First version: December 17, 1996
Revised version: May 26, 1998

EXPONENTIAL APPROXIMATIONS IN COMPLETELY REGULAR TOPOLOGICAL SPACES AND EXTENSIONS OF SANOV'S THEOREM

PETER EICHELSBACHER AND UWE SCHMOCK

Abstract. This paper is devoted to the well known transformations that preserve a large deviation principle (LDP), namely, the contraction principle with approximately continuous maps and the concepts of exponential equivalence and exponential approximations. We generalize these transformations to completely regular topological state spaces, give some examples and, as an illustration, reprove a generalization of Sanov's theorem, due to de Acosta [1]. Using partition-dependent couplings, we then extend this version of Sanov's theorem to triangular arrays and prove a full LDP for the empirical measures of exchangeable sequences with a general measurable state space.

1. Introduction and statement of the results

The question when a large deviation principle (LDP) for a family of laws can be deduced from the LDP for another family was studied for the first time in [3]. There the probability measures are defined on a metric space. The concepts nowadays called "exponential equivalence" and "exponential approximations" were formalized more generally in the exposition [9, Chapter 4.2], where the probability measures are defined on a metric space, too. A direct consequence of this concept is the proof of a contraction principle for approximately continuous maps, where the maps are defined on a Hausdorff topological space and take values in a metric space. In this paper we will generalize the concepts of "exponential equivalence" and "exponential approximations" to a completely regular topological space $(Y, \mathcal{T})$, also called Tychonoff space, and we will generalize the contraction principle accordingly. The set $\mathcal{M}_1(S)$ of all probability measures on a general measurable space $(S, \mathcal{S})$, equipped with the so-called τ-topology of setwise convergence, is an example of such a completely regular topological space. Our general results allow us to reprove Sanov's theorem in this natural setting, extend it to triangular arrays and prove a full LDP for the empirical measures of suitable exchangeable processes—without imposing restrictions on the measurable state space $(S, \mathcal{S})$. The basic tool for these extensions is the construction of couplings, which depend on finite $\mathcal{S}$-measurable partitions of S, to get exponentially good approximations.

Let us start with some topological considerations. A metric on a space Y can be regarded as providing a concept of nearness that is applicable throughout the space. When we want to consider approximations in more general topological spaces, we

1991 *Mathematics Subject Classification.* 60F10 (primary), 60G09 (secondary).

Key words and phrases. Large deviations, exponential equivalence, contraction principle, gauge space, uniform space, approximately continuous map, triangular array, exchangeable sequence.

still need such a uniformly applicable concept of nearness. Two types of topological spaces seem to be appropriate for this purpose:

- A gauge space $(Y,\mathcal{T})$, which means that the topology $\mathcal{T}$ is generated by a family $\mathcal{D}$ of pseudometrics which is separating, that is, for each pair of points $x \neq y$ in Y there exists a pseudometric $d \in \mathcal{D}$ such that $d(x,y) \neq 0$.
- A topological space $(Y,\mathcal{T})$ with a separating uniform structure, which is compatible with the topology $\mathcal{T}$.

As is well-known [13, Chap. IX, Theorems 10.6 and 11.4], the above two types of topological spaces are the same, and they coincide with the completely regular ones. Remember that a topological space $(Y,\mathcal{T})$ is called completely regular if $(Y,\mathcal{T})$ is Hausdorff and if for every closed set $C \subset Y$ and every point $y \in Y \setminus C$, there exists a continuous function $f\colon Y \to [0,1]$ such that $f(y)=1$ and $f(\tilde{y})=0$ for all $\tilde{y} \in C$. Note that every locally compact Hausdorff space and every Hausdorff topological vector space is completely regular (see [13, Chap. XI, Theorem 6.4] for the former and construct a uniform structure for the latter).

To further justify the above topological setting, notice that in a regular topological space Y, the rate function associated with the LDP is unique and Varadhan's integral lemma is applicable. For Bryc's inverse Varadhan lemma, however, the complete regularity of the topological space is needed (see [9, Chap. 4]).

For notational convenience, we will use the language of gauge spaces in the following. We will assume throughout this paper, that $(Y,\mathcal{T})$ is a gauge space and that the separating family $\mathcal{D}$ of pseudometrics generates its topology $\mathcal{T}$. Note that the collection of all balls

$$B(y,d,\delta) \equiv \{\, x \in Y \mid d(x,y) < \delta \,\}, \qquad y \in Y,\, d \in \mathcal{D},\, \delta > 0, \tag{1.1}$$

is a subbasis of the topology $\mathcal{T}$. Let $\mathcal{D}'$ be the smallest family of pseudometrics on Y which contains $\mathcal{D}$ and is closed with respect to maxima, meaning that for every choice of $d_1, d_2 \in \mathcal{D}'$, the pseudometric defined by $d(x,y) \equiv \max\{d_1(x,y), d_2(x,y)\}$ for $x,y \in Y$ also belongs to $\mathcal{D}'$. The collection of all balls $B(y,d,\delta)$ with $y \in Y$, $d \in \mathcal{D}'$ and $\delta > 0$ is a basis of $\mathcal{T}$.

In this paper we take special care to mention any connection between the topology of a space and its σ-algebra, provided we need such a connection. In applications of our results (see Theorem 1.16, for instance), it is convenient, when the σ-algebra does not need to be the Borel σ-algebra. We do not complete our probability spaces.

With respect to the gauge space Y, we assume throughout that it is equipped with a σ-algebra $\mathcal{Y}$, which contains the above collection (1.1) of balls. Note that $\mathcal{Y}$ might be smaller than the Borel σ-algebra $\sigma(\mathcal{T})$. We denote by $\mathcal{M}_1(Y)$ the set of all probability measures on $(Y,\mathcal{Y})$.

In a general gauge space with an uncountable family $\mathcal{D}$ of pseudometrics, a countable base for the uniform structure may not exist. Therefore, it is necessary to replace approximating sequences by approximating nets. For this purpose, let $(I,\preceq)$ denote a nonempty directed set, meaning that $\preceq$ is a reflexive and transitive relation such that for all $i,j \in I$ there exists a $k \in I$ satisfying $i \preceq k$ and $j \preceq k$. For a net $\{a_i\}_{i\in I} \subset \bar{\mathbb{R}}$ we define as usual $\limsup_{i\in I} a_i = \inf_{j\in I} \sup_{i\in I, i\succeq j} a_i$ and $\liminf_{i\in I} a_i = \sup_{j\in I} \inf_{i\in I, i\succeq j} a_i$.

Now we are ready to define the term "exponentially good approximation" and "exponential equivalence" in our context; illustrations are given in Section 2.

Definition 1.2. (a) A collection $\{\mu_{\varepsilon,i}\}_{\varepsilon>0,i\in I} \subset \mathcal{M}_1(Y)$ is called *$\mathcal{D}$-exponentially good approximation* of $\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0} \subset \mathcal{M}_1(Y)$, if for every $d \in \mathcal{D}'$, $\varepsilon > 0$ and $i \in I$ there exists a probability measure $\nu_{d,\varepsilon,i}$ on a σ-algebra $\mathcal{Y}_{d,\varepsilon,i}$ containing $\mathcal{Y}^{\otimes 2}$ such that the two marginals are $\mu_{\varepsilon,i}$ and $\tilde{\mu}_\varepsilon$, respectively, and

$$\limsup_{i\in I} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \nu^*_{d,\varepsilon,i}(\{\,(y,\tilde{y}) \in Y^2 \mid d(y,\tilde{y}) > \delta\,\}) = -\infty \tag{1.3}$$

for every $\delta > 0$. Here $\nu^*_{d,\varepsilon,i}$ denotes the outer measure induced by $\nu_{d,\varepsilon,i}$.

(b) If (a) holds for a collection $\{\mu_\varepsilon\}_{\varepsilon>0}$, which does not depend on $i \in I$, then $\{\mu_\varepsilon\}_{\varepsilon>0}$ is called *$\mathcal{D}$-exponentially equivalent* to $\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0}$.

Remark 1.4. If the measures $\nu_{d,\varepsilon,i}$ and σ-algebras $\mathcal{Y}_{d,\varepsilon,i}$ in Definition 1.2(a) do not depend on $d \in \mathcal{D}'$, then (1.3) for all $d \in \mathcal{D}$ implies (1.3) for all $d \in \mathcal{D}'$.

On a topological space X, a lower semicontinuous function $J\colon X \to [0,\infty]$ is called rate function. If, in addition, the level set $\{\, x \in X \mid J(x) \le L\,\}$ is quasi-compact for every $L \in [0,\infty)$, meaning that every open cover of this set has a finite subcover, then J is called good. In a Hausdorff space, the latter condition implies the lower semicontinuity.

Let $\mathcal{X}$ be a σ-algebra on a topological space X. A family $\{\mu_\varepsilon\}_{\varepsilon>0} \subset \mathcal{M}_1(X,\mathcal{X})$ of probability measures on $(X,\mathcal{X})$ is said to satisfy a weak LDP with rate function J, if the large deviations lower bound

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(A) \ge -\inf_{x\in \mathrm{int}(A)} J(x)$$

holds for all $A \in \mathcal{X}$, where $\mathrm{int}(A)$ denotes the interior of A, and if the large deviations upper bound

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu_\varepsilon(A) \le -\inf_{x\in \mathrm{cl}(A)} J(x) \tag{1.5}$$

holds for all those $A \in \mathcal{X}$ for which the closure $\mathrm{cl}(A)$ is quasi-compact. If, in addition, the upper bound (1.5) holds for all $A \in \mathcal{X}$, then $\{\mu_\varepsilon\}_{\varepsilon>0}$ is said to satisfy a full large deviation principle.

The following four results generalize [9, Theorem 4.2.13, Theorem 4.2.16, Corollary 4.2.21 and Theorem 4.2.23]. All proofs are deferred to Section 3.

