LARGE DEVIATIONS FOR INDEPENDENT RANDOM VARIABLES – APPLICATION TO ERDÖS-RENYI'S FUNCTIONAL LAW OF LARGE NUMBERS

JAMAL NAJIM[1]

Abstract. A Large Deviation Principle (LDP) is proved for the family $\frac{1}{n} \sum_1^n \mathbf{f}(x_i^n) \cdot Z_i^n$ where the deterministic probability measure $\frac{1}{n} \sum_1^n \delta_{x_i^n}$ converges weakly to a probability measure R and $(Z_i^n)_{i \in \mathbb{N}}$ are $\mathbb{R}^d$-valued independent random variables whose distribution depends on x_i^n and satisfies the following exponential moments condition:
$$\sup_{i,n} \mathbb{E} e^{\alpha^* |Z_i^n|} < +\infty \quad \text{for some} \quad 0 < \alpha^* < +\infty.$$
In this context, the identification of the rate function is non-trivial due to the absence of equidistribution. We rely on fine convex analysis to address this issue. Among the applications of this result, we extend Erdös and Rényi's functional law of large numbers.

Mathematics Subject Classification. 46E30, 60F10, 60G57.

1. INTRODUCTION

Let $(Z_i^n)_{i \leq n, \, n \in \mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued independent random variables and let $(x_i^n; \, 1 \leq i \leq n; \, n \geq 1)$ be a deterministic $\mathcal{X}$-valued sequence of elements satisfying

$$\frac{1}{n} \sum_1^n \delta_{x_i^n} \xrightarrow[n \to \infty]{weakly} R$$

where $\mathcal{X}$ is a topological space endowed with its Borel σ-field. In this article, we investigate the Large Deviations (LD) of the empirical mean

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_1^n \mathbf{f}(x_i^n) \cdot Z_i^n$$

where

$$\frac{1}{n} \sum_1^n \mathbf{f}(x_i^n) \cdot Z_i^n = \frac{1}{n} \sum_1^n \begin{pmatrix} f_1(x_i^n) \cdot Z_i^n \\ \vdots \\ f_m(x_i^n) \cdot Z_i^n \end{pmatrix},$$

---

[1] CNRS, École Nationale Supérieure des Télécommunications, 46 rue Barrault 75634 Paris Cedex 13, France; `najim@tsi.enst.fr`

the f_k's being the rows of the $m \times d$ matrix $\mathbf{f}$. Here, each f_k is a bounded continuous function from $\mathcal{X}$ to $\mathbb{R}^d$ and $\cdot$ denotes the scalar product in $\mathbb{R}^d$. In particular the random variable $\langle L_n, \mathbf{f} \rangle$ is $\mathbb{R}^m$-valued.

The random variables Z_i^n are independent and satisfy the following exponential moment condition:

$$\sup_{i,n} \mathbb{E}\, \mathrm{e}^{\alpha^* |Z_i^n|} < \infty \quad \text{for some} \quad \alpha^* > 0. \tag{1.1}$$

The law of Z_i^n (denoted by $\mathcal{L}(Z_i^n)$) depends on x_i^n in the sens that if x_i^n and x_j^n are close, then so are $\mathcal{L}(Z_i^n)$ and $\mathcal{L}(Z_j^n)$ for the following distance between probability measures:

$$d_{OW}(P,Q) = \inf_{\eta} \inf \left\{ a > 0;\, \int_{\mathbb{R}^d \times \mathbb{R}^d} \tau \left(\frac{z - z'}{a} \right)\, \eta(\mathrm{d}z \mathrm{d}z') \leq 1 \right\},$$

where η is a probability with given marginals P and Q and $\tau(z) = \mathrm{e}^{|z|} - 1$. This distance allows us to quantify easily the "exponential proximity" of two probability measures.

The large deviation principle

The main result of the article is the Large Deviations Principle (LDP) established for $\langle L_n, \mathbf{f} \rangle$ under the previously mentioned assumptions: Cramér's condition (1.1) and variables Z_i^n which are independent but not identically distributed.

The Large Deviations associated to this model are of interest in gaz theory and have been studied in the one-dimensional case by Ellis, Gough and Pulé in [15] in the case where $\mathcal{L}(Z_i^n) \propto \mathrm{e}^{g(x_i^n) \cdot z} P(\mathrm{d}z)$. Related LDPs have been established by Dembo and Zajic [10] and Gamboa and Gassiat in [17]. In both cases [15, 17], it is important to consider random variables which do not have all their exponential moments (1.1). This model is also of interest in information theory where $Z_i^n = f(x_i^n, Y_i^n)$, the Y_i^n's being i.i.d., and have been studied by Dembo and Kontoyiannis [9] and Chi [3, 4] in the case where the process Y satisfies some mixing conditions.

There are two key issues to establish the LDP.

1. Since no steepness is assumed, we are not able to use Gärtner-Ellis' theorem nor to adapt in a different way Cramér's exponential change of measure argument. Our main tools to overcome it are the technique developed in [22] in the i.i.d. setting and a coupling argument based on the distance d_{OW}.
2. The identification of the rate function is non trivial. In fact, the technique developed in [22] relies heavily on exponential approximation and thus yields to a very abstract rate function. The model being more complex here, so is the identification of the rate function. Convex Analysis provides us with very efficient tools to address this point.

Let us mention that this kind of problems often arises when dealing with LD of independent but not identically distributed random variables (Dawson and Gärtner [6, 7], Djellout, Guillin and Wu [13]). In a different fashion, let us mention the work of Zabell [29] around Mosco-convergence and Large Deviations.

Applications of the LDP

LDPs in infinite dimension. The LDP of $\langle L_n, \mathbf{f} \rangle$ is the cornerstone to get LDPs in infinite dimensional settings *via* Dawson and Gärtner's projective limit approach. We derive LDPs for the empirical measure L_n and the random walk $\bar{Z}_n$ where

$$L_n = \frac{1}{n} \sum_{i=1}^{n} Z_i^n \delta_{x_i^n} \quad \text{and} \quad \bar{Z}_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} Z_i^n,\ t \in [0,1].$$

These results extend those of [22] to the case where the random variables Z_i^n are not identically distributed and yield to rate functions with an extra term (see Ths. 4.1 and 4.3) à la Lynch and Sethuraman [21]. For instance,

the rate function of the LDP satisfied by L_n is given by:

$$I(\mu) = I_1(\mu_a) + I_2(\mu_s),$$

where $\mu = \mu_a + \mu_s$ is the Lebesgue decomposition of μ with respect to R. In particular, $I(\mu)$ can be finite for measures μ which are not absolutely continuous with respect to R. One shall notice that the LDP for $\bar{Z}_n$ has been studied by Schuette in [28] under different conditions.

Erdös and Rényi's functional law of large numbers (FLLN). Assume that the Z_i^n's are as previously, that $\mathbb{E}\, Z_i^n = 0$ and consider the process

$$\eta_{n,m} : t \mapsto \frac{1}{A_n} \sum_{i=m}^{[m+tA_n]} Z_i, \ t \in [0,1]$$

where $A_n = [\log n]$ ($[x]$ denotes the integer part of x) and m lies between 1 and $n - A_n$. Our main result is the description of the cluster points of the sets $K_n = \{\eta_{n,m}, \ 1 \leq m \leq n - A_n\}$. This problem is known as Erdös-Rényi's FLLN and has been studied by several authors among which Deheuvels [8], Borovkov [2], Sanchis [26, 27] and Gantert [18]. Let us first remind the crucial observation of Erdös and Rényi. Let

$$U_n = \max_{0 \leq m \leq n - A_n} \frac{Z_{m+1} + \cdots + Z_{m+A_n}}{A_n}.$$

If $A_n = n$, then the limit of U_n is a.s. zero (law of large numbers). If A_n is finite, say $A_n = 1$, then the limit is infinite (provided that the Z's are not bounded). Erdös and Rényi [16] have shown that for a scaling in-between, namely $A_n = [\log n]$, then the limit of U_n exists, is non trivial and might by expressed by means of the rate function associated to the LDP of the empirical mean (Cramér's theorem).

We describe the set of cluster points by means of the rate function associated to the LDP of $\bar{Z}_n$. Our results extend Deheuvels' result to the multidimensional setting and extends Borovkov's result to the case where the random variables Z_i do not have all their exponential moments. We also relax the i.i.d. assumption.

Outline of the paper

The paper is organized as follows: in Section 2, the notations and the assumptions are given, the distance d_{OW} is defined and examples fulfilling the assumptions are studied (Sect. 2.4); the LDP together with the identification of the rate function are stated and proved in Section 3. Large Deviations for the empirical measure L_n and the random walk $\bar{Z}_n$ are proved in Section 4. Finally, Section 5 is devoted to Erdös and Rényi's functional law of large numbers.

2. Notations and assumptions

2.1. **Notations**

Let $\mathcal{X}$ be a topological vector space endowed with its Borel σ-field $\mathcal{B}(\mathcal{X})$ and let R be a probability measure on $\mathcal{X}$. We denote by $C(\mathcal{X})$ (resp. $C_d(\mathcal{X})$) the set of $\mathbb{R}$-valued (resp. $\mathbb{R}^d$-valued) continuous bounded functions on $\mathcal{X}$; by $L_d^1(\mathcal{X})$ the set of $\mathbb{R}^d$-valued R-integrable functions on $\mathcal{X}$ and by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. We shall sometimes drop $\mathcal{X}$ and denote the previous sets by C, C_d, L_d^1.

Let $|\cdot|$ denote a norm on any finite-dimensional vector space (usually $\mathbb{R}$, $\mathbb{R}^d$ or $\mathbb{R}^{m \times d}$). We denote by $\|\cdot\|$ the supremum norm on the space of bounded continuous functions from $\mathcal{X}$ with values in $\mathbb{R}$, $\mathbb{R}^d$ or $\mathbb{R}^{m \times d}$, *i.e.* $\|f\| = \sup_{x \in \mathcal{X}} |f(x)|$. As usual, δ_a is the Dirac measure at a. Let $\mathbf{a}$ be a $m \times d$ matrix and let $z \in \mathbb{R}^d$.

We denote by $\cdot$ the usual matrix product, that is

$$\mathbf{a} \cdot z = \begin{pmatrix} a_1 \cdot z \\ \vdots \\ a_m \cdot z \end{pmatrix}$$

where a_j is the j-th row of the matrix $\mathbf{a}$. Hence, $\cdot$ denotes the scalar product $\lambda \cdot z$ or the matrix product $\mathbf{a} \cdot z$, depending on the context. Let $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ be a (matrix-valued) continuous bounded function and $\theta \in \mathbb{R}^m$, then

$$\mathbf{f}(x) \cdot z = \begin{pmatrix} f_1(x) \cdot z \\ \vdots \\ f_m(x) \cdot z \end{pmatrix} \quad \text{and} \quad \theta \cdot \mathbf{f}(x) = \sum_{j=1}^m \theta_j \, f_j(x)$$

where each $f_j \in C_d(\mathcal{X})$ is the j-th row of the matrix $\mathbf{f}$. Let $u : \mathcal{X} \to \mathbb{R}^d$ be a measurable function, we denote by

$$\int_{\mathcal{X}} \mathbf{f}(x) \cdot u(x) \, R(\mathrm{d}x) = \begin{pmatrix} \int_{\mathcal{X}} f_1(x) \cdot u(x) \, R(\mathrm{d}x) \\ \vdots \\ \int_{\mathcal{X}} f_m(x) \cdot u(x) \, R(\mathrm{d}x) \end{pmatrix}.$$

In the sequel, we shall follow the convention that $x \in \mathcal{X}$, y and θ are elements of $\mathbb{R}^m$ and z and λ, of $\mathbb{R}^d$.

We will denote by $\bar{A}$ (resp. $\mathrm{int}(A)$) the closure (resp. the interior) of the set A.

2.2. The Orlicz-Wasserstein distance

We introduce here a distance on the space of probability measures having some exponential moments:

$$\int_{\mathbb{R}^d} \mathrm{e}^{\alpha |z|} P(\mathrm{d}z) < \infty \quad \text{for some } \alpha > 0.$$

We call it the Orlicz-Wasserstein distance. This distance appears to be very useful to quantify the "exponential proximity" of two probability measures and is described below. It is then natural to ask to the distributions of the family (Z_i^n) to change continuously (with respect to x_i^n) with respect to this distance.

