Bernoulli **20**(4), 2014, 1845–1878
DOI: 10.3150/13-BEJ544

Large deviations for bootstrapped empirical measures

JOSÉ TRASHORRAS* and OLIVIER WINTENBERGER**

Université Paris–Dauphine, Ceremade, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France. E-mail: *xose@ceremade.dauphine.fr; **wintenberger@ceremade.dauphine.fr

We investigate the Large Deviations (LD) properties of bootstrapped empirical measures with exchangeable weights. Our main results show in great generality how the resulting rate functions combine the LD properties of both the sample weights and the observations. As an application, we obtain new LD results and discuss both conditional and unconditional LD-efficiency for many classical choices of entries such as Efron's, leave-p-out, i.i.d. weighted, k-blocks bootstraps, etc.

Keywords: exchangeable bootstrap; large deviations

1. Introduction

In 1979, in a landmark paper [14], Efron proposed the following idea: When given a realization $x_1^n, \dots, x_n^n$ of random variables $X_1^n, \dots, X_n^n$ one can easily obtain "additional data" by sampling independent and $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$-distributed random variables $X_1^*, \dots, X_m^*, \dots$. It amounts to sample with replacement from an urn which composition is described by $\frac{1}{n}\sum_{i=1}^n \delta_{x_i^n}$. Often, this is computationally cheap and theoretical studies are available to assess the quality of the distribution of $\frac{1}{m}\sum_{i=1}^m \delta_{X_i^*}$ in approximating the distribution of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}$ which makes it all worthwhile. It was soon noticed that this urn procedure is not the only possible one: It can be generalized so that

$$\mathcal{L}^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{x_i^n} \tag{1.1}$$

is the right object to consider once it is assumed that the sampling weights $(W_1^n, \dots, W_n^n)$ are positive n-exchangeable random variables such that $\sum_{i=1}^n W_i^n = n$. Let us recall that $(W_1^n, \dots, W_n^n)$ is n-exchangeable means that for every element σ of $\mathfrak{S}_n$ the set of permutations of $\{1, \dots, n\}$ both $(W_1^n, \dots, W_n^n)$ and $(W_{\sigma(1)}^n, \dots, W_{\sigma(n)}^n)$ have the same distribution. For example, Efron's bootstrap corresponds to $(W_1^n, \dots, W_n^n)$ distributed according to a Multinomial law. A rich literature started flourishing on the ground of this idea that developed into two complementary directions: "conditional" results where $x_1^n, \dots, x_n^n$ are fixed observations filling some conditions and $\mathcal{L}^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{x_i^n}$ is considered and "unconditional" results where the $x_1^n, \dots, x_n^n$ are allowed to fluctuate and $L^n = \frac{1}{n}\sum_{i=1}^n W_i^n \delta_{X_i^n}$ is considered instead, see, for example, [2,17,19].

Here we are concerned with the LD properties of $(\mathcal{L}^n)_{n\ge1}$ and $(L^n)_{n\ge1}$. Let us recall that a sequence of random variables $(Y^n)_{n\ge1}$ taking values on a topological space $\mathcal{Y}$ obeys a Large

Deviation Principle (LDP) with rate function I if I is a nonnegative, lower semi-continuous function defined on $\mathcal{Y}$ such that

$$-\inf_{y\in A^o} I(y) \le \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}\big(Y^n \in A\big) \le \limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}\big(Y^n \in A\big) \le -\inf_{y\in \bar{A}} I(y)$$

for any measurable set $A \subset \mathcal{Y}$, whose interior is denoted by A^o and closure by $\bar{A}$. If the level sets $\{y : I(y) \le \alpha\}$ are compact for every $\alpha < \infty$, I is called a good rate function. For a background on the theory of large deviations, see Dembo and Zeitouni [9] and references therein. In the present paper, we prove under natural conditions on $((W_1^n, \dots, W_n^n))_{n\ge1}$ and $((x_1^n, \dots, x_n^n))_{n\ge1}$ or $((X_1^n, \dots, X_n^n))_{n\ge1}$ that $(\mathcal{L}^n)_{n\ge1}$ and $(L^n)_{n\ge1}$ obey a LDP and show how the resulting rate functions combine the LD properties of the aforementioned random variables.

Classically a bootstrap scheme is said to be efficient when it mimics the behavior of $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}$ and one distinguishes between "conditional efficiency" and "unconditional efficiency". Following Barbe and Bertail [2], we say that a bootstrap scheme is LD-efficient when the bootstrapped empirical measure has the same LD properties as the original empirical measure. It is a very strong property: thinking of percentile bootstrap's confidence intervals, LD-efficiency says that the relative coverage accuracy tends to 1 exponentially fast. It is related to the notion of asymptotic efficiency as described in [1]. Applying our general approach, we obtain new LD results and discuss both conditional and unconditional LD-efficiency for many classical choices of $((W_i^n)_{1\le i\le n})_{n\ge1}$ such as Efron's, leave-p-out, i.i.d. weighted, k-blocks bootstraps, etc. Another possible application is the quantification of the performance of importance sampling algorithms, see Hult and Nyquist [21].

Outline of the paper

The paper is structured as follows: In Section 2, we describe our results. As an application, in Section 3 we discuss LD efficiency issues. To this end, we need sample weights LDPs which are not the main concern of our work. This is the reason why their proofs are given in the supplemental article [26]. Section 4 is devoted to the proof of the central result of our paper (Theorem 2.1 below). Section 5 is concerned with the proof of Theorem 2.2. Finally, Section 6 contains the proof of all other results.

2. Statement of the results

2.1. Notations

Wasserstein distances will play a key role in our paper. Given any Polish space (E, d) we denote by $M_1(E)$ the set of Borel probability measures on E. We further define on the so-called Wasserstein space

$$\mathcal{W}^1(E) = \left\{\rho \in M_1(E) : \text{there exists some } y \in E \text{ such that } \int_E d(x, y)\rho(\mathrm{d}x) < \infty\right\},$$

the Wasserstein distance

$$W_1^d(\rho,\gamma)=\inf_{\pi\in\mathcal{C}(\rho,\gamma)}\left\{\int_{E\times E} d(x,y)\pi(\mathrm{d}x,\mathrm{d}y)\right\},$$

where $\mathcal{C}(\rho,\gamma)$ is the subset of $M_1(E\times E)$ of couplings of ρ and γ, that is, the set of Borel probability measures π on $E\times E$ such that their first marginal π_1 is ρ and second marginal π_2 is γ. If d is a bounded distance, then $\mathcal{W}^1(E)=M_1(E)$ and W_1^d coincides with the weak-convergence topology (see, e.g., Theorem 7.12 in [27]). We shall denote by $\stackrel{w}{\to}$ (resp. $\stackrel{W_1^d}{\to}$) the weak (resp. W_1^d) convergence of probability measures on $M_1(E)$ (resp. $\mathcal{W}^1(E)$). Let us recall that $\mu^n\stackrel{w}{\to}\mu$ if and only if for every real-valued, bounded and continuous application f defined on E we have $\int_E f(x)\mu^n(\mathrm{d}x)\to\int_E f(x)\mu(\mathrm{d}x)$. Hence, by introducing W_1^ζ (resp. W_1^λ, W_1^χ), the Wasserstein distance defined on $M_1(\mathbb{R}_+)$ (resp. $M_1(E)$, $M_1(\mathbb{R}_+\times E)$) when $\mathbb{R}_+$ (resp. E, $\mathbb{R}_+\times E$) is furnished with the bounded distance

$$\zeta(x,z)=\frac{\beta(x,z)}{1+\beta(x,z)}\qquad\left(\text{resp. } \lambda(y,t)=\frac{d(y,t)}{1+d(y,t)},\chi\big((x,y),(z,t)\big)=\zeta(x,z)+\lambda(y,t)\right),$$

where $\beta(x,z)=|x-z|$ is the usual Euclidean distance on $\mathbb{R}_+$, we obtain a user-friendly distance compatible with the weak-convergence topology.

Key to our paper are the set

$$\mathcal{M}_1^1(\mathbb{R}_+\times E)=\left\{\rho\in M_1(\mathbb{R}_+\times E):\int_{\mathbb{R}_+} w\rho_1(\mathrm{d}w)=1\right\}$$

endowed with the distance $\Delta(\rho,\gamma)=W_1^\beta(\rho_1,\gamma_1)+W_1^\chi(\rho,\gamma)$ and the map F defined by

$$\begin{aligned}F:\mathcal{M}_1^1(\mathbb{R}_+\times E)&\to M_1(E),\\ \rho(\mathrm{d}w,\mathrm{d}x)&\mapsto\int_{\mathbb{R}_+} w\rho(\mathrm{d}w,\mathrm{d}x).\end{aligned}\tag{2.1}$$

Indeed, $\mathcal{L}^n=F(\frac{1}{n}\sum_{i=1}^n\delta_{(W_i^n,x_i^n)})$ and F is continuous when $\mathcal{M}_1^1(\mathbb{R}_+\times E)$ is equipped with Δ and $M_1(E)$ with the weak convergence topology. We also consider the natural set of the sampling weights distributions

$$M_1^1(\mathbb{R}_+)=\left\{\rho\in M_1(\mathbb{R}_+),\int_{\mathbb{R}_+} w\rho(\mathrm{d}w)=1\right\}.$$

Finally, for any two probabilities ρ,ν on a measurable space $(E,\mathcal{E})$ we denote by

$$H(\nu|\rho)=\begin{cases}\displaystyle\int_E \mathrm{d}\nu\log\frac{\mathrm{d}\nu}{\mathrm{d}\rho}, & \text{if } \nu\ll\rho,\\ +\infty, & \text{otherwise,}\end{cases}$$

the relative entropy of ν with respect to ρ. To any $\rho(\mathrm{d}w,\mathrm{d}x)\in M_1(\mathbb{R}_+\times E)$, we associate $\rho_x(\mathrm{d}w)\in M_1(\mathbb{R}_+)$ (resp. $\rho_w(\mathrm{d}x)\in M_1(E)$) a stochastic kernel which is the conditional distribution of the first (resp. second) marginal of ρ given the second (resp. first). We summarize this by $\rho(\mathrm{d}w,\mathrm{d}x)=\rho_x(\mathrm{d}w)\otimes\rho_2(\mathrm{d}x)$ (resp. $\rho(\mathrm{d}w,\mathrm{d}x)=\rho_1(\mathrm{d}w)\otimes\rho_w(\mathrm{d}x)$). If $\nu,\gamma\in M_1(\mathbb{R}_+\times E)$ are such that $\nu_1=\gamma_1=\theta$ (resp. $\nu_2=\gamma_2=\theta$), then

$$H(\nu|\gamma)=\int_{\mathbb{R}_+}H(\nu_w|\gamma_w)\theta(\mathrm{d}w) \tag{2.2}$$

(resp. $H(\nu|\gamma)=\int_E H(\nu_x|\gamma_x)\theta(\mathrm{d}x)$), see Lemma 1.4.3 in [13].

2.2. Main results

We are given a triangular array $((W_i^n)_{1\le i\le n})_{n\ge1}$ of $\mathbb{R}_+$-valued random variables defined on a probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that

(H1) For every $n\ge1$, we have $\sum_{i=1}^n W_i^n=n$ and $(W_1^n,\dots,W_n^n)$ is n-exchangeable.
(H2) The sequence $(\mathcal{S}^n=\frac{1}{n}\sum_{i=1}^n\delta_{W_i^n})_{n\ge1}$ satisfies a LDP on $M_1^1(\mathbb{R}_+)$ endowed with the distance W_1^β with good rate function I^W.

We are further given a triangular array $((x_i^n)_{1\le i\le n})_{n\ge1}$ of elements of (E,d) such that

(H3) $\mu^n=\frac{1}{n}\sum_{i=1}^n\delta_{x_i^n}\xrightarrow{w}\mu\in M_1(E)$.

We are first interested in the LD behavior of

$$\mathcal{V}^n=\frac{1}{n}\sum_{i=1}^n\delta_{(W_i^n,x_i^n)}.$$

Theorem 2.1. *Under* (H1–H2–H3) *the sequence* $(\mathcal{V}^n)_{n\ge1}$ *satisfies a LDP on* $\mathcal{M}_1^1(\mathbb{R}_+\times E)$ *endowed with the distance* Δ *with good rate function*

$$\mathcal{J}(\rho;\mu)=\begin{cases}H(\rho|\rho_1\otimes\mu)+I^W(\rho_1), & \text{if } \rho_2=\mu,\\ +\infty, & \text{otherwise}.\end{cases}$$

We prove Theorem 2.1 in Section 4. It is the keystone of the paper. Indeed, since

$$\mathcal{L}^n=\frac{1}{n}\sum_{i=1}^n W_i^n\delta_{x_i^n}=F(\mathcal{V}^n)$$

with F defined in (2.1) a LDP for $(\mathcal{L}^n)_{n\ge1}$ immediately follows from Theorem 2.1 by contraction (see Theorem 4.2.1 in [9]) since Δ makes F a continuous map.

Corollary 2.1. *Under* (H1–H2–H3) *the sequence* $(\mathcal{L}^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\begin{aligned}\mathcal{K}(\nu;\mu) &= \inf_{\rho:F(\rho)=\nu} \mathcal{J}(\rho;\mu)\\ &= \inf_{\rho_x:F(\rho_x\otimes\mu)=\nu}\left\{\int_E H(\rho_x|\rho_1)\mu(\mathrm{d}x)+I^W(\rho_1)\right\}.\end{aligned}$$

Next, we allow the x_i^n's to fluctuate and consider a triangular array $((X_i^n)_{1\leq i\leq n})_{n\geq 1}$ of E-valued random variables defined on $(\Omega,\mathcal{A},\mathbb{P})$ such that

(H4) The sequence $(\mathcal{O}^n=\frac{1}{n}\sum_{i=1}^n\delta_{X_i^n})_{n\geq 1}$ satisfies a LDP on $M_1(E)$ endowed with the weak convergence topology with good rate function I^X.
(H5) For every $n\geq 1$, the vectors $(X_1^n,\dots,X_n^n)$ and $(W_1^n,\dots,W_n^n)$ are independent.

A LDP for

$$V^n=\frac{1}{n}\sum_{i=1}^n\delta_{(W_i^n,X_i^n)}$$

holds as a consequence of Theorem 2.1 and Theorem 2.3 in [18].

Theorem 2.2. *Under* (H1–H2–H4–H5) *the sequence* $(V^n)_{n\geq 1}$ *satisfies a LDP on* $\mathcal{M}_1^1(\mathbb{R}_+\times E)$ *endowed with the distance* Δ *with good rate function*

$$J(\rho)=H(\rho|\rho_1\otimes\rho_2)+I^W(\rho_1)+I^X(\rho_2). \tag{2.3}$$

Theorem 2.2 is proved in Section 5. Again, by contraction, a LDP for

$$L^n=\frac{1}{n}\sum_{i=1}^n W_i^n\delta_{X_i^n}=F\big(V^n\big)$$

holds as an immediate consequence of Theorem 2.2.

