Markov Processes and Related Fields
Vol. 7, No. 3 (2001) 435–468

LARGE DEVIATIONS FOR PRODUCTS OF EMPIRICAL MEASURES OF DEPENDENT SEQUENCES

PETER EICHELSBACHER AND UWE SCHMOCK

Abstract. We prove large deviation principles (LDP) for m-fold products of empirical measures and for U-empirical measures, where the underlying sequence of random variables is a special Markov chain, an exchangeable sequence, a mixing sequence or an independent, but not identically distributed, sequence. The LDP can be formulated on a subset of all probability measures, endowed with a topology which is even finer than the usual τ-topology. The advantage of this topology is that the map $\nu \mapsto \int_{S^m} \varphi\, d\nu$ is continuous even for certain unbounded φ taking values in a Banach space. As a particular application we get large deviation results for U-statistics and V-statistics based on dependent sequences. Furthermore, we prove an LDP for products of empirical processes in a topology, which is finer than the projective limit τ-topology.

1. Introduction

Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in a Polish state space S. The corresponding empirical measures are defined by $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where δ_s denotes the probability measure concentrated at $s \in S$. Let $\mathcal{M}_1(S)$ denote the space of Borel probability measures on S.

Let us recall the definition of a large deviation principle (LDP). A sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ on a topological space $\mathcal{X}$ equipped with σ-field $\mathcal{B}$ is said to satisfy the LDP with scale $\{\varepsilon_n\}_{n\in\mathbb{N}}$ and good rate function $I\colon \mathcal{X} \to [0,\infty]$ if $\varepsilon_n \downarrow 0$ as $n \to \infty$, the level sets $\{x \in \mathcal{X} \mid I(x) \le r\}$ are compact for all $r \in [0,\infty)$, and the lower bound

$$\liminf_{n\to\infty} \varepsilon_n \log \mu_n(\Gamma) \ge -I(\mathrm{int}(\Gamma)),$$

and the upper bound

$$\limsup_{n\to\infty} \varepsilon_n \log \mu_n(\Gamma) \le -I(\mathrm{cl}(\Gamma))$$

hold for all $\Gamma \in \mathcal{B}$, where $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively, and $I(A) \equiv \inf_{x\in A} I(x)$ for $A \subset \mathcal{X}$. Normally, we choose $\varepsilon_n \equiv 1/n$. We say that a sequence of random variables satisfies the LDP if the sequence of measures induced by these variables satisfies the LDP.

The LDP for the sequence of empirical measures $\{L_n\}_{n\in\mathbb{N}}$ has been studied in several papers. In [10] and [11] *exchangeable sequences* $\{X_i\}_{i\in\mathbb{N}}$ are considered, and $\mathcal{M}_1(S)$ is endowed with the weak topology and the Borel σ-field associated with weak convergence in $\mathcal{M}_1(S)$. There are a lot of results when $\mathcal{M}_1(S)$ is endowed

1991 *Mathematics Subject Classification.* 60F10 (primary).

Key words and phrases. Large deviations, exponential approximations, contraction principle, weak topology, τ-topology, Markov chains, exchangeable sequences, mixing sequences, U-empirical measures, U-statistics, V-statistics.

with the τ-topology, i. e., the coarsest topology that makes the maps $\mathcal{M}_1(S) \ni \nu \mapsto \int_S f\,d\nu$ continuous for all f in the space $B(S,\mathbb{R})$ of bounded, real-valued, $\mathcal{S}$-measurable functions on S. Here $\mathcal{S}$ denotes the Borel σ-algebra on S. The LDP has been shown to hold for a large class of *Markov chains* (see [2], [8], [9], [14] and references therein) and for *stationary processes* satisfying strong *mixing conditions* (see [5] and references therein).

The aim of this paper is to discuss the LDP for m-fold products of the empirical measure given in (1.1) and for U-empirical measures defined in (1.9). Such LDPs can be viewed as extensions of Sanov's theorem. In [16] we discussed an LDP for U-empirical measures arising from a sequence $\{X_i\}_{i\in\mathbb{N}}$ of independent, identically distributed random variables. Within this framework we were able to work with a general measurable space $(S,\mathcal{S})$. Since we move away from the i. i. d. setting here, our methods require more structure of the state space S. Hence, unless otherwise stated, we assume that S is Polish with Borel σ-algebra $\mathcal{S}$.

For an integer $m \geq 2$ we consider the set $\mathcal{M}_1(S^m)$ of probability measures on the product space S^m, equipped with the product σ-algebra $\mathcal{S}^{\otimes m}$. If $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$ are endowed with their weak topologies, then the LDP holds for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$, defined by

$$L_n^{\otimes m} = \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} \delta_{(X_{i_1},\dots,X_{i_m})}, \qquad n \in \mathbb{N}, \tag{1.1}$$

whenever it holds for $\{L_n\}_{n\in\mathbb{N}}$. We only have to use the continuity of the map $\nu \mapsto \nu^{\otimes m}$ with respect to the weak topologies and apply the contraction principle [8, Theorem 4.2.1]. The τ-topology on $\mathcal{M}_1(S^m)$ is defined to be the one induced by $B(S^m,\mathbb{R})$, the space of all bounded, real-valued, $\mathcal{S}^{\otimes m}$-measurable functions on S^m. Taking product measures can be a discontinuous operation with respect to the τ-topologies (see [8, Exercise 7.3.18]). Moreover, the following example (see also [16, Example 1.26]) illustrates that we cannot usually expect an LDP to hold for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ arising from a sequence of dependent $\{X_i\}_{i\in\mathbb{N}}$.

Example 1.2. Let the circle $S = \mathbb{R}/\mathbb{Z}$ be equipped with the Borel σ-algebra $\mathcal{S}$ and let μ denote the Lebesgue–Borel measure on $(S,\mathcal{S})$. For every $x \in \mathbb{R}$ define the shift modulo 1 (or rotation) θ_x on S by $\theta_x(y) = x + y \bmod 1$ for all $y \in S$. Using these, define

$$S \ni \omega \mapsto L_n(\omega) = \frac{1}{2^n} \sum_{i=0}^{2^n-1} \delta_{\theta_{i2^{-n}}(\omega)} \in \mathcal{M}_1(S), \quad n \in \mathbb{N}_0.$$

Note that there is a heavy dependence between the $\theta_{i2^{-n}}(\omega)$ for $i \in \{0,\dots,2^n-1\}$. Since S is compact, it is easy to verify that $\{L_n(\omega)\}_{n\in\mathbb{N}_0}$ and $\{L_n(\omega)\otimes L_n(\omega)\}_{n\in\mathbb{N}_0}$ converge weakly to μ and $\mu\otimes\mu$, respectively, for every $\omega \in S$. Moreover, for every $\varphi \in L_1(\mu,\mathbb{R})$ it holds that

$$\mu\Big(\lim_{n\to\infty} \int_S \varphi\,dL_n = \int_S \varphi\,d\mu\Big) = 1 \tag{1.3}$$

(for a proof see [16, Example 1.26]). To show that the product measures $\{L_n \otimes L_n\}_{n\in\mathbb{N}_0}$ can go astray, we consider the $\mathcal{S}\otimes\mathcal{S}$-measurable set $A \equiv \{\,(x,y) \in S^2 \mid x - y \in \mathbb{Q}\,\}$. By Fubini's theorem, $(\mu\otimes\mu)(A) = 0$. On the other hand, the support of $L_n(\omega)\otimes L_n(\omega)$, which is $\{\,(\theta_{i2^{-n}}(\omega),\theta_{j2^{-n}}(\omega)) \mid i,j \in \{0,1,\dots,2^n-1\}\,\}$, is contained in A for every $n \in \mathbb{N}_0$ and $\omega \in S$. Therefore, the analogue of (1.3) for product measures does not even hold for the $\mathcal{S} \otimes \mathcal{S}$-measurable indicator function $\varphi \equiv 1_A$. Note that there does *not* exist a scale $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ with $\varepsilon_n \downarrow 0$ such that the random measures $\{L_n\}_{n\in\mathbb{N}_0}$ satisfy a large deviation upper bound of the form

$$\limsup_{n\to\infty} \varepsilon_n \log \mu(L_n \in C) \leq -\inf_{\nu\in C} I(\nu)$$

for all τ-closed measurable $C \subset \mathcal{M}_1(S)$, where $I\colon \mathcal{M}_1(S) \to [0,\infty]$ with $I(\mu) = 0$ and $I(\nu) = \infty$ for $\nu \neq \mu$ is the rate function which governs the large deviations of $\{L_n\}_{n\in\mathbb{N}_0}$ with respect to the weak topology on every scale $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ with $\varepsilon_n \downarrow 0$. To substantiate this claim, in [16] we consider the set $C \equiv \{\,\nu \in \mathcal{M}_1(S) \mid \nu(A) \geq \mu(A) + 1/2\,\}$, where we construct the set $A \in \mathcal{S}$ as follows: Choose a subsequence $\{\varepsilon_{n_k}\}_{k\in\mathbb{N}}$ such that $\sum_{k\in\mathbb{N}} \varepsilon_{n_k} \leq 1/2$ and define $A = \bigcup_{k\in\mathbb{N}} A_k$, where

$$A_k = \bigcup_{l=0}^{2^{n_k}-1} [l2^{-n_k}, (l+\varepsilon_{n_k})2^{-n_k}).$$

Then $\mu(A_k) = \varepsilon_{n_k}$ and $\mu(A) \leq \sum_{k\in\mathbb{N}} \varepsilon_{n_k} \leq 1/2$ as well as $L_{n_k}(A) \geq L_{n_k}(A_k) = 1$ on A_k for every $k \in \mathbb{N}$. Hence, as $k \to \infty$,

$$\varepsilon_{n_k} \log \mu(\{L_{n_k}(A) \geq \mu(A) + 1/2\}) \geq \varepsilon_{n_k} \log \mu(A_k) = \varepsilon_{n_k} \log \varepsilon_{n_k} \to 0.$$

A slight modification of [1, Example 4.1] shows that there are *ergodic, stationary* processes $\{X_i\}_{i\in\mathbb{N}}$ and bounded maps $\varphi\colon S^m \to \mathbb{R}$ such that $\int_{S^m} \varphi\, dL_n^{\otimes m}$ does not satisfy the *strong law of large numbers*. This is another reason why we cannot generally expect an LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ arising from dependent $\{X_i\}_{i\in\mathbb{N}}$. So the question is: which sequences $\{X_i\}_{i\in\mathbb{N}}$ of dependent random variables allow us to transfer the LDP for $\{L_n\}_{n\in\mathbb{N}}$ to the LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ with respect to the τ-topology?

We want to use even finer topologies than the τ-topology and for this, we need additional notation. Let $(E, \|\cdot\|_E)$ be a real separable Banach space with Borel σ-algebra $\mathcal{E}$. We always exclude the case $E = \{0\}$ to get a Hausdorff topology. Let Φ be a set of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable functions $\varphi\colon S^m \to E$ containing the set $B(S^m, E)$ of bounded measurable functions. Define the Φ-restricted set of probability measures on S^m by

$$\mathcal{M}_1^\Phi(S^m) = \left\{ \nu \in \mathcal{M}_1(S^m) \;\middle|\; \int_{S^m} \|\varphi\|_E \, d\nu < \infty \text{ for every } \varphi \in \Phi \right\}.$$

Then, the Bochner integral $\int_{S^m} \varphi\, d\nu$ is defined for every $\varphi \in \Phi$ and $\nu \in \mathcal{M}_1^\Phi(S^m)$. Let $\tau_1^\Phi(E)$ denote the coarsest topology on $\mathcal{M}_1^\Phi(S^m)$ such that the map $\mathcal{M}_1^\Phi(S^m) \ni \nu \mapsto \int_{S^m} \varphi\, d\nu$ is continuous for every $\varphi \in \Phi$. If $\Phi = B(S^m, E)$, then $\mathcal{M}_1^\Phi(S^m) = \mathcal{M}_1(S^m)$ and we write $\tau_1(E)$ instead of $\tau_1^\Phi(E)$. If, in addition, $E = \mathbb{R}$, then $\tau_1^\Phi(E)$ coincides with the usual τ-topology introduced above. The σ-algebra on $\mathcal{M}_1(S^m)$ is defined to be the smallest one, such that the set $\mathcal{M}_1^\Phi(S^m)$ and all the maps $\mathcal{M}_1(S^m) \ni \nu \mapsto \int_{S^m} f\, d\nu$ with $f \in B(S^m, E)$ are measurable. In this topological setting we will prove an LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ for certain processes $\{X_i\}_{i\in\mathbb{N}}$ already obeying an LDP for $\{L_n\}_{n\in\mathbb{N}}$ in the weak topological setting, if additional exponential moment conditions are fulfilled.

We consider the following conditions for a sequence $\{R_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S^m))$ and for the class Φ. Often, R_n will be the distribution of the U-empirical measure L_n^m defined in (1.9), the product measure $L_n^{\otimes m}$, or a modification thereof.

Condition 1.4. *There exist constants $\beta, M \in [1,\infty)$ and $n_0 \in \mathbb{N}$ as well as a reference measure $\mu \in \mathcal{M}_1(S^m)$ such that the inequality*

$$\sup_{n\ge n_0}\biggl(\int_{\mathcal{M}_1(S^m)} \exp\Bigl(n\int_{S^m}\varphi\,d\nu\Bigr) R_n(d\nu)\biggr)^{1/n} \le M\int_{S^m}\exp(\beta\varphi)\,d\mu \tag{1.5}$$

holds for all bounded measurable $\varphi\colon S^m \to [0,\infty)$.

Remark 1.6. If Condition 1.4 holds, then, by the monotone convergence theorem, (1.5) also holds for every unbounded measurable $\varphi\colon S^m \to [0,\infty)$ with the same constants and reference measure, but the right-hand side can be equal to infinity.

Condition 1.7 (Strong Cramér Condition)**.** *For every $\varphi \in \Phi$ and every $\alpha > 0$,*

$$\int_{S^m}\exp(\alpha\|\varphi\|_E)\,d\mu < \infty,$$

where μ is the reference measure of Condition 1.4*.*

Condition 1.8. *For every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that*

$$\int_{S^m}\exp(\alpha_\varphi\|\varphi\circ\pi_\tau\|_E)\,d\mu < \infty$$

for every map $\tau\colon \{1,\dots,m\} \to \{1,\dots,m\}$, where $\pi_\tau\colon S^m \to S^m$ is defined by $\pi_\tau(s) = (s_{\tau(1)},\dots,s_{\tau(m)})$ for every $s = (s_1,\dots,s_m) \in S^m$ and μ is the reference measure of Condition 1.4*.*

In the case that $\{X_i\}_{i\in\mathbb{N}}$ are dependent random variables, e. g. a Markov chain or a stationary mixing sequence, we need additional exponential moment assumptions, see (3.13) and (3.41). Conditions 1.7, 1.8 and the additional exponential moment assumptions are always satisfied in the case of $\Phi = B(S^m, E)$ and thus in the $\tau_1(E)$-topological setting. We will see that the above conditions are handy to prove an LDP in the $\tau_1^\Phi(E)$-topology for the m-fold products $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$. In order to apply our results to U-statistics, we will prove an LDP for $L_n^m\colon \Omega \to \mathcal{M}_1(S^m)$ with $n \ge m$, which are defined by

$$L_n^m = \frac{1}{n_{(m)}}\sum_{(i_1,\dots,i_m)\in I_{m,n}} \delta_{(X_{i_1},\dots,X_{i_m})}, \tag{1.9}$$

where $n_{(m)} \equiv \prod_{k=0}^{m-1}(n-k)$ and $I_{m,n} \subset \{1,\dots,n\}^m$ contains all m-tuples with pairwise different components. These L_n^m are called *U-empirical measures* of order m. The LDP for these measures requires Conditions 1.4 and 1.7; Condition 1.8 is not needed. Since $L_n^{\otimes m}$ and L_n^m take values in $\mathcal{M}_1^\Phi(S^m)$, these mappings are measurable with respect to the σ-algebras we introduced.

