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. As a particular application we get large deviation results for U-statistics and V-statistics based on dependent sequences. Furthermore we prove a 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 $(\Omega, \mathcal{A}, \mathbb{P})$ with Polish state space S. The empirical measure is defined by $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where δ_x denotes the probability measure concentrated at $x \in S$. Let $\mathcal{M}_1(S)$ denote the space of Borel probability measures on S.

Let us recall the definition of a large deviations 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 $\varepsilon_n \downarrow 0$, the level sets $\{x \in \mathcal{X} \mid I(x) \le \alpha\}$ are compact for all $\alpha \in [0,\infty)$, and the lower bound

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \ge -I(\mathrm{int}(\Gamma)),$$

and the upper bound

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma) \le -I(\mathrm{cl}(\Gamma))$$

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 $A \subset \mathcal{X}$. Normaly we choose $\varepsilon_n \equiv 1/n$. We say that a sequence of random variables satisfies the LDP provided 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 [9] and [10] *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 in case $\mathcal{M}_1(S)$ is endowed with the τ-topology, i. e. the coarsest topology that makes the maps $\mathcal{M}_1(S) \ni$

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.

$\nu \mapsto \int_S f\,d\nu$ continuous for all f in the space $B(S)$ 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], [7], [8], [13] and references therein) and for *stationary processes* satisfying strong *mixing conditions* (see [4] and references therein).

The aim of this paper is to discuss the LDP for m-fold products of the empirical measure and for U-empirical measures. In [15] we have discussed an LDP for a sequence $\{X_i\}_{i\in\mathbb{N}}$ of independent, identically distributed random variables. 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, the LDP holds for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$, defined by

$$L_n^{\otimes m} = \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} \delta_{(X_{i_1},\dots,X_{i_m})},$$

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 the contraction principle [7, Theorem 4.2.1]. The τ-topology on $\mathcal{M}_1(S^m)$ is defined to be the one induced by $B(S^m)$, the space of all bounded $\mathcal{S}^{\otimes m}$-measurable functions on S^m. Taking product measures can be a discontinuous operation with respect to the τ-topologies (see [7, Exercise 7.3.18]). Moreover the following example (see also [15, Example 1.26]) illustrates, that we cannot expect a LDP to hold for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ for a sequence of dependent $\{X_i\}_{i\in\mathbb{N}}$ in general.

Example 1.1. 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

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

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 $\varphi \in L_1(\mu, E)$ it holds that

$$\mu\Bigl(\lim_{n\to\infty} \int_S \varphi\,dL_n = \int_S \varphi\,d\mu\Bigr) = 1 \tag{1.2}$$

(for a proof see [15, Example 1.26]). To show that the product measures $\{L_n \otimes L_n\}_{n\in\mathbb{N}_0}$ can go astray, we considered 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 (1.2) 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 [15] we considered 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 $k \to \infty$,

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

Note that there are *ergodic, stationary* processes $\{X_i\}_{i\in\mathbb{N}}$ and bounded real-valued functions $\varphi\colon S^m \to \mathbb{R}$ such that $\int_{S^m} \varphi \, dL_n^{\otimes m}$ does not satisfy the *strong law of large numbers*, see [1]. This is a second reason that we cannot expect a LDP to hold for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ for a sequence of dependent $\{X_i\}_{i\in\mathbb{N}}$ in general.

So the question is: which sequences $\{X_i\}_{i\in\mathbb{N}}$ of dependent X_i 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.

In this paper S is assumed to be Polish. Let Φ be a set of $\mathcal{S}^{\otimes m}$-measurable functions $\varphi\colon S^m \to \mathbb{R}$ containing $B(S^m)$. Define

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

Let τ_Φ denote the coarsest topology on $\mathcal{M}_\Phi(S^m)$ such that the map $\mathcal{M}_\Phi(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is continuous for every $\varphi \in \Phi$. If $\Phi = B(S^m)$, then $\mathcal{M}_\Phi(S^m) = \mathcal{M}_1(S^m)$ and τ_Φ 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 $\mathcal{M}_\Phi(S^m)$ and all the maps $\mathcal{M}_1(S^m) \ni \nu \mapsto \int_{S^m} f \, d\nu$ with $f \in B(S^m)$ are measurable. In this topological setting, we will prove a LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ for processes obeying a LDP for $\{L_n\}_{n\in\mathbb{N}}$ in the weak topological setting, if in addition special moment conditions are fulfilled.

We consider the following condition for a sequence $\{R_n\}_{n\in\mathbb{N}} \subseteq \mathcal{M}_1(\mathcal{M}_1(S^m))$ and for the class Φ. For each $n \in \mathbb{N}$ the inner measure of $\mathcal{M}_\Phi(S^m)$ with respect to R_n should be 1.

Condition 1.3. *There exist constants $\beta, M \in [1, \infty)$ and a reference measure $\mu \in \mathcal{M}_1(S^m)$ such that the inequality*

$$\sup_{n\in\mathbb{N}} \left(\int_{\mathcal{M}_1(S^m)} \exp\left(n \int_{S^m} \varphi \, d\nu \right) R_n(d\nu) \right)^{1/n} \leq M \int_{S^m} \exp(\beta\varphi) \, d\mu \tag{1.4}$$

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

Remark 1.5. If Condition 1.3 holds, then, by the monotone convergence theorem, (1.4) 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 1.6. *For every $\varphi \in \Phi$ and every $\alpha > 0$,*

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

where μ is the reference measure of Condition 1.3.

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

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

for every map σ: $\{1, \dots, m\} \to \{1, \dots, m\}$, where π_σ: $S^m \to S^m$ is defined by $\pi_\sigma(s) = (s_{\sigma(1)}, \dots, s_{\sigma(m)})$ for every $s = (s_1, \dots, s_m) \in S^m$, and μ is the reference measure of Condition 1.3.

Conditions 1.6 and 1.7 are satisfied in the case $\Phi = B(S^m)$, thus in the τ-topological setting. We will see that Conditions 1.3 and 1.7 are handy to prove a LDP in the τ_Φ-topology for the m-fold products $L_n^{\otimes m}$. In order to apply our results to U-statistics we will prove a LDP for L_n^m: $\Omega \to \mathcal{M}_1(S^m)$ with $n \geq m$, which are defined by

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

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 1.3 and 1.6. Since $L_n^{\otimes m}$ and L_n^m take values in $\mathcal{M}_\Phi(S^m)$, these mappings are measurable with respect to the introduced σ-algebras.

Condition 1.3 presents the main tool to get the LDP for products we are interested in. If the condition is fulfilled and if $\{R_n\}_{n\in\mathbb{N}}$ satisfies a LDP with respect to the weak topology on $\mathcal{M}_1(S^m)$, we can infer from [8, Lemma 3.2.19 and Theorem 3.2.21], that $\{R_n\}_{n\in\mathbb{N}}$ actually satisfies a LDP in the τ-topology on $\mathcal{M}_1(S^m)$. We will prove, that under Conditions 1.3 and 1.6 (or 1.3 and 1.7), this approach can be generalized to the τ_Φ-topological setting.

Condition 1.3 describes the "amount of dependency" of the underlying process under which a LDP for the laws of the empirical measures is preserved under products in the strong topology introduced above. We observe that in general, the conditions which guarantee the LDP for $\{L_n\}_{n\in\mathbb{N}}$ are not sufficient for the LDP for the products. We will analyze Markovian, exchangeable, strong mixing and independent but not identically distributed sequences. In all these cases we want to establish the crucial estimate (1.4). In some cases we are not able to check this estimate for the law of L_n^m. Constructing a suitable modification of L_n^m and proving (1.4) therefore, the remaining part of the proof is to verify that these modifications have a large deviation behavior like L_n^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 a LDP in the weak topology with a rate function I, and if Condition 1.3 ($m = 1$) and Condition 1.6 ($m = 1$) hold for the class Φ introduced above, then the LDP holds in the τ_Φ-topology with the same rate function. Since with S the product S^m is 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 τ_Φ-topology needed to transfer the LDP for the above mentioned modifications to the LDP for L_n^m and $L_n^{\otimes m}$ respectively. In Section 3, (1.4) is verified for different dependent or independent but not identically distributed processes $\{X_i\}_{i\in\mathbb{N}}$. Finally in Section 4, we derive from the results of Section2 2 and 3 LDP results for U-statistics and V-statistics as well as LDP results for products of empirical processes.

2. Transferring LDP's to the τ_Φ-topology and exponential equivalence

There are some non-trivial problems in transferring Lemma 3.2.19 and Theorem 3.2.21 in [8] to the τ_Φ-topology. They are treated by using some technical results of [15] combined with an application of Lusin's theorem and the outline of the proof of Lemma 3.2.19 in [8]. Let Φ be a fixed set of $\mathcal{S}$-measurable functions $\varphi\colon S \to \mathbb{R}$ containing $B(S)$ and define $\mathcal{M}_\Phi(S)$ and the τ_Φ-topology as in the introduction. Then we get the following lemma:

Lemma 2.1. *Assume that $\{R_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$ and for each $n \in \mathbb{N}$ the inner measure of $\mathcal{M}_\Phi(S)$ with respect to R_n is 1. Assume that Φ satisfies Conditions 1.3 and 1.6 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 $B \subset \mathcal{M}_1(S)$*

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

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

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

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

where $\mathrm{cl}_{\tau_\Phi}(B)$ denotes the closure of the set $B \cap \mathcal{M}_\Phi(S)$ with respect to the τ_Φ-topology.

(c) *$K_r \equiv \{\, \nu \in \mathcal{M}_\Phi(S) \mid I(\nu) \leq r \,\} \subset \mathcal{M}_\Phi(S)$ and K_r is τ_Φ-compact for every $r \in [0,\infty)$.*

Proof of Lemma 2.1(c). Using Condition 1.3, we get applying [8, Lemma 3.2.7 and Lemma 3.2.19] that

$$H(\nu\,|\,\mu) \leq \beta(I(\nu) + \log(2M)), \quad \nu \in \mathcal{M}_1(S), \tag{2.2}$$

with β, M and μ as in (1.4). Here H denotes the relative entropy

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

Define $\widetilde{K}_r = \{\, \nu \in \mathcal{M}_1(S) \mid H(\nu\,|\,\mu) \leq r \,\}$ for $r \in [0,\infty)$. Since Condition 1.6 holds, $\widetilde{K}_r \subset \mathcal{M}_\Phi(S)$ and $\widetilde{K}_r$ is τ_Φ-compact by [15, Lemma 2.1(c)] for every $r \in [0,\infty)$. Since I is lower semi-continuous in the weak topology, K_r is τ_Φ-closed. Since $K_r \subset \widetilde{K}_{\beta(r+\log 2M)}$ by (2.2), the τ_Φ-compactness of K_r follows for every $r \in [0,\infty)$. $\square$

Before we will prove (a) and (b) of Lemma 2.1, we prove an additional lemma. We need some more notation. Let $\mathcal{F}$ denote the family of all finite, nonempty subsets of Φ. For every $F \in \mathcal{F}$ define

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

Lemma 2.3. *Assume the general hypotheses of Lemma 2.1. Then for every $F \in \mathcal{F}$ and every Borel set A of $\mathbb{R}^F$, we get*

$$\liminf_{n\to\infty} \frac{1}{n} \log R_n(\{ \nu \in \mathcal{M}_\Phi(S) \mid \Pi_F(\nu) \in A \}) \geq -\inf\{ I(\nu) \mid \nu \in \mathcal{M}_\Phi(S), \Pi_F(\nu) \in A^\circ \}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log R_n(\{ \nu \in \mathcal{M}_\Phi(S) \mid \Pi_F(\nu) \in A \}) \leq -\inf\{ I(\nu) \mid \nu \in \mathcal{M}_\Phi(S), \Pi_F(\nu) \in \bar{A} \},$$