Let $\tau(z) = \mathrm{e}^{|z|} - 1$, $z \in \mathbb{R}^d$ and let us consider

$$\mathcal{P}_\tau(\mathbb{R}^d) = \left\{ P \in \mathcal{P}(\mathbb{R}^d), \ \exists \, a > 0 \ \int_{\mathbb{R}^d} \tau\left(\frac{z}{a}\right) P(\mathrm{d}z) < \infty \right\}$$

$$= \left\{ P \in \mathcal{P}(\mathbb{R}^d), \ \exists \, \alpha > 0 \ \int_{\mathbb{R}^d} \mathrm{e}^{\alpha |z|} P(\mathrm{d}z) < \infty \right\}$$

$\mathcal{P}_\tau$ is the set of probability distributions having some exponential moments. We denote by $M(P, Q)$ the set of all laws on $\mathbb{R}^d \times \mathbb{R}^d$ with given marginals P and Q. Consider the following Orlicz-Wasserstein distance defined on $\mathcal{P}_\tau(\mathbb{R}^d)$ by:

$$d_{OW}(P, Q) = \inf_{\eta \in M(P,Q)} \inf \left\{ a > 0; \int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{z - z'}{a}\right) \eta(\mathrm{d}z \mathrm{d}z') \leq 1 \right\}.$$

Lemma 2.1. *d_{OW} is a distance on $\mathcal{P}_\tau(\mathbb{R}^d)$.*

Remark 2.1. Let us compare d_{OW} with the usual Wasserstein distance:

$$W(P, Q) = \inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |z - z'| \eta(\mathrm{d}z \mathrm{d}z'), \ \eta \in M(P, Q) \right\}.$$

Then, $W(P, Q) \leq d_{OW}(P, Q)$. In fact, $|z| \leq \tau(z)$ therefore

$$\int \tau \left(\frac{z - z'}{a} \right) \eta(\mathrm{d}z\mathrm{d}z') \leq 1 \Rightarrow \int |z - z'| \eta(\mathrm{d}z\mathrm{d}z') \leq a.$$

The proof of this lemma follows closely the standard proof of Dudley ([14], Lem. 11.8.3) and is omitted.

In the following, we shall endow $\mathcal{P}_\tau(\mathbb{R}^d)$ with the topology induced by d_{OW}.

2.3. Assumptions

Let us now introduce some assumptions on the model:

Assumption A-1. *R is a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfying:*

$$\textit{if } U \textit{ is a nonempty open set, then} \quad R(U) > 0.$$

Assumption A-2. *The family $(x_i^n)_{1 \leq i \leq n, \ 1 \leq n} \subset \mathcal{X}$ satisfies*

$$\frac{1}{n} \sum_1^n \delta_{x_i^n} \xrightarrow[n \to \infty]{weakly} R. \tag{2.1}$$

Remark 2.2. The combination of Assumptions (A-1) and (A-2) is standard (see [1,15,17]). It has been shown in [22] that the LDP for $\langle L_n, \mathbf{f} \rangle$ might fail to hold if Assumption (A-1) is not fulfilled.

Assumption A-3. *$\mathcal{X}$ is a compact space.*

Assumption A-4. *Let (Z_i^n) be a sequence of independent and $\mathbb{R}^d$-valued random variables. There exists a family of probability measures $(P(x, \cdot), x \in \mathcal{X})$ over $\mathbb{R}^d$ and a sequence $(x_i^n; 1 \leq i \leq n; 1 \leq n)$ with values in $\mathcal{X}$ such that the law of each Z_i^n is given by:*

$$\mathcal{L}(Z_i^n) \sim P(x_i^n, \mathrm{d}z).$$

We shall call $(P(x, \cdot), x \in \mathcal{X})$ the distribution kernel *associated to the family (Z_i^n). We will equally write $P(x, \cdot)$, P_x or $P_x(\mathrm{d}z)$.*

We can associate to each kernel $(P(x, \cdot), x \in \mathcal{X})$ a cumulant generating function defined by

$$\Lambda(x, \lambda) = \log \int_{\mathbb{R}^d} \mathrm{e}^{\lambda \cdot z} P(x, \mathrm{d}z) \quad \text{where} \quad \lambda \in \mathbb{R}^d,$$

and its convex conjugate

$$\Lambda^*(x, z) = \sup_{\lambda \in \mathbb{R}^d} \{ \lambda \cdot z - \Lambda(x, \lambda) \} \quad \text{where} \quad z \in \mathbb{R}^d.$$

Assumption A-5. *Let $(P(x, \cdot); x \in \mathcal{X}) \subset \mathcal{P}_\tau(\mathbb{R}^d)$ be a given kernel. The application $x \mapsto P(x, A)$ is measurable whenever the set $A \subset \mathbb{R}^d$ is borel. Moreover, the function*

$$\Gamma : \mathcal{X} \to \mathcal{P}_\tau(\mathbb{R}^d)$$
$$x \mapsto P(x, \cdot)$$

is continuous when $\mathcal{P}_\tau(\mathbb{R}^d)$ is endowed with the topology induced by the distance d_{OW}.

Remark 2.3. One shall notice that the two assumptions (A-3) and (A-5) yield the existence of a real number $\alpha^* > 0$ such that

$$S^* = \sup_{x \in \mathcal{X}} \int_{\mathbb{R}^d} \mathrm{e}^{\alpha^*|z|} P_x(\mathrm{d}z) < \infty. \tag{2.2}$$

In fact, $x \mapsto d_{OW}(P_x, P_{x_0})$ is a continuous application and is therefore bounded on the compact set $\mathcal{X}$, that is $d_{OW}(P_x, P_{x_0}) \leq a$ for every $x \in \mathcal{X}$. In particular, for all $x \in \mathcal{X}$, there exists a probability $\eta_x \in M(P_x, P_{x_0})$ such that

$$\int \mathrm{e}^{\frac{|z-z'|}{a}} \eta_x(\mathrm{d}z\mathrm{d}z') \leq 2.$$

Let $\alpha + \beta = 1$. Choose b large enough so that $\alpha\, b > a$ and $\int \mathrm{e}^{\frac{|z'|}{\beta b}} P_{x_0}(\mathrm{d}z') < \infty$ are satisfied. Then

$$\begin{aligned}
\int \mathrm{e}^{\frac{|z|}{b}} P_x(\mathrm{d}z) &= \int \mathrm{e}^{\frac{|z-z'+z'|}{b}} \eta_x(\mathrm{d}z\mathrm{d}z') \\
&\leq \left[\int \mathrm{e}^{\frac{|z-z'|}{\alpha b}} \eta_x(\mathrm{d}z\mathrm{d}z') \right]^\alpha \left[\int \mathrm{e}^{\frac{|z'|}{\beta b}} P_{x_0}(\mathrm{d}z') \right]^\beta \\
&\leq 2^\alpha \left[\int \mathrm{e}^{\frac{|z'|}{a}} P_{x_0}(\mathrm{d}z') \right]^\beta < \infty
\end{aligned}$$

and (2.2) is established by taking $\alpha^* = b^{-1}$.

Assumption (A-5) is probably the harder to check. Once (A-5) is fulfilled by a family of kernels $(P(x, \cdot))_x$, it is easy to build independent random variables whose law is given by $\mathcal{L}(Z_i^n) \sim P(x_i^n, \mathrm{d}z)$. In Section 2.4, Assumption (A-5) is proved to hold for the examples mentioned in the introduction.

2.4. Examples of families satisfying Assumption (A-5)

2.4.1. *A family defined by its densities with respect to some probability measure P*

Let P be such that

$$\int_{\mathbb{R}^d} \mathrm{e}^{\alpha^*|z|} P(\mathrm{d}z) < \infty \quad \text{for some } \alpha^* > 0, \tag{2.3}$$

and consider the family of probability distributions P_x defined by

$$\frac{\mathrm{d}P_x}{\mathrm{d}P}(z) = \frac{\mathrm{e}^{g(x)\cdot z}}{\int_{\mathbb{R}^d} \mathrm{e}^{g(x)\cdot z} P(\mathrm{d}z)},$$

where g is a bounded continuous function from the compact set $\mathcal{X} \subset \mathbb{R}^k$ into $\mathbb{R}^d$. We will assume that $P(\mathrm{d}z) = f(z)\,\mathrm{d}z$ where $\mathrm{d}z$ stands for the Lebesgue measure on $\mathbb{R}^d$ and that $f(z) = f(|z|)$, *i.e.* f is radial. We also assume:

$$0 < |g(z)| = \lambda < \lambda^*. \tag{2.4}$$

Under these assumptions, $\int_{\mathbb{R}^d} \mathrm{e}^{g(x)\cdot z} P(\mathrm{d}z)$ does not depend on x. In order to prove that $\lim_{x \to x_0} d_{OW}(P_x, P_{x_0}) = 0$, it is sufficient to prove that for every $\epsilon > 0$, there exists $\delta > 0$ and a joint probability η such that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{e}^{\frac{|z-z'|}{\epsilon}} \eta(\mathrm{d}z, \mathrm{d}z') \leq 2$$

where $|x - x_0| < \delta$ and $\eta \in M(P_{x_0}, P_x)$. By (2.4), there exists an orthogonal matrix O_x such that:

$$O_x g(x_0) = g(x) \quad \text{and} \quad \lim_{x \to x_0} |I - O_x| = 0, \tag{2.5}$$

where I denotes the identity matrix. Consider now the joint probability measure η associated to the random variable $(Z, O_x Z)$ where

$$\mathcal{L}(Z) \sim \frac{\mathrm{e}^{g(x_0) \cdot z}}{\int_{\mathbb{R}^d} \mathrm{e}^{g(x_0) \cdot z} P(\mathrm{d}z)} P(\mathrm{d}z).$$

Then, $\eta \in M(P_{x_0}, P_x)$. Let $A(x)$ be defined by

$$A(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{e}^{\frac{|z - z'|}{\epsilon}} \eta(\mathrm{d}z, \mathrm{d}z') = \int_{\mathbb{R}^d} \mathrm{e}^{\frac{|(I - O_x) \cdot z|}{\epsilon}} P_{x_0}(\mathrm{d}z).$$

Notice that $A(x_0) = 1$. The Dominated Convergence theorem together with (2.4) and (2.5) yields then that

$$\lim_{x \to x_0} A(x) = 1.$$

In particular, if $x - x_0$ is small enough say $|x - x_0| < \delta$ then $A(x) \leq 2$.

2.4.2. *Twisted i.i.d. random variables*

Let Z be a $\mathbb{R}^d$-valued random variable with distribution Q. We introduce the following Orlicz space (recall that $\tau(z) = \mathrm{e}^{|z|} - 1$):

$$L_\tau = \left\{ f : \mathbb{R}^d \to \mathbb{R}^d; \ \exists a > 0, \ \int_{\mathbb{R}^d} \tau\left(\frac{f(z)}{a}\right) Q(\mathrm{d}z) < \infty \right\}.$$

We endow it with the norm

$$\|f\|_\tau = \inf\left\{ a > 0; \ \int \tau\left(\frac{f}{a}\right) \mathrm{d}Q \leq 1 \right\}.$$

Then $(L_\tau, \|\cdot\|_\tau)$ is a Banach space. Note that

$$f \in L_\tau \Leftrightarrow \ \exists \alpha > 0, \ E\mathrm{e}^{\alpha |f(Z)|} < \infty.$$

Let $\mathcal{X}$ be a compact metric space and consider $C(\mathcal{X}; L_\tau)$, the set of continuous functions from $\mathcal{X}$ to L_τ. The fact that $\phi \in C(\mathcal{X}; L_\tau)$ implies two things:

- $\phi(x, \cdot) \in L_\tau \quad \forall x \in \mathcal{X}$;
- $x \mapsto \phi(x, \cdot)$ is continuous with respect to the norm $\|\cdot\|_\tau$.

Denote by P_x the distribution of $\phi(x, Z)$. Then the family $(P_x)_{x \in \mathcal{X}}$ satisfies (A-5). In fact, consider the random variables $Y = \phi(x, Z)$ and $\tilde{Y} = \phi(x', Z)$. Since ϕ is continuous,

$$E\tau\left(\frac{|Y - \tilde{Y}|}{\epsilon}\right) = E\tau\left(\frac{|\phi(x, Z) - \phi(x', Z)|}{\epsilon}\right) \leq 1$$

for x' close to x. Thus $d_{OW}(P_x, P_{x'}) \leq \epsilon$.

Here is an example of a family for which Gärtner-Ellis' theorem might not apply: let $g : \mathbb{R}^d \to \mathbb{R}$ and $f : \mathcal{X} \to \mathbb{R}$ ($\mathcal{X}$ compact) be bounded and continuous and let Z be a random variable with values in $\mathbb{R}^d$ having some exponential moments:

$$\exists \alpha > 0, \quad E\mathrm{e}^{\alpha |Z|} < \infty.$$

Then the family of laws associated to $Z_x = \frac{Z}{1 + |f(x)u(Z)|}$ satisfies (A-5).

2.4.3. *Csiszàr's example*

We study here a family of probability measures with non steep logarithmic moment generating functions. Consider the probability distribution P_a associated with the following distribution function:

$$F_a(x) = 1 - \frac{\mathrm{e}^{-ax}}{(1+x)^4}, \; x \geq 0 \quad \text{where} \quad a \in [\alpha, \beta] \subset (0, \infty).$$

This kind of distribution has been introduced by Csiszàr [5] to emphasize some side effects in conditional limit theorems (see also Léonard and Najim [20] for further information).