Condition 1.4 presents the main tool to get the LDP for products we are interested in. If Condition 1.4 is fulfilled, and if $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP with respect to the weak topology on $\mathcal{M}_1(S^m)$, we can infer from [9, Lemma 3.2.19 and Theorem 3.2.21], that $\{R_n\}_{n\in\mathbb{N}}$ actually satisfies an LDP in the τ-topology on $\mathcal{M}_1(S^m)$. We will prove in Section 2, that, under Conditions 1.4 and 1.7, this approach can be generalized to the $\tau_1^\Phi(E)$-topological setting.

Condition 1.4 describes the "amount of dependency" of the underlying process under which an LDP for the laws of the empirical measures is preserved under products in the $\tau_1(E)$-topology. We observe that the conditions which guarantee the LDP for $\{L_n\}_{n\in\mathbb{N}}$ are not, in general, sufficient for the LDP for the products. We will analyze Markovian, exchangeable, strongly mixing and independent, but not identically distributed, sequences. In all these cases we want to establish the crucial estimate (1.5). In some cases we are not able to check this estimate directly for the law of L_n^m. Having constructed a suitable modification of L_n^m and proving (1.5) for it, the remaining part of the proof will be to verify that these modifications have a large deviation behaviour like $\{L_n^m\}_{n\geq m}$.

In Section 2, we introduce our main tool: If $\{R_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$ satisfies an LDP in the weak topology with a rate function I, and if Condition 1.4 ($m = 1$) and Condition 1.7 ($m = 1$) hold for the class Φ introduced above, then the LDP holds in the $\tau_1^\Phi(E)$-topology with the same rate function. Since with S the product S^m is also Polish, we can apply this tool for S^m with $m \geq 2$, too. Moreover, Section 2 contains the concept of exponential equivalence in the $\tau_1^\Phi(E)$-topology needed to transfer the LDP for the above-mentioned modifications to the LDP for $\{L_n^m\}_{n\geq m}$ and $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ respectively. In Section 3, (1.5) is verified for different dependent or independent but not identically distributed processes $\{X_i\}_{i\in\mathbb{N}}$. Finally, in Section 4, we derive LDPs for U-statistics and V-statistics from the results of Sections 2 and 3, as well as an LDP for products of empirical processes.

2. Transferring Large Deviation Principles to the $\tau_1^\Phi(E)$-Topology and Exponential Equivalence

There are some non-trivial problems in transferring Lemma 3.2.19 and Theorem 3.2.21 in [9] to the $\tau_1^\Phi(E)$-topology. They are treated by using some technical results of [16], constructing an exponentially good approximation and using [8, Theorem 4.2.23]. The key step is an application of Lusin's theorem, for which we need the topological structure of the state space S. Let Φ be a fixed set of $\mathcal{S}$-$\mathcal{E}$-measurable functions $\varphi\colon S \to E$ containing $B(S,E)$ and define $\mathcal{M}_1^\Phi(S)$ and the $\tau_1^\Phi(E)$-topology as set out in the introduction. We then obtain the following theorem, on which our main subsequent results are based.

Theorem 2.1. *Assume that $\{R_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$ satisfies Condition* 1.4 *for $m = 1$ and that the inner measure of $\mathcal{M}_1^\Phi(S)$ with respect to R_n is* 1 *for each $n \in \mathbb{N}$. Furthermore, assume that Φ satisfies the Strong Cramér Condition* 1.7 *with $m = 1$ and $\{R_n\}_{n\in\mathbb{N}}$ satisfies the LDP with rate function I, where $\mathcal{M}_1(S)$ is endowed with the weak topology. Then:*

(a) *For every measurable $B \subset \mathcal{M}_1(S)$*

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n(B) \geq -I(\mathrm{int}_{\tau_1^\Phi(E)}(B)),$$

where $\mathrm{int}_{\tau_1^\Phi(E)}(B)$ denotes the interior of the set $B \cap \mathcal{M}_1^\Phi(S)$ with respect to the $\tau_1^\Phi(E)$-topology.

(b) *For every measurable $B \subset \mathcal{M}_1(S)$*

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(B) \leq -I(\mathrm{cl}_{\tau_1^\Phi(E)}(B)),$$

where $\mathrm{cl}_{\tau_1^\Phi(E)}(B)$ denotes the $\tau_1^\Phi(E)$-closure of the set $B \cap \mathcal{M}_1^\Phi(S)$.

(c) *$K(I,r) \equiv \{\,\nu \in \mathcal{M}_1(S) \mid I(\nu) \leq r\,\} \subset \mathcal{M}_1^\Phi(S)$ and $K(I,r)$ is $\tau_1^\Phi(E)$-compact and sequentially $\tau_1^\Phi(E)$-compact for every $r \in [0,\infty)$.*

Remark 2.2. For convenience, Theorem 2.1 is stated and proved for $m = 1$. If S is a Polish space, then $S' \equiv S^m$ with $m \in \mathbb{N}$ is also Polish and Theorem 2.1 can be rephrased accordingly.

Remark 2.3. Even in the case $m = 1$, Theorem 2.1 already improves some existing LDP results in the τ-topology for the laws of the empirical measures $\{L_n\}_{n\in\mathbb{N}}$ for some dependent sequences. The reason for this is that one can verify Condition 1.4 in several cases. For example, we get the LDP for the laws of $\{L_n\}_{n\in\mathbb{N}}$, when the $\{X_i\}_{i\in\mathbb{N}}$ are Markovian and satisfy [9, Condition (U) in Section 4.1], compare also with [9, Exercise 4.1.53]. If the sequence $\{X_i\}_{i\in\mathbb{N}}$ is stationary and satisfies Assumptions (H-1) and (H-2) in [8, Section 6.4.2], we can transfer the LDP to the $\tau_1^\Phi(E)$-topology, too (see [8, Lemma 6.4.18]).

Proof of Theorem 2.1(c). Using Condition 1.4, by applying [9, Lemma 3.2.7 and Lemma 3.2.19] we get that

$$H(\nu\,|\,\mu) \le \beta(I(\nu) + \log(2M)), \quad \nu \in \mathcal{M}_1(S), \tag{2.4}$$

with β, M and μ as in (1.5). Here, H denotes the relative entropy

$$H(\nu\,|\,\mu) \equiv \begin{cases} \int_S f \log f \, d\mu & \text{if } \nu \ll \mu \text{ and } f = \frac{d\nu}{d\mu}, \\ \infty & \text{otherwise.} \end{cases} \tag{2.5}$$

Define $K(H, r) = \{\, \nu \in \mathcal{M}_1(S) \mid H(\nu\,|\,\mu) \le r \,\}$ for every $r \in [0, \infty)$. Since Condition 1.7 holds, $K(H, r)$ has all the properties claimed for $K(I, r)$, see [16, Lemma 2.1]. Since I is lower semi-continuous in the weak topology, $K(I, r)$ is $\tau_1^\Phi(E)$-closed. Since $K(I, r) \subset K(H, \beta(r + \log 2M))$ by (2.4), Theorem 2.1(c) follows. □

Before we prove (a) and (b) of Theorem 2.1, we will prove an additional lemma. We need some more notation. Let $\mathcal{F}$ denote the family of all finite, nonempty subsets of Φ. For every $F \in \mathcal{F}$ we consider the real separable Banach space $(E^F, \|\cdot\|_{E^F})$ with $\|e\|_{E^F} \equiv \sum_{\varphi\in F} \|e_\varphi\|_E$ for $e = (e_\varphi)_{\varphi\in F} \in E^F$ and define

$$\Pi_F \colon \mathcal{M}_1^\Phi(S) \to E^F \quad \text{by} \quad \Pi_F(\nu) = \left(\int_S \varphi \, d\nu \right)_{\varphi\in F}.$$

Lemma 2.6. *Assume the general hypotheses of Theorem* 2.1 *and consider* $F \in \mathcal{F}$. *The rate function* $I_F(x) \equiv \inf\{\, I(\nu) \mid \nu \in \mathcal{M}_1^\Phi(S), \Pi_F(\nu) = x \}$ *for* $x \in E^F$ *is then good and, for every Borel set* A *of* E^F,

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n(\{\nu \in \mathcal{M}_1^\Phi(S) \mid \Pi_F(\nu) \in A\}) \ge -I_F(A^\circ)$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(\{\nu \in \mathcal{M}_1^\Phi(S) \mid \Pi_F(\nu) \in A\}) \le -I_F(\bar{A}),$$

where A° *denotes the interior and* $\bar{A}$ *the closure of* A.

Proof of Lemma 2.6. We want to approximate Π_F by continuous maps and apply [8, Theorem 4.2.23], which deals with LDPs and exponentially good approximations. Note that, by Theorem 2.1(c), the rate function I has weakly compact level sets.

Pick $\alpha > 0$. By Condition 1.7, the dominated convergence theorem and [18, Lemma V-2-4], there exists, for every $\varphi \in F$, a measurable, finitely-valued function

$\varphi_\alpha\colon S \to E$ such that

$$\int_S \exp\bigl(2\alpha\beta|F|\,\|\varphi-\varphi_\alpha\|_E\bigr)\,d\mu \le 2, \tag{2.7}$$

where $\beta \in [1,\infty)$ and $\mu \in \mathcal{M}_1(S)$ are as in Condition 1.4. If $\{e_1,\dots,e_m\} \subset E$ denotes the set of possible values of $\{\varphi_\alpha\}_{\varphi\in F}$ and $B_{\varphi,i} \equiv \{\, s\in S \mid \varphi_\alpha(s)=e_i \,\} \in \mathcal{E}$ for $i=1,\dots,m$, then $\varphi_\alpha = \sum_{i=1}^m 1_{B_{\varphi,i}} e_i$. Define $L=\max\{\|e_1\|_E,\dots,\|e_m\|_E\}$.

According to Lusin's theorem and Tietze's extension theorem (see e. g. [4, Theorem 7.3.4] and [13, Chap. VII, Theorem 5.1], respectively), there exists, for every $\varphi\in F$ and $i\in\{1,\dots,m\}$, a continuous function $\varphi_{\alpha,i}\colon S\to[0,1]$ such that

$$\mu\bigl(\{\, s\in S \bigm| \varphi_{\alpha,i}(s) \ne 1_{B_{\varphi,i}}(s) \,\}\bigr) \le \frac{1}{m}\exp(-2\alpha\beta|F|Lm).$$

Define $\varphi_{\alpha,\mathrm{c}} = \sum_{i=1}^m \varphi_{\alpha,i} e_i$ for every $\varphi\in F$. Then the map $\Pi_{F_{\alpha,\mathrm{c}}}\colon \mathcal{M}_1(S)\to E^F$ with $F_{\alpha,\mathrm{c}} \equiv \{\varphi_{\alpha,\mathrm{c}}\}_{\varphi\in F}$ is continuous with respect to the weak topology on $\mathcal{M}_1(S)$. Since $\mu(\{\, s\in S \mid \varphi_\alpha(s)\ne\varphi_{\alpha,\mathrm{c}}(s)\,\}) \le \exp(-2\alpha\beta|F|Lm)$ and $\|\varphi_\alpha-\varphi_{\alpha,\mathrm{c}}\|_E \le Lm$ on S, it follows that

$$\int_S \exp\bigl(2\alpha\beta|F|\,\|\varphi_\alpha-\varphi_{\alpha,\mathrm{c}}\|_E\bigr)\,d\mu \le 2. \tag{2.8}$$

We want to verify that $\{R_n\Pi_{F_{\alpha,\mathrm{c}}}^{-1}\}_{\alpha>0,n\in\mathbb{N}}$ approximates $\{R_n\Pi_F^{-1}\}_{n\in\mathbb{N}}$ exponentially good as $\alpha\to\infty$. Define $g_\alpha = \sum_{\varphi\in F}\|\varphi-\varphi_{\alpha,\mathrm{c}}\|_E$. Applying the exponential Chebyshev inequality and Condition 1.4, we obtain, for every $n\ge n_0$ and $\delta>0$,

$$\begin{aligned}
\Bigl(R_n\bigl(\{\,\nu\in\mathcal{M}_1^\Phi(S) \bigm| \|\Pi_F(\nu)-\Pi_{F_{\alpha,\mathrm{c}}}(\nu)\|_{E^F} > \delta\,\}\bigr)\Bigr)^{1/n} \\
&\le e^{-\alpha\delta}\biggl(\int_{\mathcal{M}_1^\Phi(S)} \exp\biggl(\alpha n\int_S g_\alpha\,d\nu\biggr)R_n(d\nu)\biggr)^{1/n} \\
&\le Me^{-\alpha\delta}\int_S \exp(\alpha\beta g_\alpha)\,d\mu.
\end{aligned} \tag{2.9}$$

By Hölder's inequality, (2.7) and (2.8),

$$\begin{aligned}
\int_S \exp(\alpha\beta g_\alpha)\,d\mu \le \prod_{\varphi\in F}\biggl(&\int_S \exp\bigl(2\alpha\beta|F|\,\|\varphi-\varphi_\alpha\|_E\bigr)\,d\mu \\
&\times\int_S \exp\bigl(2\alpha\beta|F|\,\|\varphi_\alpha-\varphi_{\alpha,\mathrm{c}}\|_E\bigr)\,d\mu\biggr)^{1/(2|F|)} \le 2.
\end{aligned} \tag{2.10}$$

By combining (2.9) and (2.10) we get, for every $\delta>0$,

$$\lim_{\alpha\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log R_n\bigl(\{\,\nu\in\mathcal{M}_1^\Phi(S) \bigm| \|\Pi_F(\nu)-\Pi_{F_{\alpha,\mathrm{c}}}(\nu)\|_{E^F}>\delta\}\bigr) = -\infty.$$

Define $K(I,r)$ and $K(H,r)$ as in Theorem 2.1(c) and its proof. Since $K(I,r)\subset K(H,\beta(r+\log 2M))$ by (2.4), it remains to show that, for every $r\in[0,\infty)$,

$$\limsup_{\alpha\to\infty}\sup_{\nu\in K(H,r)}\bigl\|\Pi_F(\nu)-\Pi_{F_{\alpha,\mathrm{c}}}(\nu)\bigr\|_{E^F} = 0. \tag{2.11}$$

Fix $r\in[0,\infty)$. Define $r' = r+1/e$ and $f_\nu = d\nu/d\mu$ for every $\nu\in K(H,r)$. Since $x\log x \ge -1/e$ for all $x\in[0,\infty)$, we have $\int_D f_\nu\log f_\nu\,d\mu \le H(\nu\,|\,\mu)+1/e \le r'$ for every $D\in\mathcal{S}$ and $\nu\in K(H,r)$. Fix $\gamma>1$ such that $r'=\gamma\log\gamma$. Choose $\varepsilon>0$

arbitrarily. Define $C_\alpha = \{\, s \in S \mid g_\alpha(s) > \varepsilon/\gamma \,\}$ and $D_\nu = \{\, s \in S \mid f_\nu(s) > \gamma \,\}$ for all $\alpha > 0$ and $\nu \in K(H, r)$. Note that

$$\bigl\|\Pi_F(\nu) - \Pi_{F_{\alpha,\mathrm{c}}}(\nu)\bigr\|_{E^F} \le \int_S g_\alpha \, d\nu = \int_S f_\nu g_\alpha \, d\mu, \quad \nu \in K(H, r).$$

