Large Deviations for a matching problem related to the ∞-Wasserstein distance

José Trashorras

Université Paris-Dauphine Ceremade
Place du maréchal de Lattre de Tassigny,
75775 Paris Cedex 16 France.
E-mail address: `trashorras@ceremade.dauphine.fr`

Abstract. Let (E, d) be a compact metric space, $X = (X_1, \ldots, X_n, \ldots)$ and $Y = (Y_1, \ldots, Y_n, \ldots)$ two independent sequences of independent E-valued random variables and $(L_n^X)_{n \geq 1}$ and $(L_n^Y)_{n \geq 1}$ the associated sequences of empirical measures. We establish a Large Deviation Principle for $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ where W_∞ is the ∞-Wasserstein distance,

$$W_\infty(L_n^X, L_n^Y) = \min_{\sigma \in \mathfrak{S}_n} \max_{1 \leq i \leq n} d(X_i, Y_{\sigma(i)})$$

where $\mathfrak{S}_n$ stands for the set of permutations of $\{1, \ldots, n\}$.

1. Introduction

There has been a lot of interest for years in considering the space $M^1(E)$ of Borel probability measures on a metric space (E, d) endowed with the so-called p-Wasserstein distances

$$W_p(\nu, \gamma) = \inf_{Q \in \mathcal{C}(\nu, \gamma)} \left\{ \left(\int_{E \times E} d(x, y)^p Q(dx, dy) \right)^{1/p} \right\}$$

where $p \in [1, \infty)$ and $\mathcal{C}(\nu, \gamma)$ stands for the set of Borel probability measures on E^2 with first marginal $Q_1 = \nu$ and second marginal $Q_2 = \gamma$, see Chapter 6 in Villani (2009) for a broad review. However, the ∞-Wasserstein distance equivalently defined by either

$$W_\infty(\nu, \gamma) = \lim_{p \to \infty} W_p(\nu, \gamma)$$

or

$$W_\infty(\nu, \gamma) = \inf_{Q \in \mathcal{C}(\nu, \gamma)} \sup \left\{ u, u \in \mathcal{S}(Q \circ d^{-1}) \right\} \tag{1.1}$$

where $\mathcal{S}(Q \circ d^{-1}) \subset E^2$ stands for the support of the probability measure $Q \circ d^{-1}$ has attracted much less attention so far.[1]

Let us give right now an elementary example in order to illustrate the difference between the W_p's and W_∞. Let $E = [0,1]$ endowed with the Euclidean distance and consider the sequence $(\nu^n = (1 - \frac{1}{n})\delta_0 + \frac{1}{n}\delta_1)_{n \geq 1}$. For every $n \geq 1$ the set $\mathcal{C}(\nu^n, \delta_0)$ has a single element $Q^n = \left(1 - \frac{1}{n}\right)\delta_{(0,0)} + \frac{1}{n}\delta_{(1,0)}$. It follows that for every $p \in [1,\infty)$ we have $W_p(\nu^n, \delta_0) = \left(\frac{1}{n}\right)^{\frac{1}{p}}$ thus $\lim_{n \to \infty} W_p(\nu^n, \delta_0) = 0$ while for every $n \geq 1$ we have $W_\infty(\nu^n, \delta_0) = 1$. In words, as it happens for the usual L^p and L^∞ norms on measurable functions spaces, perturbations so small that they do not impact W_p distances in the limit can have dramatic consequences when using the W_∞ distance instead. This naturally leads to wonder what happens at the level of asymptotic results, the Law of Large Numbers and the Large Deviation Principle, when random measures like empirical measures are involved and this was the main motivation in carrying out the present work.

The reference paper on W_∞ is Champion et al. (2008). There the authors consider measures with support on compact subsets of $\mathbb{R}^m$ ($m \geq 1$). Some of their results have been generalized from $\mathbb{R}^m$ to the setting of Polish spaces and general cost functions in Jylhä (2015). We recall them for latter use. In what follows (E, d) is a metric space.

Lemma 1.1. *(Proposition 2.1 in Champion et al., 2008 and Theorem 2.6 in Jylhä, 2015) For any two $\nu, \gamma \in M^1(E)$ with compact supports there exists at least one $Q \in \mathcal{C}(\nu, \gamma)$ such that*

$$W_\infty(\nu, \gamma) = \sup\left\{u, u \in \mathcal{S}(Q \circ d^{-1})\right\}.$$

They further establish the existence of nice solutions to the problem (1.1) which they call infinitely cyclically monotone couplings.

Definition 1.2. A probability measure $P \in M^1(E^2)$ is called infinitely cyclically monotone if and only if for every integer $n \geq 2$, every $(x_1, y_1), \ldots, (x_n, y_n) \in \mathcal{S}(P)$ and every $\sigma \in \mathfrak{S}_n$ we have

$$\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)}). \tag{1.2}$$

Lemma 1.3. *(Theorem 3.2 and Theorem 3.4 in Champion et al., 2008 and Theorem 2.16 in Jylhä, 2015) For any two $\gamma, \nu \in M^1(E)$ with compact supports there exists at least one infinitely cyclically monotone $P \in \mathcal{C}(\gamma, \nu)$ such that*

$$W_\infty(\gamma, \nu) = \sup\left\{u, u \in \mathcal{S}(P \circ d^{-1})\right\}.$$

In order to check that $P \in M^1(E^2)$ is infinitely cyclically monotone we only need to know its support. The way mass is spread over $\mathcal{S}(P)$ does not matter : this property is concerned with subsets of E^2 rather than with fully specified probability measures. This is the reason why we will also call infinitely cyclically monotone subsets of E^2 that satisfy (1.2). Infinitely cyclically monotone couplings are nice solutions to the problem (1.1) in the sense that

---

[1]Let us recall that the support $\mathcal{S}(P)$ of a Borel probability measure P on a metric space (E, d) is defined by $x \in \mathcal{S}(P)$ if and only if for every $\varepsilon > 0$, $P(B(x, \varepsilon)) > 0$ where $B(x, \varepsilon)$ stands for the open ball centered at x with radius ε. If E is separable, in particular if E is compact, $\mathcal{S}(P)$ is closed, see e.g. Theorem 2.1 in Parthasarathy (1967).

Lemma 1.4. *(Theorem 3.4 in Champion et al., 2008 and Theorem 2.17 in Jylhä, 2015) Any infinitely cyclically monotone $P \in M^1(E^2)$ satisfies*

$$W_\infty(P_1, P_2) = \sup \left\{ u, u \in \mathcal{S}(P \circ d^{-1}) \right\}.$$

Indeed, once an infinitely cyclically monotone $A \subset E^2$ is given, no matter how one spreads mass over A provided the couple that maximizes $(x,y) \mapsto d(x,y)$ over A is covered, the marginals of the resulting probability measure over E^2 will always be at the same W_∞ distance. This will show useful when constructing approximations. For any two $\gamma, \nu \in M^1(E)$ we shall denote by $\mathcal{C}_\infty(\nu, \gamma)$ the set of infinitely cyclically monotone couplings of ν and γ.

Our framework is the following: We are given a compact metric space (E, d). Without loss of generality we can assume that $\sup_{x,y \in E} d(x,y) = 1$. Let $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ be two independent sequences of E-valued independent random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that all the X_i's (resp. Y_i's) have the same distribution μ^1 (resp. μ^2). For every $n \geq 1$ we consider the empirical measures

$$L_n^X = \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{and} \quad L_n^Y = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}.$$

Since for every $n \geq 1$ every $Q \in \mathcal{C}(L_n^X, L_n^Y)$ can be represented as a bi-stochastic matrix and since, according to the Birkhoff-Von Neumann Theorem, every bi-stochastic matrix is a convex combination of permutation matrices (see e.g. Theorem 5.5.1 in Serre, 2002) we have

$$W_\infty(L_n^X, L_n^Y) = \min_{\sigma \in \mathfrak{S}_n} \max_{1 \leq i \leq n} d(X_i, Y_{\sigma(i)}) \tag{1.3}$$

where $\mathfrak{S}_n$ stands for the set of permutations of $\{1, \ldots, n\}$. Hence, computing $W_\infty(L_n^X, L_n^Y)$ is nothing but solving a minimax matching problem which is a fundamental combinatorial question. Equality (1.3) can be rephrased as

$$W_\infty(L_n^X, L_n^Y) = d_{\mathcal{H}}(\mathcal{S}(L_n^X), \mathcal{S}(L_n^Y)) \tag{1.4}$$

where $\mathcal{S}(L_n^X) = \{X_1, \ldots, X_n\}$, $\mathcal{S}(L_n^Y) = \{Y_1, \ldots, Y_n\}$ and $d_{\mathcal{H}}$ is the so-called Hausdorff distance. Let us recall that the Hausdorff distance is defined on the set of non-empty closed subsets of E by

$$d_{\mathcal{H}}(A, B) = \max \left\{ \sup_{x \in A} \inf_{y \in B} d(x,y), \sup_{y \in B} \inf_{x \in A} d(x,y) \right\}$$

see e.g. Section 6.2.2 in Delfour and Zolésio (2011). It is the essential tool in many applications like image processing Huttenlocher et al. (1993), object matching Sim et al. (1999), face detection Jesorsky et al. (2001) and evolutionary optimization Schutze et al. (2012) to name just a few.

The first results on $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ have been focused on concentration inequalities for independent $E = [0,1]^2$-valued X_i's and Y_i's with common distribution $\mu^1 = \mu^2 = \lambda$ where λ stands for the uniform distribution over $[0,1]^2$. In Leighton and Shor (1989) the authors establish that there exists a $K > 0$ such that

$$\frac{1}{K} n^{-1/2} (\log n)^{3/4} \leq W_\infty(L_n^X, L_n^Y) \leq K n^{-1/2} (\log n)^{3/4}$$

with probability $1 - o(1)$. Using majorizing measures, it is demonstrated in Talagrand (1994) that the latter result is a particular case of a general property of empirical discrepancies.

Concerning Large Deviations, in Ganesh and O'Connell (2007) the authors prove that if $E = [0,1]^2$ and $\mu^1 = \lambda$ is the uniform distribution over E then the sequence $(W_\infty(L_n^X, \lambda))_{n \geq 1}$ obeys an LDP with good rate function

$$T_\infty(x) = \inf_{\substack{\nu \in M^1(E) \\ W_\infty(\nu,\lambda)=x}} \{H(\nu|\lambda)\}.$$

where for any two $\rho, \gamma \in M^1(E)$

$$H(\rho|\gamma) = \begin{cases} \int_E \log \frac{d\rho}{d\gamma} d\rho & \text{if } \rho \ll \gamma \\ \infty & \text{otherwise.} \end{cases}$$

To obtain this result they prove that in this particular case the map $\nu \mapsto W_\infty(\nu, \lambda)$ is continuous with respect to the weak convergence topology, and then apply the contraction principle, see Theorem 4.2.1 in Dembo and Zeitouni (1998). They further show in the same framework as here, i.e. E-valued X_i's and Y_i's where (E, d) is any compact metric space, that $(W_1(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with good rate function

$$I_1(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_1(\nu^1,\nu^2)=x}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}.$$

Their proof relies on the fact that, according to the Kantorovitch-Rubinstein Theorem (see e.g. Theorem 11.8.2 in Dudley, 2002), when E is compact W_1 generates the weak convergence topology. As a consequence $(L_n^X)_{n \geq 1}$ and $(L_n^Y)_{n \geq 1}$ satisfy an LDP on $M^1(E)$ endowed with W_1 and again the contraction principle leads to the aforementioned LDP for $(W_1(L_n^X, L_n^Y))_{n \geq 1}$. Following the same approach, one can deduce from Theorem 1.1 in Wang et al. (2010) that for every $p \in [1, \infty)$ the sequence $(W_p(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with good rate function

$$I_p(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_p(\nu^1,\nu^2)=x}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}.$$

Again in Ganesh and O'Connell (2007) it is conjectured that $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ obeys an LDP with rate function

$$I_\infty(x) = \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1,\nu^2)=x}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}$$

Our main result is that if instead of I_∞ one considers its lower semi-continuous regularization J_∞ which is defined by

$$J_\infty(x) = \sup_{\delta > 0} \inf_{y \in B(x,\delta)} I_\infty(y),$$

see e.g. Chapter 1 in Rockafellar and Wets (1998), then

Theorem 1.5. *The sequence $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ satisfies an LDP on $[0,1]$ with good rate function $J_\infty(x)$.*

In Section 3.1 we show that if e.g. $E = [0,1]$ is endowed with the usual Euclidean distance and $\mu^1 = \mu^2 = \mu$ admits $f(x) = \frac{3}{2}\mathbb{1}_{[0,\frac{1}{3}]}(x) + \frac{3}{2}\mathbb{1}_{[\frac{2}{3},1]}(x)$ as a density w.r.t. the Lebesgue measure then, due to the disconnectedness of the support of μ, I_∞ fails to be lower semi-continuous. Hence, I_∞ and J_∞ do not always

coincide. This example illustrates a general feature of the problem considered here : the connectedness properties of the support of the reference measures are the key ingredient in the behavior of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$. We will say more about that in Proposition 1.13 below.

One might wonder why the idea that leads from an LDP for $(L_n^X)_{n \geq 1}$ and $(L_n^Y)_{n \geq 1}$ in $(M^1(E), W_1)$ to an LDP for $(W_1(L_n^X, L_n^Y))_{n \geq 1}$ is not applicable to the analysis of the LD properties of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$. Actually an LDP for $(L_n^X)_{n \geq 1}$ can not hold when $M^1(E)$ is endowed with W_∞, at least at this level of generality. Indeed, consider the probability measure μ^1 on $E = [0, 1]$ which density with respect to the Lebesgue measure is $g(x) = 2(\mathbb{1}_{[0, \frac{1}{4}]}(x) + \mathbb{1}_{[\frac{3}{4}, 1]}(x))$ and assume that $(L_n^X)_{n \geq 1}$ satisfies an LDP on $(M^1(E), W_\infty)$ with some rate function $\mathcal{R}$. Clearly for every odd integer n we have $\mathbb{P}(W_\infty(L_n^X, \mu^1) < 3/8) = 0$ hence for every $\nu \in B(\mu^1, 3/8)$ we should necessarily have $\mathcal{R}(\nu) = \infty$. So we would get

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, \mu^1) \leq \frac{1}{4}) = -\infty \tag{1.5}$$

but for every even n we have $\mathbb{P}(W_\infty(L_n^X, \mu^1) \leq \frac{1}{4}) \geq \binom{n}{n/2} 2^{-n}$ which contradicts (1.5). Again this is due to the lack of connectedness of the support of μ^1 and, in a sense, Theorem 1.5 is the only possible "Sanov-like" statement involving W_∞ without any additional assumption.

Now let us give some facts about the zeros of J_∞. First notice that since

$$0 \leq I_\infty(W_\infty(\mu^1, \mu^2)) \leq H(\mu^1|\mu^1) + H(\mu^2|\mu^2) = 0$$

we necessarily have $J_\infty(W_\infty(\mu^1, \mu^2)) = 0$. Actually, due to the highly discontinuous nature of W_∞, $W_\infty(\mu^1, \mu^2)$ need not be the only zero of J_∞. We shall first prove the following

Lemma 1.6. *Let $x \in [0, 1]$ be such that $J_\infty(x) = 0$. Necessarily $x \geq W_\infty(\mu^1, \mu^2)$.*

The zeros of J_∞ are related to some particular elements of E^2 as defined next

Definition 1.7. We shall say that $(a, b) \in E^2$ is a couple of directly connected points and write $a \leftrightarrow b$ if and only if $a \in \mathcal{S}(\mu^1), b \in \mathcal{S}(\mu^2)$ and there exists $Q \in \mathcal{C}_\infty(\mu^1, \mu^2)$ such that for every integer $N \geq 2$, every $(\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N) \in \mathcal{S}(Q)$ and every $\sigma \in \mathfrak{S}_N$ we have $d(a, b) \leq \max_{i=1,\ldots,N} \{d(\alpha_i, \beta_{\sigma(i)})\}$ where $(\alpha_1, \beta_1) = (a, b)$.

Notice that in this definition (a, b) need not be an element of $\mathcal{S}(Q)$. In words, $a \leftrightarrow b$ if and only if there exists $Q \in \mathcal{C}_\infty(\mu^1, \mu^2)$ such that $\{(a, b)\} \cup \mathcal{S}(Q)$ is still an infinitely cyclically monotone subset of E^2. Consider the set

$$\mathcal{Z}_{\mu^1, \mu^2} = \{x \in [0, 1] : \exists (a, b) \in E^2 \text{ such that } a \leftrightarrow b \text{ and } x = d(a, b)\}$$

Proposition 1.8. *$J_\infty(x) = 0$ if and only if $x \in \mathcal{Z}_{\mu^1, \mu^2}$.*

The notion of directly connected points takes a simpler form when $\mu^1 = \mu^2$.