Theorem 1.6. *If $\{\mu_{\varepsilon,i}\}_{\varepsilon>0,i\in I} \subset \mathcal{M}_1(Y)$ is a $\mathcal{D}$-exponentially good approximation of $\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0} \subset \mathcal{M}_1(Y)$ and if, for every $i \in I$, the family $\{\mu_{\varepsilon,i}\}_{\varepsilon>0}$ satisfies a full LDP with a (not necessarily good) rate function J_i, then the following statements hold:*

(a) *$\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0}$ satisfies a weak LDP with rate function*

$$J(y) \equiv \sup_{d\in\mathcal{D}',\delta>0} \liminf_{i\in I} \inf_{z\in B(y,d,\delta)} J_i(z), \qquad y \in Y. \tag{1.7}$$

(b) *If J is a good rate function and if for every measurable closed subset C of Y*

$$\inf_{y\in C} J(y) \le \limsup_{i\in I} \inf_{y\in C} J_i(y), \tag{1.8}$$

then the full LDP holds for $\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0}$ with the good rate function J.

Remark 1.9. In (1.7), we may replace the collection $\{B(y,d,\delta)\}_{d\in\mathcal{D}',\delta>0}$ of balls by any filterbase of neighbourhoods of y converging to y without changing J.

As an illustration, we will show in Example 2.4 how to use Theorem 1.6 to derive Sanov's theorem in the τ-topology on $\mathcal{M}_1(S)$ for a general measurable state space $(S,\mathcal{S})$ from the elementary version of Sanov's theorem for finite state spaces. This example reproves the main result of de Acosta [1]. As applications of Theorem 1.6, we will extend de Acosta's version of Sanov's theorem to triangular arrays in Theorem 1.16(b) and will prove a LDP for the empirical measures of suitable exchangeable processes. Further applications are contained in [15] and [16], which motivated this paper.

Let us first present two easy consequences of Theorem 1.6:

Corollary 1.10. *Assume the general hypotheses of Theorem* 1.6*. If the rate functions J_i are good and independent of $i\in I$, then J, given by* (1.7)*, is equal to every J_i and $\{\tilde\mu_\varepsilon\}_{\varepsilon>0}$ satisfies the full LDP with the good rate function J.*

Corollary 1.11. *Assume that $f\colon X\to Y$ is a measurable and continuous map from a topological space X with σ-algebra $\mathcal{X}$ to the gauge space Y with the σ-algebra $\mathcal{Y}$ introduced above, and assume that a family of probability measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ on $(X,\mathcal{X})$ satisfies a full LDP with a good rate function J. Furthermore, assume that for every $\varepsilon>0$ there is a measurable function $f_\varepsilon\colon X\to Y$ such that*

$$\limsup_{\varepsilon\downarrow 0}\varepsilon\log\mu_\varepsilon^*(\{\,x\in X\mid d(f(x),f_\varepsilon(x))>\delta\,\})=-\infty \tag{1.12}$$

for all $d\in\mathcal{D}$ and $\delta>0$, where μ_ε^* denotes the outer measure induced by the measure μ_ε. Then $\{\mu_\varepsilon f_\varepsilon^{-1}\}_{\varepsilon>0}$ satisfies the full LDP on Y with the good rate function $J_f(y)\equiv\inf_{x\in f^{-1}(y)}J(x)$.

Theorem 1.6 also leads to the following extension of the contraction principle for approximately continuous maps:

Theorem 1.13. *Assume that $\{\mu_\varepsilon\}_{\varepsilon>0}$ is a family of probability measures that satisfies a full LDP with a good rate function J on a topological space X with σ-algebra $\mathcal{X}$. Given $L\in[0,\infty)$, denote by $K_L\equiv\{\,x\in X\mid J(x)\le L\,\}$ the corresponding level set. For every $i\in I$ let $f_i\colon X\to Y$ be a measurable continuous map, where Y is the gauge space introduced above. Assume that there exists $f\colon X\to Y$ satisfying*

$$\limsup_{i\in I}\sup_{x\in K_L}d(f_i(x),f(x))=0 \tag{1.14}$$

for every $L\in[0,\infty)$ and $d\in\mathcal{D}$. Then every family $\{\tilde\mu_\varepsilon\}_{\varepsilon>0}\subset\mathcal{M}_1(Y)$, for which $\{\mu_\varepsilon f_i^{-1}\}_{\varepsilon>0,i\in I}$ is a $\mathcal{D}$-exponentially good approximation, satisfies the full LDP in Y with the good rate function $J_f(y)\equiv\inf_{x\in f^{-1}(y)}J(x)$.

As an application of Theorem 1.6 and its first corollary, let us extend de Acosta's version [1, Theorem 1.1] of Sanov's theorem, which we reprove in Example 2.4, to triangular arrays in part (b) of Theorem 1.16 below. Part (b) corresponds to [3, Theorem 5], which was stated in the context of Polish state spaces and the weak topology. Part (a) is the proper τ-topology version of [3, Theorem 5] for proving the full LDP for the empirical measures of suitable exchangeable processes with a general measurable state space.

In the following two theorems, let $\mathcal{M}_1(S,\mathcal{S})$, or $\mathcal{M}_1(S)$ for brevity if this is unambiguous, denote the set of probability measures on a measurable space $(S,\mathcal{S})$. The unit interval $[0,1]$ is equipped with its usual topology and the corresponding Borel σ-algebra. The τ-topology on $\mathcal{M}_1(S)$ is defined to be the coarsest topology which makes the maps $\mathcal{M}_1(S) \ni \mu \mapsto \mu(A) \in [0,1]$ continuous for every $A \in \mathcal{S}$. Note that this turns $\mathcal{M}_1(S)$ into a gauge space with the separating family $\mathcal{D} \equiv \{d_A\}_{A\in\mathcal{S}}$ of pseudometrics, where $d_A(\mu,\nu) \equiv |\mu(A) - \nu(A)|$ for $\mu, \nu \in \mathcal{M}_1(S)$. Let $\mathcal{B}(\mathcal{M}_1(S))$ be the σ-algebra generated by the maps $\mathcal{M}_1(S) \ni \mu \mapsto \mu(A) \in [0,1]$ with $A \in \mathcal{S}$. Let $\{X_k\}_{k\in\mathbb{N}}$ denote the projection maps on the product space $(\Omega, \mathcal{A}) \equiv (S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}})$ and define the empirical measure $L_n(\omega) = \frac{1}{n}\sum_{k=1}^n \delta_{X_k(\omega)} \in \mathcal{M}_1(S)$ for every $\omega \in \Omega$ and $n \in \mathbb{N}$. Note that $L_n\colon \Omega \to \mathcal{M}_1(S)$ is $\mathcal{A}$-$\mathcal{B}(\mathcal{M}_1(S))$-measurable. Given $\mu \in \mathcal{M}_1(S)$, define $\mathbb{P} = \mu^{\otimes\mathbb{N}}$ on $(\Omega, \mathcal{A})$ and $J = H(\cdot\,|\,\mu)$, where

$$H(\nu\,|\,\mu) \equiv \begin{cases} \int_S \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu}\, d\mu, & \text{if } \nu \ll \mu, \\ \infty, & \text{otherwise,} \end{cases} \tag{1.15}$$

denotes the relative entropy of $\nu \in \mathcal{M}_1(S)$ with respect to μ. By [1, Lemma 2.1], the level sets of J are τ-compact, therefore J is a good rate function. Analogously, given a collection $\{\mu_i\}_{i\in I} \subset \mathcal{M}_1(S)$, define $\mathbb{P}_i = \mu_i^{\otimes\mathbb{N}}$ on $(\Omega, \mathcal{A})$ and $J_i = H(\,\cdot\,|\,\mu_i)$ for every $i \in I$.

Theorem 1.16. (a) *Let $\{\mu_i\}_{i\in I}$ be a net of probability measures on the measurable space $(S, \mathcal{S})$ converging to a measure μ in the τ-topology. Then $\{\mathbb{P}_i L_n^{-1}\}_{n\in\mathbb{N}, i\in I}$ is a $\{d_A\}_{A\in\mathcal{S}}$-exponentially good approximation of $\{\mathbb{P} L_n^{-1}\}_{n\in\mathbb{N}}$ and the good rate functions J and $\{J_i\}_{i\in I}$ satisfy* (1.7) *and* (1.8) *for every τ-closed $C \subset \mathcal{M}_1(S)$.*

(b) *If $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(S)$ converges to μ in the τ-topology, then $\{\mathbb{P}_n L_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP in the τ-topology on $\mathcal{M}_1(S)$ with the good rate function J.*

Remark 1.17. (a) According to [17, Lemmas 1.23 and 1.24], there always exists a countably generated sub-σ-algebra $\mathcal{S}_0$ of $\mathcal{S}$ such that $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all $A \in \mathcal{S}_0$ implies convergence of $\{\mu_n\}_{n\in\mathbb{N}}$ to μ in the τ-topology.

(b) If S is a Polish space with Borel σ-algebra $\mathcal{S}$, then the τ-convergence of $\{\mu_n\}_{n\in\mathbb{N}}$ to μ is in general a stronger hypothesis than the weak convergence assumed in [3, Theorem 5]. But in this case, the assertion of Theorem 1.16(b) is also a stronger one, because the full LDP holds in the finer τ-topology of $\mathcal{M}_1(S)$.

(c) If S is a Polish space with Borel σ-algebra $\mathcal{S}$, then $\mathcal{B}(\mathcal{M}_1(S))$ coincides with the Borel σ-algebra of the weak topology of $\mathcal{M}_1(S)$, see [5, Lemma 2.1].

(d) Since, in general, S does not have a topology, the notion of separability is not available and the diagonal $\{\,(s,s') \in S^2 \mid s = s'\,\}$ need not to be in $\mathcal{S} \otimes \mathcal{S}$, see [19, Example 6.19].

Due to the measurability problem mentioned in Remark 1.17(d), the coupling used in the proof of [3, Theorem 1] does not seem to be directly extendible to our case. Using the set $\mathcal{P}$ of all finite $\mathcal{S}$-measurable partitions of S, we can explicitly construct weaker, partition-dependent couplings to verify (1.3) and, therefore, prove Theorem 1.16.