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

Proof of Lemma 2.3. For every $L \in \mathbb{N}$ and $\varphi \in F$ define $\varphi_L = 1_{\{|\varphi|\leq L\}}\varphi$. According to Lusin's theorem and Tietze's extension theorem there exists a bounded and continuous function φ_L^{c} such that $|\varphi_L^{\mathrm{c}}| \leq L$ and $\mu(\{|\varphi_L - \varphi_L^{\mathrm{c}}| \geq 1/L\}) \leq \exp(-L^2)$. Denote by $F_L^{\mathrm{c}} = \{\varphi_L^{\mathrm{c}}\}_{\varphi\in F}$, then $\Pi_{F_L^{\mathrm{c}}} \in C(\mathcal{M}_\Phi(S), \mathbb{R}^F)$ for each $L \in \mathbb{N}$, where $C((\mathcal{M}_\Phi(S), \mathbb{R}^F)$ denotes the space of all continuous, $\mathbb{R}^F$-valued measurable maps on $\mathcal{M}_\Phi(S)$. Let $V_L \equiv \sum_{\varphi\in F} |\varphi - \varphi_L|$, then $V_L \downarrow 0$ pointwise for $L \to \infty$. By Hölder's inequality and Condition 1.6 we get $\int_S \exp(\alpha|V_L|)\, d\mu < \infty$ for any $\alpha > 0$ and $L \in \mathbb{N}$. Since $|\varphi - \varphi_L| \leq |\varphi|$ for each $\varphi \in F$ and since $\int_S \exp(\alpha \sum_{\varphi\in F} |\varphi|)\, d\mu < \infty$ by applying again Hölder's inequality and Condition 1.6, the proof of [8, Lemma 3.2.7] can be adapted via the dominated convergence theorem and we get

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log R_n\big(\{ \nu \in \mathcal{M}_1(S) \mid \|\Pi_{F_L}(\nu)\| \geq \delta \}\big) = -\infty \tag{2.4}$$

for every $\delta > 0$, where $F_L \equiv \{|\varphi - \varphi_L|\}_{\varphi\in F}$ and $\|x\|_1 \equiv \sum_{i\in F} |x_i|$ for $x = (x_i)_{i\in F} \in \mathbb{R}^F$. For each $L \in \mathbb{N}$ and $\varphi \in F$ write $|\varphi_L - \varphi_L^{\mathrm{c}}| = \phi_L^1 + \phi_L^2$ with $0 \leq \phi_L^1 \leq 1/L$ and $\mu(\{\phi_L^2 \neq 0\}) \leq \exp(-L^2)$ and $\phi_L^2 \geq 0$. With $F_1 \equiv \{\phi_L^1\}_{\varphi\in F}$ we get

$$R_n\big(\{ \nu \in \mathcal{M}_1(S) \mid \|\Pi_{F_1}(\nu)\|_1 \geq \delta \}\big) = 0 \tag{2.5}$$

for every $L > |F|/\delta$ and by Condition 1.3 we get for each $\varphi \in F$ and each $\theta > 0$

$$R_n\Big(\nu \in \mathcal{M}_1(S) \Bigm| \int_S \phi_L^2\, d\nu \geq \delta\Big) \leq \Big(M \exp(-\theta\delta) \int_S \exp(\beta\theta\phi_L^2)\, d\mu\Big)^n$$

and thus

$$\frac{1}{n} \log R_n\Big(\nu \in \mathcal{M}_1(S) \Big| \int_S \phi_L^2\, d\nu \geq \delta\Big) \leq \log M - \theta\delta + \log\big(1 + \exp(L(2\beta\theta - L))\big).$$

Letting $L \to \infty$ first and then $\theta \to \infty$ and using (2.4) and (2.5) we get

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log R_n\big(\{ \nu \in \mathcal{M}_(S) \mid \|\Pi_{F-F_L^{\mathrm{c}}}(\nu)\|_1 \geq \delta \}\big) = -\infty \tag{2.6}$$

for any $\delta > 0$, where $F - F_L^{\mathrm{c}} \equiv \{|\varphi - \varphi_L^{\mathrm{c}}|\}_{\varphi\in F}$. The space $\mathcal{M}_1(S)$ equipped with the weak topology is Polish, thus we will apply the contraction principle in the form of [8, Lemma 2.1.4]: By assumption the inner measure of $\mathcal{M}_\Phi(S)$ with respect to R_n is 1 for each $n \in \mathbb{N}$. So the LDP holds for $\{R_n\}_{n\in\mathbb{N}}$ on $\mathcal{M}_\Phi(S)$ with respect to the weak topology with the same rate function I. That is a consequence of [7, Lemma 4.1.5(b)], where we are using $K_r \subset \mathcal{M}_\Phi(S)$ for each $r \in [0, \infty)$, proved in Lemma 2.1(c). Thus it is sufficient to define Π_F on $\mathcal{M}_\Phi(S)$. Now it is enough to check in addition to (2.6):

$$\limsup_{L\to\infty} \sup\{ \|\Pi_{F-F_L^c}(\nu)\|_1,\ \nu\colon I(\nu)\le C \} = 0 \tag{2.7}$$

for each $C\in(0,\infty)$. The well known estimate $xy \le e^{x-1} + y\log y$ for all $x\in\mathbb{R}$ and $y\in[0,\infty)$ leads to $\int_S |\varphi|\,d\nu < \infty$ for each $\varphi\in F$ and $\nu\in\mathcal{M}_\Phi(S)$ with $H(\nu\,|\,\mu)<\infty$. Thus $\int_S V_L\,d\nu \to 0$ for $L\to\infty$ uniformly over all measures ν with $I(\nu)\le C$, since $\{\nu\colon H(\nu\,|\,\mu)\le C\}$ is uniformly absolutely continuous with respect to μ for every $C>0$, see [8, Exercise 3.2.23]. Therefore (2.7) follows if we prove

$$\limsup_{L\to\infty} \sup\{ \|\Pi_{F_1+F_2}(\nu)\|,\ \nu\colon I(\nu)\le C \} = 0, \tag{2.8}$$

where $F_1+F_2 = \{\phi_L^1+\phi_L^2\}_{\varphi\in F}$. Clearly $\limsup_{L\to\infty}(\sum_{\nu\in F}\int_S \phi_L^1\,d\nu) = 0$ uniformly on each level set of I. With $xy \le e^{\sigma x} + \frac{y}{\sigma}\log\frac{y}{\sigma}$ for all $x,y\ge 0$ and $\sigma>0$ we obtain for $\sigma\ge 1$ and any $\varphi\in F$

$$\int_S \phi_L^2\,d\nu \le \int_{\{\phi_L^2\neq 0\}} \exp(\sigma\phi_L^2)\,d\mu + \frac{1}{\sigma}\int_S \Big|\log\frac{d\nu}{d\mu}\Big|\,d\nu.$$

By [11, Lemma 5.1 and its proof] the second term on the right-hand side in the last inequality converges to zero as $\sigma\to\infty$ uniformly on each level set of $H(\cdot\,|\,\mu)$, and therefore on each level set of I by (2.2). By construction of ϕ_L^2

$$\int_{\{\phi_L^2\neq 0\}} \exp(\sigma\phi_L^2)\,d\mu \le \exp(L(\sigma-L)),$$

which converges to zero in L for fixed σ and thus we arrive at (2.8) and the lemma is proved. $\square$

Proof of Lemma 2.1(a) and (b). (a) It suffices to consider the case $I(\mathrm{int}_{\tau_\Phi}(B)) < \infty$. By definition of the τ_Φ-topology, there exists for $\nu\in\mathrm{int}_{\tau_\Phi}(B)$ an $\varepsilon>0$ and an $F\in\mathcal{F}$ such that the τ_Φ-open set $C \equiv \big\{\tilde\nu\in\mathcal{M}_\Phi(S)\colon \|\Pi_F(\tilde\nu)-\Pi_F(\nu)\|_1 < \varepsilon\big\}$ is contained in $\mathrm{int}_{\tau_\Phi}(B)$ and $\nu\in C$. With the open set

$$A \equiv \big\{x\in\mathbb{R}^F\colon \|x-\Pi_F(\nu)\|_1<\varepsilon\big\}$$

we get $\nu\in C = \{\tilde\nu\in\mathcal{M}_\Phi(S)\mid \Pi_F(\tilde\nu)\in A\}\subset B$ and by Lemma 2.3

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\log R_n(B) &\ge \liminf_{n\to\infty}\frac{1}{n}\log R_n(C)\\ &\ge -\inf\{I(\tilde\nu)\mid \tilde\nu\in\mathcal{M}_\Phi(S),\ \Pi_F(\tilde\nu)\in A\}\ge -I(\nu).\end{aligned}$$

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

(b) It suffices to consider the case $q\equiv I(\mathrm{cl}_{\tau_\Phi}(B))>0$. Choose $r\in(0,q)$. By Lemma 2.1(c), the set K_r is contained in $\mathcal{M}_\Phi(S)$. Since $\mathrm{cl}_{\tau_\Phi}(B)\cap K_r=\varnothing$, there exist, for every $\nu\in K_r$, an $F_\nu\in\mathcal{F}$ and an open neighborhood $U_\nu\subset\mathbb{R}^{F_\nu}$ of $\Pi_{F_\nu}(\nu)$ such that $\mathrm{cl}_{\tau_\Phi}(B)\cap\Pi_{F_\nu}^{-1}(U_\nu)=\varnothing$. Since K_r is τ_Φ-compact, there exists a finite subset N of K_r such that $\bigcup_{\nu\in N}\Pi_{F_\nu}^{-1}(U_\nu)$ covers K_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 $F'\neq\varnothing$ $\Pi_{F,F'}\colon\mathbb{R}^F\to\mathbb{R}^{F'}$ denotes the canonical projection. Note that $U_\nu'\subset\mathbb{R}^F$ 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_r and is disjoint from $\mathrm{cl}_{\tau_\Phi}(B)$. Define $\varepsilon=\mathrm{dist}(\Pi_F(K_r),U^{\mathrm{c}})$. Since $\Pi_F(K_r)$ is a compact subset of the open set U, it follows that $\varepsilon>0$ and that

$$A_\varepsilon \equiv \{x\in\mathbb{R}^F \mid \mathrm{dist}(x,\Pi_F(K_r))<\varepsilon\}$$

is an open set contained in U. Therefore

$$\mathrm{cl}_{\tau_\Phi}(B) \subset \{\, \nu \in \mathcal{M}_\Phi(S) \mid \Pi_F(\nu) \in \mathbb{R}^F \setminus A_\varepsilon \,\} \subset K_r^{\mathrm{c}}. \tag{2.9}$$

By the upper bound in Lemma 2.3 it follows that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log R_n(B) &\le -\inf\{\, I(\nu) \mid \nu \in \mathcal{M}_\Phi(S),\, \Pi_F(\nu) \in \mathbb{R}^F \setminus A_\varepsilon \,\} \\ &\le -I(K_r^{\mathrm{c}}) \le -r.\end{aligned}$$

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

Remark 2.10. Lemma 2.1 holds, if we replace S by S^m, $m \in \mathbb{N}$, and assume Conditions 1.3 and 1.6. Even in the case $m = 1$ applying Lemma 2.1 we already improved some LDP results for the laws of the empirical measures L_n for some dependent sequences, which already exist in the τ-topology. The reason is that one can verify Condition 1.3 in several cases for all non-negative, bounded $\mathcal{S}$-measurable functions $f\colon S \to \mathbb{R}$. For example we get the LDP for the laws of L_n, when the $\{X_i\}_{i\in\mathbb{N}}$ are Markovian and satisfy [8, Condition (U) in Section 4.1], compare also with [8, Exercise 4.1.53]. If the sequence $\{X_i\}_{i\in\mathbb{N}}$ is stationary and satisfies Assumptions (H-1) and (H-2) in [7, Section 6.4.2], we can transfer the LDP to the τ_Φ-topology, too (see [7, Lemma 6.4.18]).