Proposition 1.9. *If $\mu^1 = \mu^2 = \mu$ then $a, b \in E$ are directly connected if and only if $a, b \in \mathcal{S}(\mu)$ and for every integer $N \geq 3$ and every sequence $\alpha_1, \ldots, \alpha_N$ of elements of $\mathcal{S}(\mu)$ such that $\alpha_1 = a$ and $\alpha_N = b$ there is at least one integer L such that $1 \leq L \leq N - 1$, and $d(a, b) \leq d(\alpha_L, \alpha_{L+1})$.*

In words, if $\mu^1 = \mu^2 = \mu$ then $a \leftrightarrow b$ if and only if one can not decompose a journey from a to b into several stages between elements of $\mathcal{S}(\mu)$ such that each of these stages has length strictly smaller than $d(a,b)$. As an example consider again $E = [0,1]$ and $\mu^1 = \mu^2 = \mu$ with density $f(x) = \frac{3}{2}\mathbb{1}_{[0,\frac{1}{3}]}(x) + \frac{3}{2}\mathbb{1}_{[\frac{2}{3},1]}(x)$ w.r.t. the Lebesgue measure. Then $\frac{1}{3} \leftrightarrow \frac{2}{3}$ hence $J_\infty(\frac{1}{3}) = 0$ as it is established through a direct computation in Section 3.1.

Finally we investigate the almost sure behaviour of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$ as $n \to \infty$ when $\mu^1 = \mu^2 = \mu$. The elements of $\mathcal{Z}_\mu$, hence the asymptotic behaviour of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$, are directly related to the connectedness properties of $\mathcal{S}(\mu)$. So let us write $\mathcal{S}(\mu) = \cup_{i=1}^D A_i$ where the A_i are the connected components of $\mathcal{S}(\mu)$. Here $D = 1$ means that $\mathcal{S}(\mu)$ is connected and we do not exclude the possibility that $D = \infty$. We have

Lemma 1.10. $\mathcal{Z}_\mu = \{0\}$ if and only if $\mathcal{S}(\mu)$ is connected.

For every $x \in \mathcal{S}(\mu)$ let us denote by $A_{i(x)}$ the connected component of $\mathcal{S}(\mu)$ that contains x and for any two closed $B, C \subset E$ we set

$$d(B,C) = \inf_{(x,y) \in B \times C} d(x,y).$$

Lemma 1.11. For every $a, b \in \mathcal{S}(\mu)$, if $a \leftrightarrow b$ then $d(a,b) = d(A_{i(a)}, A_{i(b)})$.

For any two connected components A_i, A_j of $\mathcal{S}(\mu)$ we shall write $A_i \leftrightarrow A_j$ if and only if there exists $a \in A_i$ and $b \in A_j$ such that $a \leftrightarrow b$. It directly follows from Lemma 1.11 that

$$\mathcal{Z}_\mu = \left\{ d(A_i, A_j), \ (i,j) \in \{1, \ldots, D\}^2 \text{ such that } A_i \leftrightarrow A_j \right\}.$$

In order to formulate our result we need the following

Lemma 1.12. If $\mathcal{S}(\mu)$ is not connected every $x \in \mathcal{Z}_\mu$ such that $x \neq 0$ is an isolated point.

Hence, if $\mathcal{S}(\mu)$ is not connected, we can order the elements of $\mathcal{Z}_\mu$

$$\beta_1 > \beta_2 > \cdots > \beta_n > \cdots$$

where $\beta_1 = \max \mathcal{Z}_\mu$, and for every integer $i \geq 2$

$$\beta_i = \begin{cases} \max \mathcal{Z}_\mu \setminus \{\beta_1, \ldots, \beta_{i-1}\} & \text{if } \beta_{i-1} \neq 0 \\ 0 & \text{if } \beta_{i-1} = 0. \end{cases}$$

Proposition 1.13.   (1) If $\mathcal{S}(\mu)$ is connected then $W_\infty(L_n^X, L_n^Y) \to 0$ almost surely.
   (2) If $\mathcal{S}(\mu)$ is not connected and
      (a) there exists only two connected components A_i, A_j of $\mathcal{S}(\mu)$ such that $A_i \leftrightarrow A_j$ and $d(A_i, A_j) = \beta_1$ then

$$\liminf W_\infty(L_n^X, L_n^Y) = \beta_2 \quad \text{and} \quad \limsup W_\infty(L_n^X, L_n^Y) = \beta_1 \qquad (1.6)$$

        almost surely.
      (b) there are more than two connected components A_i, A_j of $\mathcal{S}(\mu)$ such that $A_i \leftrightarrow A_j$ and $d(A_i, A_j) = \beta_1$ then $W_\infty(L_n^X, L_n^Y) \to \beta_1$ almost surely.

Hence, if $E = [0,1]$ and μ admits $f(x) = \frac{3}{2}\mathbb{1}_{[0,\frac{1}{3}]}(x) + \frac{3}{2}\mathbb{1}_{[\frac{2}{3},1]}(x)$ as a density w.r.t. the Lebesgue measure then $\limsup W_\infty(L_n^X, L_n^Y) = \frac{1}{3}$ almost surely and $\liminf W_\infty(L_n^X, L_n^Y) = 0$ almost surely.

The paper is structured as follows : Section 2 is devoted to the proof of Theorem 1.5. In Section 3 we discuss the properties of I_∞ and J_∞. Section 4 is concerned with the almost sure asymptotic behaviour of $(W_\infty(L_n^X, L_n^Y))_{n\geq 1}$ while Section 5 deals with the proof of some complementary results.

2. Proof of Theorem 1.5

The basic idea in the proof of Theorem 1.5 is to partition E in such a way that we are essentially led to consider finite set valued X_i's and Y_i's. Indeed, on the one hand we will see that W_∞ is well-behaved with respect to partitioning (see Lemma 2.1 below) and on the other hand proceeding this way reduces the computation of probabilities to simple classical combinatorial estimates. In order to go from the particular case, i.e. E finite, to the general one we shall need some results on the weak convergence of nets of probability measures. In Section 2.1 we give an account on partitions, nets and the weak convergence topology. The proof of all the statements there is postponed to Section 5. The LD lower bound is established in Section 2.2 while the LD upper bound is derived in Section 2.3.

2.1. *Some facts about partitions of E, nets and the weak convergence topology.* Let $\mathcal{P}$ be the set of finite measurable partitions of E into non-empty sets. To every $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$ we associate once for all through Section 2 a family $(s_1, \ldots, s_L) \in E^L$ such that for every $1 \leq i \leq L$ we have $s_i \in A_i$. We further associate to every $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$ a map π as follows

$$
\begin{array}{rcl}
\pi: \quad M^1(E) & \to & M^1(\{s_1, \ldots, s_L\}) \\
\nu & \mapsto & \sum_{i=1}^{L} \nu(A_i)\delta_{s_i}.
\end{array}
$$

Finally for every $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$ we define its maximal diameter as

$$
\Delta(\Pi) = \max_{1 \leq i \leq L} \sup_{x,y \in A_i} d(x,y).
$$

The following result links the W_∞ distance between elements of $M^1(E)$ to the analogue W_∞ distance between their contractions through π.

Lemma 2.1. *For every $\nu^1, \nu^2 \in M^1(E)$ and every $\Pi \in \mathcal{P}$ we have*

$$
|W_\infty(\nu^1, \nu^2) - W_\infty(\pi(\nu^1), \pi(\nu^2))| \leq 2\Delta(\Pi).
$$

Let $\Pi = (A_1, \ldots, A_L)$ and $\mathcal{K} = (B_1, \ldots, B_R)$ be two elements of $\mathcal{P}$. We say that $\mathcal{K}$ is a *refinement* of Π if and only if for every $1 \leq i \leq L$, there exists $J_i \subset \{1, \ldots, R\}$ such that $A_i = \cup_{j \in J_i} B_j$, and we denote it by $\Pi \preceq \mathcal{K}$. This makes $(\mathcal{P}, \preceq)$ a directed set. Let us recall that $(J, \trianglelefteq)$ is a *directed set* if and only if J is a non-empty set and $\trianglelefteq$ is a reflexive and transitive relation on J such that for every $i, j \in J$ there exists a $k \in J$ such that $i \trianglelefteq k$ and $j \trianglelefteq k$. We introduce a general directed set $(J, \trianglelefteq)$ since we will use both $(\mathcal{P}, \preceq)$ and $(\mathbb{N}, \leq)$ as directed sets and we do not want to be too specific in the results below. We call *net* any map $(P^j)_{j \in J}$ defined on a directed set $(J, \trianglelefteq)$. A topological space $(T, \mathcal{T})$-valued net $(P^j)_{j \in J}$ is said to *converge* to some $P \in T$ if and only if for every neighbourhood U of P there exists a $j(U) \in J$ such

that for every $k \in J$ satisfying $j(U) \trianglelefteq k$ we have $P^k \in U$. We shall denote this $\lim_{j \in J} P^j = P$. We call *subnet* of $(P^j)_{j \in J}$ any sub-family $(P^l)_{l \in L}$ parametrized by a *cofinal* L, i.e. a subset $L \subset J$ such that for every $j \in J$ there exists $l \in L$ satisfying $j \trianglelefteq l$. (While this is not the most general definition of a subnet it will be sufficient for the problem considered here). Let us recall that a topological space $(T, \mathcal{T})$ is compact if and only if every net in T admits a subnet that converges to some point in T. Finally, for every real-valued net $(P^j)_{j \in J}$ we shall consider as usual

$$\limsup_{j \in J} P^j = \inf_{j \in J} \sup_{i:j \trianglelefteq i} P^i \text{ and } \liminf_{j \in J} P^j = \sup_{j \in J} \inf_{i:j \trianglelefteq i} P^i.$$

For this and other questions related to nets we refer to e.g. Kelley (1975). The following lemmas are a consequence of a kind of Portmanteau result for nets of probability measures that will be derived in Section 5.

Lemma 2.2. *Every net $(P^j)_{j \in J}$ of probability measures supported on $\mathbb{R}$ that converges weakly to some probability measure P satisfies*

$$\limsup_{j \in J} \sup \mathcal{S}(P^j) \geq \sup \mathcal{S}(P). \tag{2.1}$$

Lemma 2.3. *Let $(P^j)_{j \in J}$ be a net of Borel probability measures on a compact metric space (Y, δ) that converges weakly to some probability measure P. For every $x \in \mathcal{S}(P)$ and every $j \in J$ there exists an $x^j \in \mathcal{S}(P^j)$ such that the net $(x^j)_{j \in J}$ converges to x.*

Lemma 2.4. *Every $P \in M^1(E^2)$ which is the limit in the weak convergence topology of a net $(P^j)_{j \in J}$ of infinitely cyclically monotone elements of $M^1(E^2)$ is infinitely cyclically monotone.*

2.2. *LD lower bound.* For every integer $n \geq 1$ we consider

$$M^{1,n}(E) = \left\{ \nu \in M^1(E) : \text{there exists } (x_1, \ldots, x_n) \in E^n \text{ such that } \nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \right\}.$$

The following lemma is the key point in the proof of the LD lower bound

Lemma 2.5. *For every $\varepsilon > 0$, every $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/2$ and every $\nu^1, \nu^2 \in M^1(E)$ there exists two sequences $(\nu^{1,n})_{n \geq 1}$ and $(\nu^{2,n})_{n \geq 1}$ such that*

(1) *For every $n \geq 1$ we have $\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E)$.*
(2) *For every $1 \leq i \leq L$ we have $\nu^{1,n}(A_i) \to \nu^1(A_i)$ and $\nu^{2,n}(A_i) \to \nu^2(A_i)$.*
(3) *There exists an N_0 such that for every $n \geq N_0$ we have*

$$|W_\infty(\nu^1, \nu^2) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon.$$

Proving Lemma 2.5 actually requires one more lemma

Lemma 2.6. *For every $Q \in M^1(E^2)$ and every $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$ there exists a sequence $(Q^n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i^n, y_i^n)})_{n \geq 1}$ of elements of $M^1(E^2)$ such that*

(1) *For every $1 \leq u, v \leq L$ we have $Q^n(A_u \times A_v) \to Q(A_u \times A_v)$.*
(2) *The sequence $(Q^n)_{n \geq 1}$ converges weakly to Q.*
(3) *If $Q(A_u \times A_v) = 0$ then for every $n \geq 1$ we have $Q^n(A_u \times A_v) = 0$.*

Proof of Lemma 2.6 Let Q be an element of $M^1(E^2)$ and $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$. Let $(X_i, Y_i)_{i \geq 1}$ be a sequence of independent and Q-identically distributed random couples. According to the strong Law of Large Numbers there exists an event B of probability 1 such that for every $1 \leq u, v \leq L$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{A_u \times A_v}(X_i, Y_i) \to Q(A_u \times A_v)$$

on B. Moreover, since E^2 is separable (it is compact), according to Varadarajan's Lemma, see e.g. Theorem 11.4.1 in Dudley (2002), there exists an event C of probability 1 such that

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_i)} \overset{w}{\to} Q$$

on C where $\overset{w}{\to}$ stands for the weak convergence of probability measures. Hence almost every realization of $\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_i)}\right)_{n \geq 1}$ can play the role of $(Q^n)_{n \geq 1}$ and the conclusions of Lemma 2.6 follow. $\square$

Proof of Lemma 2.5 Let ν^1 and ν^2 be two elements of $M^1(E)$, $\varepsilon > 0$ and $\Pi \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/2$. According to Lemma 1.3 there is a $Q \in \mathcal{C}(\nu^1, \nu^2)$ such that

$$W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(Q \circ d^{-1}).$$

Let $(Q^n)_{n \geq 1}$ be the sequence of elements of $M^1(E^2)$ associated to Q and Π by Lemma 2.6. We shall prove that $(\nu^{1,n} = Q_1^n)_{n \geq 1}$ an $(\nu^{2,n} = Q_2^n)_{n \geq 1}$ meet the conditions of Lemma 2.5. First, (1) and (2) in Lemma 2.5 are clearly satisfied with these $(\nu^{1,n})_{n \geq 1}$ and $(\nu^{2,n})_{n \geq 1}$. Now, due to the definition of W_∞, for every $n \geq 1$ we have

$$W_\infty(\nu^{1,n}, \nu^{2,n}) \leq \sup \mathcal{S}(Q^n \circ d^{-1})$$

and due to (1) and (3) in Lemma 2.6 there exists N_1 such that for every $n \geq N_1$ we have

$$\sup \mathcal{S}(Q^n \circ d^{-1}) \leq \sup \mathcal{S}(Q \circ d^{-1}) + \varepsilon$$

hence $W_\infty(\nu^{1,n}, \nu^{2,n}) \leq W_\infty(\nu^1, \nu^2) + \varepsilon$. So we are left to prove that there exists an N_2 such that for every $n \geq N_2$ we have $W_\infty(\nu^1, \nu^2) - \varepsilon \leq W_\infty(\nu^{1,n}, \nu^{2,n})$. Let us assume that this is not true i.e. that there exists a sequence $(n_k)_{k \geq 1}$ such that for every $k \geq 1$ we have

$$W_\infty(\nu^1, \nu^2) - \varepsilon > W_\infty(\nu^{1,n_k}, \nu^{2,n_k}). \tag{2.2}$$

According to Lemma 1.1 for every $k \geq 1$ there exists a $C^{n_k} \in \mathcal{C}(\nu^{1,n_k}, \nu^{2,n_k})$ such that

$$W_\infty(\nu^{1,n_k}, \nu^{2,n_k}) = \sup \mathcal{S}(C^{n_k} \circ d^{-1}).$$

The sequence $(C^{n_k})_{k \geq 1}$ admits a weakly converging subsequence since $M^1(E^2)$ is compact for the weak convergence topology, see Theorem 6.4 in Parthasarathy (1967). By a slight abuse of notation we still denote $(C^{n_k})_{k \geq 1}$ this converging subsequence. Let C be its limit. Due to (2) in Lemma 2.6 we necessarily have $C_1 = \nu^1$ and $C_2 = \nu^2$. Moreover $(C^{n_k} \circ d^{-1})_{k \geq 1}$ is a sequence of probability

measures on $\mathbb{R}$ that weakly converges to $C \circ d^{-1}$. Combining (2.2) with Lemma 2.2 we have

$$
\begin{aligned}
W_\infty(\nu^1, \nu^2) - \varepsilon & \geq \limsup_{k\to\infty} W_\infty(\nu^{1,n_k}, \nu^{2,n_k}) \\
& = \limsup_{k\to\infty} \sup \mathcal{S}(C^{n_k} \circ d^{-1}) \\
& \geq \sup \mathcal{S}(C \circ d^{-1}) \\
& \geq W_\infty(\nu^1, \nu^2)
\end{aligned}
$$

which can not be. The announced result follows. $\square$

Proof of the LD lower bound. As usual, in order to prove the LD lower bound it is sufficient to prove that for every $x \in [0,1]$ and every $\varepsilon > 0$ we have

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]x - \varepsilon, x + \varepsilon[) \geq -J_\infty(x). \tag{2.3}
$$

In particular we can assume that $J_\infty(x) < \infty$ for otherwise (2.3) trivially holds. If $J_\infty(x) < \infty$ then for every $m \geq 1$ there exists an y_m such that $|x - y_m| < 1/m$, $I_\infty(y_m) < \infty$ and $\lim_{m\to\infty} I_\infty(y_m) = J_\infty(x)$. Now let us assume that for every m and every $\nu^1, \nu^2 \in M^1(E)$ such that $W_\infty(\nu^1, \nu^2) = y_m$ we have

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4[) \geq -H(\nu^1|\mu^1) - H(\nu^2|\mu^2). \tag{2.4}
$$

It follows that for every m large enough and every $\nu^1, \nu^2 \in M^1(E)$ such that $W_\infty(\nu^1, \nu^2) = y_m$

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]x - \varepsilon, x + \varepsilon[) \geq
$$

$$
\begin{aligned}
& \geq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4[) \\
& \geq -H(\nu^1|\mu^1) - H(\nu^2|\mu^2)
\end{aligned}
$$

hence

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]x - \varepsilon, x + \varepsilon[) \geq
$$

$$
\begin{aligned}
& \geq \sup_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = y_m}} -H(\nu^1|\mu^1) - H(\nu^2|\mu^2) \\
& \geq - \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = y_m}} H(\nu^1|\mu^1) + H(\nu^2|\mu^2) \\
& \geq -I_\infty(y_m)
\end{aligned}
$$

which leads to (2.3) by letting $m \to \infty$. Hence it is sufficient to establish (2.4) and this is what we do now.