Given Theorem 1.16(a), we can derive a full LDP for the empirical measures of certain mixtures of i. i. d. sequences with a general measurable state space $(S, \mathcal{S})$. Let X be a topological space with σ-algebra $\mathcal{X}$. We assume that $\mathcal{X}$ contains a base of the topology of X, but $\mathcal{X}$ does not need to be the Borel σ-algebra of X. A particular example of such a space is any subset X of $\mathcal{M}_1(S)$ equipped with the relative τ-topology and the relative σ-algebra $\mathcal{X} \equiv \{\,A \cap X \mid A \in \mathcal{B}(\mathcal{M}_1(S))\,\}$ it inherits from $\mathcal{M}_1(S)$. For further examples, see [6, Theorem 1.15] and [21, page 57]. Let $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ be an $\mathcal{X}$-$\mathcal{B}(\mathcal{M}_1(S))$-measurable map and define $\mathbb{P}_x = \mu_x^{\otimes\mathbb{N}}$

on $(\Omega, \mathcal{A})$ for every $x \in X$. Then $X \ni x \mapsto \mathbb{P}_x(A)$ is measurable for every $A \in \mathcal{A}$. Given a probability measure Σ on $(X, \mathcal{X})$, define the probability measure $\mathbb{P}_\Sigma$ on $(\Omega, \mathcal{A})$ by

$$\mathbb{P}_\Sigma(A) = \int_X \mathbb{P}_x(A)\,\Sigma(dx), \qquad A \in \mathcal{A}. \tag{1.18}$$

The projection maps $\{X_k\}_{k\in\mathbb{N}}$ are exchangeable under $\mathbb{P}_\Sigma$.

Theorem 1.19. *Assume that the map $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ is $\mathcal{X}$-$\mathcal{B}(\mathcal{M}_1(S))$-measurable. Define $J_X(\nu) \equiv \inf_{x\in X} H(\nu\,|\,\mu_x)$ for all $\nu \in \mathcal{M}_1(S)$. Then the following statements hold:*

(a) *J_X is a rate function with respect to the τ-topology on $\mathcal{M}_1(S)$.*
(b) *For every $A \in \mathcal{B}(\mathcal{M}_1(S))$ with τ-compact closure $\mathrm{cl}_\tau(A)$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\Sigma(L_n \in A) \le - \inf_{\nu\in \mathrm{cl}_\tau(A)} J_X(\nu). \tag{1.20}$$

(c) *If the map $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ is τ-continuous and if $\Sigma(U) > 0$ for every open $U \in \mathcal{X}$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\Sigma(L_n \in A) \ge - \inf_{\nu\in \mathrm{int}_\tau(A)} J_X(\nu).$$

for every $A \in \mathcal{B}(\mathcal{M}_1(S))$, where $\mathrm{int}_\tau(A)$ denotes the τ-interior of A.
(d) *If X is quasi-compact and if $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ is τ-continuous, then the rate function J_X is good and (1.20) holds for all $A \in \mathcal{B}(\mathcal{M}_1(S))$.*

Remark 1.21. (a) We explicitly start with the mixture $\mathbb{P}_\Sigma$ given by (1.18) instead of assuming that the projection maps $\{X_k\}_{k\in\mathbb{N}}$ are an exchangeable process, because de Finetti's representation theorem [2] does not hold in our general setting. See [12, Section 2] for such an exchangeable process, whose state space $S \subset [0,1]$ is a separable metric space equipped with its Borel σ-algebra.

(b) If S is a Polish space with Borel σ-algebra $\mathcal{S}$, if X is first countable and if $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ is continuous with respect to the weak topology on $\mathcal{M}_1(S)$, then the analogue of Theorem 1.19 for the weak topology is given in [10, p. 1153].

(c) If one accepts upper bounds which are not given in terms of J_X, then one can relax the quasi-compactness assumption in part (d), see [11, Theorem 5.1].

(d) As the proofs of Theorem 1.19(c) and (d) show, the statement of Theorem 1.16(a) is the substitute for "exponential continuity" introduced in [10, (1.7)] in the context of a first countable topological space X.

(e) Note that the τ-continuity of $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ does not imply the τ-continuity of $X \ni x \mapsto \mathbb{P}_x \in \mathcal{M}_1(\Omega)$; see [9, Exercise 7.3.18] for an example with $S = [0,1]$ equipped with the Borel σ-algebra, $X = \mathcal{M}_1(S)$ and the identity map.

(f) After submitting this paper we learnt that a very similar result to Theorem 1.19 is proved in [7, Section 2] by different methods.

2. Examples and Illustrations

We give some examples and illustrations, where completely regular state spaces are involved and which are not covered by the results in [9, Section 4.2.2].

Example 2.1. As already mentioned is Section 1, the set $\mathcal{M}_1(S)$ is a gauge space with the separating family $\{d_A\}_{A\in\mathcal{S}}$ of pseudometrics, which generate the τ-topology. Obviously, the σ-algebra $\mathcal{B}(\mathcal{M}_1(S))$ contains all balls given by the pseudometrics $\{d_A\}_{A\in\mathcal{S}}$. Therefore, it satisfies our general assumption on the σ-algebra of the gauge space. Every bounded and $\mathcal{S}$-measurable $\varphi\colon S\to\mathbb{R}$ can be uniformly approximated by simple functions, hence $\mathcal{M}_1(S)\ni\mu\mapsto\int_S\varphi\,d\mu$ is measurable and τ-continuous for such functions. A special case of Corollary 1.10 is the following result, which was already proved in [21, Lemma 4.3.1] and uses $\{d_A\}_{A\in\mathcal{S}}$-exponential equivalence: Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space and assume that the map $L_n\colon\Omega\to\mathcal{M}_1(S)$ is $\mathcal{A}$-$\mathcal{B}(\mathcal{M}_1(S))$-measurable for each $n\in\mathbb{N}$. Assume that the sequence $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP in the τ-topology with a good rate function J. If the map $\tilde{L}_n\colon\Omega\to\mathcal{M}_1(S)$ is $\mathcal{A}$-$\mathcal{B}(\mathcal{M}_1(S))$-measurable for every $n\in\mathbb{N}$ and

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(|L_n(A)-\tilde{L}_n(A)|>\delta)=-\infty$$

for all $\delta>0$ and $A\in\mathcal{S}$, then $\{\mathbb{P}\tilde{L}_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP in the τ-topology with the good rate function J, too.

Let d_{v} denote the total variation distance on $\mathcal{M}_1(S)$. Since $d_A(\mu,\nu)\le d_{\mathrm{v}}(\mu,\nu)$ for every $A\in\mathcal{S}$, a full LDP for $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ in the τ-topology on $\mathcal{M}_1(S)$ with a good rate function is transferred to every d_{v}-exponentially equivalent sequence. For such an application, see [4, Proposition 1.10].

Example 2.2. The following topology is introduced in [15]. Let $(S,\mathcal{S})$ be a measurable space and let $(E,\|\cdot\|_E)$ be a separable real Banach space with Borel σ-algebra $\mathcal{E}$. Let Φ be a set of $\mathcal{S}$-$\mathcal{E}$-measurable functions $\varphi\colon S\to E$ containing $B(S,E)$, the set of all bounded E-valued, $\mathcal{S}$-$\mathcal{E}$-measurable functions on S. Define the restricted set of probability measures on $(S,\mathcal{S})$ by

$$\mathcal{M}_1^{\Phi}(S)=\Big\{\mu\in\mathcal{M}_1(S)\;\Big|\;\int_S\|\varphi\|_E\,d\mu<\infty\text{ for every }\varphi\in\Phi\Big\}.$$

Then, for every $\mu\in\mathcal{M}_1^{\Phi}(S)$ and $\varphi\in\Phi$, the Bochner integral $\int_S\varphi\,d\mu$ exists. Let $\tau_1^{\Phi}(E)$ denote the coarsest topology on $\mathcal{M}_1^{\Phi}(S)$ such that the map $\mathcal{M}_1^{\Phi}(S)\ni\mu\mapsto\int_S\varphi\,d\mu$ is continuous for every $\varphi\in\Phi$. If $\Phi=B(S,E)$, then $\mathcal{M}_1^{\Phi}(S)=\mathcal{M}_1(S)$; if, in addition, $E=\mathbb{R}$, then $\tau_1^{\Phi}(E)$ coincides with the usual τ-topology. For each $\varphi\in\Phi$ define $d_\varphi(\mu,\nu)=\|\int_S\varphi\,d\mu-\int_S\varphi\,d\nu\|_E$ for $\mu,\nu\in\mathcal{M}_1^{\Phi}(S)$. Then $\{d_\varphi\}_{\varphi\in\Phi}$ is a separating family of pseudometrics for the $\tau_1^{\Phi}(E)$-topology. In [15] and [16] the tool of $\{d_\varphi\}_{\varphi\in\Phi}$-exponential equivalence is used several times to obtain full large deviation principles in the $\tau_1^{\Phi}(E)$-topology.

Example 2.3. Using the notation of Example 2.2, let $D([0,\infty),E)$ denote the space of E-valued càdlàg functions on $[0,\infty)$ equipped with the topology of uniform convergence on compact subsets. Let $\mathcal{M}_+(S)$ be the set of all finite, nonnegative measures on $(S,\mathcal{S})$ and, similarly to Example 2.2, let $\mathcal{M}_+^{\Phi}(S)$ be the corresponding restricted set of all $\mu\in\mathcal{M}_+(S)$ satisfying $\int_S\|\varphi\|_E\,d\mu<\infty$ for all $\varphi\in\Phi$. Let $D([0,\infty),\mathcal{M}_+^{\Phi}(S))$ be the set of all those $y\colon[0,\infty)\to\mathcal{M}_+^{\Phi}(S)$, for which $[0,\infty)\ni t\mapsto\int_S\varphi\,dy_t$ is in $D([0,\infty),E)$ for every $\varphi\in\Phi$. On $D([0,\infty),\mathcal{M}_+^{\Phi}(S))$ we take the weakest topology such that all these maps from $D([0,\infty),\mathcal{M}_+^{\Phi}(S))$ to $D([0,\infty),E)$

are continuous. This topology is generated by the family of pseudometrics

$$d_{\varphi,T}(y,z) \equiv \sup_{t\in[0,T]} \Bigl\| \int_S \varphi\,dy_t - \int_S \varphi\,dz_t \Bigr\|_E, \qquad y,z \in D([0,\infty),\mathcal{M}_+^\Phi(S)),$$

with $\varphi \in \Phi$ and $T > 0$. This family is separating. For $E = \mathbb{R}$ and $\Phi = B(S,\mathbb{R})$, this topology was constructed in [8] to get a full LDP for partial sums processes and used in [14] to get a moderate deviations principle for functional partial sums processes.

In the last example we reprove a generalization of Sanov's theorem due to de Acosta [1, Theorem 1.1], in which the state space $(S,\mathcal{S})$ is an arbitrary measurable space and the set $\mathcal{M}_1(S)$ is endowed with the τ-topology.

