To appear in the Annales de l'Institut Henri Poincaré,
Probabilités et Statistiques (2002)

LARGE DEVIATIONS OF U-EMPIRICAL MEASURES IN STRONG TOPOLOGIES AND APPLICATIONS*

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 i. i. d. random variables take values in a measurable (not necessarily Polish) space $(S, \mathcal{S})$. The results can be formulated on suitable subsets of all probability measures on $(S^m, \mathcal{S}^{\otimes m})$. We endow the spaces with topologies, which are stronger than the usual τ-topology and which make integration with respect to certain unbounded, Banach-space valued functions a continuous operation. A special feature is the non-convexity of the rate function for $m \geq 2$. Improved versions of LDPs for Banach-space valued U- and V-statistics are obtained as a particular application. Some further applications concerning the Gibbs conditioning principle and a process level LDP are mentioned.

1. Introduction, statement of results, and applications

Let $(S, \mathcal{S}, \mu)$ be a probability space, let $(\Omega, \mathcal{A}, \mathbb{P}) \equiv (S^{\mathbb{N}}, \mathcal{S}^{\otimes \mathbb{N}}, \mu^{\otimes \mathbb{N}})$ be the product space, and let $\{X_i\}_{i \in \mathbb{N}}$ be the coordinate projections from Ω to S, forming an i.i.d. sequence with $\mathcal{L}(X_i) = \mu$. For every $n \in \mathbb{N}$ the empirical measure is defined by $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. The *$m$-fold products* $L_n^{\otimes m}\colon \Omega \to \mathcal{M}_1(S^m)$ of these empirical measures are also expressible as

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

The *U-empirical measure of order m* is defined by

$$L_n^m = \frac{1}{n_{(m)}} \sum_{(i_1, \ldots, i_m) \in I(m,n)} \delta_{(X_{i_1}, \ldots, X_{i_m})} \tag{1.2}$$

for all integers $n \geq m$, where $n_{(m)} \equiv \prod_{k=0}^{m-1} (n-k)$, the set $I(m,n) \subset \{1, \ldots, n\}^m$ consists of all m-tuples with pairwise different components, and δ_x denotes the probability measure concentrated at x.

In this paper we derive large deviation principles (LDP) for $\{L_n^m\}_{n \geq m}$ and $\{L_n^{\otimes m}\}_{n \in \mathbb{N}}$ in a strong topology on a restricted space $\mathcal{M}_1^{\Phi}(S^m)$ of probability measures, which is determined by a rich class of functions: Instead of bounded real-valued functions as for the usual τ-topology, we consider more general collections Φ of measurable and possibly unbounded functions taking values in a real separable Banach space E and satisfying appropriate moment conditions. These extensions

1991 *Mathematics Subject Classification.* 60F10 (primary); 62H10 and 28A35 (secondary).

Key words and phrases. Large deviations, empirical measures, Sanov's theorem, strong topology, U-statistics, von Mises statistics, Gibbs conditioning principle.

*Research partially supported by the Swiss National Foundation, Contract No. 21-298333.90.

enable us to derive LDPs for Banach-space valued U- and V-statistics as a particular application, see Subsection 1.2. These results improve the LDP for $\mathbb{R}^d$-valued U- and V-statistics obtained in [13], because we use weaker moment conditions and make no additional assumptions on the state space S. As a further application of our extension to U-empirical measures improvements of the Gibbs conditioning principle for interacting ensembles of particles and an extension of the LDP for U-empirical measures to a process level LDP are considered.

Among the techniques used, the change of measure method, the projective limit approach and the use of the existence of almost regular partitions of complete hypergraphs due to Baranyai [4] should be mentioned.

1.1. **Large deviation principles.** The large deviations of $\{L_n\}_{n\in\mathbb{N}}$ have been studied in several papers. Sanov [23] considered the problem when $S=\mathbb{R}$ and the space of probability measures $\mathcal{M}_1(S)$ is endowed with the weak topology. Donsker and Varadhan [10] and Bahadur and Zabell [3] (see also Azencott [2]) proved – with different approaches – a LDP for $\{L_n\}_{n\in\mathbb{N}}$ when S is a Polish space with Borel σ-algebra $\mathcal{S}$, again in the weak topology of $\mathcal{M}_1(S)$. Groeneboom, Oosterhoff and Ruymgaart [18] obtained a stronger result in which S is a Hausdorff topological space with Borel σ-algebra $\mathcal{S}$ and $\mathcal{M}_1(S)$ is endowed with the $\tau_1(\mathbb{R})$-topology, that is, the coarsest topology which makes the maps $\mathcal{M}_1(S)\ni\nu\mapsto\int_S\varphi\,d\nu$ continuous for all φ in the space $B(S,\mathbb{R})$ of bounded, real-valued, $\mathcal{S}$-measurable functions on S. De Acosta [1] proved the LDP for $\{L_n\}_{n\in\mathbb{N}}$ in the $\tau_1(\mathbb{R})$-topology setting when $(S,\mathcal{S})$ is a measurable space.

An extension in another direction was obtained by Wu [27, Section 2.4.4 and Section 4.1]. He considers a Polish space S with Borel σ-algebra $\mathcal{S}$ and a collection Φ consisting of all $\mathcal{S}$-measurable functions $f\colon S\to\mathbb{R}$, which are dominated in absolute value by some multiple of a single measurable reference function $\psi\colon S\to[1,\infty]$ with $\int_S\exp(\alpha\psi)\,d\mu<\infty$ for all $\alpha>0$. On a suitable subset of $\mathcal{M}_1(S)$ he defines a topology such that $\nu\mapsto\int_S\varphi\,d\nu$ is continuous for every $\varphi\in\Phi$, and proves LDP for $\{L_n\}_{n\in\mathbb{N}}$ with respect to this topology.

In this paper, we obtain large deviations results which include de Acosta's and Wu's results for general measurable spaces (not just Polish spaces), and extend them in several directions; the extension to U-empirical measures being the most important one.

To formulate our results, we need some additional notations. Let $(E,\|\cdot\|_E)$ be a separable real Banach space with Borel σ-algebra $\mathcal{E}$. We always exclude the case $E=\{0\}$. Given $m\in\mathbb{N}$ and a measurable space $(S,\mathcal{S})$ as before, we consider the set $\mathcal{M}_1(S^m)$ of probability measures on the product space S^m, equipped with the product σ-algebra $\mathcal{S}^{\otimes m}$. Let Φ be a collection of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable functions $\varphi\colon S^m\to E$ containing the set $B(S^m,E)$ of all bounded measurable ones. 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 the topology $\tau_1^\Phi(E)$ coincides with the usual $\tau_1(\mathbb{R})$-topology introduced above. The

σ-algebra on $\mathcal{M}_1(S^m)$ is defined to be the smallest one such that $\mathcal{M}_1^\Phi(S^m)$ and all the maps $\mathcal{M}_1(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ with $\varphi \in B(S^m, E)$ are measurable.

We consider the following three exponential moment conditions for Φ:

Condition 1.3 (Weak Cramér condition). *For every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that*

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

Condition 1.4 (Strong Cramér condition). *For every $\varphi \in \Phi$ and every $\alpha > 0$,*

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

Condition 1.5. *For every $\varphi \in \Phi$ there exists at least one $\alpha_\varphi > 0$ such that*

$$\int_{S^m} \exp(\alpha_\varphi \|\varphi \circ \pi_\tau\|_E) \, d\mu^{\otimes m} < \infty \tag{1.6}$$

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 $s = (s_1, \dots, s_m) \in S^m$.

Note that all three conditions are satisfied in the case $\Phi = B(S^m, E)$. Condition 1.4 is used to prove the large deviation upper bound in the $\tau_1^\Phi(E)$-topology for the U-empirical measures defined by (1.2). Due to their immediate statistical application, see (1.12) in Subsection 1.2, we named them U-empirical measures of order m. They seem to be the proper generalization of the usual empirical measures $\{L_n\}_{n\in\mathbb{N}}$, because there is no dependence within the individual terms of (1.2). With these U-empirical measures it is possible to model a weak but long-range interaction between the i. i. d. random variables $\{X_i\}_{i\in\mathbb{N}}$. This also means that we leave the realm of independence for $m \geq 2$.

We need Condition 1.5 to transfer our LDP with respect to the $\tau_1^\Phi(E)$-topology from the U-empirical measures $\{L_n^m\}_{n\geq m}$ to the m-fold products $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$.