Remark 2.11. If Condition 1.6 holds, we get from [15, Lemma 2.1(b)] and from (2.2) that $K_\infty \equiv \bigcup_{r>0} K_r \subset \mathcal{M}_\Phi(S)$, where K_r is defined as in Lemma 2.1(c). Since $I(\nu) = \infty$ for all $\nu \in \mathcal{M}_\Phi(S) \setminus K_\infty$, it follows that the rate function defined in the statement of Lemma 2.3 coincides with

$$\inf\{\, I(\nu) \mid \nu \in K_\infty,\, \Pi_F(\nu) = x \,\}, \quad x \in \mathbb{R}^F.$$

To transfer the LDP to the τ_Φ-topology, it will sometimes be necessary to cancel some summands of $L_n^{\otimes m}$ or L_n^m in order to be able to verify Condition 1.3. Thus we will establish the LDP for such reduced products in the τ_Φ-topology. Next we consider the question, 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 products. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Consider two sequences $\{S_n\}_{n\in\mathbb{N}}$ and $\{\widetilde{S}_n\}_{n\in\mathbb{N}}$ 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 $\{\widetilde{R}_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_1(\mathcal{M}_1(S))$, respectively.

Lemma 2.12. *Assume that $\{R_n\}_{n\in\mathbb{N}}$ obeys a LDP in the τ_Φ-topology on $\mathcal{M}_\Phi(S)$ and for each $n \in \mathbb{N}$ the inner measure of $\mathcal{M}_\Phi(S)$ with respect to R_n is 1. Assume that*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left| \int_S \varphi \, dS_n(\omega) - \int_S \varphi \, d\widetilde{S}_n(\omega) \right| \ge \varepsilon \right) = -\infty \tag{2.13}$$

for every $\varepsilon > 0$ and every $\varphi \in \Phi$. Then the same LDP (with the same rate function) holds for $\{\widetilde{R}_n\}_{n\in\mathbb{N}}$.

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

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

For $\mu \ne \nu$ there is an $F \in \mathcal{F}$ such that $d_F(\mu,\nu) \ne 0$. Therefore, the family $\mathcal{D} \equiv \{d_F\}_{F\in\mathcal{F}}$ is called separating. The topology having as a sub-basis the family of balls $\{B(\nu, d_F, \varepsilon) \mid \nu \in \mathcal{M}_\Phi(S),\ d_F \in \mathcal{D},\ \varepsilon > 0\}$, where $B(\nu, d_F, \varepsilon) = \{\mu \in \mathcal{M}_\Phi(S) \mid$

$d_F(\mu,\nu) < \varepsilon\}$, is the τ_Φ-topology. Therefore $(\mathcal{M}_\Phi(S), \tau_\Phi)$ is a so called gauge space and thus is a completely regular space (see for example [12, Theorem 10.6, Chapter IX]). For completely regular topological spaces, the concept of exponential equivalence was introduced in [14] and thus the lemma was proved. □

We get LDP results in the τ-topology by verifying Condition 1.3. 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 metric implies convergence in τ. Clearly convergence in the total variation metric is not sufficient for convergence in the τ_Φ-topology.

Remark 2.14. If for every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that $\int_S \exp(\alpha_\varphi|\varphi|)\, d\mu < \infty$, then the topology of convergence in information, that is $\mu_n \to \mu$ when $H(\mu_n\,|\,\mu) \to 0$ as $n \to \infty$, is finer than τ_Φ on $\mathcal{M}_\Phi(S)$. Since the level sets of $H(\cdot\,|\,\mu)$ are subsets of $\mathcal{M}_\Phi(S)$ by [15, Lemma 2.1(b)], we get this result using [5, Lemma 3.1]. Note that convergence in information implies convergence in total variation.

3. Large deviations for products

Let us first give some remarks on the case, where $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of independent, identically distributed random variables with common law $\mu \in \mathcal{M}_1(S)$. We get the LDP of L_n^m under Condition 1.6 with measure μ via the useful Hoeffding decomposition, introduced in [16]. To be more precise, 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 $\mathcal{M}_1(S^m)$ to get the LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ and thus for $\{L_n^m\}_{n\geq m}$ because

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

where $\|\cdot\|_{\mathrm{var}}$ denotes the total variation distance on $\mathcal{M}_1(S)$. For $\{L_n^m\}_{n\geq m}$ we can verify Condition 1.3 via the Hoeffding argument, if Condition 1.6 holds. Under Condition 1.7, we can prove that (2.13) holds for the laws of $L_n^{\otimes m}$ and L_n^m and thus we get the LDP for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ under Condition 1.7 in the τ_Φ-topology. The results in [15] are stronger. There S is assumed to be an arbitrary measurable space and the topology is chosen such that integration with respect to certain unbounded, Banach-space valued functions is a continuous operation. Several technical reasons lead us to the Polish setting in this paper.

3.1. **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 $\mathcal{L}(X_i) = \nu_i$. Assume that $\nu_i \ll \tilde{\mu}$ for all $i \in \mathbb{N}$ with a fixed reference measure $\tilde{\mu} \in \mathcal{M}_1(S)$. Moreover we assume that there exists a $q > 1$ such that for $f_i \equiv d\nu_i/d\tilde{\mu}$ we have

$$M \equiv \sup_{i\in\mathbb{N}} \|f_i\|_q < \infty, \tag{3.1}$$

where $\|\cdot\|_q$ denotes the q-norm in $L_q(S, \mathcal{S}, \tilde{\mu})$. For every bounded measurable $\varphi\colon S^m \to [0,\infty)$, we get by Hoeffding's decomposition [16, Section 5]

$$\mathbb{E}\Bigg[\exp\bigg(n\int_{S^m} \varphi\, dL_n^m\bigg)\Bigg] = \mathbb{E}\Bigg[\exp\bigg(\frac{1}{n!}\sum_\sigma \frac{n}{k}\sum_{i=1}^{k-1} \varphi(X_{\sigma(im+1)}, \ldots, X_{\sigma(im+m)})\bigg)\Bigg],$$

where $k \equiv \lfloor n/m \rfloor$ and σ runs through all permutations of $\{1, \ldots, n\}$. Using Jensen's inequality to handle the first convex combination, independence of the terms in the second sum, $n/k \leq 2m$, and

$$\begin{aligned}\mathbb{E}\big[\exp\big(2m\varphi(X_{\sigma(im+1)}, \ldots, X_{\sigma(im+m)})\big)\big] \\ \leq \biggl(\int_{S^m} \exp(2pm\varphi)\, d\tilde{\mu}^{\otimes m}\biggr)^{1/p} \biggl(\prod_{j=1}^{m} \int_S f^q_{\sigma(im+j)}\, d\tilde{\mu}\biggr)^{1/q},\end{aligned}$$

which follows from Hölder's inequality with $1/p + 1/q = 1$, we obtain with (3.1)

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

Since $M \geq 1$ by Jensen's inequality, $km \leq n$, and $k/p \leq n$, Condition 1.3 holds with $\mu \equiv \tilde{\mu}^{\otimes m}$ and $\beta \equiv 2mp$. Thus the LDP for $\{L_n^m\}_{n \geq m}$ is proved as in the i. i. d. case, whenever we assume that the sequence of laws of L_n, built with the $\{X_i\}_{i\in\mathbb{N}}$, already satisfies a LDP in the weak topology. If we assume Condition 1.7, then we get the result for $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ via the Chebyshev–Markov inequality.

3.2. **Markov chains.** Let $\pi\colon \mathcal{S} \times \mathcal{S} \to [0,1]$ be a probability transition kernel and let $\{\mathbb{P}_s\}_{s\in S}$ be the family of Markovian measures on the sequence space $(\Omega, \mathcal{F}) = (S^{\mathbb{N}_0}, \mathcal{S}^{\mathbb{N}_0})$ such that for every $s \in S$, the projections $\{X_i\}_{i\in\mathbb{N}_0}$ from Ω to S form a Markov chain with transition kernel π and $\mathbb{P}_s(X_0 = s) = 1$. We assume that there exist $N \in \mathbb{N}$, $l \in \{1, 2, \ldots, N\}$ and $M \in [1, \infty)$ satisfying

$$\pi^l(s, \cdot) \leq \frac{M}{N} \sum_{m=1}^{N} \pi^m(\tilde{s}, \cdot) \tag{3.2}$$

for all $s, \tilde{s} \in S$. As in [8, (4.1.39)] define the rate function $J_1\colon \mathcal{M}_1(S) \to [0,\infty]$ by

$$J_1(\mu) = -\inf_{u\in B(S,[1,\infty))} \int_S \log \frac{\pi u}{u}\, d\mu. \tag{3.3}$$

For every integer $m \geq 2$ define $J_m\colon \mathcal{M}_1(S^m) \to [0,\infty]$ by

$$J_m(\mu) = \begin{cases} J_1(\nu) & \text{if } \nu^{\otimes m} = \mu, \\ \infty & \text{otherwise.} \end{cases} \tag{3.4}$$

Let $\mu \in \mathcal{M}_1(S)$ be given by

$$\mu = \frac{1}{N} \sum_{k=1}^{N} \pi^k(s_0, \cdot) \tag{3.5}$$

with an arbitrarily chosen $s_0 \in S$. We will prove:

Theorem 3.6. *Assume that Condition 1.6 holds for Φ and with the integration with respect to the measure $\mathbb{P}_s$ uniformly in $s \in S$, then for every $m \in \mathbb{N}$ the measures $\{\mathbb{P}_s(L_n^m)^{-1}\}_{n\in\mathbb{N}, s\in S}$ satisfy a uniform LDP in the τ_Φ-topology on $\mathcal{M}_\Phi(S^m)$ with a good rate function J_m, i. e., for every measurable $B \subset \mathcal{M}_1(S^m)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{s\in S} \mathbb{P}_s(L_n^m \in B) \geq -\inf\{\, J_m(\nu) \mid \nu \in \mathcal{M}_\Phi(S^m),\, \nu \in \operatorname{int}_{\tau_\Phi}(B)\,\}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{s\in S} \mathbb{P}_s(L_n^m \in B) \leq -\inf\{\, J_m(\nu) \mid \nu \in \mathcal{M}_\Phi(S^m),\, \nu \in \operatorname{cl}_{\tau_\Phi}(B)\,\},$$

and the level set $\{\,\nu \in \mathcal{M}_\Phi(S^m) \mid J_m(\nu) \le r\,\}$ *is* τ_Φ*-compact for every* $r \in [0,\infty)$*. If we assume Condition 1.7 for* Φ *and with the integration with respect to the measure* $\mathbb{P}_s$ *uniformly in* $s \in S$*, we get the same result for the measures* $\{\mathbb{P}_s(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N},s\in S}$ *(with the same rate function).*