It is straightforward to check that $\Lambda(\lambda) = \log \int_0^\infty \mathrm{e}^{\lambda x} P_a(\mathrm{d}x)$ is not steep for each value of $a \in [\alpha, \beta]$. We shall prove here that Assumption (A-5) holds in this situation. In other words if $a_0 \in [\alpha, \beta]$ and $[\alpha, \beta] \ni a \to a_0$, then

$$\lim_{a \to a_0} d_{OW}(P_a, P_{a_0}) = 0.$$

Let F_a^{-1} be defined by

$$1 - \frac{\mathrm{e}^{-aF_a^{-1}(u)}}{(1 + F_a^{-1}(u))^4} = u, \quad u \in [0, 1). \tag{2.6}$$

It is straightforward to check that

$$F_\beta^{-1}(u) \leq F_a^{-1}(u) \leq F_\alpha^{-1}(u), \quad \forall a \in [\alpha, \beta]. \tag{2.7}$$

Consider now the joint probability distribution η associated to the random variable $(F_a^{-1}(U), F_{a_0}^{-1}(U))$ where U is uniformily distributed over $[0, 1]$. Then η's marginals are P_a and P_{a_0}. From (2.6), we deduce that

$$F_a^{-1}(u) = -\frac{1}{a} \log(1-u) - \frac{4}{a} \log(1 + F_a^{-1}(u)) \leq -\frac{1}{a} \log(1-u). \tag{2.8}$$

On the other hand, (2.7) yields

$$F_a^{-1}(u) \geq -\frac{1}{a} \log(1-u) - \frac{4}{a} \log(1 + F_\alpha^{-1}(u)) = \frac{\alpha}{a} F_\alpha^{-1}(u) \tag{2.9}$$

where the last equality follows from (2.8). Standard considerations about implicit functions (see for example Dieudonné [12], Chap. 3) yield that

$$F_\alpha^{-1}(u) = -\frac{1}{\alpha} \log(1-u) - \epsilon(\alpha, u) \log(1-u) \quad \text{where} \quad \epsilon(\alpha, u) \xrightarrow[u \to 1]{} 0. \tag{2.10}$$

In particular, equations (2.8) and (2.9) together with (2.10) yield

$$\left| F_a^{-1}(u) + \frac{1}{a} \log(1-u) \right| \leq \left| \frac{\alpha}{a} \epsilon(\alpha, u) \log(1-u) \right|. \tag{2.11}$$

Notice that $\epsilon(\alpha, u)$ does not depend on a.

Let us prove now that for every $\epsilon > 0$, there exists $\delta^* > 0$ such that $|a - a_0| \leq \delta^*$ implies that

$$\int_{\mathbb{R}^+ \times \mathbb{R}^+} \mathrm{e}^{\frac{|z-z'|}{\epsilon}} \eta(\mathrm{d}z, \mathrm{d}z') = \int_0^1 \mathrm{e}^{\frac{|F_a^{-1}(u) - F_{a_0}^{-1}(u)|}{\epsilon}} \mathrm{d}u \leq 2.$$

First choose u_0 close enough to 1 such that $\int_{u_0}^{1} e^{\frac{|F_a^{-1}(u) - F_{a_0}^{-1}(u)|}{\epsilon}} \, du \le 1$ whenever $a - a_0$ is small enough, say $|a - a_0| < \delta$ (use (2.11)). Then choose $a - a_0$ smaller enough, say $|a - a_0| < \delta^* < \delta$, such that

$$\int_{0}^{u_0} e^{\frac{|F_a^{-1}(u) - F_{a_0}^{-1}(u)|}{\epsilon}} \, du \le 1.$$

This is possible *via* the Dominated Convergence theorem since $\lim_{a \to a_0} F_a^{-1}(u) = F_{a_0}^{-1}(u)$ by the Implicit Function theorem and $|F_a^{-1}(u) - F_{a_0}^{-1}(u)| \le 2F_\alpha^{-1}(u)$ which is bounded over $[0, u_0]$.

3. The large deviation principle

We can now state the Large Deviation Principle.

Theorem 3.1 (the LDP). *Let $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ be a bounded continuous function. Assume that (A-1), (A-2), (A-3), (A-4) and (A-5) hold. Then, the family*

$$\langle L_n, \mathbf{f} \rangle = \frac{1}{n} \sum_{1}^{n} \mathbf{f}(x_i^n) \cdot Z_i^n$$

satisfies the LDP in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ with some good rate function Υ.

One can have a look at (3.8) to get more insight on the rate function Υ.

Theorem 3.2 (identification of the rate function). *Under the assumptions of Theorem 3.1, let $(P_x)_{x \in \mathcal{X}}$ be the distribution kernel associated to the family (Z_i^n) and let $I_{\mathbf{f}}$ be defined by*

$$I_{\mathbf{f}}(y) = \sup_{\theta \in \mathbb{R}^m} \left\{ \theta \cdot y - \int_{\mathcal{X}} \Lambda(x, \theta \cdot \mathbf{f}(x)) \, R(\mathrm{d}x) \right\}, \quad y \in \mathbb{R}^m,$$

where $\Lambda(x, \lambda)$ is the cumulant generating function associated to (P_x). Then

$$\Upsilon = I_{\mathbf{f}}.$$

3.1. **Proof of the LDP**

The proof follows closely proof of Theorem 2.2 in [22]. Let us briefly outline it in the real case. Consider

$$\frac{1}{n} \sum_{i=1}^{n} f(x_i^n) Z_i^n$$

where f and the Z_i^n's are real-valued and satisfy the assumptions of Theorem 3.1. Let $\tilde{f}(x) = \sum_{k=1}^{p} a_k 1_{A_k}(x)$ be a step function and let $(P_k)_{1 \le k \le p}$ be p probability measures with some exponential moments. Assume that there are independent random variables $\tilde{Z}_i^n$ such that if $x_i^n \in A_k$ then $\mathcal{L}(Z_i^n) = P_k$. Under these conditions, one can easily establish a LDP for

$$\frac{1}{n} \sum_{i=1}^{n} \tilde{f}(x_i^n) \tilde{Z}_i^n = \frac{a_1}{n} \sum_{j_1 \in \{i : x_i^n \in A_1\}} \tilde{Z}_{j_1}^n + \cdots + \frac{a_p}{n} \sum_{j_p \in \{i : x_i^n \in A_p\}} \tilde{Z}_{j_p}^n$$

via Cramér's theorem (this is the content of Lem. 3.3). The whole point of the proof of Theorem 3.1 is to build an exponential approximation of $\frac{1}{n}\sum_{i=1}^{n}f(x_i^n)Z_i^n$ based on simpler objects $\frac{1}{n}\sum_{i=1}^{n}\tilde{f}(x_i^n)\tilde{Z}_i^n$ (this is the content of Lem. 3.4) where

(1) the step function $\tilde{f}$ approximates f for the sup-norm $\|\tilde{f}-f\|\leq\epsilon$;
(2) the random variable $\tilde{Z}_i^n$ approximates Z_i^n with respect to the distance d_{OW}:

$$d_{OW}(\mathcal{L}(Z_i^n),\mathcal{L}(\tilde{Z}_i^n))\leq\epsilon. \tag{3.1}$$

One should notice that in [22], step (2) is not necessary since the variables are i.i.d.

Lemma 3.3. *Assume that (A-4) holds with the following kernel:*

$$P(x,\,\mathrm{d}z)=\sum_{k=1}^{p}\mathbf{1}_{A_k}(x)\,P_k(\mathrm{d}z)$$

where $P_k\in\mathcal{P}_\tau(\mathbb{R}^d)$ and $(A_k;1\leq k\leq p)$ is a measurable partition of $\mathcal{X}$ (in particular, $\mathcal{L}(Z_i^n)\sim P_k(\mathrm{d}z)$ if $x_i^n\in A_k$).

Assume moreover that $R(A_k)>0$ and $R(\partial A_k)=0$ where $\partial A_k=\bar{A}_k\setminus\mathrm{int}(A_k)$ and consider the (matrix valued) step function:

$$\mathbf{f}:\mathcal{X}\longrightarrow\mathbb{R}^{m\times d}$$
$$x\longmapsto\mathbf{f}(x)=\sum_{k=1}^{p}\mathbf{a}_k\mathbf{1}_{A_k}(x)$$

where $\mathbf{a}_k$ are $m\times d$ matrices. Then,

$$\langle L_n,\mathbf{f}\rangle=\frac{1}{n}\sum_{1}^{n}\mathbf{f}(x_i^n)\cdot Z_i^n$$
$$=\frac{1}{n}\sum_{j_1\in\{i:x_i^n\in A_1\}}\mathbf{a}_1\cdot Z_{j_1}^n+\cdots+\frac{1}{n}\sum_{j_p\in\{i:x_i^n\in A_p\}}\mathbf{a}_p\cdot Z_{j_1}^n$$

satisfies the LDP in $(\mathbb{R}^m,\mathcal{B}(\mathbb{R}^m))$ with the good rate function

$$I_{\mathbf{f}}(y)=\sup_{\theta\in\mathbb{R}^m}\left\{y\cdot\theta-\int_{\mathcal{X}}\Lambda\left(x,\theta\cdot\mathbf{f}(x)\right)R(\mathrm{d}x)\right\},\quad y\in\mathbb{R}^m \tag{3.2}$$

where $\Lambda(x,\cdot)$ is the cumulant generating function associated to $(P_x)_{x\in\mathcal{X}}$.

Remark 3.1 (on the assumptions $R(A_k)>0$ and $R(\partial A_k)=0$). In [22], a probability measure P with some exponential moments is built such that

(1) if Z is a random variable with $\mathcal{L}(Z)=P$, then the LDP does not hold for $\frac{Z}{n}$;
(2) if Z_i are i.i.d. P-distributed then the LDP does not hold for $\frac{1}{n}\sum_{1}^{n-1}Z_i-\frac{Z_n}{n}$;
(3) however, if $\frac{N(n)}{n}\to c>0$, then $\frac{1}{n}\sum_{1}^{N(n)}Z_i$ satisfies a LDP.

Assumptions $R(A_k) > 0$ and $R(\partial A_k) = 0$ are made in order to avoid the effects (1) and (2). Denote by $N_{A_k}(n) = \sharp\{i :, x_i^n \in A_k\}$, then $\frac{N_{A_k}(n)}{n} \to R(A_k)$ and the LDP holds for

$$\frac{1}{n} \sum_{j_k \in \{i:, x_i^n \in A_k\}} \mathbf{a}_k \cdot Z_{j_k}^n$$

by virtue of (3).

Proof of Lemma 3.3. The proof is a direct adaptation of Lemmas 5.2 and 5.6 in [22]. The LDP is established *via* Cramér's theorem (Lem. 5.2) and the identification of the rate function relies on duality arguments (Lem. 5.6). □

Lemma 3.4. *Assume that (A-1), (A-2), (A-3) and (A-5) hold and let $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ be a bounded continuous function.*

(i) *There exists a family of random variables $(Z_i^n)_{i,n}$ with distribution kernel (P_x) given by (A-5) which satisfy (A-4).*

(ii) *There exists a family of independent random variables $(Z_i^{n,\epsilon})_{i,n}$ associated to a kernel $(P_x^\epsilon)_{x \in \mathcal{X}}$ and a step function $\mathbf{f}^\epsilon$ such that*

$$\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle = \frac{1}{n} \sum_{i=1}^n \mathbf{f}^\epsilon(x_i^n) \cdot Z_i^{n,\epsilon}$$

satisfies the LDP with good rate function given by

$$I_{\mathbf{f}^\epsilon}(y) = \sup_{\theta \in \mathbb{R}^m} \left\{ y \cdot \theta - \int_{\mathcal{X}} \Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) R(\mathrm{d}x) \right\}, \quad y \in \mathbb{R}^m$$

where $\Lambda^\epsilon(x, \cdot)$ is associated to the kernel (P_x^ϵ).

(iii) *The family $\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle$ is an exponential approximation of $\langle L_n, \mathbf{f} \rangle$, i.e.:*

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log P \left\{ |\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle - \langle L_n, \mathbf{f} \rangle| > \delta \right\} = -\infty \quad \text{for every} \quad \delta > 0.$$

(iv) *The random variable $\langle L_n, \mathbf{f} \rangle$ satisfies the full LDP with good rate function*

$$\begin{aligned}
\Upsilon(y) &= \sup_{\eta > 0} \liminf_{\epsilon \to 0} \inf_{y' \in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y'), \\
&= \sup_{\eta > 0} \limsup_{\epsilon \to 0} \inf_{y' \in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y'), \quad y \in \mathbb{R}^m
\end{aligned}$$

where $B(y, \epsilon) = \{y' \in \mathbb{R}^m, |y' - y| < \epsilon\}$.

Proof.

● *The step function $\mathbf{f}^\epsilon$ and the kernel P^ϵ.* In the sequel, we shall build a step function $\mathbf{f}^\epsilon$ and a probability kernel P^ϵ

$$\mathbf{f}^\epsilon(x) = \sum_{\iota=1}^q \alpha_\iota 1_{D_\iota}(x), \quad P^\epsilon(x, \mathrm{d}z) = \sum_{\iota=1}^q \tilde{P}_\iota(\mathrm{d}z) 1_{D_\iota}(x)$$

such that

(1) $R(D_\iota) > 0$ and $R(\partial D_\iota) = 0$;
(2) $\|\mathbf{f}^\epsilon - \mathbf{f}\| < \epsilon$;
(3) $d_{OW}(P^\epsilon(x, \mathrm{d}z), P(x, \mathrm{d}z)) < \epsilon$ if $x \in D_\iota$,

in order to fulfill Assumptions of Lemma 3.3. We will rely heavily on Assumption (A-2). This is a key step.