Note that the set $\mathcal{M}_1^\Phi(S^m)$ is convex and contains the Dirac measure δ_x for every $x \in S^m$, hence $L_n^{\otimes m}$ and L_n^m take values in $\mathcal{M}_1^\Phi(S^m)$ and both mappings turn out to be measurable. Let us recall the definition of the relative entropy $H_m(\nu \,|\, \tilde{\nu})$ of $\nu \in \mathcal{M}_1(S^m)$ with respect to $\tilde{\nu} \in \mathcal{M}_1(S^m)$:

$$H_m(\nu \,|\, \tilde{\nu}) = \begin{cases} \int_{S^m} f \log f \, d\tilde{\nu}, & \text{if } \nu \ll \tilde{\nu} \text{ and } f = \frac{d\nu}{d\tilde{\nu}}, \\ \infty & \text{otherwise.} \end{cases}$$

The rate function $J_m\colon \mathcal{M}_1(S^m) \to [0, \infty]$ is defined by

$$J_m(\nu) = \begin{cases} \frac{1}{m} H_m(\nu \,|\, \mu^{\otimes m}), & \text{if } \nu = \nu_1^{\otimes m}, \\ \infty & \text{otherwise,} \end{cases}$$

where ν_1 denotes the first marginal of ν. Note that $\frac{1}{m} H_m(\nu_1^{\otimes m} \,|\, \mu^{\otimes m}) = H_1(\nu_1 \,|\, \mu)$. For every $B \subset \mathcal{M}_1(S^m)$ define $J_m(B) = \inf_{\nu \in B} J_m(\nu)$. Note that, if there exists $A \in \mathcal{S}$ satisfying $0 < \mu(A) < 1$, then the rate function J_m is non-convex for all $m \geq 2$. To see this, define the conditional probability measures $\nu = \mu(\,\cdot\,|A)$ and $\tilde{\nu} = \mu(\,\cdot\,|A^c)$ and show that $\frac{1}{2}(\nu^{\otimes m} + \tilde{\nu}^{\otimes m})$ is not a product measure. For every level $l \in [0, \infty)$ define the level set $K(J_m, l) = \{\, \nu \in \mathcal{M}_1(S^m) \mid J_m(\nu) \leq l \,\}$. Our main large deviation results, which we prove in Section 2, are the following theorems.

Theorem 1.7 (Large deviations of U-empirical measures of order m)**.**

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

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in B) \geq -J_m(\mathrm{int}_{\tau_1^\Phi(E)}(B)), \tag{1.8}$$

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

(b) *If Condition* 1.3 *holds, then* $K(J_m, l) \subset \mathcal{M}_1^\Phi(S^m)$ *for every level* $l \in [0, \infty)$*.*

(c) *If Condition* 1.4 *holds, then* $K(J_m, l)$ *is* $\tau_1^\Phi(E)$*-compact and sequentially* $\tau_1^\Phi(E)$*-compact for every* $l \in [0, \infty)$*.*

(d) *If Condition* 1.4 *holds, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n^m \in B) \leq -J_m(\mathrm{cl}_{\tau_1^\Phi(E)}(B)) \tag{1.9}$$

for every measurable $B \subset \mathcal{M}_1(S^m)$*, where* $\mathrm{cl}_{\tau_1^\Phi(E)}(B)$ *denotes the closure of the set* $B \cap \mathcal{M}_1^\Phi(S^m)$ *with respect to the* $\tau_1^\Phi(E)$*-topology.*

Theorem 1.10 (Large deviations of m-fold products of empirical measures)**.**

(a) *If Condition* 1.5 *holds, then* (1.8) *is true with* $L_n^{\otimes m}$ *in place of* L_n^m*.*

(b) *If Conditions* 1.4 *and* 1.5 *hold, then* (1.9) *is true with* $L_n^{\otimes m}$ *in place of* L_n^m*.*

Remark 1.11. Due to the following results of A. Schied, we think that the exponential moment conditions of Theorem 1.7(b) and (c) are optimal: If $\Phi = B(S, \mathbb{R}) \cup \{\varphi\}$ with a measurable $\varphi\colon S \to [0, \infty)$ and if $K(J_1, l) \subset \mathcal{M}_1^\Phi(S)$ for one $l > 0$, then Φ satisfies Condition 1.3, see [25, Proposition 1]. If $K(J_1, l)$ is a $\tau_1^\Phi(\mathbb{R})$-compact subset of $\mathcal{M}_1^\Phi(S)$ for this collection Φ and one $l > 0$, then Condition 1.4 is satisfied, see [25, Theorem 2] or [24, Satz 2.15]. The compactness of the level sets is a crucial ingredient for our proofs of the upper bounds of Theorems 1.7(d) and 1.10(b).

Theorem 1.7, specialized to $m = 1$ and $E = \mathbb{R}$, already combines and extents the aforementioned large deviation results of de Acosta and Wu for the empirical measures $\{L_n\}_{n\in\mathbb{N}}$ to the $\tau_1^\Phi(\mathbb{R})$-topology. Note that the lower bound in Theorem 1.7(a) does not make use of Condition 1.3. But if Condition 1.3 does not hold, then the lower bound might loose its strength because measures $\nu \in B$ with $J_m(\nu) < \infty$ might not be contained in $\mathcal{M}_1^\Phi(S^m)$.

A special feature of the strong topologies used in the above two theorems is the fact that the formation of product measures can be a discontinuous operation. An example suggested by Y. Peres (see [8, Exercise 7.3.18]) uses the compact unit interval $S = [0, 1]$ equipped with the Borel σ-algebra and shows that the map $\mathcal{M}_1(S) \ni \nu \mapsto \nu^{\otimes m} \in \mathcal{M}_1(S^m)$, considered at the Lebesgue measure, is discontinuous with respect to the $\tau_1(\mathbb{R})$-topologies on $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$. Therefore, the discontinuity of the product-measure operation is not an artifact of certain exotic measurable spaces or measures. Example 1.16 below may serve as a further illustration that products of "empirical measures" can exhibit a strange behaviour with respect to the $\tau_1(\mathbb{R})$-topology. Note that, due to the possible discontinuity of the map $\mathcal{M}_1(S) \ni \nu \mapsto \nu^{\otimes m}$, our large deviation results do not follow easily via the contraction principle in our general setting; see Subsection 1.3, however.

Theorems 1.7 and 1.10 show that the geometry of the Banach space $(E, \|\cdot\|)$ does not influence the LDPs. This is in contrast to the moderate deviation principle in [15], which requires an assumption on the type of $(E, \|\cdot\|)$. Without it, the moderate deviation lower bound may fail, see [15, Example 2.26].

For the proofs of the lower bounds of the above theorems, we adopt the method outlined in [9, Exercise 3.2.23(ii)] and combine it with the law of large numbers for U-statistics and V-statistics, respectively. To prove the upper bounds, we basically use de Acosta's projective limit approach contained in [1] combined with a Banach-space version of Cramér's theorem, which is due to Donsker and Varadhan. (Choosing $m = 1$, $S = E$ and $\Phi = B(S, E) \cup \{\mathrm{id}_E\}$, their result can be recovered from Theorem 1.7 by using the contraction principle [8, Theorem 4.2.1] and identifying the rate function via [10, Theorem 5.2(iv)].) Of course, we have to cope with the unbounded, Banach-space valued functions in Φ, which is one reason for the $\tau_1^\Phi(E)$-topology to be finer than the usual $\tau_1(\mathbb{R})$-topology.

There are two non-trivial problems in transferring de Acosta's approach to the case $m \geq 2$: The dependence of the different terms in (1.2) and the non-convexity of the rate function J_m. The first problem is treated by using the existence of almost regular partitions of complete hypergraphs, see Baranyai [4]. This result allows us to conveniently decompose the sum in (1.2) into an n-dependent number of partial sums, each of which consists of independent terms. To circumvent the non-convexity of J_m, we use the convexity of the relative entropy $H_m(\,\cdot\,|\,\mu^{\otimes m})$ on $\mathcal{M}_1(S^m)$ and the fact that $L_n^{\otimes m}$ is a product measure as well as a good approximation for L_n^m. Finally we have to show in Lemma 2.9 that the infimum of $H_m(\,\cdot\,|\,\mu^{\otimes m})$ over all "η-approximate product measures" of a $\tau_1^\Phi(E)$-closed set C tends to the infimum over all product measures in C as $\eta \downarrow 0$.

1.2. **Application to *U*- and *V*-statistics.** We want to derive large deviation results for Banach-space valued U-statistics and V-statistics (also called von Mises statistics) from Theorems 1.7 and 1.10, respectively. Many statistics in common use are members of these two classes and many other statistics may be approximated by a member of one of these classes. Particularly in the field of non-parametric statistics, the significance of the two classes is made plain, see for example the monographs [5] and [20] as well as the extensive list of references given there. For an $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable map $\varphi\colon S^m \to E$ the U- and V-statistics of degree m with kernel function φ are defined by

$$U_n^m(\varphi) = \int_{S^m} \varphi \, dL_n^m \quad \text{and} \quad V_n^m(\varphi) = \int_{S^m} \varphi \, dL_n^{\otimes m} \tag{1.12}$$

for all $n \geq m$ or all $n \in \mathbb{N}$, respectively. Define $\Phi = B(S^m, E) \cup \{\varphi\}$. Then the statistics defined in (1.12) are compositions of L_n^m and $L_n^{\otimes m}$, respectively, with the $\tau_1^\Phi(E)$-continuous functional $\mathcal{M}_1^\Phi(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$, and the contraction principle [8, Theorem 4.2.1] immediately leads to the following theorem which extends results obtained in [13]. Note that the following theorem uses weaker moment conditions.