Proof. By [8, Theorem 4.1.43], the theorem is true in the weak topology for $m = 1$. Furthermore, by [8, Exercise 4.1.53] the theorem is true in the strong topology for $m = 1$ and by Remark 2.10(a), the theorem is true in the τ_Φ-topology, because Condition 1.3 can be verified uniformly in $s \in S$ with respect to the reference measure μ defined in (3.5). Here we use the simple fact, that the assumption for $\mathbb{P}_s$ implies the same for μ. We may assume in the following that $m \ge 2$. For $n \ge ml$ define the empirical measure

$$\Omega \ni \omega \mapsto L_{n,m}(\omega) = \frac{1}{|A_{m,n}|} \sum_{(i_1,\dots,i_m)\in A_{m,n}} \delta_{(X_{i_1}(\omega),\dots,X_{i_m}(\omega))} \in \mathcal{M}_1(S^m),$$

where

$$A_{m,n} \equiv \{\,(i_1,\dots,i_m) \in \{l, l+1,\dots,n\}^m \mid \\ |i_j - i_k| \ge l \text{ for all } j,k \in \{1,\dots,d\} \text{ with } j \ne k\,\}. \quad (3.7)$$

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

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

as $n \to \infty$. Hence, the measures $\{\mathbb{P}_s(L_n^{\otimes m})^{-1}\}_{s\in S,n\ge ml}$ are uniformly exponentially equivalent to $\{\mathbb{P}_s(L_{n,m})^{-1}\}_{s\in S,n\ge ml}$ in the sense that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{s\in S} \mathbb{P}_s\bigl(\varrho(L_n^{\otimes m}, L_{n,m}) > \varepsilon\bigr) = -\infty$$

for every $\varepsilon > 0$, if ϱ denotes the Prohorov metric on $\mathcal{M}_1(S^m)$. That is because the Prohorov metric is bounded by the total variation metric. Now the proof of [7, Theorem 4.2.16] shows that the measures $\{\mathbb{P}_s(L_{n,m})^{-1}\}_{s\in S,n\ge ml}$ satisfy a uniform LDP in the weak topology with rate J_m, because the measures $\{\mathbb{P}_s(L_n^{\otimes m})^{-1}\}_{s\in S,n\ge ml}$ do so via the contraction principle. Next we want to establish the estimate

$$\sup_{n\ge 4ml} \left(\sup_{s\in S} \mathbb{E}_s \left[\exp\left(n \int_{S^m} \varphi \, dL_{n,m} \right) \right] \right)^{1/n} \le M^m \int_{S^m} \exp(4^m l m! \varphi) \, d\mu^{\otimes m} \quad (3.8)$$

for every bounded measurable $\varphi\colon S^m \to [0,\infty)$, where $\mu \in \mathcal{M}_1(S)$ is defined by (3.5). We will get this estimate if Condition 1.6 holds for $\mu^{\otimes m}$ and Φ, which is an immediate consequence of our assumptions. The theorem with $L_{n,m}$ in place of L_n^m (or $L_n^{\otimes m}$) then follows using Lemma 2.1. To obtain the theorem from the corresponding result for $\{L_{n,m}\}_{n\ge ml}$ use finally Lemma 2.12. Therefore we have to check for every $\varepsilon > 0$ and every $\varphi \in \Phi$ that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{s\in S} \mathbb{P}_s \left(\left| \int_{S^m} \varphi \, dL_{n,m} - \int_{S^m} \varphi \, dL_n^m \right| \ge \varepsilon \right) = -\infty$$

(the same for L_n^m replaced by $L_n^{\otimes m}$). But using the assumptions and the exponential Markov–Chebychev inequality, this is easily seen.

In the remaining part of this proof we want to establish the crucial estimate (3.8). Let us first introduce some additional notation. With $n \geq lm$ and $A_{m,n}$ given by (3.7), let

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

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

$$B_{m,n} = \{l, l+1, \ldots, 3l-1\} \times \{2l, 2l+1, \ldots, 2n\}^{m-1}.$$

Given $r = (r_1, \ldots, r_m) \in B_{m,n}$, let $C_n(r)$ denote the set of all $(i_1, \ldots, i_m) \in A'_{m,n}$ for which there exists $k \in \mathbb{N}_0$ satisfying $i_1 = r_2 - r_1 - kl$ and $i_j = r_j + (-1)^j kl$ for all $j \in \{2, 3, \ldots, m\}$. Every set $C_n(r)$ has the following two properties:

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

In order to show that

$$A'_{m,n} \subset \bigcup_{r \in B_{m,n}} C_n(r) \tag{3.9}$$

take any $(i_1, \ldots, i_m) \in A'_{m,n}$. Then there exist $r_1 \in \{l, l+1, \ldots, 3l-1\}$ and $k \in \{0, 1, \ldots, \lfloor n/2l \rfloor - 1\}$ such that $i_2 - i_1 = r_1 + 2kl$. For every $j \in \{2, 3, \ldots, m\}$ define $r_j = i_j - (-1)^j kl$. Since $i_j \geq i_2$ it follows that $r_j \geq i_2 - kl = i_1 + r_1 + kl \geq 2l$. Since $i_j \leq n$ and $k \leq n/l$, it follows that $r_j \leq n + kl \leq 2n$. Hence $(r_1, \ldots, r_m) \in B_{m,n}$.

Define $B'_{m,n} = \{ r \in B_{m,n} \mid C_n(r) \neq \varnothing \}$ and let $S(m)$ denote the set of all permutations of $\{1, \ldots, m\}$. For every $\pi \in S(m)$ and $r \in B'_{m,n}$ define $C_n(\pi, r) = \{ (i_{\pi(1)}, \ldots, i_{\pi(m)}) \mid (i_1, \ldots, i_m) \in C_n(r) \}$. Every $C_n(\pi, r)$ has the corresponding properties (a) and (b) and it follows from (3.9) that

$$A_{m,n} \subset \bigcup_{\pi \in S(m)} \bigcup_{r \in B'_{m,n}} C_n(\pi, r). \tag{3.10}$$

Starting with the expectation on the left-hand side of (3.8), Hölder's inequality yields, for every $n \geq lm$ and $s \in S$,

$$\begin{aligned}
&\mathbb{E}_s\left[\exp\left(n \int_{S^m} \varphi \, dL_{n,m}\right)\right] \\
&\quad = \mathbb{E}_s\left[\exp\left(\frac{n}{|A_{m,n}|} \sum_{(i_1,\ldots,i_m) \in A_{m,n}} \varphi(X_{i_1}, \ldots, X_{i_m})\right)\right] \\
&\quad \leq \prod_{\pi \in S(m)} \prod_{r \in B'_{m,n}} \Biggl(\mathbb{E}_s\Biggl[\exp\Biggl(\frac{n m! \, |B'_{m,n}|}{|A_{m,n}|} \\
&\qquad\qquad \times \sum_{(i_1,\ldots,i_m) \in C_n(\pi,r)} \varphi(X_{i_1}, \ldots, X_{i_m})\Biggr)\Biggr]\Biggr)^{1/(m! |B'_{m,n}|)}
\end{aligned} \tag{3.11}$$

where we used (3.10) for the last step. Since $|B'_{m,n}| \leq |B_{m,n}| \leq 2l(2n)^{m-1}$ and $|A_{m,n}| \geq (n - (2m-1)l)^m \geq (n/2)^m$ for $n \geq 4lm$, it follows that

$$\frac{n m! \, |B'_{m,n}|}{|A_{m,n}|} \leq 4^m l m! \quad \text{for } n \geq 4lm. \tag{3.12}$$

Given $\pi \in S(m)$ and $r \in B'_{m,n}$, it follows from (a) and (b) for $C_n(\pi, r)$, that the m-tuples in $C_n(\pi, r)$ consist of $p \equiv m|C_n(\pi, r)|$ different $q_1, \dots, q_p \in \{l, l+1, \dots, n\}$, which we may label such that $q_j \geq q_{j-1} + l$ for all $j \in \{2, 3, \dots, p\}$. It follows from (3.2) and (3.5) that $\pi^{q_j - q_{j-1}}(s, \cdot) \leq M\mu$ for every $j \in \{1, \dots, p\}$, where $q_0 \equiv 0$. 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 (3.12) it follows that

$$\mathbb{E}_s\left[\exp\left(\frac{nm!\,|B'_{m,n}|}{|A_{m,n}|} \sum_{(i_1,\dots,i_m) \in C_n(\pi,r)} \varphi(X_{i_1}, \dots, X_{i_m})\right)\right] \leq \left(M^m \int_{S^m} \exp(4^m l m! \varphi)\, d\mu^{\otimes m}\right)^{|C_n(\pi,r)|}. \quad (3.13)$$

Since $|C_n(\pi, r)| \leq \lfloor n/lm \rfloor \leq n$, the estimates (3.11) and (3.13) imply (3.8). ☐

Remark 3.14. Notice, that we get the results in the τ-topology without any further assumptions on the moments. Thus in the τ-topology, Assumption 3.2 suffices to transfer the LDP to products. The reason is, that we have checked Condition 1.3 (with the reference measure μ defined in (3.5)) for the reduced term $L_{n,m}$, and the result follows for L_n^m and $L_n^{\otimes m}$, respectively, using the fact that convergence in the total variation distance implies convergence in the τ-topology.

Remark 3.15. The Markov chains analyzed in [13] show that the LDP with respect to the weak topology does not transfer to the strong topology setting in general.

3.3. **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, that $\mathbb{P}$ can be represented as a μ-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 μ-mixture of $\{\mathbb{P}_\theta^n\}_{\theta \in \Theta}$, that is

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

If $\pi_\theta \equiv \mathbb{P}_\theta X_j^{-1}$, $\theta \in \Theta$, varies continuously in the weak topology on $\mathcal{M}_1(S)$ and if Θ is compact, then $\{\mathbb{P}^n\}_{n \in \mathbb{N}}$ satisfies the LDP with a rate function given by

$$\lambda_1(\nu) \equiv \inf_{\theta \in \Theta} H(\nu \,|\, \pi_\theta),$$

cf. [9, Remark (ii) after Theorem 2.3]. Notice that [9, Theorem 4.1] is an extension of this result. The condition of weak convergence of π_θ is replaced by the following two conditions:

(a) $\lambda\colon \Theta \times \mathcal{M}_1(S) \to [0, \infty]$ is jointly lower semi-continuous, where $\lambda(\theta, \nu) \equiv H(\nu \,|\, \pi_\theta)$.

(b) $\{\mathbb{P}_\theta^n\}_{n \in \mathbb{N}}$ is exponentially tight, that is: for every $L > 0$ there exists a compact set $K_L \subset \mathcal{M}_1(S)$ such that

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

is sufficiently large.

Notice that for exchangeable sequences, we can not expect that the LDP for the laws of L_n with respect to the weak topology carries over to the product in the τ-topology, because we can find an example even in the case $m = 1$, where the LDP can not be transferred from the weak to the strong topology. Thus the following example is of independent interest.

Example 3.16. Denote by λ the Lebesgue–Borel measure on $S = [0,1]$, equipped with the Borel σ-algebra. Let $\mathbb{P}_\theta \equiv (\delta_\theta)^{\otimes \mathbb{N}}$, $\theta \in [0,1]$, and define $\mathbb{P} = \int_{[0,1]} \mathbb{P}_\theta \, \lambda(d\theta)$. Note that

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

for any Borel set U in $\mathcal{M}_1([0,1])$. The expected rate function is

$$I(\mu) = \begin{cases} 0 & \text{if } \mu = \delta_\theta \text{ for a } \theta \in [0,1], \\ \infty & \text{otherwise.} \end{cases}$$

Define $U_\theta = \{\, \nu \in \mathcal{M}_1([0,1]) \mid \nu(\{\theta\}) > 1 - \varepsilon \,\}$ for $\varepsilon > 0$. We get

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

the integrand is equal to 1 for $\delta = \theta$ and 0 otherwise. Therefore $\mathbb{P}(L_n \in U_\theta) = 0$ but $\inf_{\mu \in U_\theta} I(\mu) = 0$, since $\delta_\theta \in U_\theta$, and thus the lower bound fails. Notice moreover that I has non-compact level sets, because

$$I^{-1}(\{0\}) \subset \bigcup_{\theta \in [0,1]} U_\theta.$$

In spite of this example Theorem 1.19 in [14] is a 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), condition (a) and (b) can be replaced by the following condition:

(c) The map $\theta \mapsto \pi_\theta$ vary continuously in the τ-topology and $\mu(U) > 0$ for every open $U \in \Theta \cap \mathcal{B}(\mathcal{M}_1(S))$.

