MODERATE DEVIATIONS FOR DEGENERATE U-PROCESSES

PETER EICHELSBACHER

Abstract. Sufficient conditions for a rank-dependent moderate deviations principle (MDP) for degenerate U-processes are presented. The MDP for VC classes of functions is obtained under exponential moments of the envelope. Among other techniques, randomization, decoupling inequalities and integrability of Gaussian and Rademacher chaos are used to present new Bernstein-type inequalities for U-processes which are the basis of our proofs of the MDP. We present a complete rank-dependent picture. The advantage of our approach is that we obtain in the degenerate case moderate deviations in non-Gaussian situations.

1. Introduction

Let $(S, \mathcal{S}, \mu)$ be a probability space and let $X_i : S^{\mathbb{N}} \to S$ be the coordinate functions ($\{X_i\}_{i\in\mathbb{N}}$ is thus an i.i.d. sequence with $\mathcal{L}(X_i) = \mu$). The *U-empirical measure of order m* is defined by

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

where $n_{(m)} := \prod_{k=0}^{m-1}(n-k)$ and $I_{m,n} \subset \{1,\dots,n\}^m$ contains all m-tuples with pairwise different components and δ_x denotes the probability measure degenerate at $x \in S$. Let $\mathcal{H}$ be a collection of measurable functions $h\colon S^m \to \mathbb{R}$. The *$U$-process* of order m indexed by $\mathcal{H}$ is defined for every integer $n \geq m$ as

$$U_n^m(h,\mu) := \int_{S^m} h\, dL_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\dots,i_m)\in I_{m,n}} h(X_{i_1},\dots,X_{i_m}), \quad h \in \mathcal{H}.$$

For a fixed h, the expression $U_n^m(h,\mu)$ is called a *U-statistic of order m* with kernel h based on the probability measure μ. U-processes appear in statistics frequently as unbiased estimators of the functional $\{\mu^{\otimes m} h\colon h \in \mathcal{H}\}$. For instance, Liu's simplicial depth process (Liu 1990) is a U-process. Noland and Pollard (1987), (1988) studied the law of large numbers and the central limit theorem for U-processes of order $m = 2$. Arcones and Giné (1993) developed the theory for an arbitrary $m > 1$. Regarding the law of large numbers they presented a necessary and sufficient condition for its validity in complete analogy with the results for empirical processes. The central limit theorem (CTL) and the law of iterated logarithm (LIL) are developed only for Vapnic-Červonenkis (VC) classes of sets and functions, because such conditions are unknown at present.

Wu (1994) proved necessary and sufficient conditions for the large deviation and moderate deviation estimations and the LIL of the empirical process $L_n(h) =$

1991 *Mathematics Subject Classification.* 60F10 (primary), 60B12, 60G17.

Key words and phrases. Rank-dependent moderate deviations, U-statistics, U-processes, VC classes, Bernstein-type inequality, metric entropy.

$1/n \sum_{i=1}^n h(X_i)$ with h varying in a class of uniformly bounded functions. The principles are proved for laws in the Banach space of bounded functionals on the class $\mathcal{H}$. Sufficient conditions for the large deviation principle (LDP) as well as for the moderate deviations principle (MDP) in the so-called non-degenerate case are proved in Eichelsbacher (1998). Precise definitions of LDP and MDP are given below.

The case of completely degenerate or canonical kernels is the crucial one for the MDP and in this paper the goal is to study the MDP for degenerate U-processes under certain conditions on $\mathcal{H}$. We develop the principle for VC classes. We are able to present a complete MDP ,,picture" for non-Gaussian cases.

Next we present some notation. Let us recall the definition of the MDP. A sequence of probability measures $\{\mu_n\}_{n\in\mathbb{N}}$ on a topological space $\mathcal{X}$ equipped with σ-field $\mathcal{B}$ is said to satisfy the LDP with speed $a_n \downarrow 0$ and good rate function $I(\cdot)$ if the level sets $\{x : I(x) \le \alpha\}$ are compact for all $\alpha < \infty$ and for all $\Gamma \in \mathcal{B}$ the lower bound

$$\liminf_{n\to\infty} a_n \log \mu_n(\Gamma) \ge - \inf_{x \in \mathrm{int}(\Gamma)} I(x),$$

and the upper bound

$$\limsup_{n\to\infty} a_n \log \mu_n(\Gamma) \le - \inf_{x \in \mathrm{cl}(\Gamma)} I(x)$$

hold, where $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively. We say that a sequence of random variables satisfies the LDP provided the sequence of measures induced by these variables satisfies the LDP. Let $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ be a sequence which satisfies

$$\lim_{n\to\infty} \frac{b_n}{n} = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \frac{n}{b_n^2} = 0. \tag{1.1}$$

If $\mathcal{X}$ is a topological vector space, then a sequence of random variables $\{Z_n\}_{n\in\mathbb{N}}$ satisfies the MDP with speed $\frac{n}{b_n^2}$ and with good rate function $I(\cdot)$ in case the sequence $\{\Phi\big(\frac{n}{b_n}\big) Z_n\}_{n\in\mathbb{N}}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$ and with speed $\frac{n}{b_n^2}$. Here $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex function. In our applications $\Phi(x) = x^r$ for some $r \in \mathbb{N}$.

Since we are using *Hoeffding's decomposition* of a U-statistic we state it here together with some notation. The operator $\pi_{k,m}^{\mu} = \pi_{k,m}$ (sometimes called Hoeffding projections) acts on $\mu^{\otimes m}$-integrable symmetric functions $h : S^m \to \mathbb{R}$ (symmetric in the sense that for all $x_1, \dots, x_m \in S$ and all permutations s of $\{1,\dots,m\}$ $h(x_1,\dots,x_m) = h(x_{s_1},\dots,x_{s_m})$) as follows:

$$\pi_{k,m} h(x_1,\dots,x_k) = (\delta_{x_1} - \mu) \otimes \cdots \otimes (\delta_{x_k} - \mu) \otimes \mu^{\otimes(m-k)} h,$$

where $\nu_1 \otimes \cdots \otimes \nu_m h = \int \cdots \int h(x_1,\dots,x_m) d\nu_1(x_1) \cdots d\nu_m(x_m)$ and

$$\mu^{\otimes(m-1)} h(x) = \int \cdots \int h(x_1,\dots,x_{m-1},x) d\mu(x_1) \cdots d\mu(x_{m-1}),$$

for $0 \le k \le m$. Note that $\pi_{0,m} h = \mu^{\otimes m} h$. A symmetric function h is called μ-*canonical* or *completely degenerate* if $\int h(x_1,\dots,x_m) d\mu(x_1) = 0$ for all $x_2,\dots,x_m \in S$. Note that $\pi_{k,m} h$ is a μ-canonical function of k variables. A $\mu^{\otimes m}$-integrable symmetric function $h : S^m \to \mathbb{R}$ is μ-degenerate of order $r-1$, $1 \le r \le m$, if

$$\int h(x_1,\dots,x_m)\, d\mu^{\otimes m-r+1}(x_r,\dots,x_m) = \int h\, d\mu^{\otimes m}$$

for all $x_1, \dots, x_{r-1} \in S$, whereas $\int h(x_1, \dots, x_m)\, d\mu^{\otimes m-r}(x_{r+1}, \dots, x_m)$ is not a constant function. If h is not degenerate of any positive order, we say it is *non-degenerate* or *degenerate of order zero*. We say that a U-process is μ-canonical if all the functions $h \in \mathcal{H}$ are μ-canonical. With this notation we can decompose a U-statistic with integrable h into a sum of μ-canonical U-statistics of different orders. For all $\mu^{\otimes m}$-integrable symmetric functions $h : S^m \to \mathbb{R}$ the following equation holds

$$U_n^m(h, \mu) = \sum_{k=0}^{m} \binom{m}{k} U_n^k(\pi_{k,m} h, \mu). \tag{1.2}$$

It is very easy to check that h is degenerate of order $r-1 \geq 0$ if and only if its Hoeffding expansion starts at term r, except for the constant term, that is,

$$U_n^m(h, \mu) - \mu^{\otimes m} h = \sum_{k=r}^{m} \binom{m}{k} U_n^k(\pi_{k,m} h, \mu).$$

Hoeffding's decomposition is a basic tool in the analysis of U-statistics.
For $n \geq m$ and $\{b_n\}_{n \in \mathbb{N}}$ as in (1.1) define the *moderate U-statistic of rank* $r \in \{1, \dots, m\}$ by

$$M_n^{m,r}(h, \mu) = \left(\frac{n}{b_n}\right)^r \sum_{k=r}^{m} \binom{m}{k} U_n^k(\pi_{k,m} h, \mu).$$

The *moderate U-process of rank* r indexed by $\mathcal{H}$ is defined by $\{M_n^{m,r}(h, \mu), h \in \mathcal{H}\}$. The goal of this paper is to establish sufficient conditions on the class $\mathcal{H}$ of functions to get LDP for $\{M_n^{m,r}(h, \mu), h \in \mathcal{H}\}$ for each $2 \leq r \leq m$. The case $r = 1$, e.g. the case where each $h \in \mathcal{H}$ is non-degenerate, was considered in Eichelsbacher (1998). Let $l_\infty(\mathcal{H})$ be the space of all bounded real functions on $\mathcal{H}$ with the supremum norm $\|H\|_{\mathcal{H}} := \sup_{h \in \mathcal{H}} |H(h)|$. This is, in general, a non-separable Banach space if $\mathcal{H}$ is infinite. The aim is to prove a LDP for $\{M_n^{m,r}(h, \mu), h \in \mathcal{H}\}$ in $l_\infty(\mathcal{H})$ under certain conditions on $\mathcal{H}$. The *non-separability* of $l_\infty(\mathcal{H})$ is the reason why such results do not follow easily via the contraction principle (cf. Dembo and Zeitouni (1998, Theorem 4.2.1)) from results given in Eichelsbacher and Schmock (1998).
Whereas the $\sqrt{n}$ CLT for not necessarily degenerate U-statistics was obtained by Hoeffding (1948), the CLT for degenerate U-statistics was obtained much later. The limit laws in the degenerate case are the laws of Gaussian chaos variables. It is known that the sequence $\left\{n^{m/2} U_n^m(h, \mu)\right\}$ converges in distribution if and only if $\mathbb{E}h^2$ is finite and h is canonical. A canonical U-process with class $\mathcal{H}$ of symmetric functions satisfies the CLT if the following conditions are fulfilled: for each $h \in \mathcal{H}$ denote by $K_{\mu,m}(h)$ an element of the chaos of order m associated to G_μ, the centered Gaussian process indexed by $L^2(S, \mu)$ with covariance $\mathbb{E}(G_\mu(f)\, G_\mu(g)) = \mu(f\, g) - (\mu f)(\mu g)$, $f, g \in L^2(S, \mu)$. Then a canonical U-process with class $\mathcal{H}$ satisfies the CLT if the process $\{K_{\mu,m}(h) : h \in \mathcal{H}\}$ has a version with bounded uniformly continuous paths in $(\mathcal{H}, e_{\mu,m})$, where $e_{\mu,m}(f, g) = \|f - g\|_{L^2(\mu^m)}$, and if

$$n^{m/2} U_m^n(h, \mu) \to K_{\mu,m} \circ h \quad \text{in} \quad l_\infty(\mathcal{H}), \tag{1.3}$$

where convergence in (1.3) is in the sense of Hoffmann-Jørgensen (1991).
We will obtain a LDP for the moderate U-process of rank r for every $r \in \{1, \dots, m\}$. This is a MDP for $\{U_n^m(h, \mu) - \mu^{\otimes m}, h \in \mathcal{H}\}$ where each $h \in \mathcal{H}$ is degenerate of order $r - 1$ and $\Phi(x) = x^r$. In Eichelsbacher and Schmock (1998) the MDP was checked for a finite collection $\mathcal{H}$. The advantage of our method is to obtain a MDP

in *non-Gaussian* situations. All known results on moderate deviations principles were established using Gaussian limit theorems (see, e.g. de Acosta (1992), Wu (1995)).