Let $\epsilon > 0$ be fixed and consider $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$. By Proposition A.1 in [22], there exist $\mathbf{a}_1, \ldots, \mathbf{a}_p \in \mathbb{R}^{m \times d}$ and $\epsilon_1, \ldots, \epsilon_p \in (0, \epsilon]$ such that the open sets $B_k^{\mathbf{f}} = f^{-1} B(\mathbf{a}_k, \epsilon_k)$ (where $B(\mathbf{a}_k, \epsilon_k) = \{\mathbf{a} \in \mathbb{R}^{m \times d},\ |\mathbf{a} - \mathbf{a}_k| < \epsilon_k\}$) are R-continuous and form a cover of $\mathcal{X}$:

$$\mathcal{X} \subset \cup_{k=1}^p B_k^{\mathbf{f}}.$$

Similarly, as $\Gamma : x \mapsto P_x$ is continuous and $\mathcal{X}$ is compact by (A-3), Proposition A.1 in [22] yields the existence of $P_1, \ldots, P_{p'} \in \mathcal{P}_\tau(\mathbb{R}^d)$ and $\epsilon_1', \ldots, \epsilon_{p'}' \in (0, \epsilon]$ such that the open sets $C_l^P = \Gamma^{-1} B_{OW}(P_l, \epsilon_l')$ (where $B_{OW}(P_l, \epsilon_l') = \{P \in \mathcal{P}_\tau;\ d_{OW}(P, P_l) < \epsilon_l'\}$) are R-continuous and form a cover of $\mathcal{X}$:

$$\mathcal{X} \subset \cup_{l=1}^{p'} C_l^P.$$

Thus, the family $(B_k^{\mathbf{f}} \cap C_l^P)_{k,l}$ is a cover of $\mathcal{X}$ based on R-continuous open sets. Assumption (A-1) implies that each set is either empty or with strictly positive R-measure. Let us keep the non-empty sets. The previous family satisfies the assumptions of Lemma A.2 in [22]. Therefore, there exists a partition $(D_\iota; 1 \le \iota \le q)$ satisfying the properties of Lemma A.2 in [22]: each of the D_ι is R-continuous with R-measure strictly positive. Moreover, for every ι, there exist k and l such that

$$\begin{aligned} D_\iota &\subset \overline{B_k^{\mathbf{f}}} \cap \overline{C_l^P} \\ &\subset \overline{f^{-1} B(\mathbf{a}_k, \epsilon_k)} \cap \overline{\Gamma^{-1} B_{OW}(P_l, \epsilon_l')}. \end{aligned} \tag{3.3}$$

Consider now a pairing which associates to each ι a unique couple $(k, l) = (k(\iota), l(\iota))$ such that (3.3) is satisfied. We denote by

$$\mathbf{f}^\epsilon = \sum_{\iota=1}^q \mathbf{a}_{k(\iota)}\ 1_{D_\iota}.$$

It is then straightforward to check that $\|\mathbf{f}^\epsilon - \mathbf{f}\| \le \epsilon$ and that each D_ι is R-continuous with strictly positive measure. Moreover, if $x \in D_\iota$, then

$$d_{OW}(P_x, P_{l(\iota)}) \le \epsilon.$$

We denote by P^ϵ the following kernel:

$$P^\epsilon(x,\ \mathrm{d}z) = \sum_{\iota=1}^q 1_{D_\iota}(x)\, P_{l(\iota)}(\mathrm{d}z). \tag{3.4}$$

We can now build properly the families $(Z_i^n)_{i,n}$ and $(Z_i^{n,\epsilon})_{i,n}$.

• *The families $(Z_i^n)_{i,n}$ and $(Z_i^{n,\epsilon})_{i,n}$.*
Let $x_i^n \in \mathcal{X}$. There exists a unique D_ι such that $x_i^n \in D_\iota$. In particular

$$d_{OW}(P_{x_i^n}, P_{l(\iota)}) < \epsilon,$$

where $l(\iota)$ is associated to ι *via* the pairing introduced after (3.3). Therefore, there exists a probability measure $\eta^{i,n}$ on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ with given marginals $P_{x_i^n}$ and $P_{l(\iota)}$ such that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \tau\left(\frac{x - y}{\epsilon}\right) \eta^{i,n}(\mathrm{d}x \mathrm{d}y) \le 1.$$

Endow $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ with the probability measure $\eta^{i,n}$ and set $Z_i^n(x,y) = x$ and $Z_i^{n,\epsilon}(x,y) = y$. Then Z_i^n is $P_{x_i^n}$-distributed, $Z_i^{n,\epsilon}$ is $P_{l_{(\iota)}}$-distributed and

$$E\tau\left(\frac{Z_i^n - Z_i^{n,\epsilon}}{\epsilon}\right) \leq 1 \quad \Rightarrow \quad E\mathrm{e}^{\frac{|Z_i^n - Z_i^{n,\epsilon}|}{\epsilon}} \leq 2. \tag{3.5}$$

The usual countable extension $\otimes_{i,n}\eta^{i,n}$ on the cartesian product $\Pi_{i,n}\mathbb{R}^d \times \mathbb{R}^d$ yields the existence of a family $(Z_i^n)_{i,n}$ satisfying (A-4). The previous construction also yields the fact that the distribution kernel associated to $(Z_i^{n,\epsilon})$ is given by (3.4). Thus, $\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle$ satisfies the assumptions of Lemma 3.3 and therefore satisfies the LDP with good rate function $I_{\mathbf{f}^\epsilon}$.

• *The exponential approximation.*
First notice that

$$\mathbb{P}\left\{|\langle L_n, \mathbf{f}\rangle - \langle L_n^\epsilon, \mathbf{f}^\epsilon\rangle| > \delta\right\} \leq \mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i) \cdot (Z_i^n - Z_i^{n,\epsilon})\right| > \frac{\delta}{2}\right\}$$
$$+ \mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i)) \cdot Z_i^{n,\epsilon}\right| > \frac{\delta}{2}\right\}.$$

Thus

$$\limsup_n \frac{1}{n}\log\mathbb{P}\left\{|\langle L_n, \mathbf{f}\rangle - \langle L_n^\epsilon, \mathbf{f}^\epsilon\rangle| > \delta\right\} \leq \limsup_n \frac{1}{n}\log\mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i) \cdot (Z_i^n - Z_i^{n,\epsilon})\right| > \frac{\delta}{2}\right\}$$
$$\vee \limsup_n \frac{1}{n}\log\mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i)) \cdot Z_i^{n,\epsilon}\right| > \frac{\delta}{2}\right\},$$

where $a \vee b = \max(a, b)$. Consider first:

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i) \cdot (Z_i^n - Z_i^{n,\epsilon})\right| > \frac{\delta}{2}\right\} \leq \mathbb{P}\left\{\frac{1}{n}\sum_1^n |Z_i^n - Z_i^{n,\epsilon}| > \frac{\delta}{2\|\mathbf{f}\|}\right\}$$
$$\leq \exp\left(-\frac{n\delta}{2\|\mathbf{f}\|\epsilon}\right)\prod_1^n \mathbb{E}\,\mathrm{e}^{\frac{|Z_i^n - Z_i^{n,\epsilon}|}{\epsilon}} \leq \exp\left(-\frac{n\delta}{2\|\mathbf{f}\|\epsilon}\right)2^n,$$

where the last inequality comes from (3.5). Thus

$$\limsup_n \frac{1}{n}\log\mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n \mathbf{f}(x_i) \cdot (Z_i^n - Z_i^{n,\epsilon})\right| > \frac{\delta}{2}\right\} \leq -\frac{n\delta}{2\|f\|\epsilon} + \log 2. \tag{3.6}$$

Consider now

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i)) \cdot Z_i^{n,\epsilon}\right| > \frac{\delta}{2}\right\} \leq \mathbb{P}\left\{\frac{1}{n}\sum_1^n |Z_i^{n,\epsilon}| > \frac{\delta}{2\epsilon}\right\}$$
$$\leq \exp\left(-\frac{\alpha^* n\delta}{2\epsilon}\right)\prod_1^n \mathbb{E}\,\mathrm{e}^{\alpha^*|Z_i^{n,\epsilon}|},$$

where $\alpha^* > 0$ is chosen so that (2.2) is satisfied. We therefore get

$$\limsup_n \frac{1}{n} \log \mathbb{P} \left\{ \left| \frac{1}{n} \sum_1^n (\mathbf{f}(x_i) - \mathbf{f}^\epsilon(x_i)) \cdot Z_i^{n,\epsilon} \right| > \frac{\delta}{2} \right\} \leq -\frac{\alpha^* \delta}{2\epsilon} + \log S^*, \tag{3.7}$$

where S^* is given by (2.2). Finally, combining (3.6) and (3.7) with (3.5), we get

$$\lim_{\epsilon \to 0} \limsup_n \frac{1}{n} \log \mathbb{P} \left\{ |\langle L_n, \mathbf{f} \rangle - \langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle| > \delta \right\} = -\infty.$$

Consequently, the family $(\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle; \epsilon > 0)$ is an exponential approximation of $\langle L_n, \mathbf{f} \rangle$. As $\langle L_n^\epsilon, \mathbf{f}^\epsilon \rangle$ satisfies the LDP by Lemma 3.3 with good rate function $I_{\mathbf{f}^\epsilon}$, Theorem 4.2.16 in [11] yields a weak LDP (*i.e.* the lower bound of the LDP holds for every open set while the upper bound only holds for compact sets) for $\langle L_n, \mathbf{f} \rangle$ with rate function

$$\Upsilon(y) = \sup_\eta \liminf_\epsilon \inf_{y' \in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y') = \sup_\eta \limsup_\epsilon \inf_{y' \in B(y,\eta)} I_{\mathbf{f}^\epsilon}(y'). \tag{3.8}$$

• *The full LDP: exponential tightness of $\langle L_n, \mathbf{f} \rangle$.* In order to prove that $\langle L_n, \mathbf{f} \rangle$ satisfies the full LDP, one has to check that $\langle L_n, \mathbf{f} \rangle$ is exponentially tight (see Chap. 4 in [11] for the definition of the exponential tightness):

$$\mathbb{P} \left\{ \left| \frac{1}{n} \sum_1^n \mathbf{f}(x_i) \cdot Z_i^n \right| > K \right\} \leq \mathbb{P} \left\{ \frac{1}{n} \sum_1^n |Z_i^n| > \frac{K}{\|\mathbf{f}\|} \right\}$$

$$\leq \exp \left(-\frac{n K \alpha^*}{\|\mathbf{f}\|} \right) \prod_1^n \mathbb{E} \, \mathrm{e}^{\alpha^* |Z_i^n|}.$$

As previously, if $\alpha^* > 0$ is chosen so that (2.2) is satisfied, we get:

$$\limsup_n \frac{1}{n} \log \mathbb{P} \left\{ \left| \frac{1}{n} \sum_1^n \mathbf{f}(x_i) \cdot Z_i^n \right| > K \right\} \leq -\frac{K \alpha^*}{\|\mathbf{f}\|} + \log S^* \xrightarrow[K \to \infty]{} -\infty.$$

Therefore, $\langle L_n, \mathbf{f} \rangle$ is exponentially tight and satisfies the full LDP with good rate function Υ. $\qquad \square$

Proof of Theorem 3.1. Given any kernel (P_x) satisfying (A-5), we have shown in the previous proof that there exist random variables with distribution kernel (P_x) satisfying (A-4) and such that the LDP holds for $\langle L_n, \mathbf{f} \rangle$. This is sufficient to assert that the LDP holds for any family of random variables satisfying (A-4) and (A-5). $\qquad \square$

3.2. Identification of the rate function

Let us first introduce some definitions which are customary in Convex Analysis. We say that J^ϵ **epi-converges** to J and we note $J = \text{epi-lim}_{\epsilon \to 0} J^\epsilon$ if for every $y \in \mathbb{R}^d$

$$\liminf_{\epsilon \to 0} J^\epsilon(y^\epsilon) \geq J(y) \quad \text{for every sequence } y^\epsilon \xrightarrow[\epsilon \to 0]{} y \tag{3.9}$$

$$\text{and} \quad \limsup_{\epsilon \to 0} J^\epsilon(y^\epsilon) \leq J(y) \quad \text{for one sequence } y^\epsilon \xrightarrow[\epsilon \to 0]{} y. \tag{3.10}$$

This is equivalent to:

$$\sup_{\eta > 0} \liminf_{\epsilon \to 0} \inf_{y' \in B(y,\eta)} J^\epsilon \geq J \tag{3.11}$$

$$\text{and} \quad \sup_{\eta > 0} \limsup_{\epsilon \to 0} \inf_{y' \in B(y,\eta)} J^\epsilon \leq J. \tag{3.12}$$

For the details, see the book by Rockafellar and Wets [25], Chapter 7. Epigraphical convergence has already been mentioned by Dawson and Gärtner in [6,7] to identify rate functions.

In order to prove Theorem 3.2, we will heavily rely on Wijsman's theorem (Th. 11.34, [25]) which states that the convex conjugation is continuous with respect to the epi-convergence. Otherwise stated, if J^ϵ epi-converges to J, then $J^{\epsilon,*}$ will epi-converge to J^*, where $J^{\epsilon,*}$ (resp. J^*) is the convex conjugate of J^ϵ (resp. J).