Let $\varepsilon > 0$, $y_m \in [0,1]$ be such that $I_\infty(y_m) < \infty$ and $\nu^1, \nu^2 \in M^1(E)$ be such that $W_\infty(\nu^1, \nu^2) = y_m$. According to Lemma 2.5 to any $\Pi = (A_1, \dots, A_L) \in \mathcal{P}$ with $\Delta(\Pi) \leq \varepsilon/32$ we can associate two sequences $(\nu^{1,n})_{n\geq 1}$ and $(\nu^{2,n})_{n\geq 1}$ and an integer N_0 such that for every $n \geq 1$ we have $\nu^{1,n}, \nu^{2,n} \in M^{1,n}(E)$ and for every $n \geq N_0$ one has

$$
\mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4[) =
$$

$$
\begin{aligned}
& = \mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^1, \nu^2)| < \varepsilon/4) \\
& \geq \mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8).
\end{aligned}
$$

We obviously have

$$|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq |W_\infty(L_n^X, L_n^Y) - W_\infty(\pi(L_n^X), \pi(L_n^Y))|+$$

$$+|W_\infty(\pi(L_n^X), \pi(L_n^Y)) - W_\infty(\pi(\nu^{1,n}), \pi(\nu^{2,n}))|+$$

$$+|W_\infty(\pi(\nu^{1,n}), \pi(\nu^{2,n})) - W_\infty(\nu^{1,n}, \nu^{2,n})|$$

so due to Lemma 2.1 we get

$$\{\pi(L_n^X) = \pi(\nu^{1,n})\} \cap \{\pi(L_n^Y) = \pi(\nu^{2,n})\} \subset$$

$$\subset \{|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8\}$$

hence

$$\mathbb{P}(|W_\infty(L_n^X, L_n^Y) - W_\infty(\nu^{1,n}, \nu^{2,n})| \leq \varepsilon/8) \geq$$

$$\geq \mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\} \cap \{\pi(L_n^Y) = \pi(\nu^{2,n})\})$$

$$= \mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\})\mathbb{P}(\{\pi(L_n^Y) = \pi(\nu^{2,n})\})$$

since the sequences $(X_n)_{n \geq 1}$ and $(Y_n)_{n \geq 1}$ are independent. It follows from elementary combinatorics (see e.g. Lemma 2.1.9 in Dembo and Zeitouni, 1998) that

$$\mathbb{P}(\{\pi(L_n^X) = \pi(\nu^{1,n})\}) \geq (n+1)^{-L} \exp^{-nH(\pi(\nu^{1,n})|\pi(\mu^1))}$$

and the analogue for $\mathbb{P}(\{\pi(L_n^Y) = \pi(\nu^{2,n})\})$ also holds true. As a consequence

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in]y_m - \varepsilon/4, y_m + \varepsilon/4[) \geq$$

$$\geq -\limsup_{n \to \infty}(H(\pi(\nu^{1,n})|\pi(\mu^1)) + H(\pi(\nu^{2,n})|\pi(\mu^2)))$$

$$= -H(\pi(\nu^1)|\pi(\mu^1)) - H(\pi(\nu^2)|\pi(\mu^2)) \tag{2.5}$$

$$\geq -H(\nu^1|\mu^1) - H(\nu^2|\mu^2) \tag{2.6}$$

where (2.5) is due to (2) in Lemma 2.5 and (2.6) comes from the fact that

$$H(\pi(\nu^1)|\pi(\mu^1)) \leq H(\nu^1|\mu^1)$$

for every $\Pi \in \mathcal{P}$, see e.g. Theorem 1.4.3 in Dupuis and Ellis (1997).

2.3. *LD upper bound.*

We first establish the LD upper bound under the assumption that E is finite. Then we extend this result to the general case.

2.3.1. *A particular case.*

Here we assume that $E = \{a_1, \ldots, a_N\}$. Since for every $n \geq 1$ we have $W_\infty(L_n^X, L_n^Y) \in \Gamma = \{d(a_i, a_j), 1 \leq i, j \leq N\}$ which is a finite set it is sufficient in order to establish the LD upper bound to show that for every $\delta \in \Gamma$ we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) = \delta) \leq - \inf_{\substack{\nu^1, \nu^2 \in M^1(E) \\ W_\infty(\nu^1, \nu^2) = \delta}} \{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\}.$$

$$\tag{2.7}$$

For every $n \geq 1$ we get

$$\mathbb{P}(W_\infty(L_n^X, L_n^Y) = \delta) \tag{2.8}$$

$$= \sum_{\substack{\nu^{1,n},\nu^{2,n} \in M^{1,n}(E) \\ W_\infty(\nu^{1,n},\nu^{2,n})=\delta}} \mathbb{P}(L_n^X = \nu^{1,n}, L_n^Y = \nu^{2,n})$$

$$\leq \sum_{\substack{\nu^{1,n},\nu^{2,n} \in M^{1,n}(E) \\ W_\infty(\nu^{1,n},\nu^{2,n})=\delta}} \exp\left(-nH(\nu^{1,n}|\mu^1) - nH(\nu^{2,n}|\mu^2)\right) \tag{2.9}$$

$$\leq (n+1)^{2N} \exp\left(-n \inf_{\substack{\nu^1,\nu^2 \in M^1(E) \\ W_\infty(\nu^1,\nu^2)=\delta}} \left\{H(\nu^1|\mu^1) + H(\nu^2|\mu^2)\right\}\right)$$

see e.g. Lemma 2.1.9 in Dembo and Zeitouni (1998) for the elementary combinatorial estimate (2.9). The announced (2.7) follows.

2.3.2. *The general case.* Let us introduce some more notations. For every $\Pi \in \mathcal{P}$ we consider

$$I_\infty^\Pi(x) = \inf_{\substack{\nu^1,\nu^2 \in M^1(E) \\ W_\infty(\nu^1,\nu^2)=x}} \left\{H(\pi(\nu^1)|\pi(\mu^1)) + H(\pi(\nu^2)|\pi(\mu^2))\right\}$$

and we shall denote by J_∞^Π the lower semi-continuous regularization of I_∞^Π.

Since $[0,1]$ is compact the sequence $(W_\infty(L_n^X, L_n^Y))_{n\geq 1}$ is naturally exponentially tight so it is sufficient in order to establish the LD upper bound to consider events of the form $\{W_\infty(L_n^X, L_n^Y) \in [a,b]\}$, see Lemma 1.2.18 in Dembo and Zeitouni (1998). According to Lemma 2.1 for every $a, b \in [0,1]$, $a \leq b$, and every $\Pi \in \mathcal{P}$ we have for every $n \geq 1$ that

$$\mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [a,b]) \leq \mathbb{P}(W_\infty(\pi(L_n^X), \pi(L_n^Y)) \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]).$$

Hence, due to the computation carried out assuming that E is finite we get for every $\Pi \in \mathcal{P}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in [a,b]) \leq$$

$$\leq - \inf_{\substack{\nu^1,\nu^2 \in M^1(\{s_1,\ldots,s_L\}) \\ W_\infty(\nu^1,\nu^2)\in[a-2\Delta(\Pi),b+2\Delta(\Pi)]}} \left\{H(\nu^1|\pi(\mu^1)) + H(\nu^2|\pi(\mu^2))\right\}$$

$$\leq - \inf_{\substack{\nu^1,\nu^2 \in M^1(E) \\ W_\infty(\nu^1,\nu^2)\in[a-2\Delta(\Pi),b+2\Delta(\Pi)]}} \left\{H(\pi(\nu^1)|\pi(\mu^1)) + H(\pi(\nu^2)|\pi(\mu^2))\right\}$$

$$\leq - \inf_{x\in[a-2\Delta(\Pi),b+2\Delta(\Pi)]} I_\infty^\Pi(x)$$

$$\leq - \inf_{x\in[a-2\Delta(\Pi),b+2\Delta(\Pi)]} J_\infty^\Pi(x).$$

Hence, to conclude the proof of the LD upper bound we are left to prove that

$$\sup_{\Pi\in\mathcal{P}} \inf_{x\in[a-2\Delta(\Pi),b+2\Delta(\Pi)]} J_\infty^\Pi(x) \geq \inf_{x\in[a,b]} J_\infty(x). \tag{2.10}$$

Assume for a while the following

Lemma 2.7. *For every $x \in [0,1]$ we have $\sup_\Pi J_\infty^\Pi(x) = J_\infty(x)$.*

Assume also that (2.10) does not hold : There exists a $\delta > 0$ such that for every $\Pi \in \mathcal{P}$

$$\inf_{x \in [a-2\Delta(\Pi), b+2\Delta(\Pi)]} J_\infty^\Pi(x) < \inf_{x \in [a,b]} J_\infty(x) - \delta.$$

Since for every $\Pi \in \mathcal{P}$ the map $x \mapsto J_\infty^\Pi(x)$ is lower semi-continuous there exists a net $(x_\Pi)_{\Pi \in \mathcal{P}}$ such that $x_\Pi \in [a - 2\Delta(\Pi), b + 2\Delta(\Pi)]$ and

$$J_\infty^\Pi(x_\Pi) - \delta/2 < \inf_{x \in [a,b]} J_\infty(x) - \delta.$$

Due to the log sum inequality, see e.g. Theorem 2.7.1 in Cover and Thomas (1991), for every $K, K' \in \mathcal{P}$ such that $K \preceq K'$ and every $x \in [0,1]$ we have $I_\infty^K(x) \leq I_\infty^{K'}(x)$ hence $J_\infty^K(x) \leq J_\infty^{K'}(x)$ so for every $\Pi' \in \mathcal{P}$ such that $\Pi' \preceq \Pi$ we have

$$J_\infty^{\Pi'}(x_\Pi) - \delta/2 < \inf_{x \in [a,b]} J_\infty(x) - \delta.$$

Since $(x_\Pi)_{\Pi \in \mathcal{P}}$ is a $[0,1]$-valued net it admits a converging subnet which limit we denote $x^* \in [a,b]$. Due to the fact that $J_\infty^{\Pi'}$ is lower semi-continuous we get

$$\liminf_{\Pi \in \mathcal{P}} J_\infty^{\Pi'}(x_\Pi) \geq J_\infty^{\Pi'}(x^*)$$

hence

$$J_\infty^{\Pi'}(x^*) < \inf_{x \in [a,b]} J_\infty(x) - \delta/2$$

which, according to Lemma 2.7, implies $J_\infty(x^*) < \inf_{x \in [a,b]} J_\infty(x) - \delta/2$. Since the latter can not be (2.10) holds and the LD upper bound follows.

Proof of Lemma 2.7 According to Theorem 1.4.3 in Dupuis and Ellis (1997) for every $x \in [0,1]$ and every $\Pi \in \mathcal{P}$ $I_\infty^\Pi(x) \leq I_\infty(x)$ hence $J_\infty^\Pi(x) \leq J_\infty(x)$ whence $\sup_\Pi J_\infty^\Pi(x) \leq J_\infty(x)$. Conversely let x be a fixed element of $[0,1]$. We shall assume that $\sup_\Pi J_\infty^\Pi(x) < \infty$ for otherwise the claimed equality trivially holds. Hence for every $\Pi \in \mathcal{P}$ and every $\delta > 0$ there exists $y_{\delta,\Pi}$ such that $|x - y_{\delta,\Pi}| < \delta$, $\lim_{\delta \to 0} I_\infty^\Pi(y_{\delta,\Pi}) = J_\infty^\Pi(x)$ and

$$I_\infty^\Pi(y_{\delta,\Pi}) - \delta/2 < \inf_{y \in B(x,\delta)} I_\infty^\Pi(y). \tag{2.11}$$

In particular we can assume that $I_\infty^\Pi(y_{\delta,\Pi}) < \infty$ thus there exists $\nu_1^{\Pi,\delta}, \nu_2^{\Pi,\delta} \in M^1(E)$ such that $W_\infty(\nu_1^{\Pi,\delta}, \nu_2^{\Pi,\delta}) = y_{\delta,\Pi}$ and

$$H(\pi(\nu_1^{\Pi,\delta})|\pi(\mu^1)) + H(\pi(\nu_2^{\Pi,\delta})|\pi(\mu^2)) - \delta \leq I_\infty^\Pi(y_{\delta,\Pi}) - \delta/2. \tag{2.12}$$

Next we define $\rho_1^{\Pi,\delta}, \rho_2^{\Pi,\delta} \in M^1(E)$ by

$$\rho_1^{\Pi,\delta}(F) = \sum_{A_i \in \Pi} \frac{\pi(\nu_1^{\Pi,\delta})(A_i)}{\mu^1(A_i)} \mu^1(F \cap A_i)$$

and accordingly

$$\rho_2^{\Pi,\delta}(F) = \sum_{A_i \in \Pi} \frac{\pi(\nu_2^{\Pi,\delta})(A_i)}{\mu^2(A_i)} \mu^2(F \cap A_i)$$

for every borelian $F \subset E$. Clearly $\pi(\nu_1^{\Pi,\delta}) = \pi(\rho_1^{\Pi,\delta})$ and $\pi(\nu_2^{\Pi,\delta}) = \pi(\rho_2^{\Pi,\delta})$ hence

$$
\begin{aligned}
|W_\infty(\rho_1^{\Pi,\delta}, \rho_2^{\Pi,\delta}) &- W_\infty(\nu_1^{\Pi,\delta}, \nu_2^{\Pi,\delta})| \\
&\leq \quad |W_\infty(\rho_1^{\Pi,\delta}, \rho_2^{\Pi,\delta}) - W_\infty(\pi(\rho_1^{\Pi,\delta}), \pi(\rho_2^{\Pi,\delta}))| + \\
&\quad + |W_\infty(\pi(\rho_1^{\Pi,\delta}), \pi(\rho_2^{\Pi,\delta})) - W_\infty(\pi(\nu_1^{\Pi,\delta}), \pi(\nu_2^{\Pi,\delta}))| + \\
&\quad + |W_\infty(\pi(\nu_1^{\Pi,\delta}), \pi(\nu_2^{\Pi,\delta})) - W_\infty(\nu_1^{\Pi,\delta}, \nu_2^{\Pi,\delta})| \\
&\leq \quad 4\Delta(\Pi).
\end{aligned}
$$

According to Theorem 1.4.3 in Dupuis and Ellis (1997) and Theorem 2.7.1 in Cover and Thomas (1991) we have

$$
\begin{aligned}
H(\rho_1^{\Pi,\delta}|\mu^1) &= \sup_{K \in \mathcal{P}} \sum_{B_j \in K} \rho_1^{\Pi,\delta}(B_j) \log \frac{\rho_1^{\Pi,\delta}(B_j)}{\mu^1(B_j)} \\
&= \sup_{\substack{K \in \mathcal{P} \\ K \succeq \Pi}} \sum_{A_i \in \Pi} \sum_{\substack{B_j \in K \\ B_j \subset A_i}} \rho_1^{\Pi,\delta}(B_j) \log \frac{\rho_1^{\Pi,\delta}(B_j)}{\mu^1(B_j)} \\
&= \sup_{\substack{K \in \mathcal{P} \\ K \succeq \Pi}} \sum_{A_i \in \Pi} \sum_{\substack{B_j \in K \\ B_j \subset A_i}} \frac{\pi(\nu_1^{\Pi,\delta})(A_i)}{\mu^1(A_i)} \mu^1(B_j) \log \frac{\pi(\nu_1^{\Pi,\delta})(A_i)}{\mu^1(A_i)} \\
&= \sup_{\substack{K \in \mathcal{P} \\ K \succeq \Pi}} \sum_{A_i \in \Pi} \pi(\nu_1^{\Pi,\delta})(A_i) \log \frac{\pi(\nu_1^{\Pi,\delta})(A_i)}{\mu^1(A_i)} \\
&= H(\pi(\nu_1^{\Pi,\delta})|\pi(\mu^1))
\end{aligned}
$$

and $H(\rho_2^{\Pi,\delta}|\mu^2) = H(\pi(\nu_2^{\Pi,\delta})|\pi(\mu^2))$ as well, hence

$$
H(\rho_1^{\Pi,\delta}|\mu^1) + H(\rho_2^{\Pi,\delta}|\mu^2) = H(\pi(\nu_1^{\Pi,\delta})|\pi(\mu^1)) + H(\pi(\nu_2^{\Pi,\delta})|\pi(\mu^2)).
$$

Thus it follows from the preceding, (2.11) and (2.12) that for every $\delta > 0$ and every $\Pi \in \mathcal{P}$ we have

$$
\inf_{u \in B(y_{\delta,\Pi}, 4\Delta(\Pi))} I_\infty(u) - \delta \leq \inf_{y \in B(x,\delta)} I_\infty^\Pi(y)
$$

hence

$$
\inf_{y \in B(x,\delta+4\Delta(\Pi))} I_\infty(y) - \delta \leq \inf_{y \in B(x,\delta)} I_\infty^\Pi(y)
$$

whence

$$
\begin{aligned}
\inf_{y \in B(x,\delta+4\Delta(\Pi))} I_\infty(y) - \delta &\leq \sup_{\Pi \in \mathcal{P}} \sup_{\delta > 0} \inf_{y \in B(x,\delta)} I_\infty^\Pi(y) \\
&= \sup_\Pi J_\infty^\Pi(x).
\end{aligned}
$$

But for every $\delta > 0$ we have

$$
\sup_{\Pi \in \mathcal{P}} \inf_{y \in B(x,\delta+4\Delta(\Pi))} I_\infty(y) = \sup_{\substack{\Pi \in \mathcal{P} \\ \Delta(\Pi) < \delta}} \inf_{y \in B(x,\delta+4\Delta(\Pi))} I_\infty(y)
$$

hence

$$
\inf_{y \in B(x,6\delta)} I_\infty(y) \leq \sup_{\Pi \in \mathcal{P}} J_\infty^\Pi(x)
$$

whence $\sup_\Pi J_\infty^\Pi(x) \geq J_\infty(x)$. $\square$

3. On some properties of I_∞ and J_∞.

3.1. I_∞ *can fail to be lower semi-continuous.* All through Section 3.1 we consider that $E = [0, 1]$ is endowed with the usual Euclidean distance and that $\mu^1 = \mu^2 = \mu$ admits $f(x) = \frac{3}{2}\mathbb{1}_{[0, \frac{1}{3}]}(x) + \frac{3}{2}\mathbb{1}_{[\frac{2}{3}, 1]}(x)$ as a density w.r.t. the Lebesgue measure. That in this setting I_∞ is not lower semi-continuous is an immediate consequence of the following two lemmas.