Corollary 2.2. *Under* (H1–H2–H4–H5) *the sequence* $(L^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$K(\nu)=\inf_{\rho:F(\rho)=\nu}J(\rho)=\inf_{\rho_2\in M_1(E)}\big\{\mathcal{K}(\nu;\rho_2)+I^X(\rho_2)\big\}, \tag{2.4}$$

where

$$\mathcal{K}(\nu;\rho_2)=\inf_{\rho_x:F(\rho_x\otimes\rho_2)=\nu}\left\{\int_E H(\rho_x|\rho_1)\rho_2(\mathrm{d}x)+I^W(\rho_1)\right\}.$$

It follows that for every $\nu\in M_1(E)$ *we have* $K(\nu)\leq I^X(\nu)$.

The latter inequality somehow quantifies the fact that the presence of the sampling weights offers more possibilities to the bootstrapped empirical measure L^n to get close to any ν than it is the case for $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n}$. We shall see in most examples that for classical choices of $((W_i^n)_{1\le i\le n})_{n\ge 1}$ and/or $((X_i^n)_{1\le i\le n})_{n\ge 1}$ there exists at least one $\nu \in M_1(E)$ such that $K(\nu) < I^X(\nu)$. Nevertheless, we have the following corollary.

Corollary 2.3. *Under* (H1–H2–H4–H5) *a necessary and sufficient condition on I^W to ensure that for every $((X_i^n)_{1\le i\le n})_{n\ge 1}$ and every $\nu \in M_1(E)$ we have $K(\nu) = I^X(\nu)$ is that for every $\nu, \zeta \in M_1(E)$*

$$\mathcal{K}(\nu;\zeta) = \begin{cases} 0, & \text{if } \nu = \zeta, \\ +\infty, & \text{otherwise}. \end{cases} \tag{2.5}$$

Corollary 2.3 is proved in Section 6.1.

2.3. Analysis of the rate functions

2.3.1. *Convexity issues*

It is a natural question in LD analysis to wonder if the obtained rate functions are convex.

Proposition 2.1. *If I^W is convex then for every $\mu \in M_1(E)$ the rate functions $\mathcal{J}(\cdot;\mu)$ and $\mathcal{K}(\cdot;\mu)$ are convex as well.*

Proposition 2.1 is proved in Section 6.2. Since the so-called mutual information $\rho \mapsto H(\rho|\rho_1 \otimes \rho_2)$ is neither convex nor concave (see, e.g., Theorem 2.7.4 in [8]) the convexity properties of J and K are less clear. However, if there exists some $\xi \in M_1(\mathbb{R}_+)$ and $\theta \in M_1(E)$ such that for every $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ $I^W(\rho_1) = H(\rho_1|\xi)$ and $I^X(\rho_2) = H(\rho_2|\theta)$ then $J(\rho) = H(\rho|\xi \otimes \theta)$ which is a convex function.

2.3.2. *Representation of the rate functions*

We shall often see in applications that there exists a $\xi \in M_1(\mathbb{R}_+)$ such that for every $\nu \in M_1^1(\mathbb{R}_+)$ we have $I^W(\nu) = H(\nu|\xi)$. Consider the logarithmic moment generating function of ξ defined on $\mathbb{R}$ by

$$\Lambda_\xi(\alpha) = \log \int_{\mathbb{R}_+} e^{\alpha x} \xi(dx). \tag{2.6}$$

Necessarily $\mathcal{D}_{\Lambda_\xi} = \{\alpha \in \mathbb{R}, \Lambda_\xi(\alpha) < \infty\}$ is an interval which supremum is denoted a. We further consider for every $\theta \in \mathcal{D}_{\Lambda_\xi}$ the probability measure ξ^θ defined on $\mathbb{R}_+$ by

$$\xi^\theta(dx) = \frac{e^{\theta x}}{\int_{\mathbb{R}_+} e^{\theta u}\xi(du)}\xi(dx) \quad \text{and} \quad \Lambda_\xi^*(x) = \sup_{\alpha\in\mathbb{R}}\{\alpha x - \Lambda_\xi(\alpha)\} \tag{2.7}$$

the Fenchel–Legendre transform of Λ_ξ. We obtain

Proposition 2.2. *If there exists a* $\xi \in M_1(\mathbb{R}_+)$ *such that for every* $\nu \in M_1^1(\mathbb{R}_+)$ $I^W(\nu) = H(\nu|\xi)$, *then for every* $\nu, \mu \in M_1(E)$ *we have*

$$\mathcal{K}(\nu;\mu) = \begin{cases} \int_E \Lambda_\xi^*\left(\frac{d\nu}{d\mu}(x)\right)\mu(dx) + \inf_{\substack{\rho:\rho_2=\mu \\ F(\rho)=\nu}} \int_{\mathbb{R}_+} \inf_{\theta\in\mathcal{D}_{\Lambda_\xi}} H\left(\rho_x|\xi^\theta\right)\mu(dx), \\ \qquad\qquad \textit{if } \nu \ll \mu, \\ +\infty, \qquad \textit{otherwise}. \end{cases} \tag{2.8}$$

The preceding equality reduces to

$$\mathcal{K}(\nu;\mu) = \begin{cases} \int_E \Lambda_\xi^*\left(\frac{d\nu}{d\mu}(x)\right)\mu(dx), & \textit{if } \nu \ll \mu, \\ +\infty, & \textit{otherwise}, \end{cases} \tag{2.9}$$

for every $\nu, \mu \in M_1(E)$ *if and only if* Λ_ξ *satisfies*

$$\lim_{\alpha\to-\infty} \Lambda'_\xi(\alpha) = 0 \quad \textit{and} \quad \lim_{\alpha\to a} \Lambda'_\xi(\alpha) = +\infty. \tag{2.10}$$

Proposition 2.2 is proved in Section 6.3. The rate function in (2.8) is the sum of a "regular" part – $\int_E \Lambda_\xi^*(d\nu/d\mu)$ – and a "singular" part. The latter cancels out for every ν, μ if and only for every $\nu, \mu \in M_1(E)$ one can find a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ such that $\rho_2 = \mu$, $F(\rho) = \nu$ and for μ a.e. $x \in E \rho_x$ is built on the ground of ξ via an exponential change of measure as in the proof of Cramér's theorem lower bound. Condition (2.10) is a necessary and sufficient condition on the distribution of ξ for this to hold. A typical situation where (2.10) fails to be satisfied is when ξ has bounded support.

There are references in the LD literature where, as in (2.8), the rate function can be decomposed as the sum of a "regular" part and a "singular" part. In [24], a LDP is established for $((1/n)\sum_{i=1}^n \xi_i \delta_{x_i^n})_{n\ge1}$ where $(\xi_i)_{i\ge1}$ is a sequence of independent and identically distributed random variables, $((x_i^n)_{1\le i\le n})_{n\ge1})$ satisfies (H3) and μ is a stricly positive measure. It is demonstrated there that if ξ_1 fails to have all its exponential moments finite then an extra "singular" part is added to the "regular" part of the rate function $\int_E \Lambda_{\xi_1}^*(d\nu/d\mu)$. This phenomenon is not of the same nature as the one we consider here. Indeed if, for example, ξ_1 has bounded support the rate function in [24] reduces to its regular part which is not the case here. See also [23] for a related situation with a singular component in a LD rate function.

3. Examples of applications

In this section, we investigate the LD properties of $(\mathcal{L}^n)_{n\ge1}$ and $(L^n)_{n\ge1}$ for several particular choices of $((W_i^n)_{1\le i\le n})_{n\ge1}$ and/or $((X_i^n)_{1\le i\le n})_{n\ge1}$. To this end, we need more notations: For every $\lambda, \gamma > 0$, we shall denote by $\mathcal{F}(\lambda,\gamma)$ the distribution of a random variable Y such that λY is $\mathcal{P}(\gamma)$ (Poisson)-distributed and $\mathcal{Q}(\lambda)$ denotes the distribution $\mathcal{F}(\lambda,\lambda)$. For every positive integers m and n and every n-tuple of nonnegative numbers $(p_1^n, \dots, p_n^n)$ such that $\sum_{i=1}^n p_i^n = 1$, we shall denote by $\mathrm{Mult}_n(m, (p_1^n, \dots, p_n^n))$ the (Multinomial) distribution of $(Y_1, \dots, Y_n)$ the numbers of balls found in n urns labeled $1, \dots, n$ when m balls are thrown in these urns independently, each having probability p_1^n to fall in the urn labeled 1, probability p_2^n to fall in the urn labeled 2, etc. For every positive integer n and $p \in]0, 1[$ we denote by $\mathfrak{B}(n, p)$ the Binomial distribution with parameters n and p.

3.1. Efron's bootstrap and "*m* out of *n*" bootstrap

For every $m, n \geq 1$ the weights $(W_1^n, \dots, W_n^n)$ for the "m out of n" bootstrap are defined such that $\frac{m}{n}(W_1^n, \dots, W_n^n)$ is $\mathrm{Mult}_n(m, (1/n, \dots, 1/n))$-distributed. Classical Efron's bootstrap corresponds to $m = n$. We shall assume that $m = m(n)$ and that the sequence $(\lambda_n = m(n)/n)_{n\geq 1}$ satisfies $\lim_{n\to\infty} \lambda_n = \lambda > 0$. Quite surprisingly we could not find in the literature a reference for the following, which proof is partly related to results in [15,22] and is given in [26].

Theorem 3.1. *The sequence* $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ *obeys a LDP on* $M_1^1(\mathbb{R}_+)$ *endowed with* W_1^β *with good rate function* $I^W(\gamma) = H(\gamma|\mathcal{Q}(\lambda))$.

Corollary 3.1. *Under* (H3) *the sequence* $(\mathcal{L}^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu; \mu) = \lambda H(\nu|\mu).$$

Corollary 3.1 is an immediate consequence of Proposition 2.2 since $\Lambda_{\mathcal{Q}(\lambda)}(\alpha) = \lambda(\mathrm{e}^{\alpha/\lambda} - 1)$ hence $\Lambda^*_{\mathcal{Q}(\lambda)}(x) = \lambda\Lambda^*_{\mathcal{P}(1)}(x)$.

Remark that by properly rescaling, we obtain for every $\lambda > 0$ and every measurable $A \subset M_1(E)$ the modified LD result:

$$\begin{aligned} -\inf_{\nu\in A^o} H(\nu|\mu) &\leq \liminf_{n\to\infty} \frac{1}{m(n)} \log \mathbb{P}(\mathcal{L}^n \in A) \\ &\leq \limsup_{n\to\infty} \frac{1}{m(n)} \log \mathbb{P}(\mathcal{L}^n \in A) \\ &\leq -\inf_{\nu\in\bar{A}} H(\nu|\mu). \end{aligned} \tag{3.1}$$

The "m out of n" bootstrap has the same conditional LD properties as the unconditional LD properties of the empirical measure of an $m(n)$-sample of independent and μ-distributed X_i's.

Next, we investigate the LD properties of $(L^n)_{n\geq 1}$ without any other assumption than (H4–H5). To this end, we introduce the set $\mathcal{Z} = \{\eta \in M_1(E) : I^X(\eta) = 0\}$. It follows from Corollaries 2.2 and 3.1 that produce the following corollary.

Corollary 3.2. *Under* (H4–H5) *the sequence* $(L^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function* K *such that*

$$K(\nu) = \inf_{\zeta\in M_1(E)}\left\{\lambda H(\nu|\zeta) + I^X(\zeta)\right\} \leq \lambda \inf_{\eta\in\mathcal{Z}} H(\nu|\eta).$$

Corollary 3.2 is proved in Section 6.4.

Now we consider some particular cases for $((X_i^n)_{1\leq i\leq n})_{n\geq 1}$. First, we assume that for every $n \geq 1$ the random variables $X_1^n, \dots, X_n^n$ are independent and identically μ-distributed. Then $\frac{1}{n}\sum_{i=1}^n \delta_{X_i^n} \stackrel{w}{\to} \mu$ a.s. (see Theorem 11.4.1 in [12]) and Corollary 3.1 can be interpreted as a conditional LDP. Hence, any "m out of n" bootstrap such that $\lim_{n\to\infty} m(n)/n = 1$ (in particular Efron's bootstrap) leads to a conditional LDP that coincides with the original LDP in this case. This was first established in [3] for Efron's Bootstrap and in [7] in the general case. Actually the $X_1^n, \dots, X_n^n$ need not be i.i.d., it is sufficient that the associated empirical measures satisfy a LDP with rate function $H(\cdot|\mu)$ for Efron's bootstrap to be conditionally LD-efficient.

Corollary 3.2 completes the previous result with an unconditional LDP. In this particular case $I^X(\zeta) = H(\zeta|\mu)$ and by taking, for example, $\mu = \frac{9}{10}\delta_0 + \frac{1}{10}\delta_1$, $\nu = \frac{1}{10}\delta_0 + \frac{9}{10}\delta_1$, $\zeta = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ and $\lambda = 1$ we observe that $K(\nu) \leq H(\nu|\zeta) + H(\zeta|\mu) < H(\nu|\mu)$ hence Efron's bootstrap is not unconditionally LD-efficient. Straightforward use of the same kind of arguments shows that this remark also holds true when $X_1^n, \dots, X_n^n$ is the result of sampling without replacement from an urn with suitable properties (see Theorem 7.2 in [9] for the reference LDP) or when $X_1^n, \dots, X_n^n$ are the n first components of an infinitely exchangeable sequence of random variables (see [10] for the reference LDP).

3.2. I.i.d. weighted bootstrap

The weights $(W_1^n, \dots, W_n^n)$ for an i.i.d.-weighted bootstrap are defined on the ground of a sequence $Y_1, \dots, Y_n, \dots$ of $\mathbb{R}_+$-valued independent random variables with common distribution ξ. We shall assume that for every $\alpha > 0$ we have $\Lambda_\xi(\alpha) < \infty$ and that $\Lambda_\xi^*(0) = \infty$ (or equivalently $\mathbb{P}(Y_1 = 0) = 0$). The weights $(W_1^n, \dots, W_n^n)$ are defined by

$$W_1^n = \frac{Y_1}{(1/n)\sum_{i=1}^n Y_i}, \qquad \dots, \qquad W_i^n = \frac{Y_i}{(1/n)\sum_{i=1}^n Y_i}, \qquad \dots,$$
$$W_n^n = \frac{Y_n}{(1/n)\sum_{i=1}^n Y_i}.$$

In order to describe the LD behavior of $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ we introduce the map

$$\begin{aligned}\mathcal{G}: \mathcal{W}^1(\mathbb{R}_+) \times \mathbb{R}_+^* &\to \mathcal{W}^1(\mathbb{R}_+),\\ (\gamma, m) &\mapsto \mathcal{G}(\gamma, m): A \in \mathcal{B}_{\mathbb{R}_+} \mapsto \gamma(mA).\end{aligned}$$

The continuity of $\mathcal{G}$ when $\mathcal{W}^1(\mathbb{R}_+)$ is endowed with W_1^β is the main argument in the proof of the following result given in [26].