Recall that by Lemma 3.4 (iv), the rate function of the LDP is given by $\Upsilon = \text{epi-lim}_\epsilon I_{\mathbf{f}^\epsilon}$ where $I_{\mathbf{f}^\epsilon}$ is defined in Lemma 3.4 (ii) by:

$$I_{\mathbf{f}^\epsilon}(y) = \sup_{\theta \in \mathbb{R}^m} \left\{ y \cdot \theta - \int_{\mathcal{X}} \Lambda^\epsilon\left(x, \theta \cdot \mathbf{f}^\epsilon(x)\right) R(\mathrm{d}x) \right\}, \quad y \in \mathbb{R}^m.$$

In order to prove Theorem 3.2, we will prove that $\int_{\mathcal{X}} \Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) R(\mathrm{d}x)$ epi-converges toward $\int_{\mathcal{X}} \Lambda(x, \theta \cdot \mathbf{f}(x)) R(\mathrm{d}x)$ (Lem. 3.6). Wijsman's theorem will then allow us to conclude that $I_{\mathbf{f}^\epsilon}$ converges toward the convex conjugate of $\int_{\mathcal{X}} \Lambda(x, \theta \cdot \mathbf{f}(x)) R(\mathrm{d}x)$, which is the desired result. We first need the following result:

Proposition 3.5. *Let $(Z_i^n)_{i,n}$ and $(Z_i^{n,\epsilon})_{i,n}$ be as in Lemma 3.4 and consider their distribution kernels $(P_x)_{x \in \mathcal{X}}$ and $(P_x^\epsilon)_{x \in \mathcal{X}}$. Let $x \in \mathbb{R}^d$ be fixed, then*

$$\Lambda(x, \lambda) = \text{epi-lim}_{\epsilon \to 0} \Lambda^\epsilon(x, \lambda)$$

where $\Lambda(x, \cdot)$ (resp. $\Lambda^\epsilon(x, \cdot)$) is the cumulant generating function associated to (P_x) (resp. (P_x^ϵ)).

Proof. Let x be fixed and consider a family of i.i.d. random variables $(\check{Z}_i)_{i \geq 1}$ P_x- distributed. The empirical mean $\frac{1}{n} \sum_{i=1}^n \check{Z}_i$ satisfies the LDP with good rate function

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

For every $\epsilon > 0$, the probability measure $P^\epsilon(x, \mathrm{d}z)$ satisfies

$$d_{OW}(P_x^\epsilon, P_x) \leq \epsilon.$$

Therefore one can build an exponential approximation of $\frac{1}{n} \sum_{i=1}^n \check{Z}_i$ with P_x^ϵ-distributed i.i.d. families of random variables $(\check{Z}_i^\epsilon)_{i \geq 1}$. Thus, Theorem 4.2.16 in [11] yields the following identity for the rate functions:

$$\text{let } x \text{ be fixed, } \Lambda^*(x, z) = \text{epi-lim}_{\epsilon \to 0} \Lambda_\epsilon^*(x, z).$$

Wijsman's theorem (Th. 11.34, [25]) yields then:

$$\text{let } x \text{ be fixed, } \Lambda(x, \lambda) = \text{epi-lim}_{\epsilon \to 0} \Lambda_\epsilon(x, \lambda)$$

and Proposition 3.5 is proved. □

Lemma 3.6. *Consider the following integral functionals*

$$\Gamma(\theta) = \int_{\mathcal{X}} \Lambda(x, \theta \cdot \mathbf{f}(x)) R(\mathrm{d}x) \quad and \quad \Gamma^\epsilon(\theta) = \int_{\mathcal{X}} \Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) R(\mathrm{d}x),$$

where $\theta \in \mathbb{R}^m$. Then

$$\Gamma = \text{epi-lim}_{\epsilon \to 0} \Gamma^\epsilon. \tag{3.13}$$

Proof. Assume first that $\Lambda(x, \cdot)$ and $\Lambda^\epsilon(x, \cdot)$ are non-negative.

In view of Theorem 7.17 in [25], it is necessary and sufficient to prove the pointwise convergence $\Gamma^\epsilon(\theta) \to \Gamma(\theta)$ for $\theta \in D$ where D is a dense subset of $\mathbb{R}^m$. Denote by

$$\mathcal{D}_\Gamma = \{\theta \in \mathbb{R}^m, \ \Gamma(\theta) < \infty\}.$$

We shall first deal with the case where $\theta \in (\overline{\mathcal{D}_\Gamma})^c$. In this case, $\Gamma(\theta) = \infty$ and

$$\theta \cdot \mathbf{f}^\epsilon(x) \xrightarrow[\epsilon \to 0]{} \theta \cdot \mathbf{f}(x).$$

In particular, the lower bound (3.9) of the epi-convergence yields

$$\liminf_{\epsilon \to 0} \Lambda_\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) \geq \Lambda(x, \theta \cdot \mathbf{f}(x)).$$

As the Λ_ϵ's are non-negative, we can apply Fatou's lemma:

$$\liminf_{\epsilon \to 0} \int_{\mathbb{R}^d} \Lambda_\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) \, R(\mathrm{d}x) \geq \int_{\mathbb{R}^d} \Lambda(x, \theta \cdot \mathbf{f}(x)) \, R(\mathrm{d}x).$$

Thus

$$\lim_{\epsilon \to 0} \Gamma^\epsilon(\theta) = \Gamma(\theta) = \infty \quad \text{for all} \quad \theta \in (\overline{\mathcal{D}_\Gamma})^c. \tag{3.14}$$

Let now $\theta \in \text{int}(\mathcal{D}_\Gamma)$ be fixed. We denote by $\mathcal{D}_x = \{\lambda \in \mathbb{R}^d, \ \Lambda(x, \lambda) < \infty\}$. Let us first prove the pointwise convergence:

$$\lim_{\epsilon \to 0} \Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) = \Lambda(x, \theta \cdot \mathbf{f}(x)). \tag{3.15}$$

In order to do so, we shall prove that

$$\theta \in \text{int}(\mathcal{D}_\Gamma) \Rightarrow \theta \cdot \mathbf{f}(x) \in \text{int}(\mathcal{D}_x) \ a.e. \tag{3.16}$$

If $\theta \in \text{int}(\mathcal{D}_\Gamma)$, then $\Lambda(x, \theta \cdot \mathbf{f}(x)) < \infty \ a.e.$ Assume

$$\exists x \in \mathcal{X}, \quad \theta \cdot \mathbf{f}(x) \in \partial \mathcal{D}_x. \tag{3.17}$$

Then there exists $\epsilon > 0$ and there exists $\xi \in \mathbb{R}^m$ such that

$$|\xi - \theta| < \epsilon, \quad \Lambda(x, \xi \cdot \mathbf{f}(x)) = \infty \quad \text{and} \quad \xi \in \text{int}(\mathcal{D}_\Gamma).$$

Let $\alpha + \beta + \delta = 1$ be a convex combination, then Hölder's inequality yields

$$\begin{aligned}
\Lambda(x, \xi \cdot \mathbf{f}(x)) &= \log \int_{\mathbb{R}^d} \mathrm{e}^{\xi \cdot \mathbf{f}(x) \cdot z} P(x, \mathrm{d}z) \\
&\leq \alpha \log \int_{\mathbb{R}^d} \mathrm{e}^{\frac{\xi \cdot (\mathbf{f}(x) - \mathbf{f}(y))}{\alpha} \cdot z} P(x, \mathrm{d}z) \\
&\quad + \beta \log \int_{\mathbb{R}^d} \mathrm{e}^{\frac{\xi \cdot \mathbf{f}(y)}{\beta} \cdot (z - z')} \eta(\mathrm{d}z, \mathrm{d}z') + \delta \Lambda\left(y, \frac{\xi \cdot \mathbf{f}(y)}{\delta}\right)
\end{aligned} \tag{3.18}$$

where η has given marginals P_x and P_y. Since $\xi \in \text{int}(\mathcal{D}_\Gamma)$, there exists $\delta < 1$ such that $\xi/\delta \in \text{int}(\mathcal{D}_\Gamma)$. Let $\alpha = \beta = (1 - \delta)/2$ and choose a neighborhood V_x of x such that

$$\log \mathbb{E} \, \mathrm{e}^{\frac{|\xi| |\mathbf{f}(x) - \mathbf{f}(y)|}{\alpha} |Z_x|} < \infty \quad \text{and} \quad d_{OW}(P_x, P_y) \leq \beta \quad \text{if} \ y \in V_x.$$

In particular $\log \int_{\mathbb{R}^d} \mathrm{e}^{\frac{\xi \cdot \mathbf{f}(y)}{\beta} \cdot (z - z')} \eta(\mathrm{d}z, \mathrm{d}z') < \infty$ if η is well-chosen (recall the definition of the Orlicz-Wasserstein distance) and (3.18) yields

$$y \in V_x \Rightarrow \Lambda \left(y, \frac{\xi \cdot \mathbf{f}(y)}{\delta} \right) = \infty.$$

Necessarily,

$$\Gamma(\xi/\delta) = \int_{\mathcal{X}} \Lambda \left(x, \frac{\xi \cdot \mathbf{f}(x)}{\delta} \right) R(\mathrm{d}x)$$
$$\geq \int_{V_x} \Lambda \left(x, \frac{\xi \cdot \mathbf{f}(x)}{\delta} \right) R(\mathrm{d}x) = \infty,$$

(recall that $R(V_x) > 0$ by Assumption (A-1)) which is impossible since $\xi/\delta \in \mathcal{D}_\Gamma$. Necessarily, (3.17) is false and (3.16) is proved.

Assume from now that $\theta \in \mathrm{int}(\mathcal{D}_\Gamma)$. As shown previously, there exists a neighborhood V_x of $\theta \cdot \mathbf{f}(x)$ included in $\mathrm{int}(\mathcal{D}_x)$. Since no points of the boundary $\partial \mathcal{D}_x$ belong to V_x, a caracterization of the epi-convergence (Th. 7.17, [25]) yields that:

$$\sup_{\zeta \in V_x} |\Lambda^\epsilon(x, \zeta) - \Lambda(x, \zeta)| \xrightarrow[\epsilon \to \infty]{} 0.$$

As $\lim_{\epsilon \to 0} \theta \cdot \mathbf{f}^\epsilon(x) = \theta \cdot \mathbf{f}(x)$, $\theta \cdot \mathbf{f}^\epsilon(x) \in V_x$ for ϵ small enough and

$$\lim_{\epsilon \to 0} \Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) = \Lambda(x, \theta \cdot \mathbf{f}(x)).$$

It remains to prove that $\Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x))$ is uniformly majorized by an integrable function.

$$\Lambda^\epsilon(x, \theta \cdot \mathbf{f}^\epsilon(x)) \leq \alpha \log \mathbb{E} \, \mathrm{e}^{\frac{|\theta| |\mathbf{f}^\epsilon(x) - \mathbf{f}(x)|}{\alpha} |Z_x^\epsilon|}$$
$$+ \beta \log \mathbb{E} \, \mathrm{e}^{\frac{|\theta| \|\mathbf{f}^\epsilon\|}{\beta} \cdot |Z_x^\epsilon - Z_x|} + \delta \Lambda \left(x, \frac{\theta \cdot \mathbf{f}(x)}{\delta} \right).$$

First choose δ close enough to 1 so that $\theta/\delta \in \mathcal{D}_\Gamma$. Then fix α and β so that $\alpha + \beta + \delta = 1$. For ϵ small enough, the right hand of the previous inequality can be majorized uniformily in ϵ. The Dominated convergence theorem yields then

$$\lim_{\epsilon \to 0} \Gamma^\epsilon(\theta) = \Gamma(\theta) \quad \text{for all} \quad \theta \in \mathrm{int}(\mathcal{D}_\Gamma). \tag{3.19}$$

Finally, (3.14) and (3.19) yield the pointwise convergence $\lim_{\epsilon \to 0} \Gamma^\epsilon(\theta) = \Gamma(\theta)$ for every point outside $\partial \mathcal{D}_\Gamma$. Since the complement of $\partial \mathcal{D}_\Gamma$ is dense in $\mathbb{R}^d$, Theorem 7.17 in [25] yields the epi-convergence of Γ^ϵ toward Γ and the case where Λ and Λ^ϵ are non-negative is done.