We use $S = (C_\alpha^{\mathrm{c}} \cap D_\nu^{\mathrm{c}}) \cup (C_\alpha^{\mathrm{c}} \cap D_\nu) \cup C_\alpha$ to decompose the integral. Obviously, $\int_{C_\alpha^{\mathrm{c}} \cap D_\nu^{\mathrm{c}}} f_\nu g_\alpha \, d\mu \le \varepsilon$ for all $\alpha > 0$ and $\nu \in K(H, r)$. In addition, by the choice of γ,

$$\int_{C_\alpha^{\mathrm{c}} \cap D_\nu} f_\nu g_\alpha \, d\mu \le \frac{\varepsilon}{\gamma \log \gamma} \int_{D_\nu} f_\nu \log f_\nu \, d\mu \le \varepsilon.$$

For the last term of the decomposed integral we use the well-known estimate $xy \le e^x + y \log y$ for all $x, y \ge 0$, see e. g. [16, (2.2)], to obtain

$$\int_{C_\alpha} f_\nu g_\alpha \, d\mu \le \int_{C_\alpha} \exp\Bigl(\frac{\alpha\beta}{2} g_\alpha\Bigr) \, d\mu + \frac{2}{\alpha\beta} \int_{C_\alpha} f_\nu \Bigl(\log f_\nu - \log \frac{\alpha\beta}{2}\Bigr) \, d\mu.$$

The last term is bounded by $2r'/(\alpha\beta)$ for $\alpha \ge 2$ and tends to zero as $\alpha \to \infty$. By the Cauchy–Schwarz inequality and (2.10) we get

$$\int_{C_\alpha} \exp\Bigl(\frac{\alpha\beta}{2} g_\alpha\Bigr) \, d\mu \le \sqrt{2\mu(C_\alpha)},$$

and by the exponential Chebyshev inequality, and again (2.10),

$$\mu(C_\alpha) \le e^{-\alpha\beta\varepsilon/\gamma} \int_S \exp(\alpha\beta g_\alpha) \, d\mu \le 2e^{-\alpha\beta\varepsilon/\gamma} \to 0 \quad \text{as } \alpha \to \infty.$$

This proves (2.11), hence Lemma 2.6 follows from [8, Theorem 4.2.23]. □

Proof of Theorem 2.1(a), (b). (a) It suffices to treat the case $I(\mathrm{int}_{\tau_1^\Phi(E)}(B)) < \infty$. By definition of the $\tau_1^\Phi(E)$-topology, there exist for $\nu \in \mathrm{int}_{\tau_1^\Phi(E)}(B)$ an $\varepsilon > 0$ and an $F \in \mathcal{F}$ such that the $\tau_1^\Phi(E)$-open set $C \equiv \bigl\{\tilde\nu \in \mathcal{M}_1^\Phi(S) \bigm| \|\Pi_F(\tilde\nu) - \Pi_F(\nu)\|_{E^F} < \varepsilon\bigr\}$ is contained in $\mathrm{int}_{\tau_1^\Phi(E)}(B)$. With the open set $A \equiv \bigl\{x \in E^F \bigm| \|x - \Pi_F(\nu)\|_{E^F} < \varepsilon\bigr\}$ we get $\nu \in C = \{\, \tilde\nu \in \mathcal{M}_1^\Phi(S) \mid \Pi_F(\tilde\nu) \in A \,\} \subset B$ and by Lemma 2.6

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n(B) \ge \liminf_{n\to\infty} \frac{1}{n} \log R_n(C) \ge -I_F(A) \ge -I(\nu).$$

Taking the supremum over $\nu \in \mathrm{int}_{\tau_1^\Phi(E)}(B)$, the lower bound of part (a) follows.

(b) We need only consider the case $q \equiv I(\mathrm{cl}_{\tau_1^\Phi(E)}(B)) > 0$. Choose $r \in (0, q)$. By Theorem 2.1(c), the set $K(I, r)$ is contained in $\mathcal{M}_1^\Phi(S)$. Since $\mathrm{cl}_{\tau_1^\Phi(E)}(B) \cap K(I, r) = \varnothing$, there exist, for every $\nu \in K(I, r)$, an $F_\nu \in \mathcal{F}$ and an open neighbourhood $U_\nu \subset E^{F_\nu}$ of $\Pi_{F_\nu}(\nu)$ such that $\mathrm{cl}_{\tau_1^\Phi(E)}(B) \cap \Pi_{F_\nu}^{-1}(U_\nu) = \varnothing$. Since $K(I, r)$ is $\tau_1^\Phi(E)$-compact, there exists a finite subset N of $K(I, r)$ such that $\bigcup_{\nu \in N} \Pi_{F_\nu}^{-1}(U_\nu)$ covers $K(I, r)$. Define $F = \bigcup_{\nu \in N} F_\nu$. Note that $F \in \mathcal{F}$. For every $\nu \in N$ define $U'_\nu = \Pi_{F,F_\nu}^{-1}(U_\nu)$, where for $F' \subset F$ with $F' \ne \varnothing$ the map $\Pi_{F,F'} \colon E^F \to E^{F'}$ denotes the canonical projection. Note that $U'_\nu \subset E^F$ is open and $\Pi_F^{-1}(U'_\nu) = \Pi_{F_\nu}^{-1}(U_\nu)$. Define $U = \bigcup_{\nu \in N} U'_\nu$. Then $\Pi_F^{-1}(U) = \bigcup_{\nu \in N} \Pi_{F_\nu}^{-1}(U_\nu)$, hence $\Pi_F^{-1}(U)$ covers $K(I, r)$ and is disjoint from $\mathrm{cl}_{\tau_1^\Phi(E)}(B)$. Define $\varepsilon = \mathrm{dist}(\Pi_F(K(I, r)), U^{\mathrm{c}})$. Since $\Pi_F(K(I, r))$ is a compact subset of the open set U, it follows that $\varepsilon > 0$ and that

$$A_\varepsilon \equiv \{\, x \in E^F \mid \mathrm{dist}(x, \Pi_F(K(I, r))) < \varepsilon \,\}$$

is an open set contained in U. Therefore

$$\mathrm{cl}_{\tau_1^\Phi(E)}(B) \subset \{\, \nu \in \mathcal{M}_1^\Phi(S) \mid \Pi_F(\nu) \in E^F \setminus A_\varepsilon \,\} \subset K(I,r)^{\mathrm{c}}. \tag{2.12}$$

By the upper bound in Lemma 2.6 it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(B) \le -I_F(E^F \setminus A_\varepsilon) \le -I(K(I,r)^{\mathrm{c}}) \le -r.$$

Since $r \in (0,q)$ was arbitrary, the upper bound follows. □

Remark 2.13. If Condition 1.7 holds, we get from [16, Lemma 2.1(b)] and from (2.4) that $K(I,\infty) \equiv \bigcup_{r>0} K(I,r) \subset \mathcal{M}_1^\Phi(S)$, where $K(I,r)$ is defined as in Theorem 2.1(c). Since $I(\nu) = \infty$ for all $\nu \in \mathcal{M}_1^\Phi(S) \setminus K(I,\infty)$, it follows for the rate function defined in the statement of Lemma 2.6

$$I_F(x) = \inf\{\, I(\nu) \mid \nu \in K(I,\infty), \Pi_F(\nu) = x \,\}, \quad x \in E^F,\ F \in \mathcal{F}.$$

To transfer the LDP to the $\tau_1^\Phi(E)$-topology, it will sometimes be necessary to drop some summands of $L_n^{\otimes m}$ or L_n^m in order to be able to verify Condition 1.4. Thus we will establish the LDP in the $\tau_1^\Phi(E)$-topology for such reduced empirical measures. Next we consider the question of if—and how—the LDP for the laws of $L_n^{\otimes m}$ and L_n^m, respectively, can be deduced from the LDP for the laws of the reduced empirical measures. Consider two sequences $\{S_n\}_{n\in\mathbb{N}}$ and $\{S'_n\}_{n\in\mathbb{N}}$ of $\mathcal{M}_1(S)$-valued random variables with distributions $\{R_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$ and $\{R'_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$, respectively, on a joint probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Lemma 2.14. *Assume that $\{R_n\}_{n\in\mathbb{N}}$ satisfies an LDP in the $\tau_1^\Phi(E)$-topology on $\mathcal{M}_1^\Phi(S)$ and that, for each $n \in \mathbb{N}$, the inner measure of $\mathcal{M}_1^\Phi(S)$ with respect to R_n and R'_n is 1. Assume that, for every $\varepsilon > 0$ and every $\varphi \in \Phi$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left\| \int_S \varphi\, dS_n - \int_S \varphi\, dS'_n \right\|_E \ge \varepsilon \right) = -\infty. \tag{2.15}$$

Then the same LDP (with the same good rate function) holds for $\{R'_n\}_{n\in\mathbb{N}}$.

Proof. For each $F \in \mathcal{F}$ define the pseudo-metric

$$d_F(\mu,\nu) = \max_{\varphi\in F} \left\| \int_S \varphi\, d\mu - \int_S \varphi\, d\nu \right\|_E, \qquad \mu,\nu \in \mathcal{M}_1^\Phi(S).$$

For $\mu \ne \nu$ there is an $F \in \mathcal{F}$ such that $d_F(\mu,\nu) \ne 0$, because Φ contains $B(S,E)$. Hence, the family $\mathcal{D} \equiv \{d_F\}_{F\in\mathcal{F}}$ is separating. The topology having as a sub-basis the family of balls $\{B(\nu, d_F, \varepsilon) \mid \nu \in \mathcal{M}_1^\Phi(S),\ d_F \in \mathcal{D},\ \varepsilon > 0\}$, where $B(\nu,d_F,\varepsilon) \equiv \{\, \mu \in \mathcal{M}_1^\Phi(S) \mid d_F(\mu,\nu) < \varepsilon \,\}$, is the $\tau_1^\Phi(E)$-topology. Therefore $(\mathcal{M}_1^\Phi(S), \tau_1^\Phi(E))$ is a gauge space and thus completely regular (see e. g. [13, Chap. IX, Theorem 10.6]). For these spaces, the concept of exponential equivalence was introduced in [15]; the lemma can be viewed as a special case of [15, Theorem 1.13]. □

We get LDP results in the $\tau_1(E)$-topology by verifying Condition 1.4. Now, in this topology the deduction of the laws of $L_n^{\otimes m}$ and L_n^m from the LDP results for reduced products is easier using the fact that convergence in the total variation distance implies convergence in $\tau_1(E)$. Clearly convergence in the total variation distance is not sufficient for convergence in the $\tau_1^\Phi(E)$-topology.

Remark 2.16. If for every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that $\int_S \exp(\alpha_\varphi \|\varphi\|_E)\,d\mu < \infty$, then the topology of convergence in information, i. e., $\mu_n \to \mu$ when $H(\mu_n\,|\,\mu) \to 0$ as $n \to \infty$, is finer than $\tau_1^\Phi(E)$ on $\mathcal{M}_1^\Phi(S)$. Since the level sets of $H(\cdot\,|\,\mu)$ are subsets of $\mathcal{M}_1^\Phi(S)$ by [16, Lemma 2.1(b)], we get this result using [6, Lemma 3.1]. Note that convergence in information implies convergence in total variation distance, see e. g. [9, Exercise 3.2.24].

3. Large Deviations for U-Empirical Measures and Products of Empirical Measures

3.1. **Independent, identically distributed, sequences.** Let us first show how we can get the LDP in the $\tau_1^\Phi(E)$-topology for $\{L_n^m\}_{n\ge m}$ and $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$, when the underlying sequence $\{X_i\}_{i\in\mathbb{N}}$ consists of i. i. d. random variables with common law $\mu \in \mathcal{M}_1(S)$. We will refer to the arguments in the following subsections.

Sanov's theorem (see e. g. [9, Theorem 3.2.17]) gives the LDP for $\{L_n\}_{n\in\mathbb{N}}$ with rate function $J_1 = H(\,\cdot\,|\mu)$ in the weak topology on $\mathcal{M}_1(S)$, where H denotes the relative entropy given in (2.5). We can use the continuity of the map $\nu \mapsto \nu^{\otimes m}$ with respect to the weak topologies on $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$ and the contraction principle [8, Theorem 4.2.1] to get the LDP in the weak topology for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ and thus for $\{L_n^m\}_{n\ge m}$ because

$$\|L_n^{\otimes m} - L_n^m\|_{\mathrm{var}} \le \frac{n^m - n_{(m)}}{n^m} \to 0 \quad \text{as } n \to \infty,$$

where $\|\cdot\|_{\mathrm{var}}$ denotes the total variation distance on $\mathcal{M}_1(S^m)$. The corresponding good rate function $J_m\colon \mathcal{M}_1(S^m) \to [0,\infty]$ is

$$J_m(\nu) = \begin{cases} J_1(\tilde\nu), & \text{if } \nu = \tilde\nu^{\otimes m}, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.1}$$

For $\{L_n^m\}_{n\ge m}$ we can verify Condition 1.4 via the Hoeffding decomposition introduced in [17]; this calculation is a special case of the one in Subsection 3.2 below and therefore omitted here. Hence, under Condition 1.7 with reference measure $\mu^{\otimes m}$, Theorem 2.1 implies the LDP for $\{L_n^m\}_{n\ge m}$ in the $\tau_1^\Phi(E)$-topology.

Under Condition 1.8, we will prove that the superexponential approximation (2.15) holds for $S_n \equiv L_n^m$ and $S_n' \equiv L_n^{\otimes m}$. Thus Lemma 2.14 implies the LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ in the $\tau_1^\Phi(E)$-topology. As a prerequisite for verifying (2.15) for $\varphi \in \Phi$, we want to define $\varphi_n\colon S^m \to E$ for every $n \ge m$ such that

$$\frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} \varphi(s_{i_1},\dots,s_{i_m}) = \frac{1}{n_{(m)}} \sum_{(i_1,\dots,i_m)\in I_{m,n}} \varphi_n(s_{i_1},\dots,s_{i_m}) \tag{3.2}$$

for all $s_1,\dots,s_n \in S$, which via (1.1) and (1.9) implies $\int_{S^m} \varphi\,dL_n^{\otimes m} = \int_{S^m} \varphi_n\,dL_n^m$. Using π_τ from Condition 1.8, one possibility is to define

$$\varphi_n = \sum_{j=1}^m \frac{n_{(j)}}{n^m} \sum_{\tau\in\mathcal{T}_j} \varphi\circ\pi_\tau, \tag{3.3}$$

where $\mathcal{T}_j$ denotes the set of all surjective maps $\tau\colon \{1,\dots,m\} \to \{1,\dots,j\}$ with $\tau(1) = 1$ and $\tau(k) \le 1 + \max\{\tau(1),\dots,\tau(k-1)\}$ for all $k \in \{2,\dots,m\}$. When checking (3.2), note that, given $j \in \{1,\dots,m\}$ and $(i_1,\dots,i_m) \in \{1,\dots,n\}^m$ consisting of exactly j different components $k_1,\dots,k_j$ which appear in this order, there then exist $(n-j)_{(m-j)} = n_{(m)}/n_{(j)}$ different choices for $(k_{j+1},\dots,k_m) \in \{1,\dots,n\}^{m-j}$ such that all components of $(k_1,\dots,k_m)$ are different. On the other hand, there exists exactly one $\tau \in \mathcal{T}_j$ such that $(i_1,\dots,i_m) = (k_{\tau(1)},\dots,k_{\tau(m)})$. Note that $1 - n_{(m)}/n^m \le 1 - (n-m)^m/n^m \le m^2/n$. Using (3.3), it follows that

$$n\|\varphi(s) - \varphi_n(s)\|_E \le m^2\|\varphi(s)\|_E + \sum_{j=1}^{m-1}\sum_{\tau\in\mathcal{T}_j}\|\varphi\circ\pi_\tau(s)\|_E, \qquad s \in S^m. \tag{3.4}$$

To verify (2.15) for $\varepsilon > 0$, choose α_φ according to Condition 1.8 and define $\alpha = \alpha_\varphi/(\beta(m^2 + m^{m-1}))$ with β as in Condition 1.4. By (3.2), the exponential Chebyshev inequality and Condition 1.4 with reference measure $\mu^{\otimes m}$,

$$\begin{aligned}\mathbb{P}\Big(\Big\|\int_{S^m}\varphi\,dL_n^m - \int_{S^m}\varphi\,dL_n^{\otimes m}\Big\|_E \ge \varepsilon\Big)&\\ &\le \exp(-\alpha\varepsilon n^2)\,\mathbb{E}\Big[\exp\Big(n\int_{S^m}\alpha n\|\varphi - \varphi_n\|_E\,dL_n^m\Big)\Big]\\ &\le \exp(-\alpha\varepsilon n^2)\Big(M\int_{S^m}\exp\big(\alpha\beta n\|\varphi-\varphi_n\|_E\big)\,d\mu^{\otimes m}\Big)^n. \end{aligned}\tag{3.5}$$

Using (3.4), $|\bigcup_{j=1}^{m-1}\mathcal{T}_j| \le m^{m-1}$, Hölder's inequality and Condition 1.8, it follows that the last integral in (3.5) is bounded by a real constant which does not depend on n. Therefore, (2.15) follows from (3.5). This proves the LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ in the $\tau_1^\Phi(E)$-topology.