Example 2.4. Let us return to the setting introduced just before Theorem 1.16. Our aim in this example is to apply Theorem 1.6 in order to show that the sequence $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP in the τ-topology on $\mathcal{M}_1(S)$ with the good rate function $J \equiv H(\cdot\,|\,\mu)$. This result is used to prove Theorems 1.16 and 1.19.

To define a $\mathcal{D}$-exponentially good approximation, let us introduce some more notation. By an $\mathcal{S}$-measurable partition P of the measurable space $(S,\mathcal{S})$ we mean a finite collection $\{A_1,\dots,A_n\} \subset \mathcal{S}$ of disjoint sets whose union is S. Let $\mathcal{P}$ be the set of all $\mathcal{S}$-measurable partitions of S. For $P \in \mathcal{P}$ let $\sigma(P)$ denote the σ-algebra generated by P. Fix $\mu \in \mathcal{M}_1(S)$ and define $\mathcal{P}_\mu = \{\, P \in \mathcal{P} \mid \mu(A) > 0$ for all $A \in P\,\}$. Note that $(\mathcal{P}_\mu, \preceq)$ is a directed set, provided that $P \preceq P'$ for $P, P' \in \mathcal{P}_\mu$ means $P \subset \sigma(P')$, which is equivalent to saying that P' is a refinement of P. Given $P \in \mathcal{P}$, the empirical measure $L_n(\omega)$ can also be considered as a probability measure on the measurable space $(S,\sigma(P))$. Note that $\mathcal{M}_1(S,\sigma(P))$ can be identified with $\{\,(x_A)_{A\in P} \in [0,1]^P \mid \sum_{A\in P} x_A = 1\,\}$ and that the τ-topology and the σ-algebra induced on $\mathcal{M}_1(S,\sigma(P))$ by the maps $\mathcal{M}_1(S,\sigma(P)) \ni \nu \mapsto \nu(A)$ with $A \in P$ coincides with the natural topology and Borel σ-algebra of this subset of $[0,1]^P$. It follows from Sanov's theorem for finite state spaces [9, Theorem 2.1.10] that, for every partition $P \in \mathcal{P}$, the sequence $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S,\sigma(P)))$ satisfies a full LDP. The corresponding good rate function is given by $H_P(\nu\,|\,\mu) \equiv \sum_{A\in P} \nu(A)\log(\nu(A)/\mu(A))$ for all $\nu \in \mathcal{M}_1(S,\sigma(P))$, with the understanding that $0\log 0 = 0\log(0/0) = 0$ and $a\log(a/0) = \infty$ for $a > 0$.

For every $P \in \mathcal{P}_\mu$ define

$$\mathcal{M}_1(S,\sigma(P)) \ni \nu \mapsto \Psi_P(\nu) = \sum_{A\in P} \nu(A)\frac{\mu(\,\cdot\,\cap A)}{\mu(A)} \in \mathcal{M}_1(S).$$

These maps are measurable and continuous with respect to the τ-topologies. Hence, by the elementary contraction principle (see e. g. [9, Theorem 4.2.1]), the measures $\{\mathbb{P}L_{n,P}^{-1}\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$ with $L_{n,P} \equiv \Psi_P(L_n)$ satisfy a full LDP in the τ-topology on $\mathcal{M}_1(S)$ with the good rate function $J_P\colon \mathcal{M}_1(S) \to [0,\infty]$ given by

$$J_P(\nu) \equiv \begin{cases} H_P(\tilde\nu\,|\,\mu), & \text{if } \nu = \Psi_P(\tilde\nu) \text{ with } \tilde\nu \in \mathcal{M}_1(S,\sigma(P)),\\ \infty, & \text{otherwise.}\end{cases}$$

Using Remark 1.4, let us show that $\{\mathbb{P}L_{n,P}^{-1}\}_{n\in\mathbb{N},P\in\mathcal{P}_\mu}$ is a $\{d_A\}_{A\in\mathcal{S}}$-exponentially good approximation of $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ in $\mathcal{M}_1(\mathcal{M}_1(S))$: If $A \in \mathcal{S}$ satisfies $\mu(A) \in \{0,1\}$, then $\mathbb{P}(d_A(L_n, L_{n,P}) > 0) = 0$ for all $n \in \mathbb{N}$ and $P \in \mathcal{P}_\mu$; if $\mu(A) \in (0,1)$, then this is true at least for all $P \in \mathcal{P}_\mu$ which are refinements of $\{A, A^c\}$.

To apply Theorem 1.6, it remains to show that (1.7) and (1.8) are satisfied. First note that the definitions of $H_P(\nu\,|\,\mu)$ and $\Psi_P(\nu)$ extend from $\nu\in\mathcal{M}_1(S,\sigma(P))$ to $\nu\in\mathcal{M}_1(S)$ for every $P\in\mathcal{P}$ or $P\in\mathcal{P}_\mu$, respectively. Secondly, in view of Remark 1.9, note that the collection of all sets $\{\,\tilde\nu\in\mathcal{M}_1(S)\mid \max_{A\in P} d_A(\tilde\nu,\nu)<\delta\,\}$ with $P\in\mathcal{P}$ and $\delta>0$ is a neighbourhood filterbase of $\nu\in\mathcal{M}_1(S)$.

To prove (1.8), consider any $P\in\mathcal{P}_\mu$ and $\nu\in\mathcal{M}_1(S)$ satisfying $J_P(\nu)<\infty$. Then $\nu=\Psi_P(\nu)$ and $S\ni s\mapsto\sum_{A\in P}(\nu(A)/\mu(A))1_A(s)$ is a density of ν with respect to μ, hence $J(\nu)=J_P(\nu)$. This implies (1.8).

To prove "$\geq$" in (1.7), it suffices to consider $\nu\in\mathcal{M}_1(S)$ satisfying $J(\nu)<\infty$. Note that $\nu\ll\mu$ in this case. Given $P\in\mathcal{P}$, we can join all μ-null sets of P to one $A\in P$ with $\mu(A)>0$ to obtain a $P'\in\mathcal{P}_\mu$ with $P'\subset\sigma(P)$. Define $\tilde\nu=\Psi_{P'}(\nu)$. Since $\nu\ll\mu$, it follows that $d_A(\tilde\nu,\nu)=0$ for all $A\in P$. By [18, Proposition 15.5(c)], every refinement $P''\in\mathcal{P}_\mu$ of P' satisfies $J_{P''}(\tilde\nu)=H_{P''}(\tilde\nu\,|\,\mu)\leq H(\nu\,|\,\mu)=J(\nu)$. This shows that the right-hand side of (1.7) is bounded by $J(\nu)$.

To prove "$\leq$" in (1.7) for a $\nu\in\mathcal{M}_1(S)$, we consider the following two cases:

If $\nu\not\ll\mu$, then there exists an $A\in\mathcal{S}$ satisfying $\mu(A)=0$ and $\nu(A)>0$. Hence, $B(\nu,d_A,\nu(A))\cap\Psi_P(\mathcal{M}_1(S,\sigma(P)))=\varnothing$ for every $P\in\mathcal{P}_\mu$ because $\Psi_P(\tilde\nu)(A)=0$ for all $\tilde\nu\in\mathcal{M}_1(S,\sigma(P))$. Therefore, the right-hand side of (1.7) equals infinity.

If $\nu\ll\mu$, then, given $r<H(\nu\,|\,\mu)$, there exists $P\in\mathcal{P}$ such that $r<H_P(\nu\,|\,\mu)$, see [18, Corollary 15.7]. By joining all μ-null sets of P to one $A\in P$ with $\mu(A)>0$, if necessary, we may assume that $P\in\mathcal{P}_\mu$. Since $[0,1]\ni x\mapsto x\log(x/\mu(A))$ is continuous for every $A\in P$, there exists $\delta>0$ such that $r<H_P(\tilde\nu\,|\,\mu)$ for all $\tilde\nu\in\mathcal{M}_1(S)$ with $\max_{A\in P} d_A(\nu,\tilde\nu)<\delta$. By [18, Proposition 15.5(c)], $H_P(\tilde\nu\,|\,\mu)\leq H_{P'}(\tilde\nu\,|\,\mu)$ for all refinements $P'\in\mathcal{P}_\mu$ of P. Hence, $r<J_{P'}(\tilde\nu)$ for all refinements $P'\in\mathcal{P}_\mu$ of P and all $\tilde\nu\in\mathcal{M}_1(S)$ with $\max_{A\in P} d_A(\nu,\tilde\nu)<\delta$, which implies "$\leq$" in (1.7).

3. Proofs

Proof of Theorem 1.6*:* (a) By our assumptions on the σ-algebra $\mathcal{Y}$ of Y, the basis of the topology $\mathcal{T}$ of Y consisting of all balls $B(y,d',\delta)$ with $y\in Y$, $d'\in\mathcal{D}'$ and $\delta>0$ is contained in $\mathcal{Y}$. By [9, Theorem 4.1.11], the family $\{\tilde\mu_\varepsilon\}_{\varepsilon>0}$ satisfies a weak LPD with rate function

$$\tilde J(y)\equiv-\inf_{d'\in\mathcal{D}',\,\delta>0}\liminf_{\varepsilon\downarrow0}\varepsilon\log\tilde\mu_\varepsilon(B(y,d',\delta)),\qquad y\in Y,\tag{3.1}$$

provided that we can prove the relation

$$\tilde J(y)=-\inf_{d'\in\mathcal{D}',\,\delta>0}\limsup_{\varepsilon\downarrow0}\varepsilon\log\tilde\mu_\varepsilon(B(y,d',\delta)),\qquad y\in Y.\tag{3.2}$$

While proving (3.2), we will see that $\tilde J$ from (3.1) equals J defined in (1.7).

