1) = p_{i_n, i_{n+1}}$$

holds. The matrix P is called the transition matrix *of the chain.*

Definition 2.22. (Finite state, continuous-time Markov chain (CTMC)). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ be a stochastic process with probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and state space $(\Gamma, \mathcal{P}(\Gamma))$. $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is called a* CTMC*, if for all $n \in \mathbb{N}$, all $t_j, j = 1, ..., n+1$ from $\mathbb{R}_{\geq 0}$ with $0 \leq t_1 \leq t_2 \leq \cdots \leq t_{n+1}$, and all $i_1, i_2, ..., i_{n+1} \in \Gamma$,*

$$\mathbb{P}(X(t_{n+1}) = i_{n+1} | X(t_n) = i_n, ..., X(t_1) = i_1) = \mathbb{P}(X(t_{n+1}) = i_{n+1} | X(t_n) = i_n)$$

holds. This equation is referred to as the Markovian property*.*

3. BASIC CONCEPTS AND TECHNIQUES

3.1. Weak LDP and Exponential Tightness

We introduce the somewhat weaker auxiliary concept of a *weak large deviations principle* (WLDP) first, via which an LDP can be demonstrated under some condition on the measures in question called *exponential tightness*. Namely, we say that a WLDP holds, if the upper bounds in the definition of an LDP is satisfied for all compact subsets instead:

Definition 3.1. (WLDP). *Let $\mathcal{X}$ be a Polish space, $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\rho_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$, and $I : \mathcal{X} \to [0, \infty]$ a function. The sequence $(P_n)_{n\in\mathbb{N}}$ is said to satisfy a* WLDP, *with speed ρ_n and with rate I, if:*

(i) I is a rate function.

(ii) For all open sets $G \subset \mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{\rho_n} \log P_n(G) \geq -I(G);$$

(iii) For all compact sets $K \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(K) \leq -I(K),$$

where $I(S) := \inf\{I(x); x \in S\}$ for $S \subset \mathcal{X}$, with the convention $\inf \emptyset = \infty$.

Definition 3.2. (Exponential tightness). *Let $\mathcal{X}$ be a topological space, $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and $(\rho_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$. $(P_n)_{n\in\mathbb{N}}$ is said to be* exponentially tight *in the scale of $(\rho_n)_{n\in\mathbb{N}}$, if for every $C \in \mathbb{R}_{>0}$ there exists a subset $G \subset \mathcal{X}$, whose complement is compact, such* *that*

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(G) \leq -C$$

holds true.

Here is now the technique we have mentioned above.

Theorem 3.3. (LDP via WLDP). *Let $\mathcal{X}$ be a Polish space, $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and $(\rho_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$. If $(P_n)_{n\in\mathbb{N}}$ satisfies a WLDP with speed ρ_n and with a rate function I, and if $(P_n)_{n\in\mathbb{N}}$ is in addition exponentially tight in the scale of $(\rho_n)_{n\in\mathbb{N}}$, then it satisfies the LDP with speed ρ_n and with the same rate I.*

Proof. The issue is showing the upper bounds for all closed subsets of $\mathcal{X}$. To this end, let $K \subset \mathcal{X}$ be a closed set, and assume that we consider a compact set $M \subset \mathcal{X}$. Then $K \cap M$ is compact and we get, using the WLDP upper bounds for the compact sets and the Proposition 2.10,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(K) &= \limsup_{n\to\infty} \frac{1}{\rho_n} \log \left(P_n(K\cap M) + P_n(K \cap M^C)\right) \\
&= \max\left\{\limsup_{n\to\infty} \frac{1}{\rho_n} P_n(K\cap M), \limsup_{n\to\infty} \frac{1}{\rho_n} P_n(K\cap M^C)\right\} \\
&\leq \max\left\{-I(K\cap M), \limsup_{n\to\infty} \frac{1}{\rho_n} P_n(K\cap M^C)\right\} \\
&\leq \max\left\{-I(K), \limsup_{n\to\infty} \frac{1}{\rho_n} P_n(M^C)\right\}.
\end{aligned}$$

So, choosing a set $G \subset \mathcal{X}$, whose complement is compact, with

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(G) \leq -\lambda,$$

where $\lambda \in \mathbb{R}_{>0}$ is sufficiently large using the exponential tightness, and setting $M = G^C$

above, we get

$$\limsup_{n\to\infty} \frac{1}{\rho_n} \log P_n(K) \leq -I(K),$$

which was to be demonstrated. $\square$

3.2. Exponential Approximations

Another technique for demonstrating an LDP is to approximate the initial measures or random variables by others, which are simpler to analyze, and to demonstrate the LDP for the latter. For such an approach to be possible we need approximations that respect the LDP.

Definition 3.4. (Exponentially equivalent sequences of random variables). *Consider a Polish space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ with the metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{>0}$, and two sequences of $\mathcal{X}$-valued random variables $(X_n)_{n\in\mathbb{N}}, (\tilde{X}_n)_{n\in\mathbb{N}}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further, let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$. Then, $(X_n)_{n\in\mathbb{N}}, (\tilde{X}_n)_{n\in\mathbb{N}}$ are called* exponentially equivalent *in the scale $(\rho_n)_{n\in\mathbb{N}}$, if:*

(i) For all $n \in \mathbb{N}$ and all $\delta \in \mathbb{R}_{>0}$, the set $A_{n,\delta} := \{\omega \in \Omega, d(X_n(\omega), \tilde{X}_n(\omega)) > \delta\}$ is measurable;

(ii) And for all $\delta \in \mathbb{R}_{>0}$

$$\lim_{n\to\infty} \frac{1}{\rho_n} \log \mathbb{P}(A_{n,\delta}) = -\infty.$$

Definition 3.5. (Exponentially well approximations). *Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ be a Polish space with the metric $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{>0}$, $(X_n)_{n\in\mathbb{N}}$ a sequence of $\mathcal{X}$-valued random variables, and $(X_n^{(r)})_{n\in\mathbb{N}}, r \in \mathbb{N}$ a family of sequences of $\mathcal{X}$-valued random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_{>0}$ with $\lim_{n\to\infty} \rho_n = \infty$. Then, the family $(X_n^{(r)})_{n\in\mathbb{N}}, r \in \mathbb{N}$ is called an* exponentially well approximation *of the sequence $(X_n)_{n\in\mathbb{N}}$ in the scale $(\rho_n)_{n\in\mathbb{N}}$, if:*

(i) For all $n, r \in \mathbb{N}$ and all $\delta \in \mathbb{R}_{>0}$, the set $A_{n,r,\delta} := \{\omega \in \Omega, d(X_n(\omega), X_n^{(r)}(\omega)) > \delta\}$

is measurable;

(ii) And for all $\delta \in \mathbb{R}_{>0}$

$$\lim_{r\to\infty} \limsup_{n\to\infty} \frac{1}{\rho_n} \log \mathbb{P}(A_{n,r,\delta}) = -\infty.$$

Theorem 3.6. *Exponentially equivalent sequences of random variables satisfy the same LDP, if either one of them satisfies an LDP.*

Theorem 3.7. *Let the family $(X_n^{(r)})_{n\in\mathbb{N}}, r \in \mathbb{N}$ be an exponentially well approximation of the sequence $(X_n)_{n\in\mathbb{N}}$, and let the sequence $(X_n^{(r)})_{n\in\mathbb{N}}$ satisfy an LDP for all $r \in \mathbb{N}$ with corresponding rates $I_r : \mathcal{X} \to [0,\infty]$. Then $(X_n)_{n\in\mathbb{N}}$ satisfies an LDP with rate*

$$I : \mathcal{X} \to [0,\infty],\ x \mapsto I(x) := \sup_{\delta\in\mathbb{R}_{>0}} \liminf_{r\to\infty} \inf_{y\in B_\delta(x)} I_r(y)\ .$$

Proof. See [8] for the proofs. □

3.3. The Contraction Principle

The following theorem shows that an LDP on a topological space $\mathcal{X}$ can be "contracted" to an LDP on a space $\mathcal{Y}$ under a continuous mapping $\mathcal{X} \to \mathcal{Y}$. We demonstrate this principle in Chapter 5 on Markov chains, where we e.g., follow the LDP for the empirical measures on $\mathfrak{M}(\Gamma)$ from the LDP for the pair empirical measures on $\widetilde{\mathfrak{M}}(\Gamma\times\Gamma)$ by using a canonical continuous mapping $T : \widetilde{\mathfrak{M}}(\Gamma\times\Gamma) \to \mathfrak{M}(\Gamma)$. This approach is also in agreement with the intuition since the pair empirical measures contain more information.

Theorem 3.8. (Contraction principle). *Let $\mathcal{X},\mathcal{Y}$ be Polish spaces, $T : \mathcal{X} \to \mathcal{Y}$ a continuous map, and $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ satisfying an LDP, with speed n and rate function $I : \mathcal{X} \to [0,\infty]$. Then the sequence $(Q_n)_{n\in\mathbb{N}}$ of probability measures on $(\mathcal{Y},\mathcal{B}(\mathcal{Y}))$, given by $Q_n(B) := P_n(T^{-1}(B))$ for $B \in \mathcal{B}(\mathcal{Y})$, satisfies an LDP on $\mathcal{Y}$, with speed n and rate function*

$$J : \mathcal{Y} \to [0,\infty],\ y \mapsto J(y) := \inf\{I(x); x \in \mathcal{X}, T(x) = y\}.$$

Proof. If $C \in \mathcal{B}(\mathcal{Y})$ is closed, then $T^{-1}(C) \in \mathcal{B}(\mathcal{X})$ is closed; if $O \in \mathcal{B}(\mathcal{Y})$ is open, then $T^{-1}(O) \in \mathcal{B}(\mathcal{X})$ is open by the continuity of T. So, using the LDP bounds, we get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log Q_n(C) &= \limsup_{n\to\infty} \frac{1}{n} \log P_n(T^{-1}(C)) \\
&\leq -I(T^{-1}(C)) = -\inf_{x\in T^{-1}(C)} \{I(x)\} \\
&= -\inf_{x\in\mathcal{X}, T(x)\in C} \{I(x)\} \\
&= -\inf_{y\in C} \left\{ \inf_{x\in\mathcal{X}, T(x)=y} \{I(x)\} \right\} \\
&= \inf_{y\in C} \{J(y)\} = -J(C)
\end{aligned}$$

for all closed sets C in $\mathcal{B}(\mathcal{Y})$. Similarly,

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(O) \geq -J(O)$$

for all open sets O in $\mathcal{B}(\mathcal{Y})$. It remains to show that J is a rate function. Notice that $\mathcal{D}_I := \{x \in \mathcal{X}; I(x) < \infty\} \neq \emptyset$ implies $\mathcal{D}_J := \{y \in \mathcal{Y}; J(y) < \infty\} \neq \emptyset$. And since I has compact level sets, and T is continuous, J has compact level sets: Indeed, for $c \in \mathbb{R}$

$$\begin{aligned}
J^{-1}([-\infty, c]) &= \{y \in \mathcal{Y}; J(y) \in [-\infty, c]\} \\
&= \left\{ y \in \mathcal{Y}; \inf_{x\in\mathcal{X}, T(x)=y} \{I(x)\} \in [-\infty, c] \right\} \\
&= \{T(x); x \in \mathcal{X}, I(x) \in [-\infty, c]\} \\
&= T(I^{-1}([-\infty, c])) \text{ is compact.}
\end{aligned}$$

J is thus a rate function and the result follows. $\square$

3.4. Varadhan's Lemma and Tilting an LDP

The following theorem is arguably the most important proposition of the theory, because it enables the computation of the asymptotics of exponential expectations/integrals with respect to a sequence of measures satisfying an LDP:

Theorem 3.9. (Varadhan's lemma). *Let $\mathcal{X}$ be a Polish space and $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfying an LDP with speed n and rate function $I : \mathcal{X} \to [0, \infty]$, and let further $F : \mathcal{X} \to \mathbb{R}$, $x \mapsto F(x)$ be a continuous function which is bounded from above. Then,*

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

Proof. Set $J_n := \int_S \mathrm{e}^{nF(x)} \mathrm{d}P_n$, for $S \in \mathcal{B}(\mathcal{X})$.

Step 1. (The upper bound). Let

$$a := \sup\{F(x), x \in \mathcal{X}\} \text{ and } l := \sup\{F(x) - I(x), x \in \mathcal{X}\},$$

and note that $-\infty < l \leq a < \infty$. Now let $C := F^{-1}([l, a])$. We shall show that

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \leq l,$$

and then we extend this to

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \leq l,$$

which is the "upper bound". For obtaining the former inequality we discretize $[l, a]$ by setting

$$c_j^N = l + \frac{j}{N}(a - l), \; j = 0, 1, ..., N \text{ for } N \in \mathbb{N}.$$

Then $C = \cup_{j=1}^N C_j^N$, where $C_j^N := F^{-1}([c_{j-1}^N, c_j^N])$, $j = 1, ..., N$ are disjoint. And the C_j^N are closed, since F is continuous. So, we have the LDP bounds

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C_j^N) \leq -I(C_j^N), \; j = 1, ..., N.$$

Noticing that $F(x) \leq c_j^N$ on C_j^N, and using the proposition on the exponential growth of sums, i.e. Proposition 2.10, we now get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) &= \limsup_{n\to\infty} \frac{1}{n} \log(J_n(C_1^N) + ... + J_n(C_N^N)) \\
&= \limsup_{n\to\infty} \frac{1}{n} \log \max\{J_n(C_j^N), 1 \le j \le N\} \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \max\{\mathrm{e}^{n.c_j^N} P_n(C_j^N), 1 \le j \le N\} \\
&= \max\{c_j^N + \limsup_{n\to\infty} \frac{1}{n} \log P_n(C_j^N), 1 \le j \le N\} \\
&\le \max\{c_j^N - I(C_j^N), 1 \le j \le N\}.
\end{aligned}$$

Now, since $c_j^N \le \inf_{x\in C_j^N}\{F(x)\} + \frac{1}{N}(a-l)$ for all $1 \le j \le N$, we get

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) &\le \max_{1\le j\le N} \left\{ \inf_{x\in C_j^N}\{F(x)\} - \inf_{x\in C_j^N}\{I(x)\} \right\} + \frac{1}{N}(a-l) \\
&\le \max_{1\le j\le N} \left\{ \sup_{x\in C_j^N}\{F(x) - I(x)\} \right\} + \frac{1}{N}(a-l) \\
&= \sup_{x\in C}\{F(x) - I(x)\} + \frac{1}{N}(a-l) \\
&= l + \frac{1}{N}(a-l).
\end{aligned}$$

And so,

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \le l$$

follows by taking the limit as $N \to \infty$. To extend this upper bound to $J_n(\mathcal{X})$ we make use of the estimate $J_n(\mathcal{X} \setminus C) \le \mathrm{e}^{n.l}$ (use that $F(x) \le l$ on $\mathcal{X} \setminus C$), and then apply Proposition 2.10 once more in a similar fashion.

Step 2. (The lower bound).