The LDP results in [16] for $\{L_n^m\}_{n\ge m}$ and $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ are more general, because there $(S,\mathcal{S})$ is only assumed to be an arbitrary measurable space. Several technical reasons, in particular the continuity of the map $\nu \mapsto \nu^{\otimes m}$ with respect to the weak topologies used in the above argument and Lusin's theorem used in the proof of Lemma 2.6, lead us to the Polish setting in this paper. It is an advantage that we can depart from the i. i. d. setting here.

3.2. **Independent, but not identically distributed, sequences.** Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables with values in S and laws $\mu_i \equiv \mathcal{L}(X_i)$. Assume that $\mu_i \ll \mu$ for all $i \in \mathbb{N}$ with a fixed reference measure $\mu \in \mathcal{M}_1(S)$. Moreover, we assume that there exists a $q > 1$ such that for $f_i \equiv d\mu_i/d\mu$ we have

$$M \equiv \sup_{i\in\mathbb{N}}\|f_i\|_q < \infty, \tag{3.6}$$

where $\|\cdot\|_q$ denotes the q-norm in $L_q(S,\mathcal{S},\mu)$. For every $n \ge m$ and every bounded measurable $\varphi\colon S^m \to [0,\infty)$, by Hoeffding's decomposition [17, Section 5] we get

$$\mathbb{E}\Big[\exp\Big(n\int_{S^m}\varphi\,dL_n^m\Big)\Big] = \mathbb{E}\Big[\exp\Big(\frac{1}{n!}\sum_\tau\frac{n}{k}\sum_{i=0}^{k-1}\varphi(X_{\tau(im+1)},\dots,X_{\tau(im+m)})\Big)\Big],$$

where $k \equiv \lfloor n/m\rfloor$ and τ runs through all permutations of $\{1,\dots,n\}$. Using Jensen's inequality to handle the first convex combination, using the independence of the terms in the second sum, $n/k \le 2m$, and, for every $i \in \{0, \dots, k-1\}$,

$$\begin{aligned}\mathbb{E}\bigl[\exp\bigl(2m\varphi(X_{\tau(im+1)}, \dots, X_{\tau(im+m)})\bigr)\bigr] \\ \le \biggl(\int_{S^m} \exp(2mp\varphi)\, d\mu^{\otimes m}\biggr)^{1/p} \biggl(\prod_{j=1}^{m} \int_S f^q_{\tau(im+j)}\, d\mu\biggr)^{1/q}, \quad (3.7)\end{aligned}$$

which follows from Hölder's inequality with $1/p + 1/q = 1$, we obtain with (3.6)

$$\mathbb{E}\biggl[\exp\biggl(n\int_{S^m} \varphi\, dL_n^m\biggr)\biggr] \le M^{km}\biggl(\int_{S^m} \exp(2mp\varphi)\, d\mu^{\otimes m}\biggr)^{k/p}.$$

Since $M \ge 1$ by Jensen's inequality, $km \le n$, and $k/p \le n$, Condition 1.4 holds with $\beta \equiv 2mp$, $n_0 \equiv m$ and reference measure $\mu^{\otimes m}$. Thus the LDP for $\{L_n^m\}_{n\ge m}$ in the $\tau_1^\Phi(E)$-topology is proved as in the i. i. d. case described in Subsection 3.1, whenever we assume that $\{L_n\}_{n\in\mathbb{N}}$, built with the $\{X_i\}_{i\in\mathbb{N}}$, already satisfies a LDP with a rate function J_1 in the weak topology and Condition 1.7 holds. If, in addition, we assume Condition 1.8, then we get the corresponding LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ by the superexponential approximation argument given at the end of Subsection 3.1.

3.3. **Markov chains.** Let $\pi\colon \mathcal{S} \times \mathcal{S} \to [0,1]$ be a probability transition kernel and let $\{\, \mathbb{P}_\theta \mid \theta \in \mathcal{M}_1(S)\,\}$ be the family of Markovian measures on the sequence space $(\Omega, \mathcal{A}) = (S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}})$ such that the coordinate projections $\{X_i\}_{i\in\mathbb{N}}$ from Ω to S form a Markov chain with transition kernel π and $\mathbb{P}_\theta X_1^{-1} = \theta$ for every $\theta \in \mathcal{M}_1(S)$. For $s \in S$, we write $\mathbb{P}_s$ instead of $\mathbb{P}_{\delta_s}$. We use $\mathbb{P}_\theta(A) = \int_S \mathbb{P}_s(A)\,\theta(ds)$ for $A \in \mathcal{A}$ and $\theta \in \mathcal{M}_1(S)$. We assume that there exist $N \in \mathbb{N}$, $l \in \{1, 2, \dots, N\}$ and $M \in [1, \infty)$ satisfying

$$\pi^l(s,\cdot) \le \frac{M}{N}\sum_{i=1}^{N} \pi^i(\tilde{s},\cdot) \qquad \text{for all } s, \tilde{s} \in S. \tag{3.8}$$

As in [9, (4.1.39)] define the rate function $J_1\colon \mathcal{M}_1(S) \to [0,\infty]$ by

$$J_1(\tilde{\nu}) = -\inf_{u\in B(S,[1,\infty))} \int_S \log\frac{\pi u}{u}\, d\tilde{\nu}, \tag{3.9}$$

where $B(S,[1,\infty))$ denotes the set of bounded measurable functions $u\colon S \to [1,\infty)$. For every integer $m \ge 2$ define $J_m\colon \mathcal{M}_1(S^m) \to [0,\infty]$ by

$$J_m(\nu) = \begin{cases} J_1(\tilde{\nu}) & \text{if } \tilde{\nu}^{\otimes m} = \nu, \\ \infty & \text{otherwise.}\end{cases} \tag{3.10}$$

For the remaining part of this subsection, let $\mu \in \mathcal{M}_1(S)$ be given by

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \tilde{\mu}\pi^i \tag{3.11}$$

with an arbitrarily chosen $\tilde{\mu} \in \mathcal{M}_1(S)$. We will prove the following theorem for every $m \in \mathbb{N}$.

Theorem 3.12. *Take a non-empty $\Theta \subset \mathcal{M}_1(S)$. Assume* (3.8)*, Condition* 1.7 *with reference measure $\mu^{\otimes m}$ and that, for every $\varphi \in \Phi$,*

$$\sup_{\theta\in\Theta}\sup_{\tau} \mathbb{E}_\theta\bigl[\exp\bigl(\alpha\|\varphi(X_{\tau(1)}, \dots, X_{\tau(m)})\|_E\bigr)\bigr] < \infty \qquad \textit{for all } \alpha > 0, \tag{3.13}$$

where $\sup_\tau$ *ranges over all injective* $\tau\colon \{1,\dots,m\} \to \mathbb{N}$. *Then the measures* $\{\mathbb{P}_\theta(L_n^m)^{-1}\}_{n\ge m,\theta\in\Theta}$ *satisfy a uniform LDP in the* $\tau_1^\Phi(E)$*-topology on* $\mathcal{M}_1^\Phi(S^m)$ *with the good rate function* J_m*, i. e., for every measurable* $B \subset \mathcal{M}_1(S^m)$*,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{\theta\in\Theta} \mathbb{P}_\theta(L_n^m \in B) \ge -\inf\{\, J_m(\nu) \mid \nu \in \mathcal{M}_1^\Phi(S^m),\, \nu \in \operatorname{int}_{\tau_1^\Phi(E)}(B)\,\}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\theta\in\Theta} \mathbb{P}_\theta(L_n^m \in B) \le -\inf\{\, J_m(\nu) \mid \nu \in \mathcal{M}_1^\Phi(S^m),\, \nu \in \operatorname{cl}_{\tau_1^\Phi(E)}(B)\,\},$$

and the level set $\{\, \nu \in \mathcal{M}_1^\Phi(S^m) \mid J_m(\nu) \le r\,\}$ *is* $\tau_1^\Phi(E)$*-compact for every* r *in* $[0,\infty)$*. If, in addition, we assume* (3.13) *with* $\sup_\tau$ *ranging over all* $\tau\colon \{1,\dots,m\} \to \mathbb{N}$*, then we get the same uniform LDP for the measures* $\{\mathbb{P}_\theta(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N},\theta\in\Theta}$.

Remark 3.14. Even in the case $m = 1$, as soon as there is a $\varphi \in \Phi$ which is unbounded from above, the function $S \ni s \mapsto \mathbb{E}_s[\exp(\alpha\varphi(X_1))]$ with $\alpha > 0$ is also unbounded, hence (3.13) does not hold for $\Theta = \{\delta_s\}_{s\in S}$. For this reason, we allow for other sets $\Theta \subset \mathcal{M}_1(S)$ in Theorem 3.12, Corollary 4.7 and Theorem 4.14.

Proof. By [9, Theorem 4.1.43], the uniform LDP for $\{\mathbb{P}_s L_n^{-1}\}_{n\in\mathbb{N},s\in S}$ with rate function J_1 from (3.9) holds in the weak topology on $\mathcal{M}_1(S)$. By the continuity of the map $\nu \mapsto \nu^{\otimes m}$ with respect to the weak topologies on $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$ and the contraction principle [8, Theorem 4.2.1], the uniform LDP for $\{\mathbb{P}_s(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N},s\in S}$ with rate function J_m from (3.10) holds in the weak topology on $\mathcal{M}_1(S^m)$.

For $n \ge lm+1$ define the empirical measure $L_{m,n}\colon \Omega \to \mathcal{M}_1(S^m)$ by

$$L_{m,n}(\omega) = \frac{1}{|A_{m,n}|} \sum_{(i_1,\dots,i_m)\in A_{m,n}} \delta_{(X_{i_1}(\omega),\dots,X_{i_m}(\omega))}, \tag{3.15}$$

where

$$A_{m,n} \equiv \{\, (i_1,\dots,i_m) \in \{l+1, l+2,\dots,n\}^m \mid |i_j - i_k| \ge l \text{ for all } j,k \in \{1,\dots,m\} \text{ with } j \ne k\,\}. \tag{3.16}$$

Since $|A_{m,n}| \ge (\max\{1, n-(2m-1)l\})^m$ for $n \ge lm+1$ it follows that

$$\sup_{\omega\in\Omega} \|L_n^{\otimes m}(\omega) - L_{m,n}(\omega)\|_{\mathrm{var}} \le \frac{n^m - |A_{m,n}|}{n^m} \to 0 \qquad \text{as } n \to \infty.$$

Therefore, the uniform LDP also holds for $\{\mathbb{P}_s L_{m,n}^{-1}\}_{n\ge lm+1,s\in S}$ in the weak topology on $\mathcal{M}_1(S^m)$. Below, we will establish the crucial estimate

$$\sup_{n\ge 4ml} \Bigl(\sup_{s\in S} \mathbb{E}_s\Bigl[\exp\Bigl(n\int_{S^m} \varphi\, dL_{m,n}\Bigr)\Bigr]\Bigr)^{1/n} \le M^m \int_{S^m} \exp(4^m l m! \varphi)\, d\mu^{\otimes m} \tag{3.17}$$

for every $\varphi \in B(S^m,[0,\infty))$, where $\mu \in \mathcal{M}_1(S)$ is defined by (3.11). Using Condition 1.7, an inspection of the proof of Theorem 2.1 shows that, due to the uniformity with respect to $s \in S$ in (3.17), we get the uniform LDP for $\{\mathbb{P}_s L_{m,n}^{-1}\}_{n\ge lm+1,s\in S}$ in the $\tau_1^\Phi(E)$-topology on $\mathcal{M}_1^\Phi(S^m)$. This implies the same uniform LDP for the measures $\{\mathbb{P}_\theta L_{m,n}^{-1}\}_{n\ge lm+1,\theta\in\Theta}$. To obtain Theorem 3.12 from this result via Lemma 2.14, we must check that, for every $\varepsilon > 0$ and every $\varphi \in \Phi$,

$$\lim_{n\to\infty} \frac{1}{n} \log \sup_{\theta\in\Theta} \mathbb{P}_\theta\Bigl(\Bigl\|\int_{S^m} \varphi\, dL_{m,n} - \int_{S^m} \varphi\, dL_n^m\Bigr\|_E \ge \varepsilon\Bigr) = -\infty, \tag{3.18}$$

and the same for L_n^m replaced by $L_n^{\otimes m}$.

To prove (3.18), choose $\alpha > 0$. For every $n \ge lm+1$, the exponential Chebyshev inequality yields:

$$\sup_{\theta\in\Theta} \mathbb{P}_\theta\biggl(\biggl\|\int_{S^m} \varphi\,dL_{m,n} - \int_{S^m} \varphi\,dL_n^m\biggr\|_E \ge \varepsilon\biggr) \le \exp(-\alpha\varepsilon n) \sup_{\theta\in\Theta} \mathbb{E}_\theta\Bigl[\exp\Bigl(\alpha n \int_{S^m} \|\varphi\|_E\,d|L_{m,n} - L_n^m|\Bigr)\Bigr]. \quad (3.19)$$

Using (1.9), (3.15) and $|I_{m,n} \setminus A_{m,n}| \le 2lm^2(n-1)_{(m-1)}$, it follows that

$$\alpha n \int_{S^m} \|\varphi\|_E\,d|L_{m,n} - L_n^m| \le \frac{\alpha n}{n_{(m)}} \sum_{(i_1,\dots,i_m)\in I_{m,n}\setminus A_{m,n}} \|\varphi(X_{i_1},\dots,X_{i_m})\|_E + 2\alpha l m^2 \int_{S^m} \|\varphi\|_E\,dL_{m,n}. \quad (3.20)$$

Using Jensen's inequality and (3.13), we see that

$$\sup_{n\ge lm+1} \sup_{\theta\in\Theta} \mathbb{E}_\theta\Bigl[\exp\Bigl(4\alpha l m^2 \int_{S^m} \|\varphi\|_E\,dL_{m,n}\Bigr)\Bigr] < \infty \quad (3.21)$$

and

$$\sup_{n\ge lm+1} \sup_{\theta\in\Theta} \mathbb{E}_\theta\biggl[\exp\biggl(\frac{2\alpha n}{n_{(m)}} \sum_{(i_1,\dots,i_m)\in I_{m,n}\setminus A_{m,n}} \|\varphi(X_{i_1},\dots,X_{i_m})\|_E\biggr)\biggr] < \infty. \quad (3.22)$$

Substituting (3.20) into (3.19), applying Chebyshev's inequality and using (3.21) and (3.22), we see that the left-hand side of (3.18) is bounded above by $-\alpha\varepsilon$. Since $\alpha > 0$ was arbitrary, (3.18) is proved. Using (3.13) with $\sup_\tau$ ranging over all $\tau\colon \{1,\dots,m\} \to \mathbb{N}$, a similar argument proves (3.18) with $L_n^{\otimes m}$ in place of L_n^m.