Theorem 1.13. *Assume that the kernel function $\varphi\colon S^m \to E$ satisfies the strong Cramér Condition 1.4. Then the following assertions hold:*

(a) *The U-statistics $\{U_n^m(\varphi)\}_{n \geq m}$ satisfy a full LDP with the good rate function $J_{\varphi,m}\colon E \to [0, \infty]$ given by*

$$J_{\varphi,m}(x) = \inf\Big\{ J_m(\nu) \;\Big|\; \nu \in \mathcal{M}_1^\Phi(S^m), \int_{S^m} \varphi \, d\nu = x \Big\}. \tag{1.14}$$

(b) *If, in addition, φ satisfies Condition 1.5, then the V-statistics $\{V_n^m(\varphi)\}_{n \in \mathbb{N}}$ also satisfy a full LDP with the good rate function $J_{\varphi,m}$.*

If Condition 1.3 holds for Φ, then $K(J_m,\infty) \equiv \bigcup_{l>0} K(J_m,l)$ is contained in $\mathcal{M}_1^\Phi(S^m)$ by Theorem 1.7(b). Since $J_m(\nu)=\infty$ for all $\nu\in\mathcal{M}_1^\Phi(S^m)\setminus K(J_m,\infty)$, it follows that the rate function $J_{\varphi,m}$ defined by (1.14) coincides with the one in [12, Theorem 5.1]. For special kernels $\varphi\in B(S^m,\mathbb{R})$ for which there exists a $\tilde\varphi\in B(S,\mathbb{R})$ such that $\varphi(s)=\tilde\varphi(s_1)\cdots\tilde\varphi(s_m)$ for all $s=(s_1,\dots,s_m)\in S^m$, additional representations of the rate function $J_{\varphi,m}$ can be derived by using [7, Theorem 3.1], see [21, Corollary 4.10], for example.

1.3. **Further applications.** Let us briefly show how our results can be used to improve the *Gibbs conditioning principle,* cf. [8, Section 7.3]. As before, let Φ be a set of measurable functions $\varphi\colon S^m\to E$ containing $B(S^m,E)$. Let $\{A_\delta\}_{\delta>0}$ be a nested collection of measurable subsets of $\mathcal{M}_1^\Phi(S^m)$, where nested means that $A_\delta\subset A_{\delta'}$ whenever $\delta\le\delta'$. In addition, let $\{F_\delta\}_{\delta>0}$ be nested $\tau_1^\Phi(E)$-closed subsets of $\mathcal{M}_1^\Phi(S^m)$, not necessarily measurable, such that $A_\delta\subset F_\delta$ for all $\delta>0$. Define $F_0=\bigcap_{\delta>0}F_\delta$ and $A_0=\bigcap_{\delta>0}A_\delta$.

Assume Condition 1.4 and that there exists a $\hat\nu\in A_0$ (not necessarily unique) such that $J_m(\hat\nu)=J_m(F_0)<\infty$ and

$$\lim_{n\to\infty}\hat\nu_1^{\otimes\mathbb{N}}(\{L_n^m\in A_\delta\})=1 \tag{1.15}$$

for every $\delta>0$, where $\hat\nu_1$ denotes the first marginal of $\hat\nu$. Then, by adapting the proof of [8, Theorem 7.3.3], the following result can be obtained.

(a) The set $\mathcal{E}\equiv\{\,\nu\in F_0\mid J_m(\nu)=J_m(F_0)\,\}$ is nonempty, $\tau_1^\Phi(E)$-compact and sequentially $\tau_1^\Phi(E)$-compact.

(b) If $\Gamma\subset\mathcal{M}_1(S^m)$ is measurable and $\mathcal{E}\subset\operatorname{int}_{\tau_1^\Phi(E)}(\Gamma)$, then

$$\limsup_{\delta\downarrow0}\,\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(L_n^m\notin\Gamma\,|\,L_n^m\in A_\delta)<0.$$

To apply this result, let $U\colon S^m\to\mathbb{R}$ be an $\mathcal{S}^{\otimes m}$-measurable function such that $\int_{S^m}\exp(\alpha|U|)\,d\mu^{\otimes m}<\infty$ for all $\alpha>0$. Then $\Phi\equiv B(S^m,\mathbb{R})\cup\{U\}$ satisfies Condition 1.4. Furthermore, the map $\Psi\colon\mathcal{M}_1^\Phi(S^m)\to\mathbb{R}$ with $\Psi(\nu)\equiv\int_{S^m}U\,d\nu-1$ is measurable and $\tau_1^\Phi(\mathbb{R})$-continuous. The sets $A_\delta\equiv F_\delta\equiv\{\,\nu\in\mathcal{M}_1^\Phi(S^m)\mid|\Psi(\nu)|\le\delta\,\}$ with $\delta>0$ are, therefore, nested, measurable and $\tau_1^\Phi(\mathbb{R})$-closed. Obviously, $A_0=F_0=\{\,\nu\in\mathcal{M}_1^\Phi(S^m)\mid\Psi(\nu)=0\,\}$ and $A_0\subset\operatorname{int}_{\tau_1^\Phi(\mathbb{R})}(A_\delta)$ for every $\delta>0$. If there exists a $\nu\in A_0$ with $l\equiv J_m(\nu)<\infty$, then $A_0\cap K(J_m,l)$ is $\tau_1^\Phi(\mathbb{R})$-compact by Theorem 1.7(c). Hence, the lower $\tau_1^\Phi(\mathbb{R})$-semicontinuous rate function J_m attains the infimum $J_m(A_0)$ at a $\hat\nu\in A_0$ and (1.15) holds.

Using the Dawson–Gärtner projective limit approach, we can extend Theorem 1.7 to a *LDP on process level* with respect to a projective limit $\tau_{1,\mathrm{pl}}^\Phi(E)$-topology on a restricted set $\mathcal{M}_1^\Phi(\Omega)$ of probability measures on the countable product space $(\Omega,\mathcal{A})=(S^{\mathbb{N}},\mathcal{S}^{\otimes\mathbb{N}})$. Let us introduce the necessary notation. Using (1.2), define $R_n\colon\Omega\to\mathcal{M}_1(\Omega)$ by $R_n=L_n^n\otimes\delta_{(X_{n+1},X_{n+2},\dots)}$ for every $n\in\mathbb{N}$. For every $m\in\mathbb{N}$ let $\varrho_m\colon\Omega\to S^m$ with $\varrho_m(s)=(s_1,\dots,s_m)$ for $s=(s_i)_{i\in\mathbb{N}}\in\Omega$ be the canonical projection, and let $p_m\colon\mathcal{M}_1(\Omega)\to\mathcal{M}_1(S^m)$ with $p_m(\nu)=\nu\varrho_m^{-1}$ be the corresponding map to the marginal measure. Then

$$p_m(R_n)=\begin{cases}L_n^m\,, & \text{if } m\le n,\\ L_n^n\otimes\delta_{(X_{n+1},\dots,X_m)}\,, & \text{if } m>n,\end{cases}$$

for all $m,n\in\mathbb{N}$. Let Φ be a set of $\mathcal{A}$-$\mathcal{E}$-measurable functions $\varphi\colon\Omega\to E$ such that every $\varphi\in\Phi$ only depends on finitely many coordinates, i. e., there exist $m\in\mathbb{N}$ and

$\tilde{\varphi}\colon S^m \to E$ such that $\varphi = \tilde{\varphi} \circ \varrho_m$. We assume that Φ contains the set $B_{\mathrm{f}}(\Omega, E)$ of all bounded measurable functions $\varphi\colon \Omega \to E$ depending only on finitely many coordinates. We define

$$\mathcal{M}_1^\Phi(\Omega) = \Big\{ \nu \in \mathcal{M}_1(\Omega) \,\Big|\, \int_\Omega \|\varphi\|_E \, d\nu < \infty \text{ for all } \varphi \in \Phi \Big\}.$$

The projective limit $\tau_{1,\mathrm{pl}}^\Phi(E)$-topology on $\mathcal{M}_1^\Phi(\Omega)$ is defined to be the coarsest one such that $\mathcal{M}_1^\Phi(\Omega) \ni \nu \mapsto \int_\Omega \varphi \, d\nu$ is continuous for every $\varphi \in \Phi$. Define the rate function J_∞ on $\mathcal{M}_1(\Omega)$ by

$$J_\infty(\nu) = \begin{cases} H_1(\nu_1 \,|\, \mu), & \text{if } \nu = \nu_1^{\otimes \mathbb{N}}, \\ \infty & \text{otherwise.} \end{cases}$$

Using [8, Theorem 4.6.1] and its proof, the analogue of Theorem 1.7 holds for $\{R_n\}_{n\in\mathbb{N}}$ and rate function J_∞ with respect to the $\tau_{1,\mathrm{pl}}^\Phi(E)$-topology on $\mathcal{M}_1^\Phi(\Omega)$.