Let $x \in \mathcal{X}$ and $\epsilon \in \mathbb{R}_{>0}$ be arbitrary. The set $G_{x,\epsilon} := \{y \in \mathcal{X}; F(y) > F(x) - \epsilon\}$ is an open neighborhood of x by the continuity of F. So the LDP gives

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G_{x,\epsilon}) \ge -I(G_{x,\epsilon}).$$

Using this we get,

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log J_n(\mathcal{X}) &\geq \liminf_{n\to\infty}\frac{1}{n}\log J_n(G_{x,\epsilon})\\ &\geq \liminf_{n\to\infty}\frac{1}{n}\log\left(\mathrm{e}^{n(F(x)-\epsilon)}P_n(G_{x,\epsilon})\right)\\ &= \liminf_{n\to\infty}(F(x)-\epsilon+\frac{1}{n}\log P_n(G_{x,\epsilon}))\\ &\geq F(x)-\epsilon-I(G_{x,\epsilon})\\ &\geq F(x)-\epsilon-I(x),\end{aligned}$$

for all $x\in\mathcal{X},\epsilon\in\mathbb{R}_{>0}$. Hence,

$$\liminf_{n\to\infty}\frac{1}{n}\log J_n(\mathcal{X})\geq l$$

follows by first letting $\epsilon\to 0$, and then taking the supremum over all $x\in\mathcal{X}$. The proposition follows combining the upper and the lower bounds we have presented. □

Theorem 3.10. (The tilted LDP). *Let $\mathcal{X}$ be a Polish space and $(P_n)_{n\in\mathbb{N}}$ a sequence of probability measures on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ satisfying an LDP, with speed n and rate function $I:\mathcal{X}\to[0,\infty]$, and let further $F:\mathcal{X}\to\mathbb{R}$, $x\mapsto F(x)$ be a continuous function which is bounded from above. Defining*

$$J_n(S):=\int_S \mathrm{e}^{nF(x)}\mathrm{d}P_n<\infty$$

for $S\in\mathcal{B}(\mathcal{X})$, gives rise to a sequence of probability measures on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$, given by

$$P_n^{(F)}(S):=\frac{J_n(S)}{J_n(\mathcal{X})}$$

for $S\in\mathcal{B}(\mathcal{X})$. Then the sequence $(P_n^{(F)})_{n\in\mathbb{N}}$ also satisfies an LDP on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$, with speed n and rate function

$$I^{(F)}:\mathcal{X}\to[0,\infty],\ x\mapsto I^{(F)}(x):=(I(x)-F(x))-\inf\{I(y)-F(y);y\in\mathcal{X}\}.$$

Proof.

$$\limsup_{n\to\infty} \frac{1}{n} \log J_n(C) \leq b(C), \text{ for all closed sets } C \in \mathcal{B}(\mathcal{X}),$$

$$\text{and, } \liminf_{n\to\infty} \frac{1}{n} \log J_n(O) \geq b(O), \text{ for all open sets } O \in \mathcal{B}(\mathcal{X}),$$

where $b(S) := -\inf_{x\in S}\{I(x) - F(x)\}$ for $S \subset \mathcal{X}$, are obtained analogous to the proof of Varadhan's lemma. Then using Varadhan's lemma and the above bounds we get,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log P_n^F(C) &= \limsup_{n\to\infty} \frac{1}{n} \log \frac{J_n(C)}{J_n(\mathcal{X})} \\
&= \limsup_{n\to\infty} \frac{1}{n} \log J_n(C) - \limsup_{n\to\infty} \frac{1}{n} \log J_n(\mathcal{X}) \\
&= \limsup_{n\to\infty} \frac{1}{n} \log J_n(C) + \inf_{y\in\mathcal{X}}\{I(y) - F(y)\} \\
&\leq -\inf_{x\in C}\{I(x) - F(x)\} + \inf_{y\in\mathcal{X}}\{I(y) - F(y)\} \\
&= -\inf_{x\in C}\left\{I(x) - F(x) - \inf_{y\in\mathcal{X}}\{I(y) - F(y)\}\right\} \\
&= -I^F(C),
\end{aligned}$$

for all closed sets $C \in \mathcal{B}(\mathcal{X})$. Similarly,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(O) \geq -I^F(O),$$

for all open sets $O \in \mathcal{B}(\mathcal{X})$. Finally we show that I^F is a rate function. Thence the result follows. Indeed, letting

$$B := \sup\{F(y); y \in \mathcal{X}\} < \infty,$$

we have

$$\inf\{I(y) - F(y); y \in \mathcal{X}\} \geq \inf\{I(y) - B; y \in \mathcal{X}\} \geq -B > -\infty.$$

Therefore, $I^F \not\equiv \infty$. And for every $c \in \mathbb{R}$, if some $x \in \mathcal{X}$ satisfies $I(x) - F(x) \leq c$, then $I(x) \leq c + B$. In other words,

$$(I - F)^{-1}((-\infty, c]) \subset I^{-1}((-\infty, c + B])$$

and the latter set is compact. So, it suffices to show that $(I-F)^{-1}((-\infty, c])$ is a closed set, and this follows from the semi-continuity of I. So, I has compact level sets and the result follows. □

3.5. The Gärtner-Ellis Theorem

The following theorem due J. Gärtner [9] is an elegant way to obtain the LDP when applicable, since one just has to show that a limit depending on a real parameter, defined below in the theorem, exists and defines a differentiable function to conclude, avoiding the tedious approximations needed for presenting the upper and lower bounds of the definition of an LDP separately. An extension was given by R. S. Ellis relaxing the condition that the limit should be finite. We present the former version.

Theorem 3.11. (Gärtner-Ellis theorem). *Let $d \in \mathbb{N}$ and $(Z_t)_{t \in \mathbb{R}_{>0}}$ be a family of random vectors $Z_t : \Omega_t \to \mathbb{R}^d$ on corresponding probability spaces $(\Omega_t, \mathcal{F}_t, \mathbb{P}_t)$. Let further, for all $\lambda \in \mathbb{R}^d$ the limits*

$$\lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}_t \left[\mathrm{e}^{t \cdot < \lambda, Z_t >} \right] =: \Lambda(\lambda) < \infty \text{ exist,}$$

and the map $\Lambda : \mathbb{R}^d \to \mathbb{R}$, $\lambda \mapsto \Lambda(\lambda)$ be differentiable, $\mathbb{E}_t$ standing for the corresponding expectations. Then, the measures $\mathbb{P}_t(Z_t \in \bullet)$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, which map $B \mapsto \mathbb{P}_t(Z_t^{-1}(B))$, satisfy an LDP with speed t, and with Λ^*, the Legendre transform of Λ, serving as the rate function.

Proof. We show the LDP upper bounds for closed sets. The lower bound is more technical and uses some ideas from *convex analysis*. The latter can be found in e.g., [10]. For the former it suffices to show the upper bounds for compact sets together with the exponential tightness of the measures, after Theorem 3.3. Here is some notation first for brevity: We set

$$\phi_t(\beta) := \mathbb{E}[\mathrm{e}^{<\beta, Z_t>}]$$

for $\beta \in \mathbb{R}^d, n \in \mathbb{N}$, and as above

$$\Lambda(\lambda) := \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[\mathrm{e}^{t\cdot<\lambda, Z_t>}] = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[\phi_t(t\lambda)]$$

for $\lambda \in \mathbb{R}^d$, which we assume to be a finite limit. Further, die Legendre transform of the function $\Lambda : \mathbb{R}^d \to (-\infty, \infty)$ is the function $\Lambda^* : \mathbb{R}^d \to [-\infty, \infty]$ given by

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

Finally, we introduce a notation for the measures in question by setting

$$\mu_t(B) := \mathbb{P}_t(Z_t \in B)$$

for $B \in \mathcal{B}(\mathbb{R}^d),\ t \in \mathbb{R}_{>0}$. Now we show the LDP bounds for compact sets. Let for this $K \subset \mathbb{R}^d$ be a compact set, $\epsilon \in \mathbb{R}_{>0}$ arbitrary, and define

$$\Lambda^*_\epsilon(x) := \min\left\{\Lambda^*(x) - \epsilon,\ \frac{1}{\epsilon}\right\}$$

for $x \in \mathbb{R}^d$. This is clearly a truncated version of Λ^*, with $\Lambda^*_\epsilon \to \Lambda^*$ pointwise as $\epsilon \to 0$. Since

$$\Lambda^*_\epsilon(x) \leq \Lambda^*(x) := \sup_{\lambda\in\mathbb{R}^d} \{< x, \lambda > -\Lambda(\lambda)\},$$

we follow that for all $x \in \mathbb{R}^d$ there exists a further vector, say λ_x, with

$$\Lambda^*_\epsilon(x) \leq < x, \lambda_x > -\Lambda(\lambda_x).$$

And since "$< \cdot, a >$" is a continuous function we have an open neighborhood, say A_x, of x with

$$\inf_{y \in A_x} < y, \lambda_x > \geq < x, \lambda_x > -\epsilon$$

and this is

$$\inf_{y \in A_x} < y - x, \lambda_x > \geq \ -\epsilon.$$

So we have an open cover $K \subset \cup_{x \in K} A_x$ for our compact set K, and hence a finite cover $K \subset \cup_{i=1,\dots,N} A_{x_i}$. Using the *Chebyshev's inequality* we obtain for these sets A_x the estimates:

$$\begin{aligned}
\mu_t(A_x) :&= \mathbb{P}_t(Z_t \in A_x) \\
&\leq \mathbb{P}_t(< Z_t - x, \lambda_x > \geq -\epsilon) \\
&= \mathbb{P}_t(\mathrm{e}^{t< Z_t - x, \lambda_x >} \geq \mathrm{e}^{-t\epsilon}) \\
&\leq \frac{\mathbb{E}_t[\mathrm{e}^{t< Z_t - x, \lambda_x >}]}{\mathrm{e}^{-t\epsilon}} \\
&= \mathrm{e}^{t\epsilon} \cdot \mathrm{e}^{-t< x, \lambda_x >} \cdot \mathbb{E}_t[\mathrm{e}^{t< Z_t, \lambda_x >}] \\
&= \mathrm{e}^{t\epsilon} \cdot \mathrm{e}^{-t< x, \lambda_x >} \cdot \phi_t(t\lambda_x) \ .
\end{aligned}$$

So, using this for $A_{x_i},\ i = 1, ..., N$ and using the Proposition 2.10 on the exponential rate of sums we get:

$$\begin{aligned}
\limsup_{t\to\infty} \frac{1}{t} \log \mu_t(K) &\leq \limsup_{t\to\infty} \frac{1}{t} \log \mu_t(\cup_{i=1,\dots,N} A_{x_i}) \\
&\leq \limsup_{t\to\infty} \frac{1}{t} \log(N \cdot \max_{i=1,\dots,N} \mu_t(A_{x_i})) \\
&= \limsup_{t\to\infty} \frac{1}{t} \log(\max_{i=1,\dots,N} \mu_t(A_{x_i})) \\
&= \max_{i=1,\dots,N} \{\limsup_{t\to\infty} \frac{1}{t} \log \mu_t(A_{x_i})\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \max_{i=1,\ldots,N} \left\{ \limsup_{t\to\infty} \frac{1}{t} \log \left(\mathrm{e}^{t\epsilon} \cdot \mathrm{e}^{-t<x,\lambda_x>} \cdot \phi_t(t\lambda_x) \right) \right\} \\
&= \max_{i=1,\ldots,N} \{ \epsilon - < x_i, \lambda_{x_i} > + \limsup_{t\to\infty} \frac{1}{t} \log \phi_t(t\lambda_x) \} \\
&= \epsilon - \min_{i=1,\ldots,N} \{ < x_i, \lambda_{x_i} > - \Lambda(x_i) \} \\
&\leq \epsilon - \min_{i=1,\ldots,N} \{ \Lambda^*_\epsilon(x_i) \} \\
&= \epsilon - \min_{i=1,\ldots,N} \{ \Lambda^*(x_i) - \epsilon \} \\
&\leq 2\epsilon - \inf_{x\in K} \{ \Lambda^*(x) \},
\end{aligned}
$$

for all sufficiently small $\epsilon \in \mathbb{R}_{>0}$.

$$
\limsup_{t\to\infty} \frac{1}{t} \log \mu_t(K) \leq - \inf_{x\in K} \{ \Lambda^*(x) \}
$$

follows by letting $\epsilon \to 0^+$. This was the LDP upper bound for compact sets. For the exponential tightness it suffices to show

$$
\lim_{N\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log \mu_t(\mathbb{R}^d \setminus [-N, N]^d) = -\infty.
$$

Let for this u_i denote the i-th unit vector in $\mathbb{R}^d$ for $i = 1, ..., d$. We again use Chebyshev's inequality and get for all $i = 1, ..., d$:

$$
\begin{aligned}
\mathbb{P}_t(Z_t^{(i)} \leq -N) &= \mathbb{P}_t(\mathrm{e}^{-tZ_t^{(i)}} \geq \mathrm{e}^{tN}) \\
&\leq \mathrm{e}^{-tN} \cdot \mathbb{E}_t[\mathrm{e}^{-tZ_t^{(i)}}] \\
&= \mathrm{e}^{-tN} \cdot \phi_t(-tu_i),
\end{aligned}
$$

where $Z_t^{(i)}$ denotes the i-th component of this random vector. Analogously, we have

$$
\mathbb{P}_t(Z_t^{(i)} \geq N) = \mathbb{P}_t(\mathrm{e}^{tZ_t^{(i)}} \geq \mathrm{e}^{tN}) \leq \mathrm{e}^{-tN} \cdot \mathbb{E}_t[\mathrm{e}^{tZ_t^{(i)}}] = \mathrm{e}^{-tN} \cdot \phi_t(tu_i).
$$

Using these estimates together with the Proposition 2.10, we get

$$\begin{aligned}
\limsup_{t\to\infty}\frac{1}{t}\log\mu_t(\mathbb{R}^d\setminus[-N,N]^d) &\le \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}_t(\exists i\in\{1,...,d\};Z_t^{(i)}\notin[-N,N]) \\
&= \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}_t(\cup_{i=1,\dots,d}\ \{|Z_t^{(i)}|>N\}) \\
&\le \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}_t\left(\cup_{i=1,\dots,d}\ (\{Z_t^{(i)}\ge N\}\cup\{Z_t^{(i)}\le -N\})\right) \\
&\le \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{i=1}^d\mathbb{P}_t(Z_t^{(i)}\ge N)+\sum_{i=1}^d\mathbb{P}_t(Z_t^{(i)}\le -N)\right) \\
&\le \limsup_{t\to\infty}\frac{1}{t}\log\left(\sum_{i=1}^d \mathrm{e}^{-tN}\phi_t(tu_i)+\sum_{i=1}^d\mathrm{e}^{-tN}\phi_t(-tu_i)\right) \\
&= \max_{i=1,\dots,d}\limsup_{t\to\infty}\frac{1}{t}\log\left(\max\{\mathrm{e}^{-tN}\phi_t(tu_i),\ \mathrm{e}^{-tN}\phi_t(-tu_i)\}\right) \\
&= \max_{i=1,\dots,d}\left\{\max\left\{-N+\limsup_{t\to\infty}\frac{1}{t}\log\phi_t(tu_i),\ -N+\limsup_{t\to\infty}\frac{1}{t}\log\phi_t(-tu_i)\right\}\right\} \\
&= -N+\max_{i=1,\dots,d}\left\{\max\left\{\Lambda(u_i),\ \Lambda(-u_i)\right\}\right\}.
\end{aligned}$$

$$\lim_{N\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log\mu_t(\mathbb{R}^d\setminus[-N,N]^d)=-\infty$$

follows by letting $N\to\infty$. □

Remark 3.1. The Gärtner-Ellis theorem can be viewed as an extension of Cramér's theorem. Indeed, we have for an i.i.d. sequence of $\mathbb{R}$-valued random variables $(X_i)_{i\in\mathbb{N}}$ with $\phi(t):=\mathbb{E}[\mathrm{e}^{tX_1}]<\infty$ for all $t\in\mathbb{R}$,

$$\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{n\cdot\lambda\frac{S_n}{n}}]=\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{\lambda\sum_{i=1}^n X_i}]=\frac{1}{n}\log(\mathbb{E}[\mathrm{e}^{\lambda X_1}])^n=\log\mathbb{E}[\mathrm{e}^{\lambda X_1}]$$

for all $\lambda\in\mathbb{R}$. And hence the limits

$$\Lambda(\lambda):=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\mathrm{e}^{n\cdot\lambda\frac{S_n}{n}}],\ \lambda\in\mathbb{R}$$

exist, with $\Lambda(\lambda)=\log\phi(\lambda)$ for all $\lambda\in\mathbb{R}$. In addition, the mapping $\lambda\mapsto\Lambda(\lambda)$ defines a differentiable function by the *monotone convergence theorem*, because by this theorem we have

$$\lim_{h\to 0}\frac{\mathbb{E}[\mathrm{e}^{(\lambda+h)X_1}]-\mathbb{E}[\mathrm{e}^{\lambda X_1}]}{h}=\mathbb{E}\left[\lim_{h\to 0}\frac{\mathrm{e}^{(\lambda+h)X_1}}{h}\right]=\mathbb{E}[\mathrm{e}^{\lambda X_1}\cdot X_1],$$

if $X_1\neq 0$ *almost surely*, and this case can be dealt with separately. Together with the differentiability of the logarithm function we now have the claim. This means that the assumptions of the Gärtner-Ellis theorem are met and thus the sequence $\big(\mathbb{P}^X(\frac{S_n}{n}\in\bullet)\big)_{n\in\mathbb{N}}$ satisfies an LDP with rate

$$I:\mathbb{R}\to[0,\infty],\ I(x):=\sup_{t\in\mathbb{R}}\{t\cdot x-\log\phi(t)\}.$$

Finally, Cramér's theorem follows from the LDP in the following way. By the LDP we have

$$-\inf_{x\in(a,\infty)}I(x)\leq\liminf_{n\to\infty}\log\mathbb{P}^X\left(\frac{S_n}{n}\in[a,\infty)\right),$$
$$\limsup_{n\to\infty}\log\mathbb{P}^X\left(\frac{S_n}{n}\in[a,\infty)\right)\leq-\inf_{x\in[a,\infty)}I(x),$$

and both the upper and the lower bounds in this display are equal to $I(a)$, since I is lower semi-continuous, and increasing for $a>\mathbb{E}[X_1]$. So the conclusion of Cramér's theorem follows, namely

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}^X(\frac{S_n}{n}\geq a)=-I(a),$$

for $a>\mathbb{E}[X_1]$.