Given $d'\in\mathcal{D}'$, there exists by the definition of $\mathcal{D}'$ a finite set $D\subset\mathcal{D}$ such that $d'(x,y)=\max_{d\in D}d(x,y)$ for all $x,y\in Y$. Given any $\delta>0$ and $y\in Y$, the inclusion

$$B(y,d',\delta)\times Y\subset(Y\times B(y,d',2\delta))\cup\bigcup_{d\in D}\bigl\{(z,\tilde z)\in Y^2\;\big|\;d(z,\tilde z)>\delta\bigr\}\tag{3.3}$$

is valid. Using the probability measures which bring about the $\mathcal{D}$-exponentially good approximation formulated in Definition 1.2(a), we obtain from (3.3)

$$\mu_{\varepsilon,i}(B(y,d',\delta))\leq\tilde\mu_\varepsilon(B(y,d',2\delta))+\sum_{d\in D}\nu^*_{d,\varepsilon,i}\bigl(\bigl\{(z,\tilde z)\in Y^2\;\big|\;d(z,\tilde z)>\delta\bigr\}\bigr)$$

for all $\varepsilon > 0$ and $i \in I$. Using the large deviations lower bound of $\{\mu_{\varepsilon,i}\}_{\varepsilon>0}$ for every $i \in I$ and (1.3), it follows that

$$\begin{aligned} -\liminf_{i\in I} \inf_{z\in B(y,d',\delta)} J_i(z) &\le \limsup_{i\in I} \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mu_{\varepsilon,i}(B(y,d',\delta)) \\ &\le \liminf_{\varepsilon\downarrow 0} \varepsilon \log \tilde{\mu}_\varepsilon(B(y,d',2\delta)). \end{aligned} \tag{3.4}$$

Using (3.3) in a similar way with the diameter of the balls increased by δ, we obtain

$$\tilde{\mu}_\varepsilon(B(y,d',2\delta)) \le \mu_{\varepsilon,i}(B(y,d',3\delta)) + \sum_{d\in D} \nu^*_{d,\varepsilon,i}\big(\{\, (z,\tilde{z}) \in Y^2 \;\big|\; d(z.\tilde{z}) > \delta \,\}\big)$$

for all $\varepsilon > 0$ and $i \in I$. Using (1.3) and the large deviations upper bound of $\{\mu_{\varepsilon,i}\}_{\varepsilon>0}$ for every $i \in I$, it follows that

$$\begin{aligned} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \tilde{\mu}_\varepsilon(B(y,d',2\delta)) &\le \limsup_{i\in I} \limsup_{\varepsilon\downarrow 0} \varepsilon \log \mu_{\varepsilon,i}(B(y,d',3\delta)) \\ &\le -\liminf_{i\in I} \inf_{z\in B(y,d',4\delta)} J_i(z), \end{aligned} \tag{3.5}$$

where we used that the closure of $B(y,d',3\delta)$ is contained in $B(y,d',4\delta)$. We can combine (3.4) and (3.5) to a single chain of inequalities. Taking the supremum over all $d' \in \mathcal{D}'$ and $\delta > 0$ afterwards, we see that the left- and right-hand sides agree, hence the right-hand sides of (3.1), (3.2) and (1.7) are equal.

(b) Fix a measurable set $A \in \mathcal{Y}$, let C be its closure, and define the truncation $r_n = \min\{n, \inf_{x\in C} J(x) - 1/n\}$ for all $n \in \mathbb{N}$. For every $y \in Y \setminus C$ there exist $d_y \in \mathcal{D}'$ and $\delta_y > 0$ such that $B(y,d_y,2\delta_y) \subset Y \setminus C$. Note that $B(y,d_y,2\delta_y) \in \mathcal{Y}$ by definition of $\mathcal{Y}$. By assumption on J, the level set $K \equiv \{\, y \in Y \mid J(y) \le r_n \,\}$ is compact. In addition, $K \subset Y \setminus C$ by the definition of r_n. Hence, there exists a finite subset F of K such that $V \equiv \bigcup_{y\in F} B(y,d_y,\delta_y)$ covers K. The pseudometric $d \equiv \max_{y\in F} d_y$ is in $\mathcal{D}'$. Define $\delta = \min_{y\in F} \delta_y$. Since

$$\begin{aligned} Y \times A \subset Y \times C \subset Y \times \Big(Y \setminus \bigcup_{y\in F} B(y,d_y,2\delta_y)\Big) \\ &\subset \big((Y\setminus V) \times Y\big) \cup \bigcup_{y\in F} \{\, (z,\tilde{z}) \in Y^2 \mid d_y(z,\tilde{z}) > \delta_y \,\} \\ &\subset \big((Y\setminus V) \times Y\big) \cup \{\, (z,\tilde{z}) \in Y^2 \mid d(z,\tilde{z}) > \delta \,\} \end{aligned}$$

and $V \in \mathcal{Y}$, we get, for every $\varepsilon > 0$ and $i \in I$,

$$\tilde{\mu}_\varepsilon(A) \le \mu_{\varepsilon,i}(Y\setminus V) + \nu^*_{d,\varepsilon,i}(\{\, (z,\tilde{z}) \in Y^2 \mid d(z,\tilde{z}) > \delta \,\}).$$

The large deviations upper bounds for $\{\mu_{\varepsilon,i}\}_{\varepsilon>0, i\in I}$, applied to the measurable closed set $Y \setminus V$, and (1.3) imply that

$$\limsup_{\varepsilon\downarrow 0} \varepsilon \log \tilde{\mu}_\varepsilon(A) \le -\limsup_{i\in I} \inf_{y\in Y\setminus V} J_i(y). \tag{3.6}$$

Using assumption (1.8) for the set $Y \setminus V$, we get

$$-\limsup_{i\in I} \inf_{y\in Y\setminus V} J_i(y) \le -\inf_{y\in Y\setminus V} J(y) \le -\inf_{y\in Y\setminus K} J(y) \le -r_n. \tag{3.7}$$

Combining (3.6) and (3.7) and letting $n \to \infty$, the upper bound follows. □

For proving Corollaries 1.10, 1.11 and Theorem 1.13, we need the following extension of [9, Lemma 4.1.6(b)].

Lemma 3.8. *Let J be a good rate function on the gauge space Y. Then*

$$\inf_{y\in\bar{A}} J(y) = \sup_{d\in\mathcal{D}',\delta>0} \inf_{y\in A_{d,\delta}} J(y)$$

for every $A \subset Y$, where $A_{d,\delta} \equiv \{\, y \in Y \mid$ There exists $z \in A$ with $d(y,z) \le \delta \,\}$ is called closed (d,δ)-blowup of A.

Proof. Since $\bar{A} = \bigcap_{d\in\mathcal{D}'} \bigcap_{\delta>0} A_{d,\delta}$ and $\bar{A} \subset A_{d,\delta}$ for every $d \in \mathcal{D}'$ and $\delta > 0$, it suffices to prove that

$$\gamma \equiv \sup_{d\in\mathcal{D}',\delta>0} \inf_{y\in A_{d,\delta}} J(y) \ge \inf_{y\in\bar{A}} J(y) - \eta \tag{3.9}$$

for every $\eta > 0$. Obviously, it suffices to consider the case $\gamma < \infty$. Fix $\eta > 0$, define $L = \gamma + \eta$ and let $K_L \equiv \{\, y \in Y \mid J(y) \le L \,\}$ denote the corresponding level set of J. Since J is assumed to be a good rate function, K_L is compact.

To prove (3.9) by contradiction, assume that $\bar{A} \cap K_L = \varnothing$. Since

$$\bar{A} \cap K_L = \bigcap_{d\in\mathcal{D}',\delta>0} \big(A_{d,\delta} \cap K_L\big),$$

it follows from this assumption and the finite intersection property of compact sets, that there exists a finite subcollection of $\mathcal{C} \equiv \{A_{d,\delta} \cap K_L\}_{d\in\mathcal{D}',\delta>0}$ with empty intersection. Since $(\mathcal{C}, \supset)$ is a directed set, there would exist $d \in \mathcal{D}'$ and $\delta > 0$ with $A_{d,\delta} \cap K_L = \varnothing$. By (3.9), this implies $\gamma \ge L = \gamma + \eta$, which is a contradiction. □

Proof of Corollary 1.10*:* The first part follows from Lemma 3.8. In particular, J is good. The second part then follows from Theorem 1.6, but we like to mention that in this simple case a proof of the lower bound can be given in a few lines:

Fix a measurable set $A \in \mathcal{Y}$ and let O be its open interior. For every $y \in O$ there exist $d \in \mathcal{D}'$ and $\delta > 0$ such that $B(y,d,2\delta) \subset O$. Note that

$$Y \times A \supset Y \times B(y,d,2\delta) \supset (B(y,d,\delta) \times Y) \setminus \{\, (z,\tilde{z}) \in Y^2 \mid d(z,\tilde{z}) > \delta \,\},$$

hence $\tilde{\mu}_\varepsilon(A) \ge \mu_{\varepsilon,i}(B(y,d,\delta)) - \nu^*_{d,\varepsilon,i}(\{\, (z,\tilde{z}) \in Y^2 \mid d(z,\tilde{z}) > \delta \,\})$ for all $\varepsilon > 0$ and $i \in I$. Using the large deviations lower bound of $\{\mu_{\varepsilon,i}\}_{\varepsilon>0}$ for every $i \in I$ and (1.3),

$$\liminf_{\varepsilon\downarrow 0} \varepsilon \log \tilde{\mu}_\varepsilon(A) \ge \liminf_{i\in I} \liminf_{\varepsilon\downarrow 0} \varepsilon \log \mu_{\varepsilon,i}(B(y,d,\delta)) \ge -J(y),$$

which implies the large deviations lower bound for $\{\tilde{\mu}_\varepsilon\}_{\varepsilon>0}$. □

Proof of Corollary 1.11*:* The contraction principle for continuous maps (see for example [9, Theorem 4.2.1], it holds without the Hausdorff property of X) yields the full LDP for $\{\mu_\varepsilon f^{-1}\}_{\varepsilon>0}$ with the good rate function J_f. Note that (1.12) carries over to all $d \in \mathcal{D}'$. Hence, the family $\{\mu_\varepsilon f^{-1}\}_{\varepsilon>0}$ is $\mathcal{D}$-exponentially equivalent to $\{\mu_\varepsilon f_\varepsilon^{-1}\}_{\varepsilon>0}$ and the statement follows from Theorem 1.6 and Lemma 3.8. □

Proof of Theorem 1.13*:* First note that (1.14) carries over to all $d \in \mathcal{D}'$. According to the elementary contraction principle for continuous maps, the family $\{\mu_\varepsilon f_i^{-1}\}_{\varepsilon>0}$ satisfies the full LDP with the good rate function $J_i(y) \equiv \inf_{x\in f_i^{-1}(y)} J(x)$ for every $i \in I$. By condition (1.14), the map f is continuous on each level set K_L. Therefore, $f(K_L)$ is compact for each $L \in [0,\infty)$. If $y \in Y$ satisfies $J_f(y) \le L < \infty$, then the finite intersection property, applied to the family $\{K_{L+\delta}\}_{\delta>0}$ of closed quasicompact sets, shows that the infimum in the definition of $J_f(y)$ is attained in K_L. Hence,

$\{f(K_L)\}_{L\geq 0}$ are the level sets of J_f. Therefore, J_f is a good rate function. By Theorem 1.6, it suffices to check that, for every closed subset C of Y,

$$\inf_{y\in C} J_f(y) \leq \liminf_{i\in I} \inf_{y\in C} J_i(y), \tag{3.10}$$

and that J_f has the form (1.7).