If Θ is τ-quasi-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 λ_1. A similar result is proved in [6, Section 2] by different methods.

Now we will assume in addition that $\{\pi_\theta\}_{\theta \in \Theta}$ satisfies the following condition:

(d) there exists a finite measure ν on S such that $\pi_\theta \ll \nu$ for any $\theta \in \Theta$ and for $f_\theta \equiv \frac{d\pi_\theta}{d\nu}$ we assume that there exists a $q > 0$ such that $\sup_{\theta \in \Theta} \|f_\theta\|_q < \infty$, where $\|\cdot\|_q$ denotes the norm in $L_q(S, \mathcal{S}, \nu)$.

Remark 3.17. If Θ is τ-quasi-compact, Condition (c) implies $\sup_{\theta \in \Theta} \pi_\theta(A) \le \nu(A)$ for all $A \in \mathcal{S}$ for some finite measure ν and thus (d) is fulfilled for $\{\pi_\theta\}_{\theta \in \Theta}$.

The following result is stated for a Polish space S:

Theorem 3.18. *Assume that the exchangeable sequence $\{X_i\}_{i \in \mathbb{N}}$ satisfies Condition* (a), (b) *and Θ is assumed to be weakly compact (alternatively we assume that the laws of the empirical measures of $\{X_i\}_{i \in \mathbb{N}}$ satisfy a LDP in the weak topology, not necessarily satisfying condition* (a) *and* (b)*). Moreover assume that* (d) *holds and that Condition 1.6 holds for $\nu^{\otimes m}$ and Φ. Then for every $m \in \mathbb{N}$ the measures*

$\{\mathbb{P}(L_n^m)^{-1}\}_{n\in\mathbb{N}}$ satisfy the LDP in the τ_Φ-topology on $\mathcal{M}_\Phi(S^m)$ with the good rate function

$$\lambda_m(\nu) = \begin{cases} \lambda_1(\widetilde{\nu}) & \text{if } \widetilde{\nu}^{\otimes m} = \nu, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.19}$$

If we assume Condition 1.7 for Φ and $\nu^{\otimes m}$, we get the same result for the measures $\{\mathbb{P}(L_n^{\otimes m})^{-1}\}_{n\in\mathbb{N}}$ (with the same rate function). Alternatively assume that Condition 1.6 holds for $\nu^{\otimes m}$ and Φ and the exchangeable sequence $\{X_i\}_{i\in\mathbb{N}}$ satisfies Condition (c) *and Θ is assumed to be τ-quasi-compact. Then for every $m\in\mathbb{N}$ the measures $\{\mathbb{P}(L_n^m)^{-1}\}_{n\in\mathbb{N}}$ satisfy the LDP in the τ_Φ-topology on $\mathcal{M}_\Phi(S^m)$ with the rate function λ_m.*

Proof. By [9, Theorem 4.1] and [14, Theorem 1.19], respectively, the theorem is true in the weak topology for $m=1$. For each $\theta\in\Theta$ we get by Hoeffding's result (see case 3.1), that

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_\theta}\left[\exp(m\varphi(X_1,\dots,X_m))\right] \\ &\quad\le \left(\int_{S^m} \exp(pm\varphi(X_1,\dots,X_m))\,d\nu^{\otimes m}\right)^{1/p} \left(\int_{S^m}\left(\prod_{i=1}^m f_\theta(x_i)\right)^q d\nu^{\otimes m}\right)^{1/q}. \end{aligned}$$

Thus it follows from Hölder's and Jensen's inequality

$$\begin{aligned} \sup_{n\in\mathbb{N}}&\left(\mathbb{E}_{\mathbb{P}}\left[\exp\left(n\int_{S^m}\varphi\,dL_n^m(\omega)\right)\right]\right)^{1/n} \\ &\le \left(\int_{S^m}\exp\big(pm\varphi(X_1,\dots,X_m)\big)\,d\nu^{\otimes m}\right)^{1/pm} \\ &\quad\times\int_\Theta\left(\int_{S^m}\left(\prod_{i=1}^m f_\theta(x_i)\right)^q d\nu^{\otimes m}\right)^{1/qm} d\mu(\theta). \end{aligned}$$

Using Condition (d) and arguing as in the proof of Theorem 3.6 we get the result. □

Example 3.20. A simple case is a 0-1 valued exchangeable sequence $\{X_i\}_{i\in\mathbb{N}}$. One can find a probability measure ν on $[0,1]$ such that if $\pi_\theta=(1-\theta)\delta_{\{0\}}+\theta\delta_{\{1\}}$, $\theta\in[0,1]$, the distribution $\mathbb{P}$ can be represented as $\mathbb{P}=\int_{[0,1]}\pi_\theta\,\nu(d\theta)$ (see [3, Section 4]).

Since $f(\theta,p)\equiv H(\mu_p\,|\,\pi_\theta)=p\log\frac{p}{1-\theta}+(1-p)\log\frac{1-p}{\theta}$, where $\mu_p\equiv(1-p)\delta_{\{0\}}+p\delta_{\{1\}}$, is jointly lower semi-continuous and $\sup_{\theta\in[0,1]}\pi_\theta\le\delta_{\{0\}}+\delta_{\{1\}}$, we can apply the theorem.

Example 3.21. Let $\Theta\subseteq\mathbb{R}\times\mathbb{R}_+$ be a compact set and let π_θ, $\theta=(\mu,\sigma^2)$, be normally distributed, with density $g_{\mu,\sigma^2}(x)$. We define a mixture of i.i.d. $N(\mu,\sigma^2)$ sequences with respect to λ^2, the Lebesgue-measure on $\mathbb{R}^2$, mixing over μ and σ^2 by:

$$\begin{aligned} &Q^n(A_1\times\cdots\times A_n) \\ &\quad=\int_{\Theta_1}\int_{\Theta_2}\left(\int_{\mathbb{R}^n}\left(\prod_{i=1}^n g_{\mu,\sigma^2}(x_i)1_{A_i}(x_i)\right)d\lambda^n(x_1,\dots,x_n)\right)\frac{d\lambda^2(\mu,\sigma^2)}{\lambda^2(\Theta)} \end{aligned}$$

for $A_i\in\mathcal{B}(\mathbb{R})$ for any $i\in\{1,\dots,n\}$. This mixture is exchangeable and is directed by a random measure α, a measure on the space $\mathcal{M}_1(S)$. For our parametric family of distributions we get

$$\alpha \in \{\pi_\theta, \theta \in \Theta\} =: \Theta' \quad \text{a.s.}$$

Thus the compact set $\Theta \subseteq \mathbb{R} \times \mathbb{R}_+$ can be identified with the weak-compact set of measures Θ' (see [3, Section 4]). $H(\pi \,|\, \nu_{\mu,\sigma^2})$ is jointly lower semi-continuous because the probability measures π_θ vary continuously in the weak topology. Moreover, a normal distributed random variable is q-integrable with respect to the Lebesgue measure for every $q > 1$, we get

$$\sup_{\theta \in \Theta} \|g_\theta\|_q < \infty \quad \text{with } \theta = (\mu, \sigma^2),$$

since Θ is compact and $\sigma^2 > 0$. Of course the example can be reduced to mixtures of i. i. d. $N(\mu, \sigma^2)$ sequences only mixing over μ or over σ^2 on compact subsets. Let us remark, that this example is of interest for other aspects: Schonberg's theorem says that any infinite spherically symmetric sequence of random variables is a mixture of i. i. d. $N(0, \sigma^2)$ sequences. This result fits naturally into the sufficient statistics setting, pointed out in [3, Section 3 and 18]. There the laws of special sufficient statistics can be described by mixtures (over μ and σ^2) of i. i. d. $N(\mu, \sigma^2)$ sequences, or Poisson, binomial and negative binomial sequences.

3.4. **Stationary sequences, mixing conditions.** Let $\{X_i\}_{i \in \mathbb{N}}$ be a stationary sequence of random variables which take values in S. The hypermixing conditions (H-1) and (H-2) of [7, Section 6.4.2] are as follows: For any $r \geq k \geq 2$ and $l \geq 1$ a family of functions $\{f_i,\ 1 \leq i \leq k\} \in B(S^r)$ is called l-separated, if there exist k disjoint intervals $\{a_i, a_i + 1, \ldots, b_i\}$ with $a_i \leq b_i \in \{1, \ldots, r\}$ such that $f_i(x_1, \ldots, x_r)$ is actually a bounded measurable function of $\{x_{a_i}, \ldots, x_{b_i}\}$ and for all $i \neq j$ either $a_i - b_j \geq l$ or $a_j - b_i \geq l$.

Assumption (H1). There exist $l, \alpha < \infty$ such that for all $k, r < \infty$ and any l-separated functions $f_i \in B(S^r)$

$$\begin{aligned} &\mathbb{E}_\mathbb{P}\big[|f_1(X_1, \ldots, X_r) \times \cdots \times f_k(X_1, \ldots, X_r)|\big] \\ &\qquad \leq \prod_{i=1}^k \mathbb{E}_\mathbb{P}\big[|f_i(X_1, \ldots, X_r)|^\alpha\big]^{1/\alpha}. \end{aligned} \tag{3.22}$$

Assumption (H2). There exist a constant l_0 and functions $\beta(l) \geq 1$ and $\gamma(l) \geq 0$ such that for all $l > l_0$ and all $r < \infty$ and any two l-separated function $f, g \in B(S^r)$

$$\begin{aligned} &|\mathbb{E}_\mathbb{P}[f(X_1, \ldots, X_r)]\, \mathbb{E}_\mathbb{P}[g(X_1, \ldots, X_r)] - \mathbb{E}_\mathbb{P}[f(X_1, \ldots, X_r) g(X_1, \ldots, X_r)]| \\ &\qquad \leq \gamma(l)\, \mathbb{E}_\mathbb{P}[|f(X_1, \ldots, X_r)|^{\beta(l)}]^{1/\beta(l)}\, \mathbb{E}_\mathbb{P}[|g(X_1, \ldots, X_r)|^{\beta(l)}]^{1/\beta(l)}, \end{aligned}$$

and

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

for some $\delta > 0$.

It is well known that the laws of L_n satisfy the LDP in $\mathcal{M}_1(S)$ equipped with the τ-topology, if both (H-1) and (H-2) hold [7, Theorem 6.4.14 and Lemma 6.4.18]. Furthermore by [4, Proposition 1] the Condition (H1) is unnecessary. Here we will treat only the case $m = 2$ and will assume in addition:

Assumption 3.23. There exist $l_2 \in \mathbb{N}$ and $\gamma, \beta \in [1,\infty)$ such that for all $i \in \mathbb{N}$ and $f \in B(S^2,[0,\infty))$,

$$\mathbb{E}_{\mathbb{P}}\left[\prod_{j=1}^{\lfloor i/2-l_2\rfloor} f(X_{i-j},X_j)\right] \le \gamma\,\mathbb{E}_{\mathbb{P}}\left[\prod_{j=1}^{\lfloor i/2-l_2\rfloor} f(X_{i-j},\widetilde{X}_j)^\beta\right]^{1/\beta},$$

where the process $\{\widetilde{X}_j\}_{j\in\mathbb{N}}$ is an independent copy of $\{X_j\}_{j\in\mathbb{N}}$.

