2. revised version, February 27, 2003
To appear in Probability Theory and Related Fields

RANK-DEPENDENT MODERATE DEVIATIONS OF U-EMPIRICAL MEASURES IN STRONG TOPOLOGIES*

PETER EICHELSBACHER AND UWE SCHMOCK

Abstract. We prove a rank-dependent moderate deviation principle 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 result can be formulated on a suitable subset of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$. We endow this space with a topology, which is stronger than the usual τ-topology. A moderate deviation principle for Banach-space valued U-statistics is obtained as a particular application. The advantage of our result is that we obtain in the degenerate case moderate deviations in non-Gaussian situations with non-convex rate functions.

1. Introduction

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 $\mathcal{L}(X_i) = \mu$. The *U-empirical measure of order $m \in \mathbb{N}$* is defined by

$$L_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\ldots,i_m)\in I(m,n)} \delta_{(X_{i_1},\ldots,X_{i_m})} \tag{1.1}$$

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. Let $(E, \|\cdot\|_E)$ denote a real separable Banach space with Borel σ-algebra $\mathcal{E}$. Given a measurable function $\varphi\colon S^m \to E$, the *U-statistics* of order m with kernel function φ are defined, for every $n \geq m$, as

$$U_n^m(\varphi) \equiv \int_{S^m} \varphi \, dL_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\ldots,i_m)\in I(m,n)} \varphi(X_{i_1},\ldots,X_{i_m}). \tag{1.2}$$

It is known that for every $\varphi \in L_1(\mu^{\otimes m}, E)$

$$\lim_{n\to\infty} U_n^m(\varphi) = \int_{S^m} \varphi \, d\mu^{\otimes m} \qquad \mathbb{P}\text{-a. s.}, \tag{1.3}$$

which is the *strong law of large numbers* for U-statistics, see for example [11, Theorem 3.1.1]. Note that it suffices to consider symmetric φ for the proof.

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

Key words and phrases. rank-dependent moderate deviations, empirical measures, strong topology, U-statistics.

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

Let $\mathcal{M}(S^m)$ denote the set of signed measures with finite total variation on the product space $(S^m, \mathcal{S}^{\otimes m})$. Given $\varepsilon > 0$, a number $l \in \mathbb{N}$, and bounded measurable functions $\varphi_1, \dots, \varphi_l \colon S^m \to E$, consider

$$O \equiv \Big\{ \nu \in \mathcal{M}(S^m) \;\Big|\; \Big\| \int_{S^m} \varphi_i \, d\nu \Big\|_E < \varepsilon \text{ for all } i = 1, \dots, l \Big\}. \tag{1.4}$$

The subset O of $\mathcal{M}(S^m)$ is an open neighbourhood of the zero measure in the $\tau(E)$-topology, which is defined as the coarsest topology on $\mathcal{M}(S^m)$ such that the map $\mathcal{M}(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu \in E$ is continuous for every bounded measurable function $\varphi \colon S^m \to E$. By (1.2) and the strong law of large numbers given in (1.3),

$$\mathbb{P}\big(L_n^m - \mu^{\otimes m} \notin O\big) \le \sum_{i=1}^{l} \mathbb{P}\bigg(\Big\| U_n^m(\varphi_i) - \int_{S^m} \varphi_i \, d\mu^{\otimes m} \Big\|_E \ge \varepsilon \bigg) \overset{n\to\infty}{\longrightarrow} 0. \tag{1.5}$$

The estimation of the order of magnitude of small probabilities like the ones in (1.5) falls under the theory of *large deviations.* A large deviations principle with respect to strong topologies for $\{L_n^m\}_{n \ge m}$ as random elements of the space $\mathcal{M}_1(S^m)$ of probability measures on $(S^m, \mathcal{S}^{\otimes m})$ is obtained in [17].

As is well known, the weak limits of appropriately scaled U-statistics depend on the rank of the kernel function. Let us briefly recall these results and also introduce some notation for later use. By the Hoeffding decomposition (see for example [11, (1.1.9)]), we know that for every symmetric kernel $\varphi \in L_1(\mu^{\otimes m}, E)$ the U-statistics can be decomposed into a sum of μ-canonical U-statistics of different orders:

$$U_n^m(\varphi) - \int_{S^m} \varphi \, d\mu^{\otimes m} = \sum_{k=1}^{m} \binom{m}{k} U_n^k(\pi_{k,m}\varphi) \qquad \mathbb{P}\text{-a. s.;} \tag{1.6}$$

here $\pi_{k,m} \colon L_1(\mu^{\otimes m}, E) \to L_1(\mu^{\otimes k}, E)$ is defined by

$$\pi_{k,m}\varphi(s_1, \dots, s_k) = \int_{S^m} \varphi \, d\big((\delta_{s_1} - \mu) \otimes \cdots \otimes (\delta_{s_k} - \mu) \otimes \mu^{\otimes m-k}\big) \tag{1.7}$$

for $\mu^{\otimes k}$-almost all $(s_1, \dots, s_k) \in S^k$. The $\pi_{k,m}\varphi$ are called the *Hoeffding projections* of φ. A $\varphi \in L_1(\mu^{\otimes m}, E)$ is called μ-canonical (or completely μ-degenerate) if $\mu^{\otimes m-1}$-almost surely $\int_S \varphi(s_1, \dots, s_m) \, \mu(ds_i) = 0$ for every $i \in \{1, \dots, m\}$. Note that for $\varphi \in L_1(\mu^{\otimes m}, E)$ the kernel $\pi_{k,m}\varphi$ is μ-canonical in k variables for every $k \in \{1, \dots, m\}$; and $\pi_{k,m}\varphi$ is symmetric if φ is symmetric. By definition we say that a symmetric $\varphi \in L_1(\mu^{\otimes m}, E)$ is of rank $r \in \{1, \dots, m\}$ (or μ-degenerate of order $r-1$) if $\pi_{k,m}\varphi = 0$ for all $k \in \{1, \dots, r-1\}$ and $\pi_{r,m}\varphi \neq 0$. In particular, the Hoeffding decomposition (1.6) then starts with the term $k = r$. If $\pi_{1,m}\varphi \neq 0$, then φ is called non-degenerate. One advantage of the decomposition (1.6) is the fact that the terms $U_n^k(\pi_{k,m}\varphi)$ with $k \in \{1, \dots, m\}$ are uncorrelated with variances of decreasing order in n. It is a generalisation of a projection technique used in many parts of nonparametric statistics. Often a statistic of interest is regarded as an element of a vector space and projected orthogonally onto some subspaces. In the analysis of variance, means are projected onto subspaces to obtain main effects. In the theory of rank statistics, the projection is onto the subspace of linear rank statistics, cf. [27]. If $L_2^{(k)}$ denotes the Hilbert space of all U-statistics of order k with square integrable kernel functions, then the Hoeffding decomposition is the representation of $U_n^m(\varphi) - \int_{S^m} \varphi \, d\mu^{\otimes m}$ (a member of $L_2^{(m)}$) as the sum of its

component projections onto the spaces $L_2^{(k)} \cap (L_2^{(k-1)})^\perp$ for $k = 1, \dots, m$, where $L_2^{(0)}$ denotes the constants. Note that (1.6) is true only for symmetric kernel functions φ unless a symmetrization is built into the definition of the Hoeffding projection (1.7), see (2.4) and (2.6).

The central limit theorem with $\sqrt{n}$-scaling for real-valued, not necessarily degenerate U-statistics has a simple proof and was obtained by Hoeffding [22]; the central limit theorem for degenerate U-statistics was obtained much later. The limit laws in the degenerate case are the laws of Gaussian chaos variables: if $\varphi \in L_1(\mu^{\otimes m}, \mathbb{R})$ is μ-canonical, then the sequence $\{n^{m/2} U_n^m(\varphi)\}_{n \geq m}$ converges in distribution if and only if $\int_{S^m} \varphi^2 \, d\mu^{\otimes m} < \infty$, see [30], [21], and also [20, Section 4.2], in particular the historical remarks in [20, Section 4.4].

Although it is not necessary for our results, let us construct a Gaussian chaos variable $K_\mu(\varphi)$ with the limiting distribution to give an impression how much it differs from a generic Gaussian random variable. This will also indicate that the corresponding MDP is more involved. Furthermore, the representation (1.8) motivates the approximation formulated in Lemma 3.41, which is an essential tool in the proof of our main theorem. To construct the Gaussian chaos variable $K_\mu(\varphi)$, we may assume φ to be symmetric. Let $L_2^{\mathrm{c}}(\mu^{\otimes m}, \mathbb{R})$ denote the Hilbert space of symmetric, $\mu^{\otimes m}$-square integrable, μ-canonical functions from S^m to $\mathbb{R}$ and let $\{\phi_i\}_{i \in J}$ be an orthonormal basis for $L_2^{\mathrm{c}}(\mu, \mathbb{R})$, where J is an at most countable index set. One can show that the set

$$\left\{ \binom{m}{r_j \colon j \in J}^{-1/2} \sum_{\substack{(i_1,\dots,i_m) \in J^m \\ j(i_1,\dots,i_m) = r_j}} \phi_{i_1}(s_1) \dots \phi_{i_m}(s_m) \;\middle|\; \sum_{j \in J} r_j = m,\, r_j \in \mathbb{N}_0 \right\}$$

is an orthonormal basis for $L_2^{\mathrm{c}}(\mu^{\otimes m}, \mathbb{R})$, where, for all $j \in J$ and $(i_1, \dots, i_m) \in J^m$, $j(i_1, \dots, i_m) = \sum_{k=1}^m \delta_j(i_k)$ denotes the number of occurrences of j among the coordinates of the multi-index $(i_1, \dots, i_m)$, and where $\binom{m}{r_j \colon j \in J}$ is the multinomial coefficient $m! / \prod_{j \in J} r_j!$. Let G_μ be an isonormal Gaussian process on $L_2^{\mathrm{c}}(\mu, \mathbb{R})$, that is, a centered Gaussian process with parameter space $L_2^{\mathrm{c}}(\mu, \mathbb{R})$ such that $\tilde{\mathbb{E}}\big[G_\mu(f) G_\mu(g)\big] = \int_S fg \, d\mu$ (such processes exist by Kolmogorov's consistency theorem). Since our kernel φ is in $L_2^{\mathrm{c}}(\mu^{\otimes m}, \mathbb{R})$, we obtain the L_2-expansion

$$\varphi(s_1, \dots, s_m) = \sum_{(i_1,\dots,i_m) \in J^m} a_{i_1,\dots,i_m} \, \phi_{i_1}(s_1) \dots \phi_{i_m}(s_m) \tag{1.8}$$

with the coefficients $a_{i_1,\dots,i_m} \equiv \int_{S^m} \varphi(s_1, \dots, s_m) \prod_{k=1}^m \phi_{i_k}(s_k) \, \mu^{\otimes m}(ds_1, \dots, ds_m)$, which are symmetric in their entries. For $k \in \mathbb{N}_0$ let H_k denote the Hermite polynomial of degree k with leading coefficient 1. The Gaussian chaos random variable corresponding to φ is defined in an L_2-sense by

$$K_\mu(\varphi) \equiv \sum_{(i_1,\dots,i_m) \in J^m} a_{i_1,\dots,i_m} \prod_{j \in J} H_{j(i_1,\dots,i_m)}\big(G_\mu(\phi_j)\big);$$

note that the product is actually a finite one.

Now assume that a sequence $\{b_n\}_{n \in \mathbb{N}} \subset (0, \infty)$ satisfies

$$\lim_{n \to \infty} \frac{b_n}{n} = 0 \qquad \text{and} \qquad \lim_{n \to \infty} \frac{n}{b_n^2} = 0. \tag{1.9}$$

If the functions $\varphi_1, \dots, \varphi_l$ defining O in (1.4) are symmetric, real-valued, $\mu^{\otimes m}$-square integrable and of rank $r \in \{1, \dots, m\}$, then the above Hoeffding decomposition (1.6) and the central limit theorem for completely μ-degenerate U-statistics imply for the blown-up difference of L_n^m and $\mu^{\otimes m}$ that

$$\mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r (L_n^m - \mu^{\otimes m}) \notin O\Big) \\ \leq \sum_{i=1}^{l} \sum_{k=r}^{m} \mathbb{P}\bigg(\binom{m}{k} \frac{n^{r-k/2}}{b_n^r} \big|n^{k/2} U_n^k(\pi_{k,m}\varphi_i)\big| \geq \frac{\varepsilon}{m}\bigg) \overset{n\to\infty}{\longrightarrow} 0, \quad (1.10)$$

because $n^{r-k/2}/b_n^r \leq (n/b_n^2)^{r/2} \to 0$ by the second condition in (1.9). The order of magnitude of the probability on the left-hand side of (1.10) is larger than the one in (1.5). The estimation of such probabilities is usually called a *moderate deviation problem.* Since the central limit theorem is rank dependent, we expect a related phenomenon for the large deviations of moderate U-empirical measures on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ with $\{b_n\}_{n\in\mathbb{N}}$ as in (1.9). If φ is μ-canonical on S^m, we expect a moderate deviation behaviour for $(n/b_n)^m\, U_n^m(\varphi)$ as $n \to \infty$.

This is the problem addressed in the present paper. The aim is to obtain moderate deviations in non-Gaussian situations with non-convex rate functions. The results in the literature we are familiar with deal with situations where a Gaussian behaviour is given.

We want to analyse moderate deviations for U-empirical measures. This leads to the introduction of *rank-dependent* moderate U-empirical measures in (2.1) and (2.15), which, to the best of our knowledge, are new. We prove a rank-dependent moderate deviation principle (MDP) with respect to a strong topology on a restricted space $\mathcal{M}^\Phi(S^m)$ of signed 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. This result enables us to derive rank-dependent moderate deviations for Banach-space valued U-statistics, see Subsection 2.2. As further applications in [16], a rank-dependent MDP for U-processes indexed by a uniformly bounded Vapnik–Červonenkis subgraph class of functions is obtained, and in [18] moderate deviations are derived around non-degenerate attractors of the empirical measure under mean-field Gibbs measures.

Moderate deviations for empirical measures of Markov chains and other dependent random sequences have been previously studied in [4] and [31], for example.

In Section 2 we formulate the results and applications. In Section 3 we give the proof of the main theorem. We adapt Cramér's exponential change of measure method to the moderate scale and apply the Dawson–Gärtner projective limit approach; both methods are by now classical in large deviations theory, see e. g. [14, Sections 2.2 and 4.6]. Moreover, we apply techniques like randomization, decoupling inequalities, integrability of Rademacher chaos and exponential inequalities of Bernstein-type from the theory of U-statistics. For a detailed presentation of these methods as well as historical remarks, see e. g. [20]; specific references are given in the course of our proofs. For our specific purposes, namely kernel functions with values in a Banach space E of type p, we adapt the usual exponential integrability

of a Banach space valued Rademacher chaos, and extend the Bernstein-type inequality for bounded real-valued canonical U-statistics from [8] to U-statistics with E-valued kernels of unbounded norm satisfying an exponential moment condition.