4. LARGE DEVIATIONS FOR SEQUENCES OF I.I.D. RANDOM VARIABLES

4.1. Sanov's Theorem and LDP for the Empirical Measures of Sequences of I.I.D. Random Variables

In this section we firstly present the large deviations result on empirical measures which goes back to I. N. Sanov, see [11]. Then the full LDP for the empirical measures of an i.i.d. sequence is obtained from this. In the next section the analogous result for the pair empirical measures is presented. We will be repeating the argumentation from [10].

Theorem 4.1. (Sanov's theorem). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$ and law $\mathbb{P}^X$: $\mathcal{P}(\Gamma)^{\otimes\mathbb{N}} \to [0, 1]$, where $X_1, X_2, ...$ are i.i.d. with the distribution $\rho = (\rho_s)_{s\in\Gamma}$, and $\rho_s > 0$ for all $s \in \Gamma$. Further, let $\mathbb{P}^X(L_n \in \bullet)$ stand for the distributions of the empirical measures on $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$, see Definition 2.14. Then for all $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^X(L_n \in \bar{B}_a^C(\rho)) = -\inf_{\nu\in\bar{B}_a^C(\rho)} \{I_\rho(\nu)\},$$

where $\bar{B}_a^C(\rho) = \{\nu \in \mathfrak{M}(\Gamma); d(\nu, \rho) > a\}$, and

$$I_\rho : \mathfrak{M}(\Gamma) \to [0, \infty],\ \nu = (\nu_s)_{s\in\Gamma} \mapsto I_\rho(\nu) := \sum_{s=1}^{r} \nu_s \log(\frac{\nu_s}{\rho_s}).$$

Proof. We define $K_n := \{\vec{k} = (k_1, ..., k_r) \in \mathbb{N}_0^r;\ \sum_{s=1}^r k_s = n\}$ for $n \in \mathbb{N}$, and note that $\frac{1}{n}K_n := \{\frac{1}{n}\vec{k}; \vec{k} \in K_n\} \subset \mathfrak{M}(\Gamma)$. For $\vec{k} \in K_n$,

$$\mathbb{P}^X(L_n = \frac{1}{n}\vec{k}) = \frac{n!}{k_1!...k_r!}.\rho_1^{k_1}...\rho_r^{k_r} = n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!},$$

is seen by combinatorial considerations. Now let $a \in \mathbb{R}_{>0}$ be arbitrary, and set

$$Q_n(a) := \max_{\vec{k} \in K_n; \frac{1}{n}\vec{k} \in \bar{B}_a^C(\rho)} \{\mathbb{P}^X(L_n = \frac{1}{n}\vec{k})\},$$

where $\bar{B}_a^C(\rho) = \{\nu \in \mathfrak{M}(\Gamma); d(\nu, \rho) > a\}$, $d(\cdot, \cdot)$ standing for the total variation distance. Then we have

$$\forall n \in \mathbb{N} : Q_n(a) \leq \mathbb{P}^X(L_n \in \bar{B}_a^C(\rho)) \leq |K_n|.Q_n(a), \tag{4.1}$$

see the definition of L_n for this. But, Stirling's formula, Proposition 2.9 (ii), gives

$$\frac{1}{n} \log \mathbb{P}^X(L_n = \frac{1}{n}\vec{k}) = \frac{1}{n} \log n! \prod_{s=1}^{r} \frac{\rho_s^{k_s}}{k_s!} = \sum_{s=1}^{r} \frac{k_s}{n}(\log \rho_s - \log \frac{k_s}{n}) + O(\frac{\log n}{n}),$$

where the O-constant is uniform on K_n. This is, in fact, how the function I_ρ arises, because the sum on the right hand side equals $-I_\rho(\frac{1}{n}\vec{k})$. Thus, by using (4.1) and the fact, that $|K_n| = \binom{n+r}{r-1}$, we get

$$\frac{1}{n} \log \mathbb{P}^X(L_n \in \bar{B}_a^C(\rho)) = - \min_{\vec{k} \in K_n; \frac{1}{n}\vec{k} \in \bar{B}_a^C(\rho)} \{I_\rho(\frac{1}{n}\vec{k})\} + O(\frac{\log n}{n}).$$

At this point notice that $\cup_{n \in \mathbb{N}} \frac{1}{n} K_n$ is dense in $\mathfrak{M}(\Gamma)$, and that $I_\rho : \mathfrak{M}(\Gamma) \to [0, \infty)$ is continuous. Actually, for all $\nu \in \mathfrak{M}(\Gamma)$, there exists a sequence, say $(\nu_n)_{n \in \mathbb{N}}$, with $\nu_n \in \frac{1}{n} K_n$ for all $n \in \mathbb{N}$, such that $\lim_{n \to \infty} d(\nu_n, \nu) = 0$, and then, also $\lim_{n \to \infty} I_\rho(\nu_n) = I_\rho(\nu)$ by the continuity of I_ρ. Now, since $\bar{B}_a^C(\rho)$ is open, this implies

$$\limsup_{n \to \infty} \min_{\vec{k} \in K_n; \frac{1}{n}\vec{k} \in \bar{B}_a^C(\rho)} \{I_\rho(\frac{1}{n}\vec{k})\} \leq I_\rho(\nu), \text{ for all } \nu \in \bar{B}_a^C(\rho).$$

So, taking the infimum over the ν's, we get

$$\limsup_{n \to \infty} \min_{\vec{k} \in K_n; \frac{1}{n}\vec{k} \in \bar{B}_a^C(\rho)} \{I_\rho(\frac{1}{n}\vec{k})\} \leq \inf_{\nu \in \bar{B}_a^C(\rho)} \{I_\rho(\nu)\}.$$

And the reversed inequality is clear, since $\inf_{\nu \in \bar{B}_a^C(\rho)} \{I_\rho(\nu)\}$ is a lower bound for the sequence on the left hand side, whose *limit superior* is being considered there. Hence,

$$\limsup_{n\to\infty} \min_{\vec{k}\in K_n; \frac{1}{n}\vec{k}\in \bar{B}_a^C(\rho)} \{I_\rho(\frac{1}{n}\vec{k})\} = \inf_{\nu\in \bar{B}_a^C(\rho)} \{I_\rho(\nu)\}.$$

Similarly, we get

$$\liminf_{n\to\infty} \min_{\vec{k}\in K_n; \frac{1}{n}\vec{k}\in \bar{B}_a^C(\rho)} \{I_\rho(\frac{1}{n}\vec{k})\} = \inf_{\nu\in \bar{B}_a^C(\rho)} \{I_\rho(\nu)\}.$$

And thus, we indeed have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^X(L_n \in \bar{B}_a^C(\rho)) = -\inf_{\nu\in \bar{B}_a^C(\rho)} \{I_\rho(\nu)\}.$$

□

Theorem 4.2. (LDP for the empirical measures of a sequence of i.i.d. random variables). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$ and law $\mathbb{P}^X : \mathcal{P}(\Gamma)^{\otimes\mathbb{N}} \to [0,1]$, where $X_1, X_2, ...$ are i.i.d. with the distribution $\rho = (\rho_s)_{s\in\Gamma}$, and $\rho_s > 0$ for all $s \in \Gamma$. Then the distributions of the empirical measures, $\mathbb{P}^X(L_n \in \bullet)$ on $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$, satisfy an LDP with speed n and rate function*

$$I_\rho : \mathfrak{M}(\Gamma) \to [0,\infty],\ \nu = (\nu_s)_{s\in\Gamma} \mapsto I_\rho(\nu) := \sum_{s=1}^{r} \nu_s \log(\frac{\nu_s}{\rho_s}).$$

Proof. Firstly, observe that the same argumentation in the proof of Theorem 4.1 gives us the analogous result for arbitrary open balls in the place of the complements of closed balls particularly centered at the distribution ρ. That is, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^X(L_n \in B_a(\mu)) = -\inf_{\nu\in B_a(\mu)} \{I_\rho(\nu)\},$$

for all $\mu \in \mathfrak{M}(\Gamma)$ and $a \in \mathbb{R}_{>0}$, with $B_a(\mu) := \{\nu \in \mathfrak{M}(\Gamma); d(\nu,\rho) < a\}$. This actually suffices here to follow the LDP: We explain this as a general principle. For this, let's switch to a more general notation. Consider a Polish space $\mathcal{X}$ and a sequence $(P_n)_{n\in\mathbb{N}}$

of probability measures on $\mathcal{B}(\mathcal{X})$. For the LDP lower bounds for open sets to hold it suffices to have the local estimates,

$$\lim_{\delta\to 0^+}\liminf_{n\to\infty}\frac{1}{n}\log P_n(B_\delta(x)) \geq -I(x), \text{ for all } x\in\mathcal{X}.$$

And for the LDP upper bounds for closed sets to hold it suffices to have the local estimates

$$\lim_{\delta\to 0^+}\limsup_{n\to\infty}\frac{1}{n}\log P_n(B_\delta(x)) \leq -I(x), \text{ for all } x\in\mathcal{X},$$

together with the exponential tightness of $(P_n)_{n\in\mathbb{N}}$, where the exponential tightness is used to climb up from the WLDP to LDP. Back to our case, we have the local estimates for the lower bounds already. We actually have the local estimates for the upper bounds too, but here the continuity of I_ρ is also employed. Hence we have the WLDP. Now since $\mathfrak{M}(\Gamma)$ is compact itself, every closed subset here is compact, and the LDP follows by itself. $\square$

Remark 4.1. A finite state version of Cramér's theorem follows from Theorem 4.2 by using the contraction principle with a map $T:\mathfrak{M}(\Gamma)\to\mathbb{R}$, $\nu=(\nu_s)_{s\in\Gamma}\mapsto\sum_{s\in\Gamma}\nu_s\cdot s$.

4.2. LDP for the Pair Empirical Measures of Sequences of I.I.D. Random Variables

Theorem 4.3. (LDP for the pair empirical measures of a sequence of i.i.d. random variables). *Let $r\in\mathbb{N}$, $\Gamma=\{1,...,r\}$, and let $(X_n)_{n\in\mathbb{N}}$ be a stochastic process with probability space $(\Omega,\mathcal{F},\mathbb{P})$, state space $(\Gamma,\mathcal{P}(\Gamma))$ and law $\mathbb{P}^X$, where $X_1,X_2,...$ are i.i.d. with the distribution $\rho=(\rho_s)_{s\in\Gamma}$, and $\rho_s>0$ for all $s\in\Gamma$. Then distributions of the pair empirical measures, $\mathbb{P}^X((L_n^{(2)}\in\bullet)$ on $\Big(\widetilde{\mathfrak{M}}(\Gamma\times\Gamma),\mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma\times\Gamma))\Big)$, satisfy an LDP with speed n and rate function*

$$I_\rho^{(2)}:\widetilde{\mathfrak{M}}(\Gamma\times\Gamma)\to[0,\infty],\ \nu=(\nu_{st})_{(s,t)\in\Gamma\times\Gamma}\mapsto I_\rho^{(2)}(\nu):=\sum_{(s,t)\in\Gamma\times\Gamma}[\nu_{st}\log(\frac{\nu_{st}}{\bar{\nu}_s\rho_t})],$$

where $\bar{\nu}_s := \sum_{t\in\Gamma} \nu_{st}$ *for* $s \in \Gamma$.

Proof. The argumentation is very similar to that of Theorem 4.1, but the combinatorial aspects are a bit more involved. We recapitulate the proof given in [10]. This time we define,

$$K_n := \{\vec{k} = (k_{st})_{(s,t)\in\Gamma\times\Gamma} \in \mathbb{N}_0^{r^2};\ \sum_{(s,t)\in\Gamma\times\Gamma} k_{st} = n,\ \sum_{t\in\Gamma} k_{st} = \sum_{s\in\Gamma} k_{st}\}$$

for $n \in \mathbb{N}$, giving the subsets $\frac{1}{n}K_n \subset \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma\times\Gamma))$. Further, we set $\bar{k}_s = \sum_{t\in\Gamma} k_{st}$ for $s \in \Gamma$. We want to estimate the probability

$$\mathbb{P}^X(L_n^{(2)} = \frac{1}{n}\vec{k})$$

for $\vec{k} \in K_n$. We have that

$$\mathbb{P}^X(L_n^{(2)} = \frac{1}{n}\vec{k}) = C \cdot \prod_{s\in\Gamma} \rho_s^{\bar{k}_s},$$

where C stands for the number of possible arrangements of $X_1, X_2, ..., X_n$ that give rise to $\vec{k} = (k_{st})_{(s,t)\in\Gamma\times\Gamma}$ in a certain realization of the process. For an estimation of this combinatorial factor C, it is useful to interpret the situation in *graph theoretical* terms: Let each occurrence of $(s,t) \in \Gamma\times\Gamma$ among $(X_1,X_2), (X_2,X_3), ..., (X_{n-1},X_n), (X_n,X_1)$ be thought of as an arrow from s to t. In this way we obtain an *oriented graph* having Γ as its set of *vertices*, and the arrows as its set of *oriented edges*. The condition $\sum_{t\in\Gamma} k_{st} = \sum_{s\in\Gamma} k_{st}$ says then that for each vertex s, the number of incoming arrows equals to the number of outgoing arrows. The total number of arrows in the graph is $\sum_{(s,t)\in\Gamma\times\Gamma} k_{st} = n$. We have then

$$\mathbb{P}^X(L_n^{(2)} = \frac{1}{n}\vec{k}) = C' \cdot \frac{\mathcal{E}}{\prod_{(s,t)\in\Gamma\times\Gamma} k_{st}!} \prod_{s\in\Gamma} \rho_s^{\bar{k}_s}, \tag{4.2}$$

where $\mathcal{E}$ denotes the number of *Euler circuits* on the graph (i.e. the number of looped paths respecting the arrows and making use of each arrow precisely once), and C' is the number of distinct cyclic shifts of $(X_1, ..., X_n)$. Obviously,

$$1 \le C' \le n \ .$$

And the number of Euler circuits can be controlled by

$$\prod_{s\in\Gamma,\ \bar{k}_s>0} (\bar{k}_s - 1)! \le \mathcal{E} \le \prod_{s\in\Gamma,\ \bar{k}_s>0} \bar{k}_s! \ .$$

Using these estimates and the Equation (4.2), we see that

$$\mathbb{P}^X(L_n^{(2)} = \frac{1}{n}\vec{k}) = \mathrm{e}^{O(\log n)} \frac{\prod_{s\in\Gamma} \bar{k}_s!}{\prod_{(s,t)\in\Gamma\times\Gamma} k_{st}!} \prod_{s\in\Gamma} \rho_s^{\bar{k}_s}, \text{ on } K_n \ .$$

The rest is analogous to the proofs of Theorems 4.1 and 4.2. □

5. LARGE DEVIATIONS FOR MARKOV CHAINS

We now present large deviations results related to Markov chains.