The context of the different sections is as follows. In Section 2 we consider the present situation of the MDP problem for U-statistics and U-empirical measures. We review facts on U-processes as well as on Vapnic-Červonenkis classes of functions. The main results are formulated afterwards. In Section 3 we prove some new Bernstein-type inequalities for U-processes. Therefore, techniques like *decoupling* and *randomization* offer efficient ways of proving these inequalities. In Section 4 the proofs of our main results are given. In Section 5 we discuss some improvements of Eichelsbacher (1998) for the non-degenerate case and in Section 6 we discuss some results for so-called V-processes.

2. Preliminaries and the main results

2.1. **Moderate deviations of U-empirical measures of rank $r \le m$.** To the best of our knowledge, the present situation of the MDP problem for U-statistics and U-empirical measures is as follows. Let $\mathcal{M}(S^m)$, $m \in \mathbb{N}$, denote the set of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$ with finite total variation. Given a sequence $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ satisfying (1.1) we define the moderate U-empirical measure $M_n^{m,1} : \Omega \to \mathcal{M}(S^m)$ by

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

for every $n \ge m$. If $m = 1$, then we also write M_n instead of $M_n^{1,1}$. Let $B(S^m)$ denote the space of bounded and measurable real valued functions on S^m. Let τ be the coarsest topology on $\mathcal{M}(S^m)$ such that $\mathcal{M}(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ is continuous for every $\varphi \in B(S^m)$. The large deviations of $\{M_n\}_{n\in\mathbb{N}}$ on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ have been studied by Borovkov and Mogulskii (1980, Section 3) in the τ-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 (1994, Section 3) generalized this result to a full LDP on the scale $\{b_n^2/n\}_{n\in\mathbb{N}}$ when $(S, \mathcal{S})$ is a general measurable space. Eichelsbacher and Schmock (1998) proved large deviations of the moderate U-empirical measures $\{M_n^{m,1}\}_{n\ge m}$ in stronger topologies, generated by a collection Φ of functions $\varphi : S^m \to \mathbb{R}$ satisfying appropriate moments conditions.

Eichelsbacher and Schmock (1998) introduced the following decomposition of a function $\varphi \in L_1(\mu^{\otimes m}, \mathbb{R})$. We need the decomposition, since we cannot consider only symmetric functions. This is because for $\mathcal{S} \ne \{\varnothing, S\}$ and $m \ge 2$, we are not 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 τ would loose the Hausdorff property. Given $\varphi \in L_1(\mu^{\otimes m}, \mathbb{R})$ and a nonempty subset A of $\{1, \ldots, m\}$, define $\varphi_A \in L_1(\mu^{\otimes |A|}, \mathbb{R})$ 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, $\varphi_\varnothing \equiv \int_{S^m} \varphi \, d\mu^{\otimes m} \in \mathbb{R}$. Furthermore, define $\tilde{\varphi}_A \in L_1(\mu^{\otimes |A|}, \mathbb{R})$ by

$$\tilde{\varphi}_A\big((s_i)_{i\in A}\big) = \sum_{B \subset A} (-1)^{|A\setminus B|} \varphi_B\big((s_i)_{i\in B}\big), \tag{2.2}$$

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

$$\int_{S^m} \varphi \, dL_n^m = \tilde{\varphi}_0 + \sum_{a=1}^{m} \int_{S^a} \tilde{\varphi}_a \, dL_n^a \tag{2.3}$$

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

$$\tilde{\varphi}_a \equiv \sum_{\substack{A \subset \{1,\ldots,m\} \\ |A|=a}} \tilde{\varphi}_A. \tag{2.4}$$

Note that every $\tilde{\varphi}_A$ with nonempty $A \subset \{1, \ldots, m\}$ is *completely μ-degenerate.* Moreover, the decomposition (2.3) is exactly (1.2) for every symmetric φ (see e.g. (Denker 1985, (1.2.3))).

We have already mentioned that the weak limits of appropriately scaled U-statistics depend on the rank of the kernel function Giné and de la Peña (1998, Chapter 4). A related phenomenon can be considered for the MDP. This leads to the introduction of *rank-dependent* moderate U-empirical measures, first given in Eichelsbacher and Schmock (1998): given $a \in \{1, \ldots, m\}$ and $A = \{i_1, \ldots, i_a\}$ 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 $\{1, \ldots, m\} \setminus A$. Define the permutation τ_A of $\{1, \ldots, m\}$ such that $\tau_A(i_j) = j$ for every $j \in \{1, \ldots, m\}$. Using τ_A, define the mapping $\gamma_{A,m} : \mathcal{M}(S^a) \to \mathcal{M}(S^m)$ by $\gamma_{A,m}(\nu) = (\nu \otimes \mu^{\otimes m-a})\pi_{\tau_A}^{-1}$ for all $\nu \in \mathcal{M}(S^a)$, where π_{τ_A} is defined by $\pi_{\tau_A}(s) = (s_{\tau_A(1)}, \ldots, s_{\tau_A(m)})$ for every $s = (s_1, \ldots, s_m) \in S^m$. The marginal measure of $\gamma_{A,m}(\nu)$, corresponding to the ordered indices contained in A, is then given by ν, all other one-component marginals equal μ. Related to (2.2), define $\tilde{\gamma}_{A,m} : \mathcal{M}_1(S^a) \to \mathcal{M}_1(S^m)$ by

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

where $\nu_{A,B}$ denotes the marginal $\nu_{j_1,\ldots,j_b}$ of $\nu \in \mathcal{M}_1(S^a)$ when $B = \{i_{j_1}, \ldots, i_{j_b}\}$ 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 $r \in \{1, \ldots, m\}$ by

$$M_n^{m,r} = \Big(\frac{n}{b_n}\Big)^r \Bigg(L_n^m - \mu^{\otimes m} - \sum_{\substack{A\subset\{1,\ldots,m\} \\ 1\leq|A|\leq r-1}} \tilde{\gamma}_{A,m}(L_n^{|A|})\Bigg). \tag{2.5}$$

If $r = 1$, then (2.5) reduces to (2.1). Using (2.2)–(2.4), it follows from these definitions that, for every $\varphi \in L_1(\mu^{\otimes m}, E)$,

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

which means that $M_n^{m,r}$ extracts from φ the components of higher rank.

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

$$I_{m,r}(\nu) = \frac{1}{2} \int_S \Big(\frac{d\tilde{\nu}}{d\mu} \Big)^2 \, d\mu \tag{2.6}$$

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

$$\nu = \sum_{\substack{A \subset \{1,\ldots,m\} \\ |A|=r}} \gamma_{A,m}(\tilde{\nu}^{\otimes r}); \tag{2.7}$$

and we define $I_{m,r}(\nu) = \infty$ otherwise. In the case $r = 1$, equation (2.7) 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 $\tilde{\nu}$. As considered in Eichelsbacher and Schmock (1998), $I_{m,1}$ is convex and for $r \geq 2$, the rate function $I_{m,r}$ is in general not convex.
As proved in Eichelsbacher and Schmock (1998) the following assertions hold for every $r \in \{1, \ldots, m\}$: The sequence $\{M_n^{m,r}, n \in \mathbb{N}\}$ satisfies a LDP on $\mathcal{M}(S^m)$ with respect to the τ-topology and rate function $I_{m,r}(\cdot)$ (thus $\{M_n^{m,r}/(n/b_n)^r, n \in \mathbb{N}\}$ satisfies the MDP with $\Phi(x) = x^r$).

Remark 2.8. We apply the results of Eichelsbacher and Schmock (1998) for real valued functions h in the τ-topology. Notice that in Eichelsbacher and Schmock (1998) the MDP result is proved for an even finer topology and for functions h which take values in a separable Banach space. To be more specific, let $B(S^m, E)$ denote the space of bounded functions on S^m, which take values in a separable real Banach space $(E, \|\cdot\|_E)$ with Borel σ-algebra $\mathcal{E}$ and are $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable. On $\mathcal{M}(S^m)$ consider the $\tau(E)$-topology which makes the maps $\mathcal{M}(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ continuous for all $\varphi \in B(S^m, E)$, where $\int$ denotes the Bochner integral. Let $\{b_n\}_{n\in\mathbb{N}}$ satisfy (1.1). Eichelsbacher and Schmock (1998) proved the following: If there exists a $p \in (1, 2]$ such that the Banach space E is of type p and if $\lim_{n\to\infty} n/b_n^p = 0$, we obtain the lower bound for $\{M_n^{m,r}\}$ with speed n/b_n^2 and rate $I_{m,r}(\cdot)$. If the Banach space E is of type 2, then we get the upper bound with the same speed and rate.