If S is a Polish space with Borel σ-algebra $\mathcal{S}$, then we have the weak topologies on $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$ available and we can use the continuity of the map $\nu \mapsto \nu^{\otimes m}$ with respect to the weak topologies [16, Chap. 3, Proposition 4.6]. Therefore, if the empirical measures $\{L_n\}_{n\in\mathbb{N}}$ satisfy a LDP in the weak topology, then the contraction principle implies that the products $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ satisfy a LDP in the weak topology on $\mathcal{M}_1(S^m)$. If there exist constants $\beta, M \in [1, \infty)$ and a reference measure $\tilde{\mu} \in \mathcal{M}_1(S^m)$ such that the inequality

$$\sup_{n\in\mathbb{N}} \bigg(\mathbb{E}\bigg[\exp\bigg(n \int_{S^m} V \, dL_n^{\otimes m} \bigg) \bigg] \bigg)^{1/n} \le M \int_{S^m} \exp(\beta V) \, d\tilde{\mu}$$

holds for all bounded measurable functions $V\colon S^m \to [0, \infty)$, then we can use [9, Lemma 3.2.19 and Theorem 3.2.21] to infer that $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ actually satisfy a LDP in the $\tau_1(\mathbb{R})$-topology on $\mathcal{M}_1(S^m)$. We used this approach in [14] for random variables $\{X_i\}_{i\in\mathbb{N}}$ which are dependent or not identically distributed.

Finally, let us mention that the techniques of [1, Theorem 1.2] allow the extension of LDPs like Theorems 1.7 and 1.10 to arbitrary sets of measures.

1.4. **Examples.** A special feature of the $\tau_1(\mathbb{R})$-topologies on $\mathcal{M}_1(S)$ and $\mathcal{M}_1(S^m)$ is the possible discontinuity of the map $\mathcal{M}_1(S) \ni \nu \mapsto \nu^{\otimes m} \in \mathcal{M}_1(S^m)$, see the example suggested by Y. Peres [8, Exercise 7.3.18]. The following example may serve as a further illustration that products of "empirical measures" can exhibit a strange behaviour with respect to the $\tau_1(\mathbb{R})$-topology.

Example 1.16. 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, contrary to the i. i. d. situation in Subsection 1.1, there is a heavy dependence between the $\theta_{i2^{-n}}(\omega)$ for $i \in \{0, \ldots, 2^n - 1\}$. 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 $\omega \in S$.

Next we want to show that, for every $\varphi \in L_1(\mu, E)$,

$$\mu\Big(\lim_{n\to\infty} \int_S \varphi\, dL_n = \int_S \varphi\, d\mu\Big) = 1. \tag{1.17}$$

Define the sub-σ-algebra $\mathcal{S}_n$ of $\mathcal{S}$ by $\mathcal{S}_n = \{\, A \in \mathcal{S} \mid A = \theta_{2^{-n}}(A)\,\}$ for every $n \in \mathbb{N}_0$ and note that $\mathcal{S}_{n+1} \subset \mathcal{S}_n$ and $E_\mu[\varphi|\mathcal{S}_n] = \int_S \varphi\, dL_n$. Therefore, $\{\int_S \varphi\, dL_n\}_{n\in\mathbb{N}_0}$ is a reversed martingale relative to $\{\mathcal{S}_n\}_{n\in\mathbb{N}_0}$ and it converges (strongly) μ-almost surely and in $L_1(\mu, E)$ to $g \equiv E_\mu[\varphi|\mathcal{S}_\infty]$ with $\mathcal{S}_\infty \equiv \bigcap_{n\in\mathbb{N}_0} \mathcal{S}_n$, see [6, Theorem 4]. Consider the Fourier coefficients $\hat{g}_{\psi,n} \equiv \int_S \psi(g(t))e_n(t)\,\mu(dt)$ for $n \in \mathbb{Z}$ and $\psi \in E^*$, where $e_n(t) \equiv \exp(-2\pi i n t)$. Given $n \in \mathbb{Z} \setminus \{0\}$, there exist $k \in \mathbb{N}_0$ and an odd $l \in \mathbb{Z}$ such that $n = 2^k l$. Since $g \circ \theta_{2^{-(k+1)}} = g$ and $e_{2^k l} \circ \theta_{2^{-(k+1)}} = -e_{2^k l}$, all coefficients $\hat{g}_{\psi,n}$ with $n \in \mathbb{Z} \setminus \{0\}$ vanish and, therefore, $\psi(g) = \hat{g}_{\psi,0} = \psi(\int_S \varphi\, d\mu)$ μ-almost surely [11, Chap. 1.5, Theorem 1]. Using the Hahn–Banach theorem and the separability of E, it follows that there exists a countable subset C of E^* with $\|\psi\|_{E^*} \le 1$ for every $\psi \in C$ such that $\|x\|_E = \sup_{\psi\in C} |\psi(x)|$ for all $x \in E$. Therefore, $g = \int_S \varphi\, d\mu$ μ-almost surely.

To show that the product measures $\{L_n \otimes L_n\}_{n\in\mathbb{N}_0}$ can go astray, 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 $\{\, (\theta_{i2^{-n}}(\omega), \theta_{j2^{-n}}(\omega)) \mid i, j \in \{0, 1, \ldots, 2^n - 1\}\,\}$, is contained in A for every $n \in \mathbb{N}_0$ and $\omega \in S$. Therefore, the analogue of (1.17) for product measures does not even hold for the $\mathcal{S} \otimes \mathcal{S}$-measurable indicator function $\varphi \equiv 1_A$.

Remark 1.18. There does *not* exist a scale $\{\varepsilon_n\}_{n\in\mathbb{N}_0}$ with $\varepsilon_n \downarrow 0$ such that the random measures $\{L_n\}_{n\in\mathbb{N}_0}$ from Example 1.16 satisfy a large deviation upper bound of the form

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

for all $\tau_1(\mathbb{R})$-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 $\nu \ne \mu$ 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, consider the set $C \equiv \{\, \nu \in \mathcal{M}_1(S) \mid \nu(A) \ge \mu(A) + 1/2\,\}$, where we construct the set $A \in \mathcal{S}$ as follows: Choose a subsequence $\{\varepsilon_{n_k}\}_{k\in\mathbb{N}}$ such that $\sum_{k\in\mathbb{N}} \varepsilon_{n_k} \le 1/2$ 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 $\mu(A) \le \sum_{k\in\mathbb{N}} \varepsilon_{n_k} \le 1/2$ as well as $L_{n_k}(A) \ge L_{n_k}(A_k) = 1$ on A_k for every $k \in \mathbb{N}$. Hence, as $k \to \infty$,

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

Therefore, we are still looking for an example of a sequence of random (if possible, empirical) measures $\{L_n\}_{n\in\mathbb{N}}$ which satisfy a LDP in the $\tau_1(\mathbb{R})$-topology, but for which the products $\{L_n^{\otimes m}\}_{n\in\mathbb{N}}$ for some $m \ge 2$ do *not* satisfy the corresponding LDP in the $\tau_1(\mathbb{R})$-topology on $\mathcal{M}_1(S^m)$.

2. Proofs of the large deviation principles

For every level $l \in [0,\infty)$ define the level set of the relative entropy by

$$K(H_m, l) = \{\, \nu \in \mathcal{M}_1(S^m) \mid H_m(\nu \,|\, \mu^{\otimes m}) \le l \,\}.$$

Note that $K(J_m, l) = C_m \cap K(H_m, lm)$, where $C_m \equiv \{\, \nu \in \mathcal{M}_1(S^m) \mid \nu = \nu_1^{\otimes m} \,\}$ is the set of product measures. Therefore, part (b) of the following lemma implies part (b) of Theorem 1.7.

Lemma 2.1. *Let $l \in [0,\infty)$.*

(a) *The set $K(H_m, l)$ is $\tau_1(\mathbb{R})$-compact and sequentially $\tau_1(\mathbb{R})$-compact.*
(b) *If Condition 1.3 holds, then $K(H_m, l) \subset \mathcal{M}_1^\Phi(S^m)$.*
(c) *If Condition 1.4 holds, then the identity on $K(H_m, l)$ is $\tau_1(\mathbb{R})$-$\tau_1^\Phi(E)$-continuous, hence both topologies coincide on this set and $K(H_m, l)$ is $\tau_1^\Phi(E)$-compact and also sequentially $\tau_1^\Phi(E)$-compact.*

Proof. (a) See [1, Lemma 2.1] for the $\tau_1(\mathbb{R})$-compactness and [17, Theorem 2.6] for the sequential $\tau_1(\mathbb{R})$-compactness.