In the general case, let

$$m_x = \int_{\mathbb{R}^d} z \, P(x, \mathrm{d}z) \quad \text{and} \quad m_x^\epsilon = \int_{\mathbb{R}^d} z \, P^\epsilon(x, \mathrm{d}z).$$

As $W(P, Q) \leq d_{OW}(P, Q)$ (recall Rem. 2.1), a straightforward application of the Kantorovich-Rubinstein theorem (Th. 11.8.2, [14]) yields that:

$$\lim_{\epsilon \to 0} m_x^\epsilon = m_x. \tag{3.20}$$

Since $|m_x^\epsilon| \leq K S^*$ where S^* is given by (2.2), the Dominated convergence theorem yields

$$\int_{\mathcal{X}} \mathbf{f}^\epsilon \cdot m_x^\epsilon \, R(\mathrm{d}x) \to \int_{\mathcal{X}} \mathbf{f} \cdot m_x \, R(\mathrm{d}x). \tag{3.21}$$

Let us introduce

$$\tilde{\Lambda}(x,\lambda) = \log \int_{\mathcal{X}} \mathrm{e}^{\lambda \cdot (z-m_x)} P(x,\,\mathrm{d}z) = \Lambda(x,\lambda) - \lambda \cdot m_x \quad \text{and}$$

$$\tilde{\Lambda}^{\epsilon}(x,\lambda) = \log \int_{\mathcal{X}} \mathrm{e}^{\lambda \cdot (z-m_x^{\epsilon})} P^{\epsilon}(x,\,\mathrm{d}z). = \Lambda^{\epsilon}(x,\lambda) - \lambda \cdot m_x^{\epsilon}.$$

Since the random variables $Z^x - m_x$ are centered, $\tilde{\Lambda}(x,\lambda)$ and $\tilde{\Lambda}^{\epsilon}(x,\lambda)$ are non-negative for they are convex and $\nabla \Lambda(x,0) = \nabla \Lambda^{\epsilon}(x,0) = 0$. As (3.20) holds,

$$\tilde{\Lambda}(x,\lambda) = \operatorname*{epi-lim}_{\epsilon \to 0} \tilde{\Lambda}^{\epsilon}(x,\lambda).$$

The first part of the proof yields

$$\int_{\mathcal{X}} \tilde{\Lambda}(x, \theta \cdot \mathbf{f}(x)) \, R(\mathrm{d}x) = \operatorname*{epi-lim}_{\epsilon \to 0} \int_{\mathcal{X}} \tilde{\Lambda}^{\epsilon}(x, \theta \cdot \mathbf{f}^{\epsilon}(x)) \, R(\mathrm{d}x),$$

that is

$$\int_{\mathcal{X}} \Lambda(x, \theta \cdot \mathbf{f}(x)) \, R(\mathrm{d}x) + \theta \cdot \int_{\mathcal{X}} \mathbf{f}(x) \cdot m_x \, R(\mathrm{d}x) = \operatorname*{epi-lim}_{\epsilon \to 0} \left(\int_{\mathcal{X}} \Lambda^{\epsilon}(x, \theta \cdot \mathbf{f}^{\epsilon}(x)) \, R(\mathrm{d}x) + \theta \cdot \int_{\mathcal{X}} \mathbf{f}^{\epsilon}(x) \cdot m_x^{\epsilon} \, R(\mathrm{d}x) \right).$$

The conclusion follows as $\int \mathbf{f}^{\epsilon} \cdot m_{\cdot}^{\epsilon} \, \mathrm{d}R \to \int \mathbf{f} \cdot m_{\cdot} \, \mathrm{d}R$ by (3.21) and Lemma 3.6 is proved. $\qquad \square$

Proof of Theorem 3.2. Since $\Gamma = \operatorname{epi-lim}_{\epsilon \to 0} \Gamma^{\epsilon}$, Wijsman's theorem (theorem 11.34, [25]) yields the epi-convergence for the convex conjugates, that is precisely

$$I_{\mathbf{f}} = \operatorname*{epi-lim}_{\epsilon \to 0} I_{\mathbf{f}^{\epsilon}}$$

and Theorem 3.2 is proved. $\qquad \square$

4. LARGE DEVIATIONS IN INFINITE DIMENSION

In this section, we prove a LDP for empirical measures. This result is of interest in statistical physics [15] and in statistics [17]. We also prove a LDP for random walks in the spirit of Mogul'skii's theorem. Both results rely on Theorem 3.1 and are a careful adaptation of the proofs in [22].

In the case where $\mathcal{X}$ is compact, the continuous linear forms over $C_d(\mathcal{X})$ are vector measures. We denote by $M^1(\mathcal{X})$ the set of real measures and by $M^d(\mathcal{X})$ the set of vector measures with value in $\mathbb{R}^d$, that is $\mu \in M^d(\mathcal{X})$ iff $\mu = (\mu_1, \ldots, \mu_d)$ where each $\mu_i \in M^1(\mathcal{X})$. Let $f \in C_d(\mathcal{X})$ and $\mu \in M^d(\mathcal{X})$. We denote by

$$\int_{\mathcal{X}} f(x) \cdot \mu(\mathrm{d}x) \overset{\triangle}{=} \sum_{l=1}^{d} \int_{\mathcal{X}} f_l(x) \mu_l(\mathrm{d}x). \tag{4.1}$$

In the case where $\mathbf{f} : \mathcal{X} \to \mathbb{R}^{m \times d}$ is a matrix valued bounded continuous function, we denote by

$$\int \mathbf{f} \cdot \mathrm{d}\mu \overset{\triangle}{=} \begin{pmatrix} \int_{\mathcal{X}} f_1(x) \cdot \mu(\mathrm{d}x) \\ \vdots \\ \int_{\mathcal{X}} f_m(x) \cdot \mu(\mathrm{d}x) \end{pmatrix} \tag{4.2}$$

where $f_j \in C_d(\mathcal{X})$ is the jth row of the matrix **f**. We shall endow $M^d(\mathcal{X})$ with the weak-$*$ topology $\sigma(M^d, C_d)$ which makes every linear form $\Gamma_f : \mu \mapsto \int f \cdot \mathrm{d}\mu$ continuous and with the associated Borel σ-field $\mathcal{B}(M^d)$. Let $\mu \in M^d(\mathcal{X})$. We denote by μ_a its absolutely continuous part with respect to R and by μ_s its singular part.

Following Lynch and Sethuraman [21], we introduce some notations. Let $\mathcal{B}V([0,1], \mathbb{R}^d)$ (shortened in $\mathcal{B}V$) be the space of functions of bounded variation on $[0,1]$. We identify $\mathcal{B}V$ with $M^d([0,1])$ in the usual manner: To $f \in \mathcal{B}V$ there corresponds μ^f characterized by $\mu^f([0,t]) = f(t)$. We endow $\mathcal{B}V$ with the weak-$*$ topology $\sigma(\mathcal{B}V, C_d([0,1]))$ (shortened in σ_w) and with the associated Borel σ-field $\mathcal{B}_w$. Up to this identification, $C_d([0,1])$ is the topological dual of $\mathcal{B}V$. Let $f \in \mathcal{B}V$ and μ^f be the associated measure in $M^d([0,1])$. We will denote $\mu^f = \mu_a^f + \mu_s^f$ where μ_a^f is the absolutely continuous part of μ^f with respect to the Lebesgue measure $\mathrm{d}x$ and μ_s^f its singular part.

Recall that $\mathcal{D}_x = \{\lambda \in \mathbb{R}^d, \ \Lambda(x, \lambda) < \infty\}$. The recession function $\rho(x, z)$ of $\Lambda^*(x, z)$ is defined by $\rho(x, z) = \sup\{\lambda \cdot z, \ \lambda \in \mathcal{D}_x\}$.

4.1. **Large deviations in infinite dimension**

We shall prove here a LDP for the empirical measure L_n and the random walk $\bar{Z}_n$ where

$$L_n = \frac{1}{n} \sum_{i=1}^{n} Z_i^n \delta_{x_i^n} \quad \text{and} \quad \bar{Z}_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} Z_i^n$$

where $[\cdot]$ denotes the integer part.

Theorem 4.1. *Assume that (A-1), (A-2), (A-3), (A-4) and (A-5) hold. Then the family*

$$L_n = \frac{1}{n} \sum_{1}^{n} Z_i^n \delta_{x_i^n}$$

satisfies the large deviation principle in $(M^d(\mathcal{X}), \sigma(M^d, C_d), \mathcal{B}(M^d))$ with the good rate function

$$
\begin{aligned}
I(\mu) &= \sup_{f \in C_d(\mathcal{X})} \left\{ \int_{\mathcal{X}} f(x) \cdot \mu(\mathrm{d}x) - \int_{\mathcal{X}} \Lambda(x, f(x)) \, R(\mathrm{d}x) \right\} \\
&= \int_{\mathcal{X}} \Lambda^* \left(x, \frac{\mathrm{d}\mu_a}{\mathrm{d}R}(x) \right) R(\mathrm{d}x) + \int_{\mathcal{X}} \rho \left(x, \frac{\mathrm{d}\mu_s}{\mathrm{d}\theta}(x) \right) \theta(\mathrm{d}x),
\end{aligned}
\tag{4.3}
$$

where $\rho(x, z)$ is the recession function of $\Lambda^*(x, \cdot)$ and θ is any real-valued nonnegative measure with respect to which μ_s is absolutely continuous.

Proof.
• *The LDP.* Denote by $C_d'(\mathcal{X})$ (resp. $C'(\mathcal{X})$) the algebraic dual of $C_d(\mathcal{X})$ (resp. $C(\mathcal{X})$). Denote by $\langle \ , \ \rangle$ the duality bracket between these spaces and consider the mapping

$$p_{f_1, \ldots, f_m} : C_d'(\mathcal{X}) \to \mathbb{R}^m, \quad \xi \mapsto \begin{pmatrix} \langle \xi, f_1 \rangle \\ \vdots \\ \langle \xi, f_m \rangle \end{pmatrix},$$

where $f_i \in C_d(\mathcal{X})$. Then $p_{f_1,\ldots,f_m}(L_n) = \langle L_n, \mathbf{f} \rangle$ satisfies a LDP by Theorem 3.1. By Dawson-Gärtner's theorem, L_n satisfies a LDP in $C'_d(\mathcal{X})$ endowed with the weak-$*$ topology with the good rate function

$$I(\xi) = \sup_{m \geq 1} \sup_{f_1,\ldots,f_m \in C_d(\mathcal{X})} \sup_{\theta \in \mathbb{R}^m} \left\{ \sum_1^m \theta_i \langle \xi, f_i \rangle - \int_{\mathcal{X}} \Lambda(x, \sum_1^m \theta_i f_i(x))\, R(\mathrm{d}x) \right\}$$
$$= \sup_{f \in C_d(\mathcal{X})} \left\{ \langle \xi, f \rangle - \int_{\mathcal{X}} \Lambda(x, f(x))\, R(\mathrm{d}x) \right\} \quad \text{for} \quad \xi \in C'_d(\mathcal{X}).$$

• *Restriction of the LDP.* Let us show that $I(\xi) < +\infty$ implies that ξ is a continuous linear form. Assume that $I(\xi) < +\infty$. Then for all $f \in C_d(\mathcal{X})$, $f \not\equiv 0$

$$\left\langle \xi, \frac{f}{a\|f\|} \right\rangle \leq I(\xi) + \int_{\mathcal{X}} \Lambda\left(x, \frac{f}{a\|f\|}\right) \mathrm{d}R \leq I(\xi) + \sup_{x \in \mathcal{X}} \log \int_{\mathbb{R}^d} \mathrm{e}^{\frac{|z|}{a}} P(x, \mathrm{d}z).$$

For a large enough $\sup_{x \in \mathcal{X}} \log \int_{\mathbb{R}^d} \mathrm{e}^{|z|/a} P(x, \mathrm{d}z)$ is finite by (2.2) and $\langle \xi, f \rangle \leq K\|f\|$. Considering $-f$, we get $|\langle \xi, f \rangle| \leq K\|f\|$. Thus ξ is a continuous linear form. Since $\mathcal{X}$ is compact, Riesz's representation theorem implies that ξ can be represented as a $\mathbb{R}^d$-valued measure over $\mathcal{X}$, *i.e.* $\xi \in M^d(\mathcal{X})$. We shall denote it by μ. We can now apply Lemma 4.1.5 in [11] to obtain the LDP in $(M^d(\mathcal{X}), \sigma(M^d, C_d), \mathcal{B}(M^d))$.

• *Representation of the rate function.* The representation of the rate function relies heavily on the following theorem from Rockafellar:

Theorem 4.2 (Rockafellar, Th. 5 in [24])**.** *Assume that $\mathcal{X}$ is a compact space with no nonempty open sets of measure zero, and that the multifunction $\mathcal{D} : x \mapsto \mathcal{D}_x$ is fully lower semicontinuous, with $\mathrm{int}\,\mathcal{D}_x \neq \emptyset$ for every x. Assume further that $\int_V |\Lambda(x, \lambda)| R(\mathrm{d}x) < \infty$ whenever V is an open subset of $\mathcal{X}$ and λ is a point of $\mathbb{R}^d$ having a neighborhood U such that $U \subset \mathcal{D}_x$ for all $x \in V$. Then*

$$I(\mu) = \sup_{f \in C_d(\mathcal{X})} \left\{ \int_{\mathcal{X}} f(x) \cdot \mu(\mathrm{d}x) - \int_{\mathcal{X}} \Lambda(x, f(x))\, R(\mathrm{d}x) \right\}$$
$$= \int_{\mathcal{X}} \Lambda^*\left(x, \frac{\mathrm{d}\mu_a}{\mathrm{d}R}(x)\right) R(\mathrm{d}x) + \int_{\mathcal{X}} \rho\left(x, \frac{\mathrm{d}\mu_s}{\mathrm{d}\theta}(x)\right) \theta(\mathrm{d}x),$$

where $\rho(x, \cdot)$ is the recession function of $\Lambda^*(x, \cdot)$, μ_s is the singular componant of μ with respect to R, and θ is any nonnegative measure in $M^1(\mathcal{X})$ with respect to which μ_s is absolutely continuous.

It is straightforward to check that $\mathrm{int}\,\mathcal{D}_x \neq \emptyset$ for every x holds here. We must check that the multifunction

$$\mathcal{D} : x \mapsto \mathcal{D}_x$$

is fully lower semicontinuous [24] (p. 457), that is:

(i) if $x_0 \in \mathcal{X}$ and $U \subset \mathbb{R}^d$ is an open set such that $\mathcal{D}_{x_0} \cap U \neq \emptyset$ then there exists a neighborhood V of x_0 such that $\mathcal{D}_x \cap U \neq \emptyset$ whenever $x \in V$;

(ii) if $V \times U$ is a neighborhood of $(x_0, \lambda_0) \in \mathcal{X} \times \mathbb{R}^d$, such that $\{x' \in V, U \subset \mathcal{D}_{x'}\}$ is dense in V then $\lambda_0 \in \overline{\mathcal{D}_{x_0}}$.