Remark 3.24. (a) If Assumption (H1) holds for l, then it holds for all $\tilde{l} \in \mathbb{N}$ satisfying $\tilde{l} \ge l$.
(b) Assumption 3.23 is fulfilled, if there exist $f_1, f_2 \in B(S,[0,\infty))$ such that $f(x,y) = f_1(x)f_2(y)$ for all $x,y \in S$, because it follows from the Cauchy–Schwarz inequality for $\gamma = 1$ and $\beta = 2$.

We get the following result:

Theorem 3.25. *If* (H1), (H2) *and Assumption 3.23 hold for the stationary sequence* $\{X_i\}_{i\in\mathbb{N}}$, *and if Condition 1.6 holds for* $\mathcal{L}(X_1)\otimes\mathcal{L}(X_1)$ *and* Φ *($m=2$), then* $\{\mathbb{P}(L_n^2)^{-1}\}_{n\in\mathbb{N}}$ *satisfies a LDP in the* τ_Φ*-topology on* $\mathcal{M}_\Phi(S^2)$ *with a good rate function* $I_2(\nu)$*, defined by*

$$I_2(\nu) = \begin{cases} I_1(\mu) & \text{if } \mu\otimes\mu = \nu, \\ \infty & \text{otherwise,}\end{cases} \tag{3.26}$$

where

$$I_1(\mu) \equiv \sup_{f\in B(S)}\Big(\int_S f\,d\mu - \Lambda(f)\Big), \qquad \mu \in \mathcal{M}_1(S),$$

and

$$\Lambda(f) \equiv \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(\sum_{i=1}^n f(X_i)\right)\right]. \tag{3.27}$$

In particular the limit (3.27) *exists for every* $f \in B(S)$. *If we assume Condition 1.7 for* $\mathcal{L}(X_1)\otimes\mathcal{L}(X_1)$ *and* Φ *($m=2$), we get the same result for the measures* $\{\mathbb{P}(L_n^{\otimes 2})^{-1}\}_{n\in\mathbb{N}}$ *(with the same rate function).*

Proof. By [4, Theorem 1], the theorem is true in the strong topology for $m=1$ if [4, Condition (S)] holds, because (H2) implies (S). By [7, Lemma 6.4.18], Condition 1.3 can be verified if (H1) holds, and thus by 2.10(a) we get the result in the τ_Φ-topology. Assume $m=2$ and define for a fixed $l \in \mathbb{N}$ the empirical measure

$$\Omega \ni \omega \mapsto L_{n,l}(\omega) = \frac{1}{|A_{n,l}|}\sum_{(i,j)\in A_{n,l}} \delta_{(X_i(\omega),X_j(\omega))}$$

where

$$A_{n,l} \equiv \big\{\, (i,j) \in \{1,\dots,n\}^2 \;\big|\; |i-j| \ge l \text{ for all } i \ne j \,\big\}. \tag{3.28}$$

Without loss of generality we may assume that $\varphi\colon S^2 \mapsto [0,\infty)$ is a bounded and symmetric function in Φ. Remark that $|A_{n,l}| \ge n^2/4$ for $n \ge 8l$ and thus $\frac{n(n-l)}{|A_{n,l}|} \le 4$

for $n \geq 8l$. Now, for $n \geq 8l$ we get

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(\frac{n}{|A_{n,l}|}\sum_{(i,j)\in A_{n,l}}\varphi(X_i,X_j)\biggr)\Biggr]\\
&\quad= \mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(\frac{n}{|A_{n,l}|}\sum_{i=l+2}^{2n-l}\sum_{j=\max\{1,i-n\}}^{\min\{n,i-1\}}\varphi(X_{i-j},X_j)1_{\{(i,j)\colon |i-2j|\geq l\}}\biggr)\Biggr]\\
&\quad\leq \mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(\frac{n(n-l)}{|A_{n,l}|}\sum_{j=l}^{n-l}\varphi(X_{n+1-j},X_j)\biggr)\Biggr]^{1/n-l}\\
&\qquad\times\Biggl(\prod_{\substack{i=l+2\\ i\neq n+1}}^{2n-l}\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(\frac{n2(n-l)}{|A_{n,l}|}\sum_{j=\max\{1,i-n\}}^{\min\{i-1,n\}}\varphi(X_{i-j},X_j)\\
&\qquad\qquad\times 1_{\{(i,j)\colon |i-2j|\geq l\}}\biggr)\Biggr]\Biggr)^{1/2(n-l)}\\
&\quad\leq \mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(4\sum_{j=l}^{n-l}\varphi(X_{n+1-j},X_j)\biggr)\Biggr]^{1/n-l}\\
&\qquad\times\Biggl(\prod_{i=l+2}^{n}\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(8\sum_{j=1}^{i-1}\varphi(X_{i-j},X_j)1_{\{(i,j)\colon |i-2j|\geq l\}}\biggr)\Biggr]\Biggr)^{1/n-l}\\
&\quad\leq \Biggl(\prod_{i=l+2}^{n+1}\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(8\sum_{j=1}^{i-1}\varphi(X_{i-j},X_j)1_{\{(i,j)\colon |i-2j|\geq l\}}\biggr)\Biggr]\Biggr)^{1/n-l},
\end{aligned}$$

where we have applied Hölder's inequality and used the stationarity of the sequence $\{X_i\}_{i\in\mathbb{N}}$ as well as $\varphi \geq 0$. Since φ is symmetric we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(8\sum_{j=1}^{i-1}\varphi(X_{i-j},X_j)1_{\{(i,j)\colon |i-2j|>l\}}\biggr)\Biggr]\\
&\quad\leq \mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(16\sum_{j=1}^{\lfloor\frac{i-l}{2}\rfloor}\varphi(X_{i-j},X_j)\biggr)\Biggr]\\
&\quad\leq \gamma\,\mathbb{E}_{\mathbb{P}}\Biggl[\exp\biggl(16\beta\sum_{j=1}^{\lfloor\frac{i-l}{2}\rfloor}\varphi(X_{i-j},\widetilde{X}_j)\biggr)\Biggr]^{1/\beta},
\end{aligned}$$

where the last inequality follows by Assumption 3.23 for $l_2 = l$. For each k with $k-1 \geq l$ we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\bigl[\exp(\varphi(X_1,\widetilde{X}_k)+\varphi(X_k,\widetilde{X}_1))\bigr]\\
&\quad\leq \int_{S^2}\biggl(\int_S \exp(\alpha\varphi(x_1,\widetilde{x}_k)\mu_1(dx_1)\biggr)^{1/\alpha}\\
&\qquad\times\biggl(\int_S \exp(\alpha\varphi(x_k,\widetilde{x}_1)\mu_k(dx_k)\biggr)^{1/\alpha}\quad \mu(d\widetilde{x}_1,d\widetilde{x}_k)
\end{aligned}$$

by Assumption (H1), where μ_1 and μ_k, respectively denote the marginal distribution on the first and k-th component respectively and $\mu(d\widetilde{x}_1, d\widetilde{x}_k)$ the two dimensional marginal distribution on the first and k-th component. Applying Assumption (H1) again, we get

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}}\big[\exp(\varphi(X_1,\widetilde{X}_k)+\varphi(X_k,\widetilde{X}_1)\big]\\&\qquad\le \mathbb{E}_{\mathbb{P}}\big[\exp(\alpha\varphi(X_1,\widetilde{X}_k))\big]^{1/\alpha}\,\mathbb{E}_{\mathbb{P}}\big[\exp(\alpha\varphi(X_k,\widetilde{X}_1))\big]^{1/\alpha}.\end{aligned}$$

Thus with $k\in\mathbb{N}$ and $m>0$ such that $km=\lfloor\frac{i-l}{2}\rfloor$ we get

$$\mathbb{E}_{\mathbb{P}}\left[\exp\left(16\beta\sum_{j=1}^{\lfloor\frac{i-l}{2}\rfloor}\varphi(X_{i-j},\widetilde{X}_j)\right)\right]^{1/\beta}\le \mathbb{E}_{\mathbb{P}}\big[\exp(16\beta k\alpha\varphi(X_1,\widetilde{X}_1))\big]^{m/\alpha\beta}.$$

We arrive at

$$\begin{aligned}&\frac{1}{n}\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{n}{|A_{n,l}|}\sum_{(i,j)\in A_{n,l}}\varphi(X_i,X_j)\right)\right]\\&\qquad\le\frac{1}{n}\log\gamma^{n-l}+\frac{1}{n}\log\big(\mathbb{E}_{\mathbb{P}}\big[\exp(16\beta k\alpha\varphi(X_1,\widetilde{X}_1))\big]\big)^{\frac{n-l-1}{2k\alpha\beta}}\\&\qquad\le\log\gamma+\log\mathbb{E}_{\mathbb{P}}\big[\exp(16\beta k\alpha\varphi(X_1,\widetilde{X}_1))\big]\end{aligned}$$

and thus we have verified 1.3 for the empirical measure $L_{n,l}$ for a fixed $l\in\mathbb{N}$ which is determined by the assumptions. Arguing as in the proof of Theorem 3.6 we get the result. $\square$

Example 3.29. Let $|\alpha|<1$ and $\mathcal{L}(X_0)=N(0,1/(1-\alpha^2))$ and let $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. standard normal random variables. It is well known, that the process

$$X_{n+1}\equiv\alpha X_n+Y_n$$

satisfies (H1) and (H2) but not Assumption 3.2 (see [7, Exercise 6.4.19] and the example of the continuous time Ornstein–Uhlenbeck process discussed in [8]). Denote by $g(\alpha,\sigma^2)$ the density with respect to Lebesgue measure of the normal distribution with parameters α and σ^2. Since the transition kernel is given by $\pi^1(x,dy)=g(\alpha x,1)\,dy$ for $l\in\mathbb{N}$ $\pi^l(x,dy)=g(\alpha^l x,\alpha_l)\,dy$ with $\alpha_l=\frac{1-\alpha^{2l}}{1-\alpha^2}$. Thus for $\varphi\in B(S^2,[0,\infty))$ we get

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[\varphi(X_0,X_l)]&=\int_{\mathbb{R}}\int_{\mathbb{R}}\varphi(x,y)g(\alpha^l x,\alpha_l)g\left(0,\frac{1}{1-\alpha^2}\right)dy\,dx\\&\le\int_{\mathbb{R}}\int_{\mathbb{R}}\varphi(x,y)\left(\frac{1-\alpha^2}{2\pi}\right)^{1/2}(1-\alpha^{2l})^{-1/2}\left(\frac{1-\alpha^2}{2\pi}\right)^{1/2}\\&\qquad\times\exp\left(-\frac{1}{2}\frac{1-\alpha^2}{1+\alpha^l}x^2\right)\exp\left(-\frac{1}{2}\frac{1-\alpha^2}{1+\alpha^l}y^2\right)dx\,dy,\end{aligned}$$

using $2xy \le x^2 + y^2$, $x, y \in \mathbb{R}$. With $(1+\alpha^l)^{-1} = (1+2\alpha^l)^{-1} + \frac{\alpha^l}{(1+\alpha^l)(1+2\alpha^l)}$ and Jensen's inequality we get