5.1. LDPs for the Pair Empirical Measures, the Empirical Measures and the Empirical Means of DTMCs

We will be following the plan of the argumentation in [10] in this section. The results go back to M. D. Donsker and S. R. S. Varadhan, see [12]. See the introductory remarks to the Section 5.3 for an overview of what we are doing in this chapter.

Theorem 5.1. (LDP for the pair empirical measures of a finite state DTMC). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $(X_n)_{n\in\mathbb{N}}$ be a DTMC with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$, transition matrix $P = (p_{st})_{(s,t)\in\Gamma\times\Gamma}$ and law $\mathbb{P}^X$, and let $p_{st} > 0$ for all $s, t \in \Gamma$. Then the distributions of the pair empirical measures, $\mathbb{P}^X(L_n^{(2)} \in \bullet)$ on $\Big(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma), \mathcal{B}(\widetilde{\mathfrak{M}}(\Gamma \times \Gamma))\Big)$, satisfy an LDP with speed n and rate function*

$$I_P^{(2)} : \widetilde{\mathfrak{M}}(\Gamma \times \Gamma) \to [0, \infty],\ \nu = (\nu_{st})_{(s,t)\in\Gamma\times\Gamma} \mapsto I_P^{(2)}(\nu) := \sum_{(s,t)\in\Gamma\times\Gamma} [\nu_{st} \log(\frac{\nu_{st}}{\bar{\nu}_s p_{st}})],$$

where $\bar{\nu}_s := \sum_{t\in\Gamma} \nu_{st}$ for $s \in \Gamma$.

Proof. A stationary distribution, say π, $\pi = (\pi_s)_{s\in\Gamma}$, for the Markov chain exists under the assumptions, and is unique. Furthermore it satisfies $\pi_s > 0$ for all $s \in \Gamma$, see [14]. Introduce an auxiliary sequence $(Y_n)_{n\in\mathbb{N}}$ of independent random variables, each with the distribution π. We shall tilt the LDP of $\mathbb{P}^Y(L_n^{(2)} \in \bullet)$, where $\mathbb{P}^Y$ is the law of the process $(Y_n)_{n\in\mathbb{N}}$, and the result shall follow.
For this, let now

$$[x_1, ..., x_n] := \{(\tau_n)_{n\in\mathbb{N}} \in \Gamma^{\mathbb{N}}; \tau_1 = x_1, ..., \tau_n = x_n\} \in \mathcal{P}(\Gamma)^{\otimes\mathbb{N}}$$

be an arbitrary *simple rectangle*. Then for every *path* $x \in [x_1, ..., x_n]$ we have

$$\begin{aligned}\mathbb{P}^X([x_1, ..., x_n]) &= \mathbb{P}(X_1 = x_1, ..., X_n = x_n) \\ &= \mathbb{P}(X_1 = x_1).p_{x_1,x_2}...p_{x_{n-1},x_n} \\ &= \frac{\mathbb{P}(X_1 = x_1)}{p_{x_n,x_1}} \mathrm{e}^{\sum_{i=1}^{n-1} \log p_{x_i,x_{i+1}} + \log p_{x_n,x_1}} \\ &= \frac{\mathbb{P}(X_1 = x_1)}{p_{x_n,x_1}} \mathrm{e}^{n \sum_{(s,t)\in\Gamma\times\Gamma} l^2_{n,x}(s,t).\log p_{st}},\end{aligned}$$

where $l^{(2)}_{n,x}(s,t)$ denotes the member of $L^{(2)}_n(x)$ corresponding to the pair (s,t), as in Definition 2.15. Similarly,

$$\begin{aligned}\mathbb{P}^Y([x_1, ..., x_n]) &= \prod_{i=1}^{n} \pi_{x_i} \\ &= \mathrm{e}^{n \sum_{(s,t)\in\Gamma\times\Gamma} l^{(2)}_{n,x}(s,t).\log \pi_t},\end{aligned}$$

for all $x \in [x_1, ..., x_n]$. Thus, we have

$$\mathbb{P}^X([x_1, ..., x_n]) = O(1)\mathrm{e}^{nF(L^{(2)}_n(x))}\mathbb{P}^Y([x_1, ..., x_n]),$$

for all paths $x \in [x_1, ..., x_n]$, where

$$F : \widetilde{\mathfrak{M}}(\Gamma\times\Gamma) \to \mathbb{R},\ \nu = (\nu_{st})_{(s,t)\in\Gamma\times\Gamma} \mapsto \sum_{(s,t)\in\Gamma\times\Gamma} \nu_{st}\log(\frac{p_{st}}{\pi_t}).$$

Notice that $F(L^{(2)}_n(x))$ is constant for $x \in [x_1, ..., x_n]$, so, we have now

$$\int_{[x_1,...,x_n]} 1\mathrm{d}\mathbb{P}^X = \int_{[x_1,...,x_n]} O(1)\mathrm{e}^{nF(L^{(2)}_n(x))}\mathrm{d}\mathbb{P}^Y.$$

This is then also clearly correct for all *finite rectangles* in $\mathcal{P}(\Gamma)^{\otimes\mathbb{N}}$, that is: for all finite rectangles $Q \in \mathcal{P}(\Gamma)^{\otimes\mathbb{N}}$ we have

$$\int_Q 1\mathrm{d}\mathbb{P}^X = \int_Q O(1)\mathrm{e}^{nF(L^{(2)}_n(x))}\mathrm{d}\mathbb{P}^Y,$$

where the O-constant actually does not depend on the rectangle Q. Now finally let $S \in \widetilde{\mathfrak{M}}(\Gamma \times \Gamma)$. $(L_n^{(2)})^{-1}(S)$ is a finite rectangle in $\mathcal{P}(\Gamma)^{\otimes \mathbb{N}}$ for all $n \in \mathbb{N}$. So, we have

$$\begin{aligned}
\frac{1}{n} \log \mathbb{P}^X((L_n^{(2)})^{-1}(S)) &= \frac{1}{n} \log \int_{(L_n^{(2)})^{-1}(S))} 1 \mathrm{d}\mathbb{P}^X \\
&= \frac{1}{n} \log \int_{((L_n^{(2)})^{-1}(S))} O(1) \mathrm{e}^{nF(L_n^{(2)}(x))} \mathrm{d}\mathbb{P}^Y \\
&= O(\frac{1}{n}) + \frac{1}{n} \log \int_{((L_n^{(2)})^{-1}(S))} \mathrm{e}^{nF(L_n^{(2)}(x))} \mathrm{d}\mathbb{P}^Y \\
&= O(\frac{1}{n}) + \frac{1}{n} \log \int_S \mathrm{e}^{nF(\nu)} \mathrm{d}(\mathbb{P}^Y(L_n^{(2)} \in \bullet)).
\end{aligned}$$

The proposition follows from this by Theorems 4.3 and 3.10. Indeed, on the right-hand side, the tilt of the pair empirical measures of i.i.d. random variables occur, and the $O(\frac{1}{n})$ term ensures that the LDP also holds for the measures of the left-hand side. □

Theorem 5.2. (LDP for the empirical measures of a finite state DTMC). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $(X_n)_{n \in \mathbb{N}}$ be a DTMC with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$, transition matrix $P = (p_{st})_{(s,t) \in \Gamma \times \Gamma}$ and law $\mathbb{P}^X$, and let $p_{st} > 0$ for all $s, t \in \Gamma$. Then:*

(i) The distributions of the empirical measures, $\mathbb{P}^X(L_n \in \bullet)$ on $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$, satisfy an LDP with speed n and rate function

$$I_P : \mathfrak{M}(\Gamma) \to [0, \infty], \;\; \mu = (\mu_s)_{s \in \Gamma} \mapsto I_P(\mu) := \inf\{I_P^{(2)}(\nu); \nu \in \widetilde{\mathfrak{M}}(\Gamma \times \Gamma), \bar{\nu} = \mu\},$$

where $\bar{\nu} := (\bar{\nu}_s)_{s \in \Gamma}$ for $\nu \in \widetilde{\mathfrak{M}}(\Gamma \times \Gamma)$, with $\bar{\nu}_s := \sum_{t \in \Gamma} \nu_{st}$ for $s \in \Gamma$.

(ii)

$$I_P(\mu) = \sup_{u > 0} \{- \sum_{s=1}^{r} \mu_s \log \frac{(Pu)_s}{u_s}\}, \; \textit{for all } \mu \in \mathfrak{M}(\Gamma),$$

where the supremum runs over all families $u = (u_s)_{s \in \Gamma} \in (0, \infty)^\Gamma$, and $(Pu)_s = \sum_{t \in \Gamma} P_{st} u_t$ for $s \in \Gamma$.

Proof. *(i):* Consider the map

$$T : \widetilde{\mathfrak{M}}(\Gamma \times \Gamma) \to \mathfrak{M}(\Gamma),\ \nu \mapsto \bar{\nu} := (\bar{\nu}_s)_{s\in\Gamma}.$$

T is continuous, because $d(\bar{\nu}, \bar{\mu}) \leq d(\nu, \mu)$ for all $\nu, \mu \in \widetilde{\mathfrak{M}}(\Gamma \times \Gamma)$. And we have

$$\mathbb{P}^X(L_n^{-1}(B)) = \mathbb{P}^X\left((L_n^{(2)})^{-1}(T^{-1}(B))\right),$$

for all $B \in \mathfrak{M}(\Gamma)$, since for all x in $\Gamma^{\mathbb{N}}$, $L_n(x) = T(L_n^{(2)}(x))$. So, part (i) follows from the *contraction principle*, noticing that the function I_P is the rate function described in the contraction principle corresponding to this map T. For (ii), see [10]. □

Theorem 5.3. (LDP for the empirical means of a finite state DTMC). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$, transition matrix $P = (p_{st})_{(s,t)\in\Gamma\times\Gamma}$ and law $\mathbb{P}^X$, and let $p_{st} > 0$ for all $s, t \in \Gamma$. Let further $f : \Gamma \to \mathbb{R}$, $x \mapsto f(x)$ be a deterministic function. For $n \in \mathbb{N}$, we define*

$$S_n : \Gamma^{\mathbb{N}} \to \mathbb{R},\ x = (x_k)_{k\in\mathbb{N}} \mapsto S_n(x) := \frac{1}{n}\sum_{i=1}^{n} f(x_i).$$

S_n is $(\mathcal{P}(\Gamma)^{\otimes\mathbb{N}}, \mathcal{B}(\mathbb{R}))$-measurable. This gives rise to a sequence of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $B \mapsto \mathbb{P}^X(S_n^{-1}(B))$, called the distributions of the empirical means of $(f(X_n))_{n\in\mathbb{N}}$, and denoted by $\mathbb{P}^X(S_n \in \bullet)$. Then the measures $\mathbb{P}^X(S_n \in \bullet)$, $n \in \mathbb{N}$ satisfy an LDP on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with speed n and rate function

$$I : \mathbb{R} \to [0, \infty],\ \lambda \mapsto I(\lambda) := \inf\{I_P(\nu); \nu \in \mathfrak{M}(\Gamma), \sum_{s\in\Gamma} \nu_s f(s) = \lambda\}.$$

Proof. Apply the contraction principle once more, this time with the map $T : \mathfrak{M}(\Gamma) \to \mathbb{R}$, $\nu = (\nu_s)_{s\in\Gamma} \mapsto \sum_{s\in\Gamma} \nu_s f(s)$. □

5.2. LDP for the Occupation Time Measures of CTMCs

We move over to the setting where the time parameter varies in $\mathbb{R}_{\geq 0}$. Again we will be following the line of argumentation in [10].

Theorem 5.4. (LDP for the occupation time measures of a finite state CTMC). *Let $r \in \mathbb{N}$, $\Gamma = \{1, ..., r\}$, and $X = (X_t)_{t\in\mathbb{R}_{\geq 0}}$ be a CTMC with probability space $(\Omega, \mathcal{F}, \mathbb{P})$, state space $(\Gamma, \mathcal{P}(\Gamma))$, and law $\mathbb{P}^X : \mathcal{P}(\Gamma)^{\otimes\mathbb{R}_{\geq 0}} \to [0, 1]$, and let X possess a generator matrix $G = (G_{ij})_{i,j\in\Gamma} \in \mathbb{R}^{\Gamma\times\Gamma}$, which is irreducible. Then the distributions of the occupation time measures, $\mathbb{P}^X(L_t \in \bullet)$ on $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$, satisfy an LDP, with speed t and rate function*

$$\widetilde{I}_G : \mathfrak{M}(\Gamma) \to [0, \infty],\ \nu = (\nu_i)_{i\in\Gamma} \mapsto \widetilde{I}_G(\nu) := \sup_{u>0}[-\sum_{i=1}^{r} \nu_i \frac{(Gu)_i}{u_i}],$$

where the supremum runs over all families $u = (u_i)_{i\in\Gamma} \in (0, \infty)^\Gamma$, and $(Gu)_i = \sum_{j\in\Gamma} G_{ij}u_j$ for $i \in \Gamma$.

Proof. Step 1. Let $\alpha \in \mathbb{R}_{>0}$ be fixed but arbitrary. For $t \in \mathbb{R}_{\geq 0}$ we define

$$L_t^\alpha : \Gamma^{\mathbb{R}_{\geq 0}} \to \mathfrak{M}(\Gamma),\ x = (x_s)_{s\in\mathbb{R}_{\geq 0}} \mapsto L_t^\alpha(x) := (l_{t,x}^\alpha(\gamma_1), ..., l_{t,x}^\alpha(\gamma_r)),$$

where for fixed $\alpha \in \mathbb{R}_{>0}$, $t \in \mathbb{R}_{\geq 0}$, $x = (x_s)_{s\in\mathbb{R}_{\geq 0}} \in \Gamma^{\mathbb{R}_{\geq 0}}$, and $\gamma \in \Gamma$,

$$l_{t,x}^\alpha(\gamma) := \frac{1}{[\frac{t}{\alpha}]} \sum_{k=1}^{[\frac{t}{\alpha}]} \mathbf{1}_{\{x_{k\alpha}\}}(\gamma).$$

L_t^α is $\big(\mathcal{P}(\Gamma)^{\otimes\mathbb{R}_{\geq 0}}, \mathcal{B}(\mathfrak{M}(\Gamma))\big)$-measurable for all $\alpha \in \mathbb{R}_{>0}$, and $t \in \mathbb{R}_{\geq 0}$. This gives rise to the measures $\mathbb{P}^X(L_t^\alpha \in \bullet)$ on $(\mathfrak{M}(\Gamma), \mathcal{B}(\mathfrak{M}(\Gamma)))$.

$\mathbb{P}^X(L_t^\alpha \in \bullet)$ is actually a continuous form of the distribution of the empirical measure after $[\frac{t}{\alpha}]$ steps of the Γ-valued discrete-time Markov chain $(X_{k\alpha})_{k\in\mathbb{N}}$, whose transition matrix is given by $\mathrm{e}^{\alpha G} =: P^\alpha$, see [13], where $\mathrm{e}^A := \sum_{n=0}^{\infty} \frac{1}{n!}A^n$ for square matrices A with real coefficients is the *matrix exponential function*. Notice that P^α has strictly positive coefficients since G is irreducible. Thus, using Theorem 5.2 we see that the families $(\mathbb{P}^X(L_t^\alpha \in \bullet))_{t\in\mathbb{R}_{\geq 0}}$, $\alpha \in \mathbb{R}_{>0}$, satisfy LDPs with speed t and with the corresponding rate functions $\frac{1}{\alpha}I_{P^\alpha}$, $\alpha \in \mathbb{R}_{>0}$, where

$$\frac{1}{\alpha}I_{P^\alpha} : \mathfrak{M}(\Gamma) \to [0,\infty],\ \nu = (\nu_i)_{i\in\Gamma} \mapsto \frac{1}{\alpha}I_{P^\alpha}(\nu) := \frac{1}{\alpha}\sup_{u>0}[-\sum_{i=1}^{r}\nu_i \log(\frac{(P^\alpha u)_i}{u_i})],$$

in accordance with the notation in Theorem 5.2 .