The first step is easy to check. Let us focus on the second point. Let $V \times U$ be a neighborhood of (x_0, λ_0) such that $\{x' \in V, U \subset \mathcal{D}_{x'}\}$ is dense in V. Let $0 < \beta < 1$ be such that $\frac{\lambda_0}{\beta} \in U$ and let $x \in \{x' \in V, U \subset \mathcal{D}_{x'}\}$ be such that $d_{OW}(P_x, P_{x_0}) \leq \frac{\alpha}{|\lambda_0|}$ $(\alpha + \beta = 1)$. Hölder's inequality yields

$$\Lambda(x_0, \lambda_0) \leq \alpha \log \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{e}^{\frac{|\lambda_0||z-z'|}{\alpha}} \eta(\mathrm{d}z, \mathrm{d}z') + \beta \Lambda(x, \lambda_0/\alpha),$$

where η has marginals P_{x_0} and P_x. Since $d_{OW}(P_x, P_{x_0}) \le \frac{\alpha}{|\lambda_0|}$, there exists at least a probability measure η with marginals P_{x_0} and P_x such that

$$\log \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{|\lambda_0| \, |z-z'|}{\alpha}} \eta(\mathrm{d}z, \mathrm{d}z') < \infty.$$

Therefore $\Lambda(x_0, \lambda_0) < \infty$ and $\lambda_0 \in \mathcal{D}_{x_0} \subset \overline{\mathcal{D}_{x_0}}$. Thus $\mathcal{D}$ is fully lower semicontinuous.

We must also check the following: Assume now that V is an open subset of $\mathcal{X}$ and that λ is a point of $\mathbb{R}^d$ having a neighborhood U such that $U \subset \mathcal{D}_x$ for all $x \in V$. We have to check that in this case,

$$\int_{\mathcal{X}} |\Lambda(x, \lambda)| R(\mathrm{d}x) < \infty. \tag{4.4}$$

First note that V is relatively compact. Let $x_0 \in V$ be fixed. Let $0 < \beta < 1$ be such that $\frac{\lambda}{\beta} \in U$ and let x be such that $d_{OW}(P_{x_0}, P_x) < \frac{\alpha}{|\lambda_0|}$. Then

$$|\Lambda(x, \lambda_0)| \le \alpha \left| \log \int_{\mathbb{R}^d \times \mathbb{R}^d} e^{\frac{|\lambda_0| \, |z-z'|}{\alpha}} \eta(\mathrm{d}z, \mathrm{d}z') \right| + \beta |\Lambda(x_0, \lambda_0/\beta)|,$$

where η has marginals P_{x_0} and P_x. The previous inequality implies that $|\Lambda(x, \lambda_0)|$ is dominated by a constant (the right side of the inequality) whenever $d_{OW}(P_{x_0}, P_x) < \frac{\alpha}{|\lambda_0|}$. Since V is relatively compact, $|\Lambda(\cdot, \lambda_0)|$ is dominated by a piecewise constant function and therefore is integrable.

As (A-1) and (A-3) hold true, all the assumptions of Theorem 4.2 are satisfied and

$$I(\mu) = \int_{\mathcal{X}} \Lambda^* \left[\frac{\mathrm{d}\mu_a}{\mathrm{d}R}(x) \right] R(\mathrm{d}x) + \int_{\mathcal{X}} \rho \left[\frac{\mathrm{d}\mu_s}{\mathrm{d}\theta} \right] \mathrm{d}\theta,$$

where ρ is the recession function of Λ^* and θ is any real-valued nonnegative measure with respect to which μ_s is absolutely continuous. As Λ is the convex conjugate of Λ^*, ρ is the support function of Λ [23] (Th. 13.3), that is:

$$\rho(x, z) = \sup\{\lambda \cdot z, \ \lambda \in \mathcal{D}_x\}.$$

Hence Theorem 4.1 is proved. $\qquad \square$

We shall now derive the LDP for the random function

$$t \mapsto \bar{Z}_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} Z_i, \quad t \in [0,1]$$

where $[x]$ denotes the integer part of x. This result is a corollary of Theorem 4.1 in the case where $\mathcal{X} = [0,1]$, $R(\mathrm{d}x) = \mathrm{d}x$ is the Lebesgue measure on $[0,1]$ and $(x_i^n) = \left(\frac{i}{n}\right)$. One can check that (A-1), (A-2) and (A-3) hold true in this situation.

Theorem 4.3. *Assume that (A-4) and (A-5) hold. Then the random functions $\left(\bar{Z}_n(t)\right)_{t \in [0,1]}$ satisfy the LDP in $(\mathcal{BV}, \sigma_w, \mathcal{B}_w)$ with the good rate function*

$$\Phi(f) = \int_{[0,1]} \Lambda^*(x, f_a'(x)) \, \mathrm{d}x + \int_{[0,1]} \rho(x, f_s'(x)) \, \mathrm{d}\theta(x),$$

where θ is any real-valued nonnegative measure with respect to which μ_s^f is absolutely continuous and $f_s' = \mathrm{d}\mu_s^f/\mathrm{d}\theta$.

Remark 4.1. Note that the definition of f'_s is θ-dependent. However, by homogeneity, Φ does not depend upon θ.

Proof. Consider $\Pi : M^d \to \mathcal{BV}$ where

$$\Pi(\mu) = \begin{pmatrix} \mu_1[0,x] \\ \vdots \\ \mu_d[0,x] \end{pmatrix}_{x \in [0,1]}$$

and recall that $L_n = 1/n \sum_1^n Z_i \delta_{x_i}$. Then Π is a continuous bijection and $\Pi(L_n) = \bar{Z}_n$. As L_n satisfies the LDP with the good rate function (4.3), the contraction principle yields the LDP with the desired rate function. $\square$

5. ERDÖS AND RÉNYI'S FUNCTIONAL LAW OF LARGE NUMBERS

In this section, we establish Erdös and Rényi's functional law of large numbers. Assume for the sake of simplicity that $\mathbb{E}\, Z_i^n = 0$ for every $1 \leq i \leq n, n \in \mathbb{N}$. In this case, $\Lambda^*(x_i^n, 0) = 0$ which in turn yields that $\Phi(0) = \int_{\mathcal{X}} \Lambda^*(x, 0)\, R(\mathrm{d}x) = 0$. Let

$$\eta_{n,m} : t \mapsto \frac{1}{A_n} \sum_{i=m}^{[m+tA_n]} Z_i^n, \ t \in [0,1]$$

where $A_n = [\log n]$ and $1 \leq m \leq n - A_n$ and recall that

$$\Phi(f) = \int_{[0,1]} \Lambda^*(x, f'_a(x))\, \mathrm{d}x + \int_{[0,1]} \rho(x, f'_s(x))\, \mathrm{d}\theta(x)$$

is defined over the set $\mathcal{BV}$ (we denote by $\Phi(A) = \inf_A \Phi$). Our main result is the following:

Theorem 5.1. *Under Assumptions (A-1), (A-2), (A-3), (A-4) and (A-5), There exists a set $\tilde{\Omega}$ with probability one such that for every $\omega \in \tilde{\Omega}$:*

(1) the set $\{\eta_{n,m}(\omega); \ 1 \leq m \leq n - A_n, \ n \in \mathbb{N}\}$ is relatively compact in $\mathcal{BV}$ and every cluster point belongs to

$$\Delta = \{f \in \mathcal{BV}; \ \Phi(f) \leq 1\};$$

(2) let $f \in \Delta$, then there exists a subsequence of $\{\eta_{n,m}(\omega)\}$ which converges to f.

Remark 5.1. Note that since $\mathcal{BV}$ is not first countable, the existence of a subsequence of $\eta_{n,m}$ converging to f is a stronger result than the fact that f is a cluster point of $\{\eta_{n,m}\}$ (in the later case, we would only have the existence of a subnet converging to f).

As a corollary, we get:

Theorem 5.2. *Let $\Gamma : \mathcal{BV} \to \mathbb{R}$ be a continuous functional then under the same assumptions as in Theorem 5.1,*

$$\lim_{n \to \infty} \sup_{1 \leq m \leq n - A_n} \Gamma(\eta_{n,m}) = \sup_{f \in \Delta} \Gamma(f)$$

with probability one.

In the case where the random variables are real, Theorem 5.1 has been proved by Deheuvels in [8] for i.i.d. random variables satisfying $(\exists \alpha > 0, \mathbb{E}e^{\alpha|Z_i|} < \infty)$. In the case where the random variables are i.i.d., $\mathbb{R}^d$-valued and satisfy the stronger integrability condition:

$$\mathbb{E}e^{\alpha|Z_i|} < \infty \quad \text{for all } \alpha > 0, \tag{5.1}$$

Theorem 5.1 has been proved by Borovkov in [2].

Our proof relies on Theorem 4.3 and on the technique developed by Deheuvels [8] to handle some delicate topological issues ($\mathcal{B}V$ is not first countable). In the spirit of Deheuvels, we shall first prove Theorem 5.3 hereafter, from which Theorem 5.1 is a direct corollary. Let us first introduce some notations. Denote by $|f|_v(s)$ the total variation of f, that is $|f|_v(s) = |\mu_f|([0, s])$, where $|\mu_f| = \mu_f^+ + \mu_f^-$ (Hahn-Jordan decomposition).

Let $\mathcal{B}V_C = \{f \in \mathcal{B}V, \ |f|_v(1) \leq C\}$ and consider the following distance defined on $\mathcal{B}V$:

$$d(f, g) = \int_0^1 |f - g| \, \mathrm{d}\ell + |f(1) - g(1)|,$$

where ℓ denotes the Lebesgue measure.

Then $\mathcal{B}V_C$ is a compact set of $\mathcal{B}V$ and the topology induced by the weak topology on $\mathcal{B}V_C$ is compatible with the one induced by d on $\mathcal{B}V_C$ (see [19]). We will denote by

$$K_n = \{\eta_{n,m}; \ 1 \leq m \leq n - A_n\}.$$

As usual, we will say that a condition κ_n holds ultimately if there exists N such that κ_n is true whenever $n \geq N$. Similarly, we will say that κ_n holds ultimately with probability one if there exists a set $\tilde{\Omega}$ with probability one such κ_n is true whenever $n \geq N(\omega)$ and $\omega \in \tilde{\Omega}$.

Theorem 5.3. *There exists a nonnegative real constant C such that*

$$\Delta \subset \mathcal{B}V_C \quad \text{and} \quad K_n \subset \mathcal{B}V_C \quad \text{ultimately.}$$

If $F \subset \mathcal{B}V_C$, denote by

$$F^\epsilon = \{f \in \mathcal{B}V_C, \ \exists g \in F, \ d(f, g) < \epsilon\}.$$

Then the following inclusion holds

$$\Delta \subset K_n^\epsilon \subset \Delta^{2\epsilon}$$

ultimately with probability one.

Our proof is essentially a restatement of Deheuvels' proof of Theorem 3.1 in [8]. Since our LDP is not exactly the one stated by Deheuvels in [8] (Lem. 2.14), a few details need to be added.

Proof. The guideline of the proof is to restrict ourselves to some compact set $\mathcal{B}V_c$ of $\mathcal{B}V$ which is metrisable with respect to the induced topology. We shall prove the theorem in several steps.

• **Step 1.** Choice of the constant C.

Let $A > 0$ then the complement of $\mathcal{B}V_A$ is a σ_w-open set in $\mathcal{B}V$ and

$$\{\bar{Z}_n \notin \mathcal{B}V_A\} = \left\{\frac{1}{n}\sum_{i=1}^n |Z_i| > A\right\}.$$

The LD lower bound in Theorem 4.3 yields

$$-\inf\{\Phi(f); \ f \notin \mathcal{B}V_A\} \leq \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\{\bar{Z}_n \notin \mathcal{B}V_A\}.$$

On the other hand,

$$\mathbb{P}\{\bar{Z}_n \notin \mathcal{B}V_A\} \le \mathrm{e}^{-n(A\alpha^* - \log S^*)}$$

where α^* and S^* are defined in (2.2). Thus

$$\frac{1}{n} \log \mathbb{P}\{\bar{Z}_n \notin \mathcal{B}V_A\} \le -(A\alpha^* - \log S^*) \le -2$$

for A large enough. In particular, $\inf\{\Phi(f); \ f \notin \mathcal{B}V_A\} \ge 2 > 1$ which implies that $\Delta \subset \mathcal{B}V_A$. Let us now prove that

$$\limsup_{n\to\infty} \sup_{1\le m\le n-A_n} \frac{1}{A_n} \sum_{i=m}^{m+A_n} |Z_i| \le B < \infty \tag{5.2}$$

with probability one. One shall notice that

$$\mathbb{P}\left(\sup_{1\le m\le n-A_n} \frac{1}{A_n} \sum_{i=m+1}^{m+A_n} |Z_i| > B\right) = \mathbb{P}\left(\bigcup_{m+1}^{n-A_n}\left\{\frac{1}{A_n} \sum_{i=m+1}^{m+A_n} |Z_i| > B\right\}\right)$$

$$\le n \sup_{1\le m\le n-A_n} \mathbb{P}\{\eta_{n,m} \notin \mathcal{B}V_B\} \le n \exp\{-A_n(B\alpha^* - \log S^*)\} \le \frac{1}{n^2}$$

for B large enough. Thus (5.2) follows by Borel-Cantelli's lemma. As $|\eta_{n,m}|_v(1) = \frac{1}{A_n}\sum_{i=m+1}^{m+A_n}|Z_i|$, (5.2) yields that $K_n \subset \mathcal{B}V_{B+1}$ ultimately. Let $C = \max(A, B+1)$. Then the first part of the theorem is proved. It is useful for the sequel to note that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\bar{Z}_n \notin \mathcal{B}V_C\} \le -2. \tag{5.3}$$

For the sake of clarity, we will write d-open (resp. d-closed) when working with open sets (respectively closed sets) in $\mathcal{B}V_C$ for the induced topology.