Lemma 3.1. *Let ν^1 and ν^2 be two elements of $M^1(E)$ such that $W_\infty(\nu^1, \nu^2) = \frac{1}{3}$, $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$ and $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$. Necessarily at least one of ν^1 or ν^2 has an atom.*

Lemma 3.2. *The sequence $(\rho^n)_{n \geq 4}$ of elements of $M^1(E)$ defined by the densities*

$$g^n(x) = \frac{3}{2}\mathbb{1}_{[0, \frac{1}{3} - \frac{1}{n}]}(x) + \left(\frac{3}{2} - \frac{1}{n}\right)\mathbb{1}_{[\frac{1}{3} - \frac{1}{n}, \frac{1}{3}]}(x) +$$

$$+ \left(\frac{3}{2} + \frac{1}{n}\right)\mathbb{1}_{[\frac{2}{3}, \frac{2}{3} + \frac{1}{n}]}(x) + \frac{3}{2}\mathbb{1}_{[\frac{2}{3} + \frac{1}{n}, 1]}(x)$$

is such that for every $n \geq 4$ we have

$$\frac{1}{3} < W_\infty(\rho^n, \mu) \leq \frac{1}{3} + \frac{2}{n}$$

and

$$H(\rho^n|\mu) = \frac{3}{2n}\left(1 - \frac{2}{3n}\right)\log\left(1 - \frac{2}{3n}\right) + \frac{3}{2n}\left(1 + \frac{2}{3n}\right)\log\left(1 + \frac{2}{3n}\right).$$

Indeed, let $\nu^1, \nu^2 \in M^1(E)$ be such that $W_\infty(\nu^1, \nu^2) = \frac{1}{3}$. We have two possibilities :

i) $\mathcal{S}(\nu^1)$ or $\mathcal{S}(\nu^2)$ is not included in $\mathcal{S}(\mu)$ which implies that $H(\nu^1|\mu) + H(\nu^2|\mu) = \infty$;

ii) both $\mathcal{S}(\nu^1)$ and $\mathcal{S}(\nu^2)$ are included in $\mathcal{S}(\mu)$ which, according to Lemma 3.1, implies that at least one of ν^1 or ν^2 has an atom, hence $H(\nu^1|\mu) + H(\nu^2|\mu) = \infty$.

Thus $I_\infty(\frac{1}{3}) = \infty$. Meanwhile, it follows from Lemma 3.2 that the sequence $(x_n)_{n \geq 4}$ defined by $x_n = W_\infty(\rho^n, \mu)$ is such that $x_n \to \frac{1}{3}$ and $0 \leq I_\infty(x_n) \leq H(\rho^n|\mu) + H(\mu|\mu) \to 0$ as $n \to \infty$, whence I_∞ is not lower semi-continuous. From the preceding we deduce that $J_\infty(\frac{1}{3}) = 0$.

3.1.1. *Proof of Lemma 3.1.* Lemma 3.1 is an immediate consequence of the following three lemmas

Lemma 3.3. *Let ν^1 and ν^2 be two elements of $M^1(E)$ such that $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$ and $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$. If $W_\infty(\nu^1, \nu^2) \leq \frac{1}{3}$ then necessarily at least one of these two statements is true*

i) $\nu^1([0, \frac{1}{3}]) = \nu^2([0, \frac{1}{3}])$;

ii) *at least one of ν^1 or ν^2 has an atom.*

Lemma 3.4. *Let ν^1 and ν^2 be two elements of $M^1(E)$ such that $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$, $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$, $W_\infty(\nu^1, \nu^2) \leq \frac{1}{3}$ and $\nu^1([0, \frac{1}{3}]) = \nu^2([0, \frac{1}{3}]) = \alpha$ with $\alpha \in]0, 1[$. Then at least one of these two statements is true*

i) *we have*

$$W_\infty(\nu^1, \nu^2) = \max\left\{ W_\infty(\gamma^1, \gamma^2), W_\infty(\theta^1, \theta^2) \right\}$$

where $\gamma^1, \gamma^2 \in M^1([0, \frac{1}{3}])$ *and* $\theta^1, \theta^2 \in M^1([\frac{2}{3}, 1])$ *are defined by*

$$\gamma^1(A) = \frac{\nu^1(A)}{\nu^1([0, \frac{1}{3}])}, \quad \theta^1(B) = \frac{\nu^1(B)}{\nu^1([\frac{2}{3}, 1])}$$

for every measurable $A \subset [0, \frac{1}{3}]$ *and* $B \subset [\frac{2}{3}, 1]$, γ^2 *and* θ^2 *being defined on the ground of* ν^2 *accordingly;*

ii) *at least one of* ν^1 *or* ν^2 *has an atom.*

Lemma 3.5. *Let* ζ^1 *and* ζ^2 *be two elements of* $M^1([0, \frac{1}{3}])$ *(resp.* $M^1([\frac{2}{3}, 1])$) *such that* $W_\infty(\zeta^1, \zeta^2) = \frac{1}{3}$. *Then at least one of* ζ^1 *or* ζ^2 *has an atom.*

Proof of Lemma 3.3. Let ν^1 and ν^2 be two elements of $M^1(E)$ such that $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$ and $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$. We shall prove the following, which is equivalent to the announced statement : If $\nu^1([0, \frac{1}{3}]) \neq \nu^2([0, \frac{1}{3}])$ and neither ν^1 nor ν^2 have an atom then $W_\infty(\nu^1, \nu^2) > \frac{1}{3}$. To this end assume that $\nu^1([0, \frac{1}{3}]) > \nu^2([0, \frac{1}{3}])$. Let $Q \in \mathcal{C}(\nu^1, \nu^2)$ be such that $W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(Q \circ d^{-1})$. We have

$$\begin{aligned}
\nu^1([0, \frac{1}{3}]) &= Q([0, \frac{1}{3}] \times E) \\
&= Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) + Q([0, \frac{1}{3}] \times]\frac{1}{3}, \frac{2}{3}[) + Q([0, \frac{1}{3}] \times [\frac{2}{3}, 1]) \\
&= Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) + Q([0, \frac{1}{3}] \times [\frac{2}{3}, 1])
\end{aligned}$$

since $Q([0, \frac{1}{3}] \times]\frac{1}{3}, \frac{2}{3}[) \leq \nu^2(]\frac{1}{3}, \frac{2}{3}[) = 0$ because $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu) = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. But $Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) \leq Q(E \times [0, \frac{1}{3}]) = \nu^2([0, \frac{1}{3}])$ hence

$$\begin{aligned}
Q([0, \frac{1}{3}] \times [\frac{2}{3}, 1]) &= \nu^1([0, \frac{1}{3}]) - Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) \\
&\geq \nu^1([0, \frac{1}{3}]) - \nu^2([0, \frac{1}{3}]) \\
&> 0.
\end{aligned}$$

We have $\nu^2(\{\frac{2}{3}\}) = 0$ since ν^2 has no atom, hence $Q([0, \frac{1}{3}] \times \{\frac{2}{3}\}) \leq Q(E \times \{\frac{2}{3}\}) = 0$ whence $Q([0, \frac{1}{3}] \times]\frac{2}{3}, 1]) > 0$. Since $[0, \frac{1}{3}] \times]\frac{2}{3}, 1] = \bigcup_{m>1} [0, \frac{1}{3}] \times [\frac{2}{3} + \frac{1}{m}, 1]$ there exists an $m_0 > 0$ such that $Q([0, \frac{1}{3}] \times [\frac{2}{3} + \frac{1}{m_0}, 1]) > 0$ hence $W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(Q \circ d^{-1}) \geq \frac{1}{3} + \frac{1}{m_0} > \frac{1}{3}$. $\qquad\square$

Proof of Lemma 3.4. Let ν^1 and ν^2 be two elements of $M^1(E)$ such that $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$, $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$, $W_\infty(\nu^1, \nu^2) \leq 1$ and $\nu^1([0, \frac{1}{3}]) = \nu^2([0, \frac{1}{3}]) = \alpha$ with $\alpha \in]0, 1[$. We shall prove that if ii) is false then i) is necessarily true. According to Lemma 1.3 there exists a $Q \in \mathcal{C}_\infty(\nu^1, \nu^2)$ such that $W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(Q \circ d^{-1})$. Since $\mathcal{S}(\nu^1) \subset \mathcal{S}(\mu)$ and $\mathcal{S}(\nu^2) \subset \mathcal{S}(\mu)$ we have $Q(E \times]\frac{1}{3}, \frac{2}{3}[) = Q(]\frac{1}{3}, \frac{2}{3}[\times E) = 0$. Since $\sup \mathcal{S}(Q \circ d^{-1}) \leq \frac{1}{3}$ we further have $Q([0, \frac{1}{3}] \times [\frac{2}{3}, 1] \setminus \{(\frac{1}{3}, \frac{2}{3})\}) = Q([\frac{2}{3}, 1] \times [0, \frac{1}{3}] \setminus \{(\frac{2}{3}, \frac{1}{3})\}) = 0$. But $Q(\{(\frac{1}{3}, \frac{2}{3})\}) + Q(\{(\frac{2}{3}, \frac{1}{3})\}) = 0$ for otherwise either ν^1 or ν^2 would have at least an atom. Hence $Q(E \times E) = 1 = Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) + Q([\frac{2}{3}, 1] \times [\frac{2}{3}, 1])$ whence $Q([0, \frac{1}{3}] \times [0, \frac{1}{3}]) = \nu^1([0, \frac{1}{3}]) = \nu^2([0, \frac{1}{3}]) = \alpha > 0$ and $Q([\frac{2}{3}, 1] \times [\frac{2}{3}, 1]) = \nu^1([\frac{2}{3}, 1]) = \nu^2([\frac{2}{3}, 1]) = 1 - \alpha > 0$. Since $\mathcal{S}(Q)$ is infinitely cyclically monotone ,

the probability measures on E^2 defined by

$$Q^{sw}(A) = \frac{Q(A \cap [0,\frac{1}{3}] \times [0,\frac{1}{3}])}{Q([0,\frac{1}{3}] \times [0,\frac{1}{3}])}, \quad Q^{ne}(B) = \frac{Q(B \cap [\frac{2}{3},1] \times [\frac{2}{3},1])}{Q([\frac{2}{3},1] \times [\frac{2}{3},1])}$$

for every measurable $A, B \subset E^2$ are elements of $\mathcal{C}_\infty(\gamma^1, \gamma^2)$ and $\mathcal{C}_\infty(\theta^1, \theta^2)$ respectively and $\mathcal{S}(Q) = \mathcal{S}(Q^{sw}) \cup \mathcal{S}(Q^{ne})$. Indeed, for every measurable $A \subset [0,\frac{1}{3}]$ we have

$$
\begin{aligned}
Q_1^{sw}(A) &= Q^{sw}(A \times E) \\
&= \frac{Q(A \times E \cap [0,\frac{1}{3}] \times [0,\frac{1}{3}])}{Q([0,\frac{1}{3}] \times [0,\frac{1}{3}])} \qquad (3.1) \\
&= \frac{Q(A \times E)}{\nu^1([0,\frac{1}{3}])} \qquad (3.2) \\
&= \frac{\nu^1(A)}{\nu^1([0,\frac{1}{3}])}
\end{aligned}
$$

where (3.2) follows from (3.1) due to the fact that $Q([0,\frac{1}{3}] \times]\frac{1}{3},1]) = 0$. One proves that $Q_2^{sw} = \gamma^2, Q_1^{ne} = \theta^1$ and $Q_2^{ne} = \theta^2$ the same way. Thus we have $W_\infty(\gamma^1, \gamma^2) = \sup \mathcal{S}(Q^{sw} \circ d^{-1})$ and $W_\infty(\theta^1, \theta^2) = \sup \mathcal{S}(Q^{ne} \circ d^{-1})$ hence

$$
\begin{aligned}
W_\infty(\nu^1, \nu^2) &= \sup \mathcal{S}(Q \circ d^{-1}) \\
&= \max \left\{ \sup \mathcal{S}(Q^{sw} \circ d^{-1}), \sup \mathcal{S}(Q^{ne} \circ d^{-1}) \right\} \\
&= \max \left\{ W_\infty(\gamma^1, \gamma^2), W_\infty(\theta^1, \theta^2) \right\}.
\end{aligned}
$$

$\square$

Proof of Lemma 3.5. Let ζ^1 and ζ^2 be two elements of $M^1([0,\frac{1}{3}])$ such that $W_\infty(\zeta^1, \zeta^2) = \frac{1}{3}$. (The proof when ζ^1 and ζ^2 are elements of $M^1([\frac{2}{3},1])$ is obviously the same.) There exists a $Q \in \mathcal{C}_\infty(\zeta^1, \zeta^2)$ such that $\sup \mathcal{S}(Q \circ d^{-1}) = \frac{1}{3}$ and, to fix notations, we assume that $(0,\frac{1}{3}) \in \mathcal{S}(Q)$. We shall prove that if ζ^2 has no atoms on $[0,\frac{1}{3}]$ then 0 is an atom for ζ^1. To this end it is sufficient to prove that for every $u, v \in]0,\frac{1}{3}[$ we have $\zeta^1([0,u]) \geq \zeta^2([0,v])$ since this leads to $\zeta^1([0,u]) = 1$ for every $u \in]0,\frac{1}{3}[$ hence $\zeta^1(\{0\}) = 1$. So, let us assume there exists $u^*, v^* \in]0,\frac{1}{3}[$ such that $\zeta^1([0,u^*]) < \zeta^2([0,v^*])$. Since $(0,\frac{1}{3}) \in \mathcal{S}(Q)$ we have $0 < \zeta^1([0,u^*]) < \zeta^2([0,v^*])$. Since ζ^2 has no atoms on $[0,\frac{1}{3}]$ there exists a $v^{**} \in]0,1[$ such that $\zeta^1([0,u^*]) = \zeta^2([0,v^{**}])$. We can assume that $\zeta^1([0,u^*]) < 1$ for otherwise we would have $\zeta^1([0,u^*]) = \zeta^2([0,v^{**}]) = 1$ hence $W_\infty(\zeta^1, \zeta^2) = \max\{u^*, v^{**}\} < \frac{1}{3}$. We define four probability measures on $[0,\frac{1}{3}]$ by setting for every measurable A

$$\zeta^{1,l}(A) = \frac{\zeta^1(A \cap [0,u^*])}{\zeta^1([0,u^*])}, \quad \zeta^{2,l}(A) = \frac{\zeta^2(A \cap [0,v^{**}])}{\zeta^2([0,v^{**}])}$$

and

$$\zeta^{1,r}(A) = \frac{\zeta^1(A \cap]u^*, \frac{1}{3}])}{\zeta^1(]u^*, \frac{1}{3}])}, \quad \zeta^{2,r}(A) = \frac{\zeta^2(A \cap]v^{**}, \frac{1}{3}])}{\zeta^2(]v^{**}, \frac{1}{3}])}.$$

There exist two infinitely cyclically monotone $Q^l \in \mathcal{C}(\zeta^{1,l}, \zeta^{2,l})$ and $Q^r \in \mathcal{C}(\zeta^{1,r}, \zeta^{2,r})$ and we observe that the probability measure K defined on E^2 by

$$
\begin{aligned}
K &= \zeta^1([0, u^*])Q^l + \zeta^1(]u^*, \tfrac{1}{3}])Q^r \\
&= \zeta^2([0, v^{**}])Q^l + \zeta^2(]v^{**}, \tfrac{1}{3}])Q^r
\end{aligned}
$$

is such that $K \in \mathcal{C}(\zeta^1, \zeta^2)$. Moreover

$$
\begin{aligned}
W_\infty(\zeta^1, \zeta^2) &\leq \sup \mathcal{S}(K \circ d^{-1}) \\
&= \max\{W_\infty(\zeta^{1,l}, \zeta^{2,l}), W_\infty(\zeta^{1,r}, \zeta^{2,r})\} \\
&\leq \max\{\max(u^*, v^{**}), \max(1 - u^*, 1 - v^{**})\} \\
&< \frac{1}{3}
\end{aligned}
$$

which contradicts $W_\infty(\zeta^1, \zeta^2) = \frac{1}{3}$ and concludes the proof.