Step 2.

$$\text{Claim: } \lim_{\alpha\to 0^+}\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}^X(d(L_t, L_t^\alpha) \geq \epsilon) = -\infty, \text{ for all } \epsilon \in \mathbb{R}_{>0}.$$

Indeed, firstly we have for all realizations, say $X(\omega_0)$, of the process,

$$d\left(L_t(X(\omega_0)), L_t^\alpha(X(\omega_0))\right) \leq \frac{1}{[\frac{t}{\alpha}]}\cdot\iota,$$

where ι denotes the number of jumps in the realization $X(\omega_0)$ in $[0,t]$, since every jump causes an "error" of at most $\frac{1}{[\frac{t}{\alpha}]}$. And from the theory of Markov processes, we have that this number ι is less than or equal to $N(ct)(\omega_0)$, where $N(ct) : \Omega \to \mathbb{N}_0$, $t \in \mathbb{R}_{>0}$ are Poisson random variables with corresponding means ct, and $c := \sup_{i\in\Gamma}\{\sum_{j\neq i} G_{ij}\} < \infty$. Hence, for $\frac{\epsilon}{\alpha c} > 1$ we can estimate,

$$\begin{aligned}
\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}^X\left(d(L_t, L_t^\alpha) \geq \epsilon\right) &\leq \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(N(ct) \geq \epsilon(\frac{t}{\alpha})\right) \\
&= c\cdot\limsup_{t\to\infty}\frac{1}{ct}\log\mathbb{P}\left(\frac{N(ct)}{ct} \geq \frac{\epsilon}{\alpha c}\right) \\
&= -cI(\frac{\epsilon}{\alpha c}),
\end{aligned}$$

where $I : \mathbb{R} \to [0,\infty]$, $z \mapsto I(z)$, is the rate function in of a continuous version of Cramér's theorem for the family $(\frac{N(ct)}{ct})_{t\in\mathbb{R}_{>0}}$ of $\mathbb{R}$-valued random variables satisfying the assumptions of such a Cramér's theorem; this is also where the $\frac{\epsilon}{\alpha c} > 1 = \mathbb{E}[\frac{N(ct)}{ct}]$ assumption is made use of. The rate function I can be computed explicitly in this case of Poisson distribution. Namely, it is given by $I(z) = z\log z - z + 1$ for $z \in \mathbb{R}$. So, we get

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}^X \left(d(L_t, L_t^\alpha) \geq \epsilon\right) \leq -\frac{\epsilon}{\alpha} \log(\frac{\epsilon}{\alpha c}) - \frac{\epsilon}{\alpha} - c,$$

for $\alpha < \frac{\epsilon}{c}$. The "claim" follows by letting $\alpha \to 0^+$. This will enable us to demonstrate an LDP for $(\mathbb{P}^X(L_t \in \bullet))_{t\in\mathbb{R}_{\geq 0}}$ by defining a rate function as

$$\mathfrak{M}(\Gamma) \to [0,\infty],\ \nu \mapsto \lim_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(\nu).$$

Step 3. For the mere definition of such a function we must show that $\lim_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(\nu)$ either exists, or diverges to infinity, as $\alpha \to 0^+$. Indeed, we show that $\widetilde{I}_G(\nu)$ from the statement of the theorem, equals $\lim_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(\nu)$, which also takes care of this convergence issue:

Since $P^0 = \mathcal{I}$, where $\mathcal{I}$ denotes the identity matrix, and $\frac{\mathrm{d}}{\mathrm{d}\alpha} P^\alpha = GP^\alpha$, in the sense $\lim_{h\to 0^+} \frac{1}{h}(P^{\alpha+h} - P^\alpha) = GP^\alpha$, see [13], we get

$$\begin{aligned} \frac{1}{\alpha} I_{P^\alpha}(\nu) = \sup_{u>0} \left\{ -\sum_{i=1}^r \nu_i \frac{1}{\alpha} \int_0^\alpha \frac{(GP^{\beta u})_i}{(P^\beta u)_i} \mathrm{d}\beta \right\} &\leq \frac{1}{\alpha} \int_0^\alpha \sup_{u>0} \{ -\sum_{i=1}^r \nu_i \frac{(GP^\beta u)_i}{(P^\beta u)_i} \} \mathrm{d}\beta \\ &\leq \frac{1}{\alpha} \int_0^\alpha \sup_{u>0} \{ -\sum_{i=1}^r \nu_i \frac{(Gu)_i}{(u)_i} \} \mathrm{d}\beta \\ &= \frac{1}{\alpha} \int_0^\alpha \widetilde{I}_G(\nu) \mathrm{d}\beta \\ &= \widetilde{I}_G(\nu), \end{aligned}$$

for all $\alpha \in \mathbb{R}_{>0}$. So, we have

$$\limsup_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(\nu) \leq \widetilde{I}_G(\nu). \tag{5.1}$$

On the other hand, for all $u = (u_i)_{i\in\Gamma} \in (0,\infty)^\Gamma$ there exists a constant $c(u) \in \mathbb{R}$, such that

$$-\frac{1}{\alpha} \log \left(\frac{(P^\alpha u)_i}{u_i} \right) = -\frac{1}{\alpha} \log \left(1 + \frac{1}{u_i}((P^\alpha - \mathcal{I})u)_i \right)$$

$$\geq -\frac{1}{u_i}\left(\frac{1}{\alpha}(P^\alpha - \mathcal{I})u\right)_i$$
$$= -\frac{1}{u_i}\left(\frac{1}{\alpha}(P^\alpha - P^0)u\right)_i$$
$$\geq \frac{-(Gu)_i}{u_i} - c(u)\alpha,$$

for sufficiently small $\alpha \in \mathbb{R}_{>0}$, and all $i \in \Gamma$. Using this we get

$$\begin{aligned}\frac{1}{\alpha}I_{P^\alpha}(\nu) &= \frac{1}{\alpha}\sup_{u>0}\left\{-\sum_{i=1}^r \nu_i \log(\frac{(P^\alpha u)_i}{u_i})\right\} \\ &\geq -\sum_{i=1}^r \nu_i \frac{1}{\alpha}\log\left(\frac{(P^\alpha u)_i}{u_i}\right) \\ &\geq -\left(\sum_{i=1}^r \nu_i \frac{(Gu)_i}{u_i}\right) - r\cdot c(u)\alpha,\end{aligned}$$

for sufficiently small $\alpha \in \mathbb{R}^+$, and all $u \in (0,\infty)^\Gamma$.

$$\liminf_{\alpha\to 0^+}\frac{1}{\alpha}I_{P^\alpha}(\nu) \geq -\sum_{i=1}^r \nu_i \frac{(Gu)_i}{u_i},$$

for all $u \in (0,\infty)^\Gamma$, follows. Taking the supremum over all $u \in (0,\infty)^\Gamma$ we get,

$$\liminf_{\alpha\to 0^+}\frac{1}{\alpha}I_{P^\alpha}(\nu) \geq \widetilde{I}_G(\nu).$$

Hence, together with the Inequality 5.1 we have now

$$\liminf_{\alpha\to 0^+}\frac{1}{\alpha}I_{P^\alpha}(\nu) = \widetilde{I}_G(\nu),$$

for all $\nu \in \mathfrak{M}(\Gamma)$.

Step 4. We conclude the argumentation by showing the LDP bounds. Firstly, let $O \subset \mathfrak{M}(\Gamma)$ be open. Then for all $\epsilon \in \mathbb{R}_{>0}$ we have

$$\mathbb{P}^X(L_t \in O) \geq \mathbb{P}^X(L_t^\alpha \in O_\epsilon) - \mathbb{P}^X\left(d(L_t, L_t^\alpha) \geq \epsilon\right),$$

where $O_\epsilon := O \cap \{\nu \in \mathfrak{M}(\Gamma); d(\nu, \delta O) > \epsilon\}$ is the ϵ-interior of O. Using the claims in Steps 2 and 3, we get

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{t}\mathbb{P}^X(L_t \in O) &\geq -\liminf_{\epsilon\to 0^+} \liminf_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(O_\epsilon) \\ &\geq -\liminf_{\epsilon\to 0^+} \widetilde{I}_G(O_\epsilon),\end{aligned}$$

and this last quantity is equal to $-\widetilde{I}_G(O)$, because every $\nu \in O$ falls in O_ϵ for sufficiently small $\epsilon \in \mathbb{R}^+$. Thus, $\mathbb{P}^X(L_t \in O) \geq -\widetilde{I}_G(O)$.

Secondly, let $K \subset \mathfrak{M}(\Gamma)$ be closed. Since $\mathfrak{M}(\Gamma)$ is compact, every closed set, and hence also K, is compact in $\mathfrak{M}(\Gamma)$.

We have for all $\epsilon \in \mathbb{R}^+$,

$$\mathbb{P}^X(L_t \in K) \leq \mathbb{P}^X(L_t^\alpha \in K_\epsilon) + \mathbb{P}^X(d(L_t, L_t^\alpha) \geq \epsilon),$$

where $K_\epsilon := K \cup \{\nu \in \mathfrak{M}(\Gamma); d(\nu, \delta K) \leq \epsilon\}$ is the ϵ-exterior of K; δK denoting the boundary of K. Again using Step 2 we get,

$$\limsup_{t\to\infty} \frac{1}{t}\mathbb{P}^X(L_t \in K) \leq \limsup_{\epsilon\to 0^+} \limsup_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(K_\epsilon).$$

Next, we show

$$\limsup_{\alpha\to 0^+} \frac{1}{\alpha} I_{P^\alpha}(K_\epsilon) \geq \widetilde{I}_G(K_\epsilon) \text{ for all } \epsilon \in \mathbb{R}^+. \tag{5.2}$$

Here is the explanation: K_ϵ is compact and $\frac{1}{\alpha} I_{P^\alpha}$ is continuous, so there exists a $\nu_{\epsilon,\alpha} \in K_\epsilon$ with $\frac{1}{\alpha} I_{P^\alpha}(K_\epsilon) = \frac{1}{\alpha} I_{P^\alpha}(\nu_{\epsilon,\alpha})$. The sequence $(\nu_{\epsilon,1/n})_{n\in\mathbb{N}}$ has a convergent subsequence with the limit, say $\nu_\epsilon \in K_\epsilon$. Then we have

$$\limsup_{n\to\infty} \left(n \cdot I_{P^{1/n}}(\nu_{\epsilon,1/n})\right) \geq \widetilde{I}_G(\nu_\epsilon)$$

by the same argumentation in the beginning of Step 3, using the fact that $\widetilde{I}_G$ is continuous. And this establishes the Inequality 5.2, since $\widetilde{I}_G(\nu_\epsilon) \geq \widetilde{I}_G(K_\epsilon)$.

It remains to show that

$$\limsup_{\epsilon \to 0^+} \widetilde{I}_G(K_\epsilon) \geq \widetilde{I}_G(K),$$

which we show next: Since K_ϵ is compact and $\widetilde{I}_G$ is continuous, there exists a $\bar{\nu}_\epsilon \in K_\epsilon$ such that $\widetilde{I}_G(K_\epsilon) = \widetilde{I}_G(\bar{\nu}_\epsilon)$. The sequence $(\bar{\nu}_{1/n})_{n \in \mathbb{N}}$ has a convergent subsequence with the limit, say $\bar{\nu} \in K$. Then, we get,

$$\limsup_{n \to \infty} \widetilde{I}_G(\bar{\nu}_{1/n}) = \widetilde{I}_G(\bar{\nu}).$$

And this implies

$$\limsup_{\epsilon \to 0^+} \widetilde{I}_G(K_\epsilon) \geq \widetilde{I}_G(K),$$

because $\widetilde{I}_G(\bar{\nu}) \geq \widetilde{I}_G(K)$. So, we have established

$$\mathbb{P}^X(L_t \in K) \leq -\widetilde{I}_G(K).$$

Finally, notice that $\widetilde{I}_G$ is a rate function. Thence the proposition follows. $\square$

5.3. Another Treatment of the LDP for the Occupation Time Measures of CTMCs

So far we have presented large deviations principles for Markov chains both in discrete-, and continuous-time settings. We obtained the results for the continuous-time setting from the results of the discrete-time setting via an approximation argumentation, discretising the time and using the *exponentially wellness* of the resulting approximations. And the results in the discrete-time case were based on the i.i.d. setting by applying a tilt to the LDP for an auxiliary i.i.d. sequence of random variables having the stationary distribution of the chain. We now present another treatment, which uses the Gärtner-Ellis theorem to follow the LDP directly in the continuous-time setting. For an execution of the same method in the discrete case, see [8]. The *Perron-Frobenius theorem* plays an important role here:

Theorem 5.5. (Perron-Frobenius theorem). *Let $n \in \mathbb{N}$, and $B = (b_{ij})_{i,j=1,...,n} \in \mathbb{R}^{n\times n}$ be a positive matrix, i.e., a matrix with positive entries $b_{ij} > 0,\ i,j = 1,...,n$. Then, B possesses an eigenvalue $\mu \in \mathbb{R}$, which is maximal in the sense that for all further eigenvalues $\lambda \neq \mu$ of B in $\mathbb{C}$,*

$$|\lambda| < \mu$$

holds. In addition the following statements hold true:

(i) *μ is of multiplicity 1, i.e., $\dim_{\mathbb{R}} V_\mu = 1$, where V_μ stands for the eigenspace of μ.*
(ii) *μ possesses an eigenvector with positive entries.*
(iii) *For every $i = 1,...,n$ and every vector $\Phi = (\Phi_j)_{j=1,...,n}$ with positive entries, $\Phi_j > 0,\ j = 1,...,n$, we have,*

$$\lim_{m\to\infty} \frac{1}{m} \log\left(B^m\Phi\right)_i = \log\mu.$$

Proof. For a treatment of the subject see [15]. We merely demonstrate (iii) assuming the rest. For this let $\Phi = (\Phi_j)_{j=1,..,n}$ be a vector with positive entries, $v = (v_j)_{j=1,..n}$ be an eigenvector of positive entries from part (2), and set

$$\alpha := \max_{j=1,..,n}\{v_j\},\ \beta := \min_{j=1,..,n}\{v_j\},\ \gamma := \max_{j=1,..,n}\{\Phi_j\} \text{ and } \delta := \min_{j=1,..,n}\{\Phi_j\}.$$

Then for all $i,j = 1,..,n$, we have

$$\frac{\gamma}{\beta}B^m(i,j)v_j \geq B^m(i,j)\Phi_j \geq \frac{\delta}{\alpha}B^m(i,j)v_j.$$

Therefore, indeed

$$\lim_{m\to\infty}\frac{1}{m}\log\left(B^m\Phi\right)_i = \lim_{m\to\infty}\frac{1}{m}\log\left(\sum_{j=1}^{n}B^m(i,j)\Phi_j\right)$$
$$= \lim_{m\to\infty}\frac{1}{m}\log\left(\sum_{j=1}^{n}B^m(i,j)v_j\right) = \lim_{m\to\infty}\frac{1}{m}\log(\mu^m v_i) = \log\mu\ .$$

□

Theorem 5.6. (LDP for the occupation time measures of a CTMC II). *Consider an irreducible CTMC, $(X_s)_{s\in\mathbb{R}_{>0}}$, with a finite state space Γ, and let $f:\Gamma\to\mathbb{R}$ be a deterministic function. Assume, we have the $\mathbb{R}$-valued random variables*

$$Z_t = \frac{1}{t}\int_0^t f(X_s)\mathrm{d}s,\ t\in\mathbb{R}_{>0}.$$

Then, the family of distributions $\mathbb{P}(Z_t\in\bullet)$, $t\in\mathbb{R}_{>0}$ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ satisfies an LDP with speed n.