2.2. **U-processes, metric entropy, Vapnic-Červonenkis classes.** If E is a separable Banach space of type 2, Remark 2.8 and the contraction principle (cf. Dembo and Zeitouni (1998, Theorem 4.2.1)) gives for any $h = (h_1, \ldots, h_d) : S^m \to E^d$ with $h_i \in B(S^m, E)$ a LDP for $\{M_n^{m,r}(h), n \in \mathbb{N}\}$ with corresponding speeds and rate functions. Notice that the decomposition (2.3) coincides with the Hoeffding decomposition (1.2) when we start with a symmetric φ. Since the $\tilde{\varphi}_a$ are symmetric in their arguments, we can consider symmetric h_i without loss of generality. In non-parametric statistics we need some kind of uniform estimates of $M_n^{m,r}(h)$ over a not necessarily finite class of functions $\mathcal{H}$.
One is interested in the behaviour of $\|M_n^{m,r}(h)\|_{\mathcal{H}} := \sup_{h\in\mathcal{H}} |M_n^{m,r}(h)|$ for possibly uncountable families $\mathcal{H}$ of symmetric functions $h : S^m \to \mathbb{R}$. It is well known that if the class $\mathcal{H}$ of measurable functions is countable there are no measurability problems. Otherwise we will assume that $(S, \mathcal{S})$ is a separable measurable space and the class $\mathcal{H}$ is image admissible Suslin (see Section 3.5 in Giné and de la Peña (1998)). A consequence is that the class $\pi_{k,m}\mathcal{H} := \{\pi_{k,m}h : h \in \mathcal{H}\}$ is also image admissible Suslin and if $\mathcal{H}$ is image admissible Suslin, so is $\mathcal{H}'(\delta, d) := \{f - g : f, g \in \mathcal{H}, d(f, g) \leq \delta\}$ where d is e.g. the $L_2(\mu^{\otimes m})$ distance (for proofs see Dudley (1984)). For simplicity, image admissible Suslin classes of functions

over separable measurable spaces will simply be denoted as *measurable classes*. Quantities like $\sup_{h\in\mathcal{H}}\left|\sum \varepsilon_{i_1}\cdots\varepsilon_{i_m} h(X_{i_1},\ldots,X_{i_m})\right|$, as well as other sups that appear in this paper, now are measurable, where ε_i, $i\in\mathbb{N}$, are i.i.d. Rademacher i.e. $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. To ensure measurability, we will assume without further mention that random variables $\varepsilon_i, \varepsilon_i^{(j)}, X_i, X_i^{(j)}$, $j\le m, i\in\mathbb{N}$, are the coordinate functions of an infinite product probability measure. The ε variables are Rademacher variables and the X's are all have law μ.
Moreover, we assume that the class of functions $\mathcal{H}$ considered in this paper admit everywhere finite envelopes: $H(x_1,\ldots,x_m) := \sup_{h\in\mathcal{H}}|h(x_1,\ldots,x_m)| < \infty$, for all $x_i\in S$ and, likewise, $\sup_{h\in\mathcal{H}}|\pi_{k,m}h(x_1,\ldots,x_k)| < \infty$ for all $x_i\in S$ and $k\le m$. The functions in $\mathcal{H}$ can take values in a not necessarily separable Banach space E, instead of being just real valued; the only extra assumption to be made is that random variables $x^*(h(X_1,\ldots,X_m))$ be measurable for all x^* in the dual space E^* of E, for details see Giné and de la Peña (1998, Chapter 3).
We prove rank-dependent MDPs for U-processes indexed by a class $\mathcal{H}$ which satisfies some conditions which are given in terms of metric entropy. In particular we consider the case when $\mathcal{H}$ is a Vapnik-Červonenkis subgraph class of functions. To state the results we have to introduce some additional notation. Given a pseudo-metric space (T,d), the ε-covering number $N(\varepsilon,T,d)$ is defined as

$$N(\varepsilon,T,d) = \min\{n\in\mathbb{N}:\ \text{there exists a covering of } T \text{ by } n \text{ balls of } d\text{-radius } \le\varepsilon\}.$$

The metric entropy of (T,d) is the function $\log N(\varepsilon,T,d)$. We define $N_2(\varepsilon,\mathcal{H},\mu) := N(\varepsilon,\mathcal{H},||\cdot||_{L_2(\mu)})$. Some classes of functions satisfy a uniform bound in the entropy. Let $\mathcal{H}$ admit the envelope H, and let α be a real positive number. Then $\mathcal{H}$ is a *Vapnik-Červonenkis (VC) class of functions for* H *and* α or $\mathcal{H}$ is $VC(H,\alpha)$ if there exists a Lebesgue integrable function $\lambda:\ [0,\infty)\to[0,\infty)$ such that

$$\left(\log N_2(\tau(\mu(H^2))^{1/2},\mathcal{H},\mu)\right)^{\alpha/2} \le \lambda(\tau), \quad \tau\in\mathbb{R}_+, \tag{2.9}$$

for all probability measures μ such that $\mu(H^2)<\infty$. In fact, we may take $\alpha = m$. The family of VC classes of functions is large.

(a) VC classes of sets (Dudley (1984)) are $VC(1,\alpha)$ for all $\alpha>0$. Examples in $\mathbb{R}^d$ include classes of all rectangles, all ellipsoids, and all polyhedra of at most l sides (for any fixed l).
(b) A class of real functions $\mathcal{H}$ is *VC subgraph class* if the subgraphs of the functions in the class form a VC class of sets (subgraph of $h\in\mathcal{H}$: $\{(x,t)\in S^m\times\mathbb{R}:\ 0\le t\le h(x_1,\ldots,x_m) \text{ or } h(x_1,\ldots,x_m)\le t\le 0\}$). Any finite-dimensional vector space of functions (e.g., polynomials of bounded degree on $\mathbb{R}^d$) is a VC subgraph class. VC-subgraph classes of functions are $VC(H,\alpha)$ for all α provided $\mathcal{H}$ admits an everywhere finite measurable envelope H (see Pollard (1984)). Notice, that if $\mathcal{C}$ is a VC class of sets and q a real function on $\mathcal{C}$, then the class $\{1_C/q(C):\ C\in\mathcal{C}\}$ corresponding to a weighted empirical process is a VC subgraph class.
(c) If $\mathcal{H}$ is *Euclidean for an envelope* H, then it is $VC(H,\alpha)$ for all $\alpha>0$ (see Nolan and Pollard (1987)).

Remark that by Pollard (1984, Proposition II 2.5) one obtains for a VC subgraph class admitting an envelope $H\in L_2(\mu)$, that there are finite constants A and v such that, for each probability measure μ with $\mu(H^2)<\infty$,

$$N_2(\varepsilon,\mathcal{H},\mu) \le A(\mu(H^2)^{1/2}/\varepsilon)^2 v, \tag{2.10}$$

where v denotes the index of the class of subgraphs (see also (Giné and de la Peña 1998, Theorem 5.1.15)).

2.3. **The main results.** To state our main results we define a (rate) function $J^{\mathcal{H}}_{m,r} : l_\infty(\mathcal{H}) \to \mathbb{R}$ by

$$J^{\mathcal{H}}_{m,r}(H) := \inf\{I_{m,r}(\nu) : \nu \in \mathcal{M}(S^m) \quad \text{and} \quad \nu(\mathcal{H}) = H\}, \quad H \in l_\infty(\mathcal{H}), \quad (2.11)$$

where $I_{m,r}(\cdot)$ is given by (2.6). Here $\nu(\mathcal{H}) \in l_\infty(\mathcal{H})$ is given by

$$\nu(\mathcal{H})(h) = \nu(h) = \int_{S^m} h \, d\nu \quad \text{for all} \quad h \in \mathcal{H}.$$

We consider the following exponential moment condition for the envelope H of a class $\mathcal{H}$:

Condition 2.12 (Cramér condition for the envelope)**.** *Let H be the envelope of a class $\mathcal{H}$. There exists an $\alpha_H > 0$ such that*

$$a := \int_{S^m} \exp(\alpha_H |H|^2) \, d\mu^{\otimes m} < \infty.$$

Let $K(J^{\mathcal{H}}_{m,r}, l) = \{H \in l_\infty(\mathcal{H}) : J^{\mathcal{H}}_{m,r}(H) \leq l\}$. Now we can formulate the main result of the paper:

Theorem 2.13 (Moderate deviations of U-processes)**.**
Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\mathcal{H}$ is $VC(H,m)$. The following assertions hold for every $r \in \{1,\dots,m\}$: If $\pi_{k,m}\mathcal{H}$ has an envelope $\pi_{k,m}H$ which satisfies Cramér's condition 2.12 for every $k \in \{r, r+1, \dots, m\}$, then:

(a) *$K(J^{\mathcal{H}}_{m,r}, l) \subset l_\infty(\mathcal{H})$ is compact for every $l \in [0,\infty)$.*

(b)
$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(M_n^{m,r}(\mathcal{H}) \in B) \geq -J^{\mathcal{H}}_{m,r}(\mathrm{int}_{l_\infty(\mathcal{H})}(B))$$
for every measurable $B \subset l_\infty(\mathcal{H})$, where $\mathrm{int}_{l_\infty(\mathcal{H})}(B)$ denotes the interior of the set B with respect to the $\|\cdot\|_{\mathcal{H}}$-topology.

(c)
$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(M_n^{m,r}(\mathcal{H}) \in B) \leq -J^{\mathcal{H}}_{m,r}(\mathrm{cl}_{l_\infty(\mathcal{H})}(B))$$
for every measurable B, where $\mathrm{cl}_{l_\infty(\mathcal{H})}(B)$ denotes the closure of the set B with respect to the $\|\cdot\|_{\mathcal{H}}$-topology.

We show that the result is equivalent to the following:

Theorem 2.14 (Moderate deviations of U-processes)**.**
Assume as in Theorem 2.13. Then the following assertions hold for every $r \in \{1,\dots,m\}$:

(a) *$K(J^{\mathcal{H}}_{m,r}, l) \subset l_\infty(\mathcal{H})$ is compact for every $l \in [0,\infty)$.*

(b)
$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\left(\frac{n}{b_n}\right)^r \binom{m}{r} U_n^r(\pi_{r,m}\mathcal{H}, \mu) \in B\right) \geq - J^{\mathcal{H}}_{m,r}(\mathrm{int}_{l_\infty(\mathcal{H})}(B))$$
for every measurable $B \subset l_\infty(\mathcal{H})$, where $\mathrm{int}_{l_\infty(\mathcal{H})}(B)$ denotes the interior of the set B with respect to the $\|\cdot\|_{\mathcal{H}}$-topology.

(c)

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Bigg(\left(\frac{n}{b_n}\right)^r \binom{m}{r} U_n^r(\pi_{r,m}\mathcal{H},\mu) \in B \Bigg) \le - J_{m,r}^{\mathcal{H}}(\mathrm{cl}_{l_\infty(\mathcal{H})}(B))$$

for every measurable B, where $\mathrm{cl}_{l_\infty(\mathcal{H})}(B)$ denotes the closure of the set B with respect to the $\|\cdot\|_{\mathcal{H}}$-topology.

Remark 2.15. In the case $r = 1$ our Theorem is an improvement of Theorem 3.9 in Eichelsbacher (1998). We discuss this in Section 5.