(b) By convexity, $z \le e^{z-1}$ for all $z \in \mathbb{R}$. Substituting $z = x - t$ yields $x \le e^{x-t-1} + t$ for all $t, x \in \mathbb{R}$. Multiplication with $y = e^t$ gives the well-known estimate

$$xy \le e^{x-1} + y \log y \quad \text{for all } x \in \mathbb{R} \text{ and } y \in [0,\infty). \tag{2.2}$$

If $\nu \in \mathcal{M}_1(S^m)$ satisfies $H_m(\nu \,|\, \mu^{\otimes m}) \le l$, then there exists a density f of ν with respect to $\mu^{\otimes m}$. Given $\alpha > 0$, $A \in \mathcal{S}^{\otimes m}$ and $\varphi \in \Phi$, the estimate (2.2) leads to

$$\int_A \|\varphi\| \, d\nu = \frac{1}{\alpha} \int_A (\alpha \|\varphi\|) f \, d\mu^{\otimes m} \le \frac{1}{\alpha e} \int_A e^{\alpha\|\varphi\|} \, d\mu^{\otimes m} + \frac{1}{\alpha} \int_A f \log f \, d\mu^{\otimes m}. \tag{2.3}$$

Since $y \log y \ge -1/e$ for all $y \in [0,\infty)$,

$$\int_A \|\varphi\| \, d\nu \le \frac{1}{\alpha} \Big(\frac{1}{e} \int_A e^{\alpha\|\varphi\|} \, d\mu^{\otimes m} + \frac{1}{e} + l \Big). \tag{2.4}$$

Choosing $A = S^m$ and $\alpha = \alpha_\varphi$ in (2.4), part (b) follows.

(c) Given $\varphi \in \Phi$ and $\varepsilon > 0$, define $\alpha = (3/e + l)/\varepsilon$. By Condition 1.4, the dominated convergence theorem, and [22, Lemma V-2-4], there exists a measurable, finitely-valued function $\varphi_\varepsilon \colon S^m \to E$ such that $\int_{S^m} \exp(\alpha \|\varphi - \varphi_\varepsilon\|) \, d\mu^{\otimes m} \le 2$. Using (2.4), it follows that

$$\Big\| \int_{S^m} \varphi \, d\nu - \int_{S^m} \varphi_\varepsilon \, d\nu \Big\| \le \int_{S^m} \|\varphi - \varphi_\varepsilon\| \, d\nu \le \varepsilon \tag{2.5}$$

for all $\nu \in K(H_m, l)$. Hence, $K(H_m, l) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is $\tau_1(\mathbb{R})$-continuous, because it is the uniform limit of the $\tau_1(\mathbb{R})$-continuous functions $K(H_m, l) \ni \nu \mapsto \int_{S^m} \varphi_\varepsilon \, d\nu$ as $\varepsilon \downarrow 0$. Hence, the identity on $K(H_m, l)$ is $\tau_1(\mathbb{R})$-$\tau_1^\Phi(E)$-continuous and the $\tau_1^\Phi(E)$-compactness of $K(H_m, l)$ follows from part (a). Since the identity is bijective, it is a $\tau_1(\mathbb{R})$-$\tau_1^\Phi(E)$-homeomorphism on $K(H_m, l)$ [19, Chap. 5, Theorem 8], hence both topologies coincide on $K(H_m, l)$. Therefore, the sequential $\tau_1^\Phi(E)$-compactness also follows from part (a). □

Proof of Theorem 1.7(c)*:* The set C_m of all product measures is $\tau_1(\mathbb{R})$-closed because

$$C_m = \bigcap_{A_1,\dots,A_m \in \mathcal{S}} \Big\{ \nu \in \mathcal{M}_1(S^m) \,\Big|\, \nu(A_1 \times \cdots \times A_m) = \prod_{i=1}^m \nu_1(A_i) \Big\}. \tag{2.6}$$

Hence, by Lemma 2.1(a), the set $K(J_m, l) = C_m \cap K(H_m, lm)$ is $\tau_1(\mathbb{R})$-compact and sequentially $\tau_1(\mathbb{R})$-compact. The $\tau_1^\Phi(E)$-compactness and the sequential $\tau_1^\Phi(E)$-compactness of $K(J_m, l)$ now follow from Lemma 2.1(c). □

Lemma 2.7. *Let $\nu \in \mathcal{M}_1(S)$.*

(a) *If $\varphi \in L_1(\nu^{\otimes m}, E)$, then*

$$\lim_{n\to\infty} \int_{S^m} \varphi \, dL_n^m = \int_{S^m} \varphi \, d\nu^{\otimes m} \quad \nu^{\otimes \mathbb{N}}\text{-almost surely.}$$

(b) *If $\varphi \circ \pi_\tau \in L_1(\nu^{\otimes m}, E)$ for every $\tau\colon \{1, \ldots, m\} \to \{1, \ldots, m\}$, then*

$$\lim_{n\to\infty} \int_{S^m} \varphi \, dL_n^{\otimes m} = \int_{S^m} \varphi \, d\nu^{\otimes m} \quad \nu^{\otimes \mathbb{N}}\text{-almost surely.}$$

Proof. This lemma is a reformulation of the strong law of large numbers for U- and V-statistics, see for example [5, Theorems 3.1.1 and 3.3.2]. Note that it suffices to consider symmetric φ for the proof. □

Proof of Theorem 1.7(a)*:* Let ν be a measure in the $\tau_1^\Phi(E)$-interior of $B \cap \mathcal{M}_1^\Phi(S^m)$ with $J_m(\nu) < \infty$. Then $\Phi \subset L_1(\nu, E)$, $\nu = \nu_1^{\otimes m}$ and $J_m(\nu) = H_1(\nu_1 \,|\, \mu)$. Define $f = d\nu_1/d\mu$ and $F_n(s) = \prod_{i=1}^n f(s_i)$ for all $s = (s_i)_{i\in\mathbb{N}} \in S^{\mathbb{N}}$. By the definition of the $\tau_1^\Phi(E)$-topology, there exist $\varepsilon > 0$, $k \in \mathbb{N}$ and $\varphi_1, \ldots, \varphi_k \in \Phi$ such that the $\tau_1^\Phi(E)$-open set

$$C \equiv \left\{ \tilde\nu \in \mathcal{M}_1^\Phi(S^m) \,\middle|\, \left\| \int_{S^m} \varphi_i \, d\tilde\nu - \int_{S^m} \varphi_i \, d\nu \right\| < \varepsilon \text{ for every } i \in \{1, \ldots, k\} \right\}$$

is contained in the $\tau_1^\Phi(E)$-interior of $B \cap \mathcal{M}_1^\Phi(S^m)$. Note that C is a measurable subset of $\mathcal{M}_1(S^m)$ because $\mathcal{M}_1^\Phi(S^m)$ is measurable by definition. Define $D_n = \{L_n^m \in C,\ F_n > 0\}$ and note that $a_n \equiv \nu_1^{\otimes\mathbb{N}}(D_n) = \nu_1^{\otimes\mathbb{N}}(\{L_n^m \in C\})$ for every $n \in \mathbb{N}$. It follows from Lemma 2.7(a) that

$$\lim_{n\to\infty} \nu_1^{\otimes\mathbb{N}}(\{L_n^m \in C\}) = 1. \tag{2.8}$$

Choose $n_0 \in \mathbb{N}$ such that $a_n > 0$ for all $n \geq n_0$. Then, for every $n \geq n_0$,

$$\mathbb{P}(L_n^m \in B) \geq \mathbb{P}(L_n^m \in C) \geq \int_{D_n} \frac{1}{F_n} \, d\nu_1^{\otimes\mathbb{N}}.$$

Using Jensen's inequality, we obtain

$$\begin{aligned} \log \int_{D_n} \frac{1}{F_n} \, d\nu_1^{\otimes\mathbb{N}} &\geq \log a_n - \frac{1}{a_n} \int_{D_n} \log F_n \, d\nu_1^{\otimes\mathbb{N}} \\ &= \log a_n - \frac{1}{a_n} \int_{D_n} F_n \log F_n \, d\mu^{\otimes\mathbb{N}}. \end{aligned}$$

Since $x \log x \geq -1/e$ for all $x \in [0, \infty)$, it follows that

$$\begin{aligned} \int_{D_n} F_n \log F_n \, d\mu^{\otimes\mathbb{N}} &\leq \frac{1}{e} + \int_{S^{\mathbb{N}}} F_n \log F_n \, d\mu^{\otimes\mathbb{N}} \\ &= \frac{1}{e} + nH_1(\nu_1 \,|\, \mu) = \frac{1}{e} + nJ_m(\nu). \end{aligned}$$

The last three displays together yield

$$\log \mathbb{P}(L_n^m \in B) \geq \log a_n - \frac{1}{e a_n} - \frac{nJ_m(\nu)}{a_n}.$$

Using (2.8), Theorem 1.7(a) follows. □

With a small modification the same proof applies for the large-deviations lower bound for products of empirical measures.