Proof. We shall use the Gärtner-Ellis theorem. Therefore we show that for all $\theta\in\mathbb{R}$ the limit,

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^t f(X_s)\mathrm{d}s}],$$

exists as a finite limit, and that the function which maps $\theta\in\mathbb{R}$ to this limit correspondingly is differentiable. Firstly notice that it is enough to consider

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}]$$

instead, discretizing the time, since if this limit exists, then

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^t f(X_s)\mathrm{d}s}] = \lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}].$$

Indeed, using the fact that $f : \Gamma \to \mathbb{R}$ is thus bounded, we have

$$\mathbb{E}[\mathrm{e}^{\theta\int_0^t f(X_s)\mathrm{d}s}] = \mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s + \int_{[t]}^t f(X_s)\mathrm{d}s}]$$
$$= \mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s + O(1)}]) = O(1)\cdot\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}],$$

and so,

$$\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^t f(X_s)\mathrm{d}s}] = \frac{1}{t}\log(O(1)) + \frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}],$$

from which the assertion follows. Secondly, we may even consider the limit

$$\lim_{t\to\infty}\frac{1}{[t]}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}]$$

instead of

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^t f(X_s)\mathrm{d}s}],\ \text{ or } \lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}],$$

since $\lim_{t\to\infty}\frac{t}{[t]} = 1$. Now for $m \le t < m+1$, and $\theta \in \mathbb{R}$ we compute, that

$$\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}] = \mathbb{E}[\mathrm{e}^{\theta\int_0^{m} f(X_s)\mathrm{d}s}] = \mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-1}\int_n^{n+1} f(X_s)\mathrm{d}s}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-1}\int_n^{n+1} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s + \theta\int_{m-1}^{m} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s}\cdot\mathrm{e}^{\theta\int_{m-1}^{m} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\cdot\mathrm{e}^{\theta\int_{m-1}^{m} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]\cdot\mathbb{E}[\mathrm{e}^{\theta\int_{m-1}^{m} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]$$
$$= \sum_{x_{m-1}\in\Gamma}\mathbb{E}[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}]\cdot\mathbb{E}[\mathrm{e}^{\theta\int_{0}^{1} f(X_s)\mathrm{d}s}\cdot\mathbf{1}_{\{X_0=x_{m-1}\}}]$$

$$
\begin{aligned}
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-2}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-3}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathrm{e}^{\theta\int_{m-2}^{m-1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-3}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\mathrm{e}^{\theta\int_{m-2}^{m-1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\cdot \\
&\qquad\qquad\cdot\,\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-3}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_{m-2}^{m-1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-3}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_{m-2}^{m-1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{m-3}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_{m-1}=x_{m-1}\}}\mathbf{1}_{\{X_{m-2}=x_{m-2}\}}\big]\cdot \\
&\qquad\qquad\quad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big] \\
&\qquad\qquad\qquad\vdots
\end{aligned}
$$

And so on, by induction, we have,

$$
\begin{aligned}
&= \sum_{x_{m-1}\in\Gamma}\sum_{x_{m-2}\in\Gamma}\sum_{x_{m-3}\in\Gamma}\cdots\sum_{x_2\in\Gamma} \mathbb{E}\big[\mathrm{e}^{\theta\sum_{n=0}^{1}\int_n^{n+1} f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_2=x_2\}}\big]\cdot \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_2,X_1=x_3\}}\big]\cdot \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_3,X_1=x_4\}}\big]\cdot \\
&\qquad\qquad\vdots \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-2},X_1=x_{m-1}\}}\big]\cdot \\
&\qquad\cdot\,\mathbb{E}\big[\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}\mathbf{1}_{\{X_0=x_{m-1}\}}\big].
\end{aligned}
$$

And for the first factor of the summands above, we have

$$
\begin{aligned}
&\mathbb{E}[\mathrm{e}^{\theta \sum_{n=0}^{1} \int_n^{n+1} f(X_s)\mathrm{d}s} \mathbf{1}_{\{X_2=x_2\}}] \\
&= \sum_{x_1\in\Gamma} \mathbb{E}[\mathrm{e}^{\theta \sum_{n=0}^{1} \int_n^{n+1} f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_2=x_2\}} \cdot \mathbf{1}_{\{X_1=x_1\}}] \\
&= \sum_{x_1\in\Gamma} \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1\}} \cdot \mathrm{e}^{\theta \int_1^2 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1, X_2=x_2\}}] \\
&= \sum_{x_1\in\Gamma} \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1\}}] \cdot \mathbb{E}[\mathrm{e}^{\theta \int_1^2 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1, X_2=x_2\}}] \\
&= \sum_{x_1\in\Gamma} \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1\}}] \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_0=x_1, X_1=x_2\}}].
\end{aligned}
$$

We get after these calculations, for $m \le t < m+1$,

$$
\begin{aligned}
\mathbb{E}[\mathrm{e}^{\theta \int_0^{[t]} f(X_s)\mathrm{d}s}] = \sum_{x_{m-1}, x_{m-2}, \ldots, x_1 \in\Gamma} & \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_1=x_1\}}] \cdot \\
& \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_0=x_1, X_1=x_2\}}] \cdot \\
& \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_0=x_2, X_1=x_3\}}] \cdot \\
& \quad \vdots \\
& \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_0=x_{m-2}, X_1=x_{m-1}\}}] \cdot \\
& \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \mathbf{1}_{\{X_0=x_{m-1}\}}].
\end{aligned}
$$

So, setting

$$
p_\theta(x,y) := \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \mathbf{1}_{\{X_0=x, X_1=y\}}] \text{ for } x, y \in \Gamma,
$$

we have, for $m \le t < m+1$,

$$
\begin{aligned}
&\mathbb{E}[\mathrm{e}^{\theta \int_0^{[t]} f(X_s)\mathrm{d}s}] = \\
&\quad = \sum_{x_{m-1}, x_{m-2}, \ldots, x_1 \in\Gamma} p_\theta(x_0, x_1) \cdot p_\theta(x_1, x_2) \cdots p_\theta(x_{m-2}, x_{m-1}) \cdot \mathbb{E}[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \mathbf{1}_{\{X_0=x_{m-1}\}}]
\end{aligned}
$$

$$= \sum_{x_m,\ldots,x_1\in\Gamma} p_\theta(x_0,x_1)\cdot p_\theta(x_1,x_2)\cdots p_\theta(x_{m-2},x_{m-1})\cdot p_\theta(x_{m-1},x_m)$$
$$= \left((p_\theta)^m \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)_{x_0},$$

where (p_θ) is the $(k\times k)$-matrix with entries $p_\theta(x,y)$, $x,y\in\Gamma$, and $|\Gamma|=k$. Since the chain is assumed to be irreducible, we have that (p_θ) is a positive matrix, and hence the Perron-Frobenius theorem, i.e. Theorem 5.5, in particular the part (iii), applies and we get,

$$\lim_{t\to\infty}\frac{1}{[t]}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{[t]} f(X_s)\mathrm{d}s}] = \lim_{t\to\infty}\frac{1}{[t]}\log\left((p_\theta)^{[t]} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right)_{x_0} = \log\mu_\theta,$$

where $\mu_\theta\in\mathbb{R}_{>0}$ denotes the Perron-Frobenius eigenvalue of (p_θ), showing also that

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[\mathrm{e}^{\theta\int_0^{t} f(X_s)\mathrm{d}s}] = \log\mu_\theta,$$

as we have justified at the very beginning of the proof. It remains yet to show the differentiability of this map, $\theta\mapsto\log\mu_\theta$. It suffices to show that the entries $p_\theta(x,y)$, $x,y\in\Gamma$ are differentiable functions of θ. This claim follows from the *implicit function theorem*. Indeed, the Perron-Frobenius eigenvalue is a solution of the *characteristic equation* of the matrix, with multiplicity one. It is an implicit function of the entries of the matrix. We complete the argumentation by showing the differentiability of the entries. Indeed, we have

$$\frac{d}{d\theta}p_\theta(x,y) = \lim_{h\to\infty}\mathbb{E}\left[\frac{\mathrm{e}^{(\theta+h)\int_0^1 f(X_s)\mathrm{d}s}-\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}}{h}\cdot\mathbf{1}_{\{X_0=x,X_1=y\}}\right]$$
$$= \mathbb{E}\left[\lim_{h\to\infty}\left(\frac{\mathrm{e}^{(\theta+h)\int_0^1 f(X_s)\mathrm{d}s}-\mathrm{e}^{\theta\int_0^1 f(X_s)\mathrm{d}s}}{h}\cdot\mathbf{1}_{\{X_0=x,X_1=y\}}\right)\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \lim_{h\to\infty}\left(\frac{\mathrm{e}^{h\int_0^1 f(X_s)\mathrm{d}s} - 1}{h}\right) \cdot \mathbf{1}_{\{X_0=x, X_1=y\}}\right] \\
&= \mathbb{E}\left[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \lim_{h\to\infty}\left(\frac{\mathrm{e}^{h\int_0^1 f(X_s)\mathrm{d}s} - 1}{h}\right) \cdot \mathbf{1}_{\{X_0=x, X_1=y\}}\right] \\
&= \mathbb{E}\left[\mathrm{e}^{\theta \int_0^1 f(X_s)\mathrm{d}s} \cdot \int_0^1 f(X_s)\mathrm{d}s \cdot \mathbf{1}_{\{X_0=x, X_1=y\}}\right] < \infty,
\end{aligned}
$$

using the *monotone convergence theorem* and the assumptions. The theorem follows.

□

6. LARGE DEVIATIONS RELATED TO THE ERDŐS-KAC THEOREM

6.1. The Erdős-Kac Theorem

The following theorem of P. Erdős and M. Kac is famous, in particular because it is one of those theorems which demonstrate an application of probabilistic methods in mathematics itself. It can be interpreted as a *central limit theorem* for the number of distinct prime factors of a uniformly chosen random natural number less than n. It describes deviations from the "mean" behavior, of order $\sqrt{\log\log n}$ in its context. In the next section we present large deviations corresponding to this central limit theorem.

Theorem 6.1. (Erdős-Kac theorem). *Let $v(m)$ denote the number of distinct prime factors of m for $m \in \mathbb{N}$; and $K_{n,\omega}$, for $n \in \mathbb{N}$, $\omega \in \mathbb{R}$, the number of those integers $m \in \mathbb{N}$ from $\{1, ..., n\}$ for which $v(m) < \log\log n + \omega\sqrt{2\log\log n}$ holds. Then,*

$$\lim_{n\to\infty} \frac{K_{n,\omega}}{n} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\omega} \mathrm{e}^{-x^2} \mathrm{d}x.$$

Or, equivalently,

$$\forall \omega \in \mathbb{R} : \frac{\#\{m \in \mathbb{N};\ m \le n,\ \dfrac{v(m) - \log\log n}{\sqrt{\log\log n}} < \omega\}}{n} \to \Phi(\omega) \text{ as } n \to \infty,$$

where $\Phi : \mathbb{R} \to \mathbb{R}$, $x \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2}\, \mathrm{d}u$ is the cumulative distribution function of the standard normal distribution.

Proof. For a proof see [16]. But here is the interpretation of the theorem in our context: Let $\Omega_n := \{1, \cdots, n\}$, and define for each $n \in \mathbb{N}$ a probability measure on $(\Omega_n, \mathcal{P}(\Omega_n))$ by assigning each element of Ω_n the same probability $\frac{1}{n}$. Now we consider the random variables $X_n : \Omega_n \to \mathbb{R}$, $m \mapsto X_n(m) := m$, where $\mathbb{R}$ is considered as $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then $\frac{v(X_n)-\log\log n}{\sqrt{\log\log n}}$ for $n \in \mathbb{N}$ is also a random variable $\Omega_n \to \mathbb{R}$. And then we consider the distributions of the latter. The theorem says, that the values of the distribution of $\frac{v(X_n)-\log\log n}{\sqrt{\log\log n}}$ converges to the corresponding values of the standard normal distribution on $\mathbb{R}$ as $n \to \infty$. $\square$

6.2. Large Deviations for Strongly Additive Arithmetic Functions

Let $\mathbf{P}$ denote the set of all prime numbers in this section. See [3] for the original work, which we repeat here.

Theorem 6.2. (Large deviations for Erdős-Kac theorem). *Let $\Omega_n := \{1, \cdots, n\}$, and define for each $n \in \mathbb{N}$ a probability measure on $(\Omega_n, \mathcal{P}(\Omega_n))$ by assigning each element of Ω_n the same probability $\frac{1}{n}$, and call these measures $\mathbb{P}_n$. Consider the random variables $X_n : \Omega_n \to \mathbb{R}$, $m \mapsto X_n(m) := m$, where $\mathbb{R}$ is considered as $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and let again $v : \mathbb{N} \to \mathbb{N}$, $k \mapsto v(k)$ denote the distinct prime factor count function. Then the sequence of probability measures $\mathbb{P}_n\left(\frac{v(X_n)}{\log\log n} \in \bullet\right)$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfies a large deviations principle with speed* $\log\log n$ *and rate function, $I : \mathbb{R} \to [0, \infty]$, $x \mapsto I(x)$, where*

$$I(x) := \begin{cases} x\log x - x + 1, & \textit{for} \quad x \geq 0, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. See the following exposition, which contains the claim in a more general form. $\square$

As in the case of the Erdős-Kac theorem itself, see the original article of Erdős and Kac, [16], the large deviations result holds in a more general form.

Definition 6.3. (Strongly additive arithmetic function). *Let $g : \mathbb{N} \to \mathbb{C}$ be a function. g is called a* strongly additive arithmetic function, *if:*

(i) $\forall m, n \in \mathbb{N} : \gcd(m, n) = 1 \implies g(mn) = g(m) + g(n)$.

(ii) $\forall p \in \mathbf{P}, k \in \mathbb{N} : g(p^k) = g(p)$.

Thus, by the *fundamental theorem of arithmetic*, assigning a value $g(p)$ for each prime $p \in \mathbf{P}$ gives rise to a unique strongly additive arithmetic function g.

Theorem 6.4. (Large deviations for strongly additive arithmetic functions). *Let* $\Omega_n := \{1, \cdots, n\}$, *and define for each* $n \in \mathbb{N}$ *a probability measure on* $(\Omega_n, \mathcal{P}(\Omega_n))$ *by assigning each element of* Ω_n *the same probability* $\frac{1}{n}$, *and call these measures* $\mathbb{P}_n$. *Consider the random variables* $X_n : \Omega_n \to \mathbb{R}$, $m \mapsto X_n(m) := m$, *where* $\mathbb{R}$ *is considered as* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *and let* $g : \mathbb{N} \to \mathbb{R}$ *be a strongly additive arithmetic function. Define for* $n \in \mathbb{N}$ *probability measures* ρ_n *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *by setting*

$$\rho_n(A) := \frac{\sum_{p \in \mathbf{P}, g(p) \in A, p \leq n} (1/p)}{\sum_{p \in \mathbf{P}, p \leq n} (1/p)} \text{ for } A \in \mathcal{B}(\mathbb{R}).$$

Assume further that there exists a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *say* ρ, *such that for all* $\theta \in \mathbb{R}$,

$$\int_{\mathbb{R}} \mathrm{e}^{\theta y} \rho(\mathrm{d}y) < \infty,$$

and with

$$\int_{\mathbb{R}} \mathrm{e}^{\theta y} \rho_n(\mathrm{d}y) \to \int_{\mathbb{R}} \mathrm{e}^{\theta y} \rho(\mathrm{d}y)$$

as $n \to \infty$. *Then, the sequence of probability measures* $\mathbb{P}_n \left(\frac{g(X_n)}{\log \log n} \in \bullet \right)$ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *satisfies a large deviations principle with speed* $\log \log n$ *and rate*

$$I : \mathbb{R} \to [0, \infty], \ x \mapsto I(x) := \sup\{\theta x - \int_{\mathbb{R}} (\mathrm{e}^{\theta y} - 1) \rho(\mathrm{d}y), \theta \in \mathbb{R}\} \ .$$

Proof. $g(X_n(m)) = \sum_{p \in \mathbf{P}, p \leq n} g(p) Z_p$, where $Z_p = 1$ if $X_n(m)$ is divisible by p and $Z_p = 0$ otherwise, is seen by the strong additivity.

The result is presented in 4 steps. The first three establish approximation lemmas. The final step concludes the argumentation.