Corollary 2.16. *Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\mathcal{H}$ is $VC(1,m)$. Then the three assertions in Theorem 2.13 as well as in Theorem 2.14 hold for every $r \in \{1,\dots,m\}$.*

Remark 2.17. As already mentioned the family of VC classes of functions contains VC classes of sets, VC subgraph classes and Euclidean classes. Most statistical applications involve these nice classes of sets and functions. We only mention the simplicial depth process, empirical distribution functions with a U-statistic structure or classes of uniform Hölder functions. For further examples and applications see (Arcones and Giné 1993) and (Giné and de la Peña 1998).

3. Bernstein-typ inequalities

First we will state the Bernstein-type inequality for U-processes proved in Arcones and Giné (1994) (we state the formulation of (Giné and de la Peña 1998, Theorem 5.3.14)). We say that the class $\mathcal{H}$ is uniformly bounded if there is $M < \infty$ such that $\|h\|_\infty \le M$ for all $h \in \mathcal{H}$. There is then no loss of generality to assume that $\mathcal{H}$ consists only of functions h such that $0 \le h \le 1$. The uniformly bounded assumption is often desirable, from the point of view of statistics, in which case $\mathcal{H}$ has good properties independently of the underlying probability, in which case $\mathcal{H}$ has to be uniformly bounded (see Dudley (1984)). However, the extension to the unbounded functionals is also discussed.
Regarding notation, in proofs we write $\mathbf{i}$ for $(i_1,\dots,i_m)$, $\varepsilon_{\mathbf{i}}$ for $\varepsilon_{i_1}\cdots\varepsilon_{i_m}$, $\varepsilon_{\mathbf{i}}^{\mathrm{dec}}$ for $\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}$, $h(\mathbf{X_i})$ for $h(X_{i_1},\dots,X_{i_m})$, and $h(\mathbf{X}_{\mathbf{i}}^{\mathrm{dec}})$ for $h(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)})$ ("dec" standing for "decoupled").
For any further discussion we need the following observation. If $\mathcal{H}$ is a $VC(H,m)$ class of measurable functions, then $\pi_{k,m}\mathcal{H}$, $k \in \{1,\dots,m\}$ is a $VC(\pi_{k,m}H,k)$ class of μ-canonical functions. By Arcones and Giné (1995, Remark 4.8) we know that $\pi_{k,m}\mathcal{H}$ is a VC class of functions. Applying Giné and de la Peña (1998, Lemma 5.3.4 and Lemma 5.3.5) we see that the envelope is $\pi_{k,m}H$ and that k is the right parameter. We omit the details.

Proposition 3.1 (Arcones, Giné)**.** *Let $\mathcal{H}$ be a uniformly bounded $VC(1,m)$ class of measurable functions $h: S^m \to \mathbb{R}$, symmetric in their entries. Then, for each $k \in \{1,\dots,m\}$ there exist constants c_k and d_k such that, for all μ on $(S,\mathcal{S})$, $t>0$ and $n \ge m$,*

$$\mathbb{P}\big(\big\|n^{k/2}U_n^k(\pi_{k,m}h,\mu)\big\|_{\mathcal{H}} \ge t\big) \le c_k \exp\big(-d_k\, t^{2/k}\big). \tag{3.2}$$

For a proof see Arcones and Giné (1994, Theorem 3.2) or Giné and de la Peña (1998, Theorem 5.3.14).

Analyzing the proof of Proposition 3.1 as well as the proof of Arcones and Giné (1993, Proposition 2.3(c)) (see also Giné and de la Peña (1998, Theorem 4.1.12)) it is easy to observe the following inequalities which are basic in what follows.

Lemma 3.3. *Let $\mathcal{H}$ be a measurable class of μ-canonical functions $h : S^m \to \mathbb{R}$, symmetric in their entries. Then there are constants c and c', depending on m only such that for all μ on $(S, \mathcal{S})$, $t > 0$, $\lambda > 0$ and $n \geq m$*

$$\mathbb{P}\big(\big\|n^{m/2} U_n^m(h,\mu)\big\|_{\mathcal{H}} \geq t\big) \leq c \exp\big(-\lambda\, t^{1/m}\big)\, \mathbb{E}\big(\exp(c'\, \lambda^2 M^{2/m})\big). \tag{3.4}$$

Here M is defined by

$$M = \mathbb{E}_\varepsilon \big\| n^{-m/2} \sum_{I_{m,n}} \varepsilon_{i_1} \cdots \varepsilon_{i_m}\, h(X_{i_1}, \dots, X_{i_m}) \big\|_{\mathcal{H}}. \tag{3.5}$$

Moreover there are constants c and c', depending only on m, such that for all μ on $(S, \mathcal{S})$, $t > 0$, $\lambda > 0$ and $n \geq m$

$$\mathbb{P}\big(\big\|n^{m/2} U_n^m(h,\mu)\big\|_{\mathcal{H}} \geq t\big) \leq c \exp\big(-\lambda\, t^{2/(m+1)}\big)\, \mathbb{E}\big(\exp(c'\, \lambda^{m+1} M')\big), \tag{3.6}$$

where M' is given by

$$M' = \mathbb{E}_\varepsilon \big\| n^{-m} \sum_{I_{m,n}} \varepsilon_{i_1} \cdots \varepsilon_{i_m}\, h(X_{i_1}, \dots, X_{i_m}) \big\|_{\mathcal{H}}^2. \tag{3.7}$$

We give the proof in detail. Let (T, d) be a pseudo-metric space. A *Rademacher chaos process* of order m is a stochastic process of the form

$$\big\{X_t = \sum_{(i_1, \dots, i_m) \in I_{m,n}} a_{i_1, \dots, i_m}(t)\ \varepsilon_{i_1} \cdots \varepsilon_{i_m}\big\}_{t \in T}$$

with the functions $a_{i_1,\dots,i_m}(t) \in \mathbb{R}$, symmetric in the indices, and $\{\varepsilon_i\}$ a Rademacher sequence. Assume that $\|X\| := \sup_{t \in T} |X_t| < \infty$ a.s. It is well known (see e.g. Giné and de la Peña (1998, Corollary 3.2.7)) that for every $m \in \mathbb{N}$ and $0 < \alpha < 2/m$, there exist finite positive constants c_1, c_2, depending only on m, such that

$$\mathbb{E}\big(\exp(t\, \|X\|^\alpha)\big) \leq c_1 \exp\big(c_2 (\sigma^\alpha\, t)^{1/(1-\alpha m/2)}\big), \quad t > 0, \tag{3.8}$$

where $\sigma := \big(\mathbb{E}\|X\|^2\big)^{1/2}$. Moreover, we apply a metric entropy bound for sub-Gaussian processes (see e.g. Proposition 2.6 in Arcones and Giné (1993)):

$$\mathbb{E} \sup_{s,t \in T} |X_s - X_t| \leq K_m \int_0^D \big(\log N(\varepsilon, T, \varrho)\big)^{m/2}\, d\varepsilon, \tag{3.9}$$

where $\varrho(s,t) = \big(\mathbb{E}|X_t - X_s|^2\big)^{1/2}$, $D = \sup_{s,t \in T} \varrho(s,t)$, and K_m is a universal constant.

We need a convex modification Ψ of $\exp(x^\alpha)$, $\alpha \leq 1$, $x > 0$. Note that $\exp(x^\alpha)$ is only convex in the range $x^\alpha \geq (1-\alpha)/\alpha$. Replace $y = \exp(x^\alpha)$ by $y = \exp((1-\alpha)/\alpha)$ for $0 < x^\alpha \leq (1-\alpha)/\alpha$. Hence, Ψ satisfies

$$\exp(|x|^\alpha) \leq \Psi(|x|) \leq a_\alpha\, \Psi(|x|) \tag{3.10}$$

with $a_\alpha = \exp((1-\alpha)/\alpha)$.

Proof of Lemma 3.3*:* Constants are denoted by c and c', which depend on m and may vary from line to line. Without loss of generality we will assume that $0 \in \mathcal{H}$.

We consider $\exp(x^{1/m})$ and its convex modification Ψ. We can apply the decoupling inequality (de la Peña (1992)) and (3.10) to obtain

$$\mathbb{E}\Big[\exp\Big(\lambda\big\|n^{-m/2}\sum_{I_{m,n}} h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^{1/m}\Big)\Big] \leq c\,\mathbb{E}\Big[\Psi\Big(c'\,\lambda^m\big\|n^{-m/2}\sum_{I_{m,n}} h(X_{\mathbf{i}}^{\mathrm{dec}})\big\|_{\mathcal{H}}\Big)\Big].$$

Applying symmetrization (sometimes called a randomization inequality, contained in de la Peña (1992), see also Giné and de la Peña (1998, Theorem 3.5.3)), which uses the completely μ-degeneracy of h, the decoupling inequality (Giné and de la Peña 1998, Theorem 3.5.3) as well as (3.10), it follows that

$$\mathbb{E}\Big[\Psi\Big(c'\,\lambda^m\big\|n^{-m/2}\sum_{I_{m,n}} h(X_{\mathbf{i}}^{\mathrm{dec}})\big\|_{\mathcal{H}}\Big)\Big] \leq c\mathbb{E}\Big[\exp\Big(c'\lambda\big\|n^{-m/2}\sum_{I_{m,n}} \varepsilon_{\mathbf{i}} h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^{1/m}\Big)\Big].$$

Applying (3.8) with $t = 1$ and $\alpha = 1/m$, we obtain

$$\mathbb{E}\Big[\exp\Big(\lambda\big\|n^{-m/2}\sum_{I_{m,n}} h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^{1/m}\Big)\Big] \leq$$
$$c\,\mathbb{E}\Big[\exp\Big(c'\,\lambda^2\Big(\mathbb{E}_\varepsilon\big\|n^{-m/2}\sum_{I_{m,n}} \varepsilon_{\mathbf{i}} h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^2\Big)^{1/m}\Big)\Big],$$

letting $\mathbb{E}_\varepsilon$ denote integration with respect to the Rademacher variables only. With Borell's inequality for a Rademacher chaos process (see Borell (1979)), which is

$$\big(\mathbb{E}\|X\|^p\big)^{1/p} \leq \big((p-1)/(q-1)\big)^{m/2}\big(\mathbb{E}\|X\|^q\big)^{1/q} \quad \text{for all} \quad 1 < q < p < \infty,$$

it follows that $\big(\mathbb{E}\|X\|^2\big)^{1/2} \leq 2^{3m/2}\,\mathbb{E}\|X\|$. Equation (3.4) follows by Markov's inequality applied to the function $e^{\lambda\|\cdot\|_{\mathcal{H}}^{1/m}}$.