• **Step 2.** Let us prove that ultimately, $K_n^\epsilon \subset \Delta^{2\epsilon}$ *a.e.* Due to the triangle inequality, it is enough to prove that

$$\mathbb{P}\{K_n \not\subset \Delta^\epsilon, \text{ i.o.}\} = 0, \tag{5.4}$$

where i.o. stands for infinitely often.

$$\mathbb{P}\{K_n \not\subset \Delta^\epsilon\} = \mathbb{P}\left\{\bigcup_{m=1}^{n-A_n}\{\eta_{n,m} \notin \Delta^\epsilon\}\right\} \le n \sup_{1\le m\le n-A_n} \mathbb{P}\{\eta_{n,m} \notin \Delta^\epsilon\}.$$

Denote by $\mathcal{C}$ the complement of Δ^ϵ in $\mathcal{B}V$. One can not apply directly the large deviation upper bound to the event $\{\eta_{n,m} \in \mathcal{C}\}$ since $\mathcal{C}$ is not σ_w-closed in $\mathcal{B}V$. However,

$$\mathbb{P}\{\eta_{n,m} \in \mathcal{C}\} \le \mathbb{P}\{\eta_{n,m} \in \mathcal{C} \cap \mathcal{B}V_C\} + \mathbb{P}\{\eta_{n,m} \notin \mathcal{B}V_C\}. \tag{5.5}$$

As a subset of $\mathcal{B}V_C$, Δ^ϵ is d-open. Therefore, its complement in $\mathcal{B}V_C$, namely $\mathcal{C} \cap \mathcal{B}V_C$, is d-closed in $\mathcal{B}V_C$. As $\mathcal{B}V_C$ is a σ_w-compact set of $\mathcal{B}V$, so is $\mathcal{C} \cap \mathcal{B}V_C$. As Theorem 4.3 remains valid for $\eta_{n,m}$ with the speed A_n when m is fixed, the LD upper bound yields then

$$\limsup_{n\to\infty} \frac{1}{A_n} \log \mathbb{P}\{\eta_{n,m} \in \mathcal{C} \cap \mathcal{B}V_C\} \le -\Phi(\mathcal{C} \cap \mathcal{B}V_C).$$

Moreover $\mathcal{C} \cap \mathcal{B}V_C$ being σ_w-compact and disjoint from Δ, we get

$$\Phi(\mathcal{C} \cap \mathcal{B}V_C) > 1.$$

This together with (5.3) yields then

$$\limsup_{n\to\infty} \frac{1}{A_n} \log \mathbb{P}\{\eta_{n,m} \in \mathcal{C}\} \le \max(-2, -\Phi(\mathcal{C} \cap \mathcal{B}V_C)) = -(1+\delta), \ \delta > 0.$$

Therefore, for n large enough,

$$\mathbb{P}\{\eta_{n,m} \in \mathcal{C}\} \le \mathrm{e}^{-(1+\delta/2)A_n},$$

which yields

$$\mathbb{P}\{K_n \not\subset \Delta^\epsilon\} \le n\mathrm{e}^{-(1+\delta/2)A_n}.$$

Using the fact that $A_n \ge \log n - 1$, one gets:

$$\mathbb{P}\{K_n \not\subset \Delta^\epsilon\} \le \kappa\, n^{-\delta/2}, \quad \text{where } \kappa = \exp(1 + \delta/2). \tag{5.6}$$

Introduce now the sequence of integers $n_k = \max\{n;\ A_n = k\}$. Recalling that $A_n = [\log n]$, we see that n_k is properly defined for k large enough. Moreover, $\mathrm{e}^k \le n_k < \mathrm{e}^{(k+1)}$, and $K_n \subset K_{n_k}$ for $n_{k-1} < n \le n_k$. Thus (5.4) is equivalent to

$$\mathbb{P}\{K_{n_k} \not\subset \Delta^\epsilon, \text{i.o. (in } k)\} = 0. \tag{5.7}$$

By (5.6), we obtain for k large enough (say $k \ge k_0$):

$$\sum_{k=k_0}^{\infty} P\{K_{n_k} \not\subset \Delta^\epsilon\} \le \kappa \sum_{k=k_0}^{\infty} n_k^{-\delta/2} \le \kappa \sum_{k=k_0}^{\infty} \exp\left(-\frac{k\delta}{2}\right) < \infty.$$

Therefore, (5.7) follows by Borel-Cantelli. This in turn yields (5.4).

• **Step 3.** Let us prove now that $\Delta \subset K_n^\epsilon$ ultimately with probability one. Let $f \in \Delta$ and $\epsilon > 0$. Consider the set

$$\mathcal{N}_\epsilon(f) = \{g \in \mathcal{B}V_C,\ d(f,g) < \epsilon\}.$$

The set $\mathcal{N}_\epsilon(f)$ is d-open in $\mathcal{B}V_C$. We want to prove:

$$-(1-\delta) \le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{Z}_n \in \mathcal{N}_\epsilon(f)). \tag{5.8}$$

One can not apply directly the LD lower bound since $\mathcal{N}_\epsilon(f)$ is not σ_w-open as a subset of $\mathcal{B}V$. However, there exists a σ_w-open set G in $\mathcal{B}V$ such that $G \cap \mathcal{B}V_C = \mathcal{N}_\epsilon(f)$. In particular $(1-\alpha)f \in G$ for some $\alpha > 0$. As $\Phi(0) = 0$ (recall that $\mathbb{E}Z_i^n = 0$),

$$\Phi((1-\alpha)f) \le (1-\alpha)\Phi(f) + \alpha\Phi(0) \le 1 - \alpha < 1.$$

Therefore, $\Phi(G) = 1 - \delta < 1$ for some $\delta > 0$. As $G \subset \mathcal{N}_\epsilon(f) \cup \mathcal{B}V_C^c$, the LD lower bound (Th. 4.3) yields:

$$-(1-\delta) \le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\bar{Z}_n \in G\}$$
$$\le \max\left(\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{Z}_n \in \mathcal{N}_\epsilon(f)), \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{Z}_n \notin \mathcal{B}V_C)\right).$$

Recall that $\limsup_n \frac{1}{n} \log \mathbb{P}(\bar{Z}_n \notin \mathcal{B}V_C) \le -2$ by (5.3). Thus (5.8) is proved.

Let us introduce $R_n = [(n - A_n)/A_n]$ and the corresponding set $M_n = \{\eta_{n,rA_n}, \ 1 \le r \le R_n\}$. Observe that $M_n \subset K_n$. Moreover, $\mathbb{P}\{\eta_{n,rA_n} \in \mathcal{N}_\epsilon(f)\} \ge \exp(-A_n(1 - \delta/2))$ for n large enough. Therefore,

$$
\begin{aligned}
\mathbb{P}\{f \notin M_n^{2\epsilon}\} &= \mathbb{P}\left\{\cap_{1 \le r \le R_n}\{\eta_{n,rA_n} \notin \mathcal{N}_{2\epsilon}(f)\}\right\} \\
&\le \left(1 - \inf_{1 \le r \le R_n} \mathbb{P}(\eta_{n,rA_n} \in \mathcal{N}_{2\epsilon}(f))\right)^{R_n} \\
&\le \exp\left(-R_n \inf_{1 \le r \le R_n} \mathbb{P}(\eta_{n,rA_n} \in \mathcal{N}_{2\epsilon}(f))\right) \\
&\le \exp\left(-R_n \exp(-A_n(1 - \delta/2))\right) \\
&\le \exp(-n^{\delta/3})
\end{aligned}
\tag{5.9}
$$

for n sufficiently large.

To prove the full statement, recall that Δ is d-compact in $\mathcal{B}V_C$ thus there exist $f_1, \cdots, f_k \in \Delta$ such that $\Delta \subset \cup_{i=1}^k \mathcal{N}_\epsilon(f_i)$. Therefore,

$$
\begin{aligned}
\mathbb{P}\{\Delta \not\subset M_n^\epsilon\} &\le \mathbb{P}\{\cup_{i=1}^k \mathcal{N}_\epsilon(f_i) \not\subset M_n^\epsilon\} \\
&\le \sum_{i=1}^k \mathbb{P}\{\mathcal{N}_\epsilon(f_i) \not\subset M_n^\epsilon\} \\
&\le \sum_{i=1}^k \mathbb{P}\{f_i \not\subset M_n^{2\epsilon}\} \\
&\le k \exp(-n^{\delta/3})
\end{aligned}
$$

for n sufficiently large where the last inequality follows by (5.9). Therefore Borel-Cantelli's lemma implies that

$$
\mathbb{P}\{\Delta \not\subset M_n^\epsilon, \ \text{i.o.}\} = 0.
$$

As $M_n \subset K_n$, the same conclusion holds for K_n:

$$
\mathbb{P}\{\Delta \not\subset K_n^\epsilon, \ \text{i.o.}\} = 0. \qquad \square
$$

References

[1] G. Ben Arous, A. Dembo and A. Guionnet, Aging of spherical spin glasses. *Probab. Theory Related Fields* **120** (2001) 1–67.

[2] K.A. Borovkov, The functional form of the Erdős-Rényi law of large numbers. *Teor. Veroyatnost. i Primenen.* **35** (1990) 758–762.

[3] Z. Chi, The first-order asymptotic of waiting times with distortion between stationary processes. *IEEE Trans. Inform. Theory* **47** (2001) 338–347.

[4] Z. Chi, Stochastic sub-additivity approach to the conditional large deviation principle. *Ann. Probab.* **29** (2001) 1303–1328.

[5] I. Csiszár, Sanov property, generalized I-projection and a conditionnal limit theorem. *Ann. Probab.* **12** (1984) 768–793.

[6] D.A. Dawson and J. Gärtner, Multilevel large deviations and interacting diffusions. *Probab. Theory Related Fields* **98** (1994) 423–487.

[7] D.A. Dawson and J. Gärtner, Analytic aspects of multilevel large deviations, in *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)*. North-Holland, Amsterdam (1998) 401–440.

[8] P. Deheuvels, Functional Erdős-Rényi laws. *Studia Sci. Math. Hungar.* **26** (1991) 261–295.

[9] A. Dembo and I. Kontoyiannis, The asymptotics of waiting times between stationary processes, allowing distortion. *Ann. Appl. Probab.* **9** (1999) 413–429.

[10] A. Dembo and T. Zajic, Large deviations: from empirical mean and measure to partial sums process. *Stochastic Process. Appl.* **57** (1995) 191–224.

[11] A. Dembo and O. Zeitouni, *Large Deviations Techniques And Applications.* Springer-Verlag, New York, second edition (1998).

[12] J. Dieudonné, *Calcul infinitésimal.* Hermann, Paris (1968).

[13] H. Djellout, A. Guillin and L. Wu, Large and moderate deviations for quadratic empirical processes. *Stat. Inference Stoch. Process.* **2** (1999) 195–225.

[14] R.M. Dudley, *Real Analysis and Probability.* Wadsworth and Brooks/Cole (1989).

[15] R.S. Ellis, J. Gough and J.V. Pulé, The large deviation principle for measures with random weights. *Rev. Math. Phys.* **5** (1993) 659–692.

[16] P. Erdős and A. Rényi, On a new law of large numbers. *J. Anal. Math.* **23** (1970) 103–111.

[17] F. Gamboa and E. Gassiat, Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann. Statist.* **25** (1997) 328–350.

[18] N. Gantert, Functional Erdős-Renyi laws for semiexponential random variables. *Ann. Probab.* **26** (1998) 1356–1369.

[19] G. Högnäs, Characterization of weak convergence of signed measures on $[0, 1]$. *Math. Scand.* **41** (1977) 175–184.

[20] C. Léonard and J. Najim, An extension of Sanov's theorem: application to the Gibbs conditioning principle. *Bernoulli* **8** (2002) 721–743.

[21] J. Lynch and J. Sethuraman, Large deviations for processes with independent increments. *Ann. Probab.* **15** (1987) 610–627.

[22] J. Najim, A Cramér type theorem for weighted random variables. *Electron. J. Probab.* **7** (2002) 32 (electronic).

[23] R.T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton (1970).

[24] R.T. Rockafellar, Integrals which are convex functionals, II. *Pacific J. Math.* **39** (1971) 439–469.

[25] R.T. Rockafellar and R.J-B. Wets, *Variational Analysis.* Springer (1998).

[26] G.R. Sanchis, Addendum: "A functional limit theorem for Erdős and Rényi's law of large numbers". *Probab. Theory Related Fields* **99** (1994) 475.

[27] G.R. Sanchis, A functional limit theorem for Erdős and Rényi's law of large numbers. *Probab. Theory Related Fields* **98** (1994) 1–5.

[28] P.H. Schuette, Large deviations for trajectories of sums of independent random variables. *J. Theoret. Probab.* **7** (1994) 3–45.

[29] S.L. Zabell, Mosco convergence and large deviations, in *Probability in Banach spaces, 8 (Brunswick, ME, 1991).* Birkhäuser Boston, Boston, MA, *Progr. Probab.* **30** (1992) 245–252.