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[\varphi(X_0, X_\varphi)] \\
&\quad \le \int_{\mathbb{R}}\int_{\mathbb{R}} \varphi(x,y)(1+2\alpha^l)(1-\alpha^{2l})^{1/2}\exp\Big(-\frac{1}{2}\beta_l x^2\Big)\exp\Big(-\frac{1}{2}\beta_l y^2\Big) \\
&\qquad \times g\Big(0, \frac{1+2\alpha^l}{1-\alpha^2}\Big)^2 \, dy\, dx \\
&\quad \le \Big(\int_{\mathbb{R}}\int_{\mathbb{R}} \varphi^\beta(x,y)(1+2\alpha^l)^\beta(1-\alpha^{2l})^{-\beta/2}\exp\Big(-\frac{1}{2}\beta_l\beta x^2\Big) \\
&\qquad \times \exp\Big(-\frac{1}{2}\beta_l\beta y^2\Big) g\Big(0, \frac{1+2\alpha^l}{1-\alpha^2}\Big)^2 \, dx\, dy\Big)^{1/\beta},
\end{aligned}$$

where $\beta_l \equiv (1-\alpha^2)\alpha^l(1+\alpha^l)^{-1}(1+2\alpha^l)^{-1}$ and $\beta \equiv 2(1+\alpha^l)$. The term on the right in the last inequality is easily seen to be

$$\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}} \varphi^\beta(x,y)(1+2\alpha^l)^{\beta-1}(1-\alpha^{2l})^{-\beta/2} g\Big(0, \frac{1}{1-\alpha^2}\Big)^2 \, dx\, dy \\
&\quad = (1+2\alpha^l)^{1-1/\beta}(1-\alpha^{2l})^{-1/2}\, \mathbb{E}_{\mathbb{P}}[\varphi(X_0, \widetilde{X}_l)^\beta]^{1/\beta} \qquad (3.30) \\
&\quad \le \frac{3}{\sqrt{1-\alpha^2}}\, \mathbb{E}_{\mathbb{P}}[\varphi(X_0, \widetilde{X}_l)^4]^{1/4},
\end{aligned}$$

where we have used $\beta = 2(1+\alpha^l) \le 4$. Consider

$$K_1(x, A) \equiv \mathbb{P}((X_0, X_1, \ldots, X_{k-1}) \in A \,|\, X_k = x)$$

for $A \in \mathcal{S}^{\otimes k}$ and $K_2(y, B) \equiv \mathbb{P}((X_{k+l+1}, \ldots, X_{2k+l}) \in B \,|\, X_{k+l} = y)$ for $B \in \mathcal{S}^{\otimes k}$ and $\mu \equiv \mathcal{L}(X_k) = \mathbb{P}X_k^{-1}$. In order to verify the additional Assumption 3.23 for the model, note that

$$\mathbb{E}_{\mathbb{P}}\Big[\prod_{j=0}^{k}\varphi(X_j, X_{2k+l-j})\Big] = \mathbb{E}_{\mathbb{P}}[g(X_k, X_{k+l})],$$

where

$$\begin{aligned}
g(x,y) \equiv \int_{S^k}\int_{S^k} \varphi(x,y) \prod_{j=0}^{k}\varphi(X_j, X_{2k+l-j}) \\
K_1(x, d(x_0, \ldots, x_{k-1}))K_2(y, d(x_{k+l+1}, \ldots, x_{2k+l})).
\end{aligned}$$

Now by (3.30)

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[g(X_k, X_{k+l})] \\
&\quad \le \frac{3}{\sqrt{1-\alpha^2}}\, \mathbb{E}_{\mathbb{P}}\left[g(X_k, \widetilde{X}_{k+l})^4\right]^{1/4} \\
&\quad \le \frac{3}{\sqrt{1-\alpha^2}} \biggl(\int_S \int_S \mu(dx)\, \mu(dy) \int_{S^k} \int_{S^k} \varphi^4(x,y) \prod_{j=0}^{k-1} \varphi^4(x_j, x_{2k+l-j}) \\
&\qquad\qquad \times K_1(x, d(x_0, \ldots, x_{k-1})) \times K_2(y, d(x_{k+l+1}, \ldots, x_{2k+l})) \biggr)^{1/4} \\
&\quad = \frac{3}{\sqrt{1-\alpha^2}}\, \mathbb{E}_{\mathbb{P}}\left[\prod_{j=0}^{k} \varphi^4(X_j, \widetilde{X}_{2k+l-j}) \right]^{1/4},
\end{aligned}
$$

where we have used Jensen's inequality.

4. Statistics and Processes

The aim of this section is to establish the LDP for U-statistics and V-statistics of dependent or independent but not identically distributed processes. Moreover we get a LDP result in the Markovian situation for products of empirical processes from the results of Section 3 via the concept of the projective limit approach of Dawson and Gärtner (see [7, Theorem 4.6.1]).

For a measurable map $h\colon S^m \to \mathbb{R}^d$ the U- and V-statistics of degree m with kernel function h are defined by

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

for all $n \ge m$ and all $n \in \mathbb{N}$, respectively. If the sequence $\{X_i\}_{i\in\mathbb{N}}$ is i. i. d., we have proved a LDP result for these statistics under weak moment conditions in [14, Theorem 1.13]. We can adopt the arguments. Let $h_1, \ldots, h_d$ denote the component functions of h and define $\Phi_h = B(S^m) \cup \{h_1, \ldots, h_d\}$. Now the statistics are compositions of L_n^m and $L_n^{\otimes m}$ respectively, with the τ_{Φ_h}-continuous functional $\mathcal{M}_{\Phi_h}(S^m) \ni \nu \mapsto \int h\, d\nu$. The contraction principle [7, Theorem 4.2.1] immediately leads to the following results for U-statistics and V-statistics with dependent inputs:

Theorem 4.1. (a) *The U-statistics $\{U_n^m(h)\}_{n\ge m}$ satisfy a LDP with a good rate function provided one of the following conditions is satisfied:*

(i) $\{X_i\}_{i\in\mathbb{N}}$ *is independent satisfying* (3.1) *and* $\int_{S^m} \exp(\alpha|h_i|)\, d\mu^{\otimes m} < \infty$ *for all* $i \in \{1, \ldots, d\}$, $\alpha > 0$ *and* μ *as in* (3.1).

(ii) $\{X_i\}_{i\in\mathbb{N}}$ *is a Markov chain satisfying* (3.2) *and* $\sup_{\sigma\in S} \mathbb{E}_\sigma[\exp(\alpha|h_i|)] < \infty$ *for all* $i \in \{1, \ldots, d\}$ *and* $\alpha > 0$.

(iii) $\{X_i\}_{i\in\mathbb{N}}$ *is an exchangeable sequence satisfying* (a), (b) *and* (c) *in Subsection 3.3 and the mixing parameter set* Θ *is a weak-compact subset of* $\mathcal{M}_1(S)$. *Moreover we assume* $\int_{S^m} \exp(\alpha|h_i|)\, d\nu^{\otimes m} < \infty$ *for all* $i \in \{1, \ldots, d\}$, $\alpha > 0$ *and* ν *as in (c).*

(iv) $\{X_i\}_{i\in\mathbb{N}}$ *is a stationary sequence satisfying (H1), (H2) and* (3.23) *and* $\int_{S^m} \exp(\alpha|h_i|)\, d\mu^{\otimes m} < \infty$ *for all* $i \in \{1, \ldots, d\}$, $\alpha > 0$ *and* $\mu = \mathcal{L}(X_1)$.

(b) *If Condition 1.7 is satisfied for* $h_1, \ldots, h_d$ *and every measure given in the cases* (ai)–(aiv), *then in each case the V-statistics* $\{V_n^m(h)\}_{n\in\mathbb{N}}$ *satisfy a LDP with the same good rate function.*

The rate function, defined on $\mathbb{R}^d$ with values in $[0,\infty]$ is given by

$$I(x) = \inf\Big\{F(\nu)\ \Big|\ \nu \in \mathcal{M}_{\Phi_h}(S^m),\ \int_{S^m} h\,d\nu = x\Big\},$$

where $F(\nu)$ is the rate function for the LDP of the laws of L_n^m, see (3.4), (3.19) *and* (3.26).

Remark 4.2. Actually the theorem is a direct consequence of Lemma 2.3 if we take $\mathcal{F} = \{h\}$. Thus we do not have to apply the contraction principle two times.

Next we want to derive large deviation results on the *process level* for Markov chains. Given $k \in \mathbb{N}$ with $k \geq 2$, the transition probability kernel $\pi_k \colon S^k \times \mathcal{S}^{\otimes k} \to [0,1]$ of the Markov chain $\{(X_n, X_{n+1}, \dots, X_{n+k-1})\}_{n\in\mathbb{N}_0}$ is given by

$$\pi_k((\sigma_1,\dots,\sigma_k), A) = \int_S 1_A(\sigma_2,\dots,\sigma_k,\tau)\pi(\sigma_k, d\tau)$$

for all $(\sigma_1,\dots,\sigma_k) \in S^k$ and $A \in \mathcal{S}^{\otimes k}$. Let $\{\mathbb{P}_{k,\sigma}\}_{\sigma\in S}$ denote the family of Markovian measures. The condition (3.2) for π implies that the kernel π_k satisfies

$$\pi_k^{k+l-1}(\sigma,\cdot) \leq \frac{M}{N}\sum_{m=1}^{k+N-1} \pi_k^m(\tau,\cdot)$$

for all $\sigma,\tau \in S^k$. Let

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

be the empirical measure of the above Markov chain. We define the rate function $J_{1,k} \colon \mathcal{M}_1(S^k) \mapsto [0,\infty]$ by

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

Note that [8, Lemma 4.4.9] gives an alternative expression for $J_{1,k}$ in terms of the relative entropy. Analogously to (3.4) we define for every integer $m \geq 2$ the rate function $J_{m,k} \colon \mathcal{M}_1((S^k)^m) \mapsto [0,\infty]$ by

$$J_{m,k}(\mu) = \begin{cases} J_{1,k}(\nu), & \text{if } \nu^{\otimes m} = \mu \\ \infty, & \text{otherwise.} \end{cases} \tag{4.3}$$

Let $\mu_k \in \mathcal{M}_1(S^k)$ be given by

$$\mu_k = \frac{1}{N}\sum_{m=1}^{k+N-1} \pi_k^m(\sigma_0,\cdot) \tag{4.4}$$

with an arbitrary $\sigma_0 \in S^k$. As an immediate consequence of Theorem 3.6 we get the following extension:

Corollary 4.5. *Let Φ be a set of $\mathcal{S}^{\otimes mk}$-measurable functions $\varphi \colon S^{mk} \to \mathbb{R}$ containing $B((S^k)^m)$. Assume that Condition 1.7 holds for Φ and with the integration with respect to the measure $\mathbb{P}_{k,\sigma}$ uniformly in $\sigma \in S^k$. Then for every $m,k \in \mathbb{N}$ the measures $\{\mathbb{P}_\sigma(L_{k,n}^{\otimes m})^{-1}\}_{n\in\mathbb{N},\,\sigma\in S^k}$ satisfy a uniform large deviation principle in the τ_Φ-topology on $\mathcal{M}_\Phi((S^k)^m)$ with the good rate function $J_{m,k}$.*

Remark 4.6. (a) Again (3.2) suffices to transfer the LDP to the products $L_{k,n}^{\otimes m}$, if $\mathcal{M}_1((S^k)^m)$ is endowed with the τ-topology.