2. Statement of results and applications

2.1. **Rank-dependent moderate deviation principles.** As before, let $(S,\mathcal{S})$ be a measurable space and let $\{X_i\}_{i\in\mathbb{N}}$ be the coordinate projections defined on the product space $(\Omega,\mathcal{A},\mathbb{P})$ and $m\in\mathbb{N}$. Let $(E,\|\cdot\|_E)$ be a separable real Banach space with Borel σ-algebra $\mathcal{E}$. We always exclude the case $E=\{0\}$. Let Φ be a collection of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable and Bochner $\mu^{\otimes m}$-integrable functions $\varphi\colon S^m\to E$ containing the set $B(S^m,E)$ of all bounded measurable ones. We define the Φ-restricted set of signed measures by

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

where $|\nu|$ denotes the total variation measure corresponding to ν. Let $\tau^{\Phi}(E)$ be the coarsest topology on $\mathcal{M}^{\Phi}(S^m)$ such that $\mathcal{M}^{\Phi}(S^m)\ni\nu\mapsto\int_{S^m}\varphi\,d\nu$ is continuous for every $\varphi\in\Phi$. If $\Phi=B(S^m,E)$, then we write $\tau(E)$ instead of $\tau^{\Phi}(E)$. The σ-algebra on $\mathcal{M}(S^m)$ is defined to be the smallest one containing $\mathcal{M}^{\Phi}(S^m)$ such that all the maps $\mathcal{M}(S^m)\ni\nu\mapsto\int_{S^m}\varphi\,d\nu$ with $\varphi\in B(S^m,E)$ are measurable.

We define the moderate U-empirical measure $M_n^{m,1}\colon\Omega\to\mathcal{M}(S^m)$ of rank 1 by

$$M_n^{m,1}=\frac{n}{b_n}(L_n^m-\mu^{\otimes m}) \tag{2.1}$$

for every $n\geq m$. Since $\Phi\subset L_1(\mu^{\otimes m},E)$ by assumption, the measure-valued maps $\{M_n^{m,1}\}_{n\geq m}$ actually take values in $\mathcal{M}^{\Phi}(S^m)$ and are therefore measurable. Notice that $M_n^{m,1}$ is (on the first view) the straightforward generalization of $n(L_n^1-\mu)/b_n$, which is usually considered in the literature on moderate deviations.

The large deviations of $\{M_n^{1,1}\}_{n\in\mathbb{N}}$ on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ have been studied by Borovkov and Mogulskii [10, Section 3] in the $\tau(\mathbb{R})$-topology on the space $\mathcal{M}(S)$, when S is a Hausdorff topological space with Borel σ-algebra $\mathcal{S}$. They considered special subsets of $\mathcal{M}(S)$ for the lower and the upper bound. De Acosta [3, Section 3] generalized this result to a full large deviation principle on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ when $(S,\mathcal{S})$ is a general measurable space. We prove large deviations on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ of the moderate U-empirical measures $\{M_n^{m,1}\}_{n\geq m}$ in the stronger $\tau^{\Phi}(E)$-topologies, generated by a collection Φ of functions $\varphi\colon S^m\to E$ satisfying appropriate moments conditions. Our main extension is the passage from $m=1$ to an arbitrary order $m\in\mathbb{N}$, because for $m\geq 2$ a moderate U-empirical measures L_n^m can not be rewritten as a sum of i. i. d. terms.

We now start to introduce our afore-mentioned *rank-dependent* moderate U-empirical measures. Our goal is to introduce an empirical measure to mimic the Hoeffding decomposition on the level of measures. For rank $r\in\{1,\dots,m\}$, we want to find $M_n^{m,r}\colon\Omega\to\mathcal{M}(S^m)$ such that for every $\varphi\in L_1(\mu^{\otimes m},E)$,

$$\int_{S^m}\varphi\,dM_n^{m,r}=\Big(\frac{n}{b_n}\Big)^r\sum_{a=r}^{m}\int_{S^a}\tilde{\varphi}_a\,dL_n^a\qquad\mathbb{P}\text{-a. s.}, \tag{2.2}$$

where $\tilde{\varphi}_r,\dots,\tilde{\varphi}_m$ are suitably defined (mimicking the Hoeffding projections). This means that $M_n^{m,r}$ extracts from φ the components of higher rank. Unfortunately,

we have to introduce a considerable amount of notation before we can state our corresponding results.

For the moderate deviations upper bound in Theorem 2.20(d) below, we have to formulate moment conditions for the unbounded $\varphi \in \Phi$. This in turn requires a decomposition, which for symmetric φ is closely related to the Hoeffding decomposition (1.6) for the corresponding U-statistic. Although it might seem tempting, we do not want to consider collections Φ consisting only of symmetric functions, because for $\mathcal{S} \neq \{\varnothing, S\}$ and $m \geq 2$, we would not be able to separate the zero measure from the measure $\delta_x^{\otimes m-1} \otimes \delta_y - \delta_y \otimes \delta_x^{\otimes m-1}$ with $x \in A \in \mathcal{S}$ and $y \in S \setminus A$ for example; hence the topology $\tau^{\Phi}(E)$ would loose the Hausdorff property. Due to this technical problem we need to consider not necessarily symmetric functions. In addition, building products of measures is not a symmetric operation. Hence we have to take into our consideration all combinations of products in every subspace of the projection method mimicking the Hoeffding decomposition.

Given $\varphi \in L_1(\mu^{\otimes m}, E)$ and a nonempty subset A of $\{1, \ldots, m\}$, define $\varphi_A \in L_1(\mu^{\otimes |A|}, E)$ by μ-integrating $\varphi(s_1, \ldots, s_m)$ with respect to every s_i with $i \in \{1, \ldots, m\} \setminus A$. By convention, $\varphi_{\varnothing} \equiv \int_{S^m} \varphi \, d\mu^{\otimes m} \in E$. Furthermore, define $\tilde{\varphi}_A \in L_1(\mu^{\otimes |A|}, E)$ by

$$\tilde{\varphi}_A\big((s_i)_{i\in A}\big) = \sum_{B \subset A} (-1)^{|A\setminus B|} \varphi_B\big((s_i)_{i\in B}\big), \tag{2.3}$$

for every nonempty $A \subset \{1, \ldots, m\}$, and let $\tilde{\varphi}_{\varnothing} \equiv \varphi_{\varnothing}$. Note that

$$\tilde{\varphi}_A((s_i)_{i\in A}) = \int_{S^m} \varphi \, d\bigg(\bigotimes_{i=1}^{m} \big(1_A(i)\,(\delta_{s_i} - \mu) + 1_{A^c}(i)\,\mu\big)\bigg). \tag{2.4}$$

According to the inclusion–exclusion principle or the Möbius inversion formula,

$$\varphi(s_1, \ldots, s_m) = \sum_{A \subset \{1,\ldots,m\}} \tilde{\varphi}_A\big((s_i)_{i\in A}\big)$$

for $\mu^{\otimes m}$-almost all $(s_1, \ldots, s_m) \in S^m$. Hence, for every $n \geq m$,

$$\int_{S^m} \varphi \, dL_n^m = \tilde{\varphi}_0 + \sum_{a=1}^{m} \int_{S^a} \tilde{\varphi}_a \, dL_n^a \tag{2.5}$$

$\mathbb{P}$-almost surely, where, for every $a \in \{0, 1, \ldots, m\}$,

$$\tilde{\varphi}_a \equiv \sum_{\substack{A \subset \{1,\ldots,m\} \\ |A| = a}} \tilde{\varphi}_A. \tag{2.6}$$

Due to (2.4), every $\tilde{\varphi}_A$ with nonempty $A \subset \{1, \ldots, m\}$ is completely μ-degenerate in the sense that $\mu^{\otimes |A|-1}$-almost surely $\int_S \tilde{\varphi}_A((s_i)_{i\in A})\,\mu(ds_j) = 0$ for every $j \in A$. Note that (2.5) extends to the Hoeffding decomposition (1.6) to not necessarily symmetric functions. Similarly, (2.4) and (2.6) extend the Hoeffding projection (1.7).

Note that for i. i. d. random variables taking values in a Banach space E, the central limit theorem does not hold in general unless the space E satisfies certain geometric properties, cf. [5, Chap. 3, Sec. 7, Ex. 5] and [26]. It turns out that if the space is of type 2 (meaning that it satisfies the one-sided parallelogram inequality (2.7) with $p = 2$), then the central limit theorem behaves as in the Hilbert space setting, where i. i. d. random variables $\{Y_n\}_{n\in\mathbb{N}}$ satisfy a central limit theorem if

$\mathbb{E}[\|Y_1\|_E^2] < \infty$. A Banach space $(E, \|\cdot\|_E)$ is said to be of type $p \in [1,2]$ if there is a type constant $C_{E,p} > 0$ such that for all finite sequences $x_1, \dots, x_n$ in E

$$\left\| \sum_{i=1}^n \varepsilon_i \, x_i \right\|_E^p \leq C_{E,p} \sum_{i=1}^n \|x_i\|_E^p, \tag{2.7}$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent Rademacher variables.

We can now state the moment conditions on $\varphi \in \Phi$ in terms of the corresponding completely μ-degenerate functions $\tilde{\varphi}_1, \dots, \tilde{\varphi}_m$ given via (2.3) and (2.6). For a Banach space of type 2, these take a simple form.

Condition 2.8. *Let E be a separable real Banach space of type 2. We say that Φ satisfies this condition for rank $r \in \{1, \dots, m\}$, if $\Phi \subset L_2(\mu^{\otimes m}, E)$ and if for every $a \in \{r, \dots, m\}$ with $a \geq 2$ and for every $\varphi \in \Phi$ there exists at least one $\alpha_{a,\varphi} > 0$ such that*

$$\int_{S^a} \exp\bigl(\alpha_{a,\varphi} \|\tilde{\varphi}_a\|_E^2\bigr) \, d\mu^{\otimes a} < \infty, \tag{2.9}$$

and if in the case $r = 1$ the strong Cramér condition holds for $\tilde{\varphi}_1$, i. e.,

$$\int_S \exp(\alpha \|\tilde{\varphi}_1\|_E) \, d\mu < \infty \quad \text{for every } \alpha > 0. \tag{2.10}$$

Next we strive to balance the moment conditions with the type p of the Banach space E and the growth rate of $\{b_n\}_{n\in\mathbb{N}}$ satisfying (1.9). The general version of Condition 2.8 reads as follows.

Condition 2.11. *We say that Φ satisfies this condition for rank $r \in \{1, \dots, m\}$, if the separable real Banach space E is of type $p \in [1,2]$, if $\Phi \subset L_2(\mu^{\otimes m}, E)$, if for every $a \in \{r, \dots, m\}$ with $a \geq 2$ there exists $p_a \in (1, p]$ such that*

(a) *$p_a = 2$ if $a = r$, and $\lim_{n\to\infty} n^{3/2-1/p_a-r/a} / b_n^{1-r/a} = \infty$ if $a \geq r+1$,*
(b) *for every $\varphi \in \Phi$ there exists at least one $\alpha_{a,\varphi} > 0$ such that*

$$\int_{S^a} \exp\bigl(\alpha_{a,\varphi} \|\tilde{\varphi}_a\|_E^{p_a}\bigr) \, d\mu^{\otimes a} < \infty, \tag{2.12}$$

and if in the case $r = 1$ the following two requirements are satisfied for every $\varphi \in \Phi$:

(c) *$\tilde{\varphi}_1$ satisfies the strong Cramér condition (2.10),*
(d) *$\{\mathcal{L}(n^{-1/2} \sum_{i=1}^n \tilde{\varphi}_1(X_i))\}_{n\in\mathbb{N}}$ converges weakly to a Gaussian measure on E.*

Remark 2.13. Let is discuss Condition 2.11 in detail.

(a) The conditions (a) and (b) ensure that the terms with $a \geq r+1$ in (2.5) have no influence on the scale of the moderate deviations of rank r. The limit relation of condition (a) is used in the paragraph ending with (3.47). The exponential moment (2.12) is needed for the application of the Bernstein-type inequality given in Theorem 3.26.
(b) Conditions (c) and (d) are needed to use moderate deviation upper bounds for sums of i. i. d. Banach-space valued random variables from the literature.
(c) For $r \geq 2$, the Banach space E must be of type 2; see condition (a). Otherwise our Bernstein-type inequality in Theorem 3.26 does not seem strong enough to classify the approximation error in (3.50) as negligible on the scale of moderate deviations of rank r, even for bounded φ.

(d) If $\Phi = B(S^m, E)$, then all corresponding completely μ-degenerate functions are bounded and Conditions 2.11(b) and (c) certainly hold.
(e) If the Banach space E is of type 2, then condition (d) holds, see [5, Theorem 7.5]. In the case $p_a = 2$, the limit in condition (a) reduces to the first condition in (1.9). Therefore, Condition 2.11 simplifies to Condition 2.8.
(f) For $a \geq r+1$, choosing $p_a < 2$ allows us to balance a weaker moment condition (2.12) with a slower growth of the sequence $\{b_n\}_{n\in\mathbb{N}}$. However, due to the second condition in (1.9), a necessary requirement for the limit relation in condition (a) is $p_a > 2a/(2a-r)$.
(g) Note that for the case of rank $r = 1$ we do *not* need E to be of type 2. Furthermore, if $r = m$, condition (a) puts *no* additional restriction on the sequence $\{b_n\}_{n\in\mathbb{N}}$.

Let us define the moderate U-empirical measures of general rank $r \in \{1, \dots, m\}$. Our aim is to generalise (2.1) and to obtain the characterisation (2.2) above. For completeness, we give an explicit construction.

Given $a \in \{1, \dots, m\}$ and $A = \{i_1, \dots, i_a\}$ with $1 \leq i_1 < i_2 < \cdots < i_a \leq m$, let $1 \leq i_{a+1} < \cdots < i_m \leq m$ denote the indices in $\{1, \dots, m\} \setminus A$. Define the permutation τ_A of $\{1, \dots, m\}$ such that $\tau_A(i_j) = j$ for every $j \in \{1, \dots, m\}$. For every map $\tau\colon \{1, \dots, m\} \to \{1, \dots, m\}$ we define $\pi_\tau\colon S^m \to S^m$ by $\pi_\tau(s) = (s_{\tau(1)}, \dots, s_{\tau(m)})$ for every $s = (s_1, \dots, s_m) \in S^m$. Using τ_A, define the mapping $\gamma_{A,m}\colon \mathcal{M}(S^a) \to \mathcal{M}(S^m)$ by $\gamma_{A,m}(\nu) = (\nu \otimes \mu^{\otimes m-a})\pi_{\tau_A}^{-1}$ for all $\nu \in \mathcal{M}(S^a)$. The marginal measure of $\gamma_{A,m}(\nu)$, corresponding to the ordered indices in A, is then given by ν, all other one-component marginals equal μ. Related to (2.3), define $\tilde{\gamma}_{A,m}\colon \mathcal{M}(S^a) \to \mathcal{M}(S^m)$ by

$$\tilde{\gamma}_{A,m}(\nu) = (-1)^{|A|}\mu^{\otimes m} + \sum_{B\subset A,\, B\neq\varnothing} (-1)^{|A\setminus B|}\gamma_{B,m}(\nu_{A,B}), \tag{2.14}$$

where $\nu_{A,B}$ denotes the marginal $\nu_{j_1,\dots,j_b}$ of $\nu \in \mathcal{M}_1(S^a)$, when $B = \{i_{j_1}, \dots, i_{j_b}\}$ with $1 \leq j_1 < \cdots < j_b \leq a$. For $n \geq m$ define the moderate U-empirical measure $M_n^{m,r}$ of rank $r \in \{1, \dots, m\}$ by

$$M_n^{m,r} = \Big(\frac{n}{b_n}\Big)^r \Big(L_n^m - \mu^{\otimes m} - \sum_{\substack{A\subset\{1,\dots,m\}\\ 1\leq|A|\leq r-1}} \tilde{\gamma}_{A,m}(L_n^{|A|})\Big). \tag{2.15}$$

If $r = 1$, then (2.15) reduces to (2.1). Using (2.3)–(2.6), it follows from these definitions that (2.2) holds for every $\varphi \in L_1(\mu^{\otimes m}, E)$, which means that $M_n^{m,r}$ extracts from φ the components of higher rank.