Theorem 3.2. *The sequence* $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ *satisfies a LDP on* $M_1^1(\mathbb{R}_+)$ *endowed with* W_1^β *with good rate function*

$$I^W(\gamma) = \inf_{m>0}\bigl\{H\bigl(\gamma|\mathcal{G}(\xi,m)\bigr)\bigr\}.$$

Theorem 3.2 is proved in [26]. As a consequence of Proposition 2.2, we obtain the following corollary.

Corollary 3.3. *Under* (H3) *the sequence* $(\mathcal{L}^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu;\mu) = \begin{cases} \inf_{m>0}\left\{\int_E \Lambda_\xi^*\left(m\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x) + \inf_{\substack{\rho:\rho_2=\mu\\ F(\rho)=\nu}}\int_{\mathbb{R}_+}\inf_{\theta\in\mathbb{R}} H\bigl(\rho_x|\mathcal{G}(m,\xi)^\theta\bigr)\mu(\mathrm{d}x)\right\}, & \\ \qquad \textit{if } \nu \ll \mu, & \\ +\infty, \qquad \textit{otherwise}. & \end{cases}$$

Moreover, the preceding reduces to

$$\mathcal{K}(\nu;\mu) = \begin{cases} \inf_{m>0}\int_E \Lambda_\xi^*\left(m\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x), & \textit{if } \nu \ll \mu, \\ +\infty, & \textit{otherwise}, \end{cases}$$

for every $\nu \in M_1(E)$ *if and only if*

$$\lim_{\alpha\to-\infty}\Lambda_\xi'(\alpha) = 0 \quad \textit{and} \quad \lim_{\alpha\to+\infty}\Lambda_\xi'(\alpha) = +\infty.$$

Corollary 3.3 is proved in Section 6.5. Similar expression of the rate function are also given in [6,21].

It follows from the previous corollary that there is no distribution ξ such that for every $\nu,\mu \in M_1(E)$ the identity $\mathcal{K}(\nu;\mu) = H(\nu|\mu)$ holds. Indeed, as soon as there exists $\nu,\mu \in M_1(E)$ such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = 0$ on a set A such that $\mu(A) > 0$ one has $\mathcal{K}(\nu;\mu) = \infty$ while it could be possible that $H(\nu|\mu) < \infty$. In words there is no choice of ξ for which one gets a conditional LDP that coincides with the original one for $X_1^n,\dots,X_n^n$ independent and μ-distributed. It is clearly due to the fact that $\Lambda_\xi^*(0) = \infty$ forces all the weights $W_1^n,\dots,W_n^n$ to be positive which is to be compared to, for example, Efron's bootstrap. Finally, as for Efron's bootstrap, one can construct examples to show that in most classical cases the i.i.d.-bootstrap is not unconditionally LD-efficient.

3.3. The multivariate hypergeometric bootstrap

Let K be a fixed integer number such that $K \geq 2$. The multivariate hypergeometric bootstrap emerges from the following urn scheme: Put K copies of each observed data in an urn so that the urn contains Kn elements then draw from this urn a sample of size n without replacement. The sampling weights $(W_1^n, \dots, W_n^n)$ take their values in $\{0, 1, \dots, K\}$ under the constraint $\sum_{i=1}^n W_i^n = n$ and are distributed according to

$$\mathbb{P}\big(W_1^n = w_1^n, \dots, W_n^n = w_n^n\big) = \frac{C_K^{w_1^n} \cdots C_K^{w_n^n}}{C_{nK}^n}.$$

Theorem 3.3. *The sequence* $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ *satisfies a LDP on* $M_1^1(\mathbb{R}_+)$ *endowed with* W_1^β *with good rate function*

$$I^W(\gamma) = H\big(\gamma | \mathfrak{B}\big(K, K^{-1}\big)\big).$$

Theorem 3.3 is proved in [26]. It immediately follows from the latter result and Proposition 2.2 in the following.

Corollary 3.4. *Under* (H3) *the sequence* $(\mathcal{L}^n)_{n\geq 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu; \mu) = \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \left\{ \int_E H(\rho_x | \rho_1) \mu(\mathrm{d}x) + H\big(\rho_1 | \mathfrak{B}\big(K, K^{-1}\big)\big) \right\}.$$

Notice that $\mathfrak{B}(K, K^{-1})$ does not satisfy condition (2.10). Again, there is no integer K such that for every $\nu, \mu \in M_1(E)$ the identity $\mathcal{K}(\nu; \mu) = H(\nu|\mu)$ holds. Indeed, as soon as there exists $\nu, \mu \in M_1(E)$ such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) > K$ on a set A such that $\mu(A) > 0$ one has $\mathcal{K}(\nu; \mu) = \infty$ while it could be possible that $H(\nu|\mu) < \infty$. Thus, all multivariate hypergeometric bootstraps fail to be conditionally LD-efficients for i.i.d. observations. One can construct examples to show that in most classical cases the multivariate hypergeometric bootstrap fails to be unconditionally LD-efficient.

3.4. A bootstrap generated from deterministic weights

The weights for bootstrap schemes defined from deterministic weights are given by

$$\big(W_1^n, \dots, W_n^n\big) = \big(w_{\sigma_n(1)}^n, \dots, w_{\sigma_n(n)}^n\big),$$

where for every $n \geq 1$ the $w_1^n, \dots, w_n^n$ are fixed nonnegative real numbers such that $\sum_{i=1}^n w_i^n = n$ and

$$\frac{1}{n}\sum_{i=1}^n \delta_{w_i^n} \overset{W_1^\beta}{\to} \pi \in M_1^1(\mathbb{R}_+)$$

while σ_n is an uniformly over $\mathfrak{S}_n$ distributed random variable. Clearly the sequence $(\mathcal{S}^n = \frac{1}{n}\sum_{i=1}^n \delta_{W_i^n})_{n\geq 1}$ satisfies a LDP on $M_1^1(\mathbb{R}_+)$ endowed with W_1^β with good rate function

$$I^W(\gamma) = \begin{cases} 0, & \text{if } \gamma = \pi, \\ +\infty, & \text{otherwise.} \end{cases}$$

An important special case is the leave-p-out bootstrap, or delete-p jacknife. The leave-p-out bootstrap is generated by permuting the deterministic weights

$$(w_1^n, \dots, w_n^n) = \Bigl(\underbrace{\frac{n}{n-p}, \dots, \frac{n}{n-p}}_{n-p}, \underbrace{0, \dots, 0}_{p}\Bigr).$$

We shall take $p = p(n)$ such that $\lim_{n\to\infty} p(n)/n = \alpha \in [0, 1)$ so

$$\pi = (1-\alpha)\delta_{1/(1-\alpha)} + \alpha\delta_0.$$

Corollary 3.5. *If $\alpha > 0$, under* (H3) *the sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies a LDP on $M_1(E)$ endowed with the weak convergence topology with good rate function*

$$\mathcal{K}(\nu; \mu) = \begin{cases} (1-\alpha)H(\nu|\mu) + \alpha H\Bigl(\dfrac{\mu - (1-\alpha)\nu}{\alpha}\Big|\mu\Bigr), & \textit{if } \dfrac{\mu - (1-\alpha)\nu}{\alpha} \in M_1(E), \\ +\infty, & \textit{otherwise.} \end{cases}$$

If $\alpha = 0$ the sequence $(\mathcal{L}^n)_{n\geq 1}$ satisfies a LDP on $M_1(E)$ endowed with the weak convergence topology with good rate function

$$\mathcal{K}(\nu; \mu) = \begin{cases} 0, & \textit{if } \nu = \mu, \\ +\infty, & \textit{otherwise.} \end{cases}$$

Corollary 3.5 is proved in Section 6.6. Naturally this result coincides with Theorem 7.2.1 in [9]. Combining Corollary 2.2 and Corollary 3.5, we obtain an unconditional version of the latter result. To every $\nu \in M_1(E)$, we associate

$$\mathcal{E}_\nu = \Bigl\{\zeta \in M_1(E) : \frac{\zeta - (1-\alpha)\nu}{\alpha} \in M_1(E)\Bigr\}.$$

Corollary 3.6. *If $\alpha > 0$, under* (H4–H5) *the sequence $(L^n)_{n\geq 1}$ satisfies a LDP on $M_1(E)$ endowed with the weak convergence topology with good rate function $K(\nu) = \inf_{\zeta\in\mathcal{E}_\nu} \mathcal{U}(\nu, \zeta)$ where*

$$\mathcal{U}(\nu, \zeta) = (1-\alpha)H(\nu|\zeta) + \alpha H\Bigl(\frac{\zeta - (1-\alpha)\nu}{\alpha}\Big|\zeta\Bigr) + I^X(\zeta).$$

If $\alpha = 0$, the sequence $(L^n)_{n\geq 1}$ satisfies a LDP on $M_1(E)$ endowed with the weak convergence topology with good rate function $K(\nu) = I^X(\nu)$.

Corollary 3.6 is proved in Section 6.7. Thus when $\alpha = 0$, for example, the leave-$p(n)$-out bootstrap with $p(n) = o(n)$, the bootstrap is unconditionally LD-efficient.

3.5. The k-blocks bootstrap

To conclude, let us consider the (moving or circular) k-blocks bootstrap. Weights from the "$m = n/k$ out of n" bootstrap are such that $\frac{1}{k}(W_1^n, \dots, W_n^n)$ is $\mathrm{Mult}_n(m, (1/n, \dots, 1/n))$-distributed. The k-blocks bootstrapped empirical measure is defined as in [20] via the formula

$$\widetilde{\mathcal{L}}_n = \frac{1}{n} \sum_{i=1}^{n} W_i^n \frac{1}{k} \sum_{j \sim i} \delta_{x_j^n},$$

where $j \sim i$ means that the j belong to block i. For the moving k-blocks bootstrap, $j \sim i$ if $j \in \{i - k/2, \dots, i + k/2\}$ modulo n. We could also consider the circular k-blocks bootstrap where $j \sim i$ if $j \in \{i, \dots, i + k - 1\}$ modulo n. Both schemes are asymptotically equivalent as soon as k is fixed as it is the case here. Notice that

$$\widetilde{\mathcal{L}}_n = \frac{1}{n} \sum_{i=1}^{n} \widetilde{W_i^n} \delta_{x_i^n} \qquad \text{with } \widetilde{W_i^n} = \frac{1}{k} \sum_{j \sim i} W_j^n,$$

where $(\widetilde{W_1^n}, \dots, \widetilde{W_n^n})$ fails to be exchangeable. However, our approach relies on preliminary results like Theorem 2.1 that are general enough to allow us to handle this situation under some mild additional hypothesis. Indeed, assume that the observations $((x_i^n)_{1 \le i \le n})_{n \ge 1}$ satisfy

(H6) $\frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i^n, \dots, x_{i+k-1}^n)} \xrightarrow{w} \mu^{(k)} \in M_1(E^k)$

the i's being taken modulo n. We obtain the following theorem.

Theorem 3.4. *Under* (H6) *the sequence* $(\widetilde{\mathcal{L}}_n)_{n \ge 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\widetilde{\mathcal{K}}(\nu; \mu) = \inf\left\{ \frac{1}{k} H\left(\nu^{(k)} | \mu^{(k)}\right) : \nu^{(k)} \in M_1\left(E^k\right), \frac{1}{k} \sum_{i=1}^{k} \nu_i^{(k)} = \nu \right\}.$$

Theorem 3.4 is proved in Section 6.8. Condition (H6) is a.s. satisfied with $\mu^{(k)} = \mu^{\otimes k}$ when we are given the realization $x_1^n, \dots, x_n^n$ of independent and μ-distributed random variables $X_1^n, \dots, X_n^n$.

Corollary 3.7. *Under* (H6) *with* $\mu^{(k)} = \mu^{\otimes k}$ *the sequence* $(\widetilde{\mathcal{L}}_n)_{n \ge 1}$ *satisfies a LDP on* $M_1(E)$ *endowed with the weak convergence topology with good rate function*

$$\widetilde{\mathcal{K}}(\nu; \mu) = H(\nu | \mu).$$

Corollary 3.7 is proved in Section 6.9. The k-blocks bootstrap is thus conditionnally efficient in this case but fails to be unconditionally efficient for the same reason as Efron's bootstrap, see Section 3.1.

When we are given the realization $x_1^n, \dots, x_n^n$ of $X_1^n, \dots, X_n^n$ the first n components of a stationary Markov chains $(Y_i)_{i\ge1}$ with transition probability P and stationary measure μ as in [11], we get

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(X_i^n,\dots,X_{i+k-1}^n)} \xrightarrow{w} \mu\otimes\underbrace{P\otimes\cdots\otimes P}_{k-1} \qquad \text{a.s.}$$

so (H6) is satisfied a.s. with $\mu^{(k)} = \mu\otimes\underbrace{P\otimes\cdots\otimes P}_{k-1}$. Applying Theorem 3.4 with $\nu^{(k)} = \nu\otimes P'\otimes\cdots\otimes P'$ where P' belongs to the set $MC(\nu)$ of kernels of ergodic Markov chains with stationary measure ν, we obtain

$$\widetilde{\mathcal{K}}(\nu;\mu) \le \frac{1}{k}H(\nu|\mu) + \frac{k-1}{k}\inf_{P'\in MC(\nu)}\int H\bigl(P'|P\bigr)\,\mathrm{d}\nu.$$

It is interesting to note that this upper bound tends to the rate function in the classical LD of [11] when $k\to\infty$.