3.1.2. *Proof of Lemma 3.2.* First of all notice that for every $n \geq 3$ we have $\rho^n([0, \tfrac{1}{3}]) \neq \mu([0, \tfrac{1}{3}])$ hence $W_\infty(\rho^n, \mu) > \frac{1}{3}$ according to Lemma 3.4. Next, it is not difficult to check that for every $n \geq 4$ the probability measure on E^2 denoted Q^n which density w.r.t. the Lebesgue measure on E^2 is given by

$$
h^n(x) = \frac{3}{2\left(\frac{1}{3} - \frac{1}{n}\right)} \mathbb{1}_{[0, \frac{1}{3} - \frac{1}{n}]^2}(x) + n\left(\frac{3}{2} - \frac{1}{n}\right) \mathbb{1}_{[\frac{1}{3} - \frac{1}{n}, \frac{1}{3}]^2}(x) +
$$

$$
+ \mathbb{1}_{[\frac{1}{3} - \frac{1}{n}, \frac{1}{3}] \times [\frac{2}{3}, \frac{2}{3} + \frac{1}{n}]}(x) + \frac{9}{2}\mathbb{1}_{[\frac{2}{3}, 1]^2}(x)
$$

is such that $Q_1^n = \mu$ and $Q_2^n = \rho^n$ hence $W_\infty(\mu, \rho^n) \leq \sup \mathcal{S}(Q^n \circ d^{-1}) \leq \frac{1}{3} + \frac{2}{n}$. Finally, that $H(\rho^n|\mu)$ is as given is the result of a straightforward computation.

3.2. *On the zeros of J_∞.* First we prove Lemma 1.6. In Section 3.2.2 we establish that if $J_\infty(x) = 0$ then $x \in \mathcal{Z}_{\mu^1, \mu^2}$ while the converse is proven in Section 3.2.3. Finally in Section 3.2.4 we prove Proposition 1.9

3.2.1. *Proof of Lemma 1.6.* Let $x \in [0, 1]$ be such that $J_\infty(x) = 0$: There exists a sequence $(x^n)_{n \geq 1}$ of elements of $[0, 1]$ such that $x^n \to x$ and $\lim_{n \to \infty} I_\infty(x^n) = 0$. Thus, by taking subsequences if needed, we can say that there exists a sequence $(P^n)_{n \geq 1}$ of infinitely cyclically monotone elements of $M^1(E \times E)$ such that for every $n \geq 1$ we have $x^n = \sup \mathcal{S}(P^n \circ d^{-1})$, $H(P_1^n|\mu^1) + H(P_2^n|\mu^2) \to 0$ as $n \to \infty$ and $(P^n)_{n \geq 1}$ is convergent w.r.t. the weak convergence topology on $M^1(E^2)$ since E^2 is compact. Lets denote P the limit of $(P^n)_{n \geq 1}$. We necessarily have that $P_1^n \overset{w}{\to} P_1 = \mu^1$ and $P_2^n \overset{w}{\to} P_2 = \mu^2$ since

$$
0 \leq H(P_1|\mu^1) + H(P_2|\mu^2) \leq \liminf \left(H(P_1^n|\mu^1) + H(P_2^n|\mu^2)\right) = 0
$$

for $\nu \mapsto H(\nu|\mu)$ is lower semi-continuous w.r.t. the weak convergence topology, see e.g. Theorem 1.4.3 in Dupuis and Ellis (1997). According to Lemma 2.4 $P \in \mathcal{C}_\infty(\mu^1, \mu^2)$ whence $\sup \mathcal{S}(P \circ d^{-1}) = W_\infty(\mu^1, \mu^2)$. Finally, since $P^n \overset{w}{\to} P$ Lemma 2.2 implies $x = \limsup_{n \to \infty} \sup \mathcal{S}(P^n \circ d^{-1}) \geq \sup \mathcal{S}(P \circ d^{-1}) = W_\infty(\mu^1, \mu^2)$. $\quad\square$

3.2.2. *If $J_\infty(x) = 0$ then $x \in \mathcal{Z}_{\mu^1,\mu^2}$.* Let $x \in [0,1]$ be such that $J_\infty(x) = 0$. We know from Lemma 1.6 that $x \geq W_\infty(\mu^1, \mu^2)$. Moreover there exists a sequence $(x^n)_{n\geq 1}$ of elements of $[0,1]$ such that $x^n \to x$ and $\lim_{n\to\infty} I_\infty(x^n) = 0$. Again, by taking subsequences if needed, we can say that there exists a sequence $(P^n)_{n\geq 1}$ of infinitely cyclically monotone elements of $M^1(E \times E)$ such that for every $n \geq 1$ we have $x^n = \sup \mathcal{S}(P^n \circ d^{-1}) = d(a^n, b^n)$, $(a^n, b^n) \in \mathcal{S}(P^n)$

$$H(P_1^n | \mu^1) + H(P_2^n | \mu^2) \to 0 \qquad (3.3)$$

$n \to \infty$ and $(P^n)_{n\geq 1}$ is convergent w.r.t. the weak convergence topology on $M^1(E \times E)$. We can also assume that $(a^n)_{n\geq 1}$ and $(b^n)_{n\geq 1}$ are converging sequences of elements of $\mathcal{S}(\mu^1)$ and $\mathcal{S}(\mu^2)$ respectively. Lets denote P (resp. a, b) the limit of $(P^n)_{n\geq 1}$ (resp. $(a^n)_{n\geq 1}, (b^n)_{n\geq 1}$). We obviously have that $x = d(a, b)$ and we prove that $a \leftrightarrow b$. Again, due to Lemma 2.4, Lemma 2.2 and (3.3) we necessarily have that $P_1^n \xrightarrow{w} P_1 = \mu^1$ and $P_2^n \xrightarrow{w} P_2 = \mu^2$, hence $P \in \mathcal{C}_\infty(\mu^1, \mu^2)$. We have that $a \in \mathcal{S}(\mu^1)$ and $b \in \mathcal{S}(\mu^2)$ since both $\mathcal{S}(\mu^1)$ and $\mathcal{S}(\mu^2)$ are closed so we are left to prove that for every integer $N \geq 2$, every $(\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N) \in \mathcal{S}(P)$ and every $\sigma \in \mathfrak{S}_N$ we have

$$d(a, b) \leq \max_{i=1,\ldots,N} \left\{ d(\alpha_i, \beta_{\sigma(i)}) \right\}$$

where $\alpha_1 = a$ and $\beta_1 = b$. Indeed, it follows from Lemma 2.3 that for every integer i, $2 \leq i \leq N$, and every $n \geq 1$ there exists $(\alpha_i^n, \beta_i^n) \in \mathcal{S}(P^n)$ such that $(\alpha_i^n, \beta_i^n) \to (\alpha_i, \beta_i)$ as $n \to \infty$. Moreover, $(a^n, b^n) \to (a, b)$ and $(a^n, b^n) \in \mathcal{S}(P^n)$ for every $n \geq 1$. Since P^n is infinitely cyclically monotone and $\sup \mathcal{S}(P^n \circ d^{-1}) = d(a^n, b^n)$ we have that for every $\sigma \in \mathfrak{S}_N$ $d(a^n, b^n) \leq \max_{i=1,\ldots,N} \{ d(\alpha_i^n, \beta_{\sigma(i)}^n) \}$ where $\alpha_1^n = a^n$ and $\beta_1^n = b^n$. This proves the announced inequality by taking the limit $n \to \infty$ and $x \in \mathcal{Z}_{\mu^1,\mu^2}$ since $x = d(a, b)$ and $a \leftrightarrow b$.

3.2.3. *If $x \in \mathcal{Z}_{\mu^1,\mu^2}$ then $J_\infty(x) = 0$.* First we prove that $W_\infty(\mu^1, \mu^2) \in \mathcal{Z}_{\mu^1,\mu^2}$. Indeed, according to Lemma 1.3 there exists $\mathbb{P} \in \mathcal{C}_\infty(\mu^1, \mu^2)$ such that $W_\infty(\mu^1, \mu^2) = \sup \mathcal{S}(P \circ d^{-1}) = d(a, b)$ with $(a, b) \in \mathcal{S}(P)$. Clearly $a \leftrightarrow b$ hence $W_\infty(\mu^1, \mu^2) \in \mathcal{Z}_{\mu^1,\mu^2}$. So now let $x \in \mathcal{Z}_{\mu^1,\mu^2}$ be such that $x > W_\infty(\mu^1, \mu^2)$ since we already know that $J_\infty(W_\infty(\mu^1, \mu^2)) = 0$. Since $x \in \mathcal{Z}_{\mu^1,\mu^2}$ there exists $(a, b) \in \mathcal{S}(\mu^1) \times \mathcal{S}(\mu^2)$ such that $x = d(a, b)$ and $a \leftrightarrow b$ which means that there exists $Q \in \mathcal{C}_\infty(\mu^1, \mu^2)$ such that for every integer $N \geq 2$, every $(\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N) \in \mathcal{S}(Q)$ and every $\sigma \in \mathfrak{S}_N$ we have $d(a, b) \leq \max_{i=1,\ldots,N} \left\{ d(\alpha_i, \beta_{\sigma(i)}) \right\}$ where $(\alpha_1, \beta_1) = (a, b)$. Since $d(a, b) = x > W_\infty(\mu^1, \mu^2) = \sup \mathcal{S}(Q \circ d^{-1}) \geq 0$ there exists an ε_0 small enough to ensure that for every $0 < \varepsilon < \varepsilon_0$ we can find an open neighbourhood $U_\varepsilon \times V_\varepsilon$ of (a, b) such that

(1) $U_\varepsilon \cap V_\varepsilon = \varnothing$;
(2) $U_\varepsilon \times V_\varepsilon \cap \mathcal{S}(Q) = \varnothing$;
(3) $\mu^1(U_\varepsilon) > 0$ and $\mu^2(V_\varepsilon) > 0$;
(4) $\sup_{x,y \in U_\varepsilon} d(x, y) \leq \varepsilon/2$ and $\sup_{x,y \in V_\varepsilon} d(x, y) \leq \varepsilon/2$.

Starting from those $(U_\varepsilon)_{0<\varepsilon<\varepsilon_0}$ and $(V_\varepsilon)_{0<\varepsilon<\varepsilon_0}$, for every $0 < \varepsilon < \varepsilon_0$ one can choose a finite partition $\Pi^\varepsilon = (B_1^\varepsilon, \ldots, B_{L(\varepsilon)}^\varepsilon)$ of E into non-empty measurable sets such that

(1) for every $1 \leq i \leq L(\varepsilon)$ we have $\sup_{x,y \in B_i^\varepsilon} d(x, y) \leq \varepsilon/2$;
(2) there exists two distinct integers $i(a)$ and $i(b)$ such that $1 \leq i(a), i(b) \leq L(\varepsilon), B_{i(a)}^\varepsilon = U_\varepsilon$ and $B_{i(b)}^\varepsilon = V_\varepsilon$.

To every Π^ε we associate $(s_1^\varepsilon, \ldots, s_{L(\varepsilon)}^\varepsilon) \in E^{L(\varepsilon)}$ such that for every $1 \le i \le L(\varepsilon)$ we have $s_i^\varepsilon \in B_i^\varepsilon$ with $s_{i(a)}^\varepsilon = a, s_{i(b)}^\varepsilon = b$ and we set

$$\pi^\varepsilon: \quad M^1(E) \quad \to \quad M^1(\{s_1^\varepsilon, \ldots, s_{L(\varepsilon)}^\varepsilon\})$$
$$\nu \quad \mapsto \quad \sum_{i=1}^{L(\varepsilon)} \nu(B_i^\varepsilon)\delta_{s_i^\varepsilon}.$$

For every $0 < \varepsilon < \varepsilon_0$ we define two elements $\kappa^{1,\varepsilon}$ and $\kappa^{2,\varepsilon}$ of $M^1(E)$ by

$$\kappa^{1,\varepsilon}(A) = \frac{\mu^1(A \cap B_{i(a)}^\varepsilon)}{\mu^1(B_{i(a)}^\varepsilon)} \quad \text{and} \quad \kappa^{2,\varepsilon}(A) = \frac{\mu^2(A \cap B_{i(b)}^\varepsilon)}{\mu^2(B_{i(b)}^\varepsilon)}.$$

For every $0 < \varepsilon < \varepsilon_0$ and every $\eta \in]0, 1[$ we further introduce

$$Q^{\varepsilon,\eta} = \eta\left(\kappa^{1,\varepsilon} \otimes \kappa^{2,\varepsilon}\right) + (1-\eta)Q$$

which is an element of $M^1(E \times E)$ such that $Q_1^{\varepsilon,\eta} = \eta\kappa^{1,\varepsilon} + (1-\eta)\mu^1$ and $Q_2^{\varepsilon,\eta} = \eta\kappa^{2,\varepsilon} + (1-\eta)\mu^2$. Finally, we also consider

$$C^{\varepsilon,\eta} = \eta\delta_{(a,b)} + (1-\eta) \sum_{(i,j) \in \{1,\ldots,L(\varepsilon)\}^2} Q(B_i^\varepsilon \times B_j^\varepsilon)\delta_{(s_i^\varepsilon, s_j^\varepsilon)}.$$

We have $C_1^{\varepsilon,\eta} = \pi^\varepsilon(Q_1^{\varepsilon,\eta})$ and $C_2^{\varepsilon,\eta} = \pi^\varepsilon(Q_2^{\varepsilon,\eta})$. Since $C^{\varepsilon,\eta}$ need not be infinitely cyclically monotone we have to consider

$$\widetilde{C^{\varepsilon,\eta}} = \eta\delta_{(a,b)} + (1-\eta) \sum_{(i,j) \in \{1,\ldots,L(\varepsilon)\}^2} Q(B_i^\varepsilon \times B_j^\varepsilon)\delta_{(\widetilde{s_1^\varepsilon}(i,j), \widetilde{s_2^\varepsilon}(i,j))}$$

where for every $(i,j) \in \{1,\ldots,L(\varepsilon)\}^2$ such that $Q(B_i^\varepsilon \times B_j^\varepsilon) > 0$ we have

$$(\widetilde{s_1^\varepsilon}(i,j), \widetilde{s_2^\varepsilon}(i,j)) \in B_i^\varepsilon \times B_j^\varepsilon \cap \mathcal{S}(Q)$$

and $(\widetilde{s_1^\varepsilon}(i(a), i(b)), \widetilde{s_2^\varepsilon}(i(a), i(b))) = (a, b)$. That $a \leftrightarrow b$ implies that $\widetilde{C^{\varepsilon,\eta}}$ is infinitely cyclically monotone thus, according to Lemma 1.4, $W_\infty(\widetilde{C_1^{\varepsilon,\eta}}, \widetilde{C_2^{\varepsilon,\eta}}) = d(a, b) = x$. Clearly $W_\infty(C_1^{\varepsilon,\eta}, \widetilde{C_1^{\varepsilon,\eta}}) \le \varepsilon$ and $W_\infty(C_2^{\varepsilon,\eta}, \widetilde{C_2^{\varepsilon,\eta}}) \le \varepsilon$ so $|W_\infty(C_1^{\varepsilon,\eta}, C_2^{\varepsilon,\eta}) - x| \le 2\varepsilon$ which implies that $|W_\infty(Q_1^{\varepsilon,\eta}, Q_2^{\varepsilon,\eta}) - x| \le 4\varepsilon$ according to Lemma 2.1. For every $0 < \varepsilon < \varepsilon_0$ and every $\eta \in]0, 1[$ we have

$$\begin{aligned}
I_\infty(W_\infty(Q_1^{\varepsilon,\eta}, Q_2^{\varepsilon,\eta})) \quad &\le \quad H(Q_1^{\varepsilon,\eta}|\mu^1) + H(Q_2^{\varepsilon,\eta}|\mu^2) \\
&= \quad H(\eta\kappa^{1,\varepsilon} + (1-\eta)\mu^1|\mu^1) + H(\eta\kappa^{2,\varepsilon} + (1-\eta)\mu^2|\mu^2) \\
&\le \quad \eta H(\kappa^{1,\varepsilon}|\mu^1) + (1-\eta)H(\mu^1|\mu^1) + \\
&\qquad \qquad + \eta H(\kappa^{2,\varepsilon}|\mu^2) + (1-\eta)H(\mu^2|\mu^2) \\
&= \quad \eta H(\kappa^{1,\varepsilon}|\mu^1) + \eta H(\kappa^{2,\varepsilon}|\mu^2).
\end{aligned}$$

Since for every $0 < \varepsilon < \varepsilon_0$ we have $H(\kappa^{1,\varepsilon}|\mu^1) < \infty$ and $H(\kappa^{2,\varepsilon}|\mu^2) < \infty$ one can find an $\eta_\varepsilon \in]0, 1[$ such that

$$\lim_{\varepsilon \to 0}\left(\eta_\varepsilon H(\kappa^{1,\varepsilon}|\mu^1) + \eta_\varepsilon H(\kappa^{2,\varepsilon}|\mu^2)\right) = 0.$$

Now, for every $0 < \varepsilon < \varepsilon_0$ we set $x^\varepsilon = W_\infty(Q_1^{\varepsilon,\eta_\varepsilon}, Q_2^{\varepsilon,\eta_\varepsilon})$. By construction we have $\lim_{\varepsilon \to 0} x^\varepsilon = x$ and $\lim_{\varepsilon \to 0} I_\infty(x^\varepsilon) = 0$ hence $J_\infty(x) = 0$.