To avoid measurability problems, we always make the following additional assumption when considering moderate U-empirical measures of rank $r \geq 2$:

Assumption 2.16. *For every unbounded $\varphi \in \Phi$, every nonvoid $A \subsetneq \{1, \dots, m\}$ and every $(x_i)_{i\in A} \in S^A$ the function $S^{\{1,\dots,m\}\setminus A} \ni (x_i)_{i\in\{1,\dots,m\}\setminus A} \mapsto \varphi(x_1, \dots, x_m)$ is Bochner $\mu^{\otimes m-|A|}$-integrable.*

By Fubini's theorem for Bochner integrals, this property of Φ can always be achieved by redefining every unbounded $\varphi \in \Phi$ on a $\mathcal{S}^{\otimes m}$-measurable set of $\mu^{\otimes m}$-measure zero if necessary. Under the above assumption, the moderate U-empirical measures $\{M_n^{m,r}\}_{n\geq m}$ of rank $r \in \{2, \dots, m\}$ also take values in $\mathcal{M}^\Phi(S^m)$.

We define the rate function $I_{m,r}\colon \mathcal{M}(S^m) \to [0,\infty]$ for the moderate deviations of rank $r \in \{1,\dots,m\}$ by

$$I_{m,r}(\nu) = \frac{1}{2}\int_S \Big(\frac{d\tilde\nu}{d\mu}\Big)^2 d\mu \tag{2.17}$$

if there exists a $\tilde\nu \in \mathcal{M}(S)$ satisfying $\tilde\nu(S) = 0$ and $\tilde\nu \ll \mu$ such that

$$\nu = \sum_{\substack{A\subset\{1,\dots,m\}\\|A|=r}} \gamma_{A,m}(\tilde\nu^{\otimes r}) = \sum_{\substack{A\subset\{1,\dots,m\}\\|A|=r}} \bigotimes_{i\in\{1,\dots,m\}} \big(1_A(i)\,\tilde\nu + 1_{A^c}(i)\,\mu\big); \tag{2.18}$$

and we define $I_{m,r}(\nu) = \infty$ otherwise. If $\nu \in \mathcal{M}(S^m)$ and $\tilde\nu \in \mathcal{M}(S)$ with $\tilde\nu(S) = 0$ satisfy (2.18), then $I_{m,r}(\nu) = I_{1,1}(\tilde\nu)$, hence $I_{m,r}$ is a generalization of the well-known moderate deviations rate function $I_{1,1}$, cf. [3, Section 3]. Furthermore, $I_{m,r}$ can be finite only for symmetric measures $\nu \in \mathcal{M}(S^m)$ in the sense that (2.18) implies $\nu = \nu\pi_\tau^{-1}$ for every permutation τ of $\{1,\dots,m\}$ with π_τ defined just before (2.14). In the case $r = 1$, equation (2.18) reduces to $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \tilde\nu \otimes \mu^{\otimes m-i}$ and every one-component marginal of ν is equal to $\tilde\nu$. This is the rate function we would expect in the non-degenerate case $r = 1$. Indeed, considering $M_n^{m,1}$ given in (2.1), the first term in its expansion (2.2) is a mean of i. i. d. random variables. In case the terms with $a \in \{2,\dots,m\}$ do not contribute to the moderate deviations behaviour, we expect a rate function similar to the rate function of the sequence $\{n(L_n^1 - \mu)/b_n\}_{n\in\mathbb{N}}$, which is the Fisher information $I_{1,1}$, see [3]. The difference is that in the case $m \geq 2$, the rate function must be defined on $\mathcal{M}(S^m)$. In the case $r \in \{2,\dots,m\}$ we expect the rate function to be finite only for symmetric measures $\nu \in \mathcal{M}(S^m)$ with r entries given by $\tilde\nu$ and the other ones given by the reference measure μ.

In the general case, $(\tilde\nu(B))^r = \nu(B^r \times S^{m-r})$ for every $B \in \mathcal{S}$, hence $\tilde\nu$ is uniquely determined for odd r and up to a sign for even r. Therefore, $I_{m,r}$ is well defined. Using the convexity of $\mathbb{R} \ni x \mapsto x^2$, it follows that $I_{m,1}$ is convex. For $r \geq 2$, the rate function $I_{m,r}$ is, in general, not convex, see Remark 2.26.

Let us give some heuristics for the rank-dependent moderate deviations principle formulated in Theorem 2.20 below, and for the condition (2.18) in the definition of the rate function (2.17) in particular. Basically, the U-empirical measure L_n^m is close to the m-fold product measure $(L_n^1)^{\otimes m}$, although the use of this observation in our topology for (3.51) requires strong additional assumptions. As the proof of the lower bound given in Theorem 2.20(c) shows (see the definition of F_n and $\mathbb{P}_n$ for (3.7)), the main contribution for the moderate deviations arises when L_n^1 is close to $\mu + (b_n/n)\tilde\nu$, hence when L_n^m is close to the m-fold product measure

$$\Big(\mu + \frac{b_n}{n}\tilde\nu\Big)^{\otimes m} = \sum_{r=0}^m \Big(\frac{b_n}{n}\Big)^r \sum_{\substack{A\subset\{1,\dots,m\}\\|A|=r}} \bigotimes_{i\in\{1,\dots,m\}} \big(1_A(i)\,\tilde\nu + 1_{A^c}(i)\,\mu\big), \tag{2.19}$$

where the right-hand side of (2.19) is just the expansion of the left-hand side, ordered according to the powers of b_n/n. This motivates the scaling $(n/b_n)^r$ in (2.2) to extract the relevant term. It also motivates the condition (2.18) for the rate function $I_{m,r}$ for the ranks $r \in \{1,\dots,m\}$, because the right-hand side of (2.18) appears in the expansion (2.19).

For every $l \in [0,\infty)$ let $K(I_{m,r},l) = \{\,\nu \in \mathcal{M}(S^m) \mid I_{m,r}(\nu) \le l\,\}$ denote the corresponding level set. For $B \subset \mathcal{M}(S^m)$ define the abbreviation $I_{m,r}(B) = \inf_{\nu\in B} I_{m,r}(\nu)$.

Theorem 2.20 (Rank-dependent moderate deviations of U-empirical measures)**.** *The following assertions hold for every $r \in \{1,\dots,m\}$:*

(a) *If $\Phi \subset L_2(\mu^{\otimes m},E)$, then $K(I_{m,r},l) \subset \mathcal{M}^\Phi(S^m)$ for every $l \in [0,\infty)$.*
(b) *If $\Phi \subset L_2(\mu^{\otimes m},E)$, then the level set $K(I_{m,r},l)$ is $\tau^\Phi(E)$-compact and sequentially $\tau^\Phi(E)$-compact for every $l \in [0,\infty)$.*
(c) *Assume that there exists a $p \in (1,2]$ such that the Banach space $(E,\|\cdot\|_E)$ is of type p and $\lim_{n\to\infty} n/b_n^p = 0$. If $\Phi \subset L_p(\mu^{\otimes m},E)$ in the case $m = r$, or if $\Phi \subset L_2(\mu^{\otimes m},E)$ in the case $r < m$, then*

$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(M_n^{m,r} \in B) \ge -I_{m,r}(\mathrm{int}_{\tau^\Phi(E)}(B)) \tag{2.21}$$

for every measurable $B \subset \mathcal{M}(S^m)$, where $\mathrm{int}_{\tau^\Phi(E)}(B)$ denotes the interior of the set $B \cap \mathcal{M}^\Phi(S^m)$ with respect to the $\tau^\Phi(E)$-topology.
(d) *If Condition 2.11 holds for r, then*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(M_n^{m,r} \in B) \le -I_{m,r}(\mathrm{cl}_{\tau^\Phi(E)}(B)) \tag{2.22}$$

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

The finer the topology, the larger the $\tau^\Phi(E)$-interior of $B \cap \mathcal{M}^\Phi(S^m)$ in (2.21), which gives a better lower bound. Similarly, a stronger topology gives a smaller $\tau^\Phi(E)$-closure of $B \cap \mathcal{M}^\Phi(S^m)$ in (2.22), which implies a sharper upper bound. This is an important motivation to choose the topology as strong as possible.

Example 2.27 shows that Theorem 2.20(c) does not hold without an assumption on the type of the separable Banach space E, even in the case $\Phi = B(S^m,E)$ with $m = r = 1$, $S = E$, and a set B determined by a bounded continuous function.

Theorem 2.20, specialized to $m = 1$ and $E = \mathbb{R}$, already extends the MDP of de Acosta for $\{M_n^{1,1}\}_{n\in\mathbb{N}}$ to the $\tau^\Phi(\mathbb{R})$-topology. Note that the lower bound in Theorem 2.20(c) does not make use of the condition $\Phi \subset L_2(\mu^{\otimes m},E)$ if E is of type $p \in (1,2)$ and $m = r$. But then the lower bound might loose its strength because measures $\nu \in B$ with $I_{m,r}(\nu) < \infty$ might not be contained in $\mathcal{M}^\Phi(S^m)$.

The central limit theorem for real-valued U-statistics of different degrees of degeneracy can be reduced to the completely μ-degenerate case, because each component in the Hoeffding decomposition (1.6) converges to a Gaussian chaos under a normalization of a different order; similarly for the law of the iterated logarithm [20, Theorems 4.2.4 and 4.3.6]. Therefore, one might be tempted to derive Theorem 2.20 from the special cases $m = r$. However, that would mean to derive the MDP for the moderate U-empirical measures $\{M_n^{m,1}\}_{n\ge m}$ defined in (2.1) from the MDP for $\{M_n^{1,1}\}_{n\in\mathbb{N}}$. Due to the possible discontinuity of the map $\mathcal{M}(S) \ni \nu \mapsto \nu^{\otimes m}$ even in the $\tau(\mathbb{R})$-topology for $S = [0,1]$ with the Borel σ-algebra $\mathcal{S}$ (see [14, Exercise 7.3.18] for the case of the relative $\tau_1(\mathbb{R})$-topology on $\mathcal{M}_1(S)$), such an approach is apparently not possible, in particular when there is no topological structure on $(S,\mathcal{S})$ available. Even the general large deviation principle for $\{L_n^m\}_{m\ge n}$ requires

considerable work and does not follow easily in the way alluded to above from the large deviation principle for $\{L_n\}_{n\in\mathbb{N}}$, see [17].

For the proof of the lower bound of the above theorem, we adopt the classical method of change of measure used for large deviations lower bounds in Cramér's and Sanov's theorems. For moderate deviations, this change of measure is n-dependent, and moment inequalities for Banach-space valued U-statistics from the literature (established with decoupling and hypercontractive methods) have to be invoked instead of the usual weak law of large numbers.

In the proof of the upper bound for $r = 1$, see (3.48), we will use a moderate deviation result for sums of i. i. d. random vectors with values in a certain Banach space. Therefore we prove that by Conditions 2.11(c) and (d) we can apply [1, Theorem 2.3(i)]. For a special kind of Banach spaces and for $\{b_n\}_{n\in\mathbb{N}}$ satisfying an additional condition, Ledoux [25, Théorème] obtained a necessary and sufficient condition (weaker than Condition 2.11(c), (d)) for the upper bound of the MDP. We could apply this result. To do this we would have to impose the additional assumptions on E and $\{b_n\}_{n\in\mathbb{N}}$ as in [25, Théorème]. Then Conditions 2.11(c), (d) could be replaced by the conditions (i) and (ii) in [25, Théorème]. We omit the details. In the finite-dimensional case $E = \mathbb{R}^d$, the strong Cramér condition (2.10) for $\tilde{\varphi}_1$ from Condition 2.8 can be relaxed. The weak Cramér condition for $\tilde{\varphi}_1$ suffices, i. e., $\int_S \exp(\alpha_\varphi \|\tilde{\varphi}_1\|_{\mathbb{R}^d})\, d\mu < \infty$ for some $\alpha_\varphi > 0$, because moderate deviation results are available in this case, cf. [14, Theorem 3.7.1].

In the proof of the general upper bound we apply de Acosta's projective limit approach contained in [3]; however, there are some non-trivial problems in transferring this approach. To make sure that the terms $a \geq r+1$ in (2.5) have no influence on the scale of the moderate deviations of rank r, we will generalize exponential integrability properties for Rademacher chaos variables using a Khinchin–Kahane inequality for the Rademacher chaos. Moreover we will prove a Bernstein-type inequality for unbounded functions taking values in a Banach space of type p. Both results depend on decoupling and hypercontractive methods. We apply an L_p-approximation of symmetric and completely μ-degenerate functions. To circumvent the non-convexity of $I_{m,r}$, we introduce a modified rate function I_m, see (3.1).

2.2. **Application to U-statistics.** We want to derive moderate deviation results for Banach-space valued U-statistics from Theorem 2.20. Many statistics in common use are a member of this class and many other statistics may be approximated by a member of this class. Particularly in the field of non-parametric statistics, the significance of U-statistics is made plain, see for example the monographs [11] and [27] as well as the extensive list of references given there. See also the very recent monograph [20] for properties of U-statistics related to decoupling.

For an $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable map $\varphi\colon S^m \to E$ the U-statistics of degree m with kernel function φ are defined by (1.2) for all $n \geq m$. Define $\Phi = B(S^m, E) \cup \{\varphi\}$. Since the right-hand side of (2.2) is a composition of $M_n^{m,r}$ with the $\tau^\Phi(E)$-continuous functional $\mathcal{M}^\Phi(S^m) \ni \nu \mapsto \int_{S^m} \varphi\, d\nu$, the contraction principle [14, Theorem 4.2.1] immediately leads to the following theorem which, to the best of our knowledge, is new.

Theorem 2.23. *If E is a Banach space of type $p \in (1, 2]$, if $\lim_{n\to\infty} n/b_n^p = 0$, and if Condition 2.11 holds for $\Phi = B(S^m, E) \cup \{\varphi\}$ and $r \in \{1, \ldots, m\}$, then the*

sequence

$$\left\{ \Big(\frac{n}{b_n}\Big)^r \sum_{a=r}^{m} U_n^a(\tilde{\varphi}_a) \right\}_{n\geq m}$$

satisfies a full large deviation principle on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ *with the good rate function* $I^{\varphi}_{m,r}\colon E \to [0,\infty]$ *given by*

$$I^{\varphi}_{m,r}(x) = \inf\Big\{ I_{m,r}(\nu) \;\Big|\; \nu \in \mathcal{M}^{\Phi}(S^m),\, \int_{S^m} \varphi\, d\nu = x \Big\}.$$

Remark 2.24. From Theorem 2.23 we can derive a law of the iterated logarithm (LIL) for non-degenerate U-statistics, when φ satisfies Condition 2.11 and E is a Banach space of type 2. Let $b_n = \sqrt{2n\log\log n}$ for $n \geq 3$ and $\bar{\varphi} = \varphi - \int_{S^m} \varphi\, d\mu^{\otimes m}$. Then the compact LIL holds for $\{U_n^m(\bar{\varphi})\}_{n\geq m}$, i. e., the sequence $(\sqrt{n}/b_n)U_n^m(\bar{\varphi})$, $n \geq \max\{3,m\}$, is almost surely relatively compact and its limit set coincides with $K = \{\, x \in E \mid I_{\varphi,m}(x) \leq 1/2 \,\}$. We only point out that the MDP estimates obtained in Theorem 2.23 give the crucial estimates required by the classical proof of the LIL (in the case $m = 1$ cf. [26]). Arcones and Giné proved in [9, Theorem 3.2], that the LIL holds if $\int_{S^m} \|\varphi\|_E^2\, d\mu^{\otimes m} < \infty$. Notice moreover, that for the non-degenerate case, the rate $I^{\varphi}_{m,1}$ is the usual moderate rate function for the i. i. d. mean $\int_S \widetilde{\varphi}_1\, dL_n^1$ considered for example in [1, Theorem 2.3].