Define $\gamma_i = \inf_{y\in C} J_i(y) = \inf_{x\in f_i^{-1}(C)} J(x)$ for every $i \in I$ and $\gamma = \liminf_{i\in I} \gamma_i$. If $\gamma = \infty$, then (3.10) holds. To prove (3.10) for $\gamma < \infty$, take any $\alpha > \gamma$. We may pass to a subordinated directed set $I_\alpha \subset I$ (subordinated means that for every $i \in I$ there exists a $j \in I_\alpha$ satisfying $i \preceq j$) such that $\sup_{i\in I_\alpha} \gamma_i < \alpha$. For every $i \in I_\alpha$ there exists $x_i \in f_i^{-1}(C) \cap K_\alpha$, meaning that $f_i(x_i) \in C$ and $J(x_i) \leq \alpha$. Given $d \in \mathcal{D}'$ and $\delta > 0$, it follows from (1.14) that there exists $i_{d,\delta} \in I_\alpha$ such that $f(x_{i_{d,\delta}}) \in C_{d,\delta}$, where $C_{d,\delta}$ denotes the closed (d,δ)-blowup of C defined in Lemma 3.8. Hence, $\inf_{y\in C_{d,\delta}} J_f(y) \leq J_f(f(x_{i_{d,\delta}})) \leq J(x_{i_{d,\delta}}) \leq \alpha$. Using Lemma 3.8 and the fact that $\alpha > \gamma$ was arbitrary, (3.10) follows

To prove that J_f has the form (1.7), note that Lemma 3.8 and (3.10) imply that

$$J_f(y) = \sup_{d\in\mathcal{D}',\delta>0} \inf_{z\in\overline{B(y,d,\delta)}} J_f(z) \leq \sup_{d\in\mathcal{D}',\delta>0} \liminf_{i\in I} \inf_{z\in B(y,\delta,d)} J_i(z) \tag{3.11}$$

for every $y \in Y$, because the open balls are contained in the closed ones.

The proof of the representation (1.7) for J_f is finished, if we can show that the right-hand side of (3.11) is bounded by $J_f(y)$. It suffices to consider the case $L \equiv J_f(y) < \infty$. Then $y \in f(K_L)$ and there exists $x \in K_L$ such that $f(x) = y$. Furthermore, $y_i \equiv f_i(x) \in f_i(K_L)$ and thus $J_i(y_i) \leq L$ for all $i \in I$. By condition (1.14) there exists, for given $d \in \mathcal{D}'$ and $\delta > 0$, an $i_{d,\delta} \in I$ such that $y_i \in B(y,d,\delta)$ for all $i \in I$ with $i \succeq i_{d,\delta}$. Hence, the right-hand side of (3.11) is bounded by L. □

For the identification of the rate function in the proof of Theorem 1.16(a), we need the following property of the relative entropy:

Lemma 3.12. *Let $(S,\mathcal{S})$ be a measurable space. Then the relative entropy function $H\colon \mathcal{M}_1(S)\times\mathcal{M}_1(S) \to [0,\infty]$ defined by* (1.15) *is lower semicontinuous with respect to the product topology of the τ-topologies.*

Proof. The map $[0,1]^2 \ni (x,y) \mapsto x\log(x/y)$, with the conventions $x\log(x/0) = \infty$ for $x > 0$ and $0\log 0 = 0\log(0/0) = 0$, is lower semicontinuous. Therefore, the sum $[0,1]^{2n} \ni (x_1,\dots,x_n,y_1,\dots,y_n) \mapsto \sum_{l=1}^n x_l \log(x_l/y_l)$ is lower semicontinuous, too. As in Example 2.4, let $\mathcal{P}$ denote the set of all finite $\mathcal{S}$-measurable partitions of S. Then $\mathcal{M}_1(S)\times\mathcal{M}_1(S) \ni (\mu,\nu) \mapsto H_P(\nu\,|\,\mu) \equiv \sum_{A\in P} \nu(A)\log(\nu(A)/\mu(A))$ is lower τ-semicontinuous for every $P \in \mathcal{P}$. Since $H(\nu\,|\,\mu) = \sup_{P\in\mathcal{P}} H_P(\nu\,|\,\mu)$ by [18, Corollary 15.7], the lemma follows. □

Proof of Theorem 1.16*:* (a) Define $\mathcal{D} = \{d_A\}_{A\in\mathcal{S}}$. Let us first prove the $\mathcal{D}$-exponentially good approximation. Given $d \in \mathcal{D}'$, there exist sets $A_1,\dots,A_k \in \mathcal{S}$ with $d = \max\{d_{A_1},\dots,d_{A_k}\}$. Let P be the smallest $\mathcal{S}$-measurable partition of S such that $A_1,\dots,A_k \in \sigma(P)$. For every $i \in I$ we want to construct a coupling $\bar\mu_{i,P} \in \mathcal{M}_1(S^2,\mathcal{S}^{\otimes 2})$ of μ_i and μ. For this purpose, define $m_{A,i} = \min\{\mu_i(A),\mu(A)\}$ for $A \in P$ and $\gamma_{i,P} = \sum_{A\in P} m_{A,i}$. For $A \in P$ let $\mu(\,\cdot\,|\,A) \equiv \mu(\,\cdot\cap A)/\mu(A)$, if $\mu(A) > 0$, and, for completeness, $\mu(\,\cdot\,|\,A) \equiv \mu$ otherwise. Define the conditioned measure $\mu_i(\,\cdot\,|\,A)$ analogously. We choose the coupling

$$\bar\mu_{i,P} \equiv \sum_{A,B\in P} (\delta_{A,B}\, m_{A,i} + \vartheta_{A,B,i})\, \mu_i(\,\cdot\,|\,A) \otimes \mu(\,\cdot\,|\,B), \tag{3.13}$$

where

$$\vartheta_{A,B,i} \equiv \begin{cases} \frac{1}{1-\gamma_{i,P}}(\mu_i(A) - m_{A,i})(\mu(B) - m_{B,i}), & \text{if } \gamma_{i,P} \in [0,1), \\ 0, & \text{if } \gamma_{i,P} = 1, \end{cases}$$

and $\delta_{A,B} \equiv 1$, if $A = B$, and $\delta_{A,B} \equiv 0$, otherwise. Note that only the terms with $\mu_i(A) > 0$ and $\mu(B) > 0$ contribute to the sum in (3.13). Using these definitions, it follows that the first marginal of $\bar\mu_{i,P}$ is indeed μ_i, the second marginal is μ, and

$$\sum_{A\in P} \bar\mu_{i,P}(A \times A) = \gamma_{i,P} = 1 - \tfrac{1}{2} \sum_{A\in P} |\mu_i(A) - \mu(A)|. \tag{3.14}$$

Let $(\Omega', \mathcal{A}') \equiv (\Omega^2, \mathcal{A}^{\otimes 2})$ be the product space with canonical projections π_1 and π_2. For every $i \in I$ define the coupling measure $\mathbb{P}_{i,P} = \bar\mu_{i,P}^{\otimes\mathbb{N}}$ on $(\Omega', \mathcal{A}')$ such that $(X_k \circ \pi_1, X_k \circ \pi_2)$ has distribution $\bar\mu_{i,P}$ for every $k \in \mathbb{N}$. To verify (1.3) for $\delta > 0$, it remains to show that

$$\limsup_{i\in I} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{i,P}(d(L_n \circ \pi_1, L_n \circ \pi_2) > \delta) = -\infty. \tag{3.15}$$

By the exponential Chebycheff inequality, the triangle inequality and the independence of $\{(X_k \circ \pi_1, X_k \circ \pi_2)\}_{k\in\mathbb{N}}$, it follows that, for every $\lambda \ge 0$ and $n \in \mathbb{N}$,

$$\mathbb{P}_{i,P}(d(L_n \circ \pi_1, L_n \circ \pi_2) > \delta) \le \mathrm{e}^{-\delta\lambda n} \big(\mathbb{E}_{i,P}\big[\exp(\lambda\, d(\delta_{X_1\circ\pi_1}, \delta_{X_1\circ\pi_2}))\big]\big)^n.$$

Note that $(A_l \times A_l^{\mathrm{c}}) \cup (A_l^{\mathrm{c}} \times A_l) \subset (S \times S) \setminus \bigcup_{A\in P} A \times A$, because P is a refinement of $\{A_l, A_l^{\mathrm{c}}\}$ for every $l \in \{1, \dots, k\}$. Using also (3.14), it follows that

$$\begin{aligned} \mathbb{E}_{i,P}\big[\exp(\lambda\, d(\delta_{X_1\circ\pi_1}, \delta_{X_1\circ\pi_2}))\big] &= \int_{S^2} \exp\Big(\lambda \max_{l\in\{1,\dots,k\}} 1_{(A_l\times A_l^{\mathrm{c}})\cup(A_l^{\mathrm{c}}\times A_l)}\Big)\, d\bar\mu_{i,P} \\ &\le 1 + \mathrm{e}^\lambda\, \bar\mu_{i,P}\Bigg(\bigcup_{l\in\{1,\dots,k\}} (A_l \times A_l^{\mathrm{c}}) \cup (A_l^{\mathrm{c}} \times A_l)\Bigg) \\ &\le 1 + \mathrm{e}^\lambda (1 - \gamma_{i,P}). \end{aligned}$$

Choosing $\lambda_i = -\log(1 - \gamma_{i,P})$ with the usual convention $\log 0 = -\infty$, the above estimates show that, for every $n \in \mathbb{N}$,

$$\frac{1}{n} \log \mathbb{P}_{i,P}(d(L_n \circ \pi_1, L_n \circ \pi_2) > \delta) \le -\delta\lambda_i + \log 2. \tag{3.16}$$

Since the net $\{\mu_i\}_{i\in I}$ converges to μ in the τ-topology, it follows from (3.14) that $\gamma_{i,P} \to 1$ and therefore $\lambda_i \to \infty$, hence (3.15) follows.