3.2.4. *Proof of Proposition 1.9.* The key here is that when $\mu^1 = \mu^2 = \mu$ the set $\mathcal{C}_\infty(\mu^1, \mu^2)$ is reduced to the diagonal coupling of μ with itself, which we denote Q^μ.

Let $a, b \in E$ such that $a \leftrightarrow b$. Let N be any integer larger than 3 and $\alpha_1, \ldots, \alpha_N$ be any sequence of elements of $\mathcal{S}(\mu)$ of size N, with $\alpha_1 = a$ and $\alpha_N = b$. We introduce $(\widetilde{\alpha}_i, \widetilde{\beta}_i)_{1 \leq i \leq N+1}$ by setting $(\widetilde{\alpha}_1, \widetilde{\beta}_1) = (a, b)$ and $(\widetilde{\alpha}_i, \widetilde{\beta}_i) = (\alpha_{i-1}, \alpha_{i-1}) \in \mathcal{S}(Q^\mu)$ for every integer $i, 2 \leq i \leq N + 1$. Since $a \leftrightarrow b$, for every $\sigma \in \mathfrak{S}_{N+1}$ we have $d(a, b) \leq \max_{i=1,\ldots,N+1} d(\widetilde{\alpha}_i, \widetilde{\beta}_{\sigma(i)})$. Taking σ such that $\sigma(i) = i + 1$ for every $1 \leq i \leq N$ and $\sigma(N + 1) = 1$ leads to $d(a, b) \leq d(\alpha_L, \alpha_{L+1})$ for some L such that $1 \leq L \leq N - 1$ since $d(\widetilde{\alpha}_1, \widetilde{\beta}_{\sigma(1)}) = d(a, a) = 0$ and $d(\widetilde{\alpha}_{N+1}, \widetilde{\beta}_{\sigma(N+1)}) = d(b, b) = 0$.

Conversely let $a, b \in \mathcal{S}(\mu)$ such that for every integer $N \geq 3$ and every sequence $\alpha_1, \ldots, \alpha_N$ of elements of $\mathcal{S}(\mu)$ such that $\alpha_1 = a$ and $\alpha_N = b$ there is at least one integer L such that $1 \leq L \leq N - 1$ and $d(a, b) \leq d(\alpha_L, \alpha_{L+1})$. To prove that $a \leftrightarrow b$ we prove that for every integer N larger than 2, every family $(\widetilde{\alpha}_2, \widetilde{\beta}_2) = (\widetilde{\alpha}_2, \widetilde{\alpha}_2), \ldots, (\widetilde{\alpha}_N, \widetilde{\beta}_N) = (\widetilde{\alpha}_N, \widetilde{\alpha}_N)$ of elements of $\mathcal{S}(Q^\mu)$ and every $\sigma \in \mathfrak{S}_N$ we have $d(a, b) \leq \max_{i=1,\ldots,N} d(\widetilde{\alpha}_i, \widetilde{\beta}_{\sigma(i)})$ where $(\widetilde{\alpha}_1, \widetilde{\beta}_1) = (a, b)$. If 1 is a fixed point of σ then obviously $d(a, b) \leq \max_{i=1,\ldots,N} \{d(\widetilde{\alpha}_i, \widetilde{\alpha}_{\sigma(i)})\}$. If $\sigma(1) \neq 1$ we denote $C(\sigma, 1)$ the cycle of σ which contains 1

$$C(\sigma, 1) = (1, \sigma(1), \ldots, \sigma^{k-1}(1))$$

where k is the length of $C(\sigma, 1)$. By setting $\alpha_1 = \widetilde{\alpha}_1 = a, \alpha_2 = \widetilde{\alpha}_{\sigma(1)}, \alpha_3 = \widetilde{\alpha}_{\sigma^2(1)}, \ldots, \alpha_k = \widetilde{\alpha}_{\sigma^{k-1}(1)}, \alpha_{k+1} = b$ we clearly obtain a journey from a to b through elements of $\mathcal{S}(\mu)$ divided into k stages hence

$$
\begin{aligned}
d(a, b) \quad &\leq \quad \max_{L=1,\ldots,k} d(\alpha_L, \alpha_{L+1}) \\
&= \quad \max_{i \in C(\sigma,1) \cup \{\sigma^k(1)\}} d(\widetilde{\alpha}_i, \widetilde{\beta}_{\sigma(i)}) \\
&\leq \quad \max_{i=1,\ldots,N} d(\widetilde{\alpha}_i, \widetilde{\beta}_{\sigma(i)})
\end{aligned}
$$

which achieves the proof that $a \leftrightarrow b$.

4. On the almost sure asymptotic behavior of $(W_\infty(L_n^X, L_n^Y))_{n \geq 1}$

In this section we will successively prove Lemma 1.10, 1.11, 1.12 and Proposition 1.13. To this end we need some more notations and results. For every $x, y \in \mathcal{S}(\mu)$ and every integer $n \geq 2$ we set

$$\mathcal{R}_n(x, y) = \{\bar{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathcal{S}(\mu)^n \text{ such that } \alpha_1 = x \text{ and } \alpha_n = y\},$$

$$\mathcal{C}(x, y) = \cup_{n \geq 2} \mathcal{R}_n(x, y) \text{ and } \delta^\mu(x, y) = \inf_{\bar{\alpha} \in \mathcal{C}(x,y)} \max_{i \in \{1,\ldots,n-1\}} d(\alpha_i, \alpha_{i+1}).$$

For every $a, b, c \in \mathcal{S}(\mu)$, every $\bar{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathcal{C}(a, b)$ and every $\bar{\gamma} = (\gamma_1, \ldots, \gamma_m) \in \mathcal{C}(b, c)$ we shall write

$$\bar{\alpha} \vee \bar{\gamma} = (\alpha_1, \ldots, \alpha_n, \gamma_2, \ldots, \gamma_m) \in \mathcal{C}(a, c).$$

We will use δ^μ mainly to define appropriate partitions of $\mathcal{S}(\mu)$. It has the following properties

Lemma 4.1. (1) *For every $a, b \in \mathcal{S}(\mu)$ we have $\delta^\mu(a, b) \leq d(a, b)$.*
(2) *For every $a, b, c \in \mathcal{S}(\mu)$ we have $\delta^\mu(a, b) \leq \max\{\delta^\mu(a, c), \delta^\mu(c, b)\}$.*
(3) *For every $a, b \in \mathcal{S}(\mu)$ we have $a \leftrightarrow b$ if and only if $\delta^\mu(a, b) = d(a, b)$.*

Proof of Lemma 4.1 1. For every $a, b \in \mathcal{S}(\mu)$, by considering $\bar{\alpha} = (a, b)$ we immediately get $\delta^\mu(a, b) \leq d(a, b)$. **2.** For every $a, b, c \in \mathcal{S}(\mu)$ and every $\varepsilon > 0$ there exists $\bar{\alpha}^1 = (\alpha_1^1 = a, \ldots, \alpha_n^1 = c) \in \mathscr{C}(a, c)$ such that

$$\max_{i=1,\ldots,n-1} d(\alpha_i^1, \alpha_{i+1}^1) < \delta^\mu(a, c) + \varepsilon/2$$

and there exists $\bar{\alpha}^2 = (\alpha_1^2 = c, \ldots, \alpha_m^2 = b) \in \mathscr{C}(c, b)$ such that

$$\max_{i=1,\ldots,m-1} d(\alpha_i^2, \alpha_{i+1}^2) < \delta^\mu(c, b) + \varepsilon/2.$$

Since $\bar{\alpha} = \bar{\alpha}^1 \vee \bar{\alpha}^2 = (\alpha_1^1 = a, \ldots, \alpha_n^1 = c, \alpha_2^2, \ldots, \alpha_m^2 = b) \in \mathscr{C}(a, b)$ we have

$$
\begin{aligned}
\delta^\mu(a, b) &\leq \max\{\max_{i=1,\ldots,n-1} d(\alpha_i^1, \alpha_{i+1}^1), \max_{i=1,\ldots,m-1} d(\alpha_i^2, \alpha_{i+1}^2)\} \\
&\leq \max\{\delta^\mu(a, c), \delta^\mu(c, b)\} + \varepsilon
\end{aligned}
$$

which leads to the announced inequality by taking $\varepsilon \to 0$. **3.** Let $a, b \in \mathcal{S}(\mu)$ be such that $a \leftrightarrow b$. By definition of $\leftrightarrow$ for every $n \geq 2$ and every $\bar{\alpha} \in \mathscr{R}_n(a, b)$ we have $d(a, b) \leq \max_{i \in \{1,\ldots,n-1\}} d(\alpha_i, \alpha_{i+1})$ hence $d(a, b) \leq \delta^\mu(a, b)$. It follows that if $a \leftrightarrow b$ we have $\delta^\mu(a, b) = d(a, b)$ since $\delta^\mu(a, b) \leq d(a, b)$ is always true. Conversely assume that $a, b \in \mathcal{S}(\mu)$ are such that we have $\delta^\mu(a, b) = d(a, b)$ thus $d(a, b) \leq \delta^\mu(a, b)$. It follows from the definition of δ^μ that for every integer $n \geq 2$ and every $\bar{\alpha} \in \mathscr{R}_n(a, b)$ we have $d(a, b) \leq \max_{i \in \{1,\ldots,n-1\}} d(\alpha_i, \alpha_{i+1})$ which by definition means $a \leftrightarrow b$. $\square$

Proof of Lemma 1.10 First we assume that $\mathcal{S}(\mu)$ is connected. Then it is well-chained i.e. for every $a, b \in \mathcal{S}(\mu)$ such that $a \neq b$ and every $\varepsilon > 0$ there exist an integer $n \geq 2$ and $\bar{\alpha} \in \mathscr{R}_n(a, b)$ such that for every $1 \leq i \leq n - 1$ we have $d(\alpha_i, \alpha_{i+1}) \leq \varepsilon$, see e.g. 8.2 of Chapter I in Whyburn (1942). Thus $\delta^\mu(a, b) = 0$ for every $a, b \in \mathcal{S}(\mu)$ hence, according to the third point of Lemma 4.1, $a \leftrightarrow b$ if and only if $d(a, b) = 0$ whence $\mathcal{Z}_\mu = \{0\}$. Now we prove that if $\mathcal{Z}_\mu = \{0\}$ then $\mathcal{S}(\mu)$ is connected. It is sufficient to prove that $\mathcal{S}(\mu)$ is well-chained since, according to e.g. 9.21 of Chapter I in Whyburn (1942) any metric, compact, well-chained space is connected. Actually we shall prove that if $\mathcal{Z}_\mu = \{0\}$ then necessarily for every $a, b \in \mathcal{S}(\mu)$ we have $\delta^\mu(a, b) = 0$ since this is obviously equivalent to the well-chained condition. To this end we proceed by contradiction and assume that we have $\mathcal{Z}_\mu = \{0\}$ and that there exists some $a, b \in \mathcal{S}(\mu)$ such that $\delta^\mu(a, b) > 0$. We shall construct $(u, v) \in \mathcal{S}(\mu)^2$ such that $\delta^\mu(u, v) = d(u, v) = \delta^\mu(a, b) > 0$ in contradiction with $\mathcal{Z}_\mu = \{0\}$.

According to the definition of δ^μ for every integer $n \geq 1$ there exists $\bar{\alpha}^n = (\alpha_1^n, \ldots, \alpha_{L(n)}^n) \in \mathscr{C}(a, b)$ such that

$$\delta^\mu(a, b) \leq \max_{i=1,\ldots,L(n)-1} d(\alpha_i^n, \alpha_{i+1}^n) < \delta^\mu(a, b) + \frac{1}{n}.$$

Since $\bar{\alpha}^n \in \mathscr{C}(a, b)$ it follows from the second point in Lemma 4.1 that

$$\delta^\mu(a, b) \leq \max_{i=1,\ldots,L(n)-1} \delta^\mu(\alpha_i^n, \alpha_{i+1}^n). \tag{4.1}$$

Due to the first point in Lemma 4.1 we have

$$\max_{i=1,\ldots,L(n)-1} \delta^\mu(\alpha_i^n, \alpha_{i+1}^n) \leq \max_{i=1,\ldots,L(n)-1} d(\alpha_i^n, \alpha_{i+1}^n) < \delta^\mu(a, b) + \frac{1}{n}. \tag{4.2}$$

For every $n \geq 1$ we consider $G(n) \subset \{1, \ldots, L(n)-1\}$ defined by $j \in G(n)$ if and only if $\delta^\mu(\alpha_j^n, \alpha_{j+1}^n) = \max_{i=1,\ldots,L(n)-1} \delta^\mu(\alpha_i^n, \alpha_{i+1}^n)$. We further define $i_0(n) \in G(n)$ to be such that $d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n) = \max_{i \in G(n)} d(\alpha_i^n, \alpha_{i+1}^n)$. In particular

$$d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n) \leq \max_{i=1,\ldots,L(n)-1} d(\alpha_i^n, \alpha_{i+1}^n) < \delta^\mu(a,b) + \frac{1}{n}. \tag{4.3}$$

According to the definition of $i_0(n)$, every $j \in \{1, \ldots, L(n)-1\}$ such that

$$d(\alpha_j^n, \alpha_{j+1}^n) > d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n)$$

is necessarily such that

$$\delta^\mu(\alpha_j^n, \alpha_{j+1}^n) < \delta^\mu(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n).$$

By definition of δ^μ for every such $j \in \{1, \ldots, L(n)-1\}$ there exists

$$\hat{\alpha}^j = (\hat{\alpha}_1^j, \ldots, \hat{\alpha}_{M(j)}^j) \in \mathscr{C}(\alpha_j^n, \alpha_{j+1}^n)$$

such that

$$\max_{l=1,\ldots,M(j)-1} d(\hat{\alpha}_l^j, \hat{\alpha}_{l+1}^j) \quad < \quad \delta^\mu(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n)$$

$$\leq \quad d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n). \tag{4.4}$$

By replacing every such $(\alpha_j^n, \alpha_{j+1}^n)$ ring in the $\bar{\alpha}^n$ chain by the corresponding $\hat{\alpha}^j$ chain we get $\bar{\gamma} \in \mathscr{C}(a,b)$ such that $\max d(\gamma_i, \gamma_{i+1}) = d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n)$. Since $\bar{\gamma} \in \mathscr{C}(a,b)$, according to the second point of Lemma 4.1 and (4.4) we have

$$\delta^\mu(a,b) \leq d(\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n). \tag{4.5}$$

So we have a sequence $((u^n, v^n) = (\alpha_{i_0(n)}^n, \alpha_{i_0(n)+1}^n))_{n \geq 1}$ of elements of $\mathcal{S}(\mu)^2$ such that for every $n \geq 1$

$$\delta^\mu(a,b) \leq \delta^\mu(u^n, v^n) < \delta^\mu(a,b) + \frac{1}{n}$$

according to (4.1) and (4.2) and

$$\delta^\mu(a,b) < d(u^n, v^n) \leq \delta^\mu(a,b) + \frac{1}{n}.$$

according to (4.3) and (4.5). Since $\mathcal{S}(\mu)^2$ is compact, from this sequence we can extract a sub-sequence $((u^{n_k}, v^{n_k})_{k \geq 1}$ that converges towards some $(u,v) \in \mathcal{S}(\mu)^2$. Obviously $d(u,v) = \delta^\mu(a,b)$ and if we show that $\delta^\mu(u,v) = \delta^\mu(a,b)$ too the proof of Lemma 1.10 is done. It follows from the definition of $(u^{n_k}, v^{n_k})_{k \geq 1}$ that for every $k \geq 1$ we have

$$\begin{aligned} \delta^\mu(a,b) &\leq \delta^\mu(u^{n_k}, v^{n_k}) \\ &\leq \max\{\delta^\mu(u^{n_k}, u), \delta^\mu(u,v), \delta^\mu(v, v^{n_k})\} \\ &\leq \max\{d(u^{n_k}, u), \delta^\mu(u,v), d(v, v^{n_k})\} \end{aligned}$$

hence by taking k large enough we get $\delta^\mu(a,b) \leq \delta^\mu(u,v)$ since $\delta^\mu(a,b) > 0$. On the other hand

$$\begin{aligned} \delta^\mu(u,v) &\leq \max\{\delta^\mu(u^{n_k}, u), \delta^\mu(u^{n_k}, v^{n_k}), \delta^\mu(v, v^{n_k})\} \\ &\leq \left\{d(u^{n_k}, u), \delta^\mu(a,b) + \frac{1}{n_k}, d(v, v^{n_k})\right\} \end{aligned}$$

hence $\delta^\mu(u,v) \leq \delta^\mu(a,b) + \frac{1}{n_k}$ for k large enough hence $\delta^\mu(a,b) = \delta^\mu(u,v)$. $\square$