2.3. **Concluding remarks and examples.**

Remark 2.25. For the sake of completeness, let us mention that the lower and upper bounds (2.21) and (2.22) in Theorem 2.20 can be extended to arbitrary subsets of $\mathcal{M}(S^m)$. For every subset A of $S^{\mathbb{N}}$ let $\mathbb{P}^*(A)$ and $\mathbb{P}_*(A)$ denote the outer and the inner $\mathbb{P}$-measure of A, respectively. If $\mathbb{P}(M_n^{m,r} \in B)$ is replaced by $\mathbb{P}_*((M_n^{m,r})^{-1}(B))$ in (2.21) and by $\mathbb{P}^*((M_n^{m,r})^{-1}(B))$ in (2.22), then these bounds hold for all $B \subset \mathcal{M}(S^m)$. This can be seen by transferring the proof of [2, Theorem 1.2]. We refrain from giving the details.

Remark 2.26. The rate function $I_{m,r}$, governing the deviations of the moderate U-empirical measures $\{M_n^{m,r}\}_{n\geq m}$ of rank $r \in \{2,\dots,m\}$, is not convex if there exists a partition $\{S_1,S_2,S_3\} \subset \mathcal{S}$ of S such that $\mu(S_i) > 0$ for all $i \in \{1,2,3\}$. To see this, use conditioned measures to define $\tilde{\nu}_{(i)} = \mu(\,\cdot\,|\,S_i) - \mu(\,\cdot\,|\,S_3)$ for $i \in \{1,2\}$ and, using $\tilde{\nu}_{(i)}$, define $\nu_{(i)}$ via (2.18). If we assume that there exists a $\tilde{\nu} \in \mathcal{M}(S)$ for $\nu \equiv \frac{1}{2}(\nu_{(1)} + \nu_{(2)})$ satisfying (2.18), then $(\tilde{\nu}(S_i))^r = \nu(S_i^r \times S^{m-r}) = \frac{1}{2}$ for $i \in \{1,2\}$ and $(\tilde{\nu}(S_1 \cup S_2))^r = \nu((S_1 \cup S_2)^r \times S^{m-r}) = 1$. This gives the contradiction $1 = 2^{1-1/r}$.

The following example, derived from [26, Example 7.11], shows that Theorem 2.20(c) does not hold without an assumption on the type of the Banach space E.

Example 2.27. Let $(c_0, \|\cdot\|_\infty)$ denote the separable Banach space of all real sequences tending to zero. Let $\{\xi_{i,k}\}_{i,k\in\mathbb{N}}$ be a collection of independent random variables with $\mathbb{P}(\xi_{i,k} = 1) = \mathbb{P}(\xi_{i,k} = -1) = 1/2$ for all $i,k \in \mathbb{N}$. Consider $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ such that $b_n/n \downarrow 0$ as $n \to \infty$. Define $\alpha_k = \sqrt{b_n/n}$ whenever $(n-1)2^{n-1} < k \leq n2^n$. Then $X_i \equiv (\alpha_k \xi_{i,k})_{k\in\mathbb{N}}$ for $i \in \mathbb{N}$ gives an i. i. d. sequence of symmetric, c_0-valued random variables with $\|X_i\|_\infty \leq \bar{\alpha} \equiv \sup_{k\in\mathbb{N}} \alpha_k$ for all $i \in \mathbb{N}$. Define $S = E = c_0$, let μ denote the distribution of X_1, and define $\varphi\colon S \to E$ by $\varphi(s) = s$ for $\|s\|_\infty \leq \bar{\alpha}$ and $\varphi(s) = \bar{\alpha}s/\|s\|_\infty$ otherwise. Note that

$\varphi \in C_{\mathrm{b}}(S,E) \subset \Phi \equiv B(S,E)$, hence $B \equiv \{\, \nu \in \mathcal{M}(S) \mid \| \int_S \varphi \, d\nu \|_\infty < 1 \,\}$ is $\tau(c_0)$-open. Furthermore, B contains the zero measure, which satisfies $I_{1,1}(0) = 0$.

To evaluate the left-hand side of (2.21), first note that

$$\{M_n^{1,1} \in B\} = \left\{ \left\| \sum_{i=1}^n X_i \right\|_\infty < b_n \right\} \subset \bigcap_{k=1}^{n2^n} \left\{ \left| \sum_{i=1}^n \xi_{i,k} \right| < \sqrt{b_n n} \right\},$$

where we used that $\alpha_k \geq \sqrt{b_n/n}$ for all $k \leq n2^n$. Using $\mathbb{P}(\sum_{i=1}^n \xi_{i,k} = n) = 2^{-n}$, it follows for all $n \in \mathbb{N}$ with $b_n \leq n$ that $\mathbb{P}(M_n^{1,1} \in B) \leq (1 - 2^{-n})^{n2^n} \leq e^{-n}$. Hence, the left-hand side of (2.21) equals minus infinity.

3. Proofs of the moderate deviation principles

Define $I_m \colon \mathcal{M}(S^m) \to [0,\infty]$ by

$$I_m(\nu) = \begin{cases} \frac{1}{2} \int_{S^m} \left(\frac{d\nu}{d\mu^{\otimes m}} \right)^2 d\mu^{\otimes m} & \text{if } \nu \ll \mu^{\otimes m} \text{ and } \nu(S^m) = 0, \\ \infty & \text{otherwise.} \end{cases} \tag{3.1}$$

For a level $l \in [0,\infty)$ let $K(I_m, l) \equiv \{\, \nu \in \mathcal{M}(S^m) \mid I_m(\nu) \leq l \,\}$ denote the corresponding level set. For this moderate-deviation rate function we obtain:

Lemma 3.2. *Let $l \in [0,\infty)$.*

(a) *The set $K(I_m, l)$ is $\tau(\mathbb{R})$-compact and sequentially $\tau(\mathbb{R})$-compact.*
(b) *If $\Phi \subset L_2(\mu^{\otimes m}, E)$, then $K(I_m, l) \subset \mathcal{M}^\Phi(S^m)$.*
(c) *If $\Phi \subset L_2(\mu^{\otimes m}, E)$, then the identity on $K(I_m, l)$ is $\tau(\mathbb{R})$-$\tau^\Phi(E)$-continuous, hence both topologies coincide on this set and $K(I_m, l)$ is $\tau^\Phi(E)$-compact and also sequentially $\tau^\Phi(E)$-compact.*

Proof. (a) The $\tau(\mathbb{R})$-compactness follows by adapting [2, proof of Lemma 2.1], using $H_l \equiv \{\, f \in L_1(\mu^{\otimes m}, \mathbb{R}) \mid \int_{S^m} f^2 \, d\mu^{\otimes m} \leq 2l \,\}$. The sequential $\tau(\mathbb{R})$-compactness follows from [19, Theorem 2.6].

(b) If $\nu \in K(I_m, l)$ and $\varphi \in \Phi$, then, by the Cauchy–Schwarz inequality,

$$\int_{S^m} \|\varphi\|_E \, d|\nu| = \int_{S^m} \|\varphi\|_E \left| \frac{d\nu}{d\mu^{\otimes m}} \right| d\mu^{\otimes m} \leq \sqrt{2 I_m(\nu)} \, \|\varphi\|_{L_2(\mu^{\otimes m}, E)} < \infty.$$

(c) Given $\varphi \in \Phi$ and $\varepsilon > 0$, choose a measurable, finitely-valued approximation $\varphi_\varepsilon \colon S^m \to E$ such that $\sqrt{2l} \|\varphi - \varphi_\varepsilon\|_{L_2(\mu^{\otimes m}, E)} \leq \varepsilon$ to obtain similarly as in (b)

$$\left\| \int_{S^m} \varphi \, d\nu - \int_{S^m} \varphi_\varepsilon \, d\nu \right\|_E \leq \int_{S^m} \|\varphi - \varphi_\varepsilon\|_E \, d|\nu| \leq \varepsilon \tag{3.3}$$

for all $\nu \in K(I_m, l)$. Hence, $K(I_m, l) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is $\tau(\mathbb{R})$-continuous, because it is the uniform limit of the $\tau(\mathbb{R})$-continuous functions $K(I_m, l) \ni \nu \mapsto \int_{S^m} \varphi_\varepsilon \, d\nu$ as $\varepsilon \downarrow 0$. Hence, the identity on $K(I_m, l)$ is $\tau(\mathbb{R})$-$\tau^\Phi(E)$-continuous and the $\tau^\Phi(E)$-compactness of $K(I_m, l)$ follows from part (a). Since the identity is bijective, it is a $\tau(\mathbb{R})$-$\tau^\Phi(E)$-homeomorphism on $K(I_m, l)$ [24, Chap. 5, Theorem 8], hence both topologies coincide on $K(I_m, l)$. Therefore, the sequential $\tau^\Phi(E)$-compactness also follows from part (a). □

Proof of Theorem 2.20(a)*:* If $\nu \in \mathcal{M}(S^m)$ satisfies $I_{m,r}(\nu) < \infty$, then (2.18) implies that

$$\frac{d\nu}{d\mu^{\otimes m}}(s_1,\dots,s_m) = \sum_{\substack{A\subset\{1,\dots,m\}\\ |A|=r}} \prod_{i\in A} \frac{d\tilde{\nu}}{d\mu}(s_i) \qquad \mu^{\otimes m}\text{-a. s.} \tag{3.4}$$

Using (3.1), $\tilde{\nu}(S) = 0$ and (2.17), it follows that

$$I_m(\nu) = 2^{r-1}\binom{m}{r}(I_{m,r}(\nu))^r. \tag{3.5}$$

Hence, Theorem 2.20(a) follows from Lemma 3.2(b). □

Proof of Theorem 2.20(b)*:* The subset $C_{m,r}$ of all $(\nu,\tilde{\nu}) \in \mathcal{M}(S^m)\times\mathcal{M}(S)$ satisfying (2.18) is given by

$$C_{m,r} = \bigcap_{S_1,\dots,S_m\in\mathcal{S}} \Bigg\{ (\nu,\tilde{\nu}) \in \mathcal{M}(S^m)\times\mathcal{M}(S) \;\Bigg|\; \nu(S_1\times\cdots\times S_m) = \sum_{\substack{A\subset\{1,\dots,m\}\\ |A|=r}} \Bigg(\prod_{i\in\{1,\dots,m\}\setminus A} \mu(S_i)\Bigg)\Bigg(\prod_{i\in A}\tilde{\nu}(S_i)\Bigg)\Bigg\}.$$

Therefore, $C_{m,r}$ is closed in the product topology of the $\tau(\mathbb{R})$-topologies on $\mathcal{M}(S^m)$ and $\mathcal{M}(S)$. Using Lemma 3.2, it follows that

$$C'_{m,r} \equiv C_{m,r} \cap \Big(K\big(I_m, 2^{r-1}\tbinom{m}{r}l^r\big)\times K(I_1,l)\Big)$$

is compact and sequentially compact in the product topology of the $\tau^\Phi(E)$-topology on $\mathcal{M}^\Phi(S^m)$ and the $\tau(\mathbb{R})$-topology on $\mathcal{M}(S)$. Using (3.5), it follows that $K(I_{m,r},l)$ is the continuous image of $C'_{m,r}$ under the projection onto the $\mathcal{M}^\Phi(S^m)$-marginal. This implies part (b). □

Proof of Theorem 2.20(c)*:* Consider a measure $\nu \in \mathrm{int}_{\tau^\Phi(E)}(B)$ with $I_{m,r}(\nu) < \infty$. Then $\nu \in \mathcal{M}^\Phi(S^m)$ by the definition of the $\tau^\Phi(E)$-interior of the set B, hence $\int_{S^m} \|\varphi\|_E \, d|\nu| < \infty$ for every $\varphi \in \Phi$. By the definition of the $\tau^\Phi(E)$-topology, there exist $\eta > 0$, $k \in \mathbb{N}$ and $\varphi_1,\dots,\varphi_k \in \Phi$ such that the $\tau^\Phi(E)$-open set

$$C(\nu,2\eta) = \Big\{ \nu' \in \mathcal{M}^\Phi(S^m) \;\Big|\; \Big\|\int_{S^m} \varphi_i \, d(\nu-\nu')\Big\|_E < 2\eta \text{ for every } i\in\{1,\dots,k\}\Big\}$$

is contained in $\mathrm{int}_{\tau^\Phi(E)}(B)$. Since $I_{m,r}(\nu) < \infty$, a measure $\tilde{\nu} \in \mathcal{M}(S)$ for the representation (2.18) and a density $\tilde{g} \equiv d\tilde{\nu}/d\mu$ exist, but both might not be unique as the discussion after (2.18) showed. The density $\tilde{g}$ satisfies $\int_S \tilde{g}\, d\mu = 0$ and $\int_S \tilde{g}^2 \, d\mu = 2I_{m,r}(\nu) < \infty$. Furthermore, a density of ν with respect to $\mu^{\otimes m}$ is given by (3.4) and, for every $\varphi \in \Phi$,

$$\int_{S^m} \|\varphi(s_1,\dots,s_m)\|_E \prod_{i=1}^m |\tilde{g}(s_i)| \, \mu^{\otimes m}(ds_1,\dots,ds_m) = \int_{S^m} \|\varphi\|_E \, d|\nu| < \infty$$

in the case $m = r$; otherwise, when $\Phi \subset L_2(\mu^{\otimes m}, E)$,

$$\int_{S^m} \|\varphi(s_1,\dots,s_m)\|_E \sum_{\substack{A\subset\{1,\dots,m\}\\|A|=r}} \prod_{i\in A} |\tilde g(s_i)|\, \mu^{\otimes m}(ds_1,\dots,ds_m) \le \binom{m}{r} \|\varphi\|_{L_2(\mu^{\otimes m},E)} (2I_{m,r}(\nu))^r < \infty$$

by the Cauchy–Schwarz inequality. Define $\tilde g^{\pm}(s) = \max\{\pm\tilde g(s), 0\}$ for $s \in S$. It follows from the dominated convergence theorem that there exist two constants $c^{\pm} \in [0,\infty)$ smaller or equal to $\|\tilde g^{\pm}\|_{L_\infty(\mu)}$ such that the truncated function $g \equiv \min\{c^+, \max\{-c^-, \tilde g\}\}$ satisfies $\int_S g\, d\mu = 0$ and

$$\int_{S^m} \|\varphi_i\|_E\, d|\nu - \bar\nu| \le \eta, \qquad i \in \{1,\dots,k\}, \tag{3.6}$$

where $\bar\nu \in \mathcal{M}(S^m)$ is determined by

$$\frac{d\bar\nu}{d\mu^{\otimes m}}(s_1,\dots,s_m) = \sum_{\substack{A\subset\{1,\dots,m\}\\|A|=r}} \prod_{i\in A} g(s_i), \qquad (s_1,\dots,s_m) \in S^m.$$

Since g is bounded and $\Phi \subset L_1(\mu^{\otimes m}, E)$, the measure $\bar\nu$ is actually in $\mathcal{M}^\Phi(S^m)$.