4. Proof of Theorem 2.1

First, we describe the ideas behind the proof of Theorem 2.1. Let us recall that we are given a triangular array $((x_i^n)_{1\le i\le n})_{n\ge1}$ of elements of E that satisfies (H3). Let A be a measurable subset of $\mathcal{M}_1^1(\mathbb{R}_+\times E)$. For every integer $n\ge1$, we have

$$\mathbb{P}\bigl(\mathcal{V}^n\in A\bigr) = \int_{M_1^1(\mathbb{R}_+)}\mathbb{P}\left(\mathcal{V}^n\in A\,\middle|\,\frac{1}{n}\sum_{i=1}^{n}\delta_{W_i^n} = \nu\right)Q^n(\mathrm{d}\nu),$$

where Q^n stands for the distribution of $\frac{1}{n}\sum_{i=1}^n\delta_{W_i^n}$. Let $(\sigma_n)_{n\ge1}$ be a sequence of random variables defined on $(\Omega,\mathcal{A},\mathbb{P})$ such that for every $n\ge1$ the distribution of σ_n is uniform over $\mathfrak{S}_n$ the set of permutations of $\{1,\dots,n\}$. Since (H1) holds $(W_1^n,\dots,W_n^n)$ is n-exchangeable hence for every $w_1^n,\dots,w_n^n\in\mathbb{R}_+$ such that $\sum_{i=1}^n w_i^n = n$ the distribution of $\mathcal{V}^n$ conditioned on $\frac{1}{n}\sum_{i=1}^n\delta_{W_i^n} = \frac{1}{n}\sum_{i=1}^n\delta_{w_i^n} = \nu^n$ admits the distribution of

$$T^n = \frac{1}{n}\sum_{i=1}^{n}\delta_{(w_{\sigma_n(i)}^n,x_i^n)}$$

as a regular version. Let us denote by $\mathbb{P}(T^n\in\cdot;\nu^n)$ this distribution and assume that for every converging sequence $\nu^n\xrightarrow{W_1^\beta}\nu$ it obeys a LDP in $\mathcal{M}_1^1(\mathbb{R}_+\times E)$ endowed with Δ with some good rate function $I(\cdot;\nu)$. Then the usual LD heuristic writes

$$\mathbb{P}\big(T^n \in A; \nu^n\big) \sim \exp\Big(-n \inf_{\rho\in A} I(\rho;\nu)\Big)$$

and clearly $I(\rho;\nu) = \infty$ if $\rho_1 \neq \nu$. This conditional LDP is turned into a rigorous statement in Theorem 4.1 below. It is the main step in the proof of Theorem 2.1. Indeed if we further assume that (H2) holds and that for every $\nu \in M_1^1(\mathbb{R}_+)$ there exists a sequence $(\nu^n)_{n\geq 1}$ of elements of $M_1^1(\mathbb{R}_+)$ such that $\nu^n \xrightarrow{W_1^\beta} \nu$ then we get

$$\begin{aligned}\mathbb{P}\big(\mathcal{V}^n \in A\big) &= \int_{M_1^1(\mathbb{R}_+)} \mathbb{P}\big(\mathcal{V}^n \in A; \nu\big) Q^n(\mathrm{d}\nu)\\ &\sim \int_{M_1^1(\mathbb{R}_+)} \exp\Big(-n \inf_{\rho\in A} I(\rho;\nu)\Big)\exp\big(-nI^W(\nu)\big)\,\mathrm{d}\nu \qquad (4.1)\\ &\sim \exp\Big(-n \inf_{\rho\in A}\big\{I(\rho;\rho_1) + I^W(\rho_1)\big\}\Big)\end{aligned}$$

since $I(\rho;\nu) = \infty$ if $\rho_1 \neq \nu$. Pasting LDPs as in the latter heuristic is turned into a rigorous statement by appealing to Theorem 2.3 in [18]. We detail on that in Section 4.2.1. The required approximation result on $(M_1^1(\mathbb{R}_+), W_1^\beta)$ is proved in Section 4.2.3.

So the proof of Theorem 2.1 is divided into two main steps: In Section 4.1, we establish a conditional (on the W_i^n's) LDP. Then in Section 4.2 we show how the latter result leads to Theorem 2.1.

4.1. A conditional LDP

All through Section 4.1 we are given a fixed triangular array $((w_i^n)_{1\leq i\leq n})_{n\geq 1}$ of elements of $\mathbb{R}_+$, possibly with repetition, such that

(H7) For every $n \geq 1$ we have $\sum_{i=1}^n w_i^n = n$ and $\nu^n = \frac{1}{n}\sum_{i=1}^n \delta_{w_i^n} \xrightarrow{W_1^\beta} \nu \in M_1^1(\mathbb{R}_+)$.

Let us recall that we are also given a fixed triangular array $((x_i^n)_{1\leq i\leq n})_{n\geq 1}$ of elements of E that satisfies (H3), that is, $\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^n} \xrightarrow{w} \mu \in M_1(E)$.

Theorem 4.1. *Under* (H3–H7) *the sequence* $(T^n)_{n\geq 1}$ *satisfies a LDP on* $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ *endowed with the distance* Δ *with good rate function*

$$I(\rho;\nu) = \begin{cases} H(\rho|\nu\otimes\mu), & \textit{if } \rho_1 = \nu \textit{ and } \rho_2 = \mu,\\ +\infty, & \textit{otherwise}.\end{cases} \qquad (4.2)$$

We shall proceed in two steps in proving Theorem 4.1. First, following the proof of Theorem 1 in [25], we prove in Section 4.1.1 that a LDP for $(T^n)_{n\geq 1}$ holds in $M_1(\mathbb{R}_+ \times E)$ with good rate function $I(\cdot;\nu)$ when $M_1(\mathbb{R}_+ \times E)$ is endowed with the weak convergence topology. Then, in Section 4.1.2, we strengthen this result up to a LDP in $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ furnished with the distance Δ.

4.1.1. *A LDP in the weak convergence topology*

All through Section 4.1.1 $M_1(\mathbb{R}_+ \times E)$ is endowed with the weak convergence topology. Let us introduce some more notations. For every $n \geq 1$, we denote by

$$\mathcal{P}_n = \left\{ \rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E) : \exists \sigma \in \mathfrak{S}_n, \rho = \frac{1}{n} \sum_{i=1}^n \delta_{(w_{\sigma(i)}^n, x_i^n)} \right\}$$

the set of possible values of T^n where $((x_i^n)_{1 \leq i \leq n})_{n \geq 1}$ is defined in (H3) and $((w_i^n)_{1 \leq i \leq n})_{n \geq 1}$ is defined in (H7). Since both $(\nu^n)_{n \geq 1}$ and $(\mu^n)_{n \geq 1}$ are fixed once for all in Section 4.1, $\mathcal{P}_n$ is fixed once for all in this section as well. We shall say that a sequence $(\rho^n)_{n \geq 1}$ of elements of $M_1(\mathbb{R}_+ \times E)$ satisfies Assumption (A1) if and only if

(A1) For every $n \geq 1$, we have $\rho^n \in \mathcal{P}_n$.

We further introduce two triangular arrays $((L_i^n)_{1 \leq i \leq n})_{n \geq 1}$ and $((R_i^n)_{1 \leq i \leq n})_{n \geq 1}$ of elements of $\mathbb{R}_+$ and E, respectively, defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and such that for every $n \geq 1$ the $2n$ random variables $L_1^n, \dots, L_n^n, R_1^n, \dots, R_n^n$ are mutually independent. We assume that every L_i^n (resp. R_i^n) is distributed according to ν^n (resp. μ^n). The sequence

$$\mathcal{T}^n = \frac{1}{n} \sum_{i=1}^n \delta_{(L_i^n, R_i^n)} \in M_1(\mathbb{R}_+ \times E)$$

has the following LD behavior.

Lemma 4.1. *The sequence $(\mathcal{T}^n)_{n \geq 1}$ satisfies a LDP on $M_1(\mathbb{R}_+ \times E)$ endowed with the weak convergence topology with good rate function $H(\rho|\nu \otimes \mu)$.*

Proof. Since $\nu^n \xrightarrow{W_1^\beta} \nu$ we have $\nu^n \xrightarrow{w} \nu$ according to, for example, Theorem 7.12 in [27]. Moreover, since $\mu^n \xrightarrow{w} \mu$ we have $\nu^n \otimes \mu^n \xrightarrow{w} \nu \otimes \mu$ (see [4], Chapter 1, Theorem 3.2). The announced result then follows from Theorem 5 in [3]. □

Our strategy in proving a weak convergence version of Theorem 4.1 consists in comparing T^n to random measures associated to $\mathcal{T}^n$. Comparison is possible because the $\rho \in M_1(\mathbb{R}_+ \times E)$ such that $I(\rho; \nu) < +\infty$ can be approximated in the weak convergence topology by elements of $\mathcal{P}_n$. The next two lemmas are the crucial arguments of the proof.

Lemma 4.2. *Let $\rho \in M_1(\mathbb{R}_+ \times E)$ be such that $\rho_1 = \nu$ and $\rho_2 = \mu$. There exists a sequence $(\rho^n)_{n \geq 1}$ satisfying* (A1) *and such that $\rho^n \xrightarrow{w} \rho$.*

Proof. Let $\rho \in M_1(\mathbb{R}_+ \times E)$ be such that $\rho_1 = \nu$ and $\rho_2 = \mu$. According to Varadarajan's Lemma (see [12], Chapter 11, Theorem 11.4.1) there exists a family $((u_i^n, v_i^n)_{1 \leq i \leq n})_{n \geq 1}$ of elements of $\mathbb{R}_+ \times E$ such that

$$\gamma^n = \frac{1}{n}\sum_{i=1}^n \delta_{(u_i^n, v_i^n)} \xrightarrow{w} \rho.$$

For every $n \geq 1$, we take $\varphi_n, \tau_n \in \mathfrak{S}_n$ such that

$$\sum_{i=1}^n \zeta\left(u_i^n, w_{\varphi_n(i)}^n\right) = \min_{\varphi \in \mathfrak{S}_n}\left\{\sum_{i=1}^n \zeta\left(u_i^n, w_{\varphi(i)}^n\right)\right\}$$

and

$$\sum_{i=1}^n \lambda\left(v_i^n, x_{\tau_n(i)}^n\right) = \min_{\tau \in \mathfrak{S}_n}\left\{\sum_{i=1}^n \lambda\left(v_i^n, x_{\tau(i)}^n\right)\right\}.$$

We shall prove that the sequence of measures $\rho^n = \frac{1}{n}\sum_{i=1}^n \delta_{(w_{\varphi_n(i)}^n, x_{\tau_n(i)}^n)} \in \mathcal{P}_n$ converges weakly to ρ. Indeed, according to Lemma 2.1 in [16] we have

$$W_1^\zeta\left(\rho_1^n, \gamma_1^n\right) = \min_{Q \in \mathcal{C}(\rho_1^n, \gamma_1^n)} \int_{\mathbb{R}_+ \times \mathbb{R}_+} \zeta(u, v) Q(\mathrm{d}u, \mathrm{d}v) = \min_{\kappa \in \mathfrak{S}_n} \frac{1}{n}\sum_{i=1}^n \zeta\left(u_i^n, w_{\kappa(i)}^n\right),$$

hence $\frac{1}{n}\sum_{i=1}^n \zeta(u_i^n, w_{\varphi_n(i)}^n) = W_1^\zeta(\rho_1^n, \gamma_1^n) \to 0$ since both $\rho_1^n \xrightarrow{w} \nu$ and $\gamma_1^n = \nu^n \xrightarrow{w} \nu$. One can prove the same way that $\frac{1}{n}\sum_{i=1}^n \lambda(v_i^n, x_{\tau_n(i)}^n) \to 0$ since W_1^λ is compatible with the weak convergence topology. Finally,

$$\begin{aligned} W_1^\chi\left(\rho^n, \gamma^n\right) &= \min_{Q \in \mathcal{C}(\rho^n, \gamma^n)} \int_{\mathbb{R}_+ \times E} \chi(u, v) Q(\mathrm{d}u, \mathrm{d}v) \\ &= \min_{\kappa \in \mathfrak{S}_n} \frac{1}{n}\sum_{i=1}^n \chi\left(\left(u_i^n, v_i^n\right), \left(w_{\kappa\circ\varphi_n(i)}^n, x_{\kappa\circ\tau_n(i)}^n\right)\right) \\ &\leq \frac{1}{n}\sum_{i=1}^n \zeta\left(u_i^n, w_{\varphi_n(i)}^n\right) + \frac{1}{n}\sum_{i=1}^n \lambda\left(v_i^n, x_{\tau_n(i)}^n\right) \end{aligned}$$

so $W_1^\chi(\rho^n, \gamma^n) \to 0$ hence $\rho^n \xrightarrow{w} \rho$ since $\gamma^n \xrightarrow{w} \rho$. □

To every $n \geq 1$ and every realization of $\mathcal{T}^n$, we associate two elements $\widetilde{T}^n$ and $\widehat{T}^n$ of $M_1(\mathbb{R}_+ \times E)$ by

$$W_1^\chi\left(\mathcal{T}^n, \widetilde{T}^n\right) = \min_{\nu \in \mathcal{P}_n}\left\{W_1^\chi\left(\mathcal{T}^n, \nu\right)\right\} \tag{4.3}$$

and

$$W_1^\chi\left(\mathcal{T}^n, \widehat{T}^n\right) = \max_{\nu \in \mathcal{P}_n}\left\{W_1^\chi\left(\mathcal{T}^n, \nu\right)\right\}. \tag{4.4}$$

In case there are several elements of $\mathcal{P}_n$ achieving the min (resp. the max) $\widetilde{T}^n$ (resp. $\widehat{T}^n$) is picked uniformly at random among these measures.

Lemma 4.3. *For every $n \geq 1$ the random measures $\widetilde{T}^n, \widehat{T}^n$ and T^n are identically distributed over $M_1(\mathbb{R}_+ \times E)$.*

Proof. We shall only prove that $\widetilde{T}^n$ and T^n are identically distributed since the proof with $\widehat{T}^n$ and T^n is similar. Let $n \geq 1$ be fixed. For the sake of clarity, let us assume that there is no repetition among the $w_1^n, \dots, w_n^n$ and the $x_1^n, \dots, x_n^n$. We are thus left to prove that $\widetilde{T}^n$ is uniformly distributed over $\mathcal{P}_n$. Since there are no repetitions every $\rho \in \mathcal{P}_n$ corresponds to a single $\tau \in \mathfrak{S}_n$ by

$$\rho = \frac{1}{n} \sum_{i=1}^n \delta_{(w^n_{\tau(i)}, x_i^n)}. \tag{4.5}$$

Let us consider a fixed realization $(l_i^n, r_i^n)_{1 \leq i \leq n}$ of $(L_i^n, R_i^n)_{1 \leq i \leq n}$. We denote by t^n the corresponding value of $\mathcal{T}^n$. Due to Lemma 2.1 in [16] for every $\rho \in \mathcal{P}_n$ (i.e., every $\tau \in \mathfrak{S}_n$ according to (4.5)), there exists a $\sigma \in \mathfrak{S}_n$ such that

$$\begin{aligned} W_1^\chi(t^n, \rho) &= \min_{\kappa \in \mathfrak{S}_n} \left\{ \frac{1}{n} \sum_{i=1}^n \chi\big((l_i^n, r_i^n), (w^n_{\kappa \circ \tau(i)}, x^n_{\kappa(i)})\big) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \chi\big((l_i^n, r_i^n), (w^n_{\sigma \circ \tau(i)}, x^n_{\sigma(i)})\big) \\ &= \frac{1}{n} \sum_{i=1}^n \zeta\big(l_i^n, w^n_{\sigma \circ \tau(i)}\big) + \frac{1}{n} \sum_{i=1}^n \lambda\big(r_i^n, x^n_{\sigma(i)}\big). \end{aligned}$$