Proof of Theorem 1.10(a): If $\nu \in \mathcal{M}_1(S^m)$ with $J_m(\nu) < \infty$, then $\nu = \nu_1^{\otimes m}$. Using (2.3) and Condition 1.5, it follows that $\varphi \circ \pi_\tau \in L_1(\nu_1^{\otimes m}, E)$ for all $\varphi \in \Phi$ and $\tau\colon \{1, \dots, m\} \to \{1, \dots, m\}$. Therefore, Lemma 2.7(b) is applicable and replaces Lemma 2.7(a) in the preceding proof. □

Before we can prove the upper bounds, we need two additional lemmas. For every $\eta \geq 0$ define the set of all η-approximate product measures by

$$\mathcal{A}_{m,\eta} = \bigcap_{A_1,\dots,A_m \in \mathcal{S}} \bigl\{ \nu \in \mathcal{M}_1(S^m) \bigm| |\nu(A_1 \times \dots \times A_m) - \nu_1^{\otimes m}(A_1 \times \dots \times A_m)| \leq \eta \bigr\}.$$

Note that every $\mathcal{A}_{m,\eta}$ is $\tau_1(\mathbb{R})$-closed in $\mathcal{M}_1(S^m)$ and $\nu = \nu_1^{\otimes m}$ for every $\nu \in \mathcal{A}_{m,0}$.

Lemma 2.9. *If Condition* 1.4 *holds, then*

$$\lim_{\eta \downarrow 0} \inf_{\nu \in C \cap \mathcal{A}_{m,\eta}} \frac{1}{m} H_m(\nu \,|\, \mu^{\otimes m}) = J_m(C) \tag{2.10}$$

for every $\tau_1^\Phi(E)$-closed subset C of $\mathcal{M}_1^\Phi(S^m)$.

Proof. For every $\eta \geq 0$ define the greatest lower bound l_η by

$$l_\eta = \inf_{\nu \in C \cap \mathcal{A}_{m,\eta}} \frac{1}{m} H_m(\nu \,|\, \mu^{\otimes m}).$$

If $\eta \leq \eta'$, then $\mathcal{A}_{m,\eta} \subset \mathcal{A}_{m,\eta'}$ and $l_\eta \geq l_{\eta'}$. Therefore, the limit in (2.10) exists and $l \equiv \lim_{\eta \downarrow 0} l_\eta \leq l_0$. Note that l_0 equals the right-hand side of (2.10).

To prove that $l \geq l_0$, it suffices to consider the case $l < \infty$. Choose $l' \in (l, \infty)$. By Lemma 2.1(c) the set $C_{\eta,l'} \equiv C \cap \mathcal{A}_{m,\eta} \cap K(H_m, ml')$ is $\tau_1^\Phi(E)$-compact for every $\eta \geq 0$. Furthermore, the sets $\{C_{\eta,l'}\}_{\eta \geq 0}$ are decreasing as $\eta \downarrow 0$ and $C_{\eta,l'} \neq \varnothing$ for every $\eta > 0$ because $l' > l$. Hence, there exists $\nu \in \bigcap_{\eta > 0} C_{\eta,l'}$. Obviously, $\nu \in \mathcal{A}_{m,0}$ and, therefore, $\nu \in C_{0,l'}$. This means $l' \geq l_0$, hence $l \geq l_0$. □

Let $\mathcal{F}$ denote the family of all finite, nonempty subsets of Φ. For every $F \in \mathcal{F}$ define

$$\Pi_F\colon \mathcal{M}_1^\Phi(S^m) \to E^F \quad \text{by} \quad \Pi_F(\nu) = \left(\int_{S^m} \varphi \, d\nu \right)_{\varphi \in F}. \tag{2.11}$$

For $F' \subset F$ with $F' \neq \varnothing$ let $\Pi_{F,F'}\colon E^F \to E^{F'}$ denote the canonical projection. Note that E^F with $\|y\|_{E^F} \equiv \sum_{\varphi \in F} \|y_\varphi\|_E$ for $y = (y_\varphi)_{\varphi \in F} \in E^F$ is a Banach space. We identify its topological dual $(E^F)^*$ with $(E^*)^F$. For $F \in \mathcal{F}$ and $y \in E^F$ define

$$J_F(y) = \sup_{z \in (E^*)^F} \left(\sum_{\varphi \in F} z_\varphi(y_\varphi) - \log \int_{S^m} \exp\biggl(\sum_{\varphi \in F} z_\varphi(\varphi(s)) \biggr) \, \mu^{\otimes m}(ds) \right). \tag{2.12}$$

Lemma 2.13. *If Condition* 1.4 *holds, then*

$$J_F(y) = \inf\{ H_m(\nu \,|\, \mu^{\otimes m}) \mid \nu \in \mathcal{M}_1^\Phi(S^m), \Pi_F(\nu) = y \} \tag{2.14}$$

for every $F \in \mathcal{F}$ and $y \in E^F$.

Proof. Define ψ: $S^m \to E^F$ by $\psi(s) = (\varphi(s))_{\varphi\in F}$ for all $s \in S^m$, and define the measure $\tilde\mu \in \mathcal{M}_1(E^F)$ by $\tilde\mu = \mu^{\otimes m}\psi^{-1}$. Condition 1.4 implies the moment condition [10, (5.1)] for $\tilde\mu$. Using [10, Theorem 5.2(iv)], it follows that

$$J_F(y) = \inf\Big\{ H_{E^F}(\tilde\nu\,|\,\tilde\mu) \;\Big|\; \tilde\nu \in \mathcal{M}_1(E^F),\ \int_{E^F} \|x\|_{E^F}\,\tilde\nu(dx) < \infty,\ \int_{E^F} x\,\tilde\nu(dx) = y \Big\}.$$

We now show that the right-hand side of this equality equals the one in (2.14).

If $\nu \in \mathcal{M}_1^\Phi(S^m)$ satisfies $\Pi_F(\nu) = y$, then $\tilde\nu = \nu\psi^{-1} \in \mathcal{M}_1(E^F)$ satisfies $\int_{E^F} \|x\|_{E^F}\,\tilde\nu(dx) < \infty$ and $\int_{E^F} x\,\tilde\nu(dx) = y$. Furthermore, if $H_m(\nu\,|\,\mu^{\otimes m}) < \infty$, then it follows by using Jensen's inequality for conditional expectations, that $H_{E^F}(\tilde\nu\,|\,\tilde\mu) \le H_m(\nu\,|\,\mu^{\otimes m})$, see [26, Lemma 4.2.1] for example.

On the other hand, consider a $\tilde\nu \in \mathcal{M}_1(E^F)$ satisfying $\int_{E^F} \|x\|_{E^F}\,\tilde\nu(dx) < \infty$, $\int_{E^F} x\,\tilde\nu(dx) = y$, and $H_{E^F}(\tilde\nu\,|\,\tilde\mu) < \infty$. Then $\tilde\nu \ll \tilde\mu$, hence $\tilde g \equiv d\tilde\nu/d\tilde\mu$ exists. Define $\nu \in \mathcal{M}_1(S^m)$ by $d\nu/d\mu^{\otimes m} = \tilde g \circ \psi$. Then $\nu\psi^{-1} = \tilde\nu$ and $\Pi_F(\nu) = y$. Furthermore, $H_m(\nu\,|\,\mu^{\otimes m}) = H_{E^F}(\tilde\nu\,|\,\tilde\mu) < \infty$, hence $\nu \in \mathcal{M}_1^\Phi(S^m)$ by Lemma 2.1(b). □

Proof of Theorem 1.7(d)*:* Let C denote the $\tau_1^\Phi(E)$-closure of $B \cap \mathcal{M}_1^\Phi(S^m)$. It suffices to consider only the case $J_m(C) > 0$. Choose $l \in (0, J_m(C))$. According to Lemma 2.9 there exists $n_0 \in \mathbb{N}$ with $n_0 \ge m$ such that, for all $n \ge n_0$,

$$\inf\{\, H_m(\nu\,|\,\mu^{\otimes m}) \mid \nu \in C \cap \mathcal{A}_{m,m^2/n} \,\} > lm.$$