Proving (3.6) we take $t = 1$ and $\alpha = 2/(m+2)$ in (3.8) and apply the decoupling and symmetrization inequality to the corresponding convex modification, to conclude

$$\mathbb{E}\Big[\exp\Big(\lambda\big\|n^{-m/2}\sum_{I_{m,n}} h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^{2/(m+1)}\Big)\Big] \leq$$
$$c\,\mathbb{E}\Big[\exp\Big(c'\,\lambda^{m+1} n^{-m}\mathbb{E}_\varepsilon\big\|\sum_{I_{m,n}} \varepsilon_{\mathbf{i}}\, h(X_{\mathbf{i}})\big\|_{\mathcal{H}}^2\Big)\Big].$$

Equation (3.6) follows by Markov's inequality applied to the function $e^{\lambda\|\cdot\|_{\mathcal{H}}^{2/(m+1)}}$.

□

Remark 3.11. An immediate consequence of the proof of Lemma 3.3 is that we can replace M and M', respectively, by

$$M = \mathbb{E}_\varepsilon\big\|n^{-m/2}\sum_{I_{m,n}} \varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}\, h(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)})\big\|_{\mathcal{H}}$$

and

$$M' = \mathbb{E}_\varepsilon\big\|n^{-m}\sum_{I_{m,n}} \varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}\, h(X_{i_1}^{(1)},\dots,X_{i_m}^{(m)})\big\|_{\mathcal{H}}^2.$$

We use the fact, that a decoupled Rademacher process is a particular case of a Rademacher process.

Remark 3.12. If $\mathcal{H}$ is $VC(H,m)$, then there are constants c and c' such that for all $t>0$, $\lambda>0$, and $n\ge m$ we have

$$\mathbb{P}\big(\big\|n^{m/2}U_n^m(h,\mu)\big\|_{\mathcal{H}}\ge t\big)\le c\,\exp\big(-\lambda\,t^{1/m}\big)\,\mathbb{E}\big(\exp(c'\,\lambda^2(M'')^{2/m})\big).$$

Here M'' is defined as follows. Let $\mu_n h := n^{-m}\sum_{i_1,\dots,i_m=1}^n h(X_{i_1},\dots,X_{i_m})$ and let $D^2:=\sup_{f,g\in\mathcal{H}}\mu_n\big(|f-g|^2\big)$. Then

$$M''=\int_0^D\big(\log N_2(\tau,\mathcal{H},\mu_n)\big)^{m/2}\,d\tau.$$

This follows immediately from (3.4) in Lemma 3.3 and (3.9).

Remark 3.13. If $\mathcal{H}$ is uniformly bounded and $VC(H,m)$, then M'' in Remark 3.12 is bounded uniformly in n. The statement of Lemma 3.1 follows choosing $\lambda = t^{1/m}/2c'$.

The next inequality is basic in the proof of the rank dependent MDP for degenerate U-processes over VC-classes. Given a pseudo-metric e on $\mathcal{H}$ and $\eta>0$, we set

$$\mathcal{H}'(\eta,e):=\{h_2-h_1:\ h_1,h_2\in\mathcal{H}, e(h_1,h_2)\le\eta\}.$$

Lemma 3.14. *Let $\mathcal{H}$ be a uniformly bounded measurable $VC(1,m)$ class of functions $h:\ S^m\to\mathbb{R}$, symmetric in their entries. Then, for each $k\in\{1,\dots,m\}$ there exist constants c_k,d_k and f_k, depending only on m, such that, for all μ on $(S,\mathcal{S})$, $t>0$ and all $\eta>0$ there exists a number $N(\eta)\in\mathbb{N}$ such that for all $n\ge\max(m,N(\eta))$*

$$\mathbb{P}\big(\big\|n^{k/2}U_n^k(\pi_{k,m}h,\mu)\big\|_{\mathcal{H}'(\eta,e_k)}\ge t\big)\le c_k\,\exp\Big(-d_k\,\big(t/\kappa(2\,\eta^2)\big)^{2/k}\Big), \qquad (3.15)$$

where $e_k(h_1,h_2):=\big(\mu^{\otimes k}(\pi_{k,m}h_1-\pi_{k,m}h_2)^2\big)^{1/2}$ and $\kappa(\eta^2):=f_k\int_0^{\eta\,f_k}\lambda(\varepsilon)\,d\varepsilon$ for a suitable constant f_k and for the function λ in (2.9).

For the proof of Lemma 3.14 we apply the following basic estimate, given in Arcones and Giné (1995, Lemma 2.2). Consider the Young function of α-exponential type, $0<\alpha\le 2$. For $\alpha\ge 1$ let $\psi_\alpha(x):=\exp(x^\alpha)-1$ for $x\ge 0$. For $\alpha<1$ define τ_α to coincide with $\exp(x^\alpha)$ for $x\ge((1-\alpha)/\alpha)^{1/\alpha}$, and to coincide with the tangent to this function at $((1-\alpha)/\alpha)^{1/\alpha}, \exp((1-\alpha)/\alpha)$ for $0\le x<((1-\alpha)/\alpha)^{1/\alpha}$; then define $\psi_\alpha(x)=\tau_\alpha(x)-\tau_\alpha(0)=\tau_\alpha(x)-\alpha\,\exp((1-\alpha)/\alpha)$. The ψ_α-norm of a random variable X is defined as $\|X\|_{\psi_\alpha}:=\inf\{c>0:\ \mathbb{E}\big(\psi_\alpha(|X|/c)\big)\le 1\}$.
Let $\mathcal{H}$ be a measurable $VC(H,m)$ class of functions $h:\ S^m\to\mathbb{R}$. Letting

$$\tilde e(h_1,h_2):=\mathbb{E}_\varepsilon\Bigg(\sum_{I_{m,n}}\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}\,(h_1-h_2)(x_{i_1}^{(1)},\dots,x_{i_m}^{(m)})\Bigg)^2$$

for $h_1,h_2\in\mathcal{H}$ and an arbitrary collection $\{x_i^{(j)}\}$ of points in S, we have (see Arcones and Giné (1995, Lemma 2.2))

$$\Bigg\|\Bigg\|\sum_{I_{m,n}}\varepsilon_{i_1}^{(1)}\cdots\varepsilon_{i_m}^{(m)}\,h(x_{i_1}^{(1)},\dots,x_{i_m}^{(m)})\Bigg\|_{\mathcal{H}(\delta,\tilde e)}\Bigg\|_{\psi_{2/m}}\le$$
$$\Bigg(\sum_{I_{m,n}}H^2(x_{i_1}^{(1)},\dots,x_{i_m}^{(m)})\Bigg)^{1/2}\kappa\Bigg(\delta^2/\sum_{I_{m,n}}H^2(x_{i_1}^{(1)},\dots,x_{i_m}^{(m)})\Bigg), \qquad (3.16)$$

where $\kappa(\cdot)$ is defined as in Lemma 3.14. Moreover, there is a finite constant c such that for all finite collections $\{x_i^{(j)}\}$ of points in S we obtain

$$\Big\|\Big\|\sum_{I_{m,n}} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_m}^{(m)}\, h(x_{i_1}^{(1)}, \ldots , x_{i_m}^{(m)})\Big\|_{\mathcal{H}}\Big\|_{\psi_{2/m}} \leq c\left(\sum_{I_{m,n}} H^2(x_{i_1}^{(1)}, \ldots , x_{i_m}^{(m)})\right)^{1/2}. \tag{3.17}$$

Note that for all $\alpha > 0, p > 0$, there are finite constants such that for all random variables X, $\|X\|_p \leq c_{\alpha,p}\, \|X\|_{\psi_\alpha}$.

Proof of Lemma 3.14*:* Since $\mathcal{H}$ is a $VC(1,m)$ class, we know that $\pi_{k,m}\mathcal{H}$, $k \in \{1, \ldots , m\}$ is a $VC(1, k)$ class. For every fixed $k \in \{1, \ldots , m\}$ we apply (3.4) in Lemma 3.3 with $\mathcal{H}$ replaced by $\mathcal{H}'(\eta, e_k)$. Hence we have to estimate $\mathbb{E}\big(\exp(\lambda^2(M_{\eta,k})^{2/k})\big)$ with $M_{\eta,k}$ defined by

$$M_{\eta,k} = \frac{1}{n^{k/2}}\mathbb{E}_\varepsilon \big\| \sum_{I_{k,n}} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_k}^{(k)}\, \pi_{k,m} h(X_{i_1}^{(1)}, \ldots , X_{i_k}^{(k)})\big\|_{\mathcal{H}'(\eta,e_2)}.$$

For each $n \in \mathbb{N}$ let

$$\begin{aligned} e_{k,n}^2(h_1, h_2) &:= \frac{1}{n^k}\mathbb{E}_\varepsilon\left(\sum_{I_{k,n}} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_k}^{(k)}\, (\pi_{k,m}h_1 - \pi_{k,m}h_2)(X_{i_1}^{(1)}, \ldots , X_{i_k}^{(k)})\right)^2 \\ &= \frac{1}{n^k}\sum_{I_{k,n}} (\pi_{k,m}h_1 - \pi_{k,m}h_2)^2(X_{i_1}^{(1)}, \ldots , X_{i_k}^{(k)}). \end{aligned}$$

By Arcones and Giné (1993, Theorem 3.1 and Corollary 3.3), the law of large numbers for U-processes with a $VC(1, k)$ class, we obtain that

$$\lim_{n\to\infty} \|e_{k,n}^2(h_1, h_2) - e_k^2(h_1 - h_2)\|_{\mathcal{H}\times\mathcal{H}} = 0 \quad \text{a.s.}$$

For each $\eta > 0$ let $A_{\eta,k} := \{\omega : \|e_{k,n}^2(h_1, h_2) - e_k^2(h_1, h_2)\|_{\mathcal{H}\times\mathcal{H}} \leq \eta^2\}$. Since for linear functionals Φ, $\|\Phi\|_{\mathcal{H}'(\eta,e)} \leq 2\|\Phi\|_{\mathcal{H}}$, equation (3.17) gives for the uniformly bounded case that $M_{\eta,k} \leq 2\,M \leq 2\,c$, where M is as in (3.5) (m replaced by k) and c is the constant from (3.17) (m replaced by k, h by $\pi_{k,m}h$ and H by $\pi_{k,m}H$) which depends on k and $\pi_{k,m}\mathcal{H}$ only. Therefore, for every $\lambda > 0$ and $\eta > 0$ we can choose a number $N(\eta) \in \mathbb{N}$ such that for all $n \geq \max(m, N(\eta))$

$$\mathbb{E}\Big(1_{A_{\eta,k}^c}\, \exp\big(\lambda^2(M_{\eta,k})^{2/k}\big)\Big) = 0.$$