Hence, for this realization $(l_i^n, r_i^n)_{1 \leq i \leq n}$ of $(L_i^n, R_i^n)_{1 \leq i \leq n}$, the associated $\widetilde{T}^n$ is found by computing $\eta_1, \eta_2 \in \mathfrak{S}_n$ such that

$$\sum_{i=1}^n \zeta\big(l_i^n, w^n_{\eta_1(i)}\big) = \min_{\phi \in \mathfrak{S}_n} \left\{ \sum_{i=1}^n \zeta\big(l_i^n, w^n_{\phi(i)}\big) \right\} \tag{4.6}$$

and

$$\sum_{i=1}^n \lambda\big(r_i^n, x^n_{\eta_2(i)}\big) = \min_{\phi \in \mathfrak{S}_n} \left\{ \sum_{i=1}^n \lambda\big(r_i^n, x^n_{\phi(i)}\big) \right\} \tag{4.7}$$

and taking $\frac{1}{n} \sum_{i=1}^n \delta_{(w^n_{\eta_1(i)}, x^n_{\eta_2(i)})}$ as the corresponding value of $\widetilde{T}^n$. In case several η_1 and/or η_2 realize the minima in the displays above, those defining the corresponding value of $\widetilde{T}^n$ are picked among them uniformly at random. To every possible realization $(l_i^n, r_i^n)_{1 \leq i \leq n}$ of $(L_i^n, R_i^n)_{1 \leq i \leq n}$ we associate $\mathfrak{A}_n((l_i^n, r_i^n)_{1 \leq i \leq n}) \subset \mathfrak{S}_n \times \mathfrak{S}_n$ the set of couples of permutations (η_1, η_2) associated to $(l_i^n, r_i^n)_{1\le i\le n}$ by (4.6) and (4.7). This set may not be reduced to a single element due the possible multiplicity of minimizers. Conversely, for every $(\eta_1,\eta_2)\in\mathfrak{S}_n\times\mathfrak{S}_n$ consider

$$\mathfrak{A}_n^{-1}(\eta_1,\eta_2)=\left\{\left(l_i^n,r_i^n\right)_{1\le i\le n}:(\eta_1,\eta_2)\in\mathfrak{A}_n\left(\left(l_i^n,r_i^n\right)_{1\le i\le n}\right)\right\}.$$

Let (γ_1,γ_2) and (ϕ_1,ϕ_2) be two different elements of $\mathfrak{S}_n\times\mathfrak{S}_n$: there exists $(\varphi_1,\varphi_2)\in\mathfrak{S}_n\times\mathfrak{S}_n$ such that $\gamma_1\circ\varphi_1=\phi_1$ and $\gamma_2\circ\varphi_2=\phi_2$. Now, for every $(l_i^n,r_i^n)_{1\le i\le n}\in\mathfrak{A}_n^{-1}(\gamma_1,\gamma_2)$ we have $(l_{\varphi_1(i)}^n,r_{\varphi_2(i)}^n)_{1\le i\le n}\in\mathfrak{A}_n^{-1}(\phi_1,\phi_2)$ and for every $(l_i^n,r_i^n)_{1\le i\le n}\in\mathfrak{A}_n^{-1}(\phi_1,\phi_2)$ we have $(l_{\varphi_1^{-1}(i)}^n,r_{\varphi_2^{-1}(i)}^n)_{1\le i\le n}\in\mathfrak{A}_n^{-1}(\gamma_1,\gamma_2)$. Since for every $(l_i^n,r_i^n)_{1\le i\le n}$ and every $\kappa_1,\kappa_2\in\mathfrak{S}_n$, observing $(l_{\kappa_1(i)}^n,r_{\kappa_2(i)}^n)_{1\le i\le n}$ as a realization of $(L_i^n,R_i^n)_{1\le i\le n}$ has the same probability as observing $(l_i^n,r_i^n)_{1\le i\le n}$ we conclude that all the $(\kappa_1,\kappa_2)\in\mathfrak{S}_n\times\mathfrak{S}_n$ have the same probability to be observed as minimizers in the problem (4.6)–(4.7) above. Hence, every possible value of $\widetilde{T}^n$ has the same probability to be observed whence $\widetilde{T}^n$ is uniformly distributed over $\mathcal{P}_n$. This proof extends easily to the case when there are repetitions among the $w_1^n,\dots,w_n^n$ or $x_1^n,\dots,x_n^n$. □

We start the proof of the LD bounds in the weak convergence topology by proving the following lemma.

Lemma 4.4. *We have*:

1. $I(\cdot;\nu)$ *is a good rate function.*
2. *The sequence* $(T^n)_{n\ge1}$ *is exponentially tight.*

Proof. (1) Let $\alpha\ge0$. We have

$$\begin{aligned}N_\alpha&=\left\{\rho\in M_1(\mathbb{R}_+\times E):I(\rho;\nu)\le\alpha\right\}\\&=\left\{\rho\in M_1(\mathbb{R}_+\times E):H(\rho|\nu\otimes\mu)\le\alpha\right\}\cap\left\{\rho\in M_1(\mathbb{R}_+\times E):\rho_1=\nu\text{ and }\rho_2=\mu\right\}.\end{aligned}$$

Thus, for every $\alpha\ge0$, N_α is the intersection of a compact and a closed subset of $M_1(\mathbb{R}_+\times E)$, therefore it is compact.

(2) For every measurable $A\subset M_1(\mathbb{R}_+\times E)$, we have

$$\begin{aligned}\limsup_{n\to\infty}\frac1n\log\mathbb{P}\left(T^n\in A^c\right)&=\limsup_{n\to\infty}\frac1n\log\mathbb{P}\left(\mathcal{T}^n\in A^c\,\Big|\,\frac1n\sum_{i=1}^n\delta_{L_i^n}=\nu^n,\frac1n\sum_{i=1}^n\delta_{R_i^n}=\mu^n\right)\\&\le\limsup_{n\to\infty}\frac1n\log\mathbb{P}\left(\mathcal{T}^n\in A^c\right)\\&\quad-\liminf_{n\to\infty}\frac1n\log\mathbb{P}\left(\frac1n\sum_{i=1}^n\delta_{L_i^n}=\nu^n\right)\\&\quad-\liminf_{n\to\infty}\frac1n\log\mathbb{P}\left(\frac1n\sum_{i=1}^n\delta_{R_i^n}=\mu^n\right).\end{aligned}\tag{4.8}$$

Since $(\mathcal{T}^n)_{n\geq 1}$ satisfies a LDP on $M_1(\mathbb{R}_+ \times E)$ with a good rate function it is exponentially tight (see [9], Remark a), p. 8). Thus for every $\alpha \geq 0$, we can chose a compact set $A_\alpha \subset M_1(\mathbb{R}_+ \times E)$ that makes the first term in the last display smaller than $-\alpha - 2$. Below we prove that

$$-\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{L_i^n} = \nu^n\right) \leq 1, \tag{4.9}$$

which combined with (4.8) completes the proof since the same LD inequality holds for $\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} \delta_{R_i^n} = \mu^n)$. Indeed, for every fixed integer $n \geq 1$ let us denote by $M(n)$ the number of different values taken by the $w_1^n, \dots, w_n^n$'s, for example, $M(n) = n$ when there are no ties among the $w_1^n, \dots, w_n^n$'s and $M(n) = 1$ when they are all equals. Let us further denote by $A_1, \dots, A_{M(n)}$ the number of elements of $w_1^n, \dots, w_n^n$'s of each type, for example, $A_1 = \cdots = A_{M(n)} = 1$ when there are no ties among the $w_1^n, \dots, w_n^n$'s while $A_{M(n)} = n$ when they are all equals. Notice that we always have $A_1 + \cdots + A_{M(n)} = n$. We obtain

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{L_i^n} = \nu^n\right) &= C_n^{A_1} C_{n-A_1}^{A_2} \cdots C_{n-(A_1+\cdots+A_{M(n)-1})}^{A_{M(n)}} \left(\frac{A_1}{n}\right)^{A_1} \cdots \left(\frac{A_{M(n)}}{n}\right)^{A_{M(n)}} \\
&= \frac{n!}{A_1! \cdots A_{M(n)}!} \left(\frac{A_1}{n}\right)^{A_1} \cdots \left(\frac{A_{M(n)}}{n}\right)^{A_{M(n)}} \\
&= \frac{n!}{n^{A_1+\cdots+A_{M(n)}}} \frac{A_1^{A_1}}{A_1!} \cdots \frac{A_{M(n)}^{A_{M(n)}}}{A_{M(n)}!},
\end{aligned}$$

hence

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{L_i^n} = \nu^n\right) \geq n! \frac{1}{n^n}$$

since for every integer $m \geq 1$ we have $m^m \geq m!$. Display (4.9) immediately follows. □

Proof of the lower bound. It is sufficient in order to prove the lower bound of the LDP to prove that

$$-I(\rho; \nu) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in B(\rho, \varepsilon)\big)$$

holds for every $\rho \in M_1(\mathbb{R}_+ \times E)$ and every $\varepsilon > 0$, where $B(\rho, \varepsilon)$ stands for the open ball centered at $\rho \in M_1(\mathbb{R}_+ \times E)$ of radius $\varepsilon > 0$ for the W_1^χ metric. So let $\varepsilon > 0$ and $\rho \in M_1(\mathbb{R}_+ \times E)$ be such that $I(\rho; \nu) < +\infty$. In particular $\rho_1 = \nu$ and $\rho_2 = \mu$. According to Lemma 4.2 there exists a sequence $(\rho^n)_{n\geq 1}$ of elements of $M_1(\mathbb{R}_+ \times E)$ such that for every $n \geq 1$ we have $\rho^n \in \mathcal{P}_n$ and $\rho^n \overset{w}{\to} \rho$. According to Lemma 4.3

$$\begin{aligned}
&\mathbb{P}\big(T^n \in B(\rho, \varepsilon)\big) \\
&\quad = \mathbb{P}\big(\widetilde{T}^n \in B(\rho, \varepsilon)\big)
\end{aligned}$$

$$\geq \mathbb{P}\Big(W_1^{\chi}\big(\widetilde{T}^n, \mathcal{T}^n\big) < \frac{\varepsilon}{3}, W_1^{\chi}\big(\mathcal{T}^n, \rho^n\big) < \frac{\varepsilon}{3}, W_1^{\chi}\big(\rho^n, \rho\big) < \frac{\varepsilon}{6}\Big)$$

$$\geq \mathbb{P}\Big(W_1^{\chi}\big(\rho^n, \mathcal{T}^n\big) < \frac{\varepsilon}{3}, W_1^{\chi}\big(\rho^n, \rho\big) < \frac{\varepsilon}{6}\Big)$$

since it follows from the definition of $\widetilde{T}^n$ that for every $\rho^n \in \mathcal{P}_n$ we have

$$W_1^{\chi}\big(\widetilde{T}^n, \mathcal{T}^n\big) \leq W_1^{\chi}\big(\rho^n, \mathcal{T}^n\big).$$

On the other hand since $\rho^n \xrightarrow{w} \rho$ we get that for n large enough $\{W_1^{\chi}(\rho^n, \rho) < \frac{\varepsilon}{6}\} = \Omega$. Thus, for those n's

$$\mathbb{P}\Big(W_1^{\chi}\big(\rho^n, \mathcal{T}^n\big) < \frac{\varepsilon}{3}, W_1^{\chi}\big(\rho^n, \rho\big) < \frac{\varepsilon}{6}\Big) \geq \mathbb{P}\Big(W_1^{\chi}\big(\rho, \mathcal{T}^n\big) < \frac{\varepsilon}{6}, W_1^{\chi}\big(\rho^n, \rho\big) < \frac{\varepsilon}{6}\Big)$$
$$\geq \mathbb{P}\Big(W_1^{\chi}\big(\mathcal{T}^n, \rho\big) < \frac{\varepsilon}{6}\Big).$$

Finally, it follows from Lemma 4.1 that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in B(\rho, \varepsilon)\big) \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(W_1^{\chi}\big(\rho, \mathcal{T}^n\big) < \frac{\varepsilon}{6}\Big)$$
$$\geq -H(\rho|\nu \otimes \mu) = -I(\rho; \nu).$$

Proof of the upper bound. In order to prove the upper bound of the LDP, it is sufficient to prove that it holds for compact subsets of $M_1(\mathbb{R}_+ \times E)$. Indeed, since $(T^n)_{n\geq 1}$ is an exponentially tight sequence (see Lemma 4.4) the full upper bound will follow from Lemma 1.2.18 in [9]. Let A be a compact subset of $M_1(\mathbb{R}_+ \times E)$ and let us denote by

$$A_{\nu,\mu} = \{\rho \in A : \rho_1 = \nu \text{ and } \rho_2 = \mu\},$$

which is a compact subset of $M_1(\mathbb{R}_+ \times E)$ as well. Since the weak convergence topology on $M_1(\mathbb{R}_+ \times E)$ is compatible with the W_1^{χ} metric, it makes $M_1(\mathbb{R}_+ \times E)$ a regular topological space: For every $\rho \in A$ such that $\rho \in A_{\nu,\mu}^c$ there exists $\varepsilon_\rho > 0$ such that $B(\rho, 2\varepsilon_\rho) \cap A_{\nu,\mu} = \varnothing$. In particular, $\bar{B}(\rho, \varepsilon_\rho) \cap A_{\nu,\mu} = \varnothing$ where $\bar{B}(\rho, \varepsilon)$ denotes the closed ball centered on $\rho \in M_1(\mathbb{R}_+ \times E)$ of radius $\varepsilon > 0$ for the W_1^{χ} metric. On the other hand, since $\rho \mapsto H(\rho|\nu \otimes \mu)$ is lower semi-continuous, for every $\rho \in A_{\nu,\mu}$ and every $\delta > 0$ there exists a $\varphi(\rho, \delta) > 0$ such that

$$\inf_{\gamma \in \bar{B}(\rho, \varphi(\rho,\delta))} H(\gamma|\nu \otimes \mu) \geq \big(H(\rho|\nu \otimes \mu) - \delta\big) \wedge \frac{1}{\delta}.$$

For every $\delta > 0$, we consider the coverage

$$A \subset \Big(\bigcup_{\rho \in A \cap A_{\nu,\mu}^c} B(\rho, \varepsilon_\rho)\Big) \cup \Big(\bigcup_{\rho \in A_{\nu,\mu}} B\Big(\rho, \frac{\varphi(\rho, \delta)}{8}\Big)\Big)$$

from which we extract a finite coverage

$$A \subset \Big(\bigcup_{\rho\in I_1} B(\rho,\varepsilon_\rho)\Big) \cup \Big(\bigcup_{\rho\in I_2} B\Big(\rho,\frac{\varphi(\rho,\delta)}{8}\Big)\Big),$$

where $I_1 \subset A \cap A^c_{\nu,\mu}$ and $I_2 \subset A_{\nu,\mu}$ are finite sets. According to Lemma 1.2.15 in [9]