Note that $C(\bar\nu,\eta) \subset C(\nu, 2\eta)$ by (3.6). Furthermore, $C(\bar\nu,\eta)$ is a measurable subset of $\mathcal{M}(S^m)$ because $\mathcal{M}^\Phi(S^m)$ is measurable by definition. Define the function $F_n(s) = \prod_{i=1}^n (1 + (b_n/n) g(s_i))$ for all $s = (s_i)_{i\in\mathbb{N}} \in S^{\mathbb{N}}$. Due to (1.9) there exists, for every $\varepsilon \in (0,1)$, an $n_\varepsilon \ge m$ such that $b_n c^-/n \le \varepsilon$ for all $n \ge n_\varepsilon$. Define $\mathbb{P}_n \in \mathcal{M}_1(\Omega)$ by $d\mathbb{P}_n/d\mathbb{P} = F_n$ for all these n. Then

$$\mathbb{P}(M_n^{m,r} \in B) \ge \mathbb{P}(M_n^{m,r} \in C(\bar\nu,\eta)) = \int_{D_n} \frac{1}{F_n}\, d\mathbb{P}_n, \tag{3.7}$$

where $D_n \equiv \{M_n^{m,r} \in C(\bar\nu,\eta)\}$. This is the change of measure step. In contrast to proofs of large deviations lower bounds, the change here is n-dependent. Defining $a_n = \mathbb{P}_n(D_n)$ and using Jensen's inequality, we obtain, for every $n \ge n_\varepsilon$,

$$\begin{aligned} \log \int_{D_n} \frac{1}{F_n}\, d\mathbb{P}_n &\ge \log a_n - \frac{1}{a_n}\int_{D_n} \log F_n\, d\mathbb{P}_n \\ &= \log a_n - \frac{1}{a_n}\int_{D_n} F_n \log F_n\, d\mathbb{P}. \end{aligned} \tag{3.8}$$

Since $x \log x \ge -1/e$ for $x \ge 0$ and $\int_S g\, d\mu = 0$, it follows that, for all $n \ge n_\varepsilon$,

$$\begin{aligned} \int_{D_n} F_n \log F_n\, d\mathbb{P} &\le \frac{1}{e} + \int_{S^{\mathbb{N}}} F_n \log F_n\, d\mathbb{P} \\ &= \frac{1}{e} + n\int_S \Big(1 + \frac{b_n}{n} g\Big) \log\Big(1 + \frac{b_n}{n} g\Big)\, d\mu \\ &\le \frac{1}{e} + \frac{b_n^2}{2(1-\varepsilon)n}\int_S g^2\, d\mu \le \frac{1}{e} + \frac{b_n^2}{(1-\varepsilon)n} I_{m,r}(\nu), \end{aligned} \tag{3.9}$$

where we used the estimate $(1+x)\log(1+x) \le x + x^2/(2(1-\varepsilon))$ for all $x \ge -\varepsilon$ for the second inequality. If we can show that $\lim_{n\to\infty} a_n = 1$, then (1.9), (3.7), (3.8)

and (3.9) together imply

$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(M_n^{m,r} \in B) \geq -\frac{I_{m,r}(\nu)}{1-\varepsilon},$$

which in turn implies (2.21).

To prove $\lim_{n\to\infty} a_n = 1$, we will apply moment inequalities for U-statistics instead of using the weak law of large numbers. It suffices to show that

$$\lim_{n\to\infty} \mathbb{P}_n\Big(\Big\{\Big\|\int_{S^m} \varphi\, dM_n^{m,r} - \int_{S^m} \varphi\, d\bar{\nu}\Big\|_E \leq \eta\Big\}\Big) = 1 \tag{3.10}$$

for every $\varphi \in \Phi$. Fix one $\varphi \in \Phi$ in the following and define the completely μ-degenerate functions $\tilde{\varphi}_r, \tilde{\varphi}_{r+1}, \ldots, \tilde{\varphi}_n$ as in (2.6). These functions satisfy (2.2). Define $\mu_n \in \mathcal{M}_1(S)$ by $d\mu_n/d\mu = 1 + (b_n/n)g$ for $n \geq n_\varepsilon$.

For every $n \geq n_\varepsilon$ and $a \in \{r, r+1, \ldots, m\}$ we now want to decompose $\tilde{\varphi}_a$ into a constant $\tilde{\varphi}_{a,0}^n$ and completely μ_n-degenerate functions $\tilde{\varphi}_{a,1}^n, \ldots, \tilde{\varphi}_{a,a}^n$ such that the analogue of (2.5) holds, namely

$$\int_{S^a} \tilde{\varphi}_a\, dL_n^a = \tilde{\varphi}_{a,0}^n + \sum_{c=1}^{a} \int_{S^c} \tilde{\varphi}_{a,c}^n\, dL_n^c \qquad \mathbb{P}\text{-a. s.} \tag{3.11}$$

Together with (2.2) this will lead to the representation

$$\int_{S^m} \varphi\, dM_n^{m,r} = \Big(\frac{n}{b_n}\Big)^r \sum_{a=r}^{m} \Big(\tilde{\varphi}_{a,0}^n + \sum_{c=1}^{a} \int_{S^c} \tilde{\varphi}_{a,c}^n\, dL_n^c\Big) \qquad \mathbb{P}\text{-a. s.} \tag{3.12}$$

To derive the decomposition (3.11) for $\tilde{\varphi}_a$, consider $A \subset \{1, \ldots, a\}$ and define

$$\varphi_{a,A,g}(s_1, \ldots, s_a) = \tilde{\varphi}_a(s_1, \ldots, s_a) \prod_{i \in A^{\mathrm{c}}} g(s_i), \qquad (s_1, \ldots, s_a) \in S^a, \tag{3.13}$$

where $A^{\mathrm{c}} \equiv \{1, \ldots, a\} \setminus A$. Define $\varphi_{a,A}^n$ by μ_n-integrating $\tilde{\varphi}_a$ with respect to the arguments with indices in A^{c}. Since $\tilde{\varphi}_a$ is completely μ-degenerate,

$$\tilde{\varphi}_{a,0}^n \equiv \varphi_{a,\varnothing}^n = \int_{S^a} \tilde{\varphi}_a\, d\mu_n^{\otimes a} = \Big(\frac{b_n}{n}\Big)^a \int_{S^a} \varphi_{a,\varnothing,g}\, d\mu^{\otimes a}.$$

If $|A| \in \{1, \ldots, a-1\}$, then

$$\varphi_{a,A}^n\big((s_i)_{i\in A}\big) = \left(\frac{b_n}{n}\right)^{a-|A|} \int_{S^{a-|A|}} \varphi_{a,A,g}(s_1, \ldots, s_a)\, \mu^{\otimes a-|A|}\big(d(s_i)_{i\in A^{\mathrm{c}}}\big)$$

for $\mu^{\otimes|A|}$-almost all $(s_i)_{i\in A} \in S^{|A|}$, and $\varphi_{a,A}^n = \varphi_{a,A,g} = \tilde{\varphi}_a$ if $A = \{1, \ldots, a\}$. Analogously to (2.3) we can define

$$\tilde{\varphi}_{a,A}^n\big((s_i)_{i\in A}\big) = \sum_{B\subset A} (-1)^{|A\setminus B|}\, \varphi_{a,B}^n\big((s_i)_{i\in B}\big) = \Big(\frac{b_n}{n}\Big)^{a-|A|} \tilde{\varphi}_{a,A,g}^n$$

with

$$\tilde{\varphi}_{a,A,g}^n\big((s_i)_{i\in A}\big) \equiv \sum_{B\subset A} \Big(-\frac{b_n}{n}\Big)^{|A\setminus B|} \int_{S^{a-|B|}} \varphi_{a,B,g}(s_1, \ldots, s_a)\, \mu^{\otimes a-|B|}\big(d(s_i)_{i\in B^{\mathrm{c}}}\big) \tag{3.14}$$

(drop the integral if $|B| = a$). Finally, for $c \in \{1, \dots, a\}$, similarly to (2.6), define

$$\tilde{\varphi}^n_{a,c} = \Big(\frac{b_n}{n}\Big)^{a-c} \tilde{\varphi}^n_{a,c,g} \qquad \text{with} \qquad \tilde{\varphi}^n_{a,c,g} \equiv \sum_{\substack{A \subset \{1,\dots,a\} \\ |A|=c}} \tilde{\varphi}^n_{a,A,g}. \tag{3.15}$$

Note that every $\tilde{\varphi}^n_{a,c,g}$ with $c \in \{1, \dots, a\}$ is completely μ_n-degenerate. Substituting these definitions into (3.12) shows that

$$\int_{S^m} \varphi \, dM_n^{m,r} = \sum_{a=r}^m \Big(\frac{b_n}{n}\Big)^{a-r} \varphi_{a,\varnothing,g} + \sum_{c=1}^m \Big(\frac{n}{b_n}\Big)^c \int_{S^c} \sum_{a=\max\{c,r\}}^m \Big(\frac{b_n}{n}\Big)^{a-r} \tilde{\varphi}^n_{a,c,g} \, dL_n^c.$$

Using (2.3), (2.6), (3.6), (3.13) and $\int_S g \, d\mu = 0$, it follows that $\int_{S^m} \varphi \, d\bar{\nu} = \varphi_{r,\varnothing,g}$. Since $\lim_{n\to\infty} b_n/n = 0$ by (1.9), it suffices for proving (3.10) that

$$\lim_{n\to\infty} \mathbb{P}_n\Big(\Big(\frac{n}{b_n}\Big)^c \Big\| \int_{S^c} \tilde{\varphi}^n_{a,c,g} \, dL_n^c \Big\|_E \le \frac{\eta}{m^2}\Big) = 1 \tag{3.16}$$

for every $a \in \{r, r+1, \dots, m\}$ and $c \in \{1, \dots, a\}$. Note that $X_1, \dots, X_n$ are i. i. d. with respect to $\mathbb{P}_n$. We may replace $\tilde{\varphi}^n_{a,c,g}$ in (3.16) by the symmetric version $\bar{\varphi}^n_{a,c,g} \equiv \frac{1}{a!} \sum_{\tau \in P_a} \tilde{\varphi}^n_{a,c,g} \circ \pi_\tau$, where P_a is the set of all permutations of $\{1, \dots, a\}$. Let $\mathbb{E}_n$ denote $\mathbb{P}_n$-expectation. Using [11, Theorem 2.7.1], it follows that

$$\mathbb{E}_n\Big[\Big\|\Big(\frac{n}{b_n}\Big)^c \int_{S^c} \bar{\varphi}^n_{a,c,g} \, dL_n^c\Big\|_E^p\Big] \le \Big(\frac{n}{b_n}\Big)^{cp} \frac{\big(6C_{E,p}^{1/p} c^3\big)^{cp}}{n_{(c)}^{p-1}} \int_{S^c} \|\bar{\varphi}^n_{a,c,g}\|_E^p \, d\mu_n^{\otimes c}, \tag{3.17}$$

where $C_{E,p}$ denotes the type constant of the Banach space E of type p, see (2.7). By the assumptions on $\{b_n\}_{n\in\mathbb{N}}$ in Theorem 2.20(c),

$$\lim_{n\to\infty} n_{(c)}^{1-p} \Big(\frac{n}{b_n}\Big)^{cp} = 0.$$

Hence (3.16) follows from (3.17) by using the Markov inequality, provided that

$$\limsup_{n\to\infty} \int_{S^c} \|\bar{\varphi}^n_{a,c,g}\|_E^p \, d\mu_n^{\otimes c} < \infty. \tag{3.18}$$

By Jensen's inequality, it suffices to show (3.18) for the possibly non-symmetric $\tilde{\varphi}^n_{a,c,g}$. Since $d\mu_n/d\mu = 1 + (b_n/n)g$ and since g is bounded, it suffices to show (3.18) with μ in place of μ_n. Inspection of (3.13), (3.14), (3.15), and Jensen's inequality shows that $\tilde{\varphi}_a \in L_p(\mu^{\otimes a}, E)$ for every $a \in \{r, r+1, \dots, m\}$ is the property we need, because g is bounded. Since $\varphi \in L_p(\mu^{\otimes m}, E)$, it follows via (2.3), (2.6) and Jensen's inequality that indeed $\tilde{\varphi}_a \in L_p(\mu^{\otimes a}, E)$ for every $a \in \{r, r+1, \dots, m\}$. □

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

$$\Pi_F \colon \mathcal{M}^\Phi(S^m) \to E^F \quad \text{by} \quad \Pi_F(\nu) = \Big(\int_{S^m} \varphi \, d\nu\Big)_{\varphi \in F}. \tag{3.19}$$

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 $(E^*)^F$. Using the notation from (2.6), we can define, for every $F \in \mathcal{F}$ and $y \in E^F$,

$$I_F(y) = \sup_{z \in (E^*)^F} \Big(\sum_{\varphi \in F} z_\varphi(y_\varphi) - \frac{1}{2} \int_S \Big(\sum_{\varphi \in F} z_\varphi(\tilde{\varphi}_1(s)) \Big)^2 \mu(ds) \Big). \tag{3.20}$$

Extending the argument in [3, pages 246–247] from the case $m = 1$ with the $\tau(\mathbb{R})$-topology to our situation and combining it with the separation argument from [15, page 429], we obtain the following variational characterization of the rate function, which is used once in (3.48):

Lemma 3.21. *If $\Phi \subset L_2(\mu^{\otimes m}, E)$, then, for every $F \in \mathcal{F}$ and $y \in E^F$,*

$$I_F(y) = \inf\{\, I_{m,1}(\nu) \mid \nu \in \mathcal{M}^\Phi(S^m),\, \Pi_F(\nu) = y \,\}.$$

Proof. Given $y \in E^F$ and $\nu \in \mathcal{M}^\Phi(S^m)$ satisfying $\Pi_F(\nu) = y$, we want to prove first that $I_F(y) \le I_{m,1}(\nu)$. It suffices to consider the case $I_{m,1}(\nu) < \infty$, hence $\nu_1 \ll \mu$, $\nu(S^m) = 0$ and $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$. Define $f = d\nu_1/d\mu$. Take any $z = (z_\varphi)_{\varphi \in F} \in (E^*)^F$ and define $g = \sum_{\varphi \in F} z_\varphi \circ \tilde{\varphi}_1$. Note that $g \in L_2(\mu, \mathbb{R})$. Since $fg \le \frac{1}{2}(f^2 + g^2)$, it follows with (2.17) that

$$\int_S g\, d\nu_1 - \frac{1}{2} \int_S g^2\, d\mu = \int_S \big(fg - \tfrac{1}{2} g^2 \big)\, d\mu \le \frac{1}{2} \int_S f^2\, d\mu = I_{m,1}(\nu). \tag{3.22}$$

Using (2.6), (2.3), $\nu_1(S) = 0$, and $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$,

$$\int_S \tilde{\varphi}_1\, d\nu_1 = \sum_{i=1}^m \int_S (\varphi_{\{i\}} - \varphi_\varnothing)\, d\nu_1 = \sum_{i=1}^m \int_{S^m} \varphi\, d(\mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}) = \int_{S^m} \varphi\, d\nu \tag{3.23}$$

for every $\varphi \in F$, hence, since $\Pi_F(\nu) = y$,

$$\int_S g\, d\nu_1 = \sum_{\varphi \in F} z_\varphi \Big(\int_S \tilde{\varphi}_1\, d\nu_1 \Big) = \sum_{\varphi \in F} z_\varphi(y_\varphi).$$

Together with (3.22), this implies that $I_F(y) \le I_{m,1}(\nu)$.