(b) From [8, Exercise 4.1.48], we know that condition (3.2) guaranteed that there exists precisely one $\nu \in \mathcal{M}_1(S)$ which is π-invariant: $\nu = \nu\pi$. Furthermore, Exercise 4.1.51(ii) says, that if μ in (3.5) is replaced by ν, then Condition 1.3 holds for the laws of L_n and this ν. From this observation $H(\pi \,|\, \nu) \le \beta(J_1(\pi) + \log r)$ follows immediately for $\pi \in \mathcal{M}_1(S)$. Given ν, the measure

$$\nu_k(dx) \equiv \nu(dx_1)\pi(x_1, dx_2)\cdots\pi(x_{k-1}, dx_k)$$

is the π_k-invariant measure. Thus we get Theorem 3.6 and Corollary 4.5, respectively, if we replace $\mu^{\otimes m}$ and $\mu_k^{\otimes m}$, respectively, by $\nu^{\otimes m}$ and $\nu_k^{\otimes m}$, respectively.

To extend Theorem 3.6 to the process level, we need to introduce some additional notation and a special topology. For $n \in \mathbb{N}$ let

$$\Omega \ni \omega \mapsto R_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\theta_i(\omega)} \in \mathcal{M}_1(\Omega)$$

denote the *empirical-process measure,* where θ_i: $\Omega \to \Omega$ is the shift defined by $\theta_i(\omega) = (X_i(\omega),\ X_{i+1}(\omega), \dots)$ for every $i \in \mathbb{N}_0$. For $k \in \mathbb{N}$ let $\pi_{1,k}$: $\Omega \to S^k$ with $\pi_{1,k}(\omega) = (\omega_1, \dots, \omega_k)$, $\omega = (\omega_i)_{i\in\mathbb{N}} \in \Omega$, denote the projection onto the first k components. Given $m \ge 2$, define $\pi_{m,k}$: $\Omega^m \to (S^k)^m$ by $\pi_{m,k}(\omega_1, \dots, \omega_m) = (\pi_{1,k}(\omega_1), \dots, \pi_{1,k}(\omega_m))$ for all $(\omega_1, \dots, \omega_m) \in \Omega^m$. Note that

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

Similar as in [8, p. 174] we will introduce a *projective-limit τ_Φ-topology* on a subset of the space $\mathcal{M}_1(\Omega^m)$. Given a set Φ of $\mathcal{S}^{\otimes lm}$-measurable functions from $(S^l)^m$ to $\mathbb{R}$, the set $\mathcal{M}_\Phi((S^l)^m)$ and the τ_Φ-topology are defined as in the introduction. Let $\Phi_l \equiv \Phi$ and for $k \ge l$ define Φ_k to be the set consisting of $B((S^k)^m)$ and all functions of the form $\varphi \circ \pi_{k,l}^m$, where $\varphi \in \Phi$ and the projection $\pi_{k,l}^m$: $S^{km} \to S^{lm}$ is defined by $\pi_{k,l}^m(s_1, \dots, s_m) = (\pi_{k,l}^1(s_1), \dots, \pi_{k,l}^1(s_m))$ for $S_j \in S^k$, $j \in \{1, \dots, m\}$ and $\pi_{k,l}^1$: $S^k \to S^l$ is the usual projection map. Using Φ_k the set $\mathcal{M}_{\Phi_k}((S^k)^m)$ and the τ_{Φ_k}-topology can be defined as usual. The set $\mathcal{M}_\Phi(\Omega^m)$ should now be defined to be the set of all $\nu \in \mathcal{M}_1(\Omega^m)$ such that for every $k \ge l$ and every projection $\pi_{m,k}$, the measure $\nu\pi_{m,k}^{-1}$ belongs to $\mathcal{M}_{\Phi_k}((S^k)^m)$. The projective-limit τ_Φ-topology on $\mathcal{M}_\Phi(\Omega^m)$ is now defined to be the coarsest topology which makes all projections $\mathcal{M}_1(\Omega^m) \ni \mu \mapsto \mu\pi_{m,k}^{-1} \in \mathcal{M}_1((S^k)^m)$, $k \in \mathbb{N}$, $k \ge l$, continuous with respect to the τ_{Φ_k}-topology on $\mathcal{M}_{\Phi_k}((S^k)^m)$. Note that the given construction can be generalized to projections $\widetilde{\pi}_{k,l}^1$: $S^k \to S^l$ defined by $\widetilde{\pi}_{k,l}^1(s_1, \dots, s_k) = (s_{\tau(1)}, \dots, s_{\tau(l)})$ with strictly increasing τ: $\{1, \dots, l\} \to \{1, \dots, k\}$ and to projections $\widetilde{\pi}_{1,k}$: $\Omega \to S^k$ defined by $\widetilde{\pi}_{1,k}((s_j)_{j\in\mathbb{N}}) = (s_{\tau(1)}, \dots, s_{\tau(k)})$ with strictly increasing τ: $\{1, \dots, k\} \to \mathbb{N}$. We observe that the sets $\mathcal{M}_\Phi((S^l)^m)$ and $\mathcal{M}_{\Phi_l}((S^l)^m)$ as well as their topologies coincide. If the maps in Φ have all exponential moments with respect to a measure $\mu^{\otimes lm}$, then the same is true for the maps in Φ_k with respect to $\mu^{\otimes km}$ for every $k \ge l$. If we take $\Phi = B(S^l)^m$ then $\mathcal{M}_1((S^l)^m) = \mathcal{M}_\Phi((S^l)^m)$ and τ_Φ coincides with the τ-topology on this set. The given construction yields to the projective-limit strong topology on $\mathcal{M}_1(\Omega^m)$, which is the coarsest topology such that the projections

$$\mathcal{M}_1(\Omega^m) \ni \mu \mapsto \mu\pi_{m,k}^{-1} \in \mathcal{M}_1((S^k)^m), \quad k \in \mathbb{N},$$

are continuous with respect to the strong topology on $\mathcal{M}_1((S^k)^m)$. As in [8, (4.4.11)] we define the process-level rate function $J_{1,\infty}\colon \mathcal{M}_1(\Omega) \to [0,\infty]$ by

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

An alternative representation for $J_{1,\infty}$ is given in [8, (4.4.16)]. It shows that $J_{1,\infty}(\mu) = \infty$ if μ is not shift-invariant: as above we define for every integer $m \geq 2$ the rate function $J_{m,\infty}\colon \mathcal{M}_1(\Omega^m) \to [0,\infty]$ by

$$J_{m,\infty}(\mu) = \begin{cases} J_{1,\infty}(\nu), & \text{if } \nu^{\otimes m} = \mu \\ \infty, & \text{otherwise.} \end{cases} \tag{4.7}$$

Lemma 4.8. *If* $\mu \in \mathcal{M}_1(\Omega^m)$, *then*

$$J_{m,\infty}(\mu) = \sup_{k\in\mathbb{N}} J_{m,k}(\mu\pi_{m,k}^{-1}). \tag{4.9}$$

Proof. If there exists a $\nu \in \mathcal{M}_1(\Omega)$ with $\mu = \nu^{\otimes d}$ then $\mu\pi_{m,k}^{-1} = (\nu\pi_{1,k}^{-1})^{\otimes m}$, hence $J_{m,k}(\mu\pi_{m,k}^{-1}) = J_{1,k}(\nu\pi_{1,k}^{-1})$ for all $k \in \mathbb{N}$ and (4.9) holds.

Consider $\mu \in \mathcal{M}_1(\Omega^m)$ satisfying $\mu \neq \nu^{\otimes m}$ for all $\nu \in \mathcal{M}_1(\Omega)$ and assume that, for every $k \in \mathbb{N}$, there exists $\nu_k \in \mathcal{M}_1(S^k)$ satisfying $\mu\pi_{m,k}^{-1} = \nu_k^{\otimes m}$. By Kolmogorov's consistency theorem, the consistent family $\{\nu_k\}_{k\in\mathbb{N}}$ gives rise to a measure $\widetilde{\nu} \in \mathcal{M}_1(\Omega)$ with $\widetilde{\nu}\pi_{1,k}^{-1} = \nu_k$ for all $k \in \mathbb{N}$. Hence $\mu\pi_{m,k}^{-1} = \widetilde{\nu}^{\otimes m}\pi_{m,k}^{-1}$ for all $k \in \mathbb{N}$ and therefore $\mu = \widetilde{\nu}^{\otimes m}$ because the algebra $\bigcup_{k\in\mathbb{N}} \pi_{m,k}^{-1}((\mathcal{S}^{\otimes k})^{\otimes m})$ generates $\mathcal{F}^{\otimes m}$. This contradicts the assumption on μ, hence there exists $k \in \mathbb{N}$ with $\mu\pi_{m,k}^{-1} \neq \nu^{\otimes m}$ for all $\nu \in \mathcal{M}_1(S^k)$. Since $J_{m,\infty}(\mu) = \infty$ by (4.7) and $J_{m,k}(\mu\pi_{m,k}^{-1}) = \infty$ by (4.3) the identity (4.9) holds. □

Using Lemma 4.8 and Corollary 4.5 we obtain the following result:

Theorem 4.10. *Assume that Condition 1.7 holds for* Φ *and with respect to integration with respect to* $\mathbb{P}_\sigma$ *uniformly in* $\sigma \in S$, *then for every* $m \in \mathbb{N}$ *the measures* $\{\mathbb{P}_\sigma(R_n^{\otimes m})^{-1}\}_{n\in\mathbb{N},\sigma\in S}$ *satisfy a uniform LDP in the projective-limit* τ_Φ*-topology on* $\mathcal{M}_\Phi(\Omega^m)$ *with the good rate function* $J_{m,\infty}$. *In particular the level sets* $\{\mu \in \mathcal{M}_\Phi(\Omega^m),\ J_{m,\infty}(\mu) \leq r\}$ *are* τ_Φ *compact with respect to the projective-limit* τ_Φ*-topology.*

Proof. Since 1.7 holds for every Φ_k, $k \geq l$ and $\mu^{\otimes km}$, we can apply Corollary 4.5 in combination with [8, Theorem 4.4.27] and Lemma 4.8 to get the result for modifications as constructed in the proof of Theorem 3.6: Indeed the arguments of the proof of Theorem 4.4.27 work, because the level-sets of $J_{m,k}$ are τ_{Φ_k}-sequentially compact. To see this, note that the level sets of $H(\cdot\,|\,\mu^{\otimes mk})$ are τ_{Φ_k}-sequentially compact, if Condition 1.6 holds (see [14, Lemma 2.1(c)]). Since in addition 1.3 implies

$$H(\nu\,|\,\mu^{\otimes mk}) \leq B(J_{m,k}(\nu) + \log(2M))$$

for $\nu \in \mathcal{M}_1((S^k)^m)$, the assertion holds. □

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] W. Bryc and A. Dembo, *Large deviations and strong mixing*, Ann. Inst. H. Poincaré. Probab. Statist. **32** (1996), no. 4, 549–569.

[5] I. Csiszár, *Sanov property, generalized I-projection and a conditional limit theorem*, Ann. Probab. **12** (1984), no. 3, 768–793.

[6] T. Daras, *Large and moderate deviations for the empirical measure of an exchangeable sequence*, Statist. Probab. Lett. **36** (1997), 91–100.

[7] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, second ed., Springer-Verlag, New York, 1998.

[8] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Pure and Applied Mathematics, vol. 137, Academic Press, San Diego, 1989.

[9] I. H. Dinwoodie and S. L. Zabell, *Large deviations for exchangeable random vectors*, Ann. Probab. **20** (1992), no. 3, 1147–1166.

[10] ———, *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.

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

[12] J. Dugundji, *Topology*, Allyn and Bacon, Boston, 1966.

[13] 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.

[14] 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.

[15] ———, *Large deviations of U-empirical measures in strong topologies and applications*, submitted, 1999.

[16] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.

(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`