Since for all ω in the set $A_{\eta,k}$ we obtain $\mathcal{H}'(\eta, e_k) \subseteq \mathcal{H}'(\sqrt{2}\,\eta, e_{k,n})$ for all $n \geq m$, the previous argument reduces the estimation of $\mathbb{E}\big(\exp(\lambda^2(M_{\eta,k})^{2/k})\big)$ to bounding $\mathbb{E}\big(\exp(\lambda^2(M'_{\eta,k})^{2/k})\big)$, where $M'_{\eta,k}$ is defined as

$$M'_{\eta,k} = \frac{1}{n^{m/2}}\mathbb{E}_\varepsilon \big\| \sum_{I_{k,n}} \varepsilon_{i_1}^{(1)} \cdots \varepsilon_{i_k}^{(k)}\, \pi_{k,m} h(X_{i_1}^{(1)}, \ldots , X_{i_k}^{(k)})\big\|_{\mathcal{H}'(\sqrt{2}\,\eta,e_{k,n})}.$$

Inequality (3.16) and the uniformly boundedness give

$$\big(M'_{\eta,k}\big)^{2/k} \leq \big(\kappa(2\,\eta^2)\big)^{2/k} = \left(f_k \int_0^{f_k\,\sqrt{2}\,\eta} \lambda(\varepsilon)\, d\varepsilon\right)^{2/k}.$$

Replacing these estimates in (3.4) gives for every $\eta > 0$, $\lambda > 0$ and $n \geq \max(m, N(\eta))$

$$\mathbb{P}\big(\big\|n^{k/2} U_n^k(\pi_{k,m} h, \mu)\big\|_{\mathcal{H}'(\eta, e_k)} \geq t\big) \leq c_k \exp\big(-\lambda\, t^{1/k}\big) \; \exp\Big(c_2\, \lambda^2 \big(\kappa(2\,\eta^2)\big)^{2/k}\Big).$$

Now the result is obtained by minimizing the exponent with respect to λ, which gives $\lambda = t^{1/k}/(2 c_2 (\kappa(\eta^2))^{2/k})$. □

Finally we prove a Bernstein-type inequality for a class $\mathcal{H}$ of possibly unbounded functions.

Lemma 3.18. *Let $\mathcal{H}$ be a measurable class of μ-canonical measurable functions $h : S^m \to \mathbb{R}$, symmetric in their entries, which is $VC(H, m)$ and H satisfies the Cramér condition 2.12. We set $\sigma^2 = \mathbb{E} H^2(X_1, \ldots, X_m)$. Then there exist constants c_i, $i = 1, 2$, and F_H such that*

$$\mathbb{P}\big(\big\|n^{m/2} U_n^m(h, \mu)\big\|_{\mathcal{H}} \geq t\big) \leq c_1 \, \exp\Big(-\frac{(t/\sigma)^{2/m}}{c_2 + F_H (t^{1/m}\, n^{-1/2})^{2/(m+1)}}\Big). \tag{3.19}$$

Proof. Using (3.6) we estimate $\mathbb{E}\big(\exp(c'\,\lambda^{m+1} M')\big)$. As in the proof of Lemma 3.3 by Borell's inequality for Rademacher chaos processes we obtain that $\big(\mathbb{E}\|X\|^2\big) \leq 2^{3m} \big(\mathbb{E}\|X\|\big)^2$. Therefore, we get

$$M' \leq c\, n^{-m} \Big(\mathbb{E}_\varepsilon \big\| \sum_{I_{m,n}} \varepsilon_{\mathbf{i}}^{\mathrm{dec}}\, h(X_{\mathbf{i}}^{\mathrm{dec}}) \big\|_{\mathcal{H}}\Big)^2.$$

Now since $\mathcal{H}$ is assumed to be $VC(H, m)$, we can apply (3.17) to bound $\|\cdot\|_1$ by the estimates for $\|\cdot\|_{\Psi_{2/m}}$ and conclude that

$$\mathbb{E}\big(\exp(c'\,\lambda^{m+1} M')\big) \leq \mathbb{E}\Big(\exp\Big(c\,\lambda^{m+1} \frac{1}{n^m} \sum_{I_{m,n}} H^2(X_{i_1}, \ldots, X_{i_m})\Big)\Big).$$

Now we can proceed exactly as in Arcones and Giné (1993, Proof of Proposition 2.3(c), Inequality (2.6)): The duplication of the power of n from $n^{-m/2}$ to n^{-m} allows us to use an average procedure which goes back to Hoeffding (1963) and reduces the problem to one of sums of independent centered random variables, which can be handled by Bernstein's inequality in integral form. For bounded H the proof was given in Arcones and Giné (1993) (as well as in Giné and de la Peña (1998)); if H satisfies the Cramér condition, the proof was given in Eichelsbacher and Schmock (1998). □

4. Proof of the main result

In this section we give the proofs of our main results:

Proof of Theorems 2.13 *and* 2.14*: The case of uniformly bounded $\mathcal{H}$*:
First we will prove the rank dependent MDP for the case where $\mathcal{H}$ is uniformly bounded. The reason for this is that we obtain easily the compactness of the level sets of the rate function in $l_\infty(\mathcal{H})$ in this case. To make an approximation method work which uses finite subsets (nets) of $\mathcal{H}$, the compactness is quite fundamental. The unbounded case is considered thereafter using the exponential equivalence concept in large deviation theory (see Dembo and Zeitouni (1998, Section 4.2)).

Without loss of generality we may assume that $\mathcal{H}$ is a $VC(1,m)$ class. Again we use the fact that the class of functions $\pi_{k,m}\mathcal{H}$ is a measurable $VC(1,k)$ class of symmetric μ-canonical functions for every $k \in \{1,\dots,m\}$. Remember that a $VC(1,m)$ class $\mathcal{H}$ is Donsker and with (Ledoux and Talagrand 1991, Theorem 14.6) $\mathcal{H}$ is *totally bounded* with respect to e_2.

1.Step: In the first step we will check that Theorem 2.13 and Theorem 2.14 are equivalent. Assume that $\mathcal{H}$ is $VC(1,m)$ and that we consider the case of *rank* r, $r \in \{1,\dots,m\}$ fixed. Therefore, it suffices to prove that for every $\delta > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\left\| M_n^{m,r}(h) - \left(\frac{n}{b_n}\right)^r \binom{m}{r} U_n^r(\pi_{r,m}\, h, \mu) \right\|_{\mathcal{H}} \geq \delta \right) = -\infty \qquad (4.1)$$

(see (Dembo and Zeitouni 1998, Theorem 4.2.13), the concept of exponential equivalence). With $\mathcal{H}$ being uniformly bounded and $VC(1,m)$ every class $\pi_{k,m}\mathcal{H}$ is uniformly bounded and $VC(1,k)$ for every $k \in \{1,\dots,m\}$. Thus we can apply Lemma 3.1. With (3.2) in Lemma 3.1 we get for each $r+1 \leq k \leq m$:

$$\begin{aligned}\mathbb{P}\left(\left\| \left(\frac{n}{b_n}\right)^r U_n^k(\pi_{k,m}h, \mu) \right\|_{\mathcal{H}} \geq \delta \right) &= \mathbb{P}\left(\left\| n^{k/2} U_n^k(\pi_{k,m}h, \mu) \right\|_{\mathcal{H}} \geq \frac{\delta\, n^{k/2}\, b_n^r}{n^r} \right) \\ &\leq c_k \exp\left(-c_k' \left(\frac{\delta^{2/k}\, b_n^{2r/k}}{n^{(2r/k)-1}} \right) \right).\end{aligned}$$

This implies

$$\frac{n}{b_n^2} \log \mathbb{P}\left(\left\| \left(\frac{n}{b_n}\right)^r U_n^k(\pi_{k,m}h, \mu) \right\|_{\mathcal{H}} \geq \delta \right) \leq \frac{n}{b_n^2} \log c_k - c_k' \delta^{2/k} \frac{n^{2-(2r/k)}}{b_n^{2-(2r/k)}}.$$

Since $k \in \{r+1, r+2, \dots, m\}$, the right hand side decreases to $-\infty$ by the assumptions for $\{b_n\}_{n\in\mathbb{N}}$. Since

$$\begin{aligned}&\left\{ \left\| \left(\frac{n}{b_n}\right)^r \sum_{k=r+1}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h, \mu) \right\|_{\mathcal{H}} \geq \delta \right\} \subset \\ &\qquad \bigcup_{k=r+1}^{m} \left\{ \left(\frac{n}{b_n}\right)^r \left\| \binom{m}{k} U_n^k(\pi_{k,m}h, \mu) \right\|_{\mathcal{H}} \geq \frac{\delta}{m-r} \right\},\end{aligned}$$

Dembo and Zeitouni (1998, Lemma 1.2.15) yields (4.1). Hence what follows is the prove of Theorem 2.14.

2. Step: Notice that for finite $\mathcal{H}$ the result is a direct consequence of Theorem 1.17 in Eichelsbacher and Schmock (1998) applying the classical contraction principle (Dembo and Zeitouni 1998, Theorem 4.2.1). Moreover, since this result is true even in a finer topology, we obtain the statement of the theorem for any finite collection $\mathcal{H}$ of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable functions $h : S^m \to E$, where E is a separable real Banach space of type 2 and each h satisfies the condition

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

for every $\alpha > 0$ as well as there exist $\alpha_k > 0$ for $k \in \{r+1, r+2, \dots, m\}$ such that

$$\int_{S^k} \exp(\alpha_k ||\pi_{k,m}h||_E^2)\, d\mu^{\otimes k} < \infty$$

for each such k.

3. Step: The *level sets* $K(J^{\mathcal{H}}_{m,r}, L) := \{H \in l_\infty(\mathcal{H}) : J^{\mathcal{H}}_{m,r}(H) \leq L\}, L \geq 0$, are compact in $l_\infty(\mathcal{H})$ by Arcelà-Ascoli (Dunford and Schwartz 1967, Theorem 5, Section IV.6). Therefore, we have to check that each level set is a bounded subset of $\mathcal{C}_b(\mathcal{H}, e_2)$, the set of all bounded and continuous functionals on $(\mathcal{H}, e_2)$ and that $(\mathcal{H}, e_2)$ is totally bounded (which holds, since $\mathcal{H}$ is a $VC(1,m)$-class). For every H with $J^{\mathcal{H}}_{m,r}(H) < \infty$ there exists a signed measure $\mu \in \mathcal{M}(S^m)$ such that $H(h) = \int_{S^m} h \, d\mu$ for all $h \in \mathcal{H}$. Hence H is a bounded and continuous functional on $(\mathcal{H}, e_2)$ and $K(J^{\mathcal{H}}_{m,r}, L)$ is a subset of this space. Using the uniform boundedness of $\mathcal{H}$ we obtain $\|H\|_{\mathcal{H}} \leq 1$, i.e. $K(J^{\mathcal{H}}_{m,r}, L)$ is a bounded subset of $\mathcal{C}_b(\mathcal{H}, e_2)$. Therefore, $K(J^{\mathcal{H}}_{m,r}, L)$ is compact in $l_\infty(\mathcal{H})$ for every $L > 0$.