Proof of Lemma 1.11 Let $a,b \in \mathcal{S}(\mu)$ be such that $a \leftrightarrow b$: According to Lemma 4.1 we have $\delta^\mu(a,b) = d(a,b)$. If $a = b$ then $A_{i(a)} = A_{i(b)}$ and the equality $d(a,b) = d(A_{i(a)}, A_{i(b)})$ is satisfied. If $a \neq b$ necessarily $A_{i(a)} \neq A_{i(b)}$ for if a and b were elements of the same connected component of $\mathcal{S}(\mu)$ we would have $\delta^\mu(a,b) = 0 \neq d(a,b)$ since every connected space is well-chained. By definition we have $d(a,b) \geq d(A_{i(a)}, A_{i(b)})$ and it can not be that $d(a,b) > d(A_{i(a)}, A_{i(b)})$ since this would be in contradiction with $a \leftrightarrow b$. Indeed, since $A_{i(a)}$ and $A_{i(b)}$ are both closed there exists $(x(a), x(b)) \in A_{i(a)} \times A_{i(b)}$ such that $d(x(a), x(b)) = d(A_{i(a)}, A_{i(b)})$ and according to the second point in Lemma 4.1

$$\delta^\mu(a,b) \leq \max\{\delta^\mu(a, x(a)), d(x(a), x(b)), \delta^\mu(b, x(b))\} = d(x(a), x(b)) < d(a,b)$$

since $\delta^\mu(a, x(a)) = \delta^\mu(b, x(b)) = 0$ because $A_{i(a)}$ and $A_{i(b)}$ are connected hence well-chained, and this contradicts $a \leftrightarrow b$. $\square$

Proof of Lemma 1.12 We start showing that $\beta_1 = \sup \mathcal{Z}_\mu$ is not the limit of an increasing sequence $(x_n)_{n \geq 1}$ of elements of $\mathcal{Z}_\mu$. We can assume that $\beta_1 > 0$ since otherwise the claim is clearly true. We proceed by contradiction and assume there exists an increasing sequence $(x_n)_{n \geq 1}$ of elements of $\mathcal{Z}_\mu$ such that $x_n \to \beta_1$. Then there exists a sequence $(a_n, b_n)_{n \geq 1}$ of elements of $\mathcal{S}(\mu)^2$ such that for every $n \geq 1, a_n \leftrightarrow b_n$ and $\delta^\mu(a_n, b_n) = d(a_n, b_n) = x_n$. Since E is compact, by taking sub-sequences if needed, we can assume that there exists $a, b \in \mathcal{S}(\mu)$ such that $a_n \to a, b_n \to b$ and $d(a,b) = \beta_1$. So, we can choose $\varepsilon > 0$ small enough to ensure that there exists an N_0 such that for every $n \geq N_0, \delta^\mu(a_n, a_{n+1}) \leq d(a_n, a_{n+1}) < \varepsilon, \delta^\mu(b_n, b_{n+1}) \leq d(b_n, b_{n+1}) < \varepsilon$ and

$$\varepsilon < \beta_1 - \varepsilon < \delta^\mu(a_{N_0}, b_{N_0}) = d(a_{N_0}, b_{N_0}) < d(a_n, b_n)$$

hence

$$\begin{aligned}
&\delta^\mu(a_n, b_n) \\
\leq\ & \max_{k=0,\ldots,n-N_0-1}\{\delta^\mu(a_{N_0+k}, a_{N_0+k+1}), \delta^\mu(a_{N_0}, b_{N_0}), \delta^\mu(b_{N_0+k}, b_{N_0+k+1})\} \\
=\ & \delta^\mu(a_{N_0}, b_{N_0}) \\
<\ & d(a_n, b_n)
\end{aligned}$$

in contradiction with $a_n \leftrightarrow b_n$. Thus $\beta_1 = \max \mathcal{Z}_\mu$ is an isolated point of $\mathcal{Z}_\mu$. One can repeat this to prove that every $\beta_i = \sup \mathcal{Z}_\mu \setminus \{\beta_1, \ldots, \beta_{i-1}\}$ is an isolated point as long as $\beta_i \neq 0$. $\square$

With Lemma 1.12 in hands we can start the proper proof of Proposition 1.13. We first prove that

$$\mathbb{P}(\limsup W_\infty(L_n^X, L_n^Y) \in \mathcal{Z}_\mu) = \mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) \in \mathcal{Z}_\mu) = 1. \qquad (4.6)$$

Indeed, since J_∞ is a good rate function, it follows from Theorem 1.5 that for every $\varepsilon > 0$ there exists $C(\varepsilon) > 0$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in V_\varepsilon^c) < -C(\varepsilon)$$

where $V_\varepsilon = [0, \varepsilon[\cup (\cup_{\beta \in \mathcal{Z}_\mu}]\beta - \varepsilon, \beta + \varepsilon[)$. Thus

$$\sum_{n \geq 1} \mathbb{P}(W_\infty(L_n^X, L_n^Y) \in V_\varepsilon^c) < \infty$$

hence $\mathbb{P}(\limsup\{W_\infty(L_n^X, L_n^Y) \in V_\varepsilon^c\}) = 0$ according to the Borel-Cantelli Lemma. It follows that for every $\varepsilon > 0$

$$\mathbb{P}(\limsup W_\infty(L_n^X, L_n^Y) \in \bar{V}_\varepsilon) = 1 \text{ and } \mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) \in \bar{V}_\varepsilon) = 1$$

which, because of Lemma 1.12, implies (4.6) by taking $\varepsilon \downarrow 0$. The first point in Proposition 1.13 follows immediately thanks to Lemma 1.10.

From now on we assume that $\mathcal{S}(\mu)$ has at least two connected components. We shall prove in Section 4.1 that we always have

$$\mathbb{P}(\limsup W_\infty(L_n^X, L_n^Y) = \beta_1) = 1. \tag{4.7}$$

Next we will prove in Section 4.2 that if there exists only two connected components A_i, A_j of $\mathcal{S}(\mu)$ such that $A_i \leftrightarrow A_j$ and $d(A_i, A_j) = \beta_1$ then

$$\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) = \beta_2) = 1 \tag{4.8}$$

and in Section 4.3 we will prove that in all other situations where $\mathcal{S}(\mu)$ is not connected

$$\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) = \beta_1) = 1. \tag{4.9}$$

All these results rely on the following "bins and balls" lemma which proof is postponed until the end of the section.

Lemma 4.2. *Let $U = (U_1, \ldots, U_n, \ldots)$ and $V = (V_1, \ldots, V_n, \ldots)$ be two independent sequences of independent random variables with the same law on $\{1, \ldots, L\}$ defined by $\mathbb{P}(U_1 = i) = \gamma_i$, $0 < \gamma_i < 1$ for every $i \in \{1, \ldots, L\}$. The counters*

$$U^n = \left(\sum_{i=1}^n \mathbb{1}_{\{1\}}(U_i), \ldots, \sum_{i=1}^n \mathbb{1}_{\{L\}}(U_i) \right)$$

and

$$V^n = \left(\sum_{i=1}^n \mathbb{1}_{\{1\}}(V_i), \ldots, \sum_{i=1}^n \mathbb{1}_{\{L\}}(V_i) \right)$$

are such that

(1) *If $L = 2$ then $\mathbb{P}(\limsup\{U^n \neq V^n\}) = 1$;*
(2) *If $L = 2$ then $\mathbb{P}(\limsup\{U^n = V^n\}) = 1$;*
(3) *If $L = 3$ then $\mathbb{P}(\limsup\{U^n = V^n\}) = 0$.*

4.1. *Proof of (4.7).* According to Lemma 1.11 and Lemma 1.12 there exists (at least) two $a, b \in \mathcal{S}(\mu)$ such that $a \leftrightarrow b$ and $d(a, b) = d(A_{i(a)}, A_{i(b)}) = \beta_1$. We define a partition of $\mathcal{S}(\mu)$ into two non-empty sets by taking

$$C^1 = \{y \in \mathcal{S}(\mu) \text{ such that } \delta^\mu(a, y) < \beta_1\}$$

and

$$C^2 = \{y \in \mathcal{S}(\mu) \text{ such that } \delta^\mu(a, y) \geq \beta_1\}.$$

Notice that if $x \in C^1$ and $y \in C^2$ then necessarily $d(x, y) \geq \beta_1$. Indeed if we had $d(x, y) < \beta_1$ then, according to Lemma 4.1, we would have $\delta^\mu(a, y) \leq \max\{\delta^\mu(a, x), \delta^\mu(x, y)\} < \beta_1$ in contradiction with $y \in C^2$. It follows that if

$L_n^X(C^1) \neq L_n^Y(C^1)$ then for every $\sigma \in \mathfrak{S}_n$ there exists at least one $1 \leq i \leq n$ such that $X_i \in C^1$ and $Y_{\sigma(i)} \in C^2$ hence $W_\infty(L_n^X, L_n^Y) \geq \beta_1$. Thus

$$\limsup\{L_n^X(C^1) \neq L_n^Y(C^1)\} \subset \{\limsup W_\infty(L_n^X, L_n^Y) \geq \beta_1\}$$

hence

$$\mathbb{P}(\limsup\{L_n^X(C^1) \neq L_n^Y(C^1)\}) \leq \mathbb{P}(\limsup W_\infty(L_n^X, L_n^Y) \geq \beta_1).$$

By defining for every integer $i \geq 1$

$$U_i = \left\{ \begin{array}{ll} 1 & \text{if } X_i \in C^1 \\ 2 & \text{if } X_i \in C^2 \end{array} \right.$$

and V_i accordingly on the ground of Y_i we see that

$$\mathbb{P}(\limsup\{U^n \neq V^n\}) = \mathbb{P}(\limsup\{L_n^X(C^1) \neq L_n^Y(C^1)\})$$

whence $\mathbb{P}(\limsup W_\infty(L_n^X, L_n^Y) \geq \beta_1) = 1$ according to Lemma 4.2, which combined to (4.6) proves (4.7).

4.2. *Proof of (4.8).* Here we assume that there is a unique pair $\{A_i, A_j\}$ of connected components of $\mathcal{S}(\mu)$ such that $A_i \leftrightarrow A_j$ and $d(A_i, A_j) = \beta_1$. Let $a \in A_i$ and $b \in A_j$ be such that $a \leftrightarrow b$ and $d(a, b) = \beta_1$. Since $\mathcal{Z}_\mu$ has no accumulation point there exists (at least) two A_k, A_l such that $A_k \leftrightarrow A_l, d(A_k, A_l) = \beta_2 < \beta_1$ and $(a', b') \in A_k \times A_l$ with $a' \leftrightarrow b'$ and $d(a', b') = \beta_2$. With C^1 and C^2 as defined in the previous section we necessarily have either $a', b' \in C^1$ or $a', b' \in C^2$. Indeed if we had e.g. $a' \in C^1$ and $b' \in C^2$ then we would have $\delta^\mu(a, b') \leq \max\{\delta^\mu(a, a'), \delta^\mu(a', b')\} < \beta_1$ in contradiction with the definition of C^2. So we assume that $a', b' \in C^1$, the other case works the same way by switching a and b if needed. We define a partition of C^1 into two non-empty sets by setting

$$C^{1,1} = \{y \in C^1 \text{ such that } \delta^\mu(a', y) < \beta_2\}$$

and

$$C^{1,2} = \{y \in C^1 \text{ such that } \delta^\mu(a', y) \geq \beta_2\}.$$

Again, if $x \in C^{1,1}$ and $y \in C^{1,2}$ then necessarily $d(x, y) \geq \beta_2$. It follows that if $W_\infty(L_n^X, L_n^Y) < \beta_2$ then necessarily $L_n^X(C^{1,1}) = L_n^Y(C^{1,1})$ and $L_n^X(C^{1,2}) = L_n^Y(C^{1,2})$ simultaneously hence

$$\begin{aligned} &\{\liminf W_\infty(L_n^X, L_n^Y) < \beta_2\} \\ \subset\ &\limsup\{W_\infty(L_n^X, L_n^Y) < \beta_2\} \\ \subset\ &\limsup\left\{\{L_n^X(C^{1,1}) = L_n^Y(C^{1,1})\} \cap \{L_n^X(C^{1,2}) = L_n^Y(C^{1,2})\}\right\}. \end{aligned}$$

By defining for every integer $i \geq 1$

$$U_i = \left\{ \begin{array}{ll} 1 & \text{if } X_i \in C^{1,1} \\ 2 & \text{if } X_i \in C^{1,2} \\ 3 & \text{if } X_i \in C^2 \end{array} \right.$$

and V_i accordingly on the ground of Y_i we see that

$$\begin{aligned} &\mathbb{P}(\limsup\{U^n = V^n\}) \\ =\ &\mathbb{P}(\limsup\left\{\{L_n^X(C^{1,1}) = L_n^Y(C^{1,1})\} \cap \{L_n^X(C^{1,2}) = L_n^Y(C^{1,2})\}\right\}) \end{aligned}$$

hence $\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) < \beta_2) = 0$ according to Lemma 4.2. Thus, from (4.6) we deduce that

$$\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) \in \{\beta_1, \beta_2\}) = 1. \tag{4.10}$$

Next we show that necessarily $\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) \leq \beta_2) = 1$. Indeed, it follows from Lemma 4.2 that $\mathbb{P}(\limsup\{L_n^X(C^1) = L_n^X(C^1)\}) = 1$ so we can define an increasing sequence $(T_i)_{i \geq 1}$ of diverging random times such that for every integer $i \geq 1, L_{T_i}^X(C^1) = L_{T_i}^Y(C^1)$. We further define four processes by means of

$$T_1^{X,1} = \inf\{n \geq 1, X_n \in C^1\}, \quad T_{i+1}^{X,1} = \inf\{n > T_i^{X,1}, X_n \in C^1\},$$
$$T_1^{X,2} = \inf\{n \geq 1, X_n \in C^2\}, \quad T_{i+1}^{X,1} = \inf\{n > T_i^{X,1}, X_n \in C^2\},$$
$$T_1^{Y,1} = \inf\{n \geq 1, Y_n \in C^1\}, \quad T_{i+1}^{Y,1} = \inf\{n > T_i^{Y,1}, Y_n \in C^1\},$$
$$T_1^{Y,2} = \inf\{n \geq 1, Y_n \in C^2\}, \quad T_{i+1}^{Y,1} = \inf\{n > T_i^{Y,1}, Y_n \in C^2\},$$

and

$$\tilde{X} = (X_{T_1^{X,1}}, X_{T_2^{X,1}}, \ldots, X_{T_n^{X,1}}, \ldots) \quad \tilde{Y} = (Y_{T_1^{Y,1}}, Y_{T_2^{Y,1}}, \ldots, Y_{T_n^{Y,1}}, \ldots)$$
$$\hat{X} = (X_{T_1^{X,2}}, X_{T_2^{X,2}}, \ldots, X_{T_n^{X,2}}, \ldots) \quad \hat{Y} = (Y_{T_1^{Y,2}}, Y_{T_2^{Y,2}}, \ldots, Y_{T_n^{Y,2}}, \ldots).$$

In words, we sort the X_i's and Y_i's according to whether they take values in C^1 or C^2. Clearly $\tilde{X}, \tilde{Y}, \hat{X}$ and $\hat{Y}$ are four independent processes of independent random variables. The variables $\tilde{X}_i$ and $\tilde{Y}_j$ are all distributed according to μ^1 which is μ conditioned on C^1 and the $\hat{X}_i$ and $\hat{Y}_j$ are all distributed according to μ^2 which is μ conditioned on C^2. Since for every integer $i \geq 1$ we have

$$W_\infty(L_{T_i}^X, L_{T_i}^Y) \leq \max\left\{W_\infty(L_{T_i}^{\hat{X}}, L_{T_i}^{\hat{Y}}), W_\infty(L_{T_i}^{\tilde{X}}, L_{T_i}^{\tilde{Y}})\right\}$$

we get

$$\begin{aligned}
\liminf W_\infty(L_n^X, L_n^Y) &\leq \liminf W_\infty(L_{T_i}^X, L_{T_i}^Y) \\
&\leq \limsup W_\infty(L_{T_i}^X, L_{T_i}^Y) \\
&\leq \max\left\{\limsup W_\infty(L_n^{\hat{X}}, L_n^{\hat{Y}}), \limsup W_\infty(L_n^{\tilde{X}}, L_n^{\tilde{Y}})\right\}.
\end{aligned}$$

But $\sup \mathcal{Z}_{\mu^1} \leq \beta_2$ and $\sup \mathcal{Z}_{\mu^2} \leq \beta_2$ hence

$$\mathbb{P}(\limsup W_\infty(L_n^{\hat{X}}, L_n^{\hat{Y}}) \leq \beta_2) = \mathbb{P}(\limsup W_\infty(L_n^{\tilde{X}}, L_n^{\tilde{Y}}) \leq \beta_2) = 1$$

according to (4.7) whence $\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) \leq \beta_2) = 1$ which combined with (4.10) concludes the proof of (4.8).