In the remaining part of this proof, we show estimate (3.17). Let us first introduce some additional notation. With $n \ge lm+1$ and $A_{m,n}$ given by (3.16), let

$$A'_{m,n} = \{\,(i_1,\dots,i_m) \in A_{m,n} \mid i_j \ge i_{j-1} + l \text{ for all } j \in \{2,\dots,m\}\,\}$$

denote the subset of all ordered m-tuples of $A_{m,n}$. Define

$$B_{m,n} = \{l+1, l+2, \dots, 2l\} \times \{2l+1, 2l+2, \dots, 2n\}^{m-1}. \quad (3.23)$$

Given $r = (r_1,\dots,r_m) \in B_{m,n}$, let $C_{n,r}$ denote the set of all $(i_1,\dots,i_m) \in A'_{m,n}$ for which there exists $k \in \mathbb{N}_0$ satisfying $i_j = r_j - (-1)^j kl$ for all $j \in \{1,\dots,m\}$. Every set $C_{n,r}$ has the following two properties:

(a) Every $i \in \{l+1, l+2, \dots, n\}$ occurs at most once in at most one m-tuple contained in $C_{n,r}$.
(b) If $i, i' \in \{l+1, l+2, \dots, n\}$ are components of an m-tuple contained in $C_{n,r}$ and $i \ne i'$ then $|i - i'| \ge l$.

Define $B'_{m,n} = \{\, r \in B_{m,n} \mid C_{n,r} \ne \varnothing \,\}$. In order to show that

$$A'_{m,n} \subset \bigcup_{r\in B'_{m,n}} C_{n,r} \quad (3.24)$$

take any $(i_1,\dots,i_m) \in A'_{m,n}$. Then there exist $r_1 \in \{l+1, l+2, \dots, 2l\}$ and $k \in \{0, 1, \dots, \lfloor n/l \rfloor - 1\}$ such that $i_1 = r_1 + kl$. For every $j \in \{2, 3, \dots, m\}$ define

$r_j = i_j + (-1)^j kl$. Since $i_j \geq i_1 + l$ it follows that $r_j \geq i_j - kl \geq r_1 + l \geq 2l + 1$. Since $i_j \leq n$ and $k \leq n/l$, we obtain $r_j \leq n + kl \leq 2n$. Hence $(r_1, \dots, r_m) \in B'_{m,n}$.

Define the symmetrized version of φ by $\widetilde{\varphi} = (1/m!) \sum_\tau \varphi \circ \pi_\tau$, where the sum extends over all permutations τ of $\{1, \dots, m\}$ and π_τ is defined as in Condition 1.8. Starting with the expectation on the left-hand side of (3.17), passing to the symmetrized version $\widetilde{\varphi}$, using (3.24) and Hölder's inequality, for every $n \geq lm + 1$ and $s \in S$ we get:

$$\begin{aligned}
&\mathbb{E}_s\Big[\exp\Big(n \int_{S^m} \varphi \, dL_{m,n}\Big)\Big] \\
&\quad = \mathbb{E}_s\Big[\exp\Big(\frac{nm!}{|A_{m,n}|} \sum_{(i_1,\dots,i_m) \in A'_{m,n}} \widetilde{\varphi}(X_{i_1}, \dots, X_{i_m})\Big)\Big] \\
&\quad \leq \prod_{r \in B'_{m,n}} \Big(\mathbb{E}_s\Big[\exp\Big(\frac{nm! \, |B'_{m,n}|}{|A_{m,n}|} \sum_{(i_1,\dots,i_m) \in C_{n,r}} \widetilde{\varphi}(X_{i_1}, \dots, X_{i_m})\Big)\Big]\Big)^{1/|B'_{m,n}|}.
\end{aligned} \tag{3.25}$$

Since $|B'_{m,n}| \leq |B_{m,n}| \leq l(2n)^{m-1}$ and $|A_{m,n}| \geq (n - (2m-1)l)^m \geq (n/2)^m$ for $n \geq 4lm$, it follows that

$$\frac{nm! \, |B'_{m,n}|}{|A_{m,n}|} \leq 4^m l m! \quad \text{for } n \geq 4lm. \tag{3.26}$$

Given $r \in B'_{m,n}$, it follows from (a) and (b) that the m-tuples in $C_{n,r}$ consist of $p \equiv m|C_{n,r}|$ different $q_1, \dots, q_p \in \{l+1, \dots, n\}$, which we can label such that $q_j \geq q_{j-1} + l$ for all $j \in \{2, 3, \dots, p\}$. It follows from (3.8) and (3.11) that $\pi^{q_j - q_{j-1}}(s, \cdot) \leq M\mu$ for every $j \in \{1, \dots, p\}$ and $s \in S$, where $q_0 \equiv 1$. Hence, for every $s \in S$, $\mathbb{P}_s(X_{q_1}, \dots, X_{q_p})^{-1} \leq M^p \mu^{\otimes p}$ on $\mathcal{S}^{\otimes p}$. Using (3.26) it follows that

$$\begin{aligned}
&\mathbb{E}_s\Big[\exp\Big(\frac{nm! \, |B'_{m,n}|}{|A_{m,n}|} \sum_{(i_1,\dots,i_m) \in C_{n,r}} \widetilde{\varphi}(X_{i_1}, \dots, X_{i_m})\Big)\Big] \\
&\qquad \leq \Big(M^m \int_{S^m} \exp(4^m l m! \varphi) \, d\mu^{\otimes m}\Big)^{|C_n(\pi,r)|},
\end{aligned} \tag{3.27}$$

where we used Jensen's inequality to pass back from $\widetilde{\varphi}$ to φ in the last step. Since $|C_{n,r}| \leq \lfloor n/lm \rfloor \leq n$, the estimates (3.25) and (3.27) imply (3.17). □

Remark 3.28. Note that assumption (3.8) on the transition probabilities suffices to get the uniform LDP results in the $\tau_1(E)$-topology for products of empirical measures and U-empirical measures from Theorem 3.12, because the moment assumptions in (3.13) and Condition 1.7 are always satisfied for $\Phi = B(S^m, E)$.

Remark 3.29. By [9, Exercise 4.1.48], condition (3.8) guarantees the existence of a unique π-invariant $\tilde{\mu} \in \mathcal{M}_1(S)$. Using $\tilde{\mu}$ in (3.11), we get $\mu = \tilde{\mu}$, and Theorem 3.12 then assumes Condition 1.7 with respect to the m-fold product of the π-invariant distribution.

Remark 3.30. The Markov chains analyzed in [14] show that the LDP with respect to the weak topology does not transfer to the τ-topology setting in general.

3.4. **Exchangeable sequences.** Let $\{X_i\}_{i\in\mathbb{N}}$ be an exchangeable sequence on the probability space $(\Omega,\mathcal{A},\mathbb{P})$. It follows from de Finitti's representation theorem [3], that $\mathbb{P}$ can be represented as a Σ-mixture of probability measures $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ defined on $(\Omega,\mathcal{A})$, where Θ is a subset in $\mathcal{M}_1(S)$ and for any $\theta\in\Theta$ the sequence $\{X_i\}_{i\in\mathbb{N}}$ is i. i. d. with respect to $\mathbb{P}_\theta$. If $\mathbb{P}^n\equiv\mathbb{P}L_n^{-1}$ and $\mathbb{P}_\theta^n\equiv\mathbb{P}_\theta L_n^{-1}$, then $\mathbb{P}^n$ is the Σ-mixture of $\{\mathbb{P}_\theta^n\}_{\theta\in\Theta}$, that is

$$\mathbb{P}^n(A)=\int_\Theta \mathbb{P}_\theta^n(A)\,d\Sigma(\theta),\qquad A\in\mathcal{A}.$$

If $\Theta\ni\theta\mapsto\mu_\theta\equiv\mathbb{P}_\theta X_1^{-1}$ is continuous in the weak topology on $\mathcal{M}_1(S)$ and if Θ is weakly compact, then $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ satisfies the LDP with rate function J_1 given by

$$J_1(\nu)\equiv\inf_{\theta\in\operatorname{supp}(\Sigma)}H(\nu\,|\,\mu_\theta),\qquad \nu\in\mathcal{M}_1(S),\tag{3.31}$$

where $\operatorname{supp}(\Sigma)$ denotes the support of Σ, cf. [10, Remark (ii) after Theorem 2.3]. An extension of this result is given in [10, Theorem 4.1]. The condition of weak continuity of $\theta\mapsto\mu_\theta$ is replaced by the following two conditions:

(a) λ: $\Theta\times\mathcal{M}_1(S)\to[0,\infty]$ with $\lambda(\theta,\nu)\equiv H(\nu\,|\,\mu_\theta)$ is jointly lower semi-continuous,
(b) $\{\mathbb{P}_\theta^n\}_{n\in\mathbb{N}}$ is exponentially tight, meaning that for every $L>0$ there exists a compact set $K_L\subset\mathcal{M}_1(S)$ such that

$$\sup_{\theta\in\Theta}\mathbb{P}_\theta^n(K_L^{\mathrm{c}})<\exp(-nL)\quad\text{for all sufficiently large } n\in\mathbb{N}.$$

Note that for exchangeable sequences, we cannot expect the LDP for the laws of $\{L_n\}_{n\in\mathbb{N}}$ with respect to the weak topology to carry over to the product in the τ-topology, because the following example shows that, even in the case $m=1$, the LDP cannot be transferred from the weak to the τ-topology.

Example 3.32. Denote by Σ the Lebesgue–Borel measure on $S\equiv[0,1]$, equipped with the Borel σ-algebra. Let $\mathbb{P}_\theta\equiv(\delta_\theta)^{\otimes\mathbb{N}}$ for $\theta\in[0,1]$ and $\mathbb{P}\equiv\int_{[0,1]}\mathbb{P}_\theta\,\Sigma(d\theta)$. Note that via $\theta\mapsto\delta_\theta$ we can identify $[0,1]$ with $\Theta\equiv\{\delta_\theta\}_{\theta\in[0,1]}$. We have

$$\mathbb{P}(L_n\in U)=\int_{[0,1]}1_U(\delta_\theta)\,\Sigma(d\theta)$$

for every Borel set U in $\mathcal{M}_1([0,1])$. According to (3.31), the rate function for the LDP in the weak topology is

$$J_1(\nu)=\inf_{\theta\in[0,1]}H(\nu\,|\,\delta_\theta)=\begin{cases}0 & \text{if } \nu\in\Theta,\\ \infty & \text{otherwise.}\end{cases}$$

For $\varepsilon\in(0,1]$ and $\theta\in[0,1]$ define $U_{\varepsilon,\theta}=\{\,\nu\in\mathcal{M}_1([0,1])\mid\nu(\{\theta\})>1-\varepsilon\,\}$. Then

$$\mathbb{P}(L_n\in U_{\varepsilon,\theta})=\int_{[0,1]}\mathbb{P}_\eta\bigl(L_n(\{\theta\})>1-\varepsilon\bigr)\,\Sigma(d\eta);$$

the integrand is equal to 1 for $\eta=\theta$ and 0 otherwise. Therefore $\mathbb{P}(L_n\in U_{\varepsilon,\theta})=0$ but $\inf_{\nu\in U_{\varepsilon,\theta}}J_1(\nu)=0$, since $\delta_\theta\in U_{\varepsilon,\theta}$, and thus the LDP lower bound fails with respect to the τ-topology. Moreover, note that the level sets of J_1 are not τ-compact, because

$$J_1^{-1}(\{0\})=\Theta\subset\bigcup_{\theta\in[0,1]}U_{\varepsilon,\theta}$$

and no finite subcover exists; not a single $U_{\varepsilon,\theta}$ is superfluous.

Despite this example, [15, Theorem 1.19] is an LDP result for exchangeable sequences in the τ-topology. Let $(S, \mathcal{S})$ be a general measurable state space. If we start with the mixture $\mathbb{P}^n$ instead of assuming that the projection maps $\{X_i\}_{i\in\mathbb{N}}$ are an exchangeable process (because de Finetti's representation theorem does not hold in the general setting of an arbitrary measurable space S, see [12, Section 2]), then conditions (a) and (b) above can be replaced by the following condition.

(c) The map $\Theta \ni \theta \mapsto \mu_\theta \in \mathcal{M}_1(S)$ is continuous in the τ-topology and $\Sigma(U) > 0$ for every measurable, relatively τ-open subset U of Θ.

If Θ is τ-compact and Condition (c) holds, then $\{\mathbb{P}^n\}_{n\in\mathbb{N}}$ satisfies the LDP in the τ-topology on $\mathcal{M}_1(S)$ with the rate function J_1. A similar result is proved in [7, Section 2] by different methods.

Now, we will assume in addition that $\{\mu_\theta\}_{\theta\in\Theta}$ satisfies the next condition.

(d) There exists a reference measure $\mu \in \mathcal{M}_1(S)$ such that $\mu_\theta \ll \mu$ for every $\theta \in \Theta$ and there exists a $q > 1$ such that the densities $f_\theta \equiv d\mu_\theta/d\mu$ satisfy $M \equiv \sup_{\theta\in\Theta} \|f_\theta\|_q < \infty$, where $\|\cdot\|_q$ denotes the norm in $L_q(S, \mathcal{S}, \mu)$.