We now want to prove that J and $\{J_i\}_{i\in I}$ satisfy (1.7) and (1.8) for every τ-closed $C \subset \mathcal{M}_1(S)$. According to [1, Theorem 1.1] as well as Example 2.4, the sequences $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ and $\{\mathbb{P}_i L_n^{-1}\}_{n\in\mathbb{N}}$ for $i \in I$ satisfy the full LDP in the τ-topology on $\mathcal{M}_1(S)$ with the good rate functions J and J_i, respectively. Define J' by the right-hand side of (1.7). If we can show that the hypotheses of Theorem 1.6(b) are satisfied with J' in place of J, then Theorem 1.6 implies that $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP with rate function J', too. Hence, $J = J'$ by the uniqueness of the rate function (the proof of [9, Lemma 4.1.4] applies to our situation, because $\mathcal{B}(\mathcal{M}_1(S))$ contains a base of the τ-topology of $\mathcal{M}_1(S)$).

In order to prove that J' is a good rate function, first note that J', according to its definition via (1.7), is lower τ-semicontinuous. By [1, Lemma 2.1], the level sets of J are τ-compact, hence it suffices to prove that $J \le J'$ on $\mathcal{M}_1(S)$. Consider $\nu \in \mathcal{M}_1(S)$ and $r < H(\nu\,|\,\mu)$. By Lemma 3.12, there exist $d \in \mathcal{D}'$ and $\delta > 0$ such that $H(\tilde\nu\,|\,\tilde\mu) > r$ for all $\tilde\mu \in B(\mu,d,\delta)$ and $\tilde\nu \in B(\nu,d,\delta)$. Since that net $\{\mu_i\}_{i\in I}$ converges to μ, there exists $j \in I$ such that $\mu_i \in B(\mu,d,\delta)$ for all $i \succeq j$, hence $J_i(\tilde\nu) = H(\tilde\nu\,|\,\mu_i) > r$ for all $\tilde\nu \in B(\nu,d,\delta)$ and $i \succeq j$, which implies via (1.7) that $J'(\nu) \ge r$. Since $r < H(\nu\,|\,\mu)$ was arbitrary, $J'(\nu) \ge H(\nu\,|\,\mu) = J(\nu)$.

Given a τ-closed $C \subset \mathcal{M}_1(S)$, we want to prove (1.8) with J' in place of J. For $i \in I$ define $\gamma_i = \inf_{\nu\in C} J_i(\nu)$. It suffices to consider the case $\limsup_{i\in I}\gamma_i < \infty$. Take any $r < \inf_{\nu\in C} J'(\nu)$ and any $\gamma > \limsup_{i\in I}\gamma_i$. It suffices to show that $r \le \gamma$. There exists $i_0 \in I$ such that $\gamma_i < \gamma$ for all $i \succeq i_0$. Since $H(\,\cdot\,|\,\mu_i)$ has τ-compact level sets by [1, Lemma 2.1], it follows that $\{\,\nu \in C \mid H(\nu\,|\,\mu_i) \le \gamma\,\}$ is nonempty and τ-compact for every $i \succeq i_0$, hence there exists $\nu_i \in C$ with $\gamma_i = H(\nu_i\,|\,\mu_i)$ for $i \succeq i_0$. Suppose for a moment that $\{\nu_i\}_{i\succeq i_0}$ has an accumulation point $\nu \in C$. By the definition of $J'(\nu)$ via (1.7), there exist $d \in \mathcal{D}'$, $\delta > 0$ and $i_1 \succeq i_0$ such that $r \le H(\tilde\nu\,|\,\mu_i)$ for all $i \succeq i_1$ and $\tilde\nu \in B(\nu,d,\delta)$. Since ν is assumed to be an accumulation point of $\{\nu_i\}_{i\succeq i_0}$, there exists $i \succeq i_1$ such that $\nu_i \in B(\nu,d,\delta)$. Hence, $r \le H(\nu_i\,|\,\mu_i) = \gamma_i < \gamma$.

It remains to show that $\{\nu_i\}_{i\succeq i_0}$ has an accumulation point ν in C. By Tychonoff's theorem, the product space $[0,1]^{\mathcal{S}}$ is compact. Note that the τ-topology of $\mathcal{M}_1(S)$ is the corresponding relative topology when $\mathcal{M}_1(S)$ is viewed as a subset of $[0,1]^{\mathcal{S}}$. Therefore, $\{\nu_i\}_{i\succeq i_0}$ has an accumulation point ν in $[0,1]^{\mathcal{S}}$. As such, ν is finitely additive with $\nu(S) = 1$. To conclude that ν is in the τ-closed set C, it remains to show that ν is σ-additive. For this it suffices to show that $\lim_{k\to\infty}\nu(A_k) = 0$ for every sequence $\{A_k\}_{k\in\mathbb{N}} \subset \mathcal{S}$ with $A_k \downarrow \varnothing$ as $k \to \infty$.

By convexity, $z \le e^{z-1}$ for all $z \in \mathbb{R}$. Substituting $z = x-t$ yields $x \le e^{x-t-1}+t$ for all $t, x \in \mathbb{R}$. Multiplication with $y = e^t$ gives the well-known estimate $xy \le e^{x-1} + y\log y$ for all $x \in \mathbb{R}$ and $y \in [0,\infty)$. Given $\varepsilon > 0$, there exist $\alpha, \delta > 0$ satisfying $(\gamma + 1/\mathrm{e})/\alpha \le \varepsilon$ and $2\delta\,\mathrm{e}^{\alpha-1}/\alpha \le \varepsilon$. Since $A_k \downarrow \varnothing$, there exists $l \in \mathbb{N}$ with $\mu(A_l) \le \delta$. Since the net $\{\mu_i\}_{i\in I}$ converges to μ, there exists $i_2 \succeq i_0$ such that $\mu_i(A_l) \le 2\delta$ for all $i \succeq i_2$, hence $\mu_i(A_k) \le 2\delta$ for all $i \succeq i_2$ and $k \ge l$. Since $H(\nu_i\,|\,\mu_i) = \gamma_i < \gamma < \infty$ for all $i \succeq i_0$, a density $f_i \equiv d\nu_i/d\mu_i$ exists. Using the estimate derived at the beginning of this paragraph, it follows that

$$\nu_i(A_k) = \frac{1}{\alpha}\int_{A_k} \alpha f_i\,d\mu_i \le \frac{\mathrm{e}^{\alpha-1}}{\alpha}\mu_i(A_k) + \frac{1}{\alpha}\int_{A_k} f_i\log f_i\,d\mu_i$$

for all $i \succeq i_2$ and $k \ge l$. Since $x\log x \ge -1/\mathrm{e}$ for all $x \in [0,\infty)$, it follows that $\nu_i(A_k) \le \varepsilon + (H(\nu_i\,|\,\mu_i) + 1/\mathrm{e})/\alpha \le 2\varepsilon$ for all $i \succeq i_2$ and $k \ge l$. Hence, the accumulation point ν also satisfies $\nu(A_k) \le 2\varepsilon$ for every $k \ge l$. Thus, $\lim_{k\to\infty}\nu(A_k) = 0$.

(b) By [1, Theorem 1.1] as well as Example 2.4, the sequence $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ satisfies the full LDP in the τ-topology on $\mathcal{M}_1(S)$ with the good rate function $J \equiv H(\,\cdot\,|\,\mu)$ given by (1.15). According to Corollary 1.10 it suffices to show that $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ is $\mathcal{D}$-exponentially equivalent to $\{\mathbb{P}_nL_n^{-1}\}_{n\in\mathbb{N}}$. Setting $i = n$ in (3.16), this follows from (3.16) and the τ-convergence of $\{\mu_n\}_{n\in\mathbb{N}}$ to μ, because $\lambda_n \to \infty$ as $n \to \infty$. □

Proof of Theorem 1.19: (a) It follows from Lemma 3.12 that J_X is lower semicontinuous with respect to the τ-topology on $\mathcal{M}_1(S)$, hence J_X is a rate function.

(b) *Upper bound for $A \in \mathcal{B}(\mathcal{M}_1(S))$ with τ-compact closure C:* Given $\varepsilon > 0$, define $\gamma_\varepsilon = \min\{1/\varepsilon, \inf_{\nu \in C} J_X(\nu) - \varepsilon\}$. By the lower τ-semicontinuity of J_X, there exits, for every $\nu \in C$, a τ-open neighbourhood $U_\nu \in \mathcal{B}(\mathcal{M}_1(S))$ of ν such that $J_X(\tilde{\nu}) \geq \gamma_\varepsilon$ for all $\tilde{\nu}$ in the τ-closure $\mathrm{cl}_\tau(U_\nu)$ of U_ν. By the τ-compactness of C, there exists a finite subset M of C such that $\bigcup_{\nu \in M} U_\nu$ covers C. Since $A \subset C \subset \bigcup_{\nu \in M} U_\nu$, the large deviations upper bound for the sequence $\{\mathbb{P}_x L_n^{-1}\}_{n \in \mathbb{N}}$, see [1, Theorem 1.1] or Example 2.4, implies that for every $x \in X$

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(L_n \in A) &\leq \max_{\nu \in M} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(L_n \in U_\nu) \\ &\leq -\min_{\nu \in M} \inf_{\tilde{\nu} \in \mathrm{cl}_\tau(U_\nu)} H(\tilde{\nu}\,|\,\mu_x) \leq -\gamma_\varepsilon,\end{aligned}$$

where we used the definition of the function J_X and the choice of U_ν for $\nu \in C$. Since $\mathbb{P}_\Sigma(L_n \in A) \leq \sup_{x \in X} \mathbb{P}_x(L_n \in A)$ and since $\varepsilon > 0$ is arbitrary, (1.20) follows.