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(T^n \in A\big) \le \max\Big\{\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(T^n \in \bigcup_{\rho\in I_1}\bar{B}(\rho,\varepsilon_\rho)\cap A\Big),$$
$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(T^n \in \bigcup_{\rho\in I_2} B\Big(\rho,\frac{\varphi(\rho,\delta)}{8}\Big)\Big)\Big\}.$$

For every $\rho \in I_1$ there can not be an infinite number of integers n_k such that

$$\mathbb{P}\big(T^{n_k} \in \bar{B}(\rho,\varepsilon_\rho)\cap A\big) \neq 0$$

for otherwise we would get $\bar{B}(\rho,\varepsilon_\rho)\cap A_{\nu^1,\nu^2} \neq \varnothing$. The first term in the max is then equal to $-\infty$. We are left with the second term and according to Lemmas 4.1, 4.2 and 4.3, we have

$$\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(T^n \in A\big) &\le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(T^n \in \bigcup_{\rho\in I_2} B\Big(\nu,\frac{\varphi(\rho,\delta)}{8}\Big)\Big) \\
&\le \max_{\rho\in I_2}\Big\{\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(\widehat{T}^n \in B\Big(\rho,\frac{\varphi(\rho,\delta)}{8}\Big)\Big)\Big\} \\
&\le \max_{\rho\in I_2}\Big\{\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(W_1^\chi\big(\rho^n,\widehat{T}^n\big) < \frac{\varphi(\rho,\delta)}{4}\Big)\Big\} \\
&\le \max_{\rho\in I_2}\Big\{\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\Big(W_1^\chi\big(\rho,T^n\big) < \frac{\varphi(\rho,\delta)}{2}\Big)\Big\} \\
&\le \max_{\rho\in I_2}\Big\{-\inf_{\gamma\in\bar{B}(\rho,\varphi(\rho,\delta))} H(\gamma|\nu\otimes\mu)\Big\} \\
&\le \max_{\rho\in I_2}\Big\{-\big(H(\rho|\nu\otimes\mu)-\delta\big)\wedge\frac{1}{\delta}\Big\} \\
&\le \max_{\rho\in I_2}\Big\{-\big(I(\rho;\nu)-\delta\big)\wedge\frac{1}{\delta}\Big\} \\
&\le -\inf_{\rho\in A}\Big\{\big(I(\rho;\nu)-\delta\big)\wedge\frac{1}{\delta}\Big\}.
\end{aligned}$$

By letting $\delta \to 0$, we obtain the announced upper bound, see Remark 1.2.10 in [9].

4.1.2. *Conclusion of the proof of Theorem* 4.1

First, notice that

$$\mathcal{N}_1^1(\mathbb{R}_+ \times E) = \left\{\rho \in M_1(\mathbb{R}_+ \times E) : \int_{\mathbb{R}_+} x\rho_1(\mathrm{d}x) \le 1\right\} \tag{4.10}$$

is a closed subset of $M_1(\mathbb{R}_+ \times E)$ when the latter is endowed with the weak convergence topology. Since for every $n \ge 1$ we have $\mathbb{P}(T_n \in \mathcal{N}_1^1(\mathbb{R}_+ \times E)) = 1$, Lemma 4.1.5 in [9] implies that $(T_n)_{n\ge1}$ obeys a LDP on $\mathcal{N}_1^1(\mathbb{R}_+ \times E)$ endowed with the weak convergence topology, with good rate function I.

Next, we prove that the same remains true when $\mathcal{N}_1^1(\mathbb{R}_+ \times E)$ is endowed with the distance Δ. Indeed, since $(T^n)_{n\ge1}$ satisfies a LDP on the Polish space $(M_1(\mathbb{R}_+ \times E), W_1^{\chi})$ with a good rate function it is exponentially tight: For every $L > 0$ there exists a $A \subset M_1(\mathbb{R}_+ \times E)$, compact for the weak-convergence topology, such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in A^c\big) \le -L.$$

Let us show that the set $A \cap K$ with

$$K = \left\{\rho \in M_1(\mathbb{R}_+ \times E) : \rho_1 = \nu \text{ or } \rho_1 \in \bigcup_{n=1}^{\infty}\{\nu^n\}\right\}$$

is a compact subset of $\mathcal{N}_1^1(\mathbb{R}_+ \times E)$ endowed with Δ.

$A \cap K$ *is closed*. Let $(\gamma^n)_{n\ge1}$ be a converging sequence of elements of $A \cap K$ which limit we denote γ. Since $\Delta(\gamma^n, \gamma) = W_1^{\beta}(\gamma_1^n, \gamma_1) + W_1^{\chi}(\gamma^n, \gamma) \to 0$ necessarily $\gamma_1 = \nu$ or $\gamma_1 \in \bigcup_{n=1}^{\infty}\{\nu^n\}$ due to the particular form of K and $\gamma \in A$ since A is closed for the weak-convergence topology, hence $\gamma \in A \cap K$.

$A \cap K$ *is sequentially compact*. We prove that any sequence $(\gamma^n)_{n\ge1}$ of elements of $A \cap K$ necessarily admits a Δ-converging sub-sequence. Due to the definition of K, $(\gamma_1^n)_{n\ge1}$ admits a W_1^{β}-converging sub-sequence. Along this sub-sequence $(\gamma^n)_{n\ge1}$ also admits a weakly converging (sub-)sub-sequence since A is compact for the weak convergence topology. The sub-sequence obtained by this double extraction is Δ-convergent. Moreover,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in (A \cap K)^c\big) &\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in A^c\big) \\ &\quad + \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(T^n \in K^c\big) \\ &\le -L.\end{aligned}$$

Hence, $(T_n)_{n\ge1}$ obeys a LDP on $(\mathcal{N}_1^1(\mathbb{R}_+ \times E), \Delta)$ see Corollary 4.2.6 in [9]. To conclude notice that $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a closed subset of $(\mathcal{N}_1^1(\mathbb{R}_+ \times E), \Delta)$ and that for every $n \ge 1$, we have $\mathbb{P}(T_n \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)) = 1$ so, due to Lemma 4.1.5 in [9], the sequence $(T_n)_{n\ge1}$ obeys a LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ endowed with the distance Δ with good rate function I.

4.2. From a conditional to an unconditional LDP

In this section, we give a precise meaning to (4.1). Theorem 2.1 follows from Theorem 4.1 by a simple application of Theorem 2.3 in [18]. Nevertheless, let us give a hint of the ideas standing behind the latter result. To this end, we need to introduce the sequence of sets

$$M_1^{1,n}(\mathbb{R}_+) = \left\{\theta \in M_1^1(\mathbb{R}_+) : \exists(w_1,\dots,w_n) \in (\mathbb{R}_+)^n, \theta = \frac{1}{n}\sum_{i=1}^n \delta_{w_i}\right\}.$$

It follows from Theorem 4.1 that for every closed $C \subset \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ and every $\nu \in M_1^1(\mathbb{R}_+)$ satisfying $I(C;\nu) := \inf_{\rho\in C} I(\rho;\nu) < \infty$ there exists for each $\delta > 0$ a neighborhood U_ν of ν in $(M_1^1(\mathbb{R}_+), W_1^\beta)$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu' \in U_\nu \cap M_1^{1,n}(\mathbb{R}_+)} \mathbb{P}\big(T^n \in C; \nu'\big) \le -I(C;\nu) + \delta.$$

If $I(C;\nu) = \infty$, there exists for each $L \in \mathbb{R}$ a neighborhood U_ν of ν such that the l.h.s. of the previous display is smaller than $-L$. Now, due to the goodness of I^W we get for every $L > 0$ and every $\varepsilon > 0$ a finite covering $\bigcup_{i=1}^k U_{\nu(i)}$ of $\Phi_L = \{\nu \in M_1^1(\mathbb{R}_+), I^W(\nu) < L\}$ by such neighborhoods such that

$$\mathbb{P}\big(\mathcal{V}^n \in A\big) \le \int_{\Phi_L^c} \mathbb{P}\big(\mathcal{V}^n \in A; \nu\big) Q^n(\mathrm{d}\nu) + \sum_{i=1}^k \int_{U_{\nu(i)} \cap M_1^{1,n}(\mathbb{R}_+)} \mathbb{P}\big(\mathcal{V}^n \in A; \nu\big) Q^n(\mathrm{d}\nu).$$

This finite sum leads to the upper LD bound in a straightforward way. The proof of the lower LD bound follows the same idea and is even simpler due to the local nature of such bounds.

4.2.1. *A Large Deviations System*

In order to conclude the proof of Theorem 2.1, it is sufficient to establish that the distribution of $\mathcal{V}^n$ on $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a mixture of Large Deviation Systems (LDS) in the sense of [18]. For the sake of clarity, we recover the notations of [18] when identifying the components of the LDS:

- $\mathcal{Z} = \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a Polish space when endowed with the distance Δ, see Section 4.2.2 below.
- $\mathcal{X} = M_1^1(\mathbb{R}_+)$ is a Polish space when endowed with W_1^β since it is a closed subset of the Polish space $(\mathcal{W}^1(\mathbb{R}_+), W_1^\beta)$ (see, e.g., [5]).
- For every $n \ge 1$, we note $\mathcal{X}_n = M_1^{1,n}(\mathbb{R}_+)$ and for every $\nu \in \mathcal{X}$ and every $n \ge 1$ there exists a $\nu^n \in \mathcal{X}_n$ such that $\nu^n \overset{W_1^\beta}{\to} \nu$, see Lemma 4.5 in Section 4.2.3 below.
- The map $\pi : \mathcal{Z} \to \mathcal{X}$ defined by $\pi(\nu) = \nu_1$ is continuous and surjective.
- For every $n \ge 1$ and every $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{w_i} \in \mathcal{X}_n$ let P_ν^n be the distribution of $T_n = \frac{1}{n}\sum_{i=1}^n \delta_{(w_{\sigma_n(i)}, x_i^n)}$ under $\mathbb{P}$. The family $\Pi = \{P_\nu^n, \nu \in \mathcal{X}_n, n \ge 1\}$ of finite measures on the Borel σ-field on $\mathcal{Z}$ is such that for every $n \ge 1$ and every $\nu \in \mathcal{X}_n$ we have $P_\nu^n(\pi^{-1}(\{\nu\}^c)) = 0$.

- Let Q^n be the distribution of $\frac{1}{n}\sum_{i=1}^n \delta_{W_i^n}$. For every $n \geq 1$ and every measurable $A \subset \mathcal{M}_1^1(\mathbb{R}_+ \times E)$

$$\mathbb{P}(\mathcal{V}^n \in A) = \int_{\mathcal{X}_n} P_\nu^n(A) Q^n(\mathrm{d}\nu).$$

All the requirements of Definition 2.1 in [18] are satisfied by our model thanks to Theorem 4.1. It follows from Theorem 2.3 in [18] that the sequence $(\mathcal{V}^n)_{n\geq 1}$ obeys a LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ with distance Δ with good rate function

$$\mathcal{J}(\rho) = \begin{cases} H(\rho|\rho_1 \otimes \mu) + I^W(\rho_1), & \text{if } \rho_2 = \mu, \\ +\infty, & \text{otherwise.} \end{cases}$$

4.2.2. $(\mathcal{M}_1^1(\mathbb{R}_+ \times E), \Delta)$ *is a Polish space*

$(\mathcal{M}_1^1(\mathbb{R}_+ \times E), \Delta)$ *is complete.* Let $(\rho^n)_{n\geq 1}$ be a Cauchy sequence of elements of $(\mathcal{M}_1^1(\mathbb{R}_+ \times E), \Delta)$. In particular $(\rho_1^n)_{n\geq 1}$ is a Cauchy sequence of elements of $(M_1^1(\mathbb{R}_+), W_1^\beta)$ which is a complete space as a closed subset of $(\mathcal{W}^1(\mathbb{R}_+), W_1^\beta)$ (see, e.g., [5]). So there exists a $\rho_1 \in M_1^1(\mathbb{R}_+)$ such that $W_1^\beta(\rho_1^n, \rho_1) \to 0$. Furthermore, $(\rho^n)_{n\geq 1}$ is a Cauchy sequence of elements of the Polish space $(M_1(\mathbb{R}_+ \times E), W_1^\chi)$ hence there exists a $\gamma \in M_1(\mathbb{R}_+ \times E)$ such that $W_1^\chi(\rho^n, \gamma) \to 0$. Necessarily, $\gamma_1 = \rho_1$ hence $\Delta(\rho^n, \gamma) \to 0$.

$(\mathcal{M}_1^1(\mathbb{R}_+ \times E), \Delta)$ *is separable.* Let R and S be dense countable subsets of $\mathbb{R}_+$ and E, respectively. It is sufficient to prove that

$$\bigcup_{n\geq 1}\left\{\frac{1}{n}\sum_{i=1}^n \delta_{(u_i^n, v_i^n)} : \text{for every } 1 \leq i \leq n, u_i^n \in R \text{ and } v_i^n \in S\right\}$$

is dense in $\mathcal{N}_1^1(\mathbb{R}_+ \times E)$ defined in (4.10) endowed with the distance Δ (see (4.10)). Indeed, since $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a closed subset of $\mathcal{N}_1^1(\mathbb{R}_+ \times E)$ the announced claim will follow. So let $\rho \in \mathcal{N}_1^1(\mathbb{R}_+ \times E)$. In particular, $\int_{\mathbb{R}_+} x\rho_1(\mathrm{d}x) < \infty$. Lets us denote by $(Z_1, T_1), \dots, (Z_n, T_n), \dots$ a sequence of independent random variables with common distribution ρ. According to Varadarajan's lemma,

$$\frac{1}{n}\sum_{i=1}^n \delta_{(Z_i, T_i)} \xrightarrow{w} \rho \qquad \text{almost surely} \tag{4.11}$$

and according to the Strong Law of Large Numbers

$$\frac{1}{n}\sum_{i=1}^n Z_i \to \int_{\mathbb{R}_+} x\rho_1(\mathrm{d}x) \qquad \text{almost surely.} \tag{4.12}$$

So there exists a family $((z_i,t_i)_{1\le i\le n})_{n\ge1}$ of elements of $\mathbb{R}_+\times E$ such that

$$\frac{1}{n}\sum_{i=1}^n \delta_{(z_i,t_i)} \xrightarrow{w} \rho \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n z_i \to \int_{\mathbb{R}_+} x\rho_1(\mathrm{d}x),$$

hence $\Delta(\rho, \frac{1}{n}\sum_{i=1}^n \delta_{(z_i,t_i)}) \to 0$, see Theorem 7.12 in [27]. Since $R\times S$ is dense in $\mathbb{R}_+\times E$ for every $n\ge1$ and every $1\le i\le n$ there exists $(u_i^n, v_i^n)\in R\times S$ such that $\max\{\xi((u_i^n,v_i^n),(z_i,t_i))\}\le 2^{-n}$. Clearly,

$$W_1^\chi\left(\frac{1}{n}\sum_{i=1}^n \delta_{(u_i^n,v_i^n)}, \frac{1}{n}\sum_{i=1}^n \delta_{(x_i^n,y_i^n)}\right) \le 2^{-n},$$

hence $W_1^\chi(\frac{1}{n}\sum_{i=1}^n \delta_{(u_i^n,v_i^n)}, \rho)\to 0$.