4. Step: We show, that the MDP is established once we have

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\left\| \left(\frac{n}{b_n}\right)^r \binom{m}{r} U_n^r(\pi_{r,m}\, h, \mu) \right\|_{\mathcal{H}'(\eta, e_2)} \geq \delta \right) = -\infty \tag{4.2}$$

for all $\delta > 0$. Similar to the proof of Eichelsbacher (1998, Theorem 2.4 and 2.6) we deduce the upper and lower bounds from (4.2) as follows:

Proof of the upper bound:
For every $B \subset l_\infty(\mathcal{H})$ define $J^{\mathcal{H}}_{m,r}(B) = \inf_{H \in B} J^{\mathcal{H}}_{m,r}(H)$. Let C be an arbitrary closed set in $l_\infty(\mathcal{H})$. It suffices to consider the case $J^{\mathcal{H}}_{m,r}(C) > 0$. Choose $r \in (0, J^{\mathcal{H}}_{m,r}(C))$. For every element H in $l_\infty(\mathcal{H}) \setminus C$ there exists a $\delta_H > 0$ such that

$$U(H, 2\delta_H) := \{G \in l_\infty(\mathcal{H}) : \; ||G - H||_{\mathcal{H}} < 2\delta_H\}$$

is a subset of $l_\infty(\mathcal{H}) \setminus C$. Since the level set $K(J^{\mathcal{H}}_{m,r}, r)$ is compact and $C \cap K(J^{\mathcal{H}}_{m,r}, r) = \emptyset$, there exists a finite subset N of $K(J^{\mathcal{H}}_{m,r}, r)$ such that $U := \bigcup_{H \in N} U(H, \delta_H)$ covers $K(J^{\mathcal{H}}_{m,r}, r)$. We abbreviate

$$O_n^{m,r}(h) := \left(\frac{n}{b_n}\right)^r \binom{m}{r} U_n^r(\pi_{r,m}\, h, \mu)$$

and arrive at

$$\mathbb{P}\big((O_n^{m,r})(\mathcal{H}) \in C\big) \leq \mathbb{P}\Big((O_n^{m,r})(\mathcal{H}) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(H, 2\delta_H)\Big).$$

For a fixed $\eta > 0$ let $\mathcal{H}^\eta$ denote a finite η-net of $\mathcal{H}$: $\mathcal{H}^\eta \subset \mathcal{H}$ and for all $h \in \mathcal{H}$ there exists a $g \in \mathcal{H}^\eta$ such that $d_2(h, g) < \eta$. Denote by $J_\eta(\cdot)$ (which actually depends also on m and r) the rate function corresponding to $\mathcal{H}^\eta$ defined by (2.11) and by $p_\eta(H)$ the restriction of $H \in l_\infty(\mathcal{H})$ to the net $\mathcal{H}^\eta$. Now for every $\delta > 0$ one can find a $\eta' > 0$ such that if $d_2(h, g) < \eta'$ we get $|H(f) - H(g)| < \delta$ for each $H \in N$, since on the compact level set the H are continuous. Thus for every $\delta > 0$ there is an η with $0 < \eta < \eta'$ such that for all $F \in l_\infty(\mathcal{H})$, if

$$p_\eta(F) \in \bigcup_{H \in N} U(p_\eta(H), \delta_H)$$

and

$$\sup\{|F(h) - F(g)| : \; h, g \in \mathcal{H} \quad \text{and} \quad d_2(h, g) < \eta\} < \delta,$$

then $F \in \bigcup_{H \in N} U(H, 2\delta_H)$. Hence we obtain

$$\begin{aligned}\mathbb{P}\big((O_n^{m,r})(\mathcal{H}) \in C\big) \leq \mathbb{P}\Big((O_n^{m,r})(\mathcal{H}^\eta) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H)\Big) \\ + \mathbb{P}\big(||(O_n^{m,r})(h)||_{\mathcal{H}'(\eta,e_2)} \geq \delta\big).\end{aligned} \tag{4.3}$$

By Theorem 1.17 in Eichelsbacher and Schmock (1998) we have

$$\begin{aligned}\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big((O_n^{m,r})(\mathcal{H}^\eta) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H)\Big) \leq \\ -\inf\Big\{J_\eta(p_\eta(F)) : \, p_\eta(F) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H)\Big\} \leq -r.\end{aligned}$$

Using (4.2) for the second term in (4.3) we get the upper bound.

Proof of the lower bound:
Let O be an arbitrary open set in $l_\infty(\mathcal{H})$ and let $G \in O$ with $J_{m,r}^{\mathcal{H}}(O) < \infty$. There exists a δ_G such that $U(G, \delta_G)$ is contained in O. As in the proof of the upper bound for δ_G there exists an $\eta > 0$ such that

$$\begin{aligned}\mathbb{P}\big((O_n^{m,r})(\mathcal{H}^\eta) \in U(p_\eta(G), \delta_G)\big) \leq \\ \mathbb{P}\big(||O_n^{m,r}(h)||_{\mathcal{H}'(\eta,e_2)} \geq \delta_G\big) + \mathbb{P}\big((O_n^{m,r})(\mathcal{H}) \in O\big).\end{aligned}$$

We apply the lower bound of Theorem 1.17, Eichelsbacher and Schmock (1998), to the finite subnet $\mathcal{H}^\eta$:

$$-J_\eta(G) \leq -J_\eta(U(p_\eta(G), \delta_G)) \leq \liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\big((O_n^{m,r})(\mathcal{H}^\eta) \in U(p_\eta(G), \delta_G)\big).$$

Using the fact that the map $\nu \mapsto \nu(\mathcal{H}^\eta)$ is τ-continuous for each finite net $\mathcal{H}^\eta$, we obtain by Dembo and Zeitouni (1998, Lemma 4.1.6 (a)) that

$$\lim_{\eta\to 0} J_\eta(G) = \lim_{\eta \to 0} \inf\{I_{m,r}(\nu) : \, \nu \in \mathcal{M}(S^m), \nu(\mathcal{H}^\eta) = G\} = J_{m,r}^{\mathcal{H}}(G),$$

thus together with (4.2) we get the desired lower bound.

Step 5.: Proof of (4.2).
To prove the key estimation (4.2), we will apply Lemma 3.14. The conditions of the Lemma are fulfilled, since the class $\pi_{m,r}(\mathcal{H})$ is $VC(1,r)$ (especially uniformly bounded by 1). With $c(m,r) := \binom{m}{r}$ we obtain that for all $n \geq \max(m, N(\eta))$

$$\begin{aligned}\mathbb{P}\bigg(\Big\|\big(\frac{n}{b_n}\big)^r c(m,r)\, U_n^r(\pi_{r,m}\, h, \mu)\Big\|_{\mathcal{H}'(\eta,e_2)} \geq \delta\bigg) = \\ \mathbb{P}\bigg(\Big\|\big(n^{r/2} U_n^r(\pi_{r,m}\, h, \mu)\Big\|_{\mathcal{H}'(\eta,e_2)} \geq \delta \frac{n^{r/2}\, b_n^r}{c(m,r)\, n^r}\bigg) \leq \\ c\, \exp\bigg(-c'\bigg(\frac{\delta^{2/r}\, b_n^2}{n^2 c(m,r)^{2/r}\, \big(\kappa(\eta^2)\big)^{2/r}}\bigg)\bigg).\end{aligned}$$

This implies

$$\frac{n}{b_n^2}\log \mathbb{P}\Big(\big\| \big(\frac{n}{b_n}\big)^r \, c(m,r)\, U_n^r(\pi_{r,m}h,\mu)\big\|_{\mathcal{H}'(\eta,e_2)} \geq \delta \Big) \leq \frac{n}{b_n^2}\log c - \frac{c'\,\delta^{2/r}}{c(m,r)^{2/r}\,\big(\kappa(\eta^2)\big)^{2/r}}.$$

By definition $\kappa(\eta) = c\int_0^{c\,\eta}\lambda(\varepsilon)\,d\varepsilon$ (c a suitable constant), the right hand side decreases to $-\infty$ when first letting $n \to \infty$ (using the assumptions for $\{b_n\}_{n\in\mathbb{N}}$) and after that letting $\eta \to 0$. Hence the theorem is proved for a uniformly bounded $VC(H,m)$ class.

The unbounded case:
Now we assume that $\mathcal{H}$ is $VC(H,m)$ and that $\pi_{r,m}H$ satisfies condition 2.12 if r is fixed. $\pi_{k,m}\mathcal{H}$ is $VC(\pi_{k,m}H,k)$ for every $k \in \{1,\dots,m\}$. Theorem 2.13 and Theorem 2.14 are equivalent: we prove (4.1) applying Lemma 3.18. We obtain for each $r+1 \leq k \leq m$:

$$\mathbb{P}\Big(\big\| n^{k/2} U_n^k(\pi_{k,m}h,\mu)\big\|_{\mathcal{H}} \geq \frac{\delta\, n^{k/2}\, b_n^r}{n^r} \Big) \leq c_1 \exp\Big(- \frac{\delta^{2/k}\, b_n^{2r/k}\, n^{1-(2r/k)} 1/\sigma^{2/k}}{c_2 + F_{\pi_{k,m}H}\big(\delta^{1/k} b_n^{r/k}/(n^{r/k})\big)^{2/(k+1)}} \Big).$$

This implies

$$\frac{n}{b_n^2}\log \mathbb{P}\Big(\big\| \big(\frac{n}{b_n}\big)^r U_n^k(\pi_{k,m}h,\mu)\big\|_{\mathcal{H}} \geq \delta \Big) \leq \frac{n}{b_n^2}\log c_1 - c_2\delta^{2/k} \frac{n^{2-(2r/k)}/b_n^{2-(2r/k)}\,\sigma^{2/k}}{c_2 + F_{\pi_{k,m}H}\big(\delta^{1/k} b_n^{r/k}/(n^{r/k})\big)^{2/(k+1)}}.$$

Since $k \in \{r+1, r+2, \dots, m\}$ the right hand side decreases to $-\infty$ by the assumptions for $\{b_n\}_{n\in\mathbb{N}}$. Hence as in the 1.Step (4.1) follows in the unbounded case.
We prove Theorem 2.14 by a truncation argument. Consider $h\,1_{\{H\leq M\}}$ for a fixed $M>0$ and every $h \in \mathcal{H}$. Now it suffices to prove that for every $\delta > 0$

$$\lim_{M\to\infty}\limsup_{n\to\infty} \frac{n}{b_n^2}\log \mathbb{P}\bigg(\bigg\| \big(\frac{n}{b_n}\big)^r \binom{m}{r} U_n^r(\pi_{r,m}\,(h\,1_{\{H>M\}}),\mu)\bigg\|_{\mathcal{H}} \geq \delta \bigg) = -\infty \tag{4.4}$$