Step 1. Claim: For all $\epsilon \in \mathbb{R}_{>0}$ we have,

$$\limsup_{C\to\infty} \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(|\sum_{p\in\mathbf{P}, |g(p)|>C, p\le n} g(p) Z_p| \ge \epsilon \log\log n \right) = -\infty.$$

Indeed, we show this by showing:

$$\limsup_{C\to\infty} \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(|\sum_{p\in\mathbf{P}, g(p)>C, p\le n} g(p) Z_p| \ge (\epsilon/2) \log\log n \right) = -\infty,$$

and

$$\limsup_{C\to\infty} \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(|\sum_{p\in\mathbf{P}, g(p)<-C, p\le n} g(p) Z_p| \ge (\epsilon/2) \log\log n \right) = -\infty.$$

For these, construct independent *Bernoulli random variables* Y_p for $p \in \mathbf{P}$ with $Y_p = 1$ with probability $\frac{1}{p}$, and $Y_p = 0$ with probability $1 - \frac{1}{p}$. Then for distinct primes $p_1, p_2, \cdots, p_l$ we have,

$$\mathbb{E}[Z_{p_1} Z_{p_2} \cdots Z_{p_l}] = \frac{1}{n} \left\lfloor \frac{n}{p_1 p_2 \cdots p_l} \right\rfloor \le \frac{1}{p_1 p_2 \cdots p_l} = \mathbb{E}[Y_{p_1} Y_{p_2} \cdots Y_{p_l}].$$

Using this and the series definition of the exponential function, we get for non-negative real numbers θ_p, $p \in \mathbf{P}$,

$$\mathbb{E}\left[\mathrm{e}^{\sum_{p\in\mathbf{P}, p\le n} \theta_p Z_p} \right] \le \mathbb{E}\left[\mathrm{e}^{\sum_{p\in\mathbf{P}, p\le n} \theta_p Y_p} \right].$$

Now for $\theta \in \mathbb{R}_{>0}$, we have using Chebyshev's inequality,

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(|\sum_{p\in\mathbf{P}, g(p)>C, p\le n} g(p) Z_p| \ge \epsilon \log\log n \right) \\ &= \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(\mathrm{e}^{\theta \sum_{p\in\mathbf{P}, g(p)>C, p\le n} g(p) Z_p} \ge \mathrm{e}^{\theta\epsilon \log\log n} \right) \end{aligned}$$

$$
\begin{aligned}
&\leq \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{E}\left[\mathrm{e}^{\theta \sum_{p\in\mathbf{P},g(p)>C,p\leq n} g(p)Z_p}\right] - \theta\epsilon \\
&\leq \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{E}\left[\mathrm{e}^{\theta \sum_{p\in\mathbf{P},g(p)>C,p\leq n} g(p)Y_p}\right] - \theta\epsilon \\
&= \limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{E}\left[\prod_{p\in\mathbf{P},g(p)>C,p\leq n} \mathrm{e}^{\theta g(p)Y_p}\right] - \theta\epsilon \\
&= \limsup_{n\to\infty} \frac{1}{\log\log n} \log \prod_{p\in\mathbf{P},g(p)>C,p\leq n} \mathbb{E}\left[\mathrm{e}^{\theta g(p)Y_p}\right] - \theta\epsilon \\
&= \limsup_{n\to\infty} \frac{1}{\log\log n} \sum_{p\in\mathbf{P},g(p)>C,p\leq n} \log \mathbb{E}\left[\mathrm{e}^{\theta g(p)Y_p}\right] - \theta\epsilon \\
&= \limsup_{n\to\infty} \frac{1}{\log\log n} \sum_{p\in\mathbf{P},g(p)>C,p\leq n} \log[(\mathrm{e}^{\theta g(p)} - 1)\frac{1}{p} + 1] - \theta\epsilon \\
&\leq \limsup_{n\to\infty} \frac{1}{\log\log n} \sum_{p\in\mathbf{P},g(p)>C,p\leq n} [(\mathrm{e}^{\theta g(p)} - 1)\frac{1}{p}] - \theta\epsilon.
\end{aligned}
$$

On the other hand, using the assumptions we have,

$$
\begin{aligned}
&\int_{y>C} (\mathrm{e}^{\theta y} - 1)\rho(\mathrm{d}y) = \lim_{n\to\infty} \int_{y>C} (\mathrm{e}^{\theta y} - 1)\rho_n(\mathrm{d}y) \\
&= \lim_{n\to\infty} \sum_{p\in\mathbf{P},g(p)>C} (\mathrm{e}^{\theta g(p)} - 1) \cdot \left(\frac{\frac{1}{p}\cdot\delta_{\{p\leq n\}}}{\sum_{p\in\mathbf{P},p\leq n}\frac{1}{p}}\right) \\
&= \lim_{n\to\infty} \frac{\sum_{p\in\mathbf{P},g(p)>C,p\leq n}(\mathrm{e}^{\theta g(p)} - 1)}{\sum_{p\in\mathbf{P},p\leq n}\frac{1}{p}} \\
&= \lim_{n\to\infty} \frac{\sum_{p\in\mathbf{P},g(p)>C,p\leq n}(\mathrm{e}^{\theta g(p)} - 1)}{\log\log n} \ ,
\end{aligned}
$$

since $\sum_{p\in\mathbf{P},p\leq n}\frac{1}{p} \sim \log\log n$, see e.g., [17]. So, we have for all $\theta \in \mathbb{R}_{>0}$,

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n\left(|\sum_{p\in\mathbf{P},g(p)>C,p\leq n} g(p)Z_p| \geq \epsilon\log\log n\right) \\
&\leq \int_{y>C} (\mathrm{e}^{\theta y} - 1)\rho(\mathrm{d}y) - \theta\epsilon \leq \int_{y>C} \mathrm{e}^{\theta y}\rho(\mathrm{d}y) - \theta\epsilon \ ,
\end{aligned}
$$

and this last quantity tends to $-\theta\epsilon$ as $C \to \infty$.

So we have,

$$\limsup_{C\to\infty}\limsup_{n\to\infty}\frac{1}{\log\log n}\log\mathbb{P}_n\left(|\sum_{p\in\mathbf{P},g(p)>C,p\le n}g(p)Z_p|\ge(\epsilon/2)\log\log n\right)=-\infty,$$

by letting $\theta\to\infty$ and observing that $\epsilon\in\mathbb{R}_{>0}$ is arbitrary. Similarly,

$$\limsup_{n\to\infty}\frac{1}{\log\log n}\log\mathbb{P}_n\left(|\sum_{p\in\mathbf{P},g(p)<-C,p\le n}g(p)Z_p|\ge\epsilon\log\log n\right)\le\int_{y<-C}\mathrm{e}^{\theta y}\rho(\mathrm{d}y)+\theta\epsilon,$$

and so,

$$\limsup_{C\to\infty}\limsup_{n\to\infty}\frac{1}{\log\log n}\log\mathbb{P}_n\left(|\sum_{p\in\mathbf{P},g(p)<-C,p\le n}g(p)Z_p|\ge(\epsilon/2)\log\log n\right)=-\infty,$$

by letting $\theta\to-\infty$. The claim follows.

Step 2. Let $k_n:=n^{\frac{1}{(\log\log n)^2}}$ for $n\in\mathbb{N}$.

Claim: For all $\epsilon\in\mathbb{R}_{>0},C\in\mathbb{R}_{\ge0}$ we have,

$$\limsup_{n\to\infty}\frac{1}{\log\log n}\log\mathbb{P}_n\left(|\sum_{p\in A(n,C)}g(p)Z_p|\ge\epsilon\log\log n\right)=-\infty,$$

where $A(n,C):=\{p\in\mathbf{P};|g(p)|\le C,k_n\le p\le n\}$, for $n\in\mathbb{N},C\in\mathbb{R}_{\ge0}$.

Indeed, for any $\theta\in\mathbb{R}_{>0}$, we have,

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{\theta|\sum_{p\in A(n,C)}g(p)Z_p|}]\le&\mathbb{E}[\mathrm{e}^{\theta\sum_{p\in A(n,C)}|g(p)|Z_p}]\\\le&\mathbb{E}[\mathrm{e}^{\theta C\sum_{p\in A(n,C)}Z_p}]\\\le&\mathbb{E}[\mathrm{e}^{\theta C\sum_{p\in A(n,C)}Y_p}].\end{aligned}$$

Therefore, for any $\theta\in\mathbb{R}_{>0}$, we have,

$$\log\mathbb{E}[\mathrm{e}^{\theta|\sum_{p\in A(n,C)}g(p)Z_p|}]\le\log\mathbb{E}[\mathrm{e}^{\theta C\sum_{p\in A(n,C)}Y_p}]$$

$$
\begin{aligned}
&= \sum_{p \in A(n,C)} \log((\mathrm{e}^{\theta C} - 1)\frac{1}{p} + 1) \\
&\leq \sum_{p \in A(n,C)} (\mathrm{e}^{\theta C} - 1)\frac{1}{p} \\
&= (\mathrm{e}^{\theta C} - 1) \sum_{p \in \mathbf{P}, k_n \leq p \leq n} \frac{1}{p}.
\end{aligned}
$$

And notice that

$$
\begin{aligned}
\sum_{p \in \mathbf{P}, k_n \leq p \leq n} \frac{1}{p} = \sum_{p \in \mathbf{P}, p \leq n} \frac{1}{p} - \sum_{p \in \mathbf{P}, p < k_n} \frac{1}{p} &\sim \log\log n - \log\log k_n \\
&= \log\log n - \log(\frac{1}{(\log\log n)^2} \log n) = 2\log\log\log n.
\end{aligned}
$$

So, again using Chebyshev's inequality in the appropriate way, we get,

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log\log n} \log \mathbb{P}_n \left(|\sum_{p \in A(n,C)} g(p) Z_p| \geq \epsilon \log\log n \right) \\
&\leq \limsup_{n\to\infty} \frac{1}{\log\log n} \log \left(\frac{(\mathrm{e}^{\theta C} - 1)}{\mathrm{e}^{\theta \epsilon \log\log n}} \sum_{p \in \mathbf{P}, k_n \leq p \leq n} \frac{1}{p} \right) \\
&= -\theta\epsilon.
\end{aligned}
$$

The claim follows now by letting $\theta \to \infty$.

Step 3. Claim: For all $\theta \in \mathbb{R}, C \in \mathbb{R}_{\geq 0}$,

$$
\lim_{n\to\infty} \frac{1}{\log\log n} \log |\mathbb{E}[\mathrm{e}^{\theta \sum_{p \in B(n,C)} g(p) Z_p}] - \mathbb{E}[\mathrm{e}^{\theta \sum_{p \in B(n,C)} g(p) Y_p}]| = -\infty,
$$

where $B(n,C) := \{p \in \mathbf{P}; |g(p)| \leq C, p \leq k_n\}$ for $n \in \mathbb{N}, C \in \mathbb{R}_{\geq 0}$.

Indeed, we firstly set

$$
S_n := \sum_{p \in B(n,C)} g(p) Z_p, \; \tilde{S}_n := \sum_{p \in B(n,C)} g(p) Y_p \text{ for } n \in \mathbb{N}
$$

for the simplification of the following expressions. Then we have for all $\theta \in \mathbb{R}, \; C \in \mathbb{R}$,

$K \in \mathbb{R}$,

$$
\begin{aligned}
&|\mathbb{E}[e^{\theta \sum_{p\in B(n,C)} g(p)Z_p}] - \mathbb{E}[e^{\theta \sum_{p\in B(n,C)} g(p)Y_p}]| \\
&= |\mathbb{E}[\sum_{r=0}^{\infty} \frac{(\theta S_n)^r}{r!}] - \mathbb{E}[\sum_{r=0}^{\infty} \frac{(\theta \tilde{S}_n)^r}{r!}]| \\
&= |\mathbb{E}[\sum_{r\le K(\log\log n)^{\frac{3}{2}}} \frac{(\theta S_n)^r}{r!} + \sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{(\theta S_n)^r}{r!}] \\
&\qquad - \mathbb{E}[\sum_{r\le K(\log\log n)^{\frac{3}{2}}} \frac{(\theta \tilde{S}_n)^r}{r!} + \sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{(\theta \tilde{S}_n)^r}{r!}]| \\
&= |\sum_{r\le K(\log\log n)^{\frac{3}{2}}} \frac{\theta^r(\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S}_n)^r])}{r!} + \mathbb{E}[\sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{\theta^r}{r!}(S_n)^r] \\
&\qquad - \mathbb{E}[\sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{\theta^r}{r!}(\tilde{S}_n)^r]| \\
&\le \sum_{r\le K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}|\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S}_n)^r]| + \mathbb{E}[\sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}(S_n)^r] \\
&\qquad + \mathbb{E}[\sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}(\tilde{S}_n)^r] \\
&= \sum_{r\le K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}|\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S}_n)^r]| + \sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}\mathbb{E}[(S_n)^r] \\
&\qquad + \sum_{r> K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!}\mathbb{E}[(\tilde{S}_n)^r], \qquad (6.1)
\end{aligned}
$$

by the monotone convergence theorem. Now for an estimation of (6.1), we firstly show

$$
|\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S}_n)^r]| \le \frac{(Ck_n)^r}{n}.
$$

For this notice that

$$
\begin{aligned}
&\mathbb{E}[(S_n)^r] = \mathbb{E}[(\sum_{p\in B(n,C)} g(p)Z_p)^r] \\
&= \sum_{k=1}^{r} \sum_{r_1+\cdots+r_k=r, r_i>0} \frac{r!}{r_1!\cdots r_k!}\cdot\frac{1}{k!} \sum_{p_1,\cdots,p_k\in B(n,C)} g(p_1)^{r_1}\cdots g(p_k)^{r_k}\mathbb{E}[Z_{p_1}^{r_1}\cdots Z_{p_k}^{r_k}],
\end{aligned}
$$

and that,

$$\mathbb{E}[(\tilde{S_n})^r] = \sum_{k=1}^{r} \sum_{r_1+\cdots+r_k=r, r_i>0} \frac{r!}{r_1!\cdots r_k!} \cdot \frac{1}{k!} \sum_{p_1,\cdots,p_k \in B(n,C)} g(p_1)^{r_1}\cdots g(p_k)^{r_k} \mathbb{E}[Y_{p_1}^{r_1}\cdots Y_{p_k}^{r_k}].$$

Also observe,

$$\mathbb{E}[Z_{p_1}^{r_1}\cdots Z_{p_k}^{r_k}] = \mathbb{E}[Z_{p_1}\cdots Z_{p_k}] = \frac{1}{n}[\frac{n}{p_1\cdots p_k}],$$
$$\mathbb{E}[Y_{p_1}^{r_1}\cdots Y_{p_k}^{r_k}] = \mathbb{E}[Y_{p_1}^{r_1}\cdots Y_{p_k}^{r_k}] = \frac{1}{p_1\cdots p_k},$$

for $r_1,\cdots r_k > 0$, and so, that $\mathbb{E}[Z_{p_1}^{r_1}\cdots Z_{p_k}^{r_k}]$ and $\mathbb{E}[Y_{p_1}^{r_1}\cdots Y_{p_k}^{r_k}]$ differ by at most $\frac{1}{n}$. Therefore, indeed,

$$|\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S_n})^r]| \leq \sum_{k=1}^{r} \sum_{r_1+\cdots+r_k=r,\ r_i>0} \frac{r!}{r_1!\cdots r_k!} \cdot \frac{1}{k!} \sum_{p_1,\cdots,p_k \in B(n,C)} \frac{(C)^r}{n}$$
$$= \frac{1}{n}(\sum_{p \in B(n,C)} C)^r \leq \frac{(Ck_n)^r}{n}.$$

Thus, we get,

$$\sum_{r \leq K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!} |\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S_n})^r]|$$
$$\leq \sum_{r \leq K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!} \cdot \frac{(Ck_n)^r}{n}$$
$$= \sum_{r \leq K(\log\log n)^{\frac{3}{2}}} \frac{1}{r!} \mathrm{e}^{r(log|\theta|+\log C+\log k_n)-\log n}$$
$$\leq C_0 \sum_{r \leq K(\log\log n)^{\frac{3}{2}}} \frac{\mathrm{e}^r}{r!} \cdot \mathrm{e}^{r(log|\theta|+\log C+\log k_n)-\log n}$$
$$\leq C_0 \mathrm{e}^{-\log n} \cdot K(\log\log n)^{\frac{3}{2}} \cdot \max_{r \leq K(\log\log n)^{\frac{3}{2}}} \mathrm{e}^{r(log|\theta|+\log C+\log k_n+1)-r\log r}.$$