(c) *Lower bound:* Let $A \subset \mathcal{M}_1(S)$ be measurable and let $O \subset A$ be its τ-open interior. Given $(\nu, x) \in O \times X$, we have to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\Sigma(L_n \in A) \geq -H(\nu\,|\,\mu_x). \tag{3.17}$$

It suffices to consider the case $H(\nu\,|\,\mu_x) < \infty$. We claim that, for every $\varepsilon > 0$, there exist a measurable open neighbourhood U_x of x and $n_0 \in \mathbb{N}$ such that, for every $y \in U_x$ and $n \geq n_0$,

$$\mathbb{P}_y(L_n \in A) \geq \exp\bigl(-n(H(\nu\,|\,\mu_x) + \varepsilon)\bigr). \tag{3.18}$$

Assume that this claim were not true. Then there exists a filterbase $\{U_i\}_{i \in I} \subset \mathcal{X}$ of open neighbourhoods of x converging to x and $x_i \in U_i$ for every $i \in I$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{x_i}(L_n \in A) \leq -H(\nu\,|\,\mu_x) - \varepsilon. \tag{3.19}$$

Since $X \ni y \mapsto \mu_y$ is τ-continuous, the net $\{\mu_{x_i}\}_{i \in I}$ converges to μ_x. Hence, we are in the setting of Theorem 1.16(a) and obtain via (1.7) that

$$H(\nu\,|\,\mu_x) \geq \liminf_{i \in I} \inf_{\tilde{\nu} \in O} H(\tilde{\nu}\,|\,\mu_{x_i}). \tag{3.20}$$

Using the large deviations lower bound of $\{\mathbb{P}_{x_i} L_n^{-1}\}_{n \in \mathbb{N}}$ for every $i \in I$,

$$\limsup_{i \in I} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{x_i}(L_n \in A) \geq -\liminf_{i \in I} \inf_{\tilde{\nu} \in O} H(\tilde{\nu}\,|\,\mu_{x_i}). \tag{3.21}$$

The estimates (3.19)–(3.21) lead to the contradiction $H(\nu\,|\,\mu_x) \geq H(\nu\,|\,\mu_x) + \varepsilon$.

It follows from (3.18) that, for every $n \geq n_0$,

$$\mathbb{P}_\Sigma(L_n \in A) \geq \int_{U_x} \mathbb{P}_y(L_n \in A)\,\Sigma(dy) \geq \exp\bigl(-n(H(\nu\,|\,\mu_x) + \varepsilon)\bigr)\,\Sigma(U_x).$$

Since $\Sigma(U_x) > 0$ by assumption and since $\varepsilon > 0$ is arbitrary, (3.17) follows.

(d) *Upper bound for general $A \in \mathcal{B}(\mathcal{M}_1(S))$:* Let C be the τ-closure of A. For every $x \in X$ define $\gamma_x = \inf_{\nu \in C} H(\nu\,|\,\mu_x)$. Given $\varepsilon > 0$ and $x \in X$, we want to show that there exist a measurable open neighbourhood U_x of x and $n_x \in \mathbb{N}$ such that $\mathbb{P}_y(L_n \in A) \leq \exp(-n\gamma_{\varepsilon,x})$ for all $n \geq n_x$ and $y \in U_x$, where $\gamma_{\varepsilon,x} \equiv \min\{1/\varepsilon, \gamma_x - \varepsilon\}$.

Assume that this claim were not true. Then there exists a filterbase $\{U_i\}_{i\in I} \subset \mathcal{X}$ of open neighbourhoods of x converging to x and $x_i \in U_i$ for every $i \in I$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{x_i}(L_n \in A) \ge -\gamma_{\varepsilon,x}. \tag{3.22}$$

Using the large deviations upper bound for $\{\mathbb{P}_{x_i} L_n^{-1}\}_{n\in\mathbb{N}}$ for every $i \in I$,

$$\liminf_{i\in I} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{x_i}(L_n \in A) \le -\limsup_{i\in I} \gamma_{x_i}. \tag{3.23}$$

By Theorem 1.16(a), we can use (1.8) and obtain that $\gamma_x \le \limsup_{i\in I} \gamma_{x_i}$, which via (3.22) and (3.23) gives the contradiction $\gamma_{\varepsilon,x} \ge \gamma_x$.

By the quasi-compactness of X, there exists a finite subset M of X such that the collection $\{U_x\}_{x\in M}$ covers X. It follows for $n \ge \max_{x\in M} n_x$ that

$$\mathbb{P}_\Sigma(L_n \in A) \le \sum_{x\in M} \int_{U_x} \mathbb{P}_y(L_n \in A)\, \Sigma(dy) \le \sum_{x\in M} \exp(-n\gamma_{\varepsilon,x}),$$

hence

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\Sigma(L_n \in A) \le -\min_{x\in M} \gamma_{\varepsilon,x} \le -\min\{1/\varepsilon, \inf_{\nu\in C} J_X(\nu) - \varepsilon\},$$

which implies the large deviations upper bound.

Compactness of the level sets: It suffices to prove that, for every $r \ge 0$, the level set $M_r \equiv \{\, (\nu, x) \in \mathcal{M}_1(S) \times X \mid H(\nu \,|\, \mu_x) \le r \,\}$ is quasi-compact, because $\bigcap_{\varepsilon>0} \pi_1(M_{r+\varepsilon}) = \{\, \nu \in \mathcal{M}_1(S) \mid J_X(\nu) \le r \,\}$, where $\pi_1\colon \mathcal{M}_1(S) \times X \to \mathcal{M}_1(S)$ is the projection.

According to [20, Chap. 5, Theorem 2], given a net $\{(\nu_i, x_i)\}_{i\in I}$ in M_r, we have to show that it has an accumulation point $(\tilde\nu, \tilde x)$ in M_r. Since X is quasi-compact, there exists $\tilde x \in X$ and a subordinated net $\{x_i\}_{i\in I'}$ converging to $\tilde x$. (Subordinated means that $I' \subset I$ and that for every $i \in I$ there exists $j \in I'$ with $i \preceq j$.) Since $X \ni x \mapsto \mu_x \in \mathcal{M}_1(S)$ is τ-continuous, $\{\mu_{x_i}\}_{i\in I'}$ converges to $\mu_{\tilde x}$ in the τ-topology. For every $j \in I'$ let C_j denote the τ-closure of $\{\nu_i\}_{i\in I', i\succeq j}$. With the subordinated net $\{\mu_{x_i}\}_{i\in I'}$ we are in the setting of Theorem 1.16(a), hence we obtain via (1.8) that, for every $j \in I'$,

$$\inf_{\nu\in C_j} H(\nu \,|\, \mu_{\tilde x}) \le \limsup_{i\in I'} \inf_{\nu\in C_j} H(\nu \,|\, \mu_{x_i}) \le r.$$

By [1, Lemma 2.1], the level set $K_\alpha \equiv \{\, \nu \in \mathcal{M}_1(S) \mid H(\nu \,|\, \mu_{\tilde x}) \le \alpha \,\}$ is τ-compact for every $\alpha \in [0, \infty)$. Thus, $C_j \cap K_r = \bigcap_{\varepsilon>0}(C_j \cap K_{r+\varepsilon}) \ne \varnothing$ by the finite intersection property of compact sets. By the same property, $C \equiv \bigcap_{j\in I'}(C_j \cap K_r) \ne \varnothing$, because $C_j \subset C_k$ for $j \succeq k$. Every $\tilde\nu \in C$ is an accumulation point of $\{\nu_i\}_{i\in I'}$. □

References

1. A. de Acosta, *On large deviations of empirical measures in the τ-topology*, J. Appl. Probab. **31A** (1994), 41–47, (special volume).
2. D. J. Aldous, *Exchangeability and related topics*, Ecole d'Eté de Probabilités de Saint-Flour XIII-1983 (P. L. Hennequin, ed.), Lecture Notes in Mathematics, vol. 1117, Springer-Verlag, Berlin, 1985, pp. 1–198.
3. J. R. Baxter and N. C. Jain, *A comparison principle for large deviations*, Proc. Amer. Math. Soc. **103** (1988), no. 4, 1235–1240.
4. E. Bolthausen, *Markov process large deviations in τ-topology*, Stochastic Process. Appl. **25** (1987), 95–108.

5. E. Bolthausen and U. Schmock, *On the maximum entropy principle for uniformly ergodic Markov chains*, Stochastic Process. Appl. **33** (1989), 1–27.
6. ______, *On self-attracting d-dimensional random walks*, Ann. Probab. **25** (1997), no. 2, 531–572.
7. T. Daras, *Large and moderate deviations for the empirical measure of an exchangeable sequence*, Statist. Probab. Lett. **36** (1997), 91–100.
8. A. Dembo and T. Zajic, *Uniform large and moderate deviations for functional empirical processes*, Stochastic Process. Appl. **67** (1997), no. 2, 195–211.
9. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, MA, 1993.
10. I. H. Dinwoodie and S. L. Zabell, *Large deviations for exchangeable random vectors*, Ann. Probab. **20** (1992), no. 3, 1147–1166.
11. ______, *Large deviations for sequences of mixtures*, Statistics and Probability: A Raghu Raj Bahadur Festschrift (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy, and B. L. S. Prakasa Rao, eds.), Wiley Eastern Limited, New Delhi, 1993, pp. 171–190.
12. L. E. Dubins and D. A. Freedman, *Exchangeable processes need not be mixtures of independent, identically distributed random variables*, Z. Wahrsch. verw. Gebiete **48** (1979), 115–132.
13. J. Dugundji, *Topology*, Allyn and Bacon, Boston, 1966.
14. P. Eichelsbacher, *Moderate deviations for functional U-processes*, preprint, 1997.
15. P. Eichelsbacher and U. Schmock, *Large and moderate deviations of products of empirical measures and U-empirical measures in strong topologies*, SFB-preprint 97-082, Universität Bielefeld, 1997.
16. ______, *Large deviations for products of empirical measures of dependent random sequences in strong topologies*, preliminary version, 1997.
17. P. Gänssler, *Compactness and sequential compactness in spaces of measures*, Z. Wahrscheinlichkeitstheorie verw. Geb. **17** (1971), 124–146.
18. H.-O. Georgii, *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin, 1988.
19. E. B. Hall and G. L. Wise, *Counterexamples in Probability and Real Analysis*, Oxford University Press, New York, 1993.
20. J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.
21. U. Schmock, *On the Maximum Entropy Principle for Markov Chains and Processes*, Ph.D. thesis, Technische Universität Berlin, Germany, 1990.

(P. Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany

E-mail address: peter@mathematik.uni-bielefeld.de

(U. Schmock) Departement Mathematik, ETH Zentrum, CH-8092 Zürich, Switzerland

E-mail address: schmock@math.ethz.ch