To complete the proof, we proceed by contradiction. Assume that there exists $y \in E^F$ with $I_F(y) < \infty$ such that $I_{m,1}(\nu) > l \equiv I_F(y)$ for all $\nu \in \mathcal{M}^\Phi(S^m)$ with $\Pi_F(\nu) = y$. Define $C = \{\, \Pi_F(\nu) \mid \nu \in K(I_{m,1}, l) \,\}$. Since $I_{m,1}$ is convex, the same is true for $K(I_{m,1}, l)$ and C. By Theorem 2.20(b), the set $K(I_{m,1}, l)$ is $\tau^\Phi(E)$-compact, hence C is compact in E^F. By the above assumption, $y \notin C$. Hence, by the Hahn–Banach theorem, there exist $z \in (E^*)^F$ and $\gamma \in \mathbb{R}$ such that $z(y) > \gamma \ge z(\Pi_F(\nu))$ for all $\nu \in K(I_{m,1}, l)$. Note that $z(y) > 0$, because the zero measure is in $K(I_{m,1}, l)$. Define $g = \sum_{\varphi \in F} z_\varphi \circ \tilde{\varphi}_1$. Since $z(y) > 0$ and $I_F(y) < \infty$, it follows from (3.20) by considering αz for $\alpha > 0$ that $\int_S g^2\, d\mu > 0$. Since $F \subset L_2(\mu^{\otimes m}, E)$, it follows that $g \in L_2(\mu, \mathbb{R})$. Hence, there exists $\alpha > 0$ such that $z(y) = \int_S \alpha g^2\, d\mu$. Define $\nu \in \mathcal{M}(S^m)$ by $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$, where $d\nu_1/d\mu = \alpha g$. Since $\int_S \tilde{\varphi}_1\, d\mu = 0$ for every $\varphi \in F$, it follows that $\nu(S^m) = 0$. Using (3.20), it follows that

$$I_F(y) \ge \alpha z(y) - \frac{1}{2} \int_S \Big(\sum_{\varphi \in F} \alpha z_\varphi(\tilde{\varphi}_1(s)) \Big)^2 \mu(ds) = \frac{1}{2} \int_S (\alpha g)^2\, d\mu = I_{m,1}(\nu),$$

hence $\nu \in K(I_{m,1}, l)$. Therefore, $z(\Pi_F(\nu)) \le \gamma$. But, by (3.23) and the choice of α,

$$z(\Pi_F(\nu)) = \int_{S^m} \sum_{\varphi\in F} z_\varphi(\varphi(s))\,\nu(ds) = \int_S \sum_{\varphi\in F} z_\varphi \circ \tilde{\varphi}_1\, d\nu_1 = \int_S \alpha g^2\, d\mu = z(y) > \gamma,$$

which is the desired contradiction. □

The following result generalizes [8, inequality (2.4)] from $\mathbb{R}$ to a Banach space E of type $p \in (1,2]$. A type-independent Banach space version is given in [20, Corollary 3.2.7].

Lemma 3.24. *Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be independent, symmetric, ±1-valued random variables. Let $(E, \|\cdot\|_E)$ be a Banach space of type $p \in (1,2]$, let $r \in \mathbb{N}$ and $\eta \in (0,2]$. Then there exists a constant $c_1 > 0$ such that, for every $\gamma \in (0, \eta/r)$, there exists $c_2 \in [1,\infty)$ satisfying*

$$\mathbb{E}\big[\exp\big(t\|Y\|_E^\gamma\big)\big] \le c_2 \exp\Big(\big((s/c_1)^\gamma\, t\big)^{1/(1-\gamma r/\eta)}\Big)$$

for all $t \ge 0$, $n \in \mathbb{N}$ and $x_{i_1,\dots,i_r} \in E$ with $(i_1,\dots,i_r) \in I(r,n)$, where

$$Y \equiv \sum_{(i_1,\dots,i_r)\in I(r,n)} \varepsilon_{i_1}\dots\varepsilon_{i_r} x_{i_1,\dots,i_r} \quad \text{and} \quad s \equiv \bigg(\sum_{(i_1,\dots,i_r)\in I(r,n)} \|x_{i_1,\dots,i_r}\|_E^p\bigg)^{1/p}.$$

Proof. If suffices to consider the case $s > 0$. By taking the logarithm of both sides and using its concavity, one can verify the well-known inequality $ab \le a^q/q + b^{q'}/q'$ for all $a, b \in [0,\infty)$ and $q, q' \in (1,\infty)$ satisfying $1/q + 1/q' = 1$. Hence

$$t\|Y\|_E^\gamma = \frac{s^\gamma t}{c_1^\gamma}\,\Big\|\frac{c_1}{s}Y\Big\|_E^\gamma \le \Big(1 - \frac{\gamma r}{\eta}\Big)\Big(\frac{s^\gamma t}{c_1^\gamma}\Big)^{1/(1-\gamma r/\eta)} + \frac{\gamma r}{\eta}\Big\|\frac{c_1}{s}Y\Big\|_E^{\eta/r}$$

for every $c_1 > 0$, which implies, since $\gamma r/\eta < 1$,

$$\mathbb{E}\big[\exp\big(t\|Y\|_E^\gamma\big)\big] \le \exp\Big(\Big(\frac{s^\gamma t}{c_1^\gamma}\Big)^{1/(1-\gamma r/\eta)}\Big)\mathbb{E}\big[\exp\big(\|c_1 Y/s\|_E^{\eta/r}\big)\big].$$

Choose $c_1 > 0$ so small that $C_{E,p}^{\eta/p} c_1^{\eta/r} e\eta r^{\eta-1} < p-1$; here $C_{E,p}$ denotes the type constant of $(E, \|\cdot\|_E)$, see (2.7).

To finish the proof, it suffices to show that there exists a constant $c_2 \in [1,\infty)$, depending only on c_1, $C_{E,p}$, γ, η, p and r, such that

$$\mathbb{E}\big[\exp\big(\|c_1 Y/s\|_E^{\eta/r}\big)\big] = 1 + \sum_{k=1}^\infty \frac{(c_1/s)^{\eta k/r}}{k!}\,\mathbb{E}\big[\|Y\|_E^{\eta k/r}\big] \le c_2. \tag{3.25}$$

If $k \le pr/\eta$, then $\mathbb{E}[\|Y\|_E^{\eta k/r}] \le \mathbb{E}[\|Y\|_E^p]^{\eta k/(pr)}$ by Jensen's inequality. If $k \ge pr/\eta$, then by Borell's inequality (a Khintchine-type inequality for the Rademacher chaos, see [11, (2.6.5)]),

$$\mathbb{E}\big[\|Y\|_E^{\eta k/r}\big] \le \Big(\frac{\eta k/r - 1}{p-1}\Big)^{\eta k/2}\,\mathbb{E}\big[\|Y\|_E^p\big]^{\eta k/(pr)}.$$

Let $\{\varepsilon_i^{(j)}\}_{i\in\mathbb{N}}$ for $j \in \{1,\dots,r\}$ be independent copies of $\{\varepsilon_i\}_{i\in\mathbb{N}}$. By decoupling Y, see [11, Theorem 2.5.4], and using r times the fact that $(E, \|\cdot\|_E)$ is of type p, see (2.7), it follows that

$$\mathbb{E}\big[\|Y\|_E^p\big] \le r^{pr}\,\mathbb{E}\bigg[\bigg\|\sum_{(i_1,\dots,i_r)\in I(r,n)} \varepsilon_{i_1}^{(1)}\dots\varepsilon_{i_r}^{(r)} x_{i_1,\dots,i_r}\bigg\|_E^p\bigg] \le C_{E,p}^r\, r^{pr} s^p.$$

Since $\eta k/r - 1 \le \eta k/r$ and $\eta/2 \le 1$ as well as $k^k \le k!\, e^k$, we obtain

$$\mathbb{E}\big[\exp\big(\|c_1 Y/s\|_E^{\eta/r}\big)\big] \le 1 + \sum_{k=1}^{\lfloor pr/\eta \rfloor} \frac{\big(C_{E,p}^{\eta/p} c_1^{\eta/r} r^\eta\big)^k}{k!} + \sum_{k=\lceil pr/\eta \rceil}^{\infty} \left(\frac{C_{E,p}^{\eta/p} c_1^{\eta/r} e \eta r^{\eta-1}}{p-1}\right)^k \equiv c_2$$

by inserting the estimates of $\mathbb{E}[\|Y\|_E^{\eta k/r}]$ into (3.25). □

One technical key of the proof of Theorem 2.20(d) will be a Bernstein-type inequality for completely μ-degenerate U-statistics of degree r with $1 \le r \le m$. For bounded $\mathbb{R}$-valued kernel functions, such an inequality is given in [8, Proposition 2.3]; for bounded kernel functions with values in a Banach space of type 2, a Bernstein-type inequality is given in [11, Theorem 8.15, Corollary 8.1.5]. Along the lines of [20, proof of Theorem 4.1.12], we derive an inequality for unbounded functions taking values in a Banach space of type p, whose norm raised to the p-th power satisfies the so-called weak Cramér Condition. The decoupling result, which is an essential tool in our proof, was originally obtained in [28] and [29].

Theorem 3.26 (Bernstein-type inequality)**.** *Let $(E, \|\cdot\|_E)$ be a separable Banach space of type $p \in (1,2]$ and let $r \in \mathbb{N}$. Then there exist constants $c_3, c_4 \in [1,\infty)$ such that the following holds: For every symmetric, Bochner $\mu^{\otimes r}$-integrable, completely μ-degenerate function $\varphi\colon S^r \to E$ with an $\alpha > 0$ satisfying*

$$\tilde{a} \equiv \int_{S^r} \exp\big(\alpha \|\varphi\|_E^p\big)\, d\mu^{\otimes r} < \infty \tag{3.27}$$

there exists a constant $c_\varphi \in (0,\infty)$ such that, for all $x > 0$ and all integers $n \ge r$,

$$\mathbb{P}\big(\|n^\varrho U_n^r(\varphi)\|_E \ge x\big) \le c_3 \exp\left(-\frac{x^{2/r}}{c_4 \sigma^{2/r} + c_\varphi (x^{2/r}/n)^{1/(1+pr/2)}}\right), \tag{3.28}$$

where $\sigma \equiv \|\varphi\|_{L_p(\mu^{\otimes r}, E)}$ and $\varrho \equiv r(1-1/p)$.

Proof. By Chebychev's inequality, for all $t, x \in [0,\infty)$ and integers $n \ge r$,

$$\mathbb{P}\big(\|n^\varrho U_n^r(\varphi)\|_E \ge x\big) \le \exp(-tx^\gamma)\, \mathbb{E}\big[\exp\big(t\|n^\varrho U_n^r(\varphi)\|_E^\gamma\big)\big], \tag{3.29}$$

where $\gamma \equiv p/\kappa$ with $\kappa \equiv 1 + pr/2$. The function $\Psi(y) \equiv \exp(\max\{y^\gamma, (1-\gamma)/\gamma\})$ for $y \ge 0$ is convex, increasing and satisfies

$$\exp(y^\gamma) \le \Psi(y) \le e^{(1-\gamma)/\gamma} \exp(y^\gamma) \tag{3.30}$$

for all $y \ge 0$. Using (3.30) and [11, Theorem 2.5.4], we obtain

$$\mathbb{E}\big[\exp\big(t\|n^\varrho U_n^r(\varphi)\|_E^\gamma\big)\big] \le \mathbb{E}\big[\Psi\big(r^r t^{1/\gamma} n^\varrho \|U_n^{r,\mathrm{dec}}(\varphi)\|_E\big)\big] \tag{3.31}$$

where the decoupled U-statistic is defined by

$$U_n^{r,\mathrm{dec}}(\varphi) \equiv \frac{1}{n_{(r)}} \sum_{(i_1,\ldots,i_r) \in I(r,n)} \varphi(X_{i_1}^{(1)}, \ldots, X_{i_r}^{(r)})$$

and $\{X_i^{(j)}\}_{i\in\mathbb{N}}$ for $j \in \{1,\ldots,r\}$ are independent copies of $\{X_i\}_{i\in\mathbb{N}}$, see (1.1) and (1.2) for the remaining notation. Let $\{\varepsilon_i^{(j)}\}_{i\in\mathbb{N}}$ for $j \in \{1,\ldots,r\}$ be sequences of independent, symmetric, $\{-1,1\}$-valued random variables, independent of one another and all the "X-variables." Applying r times the symmetrization lemma [11, Lemma 2.4.5], which uses the complete μ-degeneracy of φ, it follows that

$$\mathbb{E}\big[\Psi\big(r^r t^{1/\gamma} n^\varrho \|U_n^{r,\mathrm{dec}}(\varphi)\|_E\big)\big] \le \mathbb{E}\big[\Psi\big(2^r r^r t^{1/\gamma} n^\varrho \|U_{n,\mathrm{sym}}^{r,\mathrm{dec}}(\varphi)\|_E\big)\big] \tag{3.32}$$

with

$$U^{r,\mathrm{dec}}_{n,\mathrm{sym}}(\varphi) \equiv \frac{1}{n_{(r)}} \sum_{(i_1,\ldots,i_r)\in I(r,n)} \varepsilon^{(1)}_{i_1} \ldots \varepsilon^{(r)}_{i_r} \varphi(X^{(1)}_{i_1}, \ldots, X^{(r)}_{i_r}).$$

Using [11, Theorem 2.5.4] to couple the independent copies and using the upper bound from (3.30) afterwards, we obtain

$$\mathbb{E}\big[\Psi\big(2^r r^r t^{1/\gamma} n^{\varrho} \|U^{r,\mathrm{dec}}_{n,\mathrm{sym}}(\varphi)\|_E\big)\big] \le e^{(1-\gamma)/\gamma}\, \mathbb{E}\big[\exp\big(C_r^{\gamma} t n^{\gamma\varrho} \|U^{r}_{n,\mathrm{sym}}(\varphi)\|^{\gamma}_E\big)\big] \quad (3.33)$$

with

$$U^{r}_{n,\mathrm{sym}}(\varphi) \equiv \frac{1}{n_{(r)}} \sum_{(i_1,\ldots,i_r)\in I(r,n)} \varepsilon_{i_1} \ldots \varepsilon_{i_r} \varphi(X_{i_1}, \ldots, X_{i_r})$$

and $C_r \equiv 2^r r^r \sum_{k=1}^r k^r/((r-k)!\,k!)$. Taking the expectation with respect to $\{\varepsilon_i\}_{i\in\mathbb{N}}$, an application of Lemma 3.24 with $\eta = 2$ gives

$$\mathbb{E}\big[\exp\big(C_r^{\gamma} t n^{\gamma\varrho} \|U^{r}_{n,\mathrm{sym}}(\varphi)\|^{\gamma}_E\big)\big] \le c_2\, \mathbb{E}\big[\exp\big(c_1^{-p}(C_r^{\gamma} t n^{\gamma\varrho})^{\kappa} n_{(r)}^{1-p} U^{r}_{n,p}(\varphi)\big)\big] \quad (3.34)$$

with

$$U^{r}_{n,p}(\varphi) \equiv \frac{1}{n_{(r)}} \sum_{(i_1,\ldots,i_r)\in I(r,n)} \|\varphi(X_{i_1}, \ldots, X_{i_r})\|^p_E.$$