4.3. *Proof of (4.9).* Finally let us assume that there are (at least) two different $\{A_i, A_j\}$ and $\{A_k, A_l\}$ such that $A_i \leftrightarrow A_j, A_k \leftrightarrow A_l$ and $d(A_i, A_j) = d(A_k, A_l) = \beta_1$. There exists a, b, a', b' such that $a \leftrightarrow b, a' \leftrightarrow b', a \in A_i, b \in A_j, a' \in A_k$ and $b' \in A_l$. Since at least one of A_k or A_l is not A_i or A_j, at least one of a' or b' is not a or b, let us denote it c. We define a partition of $\mathcal{S}(\mu)$ into three non-empty sets by setting

$$D^1 = \{y \in \mathcal{S}(\mu) \text{ such that } \delta^\mu(a, y) < \beta_1\}$$

and

$$D^2 = \{y \in \mathcal{S}(\mu) \text{ such that } \delta^\mu(a, y) \geq \beta_1, \delta^\mu(c, y) < \beta_1\}$$

and

$$D^3 = \{y \in \mathcal{S}(\mu) \text{ such that } \delta^\mu(a, y) \geq \beta_1, \delta^\mu(c, y) \geq \beta_1\}.$$

Again, for any two x and y in two different elements of the partition D^1, D^2 and D^3 we necessarily have $d(x, y) \geq \beta_1$. Hence

$$\{W_\infty(L_n^X, L_n^Y) < \beta_1\} \subset \{L_n^X(D^1) = L_n^Y(D^1)\} \cap \{L_n^X(D^2) = L_n^Y(D^2)\}$$

whence

$$\mathbb{P}(\liminf W_\infty(L_n^X, L_n^Y) < \beta_1) \leq \mathbb{P}(\limsup\{W_\infty(L_n^X, L_n^Y) < \beta_1\})$$

$$\leq \mathbb{P}(\limsup\left\{\{L_n^X(D^1) = L_n^Y(D^1)\} \cap \{L_n^X(D^2) = L_n^Y(D^2)\}\right\}) = 0$$

according to Lemma 4.2.

Proof of Lemma 4.2 We shall proceed by means of an elementary coupling. To the sequences U and V we associate the sequences $\widetilde{U}$ and $\widetilde{V}$ by setting $\widetilde{U}_i = (u_1, \dots, u_L) \in \{0,1\}^L$ with $u_k = 1$ if $U_i = k$ all other components of $\widetilde{U}_i$ being 0. The random variables $\widetilde{V}_i$ are defined accordingly on the ground of the V_i's, and we set $Z_i = \widetilde{U}_i - \widetilde{V}_i$. If $L = 2$ then $Z_i \in \{(0,0), e_1, -e_1\}$ with $e_1 = (1,-1)$ and $\mathbb{P}(Z_i = e_1) = \mathbb{P}(Z_i = -e_1)$. If $L = 3$ then $Z_i \in \{(0,0,0), e_1, e_2, e_3, -e_1, -e_2, -e_3\}$ with $e_1 = (1,-1,0), e_2 = (1,0,-1), e_3 = (0,1,-1)$ and for every $j = 1,2,3$ we have $\mathbb{P}(Z_i = e_j) = \mathbb{P}(Z_i = -e_j)$. Hence $\mathbf{Z}_n = \sum_{l=1}^n Z_i$, $\mathbf{Z}_0 = (0, \dots, 0)$ corresponds to the position of a symmetric random walk on $\mathbb{Z}^d$ with $d = \frac{L(L-1)}{2}$, and $\{U^n = V^n\} = \{\mathbf{Z}_n = 0\}$, while $\{U^n \neq V^n\} = \{\mathbf{Z}_n \neq 0\}$. Thus the first claim corresponds to the fact that a one-dimensional symmetric random walk is infinitely often away from zero with probability one. The second claim corresponds to the fact that one-dimensional symmetric random walks are recurrent. The third claim corresponds to the fact that three-dimensional symmetric random walks are transient. $\qquad\square$

5. Additional proofs

Here we give the proofs of the lemmas presented in Section 2.1. They are ordered so as to minimize the total length of the paper and not according to the expository order. Lemmas 2.2, 2.3 and 2.4 are already established in Champion et al. (2008) and Jylhä (2015) for sequences rather than nets. We give their proofs as well as a proof of (1.4) for the paper to be self-contained.

5.1. *One more lemma.* We shall employ the following Portmanteau-type result several times in the present section

Lemma 5.1. *Let $(P^j)_{j \in J}$ be a net of Borel probability measures on a metric space (Y, δ) that converges weakly to some probability measure P. For every open $U \subset Y$ we have*

$$\liminf_{j \in J} P^j(U) \geq P(U).$$

Proof of Lemma 5.1. We follow the proof of Theorem 11.1.1 in Dudley (2002). Let $F = U^c$. For every $x \in Y$ we define $\delta(x, F) = \inf_{y \in F} \delta(x, y)$ and for every $m \geq 1$, $f_m(x) = \min(1, m\delta(x, F))$ and $F_m = f_m^{-1}(\{1\})$. We see that f_m is an increasing sequence of bounded and continuous functions that converges to $\mathbb{1}_U$. For every $\varepsilon > 0$ there exists an M_0 such that for every $m \geq M_0$ we have $P(F_m) > P(U) - \varepsilon$ and for those m we have

$$
\begin{aligned}
\liminf_{j \in J} P^j(U) &\geq \liminf_{j \in J} \int_Y f_m(x) P^j(dx) \\
&= \int_Y f_m(x) P(dx) \\
&\geq P(F_m) \\
&> P(U) - \varepsilon
\end{aligned}
$$

the conclusion follows by letting $\varepsilon \to 0$. $\square$

5.2. *Proof of Lemma 2.3.* Assume that the announced statement is not true: There exist an $x \in \mathcal{S}(P)$ such that for every net $(x^j)_{j \in J}$ satisfying $x^j \in \mathcal{S}(P^j)$ the net does not converge to x. Consider $(x^j)_{j \in J}$ with x^j defined by $d(x, x^j) = \inf_{y \in \mathcal{S}(P^j)} d(x, y)$. Then there exist an $\varepsilon > 0$ and a cofinal $L \subset J$ such that for every $l \in L$ $d(x, x^l) > \varepsilon$. For every $l \in L$ we have $P^l(B(x, \varepsilon/2)) = 0$ while $P(B(x, \varepsilon/2)) > 0$ by definition. This is impossible according to Lemma 5.1 and the conclusion follows. $\square$

5.3. *Proof of Lemma 2.4.* Let $n \geq 2, (x_1, y_1), \ldots, (x_n, y_n) \in \mathcal{S}(P)$ and $\sigma \in \mathfrak{S}_n$. Due to Lemma 2.3 for every $1 \leq i \leq n$ there exists $(x_i^j, y_i^j)_{j \in J}$ such that for every $j \in J$ $(x_i^j, y_i^j) \in \mathcal{S}(P^j)$ and $\lim_{j \in J}(x_i^j, y_i^j) = (x_i, y_i)$. Since for every $j \in J$, P^j is infinitely cyclically monotone we have $\max_{1 \leq i \leq n} d(x_i^j, y_i^j) \leq \max_{1 \leq i \leq n} d(x_i^j, y_{\sigma(i)}^j)$ and passing to the limit we get $\max_{1 \leq i \leq n} d(x_i, y_i) \leq \max_{1 \leq i \leq n} d(x_i, y_{\sigma(i)})$. $\square$

5.4. *Proof of Lemma 2.2.* For every $j \in J$ we denote $x^j = \sup \mathcal{S}(P^j)$ and $x = \sup \mathcal{S}(P)$. Let us assume that (2.1) does not hold i.e. there exists an $\eta > 0$ such that $\limsup_{j \in J} x^j < x - \eta$. As a consequence there exists a subnet $(P^l)_{l \in L}$ of $(P^j)_{j \in J}$ such that for every $l \in L$ we have $P^l(]x - \eta/2, x + \eta/2[) = 0$ while by definition $P(]x - \eta/2, x + \eta/2[) > 0$. But according to Lemma 5.1 we should have $\liminf_{l \in L} P^l(]x - \eta/2, x + \eta/2[) \geq P(]x - \eta/2, x + \eta/2[)$. The conclusion follows.$\square$

5.5. *Proof of Lemma 2.1.* Let $\nu^1, \nu^2 \in M^1(E)$ and $\Pi = (A_1, \ldots, A_L) \in \mathcal{P}$. According to Lemma 1.3 there exists $P \in \mathcal{C}_\infty(\nu^1, \nu^2)$ such that $W_\infty(\nu^1, \nu^2) = \sup \mathcal{S}(P \circ d^{-1})$. Consider $P^\Pi \in M^1(\{s_1, \ldots, s_L\}^2)$ defined by

$$P^\Pi = \sum_{i,j=1}^L P(A_i \times A_j) \delta_{(s_i, s_j)}.$$

Since $P \in \mathcal{C}_\infty(\nu^1, \nu^2)$ necessarily $P^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$ hence

$$
\begin{aligned}
W_\infty(\pi(\nu^1), \pi(\nu^2)) &\leq \sup \mathcal{S}(P^\Pi \circ d^{-1}) \\
&\leq \sup \mathcal{S}(P \circ d^{-1}) + 2\Delta(\Pi) \\
&\leq W_\infty(\nu^1, \nu^2) + 2\Delta(\Pi).
\end{aligned}
$$

We are left to prove that $W_\infty(\nu^1, \nu^2) \leq W_\infty(\pi(\nu^1), \pi(\nu^2)) + 2\Delta(\Pi)$. Again, according to Lemma 1.3 there exists $Q^\Pi \in \mathcal{C}_\infty(\pi(\nu^1), \pi(\nu^2))$ such that

$$W_\infty(\pi(\nu^1), \pi(\nu^2)) = \sup \mathcal{S}(Q^\Pi \circ d^{-1}).$$

Clearly Q^Π is of the form $Q^\Pi = \sum_{i,j=1}^L Q_{i,j} \delta_{(s_i, s_j)}$. Now we consider $Q \in M^1(E^2)$ defined by

$$Q(F) = \sum_{i,j=1}^L \frac{Q_{i,j}}{\nu^1(A_i)\nu^2(A_j)} \nu^1 \otimes \nu^2(F \cap (A_i \times A_j)). \qquad (5.1)$$

Since $Q^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$ we see that $Q_{i,j} = 0$ as soon as $\nu^1(A_i) = 0$ or $\nu^2(A_j) = 0$ so Q is well defined and we can assume that the sum in (5.1) runs over i, j such

that $\nu^1(A_i)\nu^2(A_j) \neq 0$. Let us check that e.g. $Q_1 = \nu^1$. For every measurable $U \subset E$ we have

$$
\begin{aligned}
Q(U \times E) &= \sum_{i,j=1}^{L} \frac{Q_{i,j}}{\nu^1(A_i)\nu^2(A_j)} \nu^1 \otimes \nu^2((U \times E) \cap (A_i \times A_j)) \\
&= \sum_{i,j=1}^{L} \frac{Q_{i,j}}{\nu^1(A_i)\nu^2(A_j)} \nu^1(U \cap A_i)\nu^2(A_j) \\
&= \sum_{i,j=1}^{L} \frac{Q_{i,j}}{\nu^1(A_i)} \nu^1(U \cap A_i) \\
&= \sum_{i=1}^{L} \frac{1}{\nu^1(A_i)} \left(\sum_{j=1}^{L} Q_{i,j} \right) \nu^1(U \cap A_i) \\
&= \sum_{i=1}^{L} \nu^1(U \cap A_i) \qquad\qquad\qquad\qquad\qquad (5.2) \\
&= \nu^1(U)
\end{aligned}
$$

where (5.2) is due to the fact that $Q^\Pi \in \mathcal{C}(\pi(\nu^1), \pi(\nu^2))$. Now

$$
\begin{aligned}
W_\infty(\nu^1, \nu^2) &\leq \sup \mathcal{S}(Q \circ d^{-1}) \\
&\leq \sup \mathcal{S}(Q^\Pi \circ d^{-1}) + 2\Delta(\Pi) \\
&\leq W_\infty(\pi(\nu^1), \pi(\nu^2)) + 2\Delta(\Pi).
\end{aligned}
$$

which concludes the proof. $\square$

5.6. *On W_∞ and the Hausdorff distance.* Here we prove (1.4) i.e. for any integer $n \geq 1$ and every $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ in E^n we have

$$
W_\infty(L_n^X, L_n^Y) = d_{\mathcal{H}}(\mathcal{S}(L_n^X), \mathcal{S}(L_n^Y))
$$

where $L_n^X = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ and $L_n^Y = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$. Indeed, for every integer $1 \leq i \leq n$ and every $\sigma \in \mathfrak{S}_n$ we have

$$
d(x_i, y_{\sigma(i)}) \geq \inf_{y \in \mathcal{S}(L_n^Y)} d(x_i, y)
$$

hence

$$
\max_{i=1,\ldots,n} d(x_i, y_{\sigma(i)}) \geq \sup_{x \in \mathcal{S}(L_n^X)} \inf_{y \in \mathcal{S}(L_n^Y)} d(x, y)
$$

whence

$$
W_\infty(L_n^X, L_n^Y) = \min_{\sigma \in \mathfrak{S}_n} \max_{i=1,\ldots,n} d(x_i, y_{\sigma(i)}) \geq \sup_{x \in \mathcal{S}(L_n^X)} \inf_{y \in \mathcal{S}(L_n^Y)} d(x, y).
$$

The proof that $W_\infty(L_n^X, L_n^Y) \geq \sup_{y \in \mathcal{S}(L_n^Y)} \inf_{x \in \mathcal{S}(L_n^X)} d(x, y)$ is the same once one has noticed that $W_\infty(L_n^X, L_n^Y) = \min_{\sigma \in \mathfrak{S}_n} \max_{i=1,\ldots,n} d(x_{\sigma(i)}, y_i)$.

To prove the reverse inequality notice that there exists $x^* \in \mathcal{S}(L_n^X)$ and $y^* \in \mathcal{S}(L_n^Y)$ such that $W_\infty(L_n^X, L_n^Y) = d(x^*, y^*)$ and without loss of generality we can claim that for every $y \in \mathcal{S}(L_n^Y)$ we have $d(x^*, y) \geq d(x^*, y^*)$. In words, transporting

the mass at x^* somewhere else than y^* would have been at least as expensive. We get $\inf_{y \in \mathcal{S}(L_n^Y)} d(x^*, y) \geq d(x^*, y^*)$ hence

$$\sup_{x \in \mathcal{S}(L_n^X)} \inf_{y \in \mathcal{S}(L_n^Y)} d(x, y) \geq d(x^*, y^*)$$

whence $d_{\mathcal{H}}(\mathcal{S}(L_n^X), \mathcal{S}(L_n^Y)) \geq W_\infty(L_n^X, L_n^Y)$ which concludes the proof of the announced claim.

Acknowledgements

References

T. Champion, L. De Pascale and P. Juutinen. The ∞-Wasserstein distance: local solutions and existence of optimal transport maps. *SIAM J. Math. Anal.* **40** (1), 1–20 (2008). MR2403310.

T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley Series in Telecommunications. John Wiley & Sons, Inc., New York (1991). ISBN 0-471-06259-6. MR1122806.

M. C. Delfour and J.-P. Zolésio. *Shapes and geometries*, volume 22 of *Advances in Design and Control.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition (2011). ISBN 978-0-898719-36-9. MR2731611.

A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition (1998). ISBN 0-387-98406-2. MR1619036.

R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge (2002). ISBN 0-521-00754-2. MR1932358.

P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York (1997). ISBN 0-471-07672-4. MR1431744.

A. Ganesh and N. O'Connell. Large and moderate deviations for matching problems and empirical discrepancies. *Markov Process. Related Fields* **13** (1), 85–98 (2007). MR2321752.

D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence* **15**, 850–863 (1993). DOI: 10.1109/34.232073.

O. Jesorsky, K. J. Kirchberg and R. W. Frischholz. Robust face detection using the hausdorff distance. pages 90–95. Springer (2001). DOI: 10.1007/3-540-45344-X_14.

H. Jylhä. The L^∞ optimal transport: infinite cyclical monotonicity and the existence of optimal transport maps. *Calc. Var. Partial Differential Equations* **52** (1-2), 303–326 (2015). MR3299183.

J. L. Kelley. *General topology.* Springer-Verlag, New York-Berlin (1975). MR0370454.

T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9** (2), 161–187 (1989). MR1030371.

K. R. Parthasarathy. *Probability measures on metric spaces.* Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London (1967). MR0226684.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (1998). ISBN 3-540-62772-3. MR1491362.

O. Schutze, X. Esquivel, A. Lara and C. A. C. Coello. Using the averaged hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation* **16**, 504–522 (2012). DOI: 10.1109/TEVC.2011.2161872.

D. Serre. *Matrices*, volume 216 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (2002). ISBN 0-387-95460-0. MR1923507.

D.-G. Sim, O.-K. Kwon and R.-H. Park. Object matching algorithms using robust hausdorff distance measures. *IEEE Transactions on Image Processing* **8**, 425–429 (1999). DOI: 10.1109/83.748897.

M. Talagrand. Matching theorems and empirical discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* **7** (2), 455–537 (1994). MR1227476.

C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin (2009). ISBN 978-3-540-71049-3. MR2459454.

R. Wang, X. Wang and L. Wu. Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition. *Statist. Probab. Lett.* **80** (5-6), 505–512 (2010). MR2593592.

G. T. Whyburn. *Analytic Topology.* American Mathematical Society Colloquium Publications, v. 28. American Mathematical Society, New York (1942). MR0007095.