4.2.3. *An approximation result in* $(M_1^1(\mathbb{R}_+), W_1^\beta)$

Lemma 4.5. *For every* $\gamma\in M_1^1(\mathbb{R}_+)$ *there exists a sequence* $(\gamma^n)_{n\ge1}$ *of elements of* $M_1^1(\mathbb{R}_+)$ *such that for every* $n\ge1$ $\gamma^n = \frac{1}{n}\sum_{i=1}^n \delta_{v_i^n}$ *and* $\gamma^n \xrightarrow{W_1^\beta} \gamma$.

Proof. By the same kind of argument as in the separability proof above, one can construct a sequence $(\frac{1}{n}\sum_{i=1}^n \delta_{u_i^n})_{n\ge1}$ such that $\frac{1}{n}\sum_{i=1}^n \delta_{u_i^n} \xrightarrow{W_1^\beta} \rho$. In particular $\frac{1}{n}\sum_{i=1}^n u_i^n \to 1$. So we only need to modify the u_i^n's in such a way that for every $n\ge1$ their total sum equals n. For a fixed n, we have three possibilities:

- *If* $\sum_{i=1}^n u_i^n = n$, we take $v_i^n = u_i^n$ for every $1\le i\le n$.
- *If* $\sum_{i=1}^n u_i^n > n$, we look at the u_i^n's as the occupation masses of n cells by a mass u_i^n each. We pick uniformly at random the excess of mass until we get new occupation masses $v_1^n,\dots,v_n^n$ such that $\sum_{i=1}^n v_i^n = n$.
- *If* $\sum_{i=1}^n u_i^n < n$ again, we look at the u_i's as the occupation masses and add mass uniformly at random into the n cells until they contain a total mass of n. We call $v_1^n,\dots,v_n^n$ the final occupation masses.

Due to Lemma 2.1 in [16], in all the cases considered above we have

$$\begin{aligned}
W_1^\beta\left(\frac{1}{n}\sum_{i=1}^n \delta_{u_i^n}, \frac{1}{n}\sum_{i=1}^n \delta_{v_i^n}\right) &= \min_{\tau\in\mathfrak{S}_n}\frac{1}{n}\sum_{i=1}^n |u_i^n - v_{\tau(i)}^n| \le \frac{1}{n}\sum_{i=1}^n |u_i^n - w_i^n| \\
&= \left|\frac{1}{n}\sum_{i=1}^n u_i^n - \frac{1}{n}\sum_{i=1}^n w_i^n\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^n u_i^n - 1\right| \to 0,
\end{aligned}$$

which conclude the proof. $\square$

5. Proof of Theorem 2.2

In order to prove Theorem 2.2 it is sufficient to establish that the distribution of V^n on $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a mixture of LDS. Again we recover the notations of [18] when identifying the components of the LDS:

- $\mathcal{Z} = \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ is a Polish space when endowed with the distance Δ.
- $\mathcal{X} = M_1(E)$ is a Polish space when endowed with the weak convergence topology.
- For every $n \geq 1$, we note

$$\mathcal{X}_n = \left\{ \nu \in M_1^1(E) : \exists (x_1, \dots, x_n) \in E^n, \nu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \right\}$$

  and according to Varadarajan's lemma for every $\nu \in \mathcal{X}$ and every $n \geq 1$ there exists a $\nu^n \in \mathcal{X}_n$ such that $\nu^n \xrightarrow{w} \nu$.
- The map $\pi : \mathcal{Z} \to \mathcal{X}$ defined by $\pi(\nu) = \nu_2$ is continuous and surjective.
- For every $n \geq 1$ and every $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \in \mathcal{X}_n$ let P_ν^n be the distribution of $T_n = \frac{1}{n} \sum_{i=1}^n \delta_{W_i^n, x_i}$ under $\mathbb{P}$. The family $\Pi = \{P_\nu^n, \nu \in \mathcal{X}_n, n \geq 1\}$ of finite measures on the Borel σ-field on $\mathcal{Z}$ is such that for every $n \geq 1$ and every $\nu \in \mathcal{X}_n$ we have $P_\nu^n(\pi^{-1}(\{\nu\}^c)) = 0$.
- Let Q^n be the distribution of $\frac{1}{n} \sum_{i=1}^n \delta_{X_i^n}$. For every $n \geq 1$ and every measurable $A \subset \mathcal{M}_1^1(\mathbb{R}_+ \times E)$

$$\mathbb{P}\big(\mathcal{V}^n \in A\big) = \int_{\mathcal{X}_n} P_\nu^n(A) Q^n(\mathrm{d}\nu).$$

All the requirements of Definition 2.1 in [18] are satisfied by our model thanks to Theorem 2.1. It follows from Theorem 2.3 in [18] that the sequence $(V^n)_{n \geq 1}$ obeys a LDP on $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ with distance Δ with good rate function

$$J(\rho) = H(\rho | \rho_1 \otimes \rho_2) + I^W(\rho_1) + I^X(\rho_2).$$

6. Some more proofs

6.1. Proof of Corollary 2.3

In view of (2.4), a necessary condition on $((W_i^n)_{1 \leq i \leq n})_{n \geq 1}$ for $K = I^X$ is that for every $\nu, \zeta \in M_1(E)$ and every $((X_i^n)_{1 \leq i \leq n})_{n \geq 1}$

$$I^X(\nu) - I^X(\zeta) \leq \mathcal{K}(\nu, \zeta).$$

If we consider $X_1^n, \dots, X_n^n$ resulting from sampling without replacement on an urn which composition $x_1^n, \dots, x_n^n$ satisfies $\frac{1}{n} \sum_{i=1}^n \delta_{x_i^n} \xrightarrow{w} \zeta$ we know that $\frac{1}{n} \sum_{i=1}^n \delta_{X_i^n}$ satisfies a LDP with good rate function

$$I^X(\theta) = \begin{cases} 0, & \text{if } \theta = \zeta, \\ \infty, & \text{otherwise,} \end{cases}$$

so the announced condition is necessary. It is also clearly sufficient and the claimed result follows.

6.2. Proof of Proposition 2.1

Let $\rho, \gamma \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ and $\lambda \in]0, 1[$ be such that $\lambda \mathcal{J}(\rho; \mu) + (1-\lambda)\mathcal{J}(\gamma; \mu) < \infty$ for otherwise the inequality to be proved trivially holds. In particular, $\rho_2 = \gamma_2 = (\lambda\rho + (1-\lambda)\gamma)_2 = \mu$. Since $H(\cdot|\cdot)$ is convex in its two arguments (see, e.g., Lemma 1.4.3 in [13]) we have

$$\begin{aligned}
\mathcal{J}\big(\lambda\rho + (1-\lambda)\gamma\big) &= H\big(\lambda\rho + (1-\lambda)\gamma \big| \big(\lambda\rho_1 + (1-\lambda)\gamma_1\big) \otimes \mu\big) + I^W\big(\lambda_1\rho + (1-\lambda)\gamma_1\big) \\
&\le H\big(\lambda\rho + (1-\lambda)\gamma \big| \lambda(\rho_1 \otimes \mu) + (1-\lambda)(\gamma_1 \otimes \mu)\big) + I^W\big(\lambda_1\rho + (1-\lambda)\gamma_1\big) \\
&= \lambda\big(H(\rho|\rho_1 \otimes \mu) + I^W(\rho_1)\big) + (1-\lambda)\big(H(\gamma|\gamma_1 \otimes \mu) + I^W(\gamma_1)\big).
\end{aligned}$$

Proving that $\mathcal{K}(\cdot; \mu)$ is convex works the same way.

6.3. Proof of Proposition 2.2

We start the proof of Proposition 2.2 by proving (2.8). To this end, we first establish that for every $\nu, \mu \in M_1(E)\mathcal{K}(\nu; \mu) < \infty$ implies $\nu \ll \mu$. Indeed, if the former condition holds there necessarily exists a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ such that $\rho_2 = \mu$ and $F(\rho) = \nu$. For every such ρ the latter reads

$$\nu(A) = \int_{\mathbb{R}_+ \times A} w \rho_x(\mathrm{d}w)\mu(\mathrm{d}x)$$

for every measurable $A \subset E$, hence $\nu \ll \mu$ and

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = \int_{\mathbb{R}_+} w \rho_x(\mathrm{d}w) \tag{6.1}$$

μ a.s. Thus, if $\nu \ll \mu$ does not hold we necessarily have $\mathcal{K}(\nu; \mu) = \infty$.

Now let us assume that $\nu \ll \mu$. First, we show that

$$\mathcal{H}_{\nu,\mu} = \big\{\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E) : \rho_2 = \mu \text{ and } F(\rho) = \nu\big\}$$

is not empty. Consider $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ defined by $\rho_2 = \mu$ while the regular conditional distribution of its first marginal given the second is

$$\rho_x(\mathrm{d}w) = \begin{cases} \mathcal{P}\left(\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)(\mathrm{d}w), & \text{if } \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) > 0, \\ \delta_0(\mathrm{d}w), & \text{otherwise.} \end{cases}$$

Let us check that $F(\rho)=\nu$. Indeed, every measurable $A\subset E$ can be decomposed into $A=A_\nu\cup A_\nu^\perp$ where $A_\nu=A\cap\mathrm{Support}(\nu)$ and

$$\begin{aligned}
F(\rho)(A)&=\int_{\mathbb{R}_+\times A} w\rho_x(\mathrm{d}w)\mu(\mathrm{d}x)\\
&=\int_{A_\nu}\left(\int_{\mathbb{R}_+} w\mathcal{P}\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)(\mathrm{d}w)\right)\mu(\mathrm{d}x)+\int_{A_\nu^\perp}\left(\int_{\mathbb{R}_+} w\delta_0(\mathrm{d}w)\right)\mu(\mathrm{d}x)\\
&=\int_{A_\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\mu(\mathrm{d}x)\\
&=\nu(A_\nu)=\nu(A).
\end{aligned}$$

For every $\rho\in\mathcal{H}_{\nu,\mu}$, every $\theta\in\mathcal{D}_{\Lambda_\xi}$ and μ a.e. $x\in E$, we have

$$\begin{aligned}
H(\rho_x|\xi)&=H\left(\rho_x|\xi^\theta\right)+\int_{\mathbb{R}_+}\log\frac{\mathrm{d}\xi^\theta}{\mathrm{d}\xi}(w)\rho_x(\mathrm{d}w)\\
&=H\left(\rho_x|\xi^\theta\right)+\theta\int_{\mathbb{R}_+} w\rho_x(\mathrm{d}w)-\log\int_{\mathbb{R}_+}\mathrm{e}^{\theta u}\xi(\mathrm{d}u)\\
&=H\left(\rho_x|\xi^\theta\right)+\theta\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)-\log\int_{\mathbb{R}_+}\mathrm{e}^{\theta u}\xi(\mathrm{d}u),
\end{aligned}\tag{6.2}$$

hence

$$H(\rho_x|\xi)-\sup_{\theta\in\mathcal{D}_{\Lambda_\xi}}\left\{\theta\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)-\log\int_{\mathbb{R}_+}\mathrm{e}^{\theta u}\xi(\mathrm{d}u)\right\}=\inf_{\theta\in\mathcal{D}_{\Lambda_\xi}}H\left(\rho_x|\xi^\theta\right)$$

whence

$$\int_E H(\rho_x|\xi)\mu(\mathrm{d}x)-\int_E\Lambda_\xi^*\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x)=\int_E\inf_{\theta\in\mathcal{D}_{\Lambda_\xi}}H\left(\rho_x|\xi^\theta\right)\mu(\mathrm{d}x).\tag{6.3}$$

Since $\mathcal{J}(\rho;\mu)=H(\rho|\rho_1\otimes\mu)+H(\rho_1|\xi)=H(\rho|\xi\otimes\mu)=\int_E H(\rho_x|\xi)\mu(\mathrm{d}x)$ it follows from (6.3) that if $\nu\ll\mu$ then

$$\mathcal{K}(\nu;\mu)=\int_E\Lambda_\xi^*\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)\right)\mu(\mathrm{d}x)+\inf_{\rho\in\mathcal{H}_{\nu,\mu}}\int_E\inf_{\theta\in\mathcal{D}_{\Lambda_\xi}}H\left(\rho_x|\xi^\theta\right)\mu(\mathrm{d}x).$$

Next, we prove that (2.8) reduces to (2.9) if and only if (2.10) holds. In view of (6.2) it amounts to show that (2.10) is a necessary and sufficient condition for the following to hold: for every $\nu,\mu\in M_1(E)$ there exists a $\rho\in\mathcal{M}_1^1(\mathbb{R}_+\times E)$ such that

1. $\rho_2=\mu$,
2. for μ a.e. $x\in E$ $\rho_x=\xi^{\theta_x}$ for some $\theta_x\in\mathcal{D}_{\Lambda_\xi}$,
3. for μ a.e. $x\in E$ such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)>0$ we have $\int_{\mathbb{R}_+} w\rho_x(\mathrm{d}w)=\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$.

Let us recall that Λ_ξ is C^∞ in $\overset{\circ}{\mathcal{D}}_{\Lambda_\xi}$, that for every $\alpha \in \overset{\circ}{\mathcal{D}}_{\Lambda_\xi}$

$$\Lambda_\xi'(\alpha) = \frac{\int_{\mathbb{R}_+} w\mathrm{e}^{\alpha w}\xi(\mathrm{d}w)}{\int_{\mathbb{R}_+} \mathrm{e}^{\alpha w}\xi(\mathrm{d}w)} = \frac{1}{Z}\int_{\mathbb{R}_+} w\mathrm{e}^{\alpha w}\xi(\mathrm{d}w)$$

and that $\Lambda_\xi''(\alpha) > 0$ for every $\alpha \in \overset{\circ}{\mathcal{D}}_{\Lambda_\xi}$, see Section 2.2.1 in [9]. So finally it appears that

$$\lim_{\alpha\to-\infty} \Lambda_\xi'(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha\to a} \Lambda_\xi'(\alpha) = +\infty$$

is a necessary and sufficient condition to get for every $\nu, \mu \in M_1(E)$ and every $x \in E$ such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) > 0$ a ρ_x of the form ξ^θ with $\int_{\mathbb{R}_+} w\rho_x(\mathrm{d}w) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$ and $H(\rho_x|\xi) = \Lambda_\xi^*(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x))$. If $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = 0$, then one takes $\rho_x = \delta_0$ and still gets $\int_{\mathbb{R}_+} w\rho_x(\mathrm{d}w) = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)$ and $H(\rho_x|\xi) = \Lambda_\xi^*(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x)) = \Lambda_\xi^*(0)$. Finally, by properly choosing $\nu, \mu \in M_1(E)$ one can see that (2.10) is also necessary to ensure that (2.9) holds for every $\nu, \mu \in M_1(E)$.