Define $C_0 = C \cap \mathcal{A}_{m,m^2/n_0}$, which is $\tau_1^\Phi(E)$-closed. Note that $(L_n^m)_1 = L_n$ and

$$\|L_n^m - L_n^{\otimes m}\|_{\mathrm{var}} \le 1 - \frac{n_{(m)}}{n^m} \le \frac{n^m - (n-m)^m}{n^m} \le \frac{m^2}{n}, \tag{2.15}$$

hence $L_n^m \in \mathcal{A}_{m,m^2/n}$ for all $n \ge m$. Since also $L_n^m \in \mathcal{M}_1^\Phi(S^m)$, it follows that

$$\{L_n^m \in B\} \subset \{L_n^m \in C_0\} \tag{2.16}$$

for all $n \ge n_0$. By Lemma 2.1(b), the set $K(H_m, lm)$ is contained in $\mathcal{M}_1^\Phi(S^m)$. Since C_0 is $\tau_1^\Phi(E)$-closed and $C_0 \cap K(H_m, lm) = \varnothing$, there exist, for every $\nu \in K(H_m, lm)$, an $F_\nu \in \mathcal{F}$ and an open neighbourhood $U_\nu \subset E^{F_\nu}$ of $\Pi_{F_\nu}(\nu)$ such that $C_0 \cap \Pi_{F_\nu}^{-1}(U_\nu) = \varnothing$. Since $K(H_m, lm)$ is $\tau_1^\Phi(E)$-compact by Lemma 2.1(c), there exists a finite subset N of $K(H_m, lm)$ such that $\bigcup_{\nu\in N} \Pi_{F_\nu}^{-1}(U_\nu)$ covers $K(H_m, lm)$. Define $F = \bigcup_{\nu\in N} F_\nu$. Note that $F \in \mathcal{F}$. For every $\nu \in N$ define $U'_\nu = \Pi_{F,F_\nu}^{-1}(U_\nu)$. 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 $K(H_m, lm)$ and is disjoint from C_0. Define $\varepsilon = \mathrm{dist}(\Pi_F(K(H_m, lm)), U^{\mathrm{c}})$. Since $\Pi_F(K(H_m, lm))$ is a compact subset of the open set U, it follows that $\varepsilon > 0$ and that

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

is an open set contained in U. Therefore, for all $n \ge n_0$,

$$\{L_n^m \in B\} \subset \{L_n^m \in C_0\} \subset \{\Pi_F(L_n^m) \in E^F \setminus A_\varepsilon\,\}. \tag{2.17}$$

For $n \in \mathbb{N}$ let G_n denote the group of all permutations of $\{1, \dots, n\}$. Consider $k, n \in \mathbb{N}$ satisfying $km \le n$. For $\sigma \in G_n$ define

$$\Omega \ni \omega \mapsto L_{k,n,\sigma}(\omega) = \frac{1}{k} \sum_{j=0}^{k-1} \delta_{(X_{\sigma(jm+1)}(\omega),\dots,X_{\sigma(jm+m)}(\omega))} \in \mathcal{M}_1(S^m). \tag{2.18}$$

This map is measurable and $\mathbb{P}L_{k,n,\sigma}^{-1}$ does not depend on σ, because $\{X_i\}_{i\in\mathbb{N}}$ are i. i. d. We write $L_{k,n}$ for $L_{k,n,\sigma}$, if σ is the identity on $\{1,\dots,n\}$. Note that

$$\Pi_F(L_{k,n}) = \frac{1}{k}\sum_{j=0}^{k-1}(\varphi(X_{jm+1},\dots,X_{jm+m}))_{\varphi\in F}$$

is a mean of k independent and identically distributed E^F-valued random variables. Using the upper bound of Cramér's theorem for Banach spaces (which is due to Donsker and Varadhan [10, Theorem 5.3]),

$$\limsup_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\Pi_F(L_{k,km})\in E^F\setminus A_\varepsilon)\le -J_F(E^F\setminus A_\varepsilon). \tag{2.19}$$

By Lemma 2.13,

$$\begin{aligned} J_F(E^F\setminus A_\varepsilon) &= \inf\{\, H_m(\nu\,|\,\mu^{\otimes m}) \mid \nu\in\mathcal{M}_1^\Phi(S^m),\, \Pi_F(\nu)\in E^F\setminus A_\varepsilon\,\}\\ &\ge \inf\{\, H_m(\nu\,|\,\mu^{\otimes m}) \mid \nu\in\mathcal{M}_1^\Phi(S^m)\setminus K(H_m,lm)\,\}\ge lm. \end{aligned}\tag{2.20}$$

It remains to transfer the upper bound obtainable from (2.19) and (2.20) into a corresponding result for the U-empirical measures $\{L_n^m\}_{n\ge m}$. Consider $n\in\mathbb{N}$ satisfying $n\ge m$. There are $l_n\equiv\binom{n}{m}$ different ordered m-tuples in the index set $I(m,n)$ appearing in (1.2). If $m=1$, define $k_n=n$, $p_n=1$ and $q_n=0$. If $m\ge 2$, define $k_n=\lfloor n/m\rfloor$, $p_n=l_n-(k_n-1)\lceil l_n/k_n\rceil$ and $q_n=k_n\lceil l_n/k_n\rceil-l_n$. Note that in both cases $l_n=k_np_n+(k_n-1)q_n$ and $q_n\in\{0,1,\dots,k_n-1\}$. If $m\ge 2$ and $n\ge 2m$, then $l_n\ge n(n-1)/2\ge k_n(k_n-1)$, hence p_n is nonnegative because

$$p_n\ge l_n-(k_n-1)\Big(\frac{l_n}{k_n}+1\Big)\ge\frac{l_n}{k_n}-(k_n-1)\ge 0.$$

Therefore, we restrict ourselves to $n\ge 2m$ in the following. Let $I'(m,n)$ denote the set of all ordered m-tuples in $I(m,n)$. According to a result on complete uniform hypergraphs due to Baranyai [4, Theorem 1], there exists a partition of $I'(m,n)$ into p_n sets with k_n elements and q_n sets with k_n-1 elements such that each number from $\{1,\dots,n\}$ is a component of at most one m-tuple in each of the p_n+q_n sets. Hence, there exist two (disjoint) subsets G_n' and G_n'' of G_n with $|G_n'|=m!\,p_n$ and $|G_n''|=m!\,q_n$ such that for every m-tuple $(i_1,\dots,i_m)\in I(m,n)$ there exists exactly one pair $(\sigma,j)\in(G_n'\times\{0,1,\dots,k_n-1\})\cup(G_n''\times\{0,1,\dots,k_n-2\})$ such that $(i_1,\dots,i_m)=(\sigma(jm+1),\dots,\sigma(jm+m))$. Thus, we obtain the representation

$$L_n^m=\frac{k_n}{n_{(m)}}\sum_{\sigma\in G_n'}L_{k_n,n,\sigma}+\frac{k_n-1}{n_{(m)}}\sum_{\sigma\in G_n''}L_{k_n-1,n,\sigma}. \tag{2.21}$$

Since $K(H_m,lm)$ is convex, A_ε is convex, too. Hence, for every $n\ge 2m$,

$$\begin{aligned}\{\Pi_F(L_n^m)\in E^F\setminus A_\varepsilon\}\subset &\bigcup_{\sigma\in G_n'}\{\Pi_F(L_{k_n,n,\sigma})\in E^F\setminus A_\varepsilon\}\\ &\cup\bigcup_{\sigma\in G_n''}\{\Pi_F(L_{k_n-1,n,\sigma})\in E^F\setminus A_\varepsilon\}.\end{aligned}$$

Since the distributions of $L_{k_n,n,\sigma}$ and $L_{k_n-1,n,\sigma}$ do not depend on σ,

$$\begin{aligned}\mathbb{P}(\Pi_F(L_n^m)\in E^F\setminus A_\varepsilon)\le{}& m!\,p_n\,\mathbb{P}(\Pi_F(L_{k_n,k_nm})\in E^F\setminus A_\varepsilon)\\ &+m!\,q_n\,\mathbb{P}(\Pi_F(L_{k_n-1,(k_n-1)m})\in E^F\setminus A_\varepsilon).\end{aligned}$$

Note that $p_n \le n^m$ and $q_n \le n$. Hence, using (2.17), (2.19) and (2.20),

$$\begin{aligned}\limsup_{n\to\infty} &\frac{1}{n}\log \mathbb{P}(L_n^m \in B)\\ &\le \limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\Pi_F(L_n^m)\in E^F\setminus A_\varepsilon)\\ &\le \frac{1}{m}\limsup_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\Pi_F(L_{k,km})\in E^F\setminus A_\varepsilon)\le -l.\end{aligned}\tag{2.22}$$

Since $l \in (0, J_m(C))$ was arbitrary, the upper bound follows. □

The proof of the upper bound for product measures is similar to the previous one, but we need a superexponential estimate to handle the "diagonal terms" in $L_n^{\otimes m}$. Since $L_n^{\otimes m}$ already is a product measure, we do not need Lemma 2.9 here.