The following result is stated for a Polish state space S:

Theorem 3.33. *Consider $m \in \mathbb{N}$ and an exchangeable sequence $\{X_i\}_{i\in\mathbb{N}}$ satisfying Condition* (d)*.*

(a) *Assume Conditions* (a)*,* (b) *and the weak compactness of Θ (or, alternatively, assume that the laws of the empirical measures of $\{X_i\}_{i\in\mathbb{N}}$ satisfy an LDP in the weak topology, not necessarily satisfying Conditions* (a) *and* (b)*).*

  (i) *If Condition* 1.7 *holds for $\mu^{\otimes m}$ and Φ, then the measures $\{\mathbb{P}(L_n^m)^{-1}\}_{n\geq m}$ satisfy the LDP in the $\tau_1^\Phi(E)$-topology on $\mathcal{M}_1^\Phi(S^m)$ with the good rate function*

$$J_m(\nu) = \begin{cases} J_1(\tilde{\nu}) & \text{if } \tilde{\nu}^{\otimes m} = \nu, \\ \infty & \text{otherwise.} \end{cases} \tag{3.34}$$

  (ii) *If Condition* 1.8 *holds for Φ and $\mu^{\otimes m}$, then the same result is true for the measures $\{\mathbb{P}(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N}}$ with the same good rate function J_m.*

(b) *If Condition* (c) *holds and Θ is τ-compact, then* (ai) *and* (aii) *are also true.*

Proof. By [10, Theorem 4.1] and [15, Theorem 1.19], respectively, the theorem is true in the weak topology for $m = 1$. Similarly to (3.7), for each $\theta \in \Theta$, by Hölder's inequality with $p > 1$ satisfying $1/p + 1/q = 1$ and Condition (d) we get:

$$\mathbb{E}_{\mathbb{P}_\theta}\big[\exp(2m\varphi(X_1, \ldots, X_m))\big] \leq M^m \left(\int_{S^m} \exp(2mp\varphi)\, d\mu^{\otimes m} \right)^{1/p}.$$

Since

$$\mathbb{E}_{\mathbb{P}}\Big[\exp\Big(n \int_{S^m} \varphi\, dL_n^m\Big)\Big] \leq \sup_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}_\theta}\Big[\exp\Big(n \int_{S^m} \varphi\, dL_n^m\Big)\Big],$$

we can use the same arguments as in Subsections 3.1 and 3.2 to get the result. □

Example 3.35. A simple case is a 0-1-valued exchangeable sequence $\{X_i\}_{i\in\mathbb{N}}$. One can find a $\Sigma \in \mathcal{M}_1([0,1])$ such that with $\mu_\theta \equiv (1-\theta)\delta_0 + \theta\delta_1$ for $\theta \in [0,1]$ the distribution $\mathbb{P}$ of $\{X_i\}_{i\in\mathbb{N}}$ can be represented as $\mathbb{P} = \int_{[0,1]} \mu_\theta^{\otimes\mathbb{N}}\, \Sigma(d\theta)$ (see [3, Section 4]).

With the conventions $0\log(0/a) = 0$ for $a \geq 0$ and $a\log(a/0) = \infty$ for $a > 0$, the function

$$\lambda(\theta,\nu) \equiv H(\nu\,|\,\mu_\theta) = \nu(\{0\})\log\frac{\nu(\{0\})}{1-\theta} + \nu(\{1\})\log\frac{\nu(\{1\})}{\theta}$$

is jointly lower semi-continuous on $[0,1]\times\mathcal{M}_1(\{0,1\})$, and Condition (a) holds. Since $\mathcal{M}_1(\{0,1\})$ is compact in the weak topology and $f_\theta \equiv d\mu_\theta/d\mu_{1/2} \leq 2$ for all $\theta\in[0,1]$, Conditions (b) and (d) hold. Hence, we can apply Theorem 3.33.

Example 3.36 (Mixtures of normal distributions)**.** Let $\Theta' \subset \mathbb{R}^d\times\mathbb{R}^{d\times d}$ be a compact set such that, for every $\theta = (m,C)\in\Theta'$, the $(d\times d)$-matrix C is positive definite. Let μ_θ be the corresponding d-dimensional normal distribution with mean m and covariance matrix C. Define $\Theta = \{\mu_\theta\}_{\theta\in\Theta'} \subset \mathcal{M}_1(\mathbb{R}^d)$.

Let us verify the conditions on $\theta\mapsto\mu_\theta$ for the applicability of Theorem 3.33. By the compactness of Θ' there exist $r,\sigma_1,\sigma_2 > 0$ such that every $(m,C)\in\Theta'$ satisfies $\|m\|_2\leq r$ and $\|\sigma_1 x\|_2^2 \leq \langle x, Cx\rangle \leq \|\sigma_2 x\|_2^2$ for all $x\in\mathbb{R}^d$. Let I_d denote the d-dimensional identity matrix. Then

$$\frac{d\mu_{(m,C)}}{d\mu_{(0,\sigma_2^2 I_d)}}(x) \leq \Bigl(\frac{\sigma_2}{\sigma_1}\Bigr)^d \exp\biggl(\frac{r\|x\|_2}{\sigma_1^2}\biggr), \qquad \text{for all } x\in\mathbb{R}^d,$$

and every power $q > 0$ of the right-hand side is integrable with respect to the normal distribution $\mu_{(0,\sigma_2^2 I_d)}$. Hence, by the dominated convergence theorem, the map $\Theta'\ni\theta\mapsto\mu_\theta(A)$ is continuous for every $A\in\mathcal{B}^{\otimes d}$, hence $\Theta'\ni\theta\mapsto\mu_\theta$ is continuous in the τ-topology and $\Theta = \{\mu_\theta\}_{\theta\in\Theta'}$ is τ-compact. Furthermore, Condition (d) is satisfied with reference measure $\mu_{(0,\sigma_2^2 I_d)}$.

Let us remark that Example 3.36 is of interest in other aspects: Schoenberg's theorem (see, e. g., [3, Theorem 3.6]) states that every infinite spherically symmetric sequence of real-valued random variables is a mixture of i. i. d. $\mathcal{N}(0,\sigma^2)$ sequences. This result fits naturally into the sufficient statistics setting, discussed in [3, Section 18]. There, the laws of special sufficient statistics can be described by mixtures of i. i. d. $\mathcal{N}(m,\sigma^2)$ sequences, or Poisson, binomial and negative binomial sequences.

3.5. **Stationary sequences, mixing conditions.** Let $\{X_i\}_{i\in\mathbb{N}}$ be a stationary sequence of S-valued random variables with $\mu\equiv\mathcal{L}(X_1)$. To formulate the hypermixing conditions (H-1) and (H-2) of [8, Section 6.4.2], the following notation is needed: For any integers $n\geq k\geq 2$ and $l\in\mathbb{N}$, a family of functions $f_1,\dots,f_k\colon S^n\to\mathbb{R}$ is called l-separated if there exist k disjoint discrete intervals $\{a_i,a_i+1,\dots,b_i\}$ with $a_i,b_i\in\{1,\dots,n\}$, $a_i\leq b_i$, such that $f_i(s_1,\dots,s_n)$ is actually a function of $\{s_{a_i},\dots,s_{b_i}\}$ and either $a_i-b_j\geq l$ or $a_j-b_i\geq l$ for all $i,j\in\{1,\dots,k\}$ with $i\neq j$.

Assumption (H-1). There exist $\alpha\in[1,\infty)$ and $l\in\mathbb{N}$ such that for all integers $n\geq k\geq 2$ and any l-separated functions $f_1,\dots,f_k\in B(S^n,\mathbb{R})$

$$\mathbb{E}\bigl[|f_1(X_1,\dots,X_n)\dots f_k(X_1,\dots,X_n)|\bigr] \leq \prod_{i=1}^k \mathbb{E}\bigl[|f_i(X_1,\dots,X_n)|^\alpha\bigr]^{1/\alpha}. \tag{3.37}$$

Assumption (H-2). There exist a constant $l_0\in\mathbb{N}$ and functions $\beta\colon\mathbb{N}\to[1,\infty)$, $\gamma\colon\mathbb{N}\to[0,\infty)$ such that, for all integers $l\geq l_0$ and $n\geq 2$ and any two l-separated functions $f, g \in B(S^n, \mathbb{R})$,

$$\big|\mathbb{E}[f(X_1, \dots, X_n)]\,\mathbb{E}[g(X_1, \dots, X_n)] - \mathbb{E}[f(X_1, \dots, X_n)g(X_1, \dots, X_n)]\big| \\ \le \gamma(l)\,\mathbb{E}\big[|f(X_1, \dots, X_n)|^{\beta(l)}\big]^{1/\beta(l)}\,\mathbb{E}\big[|g(X_1, \dots, X_n)|^{\beta(l)}\big]^{1/\beta(l)},$$

where

$$\lim_{l\to\infty} \gamma(l) = 0 \quad \text{and} \quad \limsup_{l\to\infty} (\beta(l) - 1) l (\log l)^{1+\delta} < \infty$$

for some $\delta > 0$.

It is well known that the laws of $\{L_n\}_{n\in\mathbb{N}}$ satisfy the LDP in $\mathcal{M}_1(S)$ equipped with the τ-topology, if (H-1) and (H-2) hold [8, Theorem 6.4.14 and Lemma 6.4.18]. Furthermore, by [5, Proposition 1], the Condition (H-1) is unnecessary. For our purpose, we use a related version of Assumption (H-1), for which we need additional notation. For $l, m, n \in \mathbb{N}$ with $n \ge lm$ a function $f\colon S^n \to \mathbb{R}$ is called m-block l-separated, if there exist m disjoint discrete intervals $\{a_i, a_i + 1, \dots, b_i\}$ with $a_i, b_i \in \{1, \dots, n\}$, $a_i \le b_i$, such that $f(s_1, \dots, s_n)$ is actually a function of the blocks $\{s_{a_i}, \dots, s_{b_i}\}$ for $i = 1, \dots, m$ and these m blocks are l-separated, meaning that either $a_i - b_j \ge l$ or $a_j - b_i \ge l$ for all $i, j \in \{1, \dots, m\}$ with $i \ne j$. We will assume the following for a specific $m \in \mathbb{N}$ under consideration in Theorem 3.40.

Assumption (H-3). There exist $l \in \mathbb{N}$ and $\beta, \gamma \in [1, \infty)$ such that for all integers $n \ge lm$ and all symmetric m-block l-separated $f \in B(S^n, [0, \infty))$,

$$\mathbb{E}[f(X_1, \dots, X_n)] \le \gamma\,\mathbb{E}\big[f^\beta\big(\widetilde{X}_1^{(1)}, \dots, \widetilde{X}_{b_1}^{(1)}, \widetilde{X}_{b_1+1}^{(2)}, \dots, \widetilde{X}_{b_2}^{(2)}, \dots, \\ \widetilde{X}_{b_{m-1}+1}^{(m)}, \dots, \widetilde{X}_n^{(m)}\big)\big]^{1/\beta}, \quad (3.38)$$

where the processes $\{\widetilde{X}_i^{(1)}\}_{i\in\mathbb{N}}, \dots, \{\widetilde{X}_i^{(m)}\}_{i\in\mathbb{N}}$ are independent copies of $\{X_i\}_{i\in\mathbb{N}}$ and $1 \le b_1 < b_2 < \cdots < b_m \le n$ denote the right endpoints of the m blocks of f.

Remark 3.39. (a) If Assumption (H-1) holds for l, then it holds for all integers $l' \ge l$; a similar statement is true for Assumption (H-3).

(b) If $f \in B(S^n, [0, \infty))$ can be represented as a product in the form $f = f_1 \dots f_m$ with l-separated $f_1, \dots, f_m \in B(S^n, [0, \infty))$, then Hölder's inequality implies (3.38) with $\beta = m$ and $\gamma = 1$.

(c) If $\{X_i\}_{i\in\mathbb{N}}$ is i. i. d., then Assumption (H-3) holds for all $l \in \mathbb{N}$ with $\beta = \gamma = 1$. See Example 3.48 for a stationary Markov chain satisfying Assumption (H-3).

(d) If $m = 1$, then Assumption (H-3) always holds with $\beta = \gamma = 1$.

We get the following result.

Theorem 3.40. *Assume* (H-1), (H-2), (H-3) *for the stationary sequence* $\{X_i\}_{i\in\mathbb{N}}$, *Condition* 1.7 *for* Φ *and reference measure* $\mu^{\otimes m}$, *and, for every* $\varphi \in \Phi$,

$$\sup\nolimits_\tau \mathbb{E}\big[\exp\big(\alpha\|\varphi(X_{\tau(1)}, \dots, X_{\tau(m)})\|_E\big)\big] < \infty \qquad \text{for all } \alpha > 0, \quad (3.41)$$

where $\sup_\tau$ *ranges over all injective* $\tau\colon \{1, \dots, m\} \to \mathbb{N}$. *Then* $\{\mathbb{P}(L_n^m)^{-1}\}_{n\ge m}$ *satisfies the LDP in the* $\tau_1^\Phi(E)$*-topology on* $\mathcal{M}_1^\Phi(S^m)$ *with the good rate function* $I_m\colon \mathcal{M}_1(S^m) \to [0, \infty]$ *defined by*

$$I_m(\nu) = \begin{cases} I_1(\tilde{\nu}) & \text{if } \tilde{\nu}^{\otimes m} = \nu, \\ \infty & \text{otherwise}, \end{cases} \quad (3.42)$$

where

$$I_1(\tilde\nu) \equiv \sup_{f\in B(S,\mathbb{R})} \Bigl(\int_S f\,d\tilde\nu - \Lambda(f)\Bigr), \qquad \tilde\nu \in \mathcal{M}_1(S),$$

and

$$\Lambda(f) \equiv \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\Bigl[\exp\Bigl(\sum_{i=1}^n f(X_i)\Bigr)\Bigr], \qquad f \in B(S,\mathbb{R}). \tag{3.43}$$

In particular, the limit (3.43) *exists for every* $f \in B(S,\mathbb{R})$. *If, in addition, we assume* (3.41) *for all* $\varphi \in \Phi$ *with* $\sup_\tau$ *ranging over all* $\tau\colon \{1,\dots,m\} \to \mathbb{N}$, *then we get the same LDP for the measures* $\{\mathbb{P}(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N}}$ *(with the same rate function).*

Proof. By [5, Theorem 1], the LDP for $\{\mathbb{P}L_n^{-1}\}_{n\in\mathbb{N}}$ holds in the τ-topology, because Assumption (H-2) implies [5, Condition (S)]. Therefore, we can adapt the arguments from the proof of Theorem 3.12; only the crucial estimate (3.17) (without the supremum over $s \in S$), which corresponds to (1.5) in Condition 1.4, has to be verified in the present setting with $\mu \equiv \mathcal{L}(X_1)$.