6.4. Proof of Corollary 3.2

Let $\nu \in M_1(E)$ be such that $\inf_{\eta\in\mathcal{Z}} H(\nu|\eta) < \infty$. Necessarily for every $\eta \in \mathcal{Z}$ such that $H(\nu|\eta) < \infty$ we have $\nu \ll \eta$. Now for every such η consider $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ defined by $\rho_2 = \eta$ while the regular conditional distribution of its first marginal given the second is

$$\rho_x(\mathrm{d}w) = \begin{cases} \mathcal{F}\Big(\lambda, \lambda\dfrac{\mathrm{d}\nu}{\mathrm{d}\eta}(x)\Big)(\mathrm{d}w), & \text{if } \dfrac{\mathrm{d}\nu}{\mathrm{d}\eta}(x) > 0, \\ \delta_0(\mathrm{d}w), & \text{otherwise.} \end{cases}$$

Following the proof of Proposition 2.2, we get $F(\rho) = \nu$. Moreover by taking $E_1 = E \cap \{\frac{\mathrm{d}\nu}{\mathrm{d}\eta}(x) > 0\}$ and $E_2 = E \cap \{\frac{\mathrm{d}\nu}{\mathrm{d}\eta}(x) = 0\}$, we get

$$\begin{aligned}
&H(\rho|\rho_1 \otimes \eta) + H\big(\rho_1|\mathcal{Q}(\lambda)\big) + I^X(\eta) \\
&\quad = H\big(\rho|\mathcal{Q}(\lambda) \otimes \eta\big) \\
&\quad = \int_E H\big(\rho_x(\cdot)|\mathcal{Q}(\lambda)\big)\eta(\mathrm{d}x) \\
&\quad = \int_{E_1} H\Big(\mathcal{P}\Big(\lambda\frac{\mathrm{d}\nu}{\mathrm{d}\eta}(x)\Big)\Big|\mathcal{P}(\lambda)\Big)\eta(\mathrm{d}x) + \int_{E_2} H\big(\delta_0|\mathcal{P}(\lambda)\big)\eta(\mathrm{d}x) \\
&\quad = \int_{E_1} \lambda\Big(1 - \frac{\mathrm{d}\nu}{\mathrm{d}\eta}(x) + \frac{\mathrm{d}\nu}{d\eta}(x)\log\frac{\mathrm{d}\nu}{\mathrm{d}\eta}(x)\Big)\eta(\mathrm{d}x) + \lambda\eta(E_2) \\
&\quad = \lambda H(\nu|\eta),
\end{aligned}$$

hence $K(\nu) \le \lambda \inf_{\eta\in\mathcal{Z}} H(\nu|\eta)$.

6.5. Proof of Corollary 3.3

Let $\nu \in M_1(E)$ be such that $\nu \ll \mu$ for otherwise we already know that $\mathcal{K}(\nu;\mu) = +\infty$. We have

$$\begin{aligned}
\mathcal{K}(\nu;\mu) &= \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \Bigl\{ H(\rho|\rho_1 \otimes \mu) + \inf_{m>0} \bigl\{ H\bigl(\rho_1 | \mathcal{G}(\xi, m)\bigr)\bigr\}\Bigr\} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \bigl\{ H(\rho|\rho_1 \otimes \mu) + H\bigl(\rho_1 | \mathcal{G}(\xi, m)\bigr)\bigr\} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \bigl\{ H\bigl(\rho | \mathcal{G}(\xi, m) \otimes \mu\bigr)\bigr\} \\
&= \inf_{m>0} \inf_{\rho_x : F(\rho_x \otimes \mu) = \nu} \biggl\{ \int_E H\bigl(\rho_x | \mathcal{G}(\xi, m)\bigr)\mu(\mathrm{d}x)\biggr\},
\end{aligned}$$

where, to establish the result, we proceed as in the proof of Proposition 2.2.

6.6. Proof of Corollary 3.5

First, we consider $\alpha > 0$. Let $\nu \in M_1(E)$ be such that $\frac{\mu-(1-\alpha)\nu}{\alpha} \in M_1(E)$. Then $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ defined by $\rho_1 = \pi$ while the regular conditional distribution of its second marginal given the first is

$$\rho_{1/(1-\alpha)}(\mathrm{d}x) = \nu(\mathrm{d}x) \quad \text{and} \quad \rho_0(\mathrm{d}x) = \frac{\mu(\mathrm{d}x) - (1-\alpha)\nu(\mathrm{d}x)}{\alpha}$$

is the only element of $\mathcal{M}_1^1(\mathbb{R}_+ \times E)$ that satisfies $F(\rho) = \nu$, $\rho_1 = \pi$ and $\rho_2 = \mu$. Since for this particular ρ we have

$$H(\rho|\rho_1 \otimes \rho_2) = (1-\alpha)H(\nu|\mu) + \alpha H\biggl(\frac{\mu - (1-\alpha)\nu}{\alpha}\Big|\mu\biggr)$$

we obtain an upper-bound on $\mathcal{K}$ as announced. To prove the reverse inequality let $\nu \in M_1(E)$ be such that $\mathcal{K}(\nu;\mu) < \infty$ for otherwise the announced result trivially holds. Necessarily there exists a $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ such that $F(\rho) = \nu$, $\rho_1 = \pi$ and $\rho_2 = \mu$. These conditions are only met by the probability measure ρ introduced above. In particular $\frac{\mu-(1-\alpha)\nu}{\alpha}$ must be a probability and the reverse inequality holds.

If $\alpha = 0$, then $\pi = \delta_1$ so the only $\nu \in M_1(E)$ such that there exists $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ such that $F(\rho) = \nu$ and $\rho_2 = \mu$ is μ and necessarily $\rho = \delta_1 \otimes \mu$. The announced result follows.

6.7. Proof of Corollary 3.6

First, we consider $\alpha > 0$. Let $\nu \in M_1(E)$ be such that $\inf_{\zeta \in \mathcal{E}_\nu} \mathcal{U}(\nu, \zeta) < \infty$ holds. Then for every $\zeta \in \mathcal{E}_\nu$ we have $\frac{\zeta-(1-\alpha)\nu}{\alpha} \in M_1(E)$ and $\rho \in \mathcal{M}_1^1(\mathbb{R}_+ \times E)$ defined by $\rho_1 = \pi$ and

$$\rho_{1/(1-\alpha)}(\mathrm{d}x) = \nu(\mathrm{d}x) \quad \text{and} \quad \rho_0(\mathrm{d}x) = \frac{\zeta(\mathrm{d}x) - (1-\alpha)\nu(\mathrm{d}x)}{\alpha}$$

satisfies $F(\rho)=\nu$, $\rho_1=\pi$ and $\rho_2=\zeta$ and

$$H(\rho|\rho_1\otimes\rho_2)+I^W(\rho_1)+I^X(\rho_2)=(1-\alpha)H(\nu|\zeta)+\alpha H\biggl(\frac{\zeta-(1-\alpha)\nu}{\alpha}\Big|\zeta\biggr)+I^X(\zeta)$$

hence $K(\nu)\le\inf_{\zeta\in\mathcal{E}_\nu}\mathcal{U}(\nu,\zeta)$. To prove the reverse inequality, let us assume that $K(\nu)<\infty$. Then there exists a $\rho\in\mathcal{M}_1^1(\mathbb{R}_+\times E)$ such that $F(\rho)=\nu$ and $\rho_1=\pi$. Necessarily $\rho_{1/(1-\alpha)}(\mathrm{d}x)=\nu(\mathrm{d}x)$ and ρ_2 is such that $\rho_0(\mathrm{d}x)=\frac{\rho_2(\mathrm{d}x)-(1-\alpha)\nu(\mathrm{d}x)}{\alpha}\in M_1(E)$. Thus, $\mathcal{E}_\nu$ is non-empty and

$$\begin{aligned}H(\rho|\rho_1\otimes\rho_2)+I^W(\rho_1)+I^X(\rho_2)&=(1-\alpha)H(\nu|\rho_2)+\alpha H\biggl(\frac{\rho_2-(1-\alpha)\nu}{\alpha}\Big|\rho_2\biggr)+I^X(\rho_2)\\&\ge\inf_{\rho_2\in\mathcal{E}_\nu}\mathcal{U}(\nu,\rho_2).\end{aligned}$$

If $\alpha=0$, then $\pi=\delta_1$ and for every $\nu\in M_1(E)$ there is only one $\rho\in\mathcal{M}_1^1(\mathbb{R}_+\times E)$ such that $F(\rho)=\nu$ which is $\rho=\delta_1\otimes\nu$ and the announced result immediately follows.

6.8. Proof of Theorem 3.4

According to Corollary 3.1, Theorem 3.1 and Proposition 2.2 the sequence $(\mathcal{L}^n=\frac{1}{n}\times\sum_{i=1}^n W_i^n\delta_{(x_i^n,\dots,x_{i+k-1}^n)})_{n\ge1}$ obeys a LDP on $M_1(E^k)$ endowed with the weak convergence topology with good rate function $\mathcal{K}(\nu^{(k)};\mu^{(k)})=\frac{1}{k}H(\nu^{(k)}|\mu^{(k)})$. The announced result follows since the map $\mathcal{H}$ defined on $M_1(E^k)$ by $\mathcal{H}(\rho^{(k)})=\frac{1}{k}\sum_{i=1}^k\rho_i^{(k)}$ is continuous.

6.9. Proof of Corollary 3.7

Let $\nu\in M_1(E)$. Since $\mathcal{H}(\nu^{\otimes k})=\nu$ and in this particular case $\mu^{(k)}=\mu^{\otimes k}$ we get $\widetilde{\mathcal{K}}(\nu;\mu)\le\frac{1}{k}H(\nu^{\otimes k}|\mu^{\otimes k})=H(\nu|\mu)$. On the other hand for every $\rho^{(k)}$ such that $\mathcal{H}(\rho^{(k)})=\nu$ we have $H(\nu|\mu)=H(\frac{1}{k}\sum_{i=1}^k\rho_i^{(k)}|\mu)\le\frac{1}{k}\sum_{i=1}^k H(\rho_i^{(k)}|\mu)$ since $H(\cdot|\mu)$ is convex. To conclude the proof, just notice that $H(\rho^{(k)}|\mu^{\otimes k})=H(\rho^{(k)}|\bigotimes_{i=1}^k\rho_i^{(k)})+H(\bigotimes_{i=1}^k\rho_i^{(k)}|\mu^{\otimes k})\ge H(\bigotimes_{i=1}^k\rho_i^{(k)}|\mu^{\otimes k})=\sum_{i=1}^k H(\rho_i^{(k)}|\mu)$.

Supplementary Material

Large deviations for bootstrapped empirical measures (DOI: 10.3150/13-BEJ544SUPP; .pdf). In the supplemental article [26], we give the proofs of the sample weights LDPs stated in Section 3: Theorems 3.1, 3.2 and 3.3.

References

[1] Bahadur, R.R. (1969). *Some Limit Theorems in Statistics*. Philadelphia, PA: SIAM. MR0315820

[2] Barbe, P. and Bertail, P. (1995). *The Weighted Bootstrap*. *Lecture Notes in Statistics* **98**. New York: Springer. MR2195545

[3] Baxter, J.R. and Jain, N.C. (1988). A comparison principle for large deviations. *Proc. Amer. Math. Soc.* **103** 1235–1240. MR0955016

[4] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[5] Bolley, F. (2008). Separability and completeness for the Wasserstein distance. In *Séminaire de Probabilités XLI*. *Lecture Notes in Math.* **1934** 371–377. Berlin: Springer. MR2483740

[6] Broniatowski, M. and Cao, Z. (2013). Weighted sampling, maximum likelihood and minimum divergence estimators. Preprint, Université Pierre et Marie Curie, Paris, France.

[7] Chaganty, N.R. and Karandikar, R.L. (1996). Some properties of the Kullback–Leibler number. *Sankhyā A* **58** 69–80. MR1659055

[8] Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*. *Wiley Series in Telecommunications*. New York: Wiley. MR1122806

[9] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. New York: Springer. MR1619036

[10] Dinwoodie, I.H. and Zabell, S.L. (1992). Large deviations for exchangeable random vectors. *Ann. Probab.* **20** 1147–1166. MR1175254

[11] Donsker, M.D. and Varadhan, S.R.S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461. MR0428471

[12] Dudley, R.M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. Revised reprint of the 1989 original. MR1932358

[13] Dupuis, P. and Ellis, R.S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. New York: Wiley. MR1431744

[14] Efron, B. (1982). *The Jackknife*, *the Bootstrap and Other Resampling Plans*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. Philadelphia, PA: SIAM. MR0659849

[15] Eichelsbacher, P. and Schmock, U. (2002). Large deviations of U-empirical measures in strong topologies and applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 779–797. MR1931586

[16] Ganesh, A. and O'Connell, N. (2007). Large and moderate deviations for matching problems and empirical discrepancies. *Markov Process. Related Fields* **13** 85–98. MR2321752

[17] Gine, E. (1997). Lectures on some aspects of the bootstrap. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1996). *Lecture Notes in Math.* **1665** 37–151. Berlin: Springer. MR1490044

[18] Grunwald, M. (1996). Sanov results for Glauber spin-glass dynamics. *Probab. Theory Related Fields* **106** 187–232. MR1410687

[19] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. *Springer Series in Statistics*. New York: Springer. MR1145237

[20] Holmes, S. and Reinert, G. (2004). Stein's method for the bootstrap. In *Stein's Method*: *Expository Lectures and Applications*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **46** 95–136. Beachwood, OH: IMS. MR2118605

[21] Hult, H. and Nyquist, P. (2013). Large deviations for weighted empirical measures arising in importance sampling. Preprint, Royal Institute of Technology, Stockholm, Sweden.

[22] Léonard, C. and Najim, J. (2002). An extension of Sanov's theorem: Application to the Gibbs conditioning principle. *Bernoulli* **8** 721–743. MR1963659

[23] Lynch, J. and Sethuraman, J. (1987). Large deviations for processes with independent increments. *Ann. Probab.* **15** 610–627. MR0885133

[24] Najim, J. (2002). A Cramér type theorem for weighted random variables. *Electron*. *J*. *Probab*. **7** no. 4, 32 pp. (electronic). MR1887624

[25] Trashorras, J. (2008). Large deviations for symmetrised empirical measures. *J*. *Theoret*. *Probab*. **21** 397–412. MR2391251

[26] Trashorras, J. and Wintenberger, O. (2014). Supplement to "Large deviations for bootstrapped empirical measures." DOI:10.3150/13-BEJ544SUPP.

[27] Villani, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Providence, RI: Am. Math. Soc. MR1964483

Received September 2011 and revised June 2013