Proof of Theorem 1.10(b)*:* Let C denote the $\tau_1^\Phi(E)$-closure of $B\cap\mathcal{M}_1^\Phi(S^m)$. It suffices to consider only the case $J_m(C) > 0$. Choose $l \in (0, J_m(C))$. Define $C_0 = C\cap C_m$, where $C_m \equiv \{\nu\in\mathcal{M}_1(S^m) \mid \nu = \nu_1^{\otimes m}\}$. Due to (2.6) the set C_0 is $\tau_1^\Phi(E)$-closed. Furthermore, $L_n^{\otimes m}\in\mathcal{M}_1^\Phi(S^m)$ and $L_n^{\otimes m}\in C_m$, hence $\{L_n^{\otimes m}\in B\}\subset\{L_n^{\otimes m}\in C_0\}$ for all $n\in\mathbb{N}$. As in the preceding proof we can find $F\in\mathcal{F}$ and an open convex ε-neighbourhood $A_\varepsilon\subset E^F$ of $\Pi_F(K(H_m, lm))$ such that, for all $n\in\mathbb{N}$,

$$\{L_n^{\otimes m}\in B\}\subset\{L_n^{\otimes m}\in C_0\}\subset\{\Pi_F(L_n^{\otimes m})\in E^F\setminus A_\varepsilon\}.\tag{2.23}$$

For every $n\ge m$ we want to define a collection $F_n = \{\varphi_n\}_{\varphi\in F}$ of measurable functions such that

$$\Pi_F(L_n^{\otimes m}) = \Pi_{F_n}(L_n^m).\tag{2.24}$$

Using π_τ from Condition 1.5, one possibility is to define, for all $\varphi\in F$,

$$\varphi_n = \sum_{j=1}^m \frac{n_{(j)}}{n^m}\sum_{\tau\in\mathcal{T}_j}\varphi\circ\pi_\tau,\tag{2.25}$$

where $\mathcal{T}_j$ denotes the set of all surjectiv maps τ: $\{1,\dots,m\}\to\{1,\dots,j\}$ 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 (2.24), 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, then there 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 $(i_1,\dots,i_m) = (k_{\tau(1)},\dots,k_{\tau(m)})$.

For $k, n\in\mathbb{N}$ satisfying $km\le n$ and $\sigma\in G_n$, define $L_{k,n,\sigma}$ and $L_{k,n}$ as in the preceding proof. Define $\alpha = (|F|(m^{m-1}+m^2))^{-1}\min_{\varphi\in F}\alpha_\varphi$ with α_φ as in Condition 1.5. By the exponential Chebychev inequality,

$$\begin{aligned}&\mathbb{P}(\|\Pi_F(L_{k,km})-\Pi_{F_n}(L_{k,km})\|_{E^F}\ge\varepsilon/2)\\ &\qquad\le e^{-\alpha\varepsilon kn/2}\,\mathbb{E}[\exp(\alpha kn\|\Pi_F(L_{k,km})-\Pi_{F_n}(L_{k,km})\|_{E^F})].\end{aligned}$$

Using independence and Hölder's inequality, it follows that

$$\begin{aligned}\mathbb{E}[\exp(\alpha kn\|\Pi_F(L_{k,km})-\Pi_{F_n}(L_{k,km})\|_{E^F})]&\\ \le \left(\int_{S^m}\prod_{\varphi\in F}\exp(\alpha n\|\varphi-\varphi_n\|_E)\,d\mu^{\otimes m}\right)^k&\\ \le \left(\prod_{\varphi\in F}\int_{S^m}\exp(\alpha|F|n\|\varphi-\varphi_n\|_E)\,d\mu^{\otimes m}\right)^{k/|F|}&.\end{aligned}\tag{2.26}$$

Note that $1-n_{(m)}/n^m\le m^2/n$ by (2.15). Using (2.25), it follows that

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

for all $s\in S^m$. Using this estimate, $|\bigcup_{j=1}^{m-1}\mathcal{T}_j|\le m^{m-1}$, Hölder's inequality and Condition 1.5, it follows that the product of the integrals in (2.26) is bounded by a constant which does not depend on k or n. Hence, for any sequence $\{n_k\}_{k\in\mathbb{N}}$ satisfying $km\le n_k$ for all $k\in\mathbb{N}$,

$$\limsup_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\|\Pi_F(L_{k,km})-\Pi_{F_{n_k}}(L_{k,km})\|_{E^F}\ge\varepsilon/2)=-\infty.$$

Therefore, using (2.19) with $A_{\varepsilon/2}$ instead of A_ε,

$$\limsup_{k\to\infty}\frac{1}{k}\log\mathbb{P}(\Pi_{F_{n_k}}(L_{k,km})\in E^F\setminus A_\varepsilon)\le -J_F(E^F\setminus A_{\varepsilon/2})\tag{2.27}$$

for every sequence $\{n_k\}_{k\in\mathbb{N}}$ satisfying $km\le n_k$ for all $k\in\mathbb{N}$. The remaining part of the proof follows along the lines of the preceding proof by using (2.23), (2.24), (2.27) and F_{n_k} instead of F. □

Acknowledgement. We would like to thank Dr. A. Schied for discussions concerning Lemma 2.1(c); see [24, pp. 59–60] and the reference given there.

References

1. A. de Acosta, *On large deviations of empirical measures in the τ-topology*, J. Appl. Probab. **31A** (1994), 41–47, (special volume).
2. R. Azencott, *Grandes Déviations et Applications*, Lectures Notes in Mathematics, vol. 774, Springer-Verlag, Berlin, 1980.
3. R. R. Bahadur and S. L. Zabell, *Large deviations of the sample mean in general vector spaces*, Ann. Probab. **7** (1979), 587–621.
4. Z. Baranyai, *On the factorisation of the complete uniform hypergraph*, Infinite and Finite Sets (A. Hajnal, T. Rado, and V. T. Sós, eds.), North-Holland, Amsterdam, 1975, pp. 91–108.
5. Yu. V. Borovskikh, *U-Statistics in Banach Spaces*, VSP, Utrecht, 1996.
6. S. D. Chatterji, *Martingale convergence and the Radon–Nikodým theorem in Banach spaces*, Math. Scand. **22** (1968), no. 1, 21–41.
7. I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab. **3** (1975), no. 1, 146–158.
8. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd 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. 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.
11. H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic Press, New York and London, 1972.

12. P. Eichelsbacher, *Large deviations for products of empirical probability measures in the τ-topology*, J. Theoret. Probab. **10** (1997), no. 3, 903–929.
13. P. Eichelsbacher and M. Löwe, *Large deviation principle for m-variate von Mises-statistics and U-statistics*, J. Theoret. Probab. **8** (1995), 807–824.
14. P. Eichelsbacher and U. Schmock, *Large deviations for products of empirical measures of dependent sequences*, Markov Process. Related Fields **7** (2001), 435–468.
15. ———, *Rank-dependent moderate deviations of U-empirical measures in strong topologies*, submitted to Probab. Theor. Relat. Fields, 2002.
16. S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley & Sons, New York, 1986.
17. P. Gänssler, *Compactness and sequential compactness in spaces of measures*, Z. Wahrscheinlichkeitstheorie verw. Geb. **17** (1971), 124–146.
18. P. Groeneboom, J. Oosterhoff, and F. H. Ruymgaart, *Large deviation theorems for empirical probability measures*, Ann. Probab. **7** (1979), 553–586.
19. J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.
20. A. J. Lee, *U-Statistics: Theory and Practice*, Marcel Dekker, New York, 1990.
21. M. Löwe, *Exponential Inequalities and Principles of Large Deviations for U-Statistics and von Mises-Statistics*, Ph.D. thesis, Universität Bielefeld, Germany, 1992.
22. J. Neveu, *Discrete-Parameter Martingales*, North-Holland, Amsterdam, 1975.
23. I. N. Sanov, *On the probability of large deviations of random variables*, Selected Transl. Math. Statist. and Prob. **I** (1961), 214–244.
24. A. Schied, *Große Abweichungen für die Pfade der Super-Brownschen Bewegung*, Ph.D. thesis, Universität Bonn, Germany, 1994.
25. ———, *Cramér's condition and Sanov's theorem*, Statist. Probab. Lett. **39** (1998), no. 1, 55–60.
26. U. Schmock, *On the Maximum Entropy Principle for Markov Chains and Processes*, Ph.D. thesis, Technische Universität Berlin, Germany, 1990.
27. L. Wu, *Some general methods of large deviations and applications*, preprint, 1993.

(P. Eichelsbacher) Fakultät für Mathematik, Ruhr-Universität Bochum, Gebäude NA 3/68, D-44780 Bochum, Germany

E-mail address: peich@math.ruhr-uni-bochum.de

(U. Schmock) Departement Mathematik, ETH Zentrum, HG F 42.1, CH-8092 Zürich, Switzerland

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

URL: http://www.math.ethz.ch/~schmock/