Using Remark 3.39(a), choose $l \in \mathbb{N}$ according to Assumptions (H-1) and (H-3). For brevity, we will use some notation from the proof of Theorem 3.12. For every $n \ge lm + 1$, define the empirical measure $L_{m,n}\colon \Omega \to \mathcal{M}_1(S^m)$ by (3.15). We want to verify Condition 1.4 for all $R_n \equiv \mathbb{P}L_{m,n}^{-1}$ with $n \ge n_0 \equiv 4lm$ and reference measure $\mu^{\otimes m}$. Let $A_{m,n}$ be given by (3.16), and let $B'_{m,n}$ and $C_{m,n}$ have the same meaning as in (3.24). Consider $\varphi \in B(S^m,[0,\infty))$ and let $\widetilde\varphi$ denote its symmetrized version. Then (3.25) and (3.26) imply, for all $n \ge n_0$,

$$\begin{aligned}&\mathbb{E}\Bigl[\exp\Bigl(n\int_{S^m} \varphi\,dL_{m,n}\Bigr)\Bigr]\\ &\quad\le \prod_{r\in B'_{m,n}} \Bigl(\mathbb{E}\Bigl[\exp\Bigl(4^m l m! \sum_{(i_1,\dots,i_m)\in C_{n,r}} \widetilde\varphi(X_{i_1},\dots,X_{i_m})\Bigr)\Bigr]\Bigr)^{1/|B'_{m,n}|}.\end{aligned} \tag{3.44}$$

Given $r \in B'_{m,n}$, it follows from the definition of $C_{n,r}$ just below (3.23) that

$$f(s_1,\dots,s_n) \equiv \sum_{(i_1,\dots,i_m)\in C_{n,r}} \widetilde\varphi(s_{i_1},\dots,s_{i_m}), \qquad (s_1,\dots,s_n) \in S^n,$$

is symmetric m-block l-separated. Therefore, according to Assumption (H-3), the expectation in the right-hand side of 3.44 is bounded above by

$$\gamma\Bigl(\mathbb{E}\Bigl[\exp\Bigl(4^m\beta l m! \sum_{(i_1,\dots,i_m)\in C_{n,r}} \widetilde\varphi(X^{(1)}_{i_1},\dots,X^{(m)}_{i_m})\Bigr)\Bigr]\Bigr)^{1/\beta}. \tag{3.45}$$

Using Assumption (H-1) for $k \equiv |C_{n,r}|$ and expectation with respect to the process $\{X_i^{(1)}\}_{i\in\mathbb{N}}$, the expectation in (3.45) is bounded above by

$$\mathbb{E}\Bigl[\prod_{(i_1,\dots,i_m)\in C_{n,r}} \Bigl(\int_S \exp\bigl(4^m\alpha\beta l m! \widetilde\varphi(s, X^{(2)}_{i_2},\dots,X^{(m)}_{i_m})\bigr)\,\mu(ds)\Bigr)^{1/\alpha}\Bigr]. \tag{3.46}$$

Applying Assumption (H-1) iteratively for the processes $\{X_i^{(j)}\}_{i\in\mathbb{N}}$ for $j = 2,\dots,m$, we obtain

$$\prod_{(i_1,\dots,i_m)\in C_{n,r}} \Bigl(\int_{S^m} \exp\bigl(4^m\alpha\beta l m! \widetilde\varphi\bigr)\,d\mu^{\otimes m}\Bigr)^{1/\alpha} \tag{3.47}$$

as an upper estimate for (3.46). Putting (3.44)–(3.47) together yields

$$\mathbb{E}\left[\exp\left(n\int_{S^m}\varphi\,dL_{m,n}\right)\right]\le\gamma\left(\int_{S^m}\exp\bigl(4^m\alpha\beta l m!\widetilde{\varphi}\bigr)\,d\mu^{\otimes m}\right)^{|C_{m,r}|/(\alpha\beta)}.$$

Using $|C_{m,r}|/(\alpha\beta)\le n$ and Jensen's inequality to pass back from $\widetilde{\varphi}$ to φ, estimate (1.5) follows. □

Example 3.48. Let $|\alpha|<1$ and $\mathcal{L}(X_1)=\mathcal{N}(0,1/(1-\alpha^2))$ and let $\{Y_i\}_{i\in\mathbb{N}}$ be a sequence of independent standard normal random variables. It is well known that

$$X_{i+1}\equiv\alpha X_i+Y_i,\qquad i\in\mathbb{N},$$

defines a stationary Gaussian process $\{X_i\}_{i\in\mathbb{N}}$ satisfying the hypermixing conditions (H-1) and (H-2) but not condition (3.8) for Markov chains (see [8, Exercise 6.4.19] and the example of the continuous-time Ornstein–Uhlenbeck process discussed in [9]). We want to verify Assumption (H-3) for the process $\{X_i\}_{i\in\mathbb{N}}$.

Denote by $g(\mu,\sigma^2,\cdot)$ the continuous density with respect to Lebesgue measure of the one-dimensional normal distribution $\mathcal{N}(\mu,\sigma^2)$ with expectation $\mu\in\mathbb{R}$ and variance $\sigma^2>0$. The l-step transition kernel $\pi^l(x,\cdot)$ for $\{X_i\}_{i\in\mathbb{N}}$ is given by $\mathcal{N}(\alpha^l x,\alpha_l)$ with $\alpha_l\equiv(1-\alpha^{2l})/(1-\alpha^2)$ for all $x\in\mathbb{R}$ and $l\in\mathbb{N}$. Thus, for $\varphi\in B(\mathbb{R}^2,\mathbb{R})$ and $k\in\mathbb{N}$, using $2\alpha^l y_1y_2\le|\alpha|^l\|y\|_2^2$ for $y=(y_1,y_2)\in\mathbb{R}^2$, we get

$$\begin{aligned}\mathbb{E}[\varphi(X_k,X_{k+l})]&=\int_{\mathbb{R}^2}\varphi(y)\,g\Bigl(0,\frac{1}{1-\alpha^2},y_1\Bigr)\,g(\alpha^l y_1,\alpha_l,y_2)\,dy\\&\le\frac{1-\alpha^2}{2\pi\sqrt{1-\alpha^{2l}}}\int_{\mathbb{R}^2}\varphi(y)\exp\Bigl(-\frac{1-\alpha^2}{1+|\alpha|^l}\frac{\|y\|_2^2}{2}\Bigr)\,dy.\end{aligned}$$

With $\beta_l\equiv2(1+|\alpha|^l)$ and the decomposition

$$\frac{1-\alpha^2}{1+|\alpha|^l}=\eta_l+\frac{1}{\kappa_l}\quad\text{with}\quad\eta_l\equiv\frac{|\alpha|^l}{(1+|\alpha|^l)\kappa_l}\quad\text{and}\quad\kappa_l\equiv\frac{1+2|\alpha|^l}{1-\alpha^2},$$

we can rearrange the right-hand side and then use Jensen's inequality as follows:

$$\begin{aligned}&=\frac{1+2|\alpha|^l}{\sqrt{1-\alpha^{2l}}}\int_{\mathbb{R}^2}\varphi(y)\exp\Bigl(-\eta_l\frac{\|y\|_2^2}{2}\Bigr)g(0,\kappa_l,y_1)\,g(0,\kappa_l,y_2)\,dy\\&\le\frac{1+2|\alpha|^l}{\sqrt{1-\alpha^{2l}}}\biggl(\int_{\mathbb{R}^2}\varphi(y)^{\beta_l}\exp\Bigl(-\beta_l\eta_l\frac{\|y\|_2^2}{2}\Bigr)g(0,\kappa_l,y_1)\,g(0,\kappa_l,y_2)\,dy\biggr)^{1/\beta_l}.\end{aligned}$$

Rearranging, using Jensen's inequality and $\beta_l\le4$ for the last step, we get

$$\begin{aligned}&=\frac{(1+2|\alpha|^l)^{1-1/\beta_l}}{\sqrt{1-\alpha^{2l}}}\biggl(\int_{\mathbb{R}^2}\varphi(y)^{\beta_l}g\Bigl(0,\frac{1}{1-\alpha^2},y_1\Bigr)g\Bigl(0,\frac{1}{1-\alpha^2},y_2\Bigr)\,dy\biggr)^{1/\beta_l}\\&=\frac{(1+2|\alpha|^l)^{1-1/\beta_l}}{\sqrt{1-\alpha^{2l}}}\,\mathbb{E}\bigl[\varphi(X_k,\widetilde{X}_{k+l})^{\beta_l}\bigr]^{1/\beta_l}\\&\le\gamma_\alpha\,\mathbb{E}\bigl[\varphi(X_k,\widetilde{X}_{k+l})^4\bigr]^{1/4}\qquad\text{with}\qquad\gamma_\alpha\equiv3/\sqrt{1-\alpha^2},\end{aligned}\tag{3.49}$$

where $\{\widetilde{X}_i\}_{i\in\mathbb{N}}$ denotes an independent copy of $\{X_i\}_{i\in\mathbb{N}}$.

To reduce the verification of Assumption (H-3) for $m=2$ to the above calculation, consider $n\ge l+1$ and $k\in\mathbb{N}$ such that $1<k<n-l$ (the case $k\in\{1,n-l\}$

is easier). Define two stochastic transition kernels $\Pi_1\colon \mathbb{R}\times\mathcal{B}^{\otimes k-1}\to[0,1]$ and $\Pi_2\colon \mathbb{R}\times\mathcal{B}^{\otimes n-k-l}\to[0,1]$ such that

$$\Pi_1(y_1,A)=\mathbb{P}((X_1,\dots,X_{k-1})\in A\,|\,X_k=y_1),$$
$$\Pi_2(y_2,B)=\mathbb{P}((X_{k+l+1},\dots,X_n)\in B\,|\,X_{k+l}=y_2)$$

for all Borel sets $A\subset\mathbb{R}^{k-1}$ and $B\subset\mathbb{R}^{n-k-l}$ and for almost all $y_1,y_2\in\mathbb{R}$. Due to the Markov property of $\{X_i\}_{i\in\mathbb{N}}$, it follows that for any $g\in B(S^{n-l+1})$

$$\mathbb{E}\big[g(X_1,\dots,X_k,X_{k+l},\dots,X_n)\big]=\mathbb{E}[\varphi(X_k,X_{k+l})],$$

where

$$\varphi(y)\equiv\int_{S^{n-k-l}}\int_{S^{k-1}}g(x,y,z)\,\Pi_1(y_1,dx)\,\Pi_2(y_2,dz)$$

for all $y=(y_1,y_2)\in\mathbb{R}^2$. Using the calculation leading to (3.49) for the first step and Jensen's inequality for the second step, we obtain

$$\begin{aligned}\mathbb{E}[\varphi(X_k,X_{k+l})]&\le\gamma_\alpha\,\mathbb{E}\big[\varphi(X_k,\widetilde{X}_{k+l})^4\big]^{1/4}\\&\le\gamma_\alpha\,\mathbb{E}\bigg[\int_{S^{n-k-l}}\int_{S^{k-1}}g(x,X_k,\widetilde{X}_{k+l},z)^4\,\Pi_1(X_k,dx)\,\Pi_2(\widetilde{X}_{k+l},dz)\bigg]^{1/4}\\&=\gamma_\alpha\,\mathbb{E}\big[g(X_1,\dots,X_k,\widetilde{X}_{k+l},\dots,\widetilde{X}_n)^4\big]^{1/4},\end{aligned}$$

which implies Assumption (H-3) for $m=2$. To verify Assumption (H-3) for an integer $m\ge3$, we iterate the above result, obtaining $\beta=4^m$ and $\gamma=\prod_{p=0}^{m-2}\gamma_\alpha^{4^{-p}}$.

4. Applications to Statistics and Processes

4.1. **Large deviations for *U*- and *V*-statistics.** The aim of this subsection is to establish the LDP for Banach space valued U-statistics and V-statistics of dependent or independent, but not identically distributed, processes.

For a measurable map $\varphi\colon S^m\to E$, the U- and V-statistics of degree m with kernel function φ are defined by

$$U_n^m(\varphi)=\int_{S^m}\varphi\,dL_n^m\quad\text{and}\quad V_n^m(\varphi)=\int_{S^m}\varphi\,dL_n^{\otimes m}\tag{4.1}$$

for all $n\ge m$ and all $n\in\mathbb{N}$, respectively. For a sequence $\{X_i\}_{i\in\mathbb{N}}$ of i. i. d. random variables taking values in a general measurable space S, we proved an LDP for these statistics under exponential moment conditions in [16, Theorem 1.13]. We can adopt the arguments to the present non-i. i. d. setting, where S is assumed to be a Polish space. Define $\Phi=B(S^m,E)\cup\{\varphi\}$. Then the statistics in (4.1) are compositions of L_n^m and $L_n^{\otimes m}$, respectively, with the $\tau_1^\Phi(E)$-continuous functional $\mathcal{M}_1^\Phi(S^m)\ni\nu\mapsto\int_{S^m}\varphi\,d\nu\in E$. Hence, the LDPs of Section 3 combined with the contraction principle [8, Theorem 4.2.1] immediately lead to the following results for U-statistics and V-statistics with dependent inputs.

Theorem 4.2. (a) *The U-statistics $\{U_n^m(\varphi)\}_{n\ge m}$ satisfy an LDP on E with the good rate function J_φ provided one of the following conditions is satisfied:*

(i) *$\{X_i\}_{i\in\mathbb{N}}$ is a sequence of independent random variables satisfying* (3.6),

$$\int_{S^m}\exp(\alpha\|\varphi\|_E)\,d\mu^{\otimes m}<\infty\quad\textit{for all }\alpha>0\tag{4.3}$$

with μ as in (3.6)*, and the corresponding empirical process $\{L_n\}_{n\in\mathbb{N}}$ satisfies an LDP with a rate function J_1 in the weak topology on $\mathcal{M}_1(S)$.*

(ii) $\{X_i\}_{i\in\mathbb{N}}$ *is a Markov chain satisfying* (3.8), (4.3) *with μ given by* (3.11), *and* (3.13) *for all $\alpha > 0$ with $\Theta \equiv \{\mathcal{L}(X_1)\}$ and $\sup_\tau$ ranging over all injective $\tau\colon \{1,\dots,m\} \to \mathbb{N}$.*

(iii) $\{X_i\}_{i\in\mathbb{N}}$ *is an exchangeable sequence satisfying* (a), (b) *and* (d) *of Subsection* 3.4, (4.3) *with μ as in* (d)*, and the mixing parameter set Θ is a weakly compact subset of $\mathcal{M}_1(S)$.*

(iv) $\{X_i\}_{i\in\mathbb{N}}$ *is a stationary sequence satisfying Assumptions* (H-1), (H-2) *and* (H-3) *of Subsection* 3.5, (4.3) *with $\mu = \mathcal{L}(X_1)$, and* (3.41) *with $\sup_\tau$ ranging over all injective $\tau\colon \{1,\dots,m\} \to \mathbb{N}$.*

(b) *The V-statistics $\{V_n^m(\varphi)\}_{n\in\mathbb{N}}$ satisfy an LDP on E with the good rate function J_φ if, in addition, Condition* 1.8 *holds in the cases* (ai)*,* (aiii) *and if* (3.13) *holds in* (aii) *and* (3.41) *in* (aiv) *where $\sup_\tau$ ranges over all maps $\tau\colon \{1,\dots,m\} \to \mathbb{N}$.*

The rate function $J_\varphi\colon E \to [0,\infty]$ is given by

$$J_\varphi(x) = \inf\Big\{ J_m(\nu) \ \Big|\ \nu \in \mathcal{M}_1^\Phi(S^m), \int_{S^m} \varphi\, d\nu = x \Big\}, \qquad x \in E,$$

where J_m is the rate function for the LDP of the laws of $\{L_n^m\}_{n\ge m}$, see (3.1), (3.10)*,* (3.34) *and* (3.42) *for the cases* (ai)–(aiv).

4.2. **Process-level large deviations for Markov chains.** From the results of Subsection 3.3, we get an LDP in the Markovian situation for products of empirical processes via the projective limit approach of Dawson and Gärtner (see [8, Theorem 4.6.1]). Using the setting of Subsection 3.3, we introduce some more notation.