Since $1/n_{(r)} \le r!/n^r$ and $\gamma\varrho\kappa = r(p-1)$, it follows that $n^{\gamma\varrho\kappa} n_{(r)}^{1-p} \le r!$ for all $n \ge r$. Using Hoeffding's formula [23], see e. g. [11, p. 14], and Jensen's inequality, we obtain

$$\mathbb{E}\big[\exp\big(c_1^{-p} C_r^p r!\, t^{\kappa} U^{r}_{n,p}(\varphi)\big)\big] \le \exp\big(c_5 \sigma^p t^{\kappa}\big)\, \mathbb{E}\big[\exp\big(c_5 t^{\kappa} S_{\lfloor n/r\rfloor}\big)\big], \quad (3.35)$$

where $S_k \equiv (1/k)\sum_{i=1}^k Y_i$ with $Y_i \equiv \|\varphi(X_{(i-1)r+1}, \ldots, X_{ir})\|^p_E - \sigma^p$ for $i, k \in \mathbb{N}$ and $c_5 \equiv c_1^{-p} C_r^p r!$. Using $\mathbb{E}[Y_i] = 0$, the estimate $y^k \le k! e^y$ for $y \in \mathbb{R}$, condition (3.27), $a \equiv \tilde{a}\exp(-\alpha\sigma^p)$, the geometric series, and $1 + y \le e^y$ for $y \in \mathbb{R}$, we get

$$\mathbb{E}[\exp(hY_1)] = 1 + \sum_{k=2}^{\infty} \Big(\frac{h}{\alpha}\Big)^k \frac{\mathbb{E}[(\alpha Y_1)^k]}{k!} \le 1 + a \sum_{k=2}^{\infty} \Big(\frac{h}{\alpha}\Big)^k \le \exp\Big(\frac{ah^2}{\alpha^2 - \alpha h}\Big)$$

for all $h \in [0, \alpha)$, hence, due to independence,

$$\mathbb{E}[\exp(hS_k)] = \big(\mathbb{E}[\exp(hY_1/k)]\big)^k \le \exp\Big(\frac{ah^2}{\alpha^2 k - \alpha h}\Big) \quad (3.36)$$

for all $k \in \mathbb{N}$ and $h \in [0, \alpha k)$. Combination of (3.29) and (3.31)–(3.36) yields

$$\mathbb{P}\big(\|n^{\varrho} U^r_n(\varphi)\|_E \ge x\big) \le c_3 \exp\bigg(-tx^{\gamma} + c_5\sigma^p t^{\kappa} + \frac{a c_5^2 t^{2\kappa}}{\alpha^2 \lfloor n/r\rfloor - \alpha c_5 t^{\kappa}}\bigg) \quad (3.37)$$

for all integers $n \ge r$ and all $t, x \in [0, \infty)$ such that $c_5 t^{\kappa} < \alpha\lfloor n/r\rfloor$, where $c_3 \equiv c_2 \exp((1-\gamma)/\gamma)$. The choice $t_0 \equiv (x^{\gamma}/(c_5\kappa\sigma^p))^{2/(pr)}$ minimizes the sum of the first two terms in the exponent of (3.37) and yields

$$-t_0 x^{\gamma} + c_5\sigma^p t_0^{\kappa} = -\frac{2}{c_4}\Big(\frac{x}{\sigma}\Big)^{2/r} \qquad \text{with} \qquad c_4 \equiv \frac{4 c_5^{2/(pr)} \kappa^{1+2/(pr)}}{pr}.$$

Note that it suffices to consider $\sigma > 0$, because (3.28) holds if $\sigma = 0$. If t_0 satisfies

$$c_5 t_0^{\kappa} \le \alpha c_6 \lfloor n/r\rfloor \qquad \text{with} \qquad c_6 \equiv \min\Big\{\frac{1}{2}, \frac{\alpha pr\sigma^p}{8a}\Big\}, \quad (3.38)$$

then the last term of (3.37) is bounded by $(1/c_4)(x/\sigma)^{2/r}$ and (3.28) follows.

If t_0 does not satisfy (3.38), then consider $t_1 > 0$ determined by the equation $t_1^\kappa = \alpha c_7 \lfloor n/r \rfloor / c_5$, where $c_7 \in (0,1)$ denotes the solution of

$$1 + \frac{ac_7}{\alpha\sigma^p(1-c_7)} = \frac{\kappa}{2}\Big(\frac{c_6}{c_7}\Big)^{pr/(2\kappa)}. \tag{3.39}$$

We can rewrite the exponent from (3.37) and then use (3.39) to obtain

$$-t_1x^\gamma + c_5\Big(\sigma^p + \frac{ac_7}{\alpha(1-c_7)}\Big)t_1^\kappa = -t_1\Big(x^\gamma - \frac{c_5\kappa\sigma^p}{2}\Big(\frac{c_6}{c_7}\Big)^{pr/(2\kappa)} t_1^{pr/2}\Big). \tag{3.40}$$

Since t_0 does not satisfy (3.38), it follows that $t_0^\kappa > c_6 t_1^\kappa / c_7$, hence

$$x^\gamma = c_5\kappa\sigma^p t_0^{pr/2} > c_5\kappa\sigma^p (c_6/c_7)^{pr/(2\kappa)} t_1^{pr/2}$$

and $-t_1x^\gamma/2$ is an upper estimate of (3.40). This implies (3.28). □

Next we give an L_p-approximation of symmetric and completely μ-degenerate functions. This is a key step in our proof of the upper bound.

Lemma 3.41. *Let $\sigma > 0$, $p \geq 1$, $r \in \mathbb{N}$, and let $\varphi \in L_p(\mu^{\otimes r}, E)$ be a symmetric and completely μ-degenerate function. Then there exist $k \in \mathbb{N}$, vectors $\beta_1, \dots, \beta_k \in E$ and bounded measurable functions $f_1, \dots, f_k \colon S \to \mathbb{R}$ with $\int_S f_i \, d\mu = 0$ for all $i \in \{1, \dots, k\}$ such that $\varphi_\sigma \equiv \sum_{i=1}^k \beta_i f_i^{\otimes r}$ satisfies $\|\varphi - \varphi_\sigma\|_{L_p(\mu^{\otimes r}, E)} \leq \sigma$, where $f_i^{\otimes r}(s_1, \dots, s_r) \equiv \prod_{j=1}^r f_i(s_j)$ for all $s_1, \dots, s_r \in S$.*

Proof. Going through the proof of [6, Lemma 3.2] shows that there exist $n \in \mathbb{N}$, vectors $\alpha_1, \dots, \alpha_n \in E$ and sets $A_{i,1}, \dots, A_{i,r} \in \mathcal{S}$ for $i \in \{1, \dots, n\}$ such that

$$\varphi_\sigma \equiv \frac{1}{r!} \sum_\tau \sum_{i=1}^n \alpha_i \, (g_{i,1} \otimes \cdots \otimes g_{i,r}) \circ \pi_\tau$$

satisfies $\|\varphi - \varphi_\sigma\|_{L_p(\mu^{\otimes r}, E)} \leq \sigma$, where the sum extends over all permutations τ of $\{1, \dots, r\}$, the map π_τ is as in (2.14), and $g_{i,j} \equiv 1_{A_{i,j}} - \mu(A_{i,j})$. Note that φ_σ is symmetric and completely μ-degenerate. By polarization,

$$\sum_\tau (g_{i,1} \otimes \cdots \otimes g_{i,r}) \circ \pi_\tau = \frac{1}{2^r} \sum_{\varepsilon_1, \dots, \varepsilon_r \in \{-1,1\}} \varepsilon_1 \dots \varepsilon_r \Big(\sum_{i=1}^r \varepsilon_i \, g_{i,1}\Big) \otimes \cdots \otimes \Big(\sum_{i=1}^r \varepsilon_i \, g_{i,r}\Big)$$

for every $i \in \{1, \dots, n\}$, which proves that φ_σ has the desired form. Notice that a similar argument for $L_2(\mu^{\otimes r}, \mathbb{R})$ is contained in [7, p. 660]; the argument appeared first implicitly in [12]. □

Proof of Theorem 2.20(d)*:* Let C denote the $\tau^\Phi(E)$-closure of $B \cap \mathcal{M}^\Phi(S^m)$. Due to Assumption 2.16, the moderate U-empirical measures $\{M_n^{m,r}\}_{n \geq m}$ take values in $\mathcal{M}^\Phi(S^m)$. It suffices to consider only the case $I_{m,r}(C) > 0$. Choose $l \in (0, I_{m,r}(C))$. By (a) and (b), the set $K(I_{m,r}, l)$ is contained in $\mathcal{M}^\Phi(S^m)$ and $\tau^\Phi(E)$-compact.

Let us use the notation introduced just before Lemma 3.21. Since C is $\tau^\Phi(E)$-closed and $C \cap K(I_{m,r}, l) = \varnothing$, there exist, for every $\nu \in K(I_{m,r}, l)$, an $F_\nu \in \mathcal{F}$ and an open neighbourhood $U_\nu \subset E^{F_\nu}$ of $\Pi_{F_\nu}(\nu)$ such that $C \cap \Pi_{F_\nu}^{-1}(U_\nu) = \varnothing$. Since $K(I_{m,r}, l)$ is $\tau^\Phi(E)$-compact by (b), there exists a finite subset N of $K(I_{m,r}, l)$ such that $\bigcup_{\nu \in N} \Pi_{F_\nu}^{-1}(U_\nu)$ covers $K(I_{m,r}, l)$. 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 $K(I_{m,r},l)$ and is disjoint from C. Define $\varepsilon = \operatorname{dist}(\Pi_F(K(I_{m,r},l)), U^{\mathrm{c}})$. Since $\Pi_F(K(I_{m,r},l))$ is a compact subset of the open set U, it follows that $\varepsilon > 0$ and that

$$A_\varepsilon \equiv \{\, x \in E^F \mid \operatorname{dist}(x, \Pi_F(K(I_{m,r},l))) < \varepsilon \,\}$$

is an open set contained in U. Thus we can therefore find $F \in \mathcal{F}$ and an open (possibly non-convex) ε-neighbourhood $A_\varepsilon \subset E^F$ of $\Pi_F(K(I_{m,r},l))$ such that

$$\{M_n^{m,r} \in B\} \subset \{M_n^{m,r} \in C\} \subset \{\Pi_F(M_n^{m,r}) \in E^F \setminus A_\varepsilon\} \tag{3.42}$$

for all $n \geq m$. Using (2.2), it follows that

$$\begin{aligned} \mathbb{P}(\,\Pi_F(M_n^{m,r}) \in E^F \setminus A_\varepsilon\,) \leq \sum_{a=r+1}^{m} \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|\Pi_{F_a}(L_n^a)\big\|_{E^F} \geq \frac{\varepsilon}{4m}\Big) \\ + \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \Pi_{F_r}(L_n^r) \in E^F \setminus A_{3\varepsilon/4}\Big), \end{aligned} \tag{3.43}$$

where $F_a \equiv \{\tilde\varphi_a\}_{\varphi\in F}$ for $a \in \{r, r+1, \ldots, m\}$ with the notation from (2.6).

In view of (1.1), we may and will assume in the following (without loss of generality) that F, and therefore F_r up to F_m, consist of symmetric functions: If we replace a φ by its symmetrized version, then (2.3) and (2.6) lead to the symmetrized version of $\tilde\varphi_a$, Assumption 2.16 holds for the symmetrized version of φ, and, due to Jensen's inequality, (2.12) also holds for the symmetrized version.

In order to see that the terms with $a \geq r+1$ in (3.43) do not contribute to the moderate deviations upper bound (2.22), it suffices to prove that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|U_n^a(\tilde\varphi_a)\big\|_E \geq \eta\Big) = -\infty \tag{3.44}$$

for every $a \in \{r+1, \ldots, m\}$, $\eta > 0$ and $\tilde\varphi_a \in F_a$. Consider $a \in \{r, \ldots, m\}$. Define $\varrho_a = a(1 - 1/p_a)$ with $p_a \in (1, 2]$ according to Condition 2.11. Then

$$\mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|U_n^a(\tilde\varphi_a)\big\|_E \geq \eta\Big) = \mathbb{P}\big(\big\|n^{\varrho_a} U_n^a(\tilde\varphi_a)\big\|_E \geq \eta x_{a,n}\big) \tag{3.45}$$

with $x_{a,n} \equiv b_n^r / n^{r-\varrho_a}$ for all $n \geq a$. Using $\varrho_a \leq a/2$ and (1.9), it follows that

$$\frac{x_{a,n}^{2/a}}{n} = \Big(\frac{b_n^r}{n^{r-\varrho_a+a/2}}\Big)^{2/a} \to 0 \qquad \text{as } n \to \infty. \tag{3.46}$$

By Condition 2.11(a), for every $a \in \{r+1, \ldots, m\}$,

$$\frac{n}{b_n^2} x_{a,n}^{2/a} = \Big(\frac{n^{3/2-1/p_a-r/a}}{b_n^{1-r/a}}\Big)^2 \to \infty \qquad \text{as } n \to \infty. \tag{3.47}$$

Hence, for every $a \in \{r+1, \ldots, m\}$, (3.44) follows from (3.45) and an application of the Bernstein-type inequality (3.28).

To treat the last term in (3.43), let us first consider the easier case $r = 1$. Note that $(n/b_n)\Pi_{F_1}(L_n^1) = \Pi_{F_1}(M_n^{1,1})$ is the sum of n independent and identically distributed E^F-valued random vectors of mean zero. Condition 2.11(c) and Hölder's inequality imply that $\int_S \exp(\alpha\|(\tilde\varphi_1)_{\varphi\in F}\|_{E^F})\, d\mu < \infty$ for every $\alpha > 0$. Condition 2.11(d) implies that the sequence of the distributions of $n^{-1/2}\sum_{i=1}^n (\tilde\varphi_1(X_i))_{\varphi\in F}$

converges weakly to a Gaussian measure on E^F as $n \to \infty$. Therefore, it follows from [1, Theorem 2.3(i)] and Lemma 3.21 that

$$\begin{aligned}\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\big(\Pi_{F_1}(M_n^{1,1}) \in E^F \setminus A_{3\varepsilon/4}\big) &\le -I_F\big(E^F \setminus A_{3\varepsilon/4}\big) \\ &= -I_{m,1}\big(\Pi_F^{-1}(E^F \setminus A_{3\varepsilon/4})\big) \le -I_{m,1}\big(\mathcal{M}^\Phi(S^m) \setminus K(I_{m,1}, l)\big) \le -l. \quad (3.48)\end{aligned}$$

Together with (3.42), (3.43) and (3.44), this estimate implies (2.22) for $r = 1$.