Let now $F(r) := r(log|\theta| + \log C + \log k_n + 1) - r\log r$, for $r \in \mathbb{R}_{\geq 0}$. Then $F'(r) = \log|\theta| + \log C + \log k_n + 1 - \log r - 1 > 0$, for all $r \leq K(\log\log n)^{\frac{3}{2}}$, for sufficiently large n. Therefore the maximum of $F(r)$, $r \leq K(\log\log n)^{\frac{3}{2}}$ is achieved at $r = K(\log\log n)^{\frac{3}{2}}$

for such large n, and hence, now

$$\begin{aligned}
&\sum_{r \le K(\log\log n)^{\frac{3}{2}}} \frac{|\theta|^r}{r!} |\mathbb{E}[(S_n)^r] - \mathbb{E}[(\tilde{S}_n)^r]| \\
\le & C_0 \cdot \mathrm{e}^{-\log n} K(\log\log n)^{\frac{3}{2}} \cdot \mathrm{e}^{K(\log\log n)^{\frac{3}{2}}(\log|\theta| + \log C + \log k_n + 1 - \log(K(\log\log n)^{\frac{3}{2}}))} \\
\le & \mathrm{e}^{-\log n} \mathrm{e}^{C_1(\log\log n)^{\frac{3}{2}} \cdot (\frac{\log n}{(\log\log n)^2} - \frac{3}{2}\log\log\log n)} \\
\le & \mathrm{e}^{-\frac{1}{2}\log n},
\end{aligned}$$

for sufficiently large $n \in \mathbb{N}$. This was an estimation of the first term in (6.1). Notice now, that the second term in (6.1) is bounded from above by the third term there. To bound the third term, first observe

$$\mathbb{E}[(\tilde{S}_n)^r] \le C^r[(\sum_{p \in \mathbf{P}, p \le n} \frac{1}{p}) + (\sum_{p \in \mathbf{P}, p \le n} \frac{1}{p})^2 + \cdots + (\sum_{p \in \mathbf{P}, p \le n} \frac{1}{p})^r] \le rC^r(\sum_{p \in \mathbf{P}, p \le n} \frac{1}{p})^r.$$

So, for $r > K(\log\log n)^{\frac{3}{2}}$, we can estimate:

$$\begin{aligned}
\frac{|\theta|^r}{r!} \mathbb{E}[(\tilde{S}_n)^r] &\le \mathrm{e}^{r\log|\theta| - \log r! + \log r + r\log C + r\log\sum_{p \in \mathbf{P}, p \le n} \frac{1}{p}} \\
&\le \mathrm{e}^{r\log|\theta| + r - r\log r + \log r + r\log C + r \cdot \frac{3}{2}\log\log\log n} \\
&< \mathrm{e}^{r(\log|\theta| + 1 + \log C - \log K) + \log r} \\
&\le \mathrm{e}^{r(-\frac{1}{2}\log K)},
\end{aligned}$$

for all sufficiently large $K \in \mathbb{R}_{>0}$, sufficiently large $n \in \mathbb{N}$, and all corresponding $r > K(\log\log n)^{\frac{3}{2}}$. Hence, the third term in (6.1) is bounded above by

$$\sum_{r > K(\log\log n)^{\frac{3}{2}}} \mathrm{e}^{-\frac{1}{2}r\log K} \le C_2 \mathrm{e}^{-\frac{1}{2}\log K \cdot \log\log n}.$$

So, now, for sufficiently large $K \in \mathbb{R}_{>0}$, $n \in \mathbb{N}$,

$$|\mathbb{E}[\mathrm{e}^{\theta\sum_{p \in B(n,C)} g(p)Z_p}] - \mathbb{E}[\mathrm{e}^{\theta\sum_{p \in B(n,C)} g(p)Y_p}]| \le \mathrm{e}^{-\frac{1}{2}\log n} + C_3 \cdot \mathrm{e}^{-\frac{1}{2}\log K\log\log n}$$

$$\leq C_4 \cdot \mathrm{e}^{-\frac{1}{2}\log K \log\log n}.$$

And, thus,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\log\log n} \log |\mathbb{E}[\mathrm{e}^{\theta \sum_{p\in B(n,C)} g(p)Z_p}] - \mathbb{E}[\mathrm{e}^{\theta \sum_{p\in B(n,C)} g(p)Y_p}]| \\
&\leq \limsup_{n\to\infty} \frac{1}{\log\log n} \log \left(C_4 \cdot \mathrm{e}^{-\frac{1}{2}\log K \log\log n} \right) \\
&\leq C_5 \cdot \log K.
\end{aligned}$$

Letting $K \to \infty$ the claim follows.

Step 4. We had $B(n,C) := \{p \in \mathbf{P}; |g(p)| \leq C, p \leq k_n\}$. For $\theta \in \mathbb{R}$, we have

$$\log \mathbb{E}[\mathrm{e}^{\theta \sum_{p\in B(n,C)} g(p)Y_p}] = \log \prod_{p\in B(n,C)} \mathbb{E}[\mathrm{e}^{\theta g(p) Y_p}] = \sum_{p\in B(n,C)} \log(\frac{1}{p}\mathrm{e}^{\theta g(p)} + 1 - \frac{1}{p}).$$

Using the representation

$$\log(1+x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}, \text{ for } |x| < 1,$$

we see that for sufficiently large $p \in B(n,C)$, also n large accordingly,

$$\log(\frac{1}{p}\mathrm{e}^{\theta g(p)} + 1 - \frac{1}{p}) = (\mathrm{e}^{\theta g(p)} - 1)\frac{1}{p} + O(\frac{1}{p^2}),$$

and so,

$$\log \mathbb{E}[\mathrm{e}^{\theta \sum_{p\in B(n,C)} g(p)Y_p}] = \left(\sum_{p\in B(n,C)} (\mathrm{e}^{\theta g(p)} - 1)\frac{1}{p} \right) + O(1) \tag{6.2}$$

for sufficiently large $n \in \mathbb{N}$. Now noticing

$$\lim_{n\to\infty} \frac{1}{\log\log n} \sum_{p\in\mathbf{P}, p\leq n} \frac{1}{p} = 1 \text{ and } \lim_{n\to\infty} \frac{\log\log k_n}{\log\log n} = \frac{\log\log n - 2\log\log\log n}{\log\log n} = 1,$$

we have

$$\begin{aligned}
&\int_{-C}^{C}(\mathrm{e}^{\theta y}-1)\rho(\mathrm{d}y) = \lim_{n\to\infty}\int_{-C}^{C}(\mathrm{e}^{\theta y}-1)\rho_n(\mathrm{d}y)\\
&= \lim_{n\to\infty}\sum_{p\in\mathbf{P},|g(p)|<C}(\mathrm{e}^{\theta g(p)}-1)\cdot\left(\frac{\frac{1}{p}\cdot\mathbf{1}_{\{p\le n\}}}{\sum_{p\in\mathbf{P},p\le n}\frac{1}{p}}\right)\\
&= \lim_{n\to\infty}\sum_{p\in\mathbf{P},|g(p)|\le C}(\mathrm{e}^{\theta g(p)}-1)\cdot\left(\frac{\frac{1}{p}\cdot\mathbf{1}_{\{p\le n\}}}{\sum_{p\in\mathbf{P},p\le n}\frac{1}{p}}\right)\\
&= \lim_{n\to\infty}\frac{\sum_{p\in\mathbf{P},|g(p)|\le C,p\le n}(\frac{\mathrm{e}^{\theta g(p)}-1}{p})}{\sum_{p\in\mathbf{P},p\le n}\frac{1}{p}}\\
&= \lim_{n\to\infty}\frac{\sum_{p\in\mathbf{P},|g(p)|\le C,p\le k_n}(\frac{\mathrm{e}^{\theta g(p)}-1}{p})}{\sum_{p\in\mathbf{P},p\le k_n}\frac{1}{p}}\\
&= \lim_{n\to\infty}\frac{\sum_{p\in B(n,C)}(\frac{\mathrm{e}^{\theta g(p)}-1}{p})}{\log\log k_n} = \lim_{n\to\infty}\frac{\sum_{p\in B(n,C)}(\frac{\mathrm{e}^{\theta g(p)}-1}{p})}{\log\log n},
\end{aligned}$$

provided that $g(p)$ skips the values $C,-C$, using also the fact that the convergence of *moment generating functions* implies *weak convergence*. Notice here, when we shall send C to infinity, we will be sending such C's, say the C's coming from a set $\mathfrak{C}$ for sake of notation. After this calculation we have by the Equation 6.2,

$$\lim_{n\to\infty}\frac{1}{\log\log n}\log\mathbb{E}[\mathrm{e}^{\theta\sum_{p\in B(n,C)}g(p)Y_p}] = \int_{-C}^{C}(\mathrm{e}^{\theta y}-1)\rho(\mathrm{d}y)\in\mathbb{R}.$$

And, this mapping

$$\Lambda:\mathbb{R}\to\mathbb{R},\ \theta\mapsto\Lambda(\theta) := \int_{-C}^{C}(\mathrm{e}^{\theta y}-1)\rho(\mathrm{d}y)$$

is differentiable: Indeed, we have

$$\begin{aligned}
&\lim_{h\to 0^-}\int_{-C}^{C}\frac{\mathrm{e}^{(\theta+h)y}-\mathrm{e}^{\theta y}}{h}\rho(\mathrm{d}y)\\
&= \lim_{h\to 0^-}\int_{-C}^{C}\mathrm{e}^{\theta y}\left(\frac{\mathrm{e}^{hy}-1}{h}\right)\rho(\mathrm{d}y) = \int_{-C}^{C}\mathrm{e}^{\theta y}\lim_{h\to 0^-}\left(\frac{\mathrm{e}^{hy}-1}{h}\right)\rho(\mathrm{d}y)\\
&= \int_{-C}^{C}y\mathrm{e}^{\theta y}\rho(\mathrm{d}y)\in\mathbb{R},
\end{aligned}$$

for $y \geq 0$, where taking the limit inside is justified after the *monotone convergence theorem*, since $h \to \frac{e^{hy}-1}{h}$ defines an increasing function. A similar argumentation yields the same limit for $h \to 0^+$. Also the case $y \leq 0$ is analogous, and hence Λ is differentiable. We conclude after the Gärtner-Ellis theorem, that the sequence $\mathbb{P}_n\left(\frac{1}{\log\log n}\sum_{p\in B(n,C)} g(p)Y_p \in \bullet\right)$ satisfies an LDP with speed $\log\log n$ and rate

$$I_C : \mathbb{R} \to [0,\infty] : x \mapsto I_C(x) := \sup_{\theta\in\mathbb{R}}\{\theta x - \int_{-C}^{C}(e^{\theta y} - 1)\rho(dy)\}.$$

By Step 3, the sequence $\mathbb{P}_n\left(\frac{1}{\log\log n}\sum_{p\in B(n,C)} g(p)Z_p \in \bullet\right)$ satisfies the same LDP. And by Step 2, the sequence $\mathbb{P}_n\left(\frac{1}{\log\log n}\sum_{p\in \mathbf{P},|g(p)|\leq C} g(p)Z_p \in \bullet\right)$ satisfies the same LDP, since the sequences

$$\left(\frac{1}{\log\log}\sum_{p\in B(n,C)} g(p)Z_p\right)_{n\in\mathbb{N}}, \left(\frac{1}{\log\log}\sum_{p\in \mathbf{P},|g(p)|\leq C} g(p)Z_p\right)_{n\in\mathbb{N}}$$

are *exponentially equivalent* in the sense of its definition. Finally, Step 1 says precisely that the family

$$\left(\frac{1}{\log\log}\sum_{p\in \mathbf{P},|g(p)|\leq C} g(p)Z_p\right)_{n\in\mathbb{N}}, \ C \in \mathfrak{C}$$

is an *exponentially well approximation* of $\left(\frac{g(X_n)}{\log\log n}\right)_{n\in\mathbb{N}}$, and hence after Theorem 3.7, the latter sequence satisfies an LDP with speed $\log\log n$ and rate

$$\begin{aligned} I : \mathbb{R} \to [0,\infty],\ x \mapsto I(x) :&= \sup_{\delta\in\mathbb{R}_{>0}} \liminf_{C\to\infty} \inf_{x\in B_\delta(x)} \left\{\sup_{\theta\in\mathbb{R}}\{\theta x - \int_{-C}^{C}(e^{\theta y} - 1)\rho(dy)\}\right\} \\ &= \sup_{\delta\in\mathbb{R}_{>0}} \liminf_{C\to\infty} \sup_{\theta\in\mathbb{R}}\{\theta(x-\delta) - \int_{-C}^{C}(e^{\theta y} - 1)\rho(dy)\} \\ &= \sup_{\delta\in\mathbb{R}_{>0}} \sup_{\theta\in\mathbb{R}}\{\theta(x-\delta) - \int_{-\infty}^{\infty}(e^{\theta y} - 1)\rho(dy)\} \\ &= \sup_{\theta\in\mathbb{R}}\{\theta x - \int_{-\infty}^{\infty}(e^{\theta y} - 1)\rho(dy)\}. \end{aligned}$$

$\square$

REFERENCES

1. Cramér, H., "Sur un Nouveau Théorème-Limite de la Théorie des Probabilités," *Actualités Scientifiques et Industrielles*, Vol. 736, pp. 5-23, 1938.

2. Varadhan, S. R. S., "Asymptotic Probabilities and Differential Equations," *Communications on Pure and Applied Mathematics*, Vol. 19, pp. 261-286, 1966.

3. Mehrdad, B., L. Zhu, "Moderate and Large Deviations for Erdős-Kac Theorem," *arXiv*, 2013, http://arxiv.org/pdf/1311.6180v2.pdf, [Accessed August 2015].

4. Bourbaki, N., *General Topology*, Elements of Mathematics, Springer, Berlin-Heidelberg, 1989.

5. Dupuis, P., R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley, New York, 1997.

6. Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edition, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1968.

7. Seneta, E., "Markov and the Birth of Chain Dependence Theory," *International Statistical Review*, Vol. 64, No. 3, pp. 255-263, International Statistical Institute, 1996.

8. Dembo, A., O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd edition, Applications of Mathematics, Springer, Berlin-Heidelberg, 2010.

9. Gärtner, J., "On Large Deviations from the Invariant Measure," *Theory of Probability and Its Applications*, Vol. 22, No. 1, pp. 24-39, 1977.

10. den Hollander, F., *Large Deviations*, Fields Institute Monographs, American Math-

ematical Society, Toronto, 2000.

11. Sanov, I. N., "On the Probability of Large Deviations of Random Variables," *Selected Translations in Mathematical Statistics and Probability*, Vol. 1, pp. 213-244, 1961.

12. Donsker, M. D., S. R. S. Varadhan, "Asymptotic Evaluation of Certain Markov Process Expectations for Large Time," *Communications on Pure and Applied Mathematics*, Vol. 28, pp. 1-47, 1975.

13. Grimmett, G. R., D. R. Stirzaker, *Probability Theory and Random Processes*, 2nd edition, Clarendon Press, Oxford, 1995.

14. Serfozo, R., *Basics of Applied Stochastic Processes*, Probability and Its Applications, Springer, Berlin-Heidelberg, 2009.

15. Seneta, E., *Non-negative Matrices and Markov Chains*, Springer Series in Statistics, Springer, Berlin-Heidelberg, 2006.

16. Erdős, P., M. Kac, "On the Gaussian Law of Errors in the Theory of Additive Functions," *American Journal of Mathematics*, Vol. 62, pp. 738-742 , 1940.

17. Hardy, G. H., E. M. Wright, *An Introduction to the Theory of Numbers*, 5th edition, Clarendon Press, Oxford, 2003.