Given $k \in \mathbb{N}$ with $k \ge 2$, the transition probability kernel $\pi_k\colon S^k \times \mathcal{S}^{\otimes k} \to [0,1]$ of the Markov chain $\{(X_i, X_{i+1}, \dots, X_{i+k-1})\}_{i\in\mathbb{N}}$ is given by

$$\pi_k((s_1,\dots,s_k), A) = \int_S 1_A(s_2,\dots,s_k,\sigma)\pi(s_k, d\sigma)$$

for all $(s_1,\dots,s_k) \in S^k$ and $A \in \mathcal{S}^{\otimes k}$. The condition (3.8) for π implies that the kernel π_k satisfies

$$\pi_k^{k+l-1}(s,\cdot) \le \frac{M}{N}\sum_{i=1}^N \pi_k^{i+k-1}(\tilde{s},\cdot) \qquad \text{for all } s,\tilde{s} \in S^k. \tag{4.4}$$

Hence π_k also satisfies (3.8) with $l' \equiv k+l-1$, $M' \equiv M(N+k-1)/N$, $N' \equiv N+k-1$ and $S' \equiv S^k$, but (4.4) is slightly more specific than this version of (3.8). Let

$$\Omega \ni \omega \mapsto L_{k,n}(\omega) \equiv \frac{1}{n}\sum_{i=1}^n \delta_{(X_i(\omega),\dots,X_{i+k-1}(\omega))} \in \mathcal{M}_1(S^k), \qquad n \in \mathbb{N},$$

denote the empirical measures of the above Markov chain. Similar to (3.9) we define the rate function $J_{1,k}\colon \mathcal{M}_1(S^k) \to [0,\infty]$ by

$$J_{1,k}(\tilde{\nu}) = -\inf_{u\in B(S^k,[1,\infty))} \int_{S^k} \log\frac{\pi_k u}{u}\, d\tilde{\nu}.$$

Note that [9, Lemma 4.4.9] gives an alternative expression for $J_{1,k}$ in terms of the relative entropy. Analogously to (3.10), for every integer $m \ge 2$ we define the rate function $J_{m,k}\colon \mathcal{M}_1((S^k)^m) \to [0,\infty]$ by

$$J_{m,k}(\nu) = \begin{cases} J_{1,k}(\tilde\nu), & \text{if } \tilde\nu^{\otimes m} = \nu \\ \infty, & \text{otherwise.} \end{cases} \tag{4.5}$$

In view of (4.4), let $\mu_k \in \mathcal{M}_1(S^k)$ be given by

$$\mu_k = \frac{1}{N}\sum_{i=1}^{N} \tilde\mu_k \pi_k^{i+k-1} \tag{4.6}$$

with an arbitrary $\tilde\mu_k \in \mathcal{M}_1(S^k)$. Given $m \in \mathbb{N}$, let Φ be a set of $(\mathcal{S}^{\otimes k})^{\otimes m}$-$\mathcal{E}$-measurable functions $\varphi\colon (S^k)^m \to E$ containing $B((S^k)^m, E)$. As an immediate consequence of Theorem 3.12 we get the following extension:

Corollary 4.7. *Take a non-empty $\Theta \subset \mathcal{M}_1(S)$. Assume* (3.8)*, Condition* 1.7 *with reference measure $\mu_k^{\otimes m}$ given via* (4.6)*, and, for every $\varphi \in \Phi$,*

$$\sup_{\theta\in\Theta}\sup_{\tau} \mathbb{E}_\theta\bigl[\exp\bigl(\alpha\bigl\|\varphi((X_{\tau(1)},\ldots,X_{\tau(1)+k-1}),\ldots,(X_{\tau(m)},\ldots,X_{\tau(m)+k-1}))\bigr\|_E\bigr)\bigr] < \infty \qquad \text{for all } \alpha > 0, \tag{4.8}$$

where $\sup_\tau$ ranges over all functions $\tau\colon \{1,\ldots,m\} \to \mathbb{N}$. Then the measures $\{\mathbb{P}_\theta(L_{k,n}^{\otimes m})^{-1}\}_{n\in\mathbb{N},\theta\in\Theta}$ satisfy a uniform LDP in the $\tau_1^\Phi(E)$-topology on $\mathcal{M}_1^\Phi((S^k)^m)$ with the good rate function $J_{m,k}$ given by (4.5)*.*

Remark 4.9. Again, condition (3.8) suffices to transfer the LDP to the products $\{L_{k,n}^{\otimes m}\}_{n\in\mathbb{N}}$ if $\mathcal{M}_1((S^k)^m)$ is endowed with the $\tau_1(E)$-topology, see Remark 3.28.

Remark 4.10. As in Remark 3.29, condition (3.8) implies the existence of a unique π-invariant $\tilde\mu \in \mathcal{M}_1(S)$. Then $\tilde\mu_k(ds) \equiv \tilde\mu(ds_1)\pi(s_1, ds_2)\ldots\pi(s_{k-1}, ds_k)$ is the unique π_k-invariant measure and (4.6) gives $\mu_k = \tilde\mu_k$. Hence, Corollary 4.7 assumes Condition 1.7 with respect to the m-fold product of the π_k-invariant distribution.

To extend Theorem 3.12 to the process level, we need to introduce some additional notation and a special topology. Recall that $(\Omega, \mathcal{A}) = (S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}})$. For $n \in \mathbb{N}$ let

$$\Omega \ni \omega \mapsto Q_n(\omega) \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{\theta^{\circ i}(\omega)} \in \mathcal{M}_1(\Omega)$$

denote the *empirical-process measure,* where $\theta\colon \Omega \to \Omega$ is the shift defined by $\theta(\omega) = (\omega_{i+1})_{i\in\mathbb{N}}$ for every $\omega = (\omega_i)_{i\in\mathbb{N}}$. For $k \in \mathbb{N}$ let $\varrho_{1,k}\colon \Omega \to S^k$ with $\varrho_{1,k}(\omega) \equiv (\omega_1,\ldots,\omega_k)$, $\omega = (\omega_i)_{i\in\mathbb{N}} \in \Omega$, denote the projection onto the first k components. Given $m \ge 2$, define $\varrho_{m,k}\colon \Omega^m \to (S^k)^m$ by $\varrho_{m,k}(\omega_1,\ldots,\omega_m) = (\varrho_{1,k}(\omega_1),\ldots,\varrho_{1,k}(\omega_m))$ for all $(\omega_1,\ldots,\omega_m) \in \Omega^m$. Note that

$$Q_n^{\otimes m}\varrho_{m,k}^{-1} = L_{k,n}^{\otimes m} \quad \text{for all } k, m, n \in \mathbb{N}.$$

Similar to what was done in [9, p. 174], we will introduce a *projective-limit $\tau_1^\Phi(E)$-topology* on a subset of the space $\mathcal{M}_1(\Omega^m)$. Given $p \in \mathbb{N}$ and a set Φ of $(\mathcal{S}^{\otimes p})^{\otimes m}$-$\mathcal{E}$-measurable functions from $(S^p)^m$ to E containing $B((S^p)^m, E)$, the set $\mathcal{M}_1^\Phi((S^p)^m)$ and the $\tau_1^\Phi(E)$-topology are defined as in the introduction. Let $\Phi_p \equiv \Phi$, and for every integer $k \ge p$ define Φ_k to be the set consisting of $B((S^k)^m, E)$ and all functions of the form $\varphi \circ \varrho_{k,p}^m$, where $\varphi \in \Phi$ and the projection $\varrho_{k,p}^m\colon S^{km} \to S^{pm}$ is defined by $\varrho_{k,p}^m(s_1,\ldots,s_m) = (\varrho_{k,p}^1(s_1),\ldots,\varrho_{k,p}^1(s_m))$ for $s_1,\ldots,s_m \in S^k$ and

$\varrho_{k,p}^1$: $S^k \to S^p$ is the usual projection map onto the first p coordinates. Using Φ_k, the set $\mathcal{M}_1^{\Phi_k}((S^k)^m)$ and the $\tau_1^{\Phi_k}(E)$-topology can be defined as usual. We observe that the sets $\mathcal{M}_1^{\Phi}((S^p)^m)$ and $\mathcal{M}_1^{\Phi_p}((S^p)^m)$, and their topologies, coincide. The set $\mathcal{M}_1^{\Phi}(\Omega^m)$ is now defined as the set of all $\nu \in \mathcal{M}_1(\Omega^m)$ such that, for every integer $k \geq p$, the measure $\nu\varrho_{m,k}^{-1}$ belongs to $\mathcal{M}_1^{\Phi_k}((S^k)^m)$. The projective-limit $\tau_1^{\Phi}(E)$-topology on $\mathcal{M}_1^{\Phi}(\Omega^m)$ is then defined to be the coarsest topology which, for every integer $k \geq p$, makes the projection $\mathcal{M}_1^{\Phi}(\Omega^m) \ni \nu \mapsto \nu\varrho_{m,k}^{-1} \in \mathcal{M}_1^{\Phi_k}((S^k)^m)$ continuous with respect to the $\tau_1^{\Phi_k}(E)$-topology on $\mathcal{M}_1^{\Phi_k}((S^k)^m)$.

If Φ satisfies Condition 1.7 with respect to the reference measure $\mu_p^{\otimes m}$ given via (4.6), then the same is true for Φ_k with reference measure $\mu_k^{\otimes m}$ for every integer $k \geq p$, where $\mu_k \equiv \mu_p \otimes \pi_p^{\otimes k-p}$.

As in [9, (4.4.11)], define the process-level rate function $J_{1,\infty}$: $\mathcal{M}_1(\Omega) \to [0,\infty]$ by

$$J_{1,\infty}(\nu) = \sup_{k\in\mathbb{N}} J_{1,k}(\nu\varrho_{1,k}^{-1}).$$

An alternative representation for $J_{1,\infty}$ is given in [9, (4.4.16)]. It shows that $J_{1,\infty}(\nu) = \infty$ if ν is not shift-invariant. Similarly to (4.5), for every integer $m \geq 2$, we define the rate function $J_{m,\infty}$: $\mathcal{M}_1(\Omega^m) \to [0,\infty]$ by

$$J_{m,\infty}(\nu) = \begin{cases} J_{1,\infty}(\tilde{\nu}), & \text{if } \tilde{\nu}^{\otimes m} = \nu, \\ \infty, & \text{otherwise.} \end{cases} \tag{4.11}$$

Lemma 4.12. *If $\nu \in \mathcal{M}_1(\Omega^m)$, then*

$$J_{m,\infty}(\nu) = \sup_{k\in\mathbb{N}} J_{m,k}(\nu\varrho_{m,k}^{-1}). \tag{4.13}$$

Proof. If there exists a $\tilde{\nu} \in \mathcal{M}_1(\Omega)$ with $\nu = \tilde{\nu}^{\otimes m}$, then $\nu\varrho_{m,k}^{-1} = (\tilde{\nu}\varrho_{1,k}^{-1})^{\otimes m}$, hence $J_{m,k}(\nu\varrho_{m,k}^{-1}) = J_{1,k}(\tilde{\nu}\varrho_{1,k}^{-1})$ for all $k \in \mathbb{N}$ and (4.13) holds.

Consider $\nu \in \mathcal{M}_1(\Omega^m)$ satisfying $\nu \neq \tilde{\nu}^{\otimes m}$ for all $\tilde{\nu} \in \mathcal{M}_1(\Omega)$ and assume that, for every $k \in \mathbb{N}$, there exists $\tilde{\nu}_k \in \mathcal{M}_1(S^k)$ satisfying $\nu\varrho_{m,k}^{-1} = \tilde{\nu}_k^{\otimes m}$. By Kolmogorov's consistency theorem, the consistent family $\{\tilde{\nu}_k\}_{k\in\mathbb{N}}$ gives rise to a measure $\tilde{\nu}_\infty \in \mathcal{M}_1(\Omega)$ with $\tilde{\nu}_\infty\varrho_{1,k}^{-1} = \tilde{\nu}_k$ for all $k \in \mathbb{N}$. Hence $\nu\varrho_{m,k}^{-1} = \tilde{\nu}_\infty^{\otimes m}\varrho_{m,k}^{-1}$ for all $k \in \mathbb{N}$ and therefore $\nu = \tilde{\nu}_\infty^{\otimes m}$, because the algebra $\bigcup_{k\in\mathbb{N}} \varrho_{m,k}^{-1}((\mathcal{S}^{\otimes k})^{\otimes m})$ generates $\mathcal{A}^{\otimes m}$. This contradicts the assumption on ν, hence there exists $k \in \mathbb{N}$ with $\nu\varrho_{m,k}^{-1} \neq \tilde{\nu}_k^{\otimes m}$ for all $\tilde{\nu}_k \in \mathcal{M}_1(S^k)$. Since $J_{m,\infty}(\nu) = \infty$ by (4.11) and $J_{m,k}(\nu\varrho_{m,k}^{-1}) = \infty$ by (4.5), the identity (4.13) holds. □

Using Corollary 4.7 and Lemma 4.12, we obtain the following result.

Theorem 4.14. *Under the assumptions of Corollary 4.7 for $k = p$, the empirical process-level product measures $\{\mathbb{P}_\theta(Q_n^{\otimes m})^{-1}\}_{n\in\mathbb{N},\theta\in\Theta}$ satisfy a uniform LDP in the projective-limit $\tau_1^{\Phi}(E)$-topology on $\mathcal{M}_1^{\Phi}(\Omega^m)$ with the good rate function $J_{m,\infty}$. In particular, for every $r \in [0,\infty)$, the level set $\{\nu \in \mathcal{M}_1^{\Phi}(\Omega^m) \,|\, J_{m,\infty}(\nu) \leq r\}$ is compact with respect to the projective-limit $\tau_1^{\Phi}(E)$-topology.*

Proof. For every integer $k \geq p$, (4.8) holds for every $\varphi \in \Phi_k$ and Condition 1.7 holds for Φ_k with reference measure $\mu_k^{\otimes m}$ given via $\mu_k \equiv \mu_p \otimes \pi_p^{\otimes k-p}$, hence we can apply Corollary 4.7. In particular, for every $k \geq p$, the rate function $J_{m,k}$ is good with respect to the $\tau_1^{\Phi_k}(E)$-topology on $\mathcal{M}_1^{\Phi_k}((S^k)^m)$, meaning that is has $\tau_1^{\Phi_k}(E)$-compact level sets. Given the representation of $J_{m,\infty}$ from Lemma 4.12, the proof of [9, Theorem 4.4.27] can be transferred to the present setting without difficulties. □

Remark 4.15. If we take $\Phi = B((S^p)^m, E)$, then $\mathcal{M}_1((S^p)^m) = \mathcal{M}_1^\Phi((S^p)^m)$ and the $\tau_1^\Phi(E)$-topology coincides with the $\tau_1(E)$-topology on this set. The above construction leads to the projective-limit $\tau_1(E)$-topology on $\mathcal{M}_1(\Omega^m)$, which is the coarsest topology, such that the projection $\mathcal{M}_1(\Omega^m) \ni \nu \mapsto \nu\varrho_{m,k}^{-1} \in \mathcal{M}_1((S^k)^m)$ is continuous with respect to the $\tau_1(E)$-topology on $\mathcal{M}_1((S^k)^m)$ for every $k \geq p$. In this setting, Theorem 4.14 applies without moment conditions.

References

1. J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill, and B. Weiss, *Strong laws for L- and U-statistics*, Trans. Amer. Math. Soc. **348** (1996), no. 7, 2845–2866.
2. A. de Acosta, *Large deviations for empirical measures of Markov chains*, J. Theoret. Probab. **3** (1990), no. 3, 395–431.
3. 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.
4. H. Bauer, *Probability Theory and Elements of Measure Theory*, second ed., Academic Press, London, 1981.
5. W. Bryc and A. Dembo, *Large deviations and strong mixing*, Ann. Inst. H. Poincaré. Probab. Statist. **32** (1996), no. 4, 549–569.
6. I. Csiszár, *Sanov property, generalized I-projection and a conditional limit theorem*, Ann. Probab. **12** (1984), no. 3, 768–793.
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 O. Zeitouni, *Large Deviations Techniques and Applications*, second ed., Springer-Verlag, New York, 1998.
9. J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Pure and Applied Mathematics, vol. 137, Academic Press, San Diego, 1989.
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. Dupuis and O. Zeitouni, *A nonstandard form of the rate function for the occupation measure of a Markov chain*, Stochastic Process. Appl. **61** (1996), 249–261.
15. P. Eichelsbacher and U. Schmock, *Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem*, Stochastic Process. Appl. **77** (1998), 233–251.
16. ———, *Large deviations of U-empirical measures in strong topologies and applications*, Ann. Inst. H. Poincaré. Probab. Statist. (2002), to appear.
17. W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.
18. J. Neveu, *Discrete-Parameter Martingales*, North-Holland, Amsterdam, 1975.

(Peter Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Postfach 10-0131, D-33501 Bielefeld, Germany
E-mail address: peter@mathematik.uni-bielefeld.de

(Uwe Schmock) Mathematik Departement, ETH Zentrum, CH-8092 Zürich, Switzerland
E-mail address: schmock@math.ethz.ch
URL: http://www.math.ethz.ch/~schmock