To treat the last term in (3.43) in the case $r \in \{2, 3, \dots, m\}$, choose $\sigma > 0$ so that $\varepsilon/(4|F|\sigma) > (\max\{c_4, 2\}l)^{r/2}$ with c_4 as in (3.28). For every $\tilde\varphi_r$ corresponding to a $\varphi \in F$ we can find a function $\tilde\varphi_{r,\sigma}$ as described in Lemma 3.41 such that $\|\tilde\varphi_r - \tilde\varphi_{r,\sigma}\|_{L_2(\mu^{\otimes r},E)} \le \sigma$. Define $F_{r,\sigma} = \{\tilde\varphi_{r,\sigma}\}_{\varphi\in F}$. Then

$$\begin{aligned}\mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \Pi_{F_r}(L_n^r) \in E^F \setminus A_{3\varepsilon/4}\Big) \le \sum_{\varphi\in F} \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|U_n^r(\tilde\varphi_r - \tilde\varphi_{r,\sigma})\big\|_E \ge \frac{\varepsilon}{4|F|}\Big) \\ + \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \Pi_{F_{r,\sigma}}(L_n^r) \in E^F \setminus A_{\varepsilon/2}\Big). \quad (3.49)\end{aligned}$$

In order to show that, for every $\varphi \in F$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|U_n^r(\tilde\varphi_r - \tilde\varphi_{r,\sigma})\big\|_E \ge \frac{\varepsilon}{4|F|}\Big) \le -l, \quad (3.50)$$

apply (3.45) with $a = r$, $p_r = 2$, $\varrho_r = r/2$, $\eta = \varepsilon/(4|F|)$, and $\tilde\varphi_a$ replaced by $\tilde\varphi_r - \tilde\varphi_{r,\sigma}$. Then use the Bernstein-type inequality (3.28), note that (3.46) holds, and that the expression in (3.47) is always 1. Hence, the upper bound in (3.50) is determined by the choice of σ, which measures the quality of the approximation.

To treat the last term in (3.49), we want to prove that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big\|\Pi_{F_{r,\sigma}}(L_n^r - L_n^{\otimes r})\big\|_{E^F} \ge \frac{\varepsilon}{4}\Big) = -\infty. \quad (3.51)$$

Due to the form of the functions in $F_{r,\sigma}$, as described in Lemma 3.41, it suffices to show

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \big|U_n^r(f^{\otimes r}) - (U_n^1(f))^r\big| \ge \eta\Big) = -\infty \quad (3.52)$$

for every $\eta > 0$ and every bounded measurable function $f\colon S \to \mathbb{R}$ satisfying $\int_S f\,d\mu = 0$. We start be recalling Newton's formula (see, e. g., [13, p. 179]): Given $n \in \mathbb{N}$ and variables $x_1, \dots, x_n$, define the elementary symmetric polynomial $e_k = \sum_{1\le i_1<\dots<i_k\le n} x_{i_1} \dots x_{i_k}$ and the power sum $p_k = \sum_{i=1}^n x_i^k$ for every $k \in \{1, \dots, n\}$. Newton's formula asserts that

$$ke_k = (-1)^{k-1}p_k + \sum_{j=1}^{k-1} (-1)^{j-1} e_{k-j}p_j \qquad \text{for } k \in \{1, \dots, n\}.$$

Using this formula recursively for $k \in \{r, \dots, 2, 1\}$, it follows that there exists a polynomial R_r of degree r with r indeterminates such that $r!e_r - p_1^r = R_r(p_1, \dots, p_r)$; furthermore, every monomial of $R_r(p_1, \dots, p_r)$ is of the form $ap_1^{k_1} \dots p_r^{k_r}$ with $a \in \mathbb{Z}$ and $k_1, \dots, k_r \in \mathbb{N}_0$ satisfying $\sum_{j=1}^r jk_j = r$ and $k_1 \le r - 1$, hence actually

$k_1 \le r-2$. Setting $x_i = f(X_i)$ for $i \in \{1,\dots,n\}$, we see that $r!e_r = n_{(r)}U_n^r(f^{\otimes r})$ and $p_k = nU_n^1(f^k)$ for $k \in \{1,\dots,r\}$, hence

$$\Big(\frac{n}{b_n}\Big)^r \big(U_n^r(f^{\otimes r}) - (U_n^1(f))^r\big) = \Big(\frac{n}{b_n}\Big)^r \Big(1 - \frac{n_{(r)}}{n^r}\Big)U_n^r(f^{\otimes r}) + R_r\Big(\frac{n}{b_n}U_n^1(f), \frac{n}{b_n^2}U_n^1(f^2), \dots, \frac{n}{b_n^r}U_n^1(f^r)\Big). \tag{3.53}$$

Since $f^{\otimes r}$ is bounded and completely μ-degenerate and $1 - n_{(r)}/n^r \to 0$ as $n \to \infty$, it follows in a similar way as (3.50) from the Bernstein-type inequality (3.28) that we can neglect the first term on the right-hand side of (3.53). Every monomial of the last term of (3.53) contains at least one factor of the form $nb_n^{-k}U_n^1(f^k)$ with $k \in \{2,3,\dots,r\}$. Since f is bounded, $\lim_{n\to\infty} b_n = \infty$ and $\lim_{n\to\infty} n/b_n^2 = 0$ by (1.9), it follows that there exists a constant $C > 0$ such that, for all $n \ge r$,

$$\Big|R_r\Big(\frac{n}{b_n}U_n^1(f), \dots, \frac{n}{b_n^r}U_n^1(f^r)\Big)\Big| \le C\sum_{j=0}^{r-1} \frac{n}{b_n^2}\Big(\frac{n}{b_n}\big|U_n^1(f)\big|\Big)^j. \tag{3.54}$$

By applying the Bernstein-type inequality (3.28) to every term on the right-hand of (3.54), we see that we can neglect the last term in (3.53). This proves (3.52).

Combining (3.43), (3.44), (3.49), (3.50), and (3.51), it remains to show that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\Big(\frac{n}{b_n}\Big)^r \Pi_{F_{r,\sigma}}(L_n^{\otimes r}) \in E^F \setminus A_{\varepsilon/4}\Big) \le -l. \tag{3.55}$$

For this purpose we assume without loss of generality that there exists a joint $k \in \mathbb{N}$ such that every $\tilde\varphi_{r,\sigma}$ with $\varphi \in F$ used in (3.49) and described in Lemma 3.41 is of the form $\tilde\varphi_{r,\sigma} = \sum_{i=1}^k \beta_{\varphi,i} f_{\varphi,i}^{\otimes r}$ where $\beta_{\varphi,i} \in E$ and $f_{\varphi,i}\colon S \to \mathbb{R}$ is bounded and μ-degenerate for all $i \in \{1,\dots,k\}$ and $\varphi \in F$. Define the map $\psi\colon S \to (\mathbb{R}^k)^F$ by $\psi(s) = (f_{\varphi,1}(s), \dots, f_{\varphi,k}(s))_{\varphi\in F}$ and the map $T\colon (\mathbb{R}^k)^F \to E^F$ by $T((x_{\varphi,1},\dots,x_{\varphi,k})_{\varphi\in F}) = \big(\sum_{i=1}^k \beta_{\varphi,i}x_{\varphi,i}^r\big)_{\varphi\in F}$. Then, for every $n \in \mathbb{N}$,

$$\Big\{\Big(\frac{n}{b_n}\Big)^r \Pi_{F_{r,\sigma}}(L_n^{\otimes r}) \in E^F \setminus A_{\varepsilon/4}\Big\} = \Big\{\int_S \psi\, dM_n^{1,1} \in T^{-1}\big(E^F \setminus A_{\varepsilon/4}\big)\Big\}. \tag{3.56}$$

Since T is continuous, $T^{-1}(E^F \setminus A_{\varepsilon/4})$ is closed in $(\mathbb{R}^k)^F$ and the moderate deviations upper bound in the τ-topology on $\mathcal{M}(S)$ from [3, Theorem 3.1] implies that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(\int_S \psi\, dM_n^{1,1} \in T^{-1}\big(E^F \setminus A_{\varepsilon/4}\big)\Big) \le -\inf\Big\{ I_1(\tilde\nu) \;\Big|\; \tilde\nu \in \mathcal{M}(S), \int_S \psi\, d\tilde\nu \in T^{-1}\big(E^F \setminus A_{\varepsilon/4}\big)\Big\} \tag{3.57}$$

with I_1 as in (3.1). Note that $T(\int_S \psi\, d\tilde\nu) = \Pi_{F_{r,\sigma}}(\tilde\nu^{\otimes r})$ for all $\tilde\nu \in \mathcal{M}(S)$. Hence, to derive (3.55) from (3.56) and (3.57), if suffices to show that $\Pi_{F_{r,\sigma}}(\tilde\nu^{\otimes r}) \in A_{\varepsilon/4}$ for all $\tilde\nu \in K(I_1, l)$.

Consider any $\tilde\nu \in K(I_1, l)$. Define $\nu \in \mathcal{M}(S^m)$ by (2.18). According to (2.17) and (3.1), $I_{m,r}(\nu) = I_1(\tilde\nu)$, hence $\nu \in K(I_{m,r}, l) \subset \mathcal{M}^\Phi(S^m)$. For $\varphi \in \Phi$, using

(2.18) and the notation introduced before (2.14),

$$\int_{S^m} \varphi \, d\nu = \sum_{\substack{A,B\subset\{1,\dots,m\}\\ |A|=r}} \int_{S^m} \tilde{\varphi}_B((s_{\tau_A(i)})_{i\in B}) \, (\tilde{\nu}^{\otimes r} \otimes \mu^{\otimes m-r})(ds_1,\dots,ds_m). \tag{3.58}$$

Note that $\{\tau_A(i)\}_{i\in A} = \{1,\dots,r\}$. If there is an $i \in A \setminus B$, then the integral in (3.58) vanishes because $\tilde{\nu}(S) = 0$. If there is an $i \in B \setminus A$, then the integral vanishes because $\tilde{\varphi}_B$ is completely μ-degenerate. Therefore, $\int_{S^m} \varphi \, d\nu = \int_{S^m} \tilde{\varphi}_r \, d\tilde{\nu}^{\otimes r}$ by (2.6), hence $\Pi_F(\nu) = \Pi_{F_r}(\tilde{\nu}^{\otimes r})$. Using (3.1), we see that $I_r(\tilde{\nu}^{\otimes r}) = 2^{r-1}(I_1(\tilde{\nu}))^r \le 2^{r-1} l^r$. Using the Cauchy–Schwarz inequality,

$$\Big\| \int_{S^r} (\tilde{\varphi}_r - \tilde{\varphi}_{r,\sigma}) \, d\tilde{\nu}^{\otimes r} \Big\|_E \le \int_{S^r} \|\tilde{\varphi}_r - \tilde{\varphi}_{r,\sigma}\|_E \Big| \frac{d\tilde{\nu}^{\otimes r}}{d\mu^{\otimes r}} \Big| \, d\mu^{\otimes r} \le \sigma \sqrt{2 I_r(\tilde{\nu}^{\otimes r})} \le \sigma (2l)^{r/2},$$

hence $\|\Pi_F(\nu) - \Pi_{F_{r,\sigma}}(\tilde{\nu}^{\otimes r})\|_{E^F} \le \sigma |F| (2l)^{r/2} < \varepsilon/4$. This shows $\Pi_{F_{r,\sigma}}(\tilde{\nu}^{\otimes r}) \in A_{\varepsilon/4}$, which finishes the proof of the moderate deviations upper bound (2.22). □

References

1. A. de Acosta, *Moderate deviations and associated Laplace approximations for sums of independent random vectors*, Trans. Amer. Math. Soc. **329** (1992), no. 1, 357–375.
2. ———, *On large deviations of empirical measures in the τ-topology*, J. Appl. Probab. **31A** (1994), 41–47, (special volume).
3. ———, *Projective systems in large deviation theory II: some applications*, Probability in Banach Spaces, 9 (J. Hoffman-Jørgensen, J. Kuelbs, and M. B. Marcus, eds.), Progress in Probability, vol. 35, Birkhäuser, Boston, 1994, pp. 241–250.
4. ———, *Moderate deviations for empirical measures of Markov chains: lower bounds*, Ann. Probab. **25** (1997), no. 1, 259–284.
5. A. Araujo and E. Giné, *The Central Limit Theorem for Real and Banach Valued Random Variables*, John Wiley & Sons, New York, 1980.
6. M. A. Arcones, *Limits of canonical U-processes and B-valued U-statistics*, J. Theoret. Probab. **7** (1994), no. 2, 339–349.
7. M. A. Arcones and E. Giné, *On the bootstrap of U and V statistics*, Ann. Statist. **20** (1992), no. 2, 655–674.
8. ———, *Limit theorems for U-processes*, Ann. Probab. **21** (1993), no. 3, 1494–1542.
9. ———, *On the law of the iterated logarithm for canonical U-statistics and processes*, Stochastic Process. Appl. **58** (1995), 217–245.
10. A. A. Borovkov and A. A. Mogulskii, *Probabilities of large deviations in topological spaces II*, Siberian Math. J. **21** (1980), 653–664.
11. Yu. V. Borovskikh, *U-Statistics in Banach Spaces*, VSP, Utrecht, 1996.
12. J. Bretagnolle, *Lois limites du bootstrap de certaines fonctionnelles*, Ann. Inst. H. Poincaré, Sect. B: Calcul des Probabilités et Statistique **19** (1983), no. 3, 281–296.
13. P. M. Cohn, *Algebra*, second ed., vol. 1, John Wiley & Sons, Chichester, 1982.
14. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Springer-Verlag, New York, 1998.
15. 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.
16. P. Eichelsbacher, *Moderate deviations for degenerate U-processes*, Stochastic Process. Appl. **87** (2000), no. 2, 255–279.
17. P. Eichelsbacher and U. Schmock, *Large deviations of U-empirical measures in strong topologies and applications*, Ann. Inst. H. Poincaré. Probab. Statist. **38** (2002), no. 5, 779–797.
18. P. Eichelsbacher and T. Zajic, *Moderate deviations for mean field Gibbs measures*, Bernoulli **9** (2003), no. 1, 67–95.
19. P. Gänssler, *Compactness and sequential compactness in spaces of measures*, Z. Wahrscheinlichkeitstheorie verw. Geb. **17** (1971), 124–146.

20. E. Giné and V. H. de la Peña, *Decoupling: From Dependence to Independence*, Springer-Verlag, New York, 1999.
21. E. Giné and J. Zinn, *A remark on convergence in distribution of U-statistics*, Ann. Probab. **22** (1994), no. 1, 117–125.
22. W. Hoeffding, *A class of statistics with asyptotically normal distribution*, Ann. Mat. Statist. **19** (1948), 293–325.
23. _____, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.
24. J. L. Kelley, *General Topology*, D. Van Nostrand, New York, 1955.
25. M. Ledoux, *Sur les déviations modérées des sommes de variables aléatoires vectorielles indédendantes de même loi*, Ann. Inst. H. Poincaré. Probab. Statist. **28** (1992), no. 2, 267–280.
26. M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.
27. A. J. Lee, *U-Statistics: Theory and Practice*, Marcel Dekker, New York, 1990.
28. S. Montgomery-Smith and V. H. de la Peña, *Bounds for the tail probabilities of U-statistics and quadratic forms*, Bull. Amer. Math. Soc. **31** (1994), 223–227.
29. _____, *Decoupling inequalities for the tail probabilities of multivariate U-statistics*, Ann. Probab. **23** (1995), 806–816.
30. M. Rubin and R. A. Vitale, *Asymptotic distribution of symmetric statistics*, Ann. Statist. **8** (1980), no. 1, 165–170.
31. L. Wu, *Moderate deviations of dependent random variables related to CLT*, Ann. Probab. **23** (1995), no. 1, 420–445.

(P. Eichelsbacher) Fakultät für Mathematik, Ruhr-Universität Bochum, Gebäude NA 3/68, D-44780 Bochum, Germany
E-mail address: peter.eichelsbacher@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/