(see Dembo and Zeitouni (1998, Theorem 4.2.16 and Remark 4.2.16 (b)), the concept of exponentially good approximations). We apply Lemma 3.18 and obtain

$$\frac{n}{b_n^2}\log \mathbb{P}\Big(\big\| \big(\frac{n}{b_n}\big)^r U_n^r(\pi_{r,m}(h\,1_{\{H>M\}}),\mu)\big\|_{\mathcal{H}} \geq \delta \Big) \leq \frac{n}{b_n^2}\log c_1 - \frac{\delta^{2/k}}{|H\,1_{\{H>M\}}|^r_{L_2(\mu^{\otimes r})}\,\big(c_2 + F_{\pi_{r,m}H}\delta^{2/k(k+1)}\big)}.$$

Since the right hand side decreases to $-\infty$ when first letting $n \to \infty$ (using the assumptions for $\{b_n\}_{n\in\mathbb{N}}$) and after that letting $M \to \infty$, the theorem is proved. $\square$

5. The non-degenerate case

Let us consider the non-degenerate case i.e., when $r = 1$ in Theorem 2.13. Write $e_1(h,g) := \big(\mu(\pi_{1,m}h - \pi_{1,m}g)^2\big)^{1/2}$ (see Lemma 3.14) and

$$H(n,\eta) = \mathbb{E}\Big\|\sum_{k=1}^{n} \varepsilon_k \ \pi_{1,m}\, h(X_k)\Big\|_{\mathcal{H}'(\eta,e_1)}$$

Applying (Wu 1994, Theorem 2) we obtain the following improvement of (Eichelsbacher 1998, Theorem 3.9):

Theorem 5.1 (MDP of U-processes, the non-degenerate case)**.**
Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\mathcal{H}$ is $VC(1,m)$. Then the following three assertions hold and are equivalent:

(a) *The sequence $\{M_n^{m,1}(\mathcal{H}), n \in N\}$ satisfies a LDP in $l_\infty(\mathcal{H})$ with speed n/b_n^2 and good rate function $J^{\mathcal{H}}_{m,1}(\cdot)$, defined in 2.11.*

(b) *$(\pi_{1,m}\mathcal{H}, e_1)$ is totally bounded and*

$$\lim_{\eta\to 0}\lim_{n\to\infty} \frac{H(n,\eta)}{b_n} = 0.$$

(c) *$(\pi_{1,m}\mathcal{H}, e_1)$ is totally bounded and $M_n^{m,1}(\mathcal{H}) \to 0$ in probability in $l_\infty(\mathcal{H})$.*

Assume that the class $\mathcal{H}$ is $VC(H,m)$ and $\pi_{k,m}\mathcal{H}$ has an envelope $\pi_{k,m}H$ which satisfies Cramér's condition 2.12 for every $k \in \{1,\dots,m\}$, then, the three assertions are equivalent, too.

Proof. The first assertion is Theorem 2.13 for $r = 1$. By Theorem 2.14 we know that this is equivalent to the statement that the sequence $\{(n/b_n^2)\, m\, (U_n^1(\pi_{1,m}(\mathcal{H}) - \mu^{\otimes m}\mu), n \in N\}$ satisfies a LDP in $l_\infty(\mathcal{H})$ with speed n/b_n^2 and good rate function $J^{\mathcal{H}}_{m,1}(\cdot)$, defined in 2.11. Apart the factor m, this is the MDP for an empirical process stated in Wu (1994, Theorem 2) and the same Theorem proves the equivalence of the three assertions. ☐

Remark 5.2. Apart from the equivalence given in Theorem 5.1 we improve Theorem 3.9 in Eichelsbacher (1998) in dealing with the unbounded case. Notice that on the MDP scale there is no need to apply Talagrand's isoperimetric inequalities for empirical processes (see Talagrand (1994, Theorem 3.5)) used in the proof of Theorem 3.9 in Eichelsbacher (1998). These quite sharp and very general results are the effective tool to obtain the *necessary and sufficient* conditions considered in (Wu 1994). However, if the class of functions $\mathcal{H}$ is VC, the Bernstein type inequalities presented in Section 3 and in (Arcones and Giné 1994) are sharp enough. Best possible bounds for degenerate U-processes in general seem to be out of reach at present.

6. V-processes

Finally, we consider the MDP for V-processes $\{V_n^m(h,\mu) = L_n^{\otimes m}(h) : h \in \mathcal{H}\}$. For a fixed h V-statistics $V_n^m(h,\mu)$ appear in the Taylor-von Mises development of smooth statistics. For $m = 2$ we obtain the following result:

Theorem 6.1 (Moderate deviations of V-processes, $m = 2$)**.**
Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\mathcal{H}$ is $VC(1,2)$. Then the same assertions as in Theorem 2.14 hold when $U_n^r(\pi_{r,2}h,\mu)$ is replaced by $V_n^r(\pi_{r,2}h,\mu)$.

Remark 6.2. Fix $r \in \{1,2\}$. Assume that there exists a $\alpha_{r,m} > 0$ such that

$$\int_{S^r} \exp\big(\alpha_{r,m} |\pi_{r,m} H \circ \pi_\tau|^2\big)\, d\mu^{\otimes r} < \infty$$

for every map $\tau : \{1,\dots,r\} \to \{1,\dots,r\}$, where $\pi_\tau : S^r \to S^r$ is defined by $\pi_\tau(s) = (s_{\tau(1)},\dots,s_{\tau(r)})$ for every $s = (s_1,\dots,s_r) \in S^r$. Then for every fixed $r \in \{1,2\}$ the three assertions of Theorem 2.14 hold true for a $VC(H,2)$ class when $U_n^r(\pi_{r,2}h,\mu)$ is replaced by $V_n^r(\pi_{r,2}h,\mu)$. We omit the details.

Proof of Theorem 6.1*:* It suffices to check that for every $\delta > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^2 \big\| L_n^{\otimes 2}(\pi_{2,2}\, h) - L_n^2(\pi_{2,2}\, h)\big\|_{\mathcal{H}} \geq \delta\bigg) = -\infty.$$

Using

$$L_n^{\otimes 2}(\pi_{2,2}\, h) = \frac{1}{n^2}\sum_{j=1}^n \pi_{2,2}\, h(X_j,X_j) + \frac{n_{(2)}}{n^2}\, U_n^2(\pi_{2,2}\, h,\mu),$$

we apply twice Proposition 3.1 and observe

$$\mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^2 \Big\|\Big(\frac{n_{(2)}}{n^2}-1\Big) U_n^2(\pi_{2,2}\, h,\mu)\Big\|_{\mathcal{H}} \geq \delta/2\bigg) \leq c\, \exp\bigg(-d\, \frac{\delta\, n\, b_n^2}{2(n^2-(n)_2)}\bigg)$$

and

$$\mathbb{P}\bigg(\Big\|\frac{1}{b_n^2}\sum_{j=1}^n \pi_{2,2}\, h(X_j,X_j)\Big\|_{\mathcal{H}} \geq \delta/2\bigg) \leq c'\, \exp\bigg(-d'\, \frac{\delta^2\, b_n^4}{4\, n}\bigg),$$

where c, c' and d, d' are some constants. Thus the Theorem is proved. □

We do not consider $m = 2$ only for notational convenience. Actually in the case $m \geq 3$ the Bernstein type inequalities we developed in Section 3 are not sharp enough to get

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^a \big\| L_n^{\otimes a}(\pi_{a,m}\, h) - L_n^a(\pi_{a,m}\, h)\big\|_{\mathcal{H}} \geq \delta\bigg) = -\infty \tag{6.3}$$

for every $\delta > 0$ and for every $a \in \{s, s+1,\dots,m\}$, where $s \in \mathbb{N}$ is the rank: We use the well known decomposition

$$L_n^{\otimes a}(\pi_{a,m}\, h) = \frac{1}{n^a}\sum_{j=1}^a j!S_a^{(j)} \binom{n}{j} U_n^j(h_{a,j},\mu).$$

Here $h_{a,j}$ is a function on S^j defined by

$$h_{a,j}(x_1,\dots,x_j) := \frac{1}{j!S_a^{(j)}} \sum_{(j)}^{*} \pi_{a,m} h(x_{i_1},\dots,x_{i_a}),$$

and the sum $\sum_{(j)}^*$ is taken over all a-tuples $(i_1,\dots,i_a)$ formed from $\{1,2,\dots,j\}$ having *exactly* j indices distinct, where the quantities $S_a^{(j)}$ are Stirling numbers of the second kind (see Lee (1990, Chapter 4.2, Theorem 1)). Applying Proposition 3.1 it follows easily that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^a \Big\|\Big(\frac{n_{(a)}}{n^a}-1\Big) U_n^a(\pi_{a,m}\, h,\mu)\Big\|_{\mathcal{H}} \geq \delta\bigg) = -\infty.$$

In general the diagonals $h_{a,j}$ $(1 \leq j \leq a-1)$ are not μ-canonical functions. Assume that $h_{a,j}$ is degenerate of order $t-1 \geq 0$ with $1 \leq t \leq j$ and denote by h_t the kernel function of the first summand of the Hoeffding decomposition (1.2) of $h_{a,j}$ (h_t is a function in t variables and actually depends on a and j). We consider with Proposition 3.1 that

$$\frac{n}{b_n^2}\log\mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^a \big\|U_n^t(h_t,\mu)\big\|_{\mathcal{H}} \geq \frac{\delta\, n^a}{n^j}\bigg) \leq -\delta^{2/t}\frac{b_n^{(2a/t)-2}}{n^{(2j/t)-2}}.$$

But the right hand side decreases to $-\infty$ with $n\to\infty$ only for special sequences $(b_n)_{n\in\mathbb{N}}$.

For a special class of functions we obtain a result for every $m\geq 2$:

Theorem 6.4. *Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\pi_{a,m}\mathcal{H}$ is $VC(1,a)$. Assume that for every $h\in\pi_{a,m}\mathcal{H}$ there exist a bounded measurable function $f: S\to\mathbb{R}$ with $\int_S f\,d\mu = 0$ such that $h(x_1,\ldots,x_a) = f^{\otimes a}(x_1,\ldots,x_a) := \prod_{j=1}^a f(x_j)$ for all $x_1,\ldots,x_a\in S$. Then*

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\bigg(\Big(\frac{n}{b_n}\Big)^a \big\|U_n^a(f^{\otimes a},\mu) - V_n^a(f^{\otimes a},\mu)\big\|_{\mathcal{H}} \geq \delta\bigg) = -\infty.$$

Proof. For one function $f^{\otimes a}$ the proof was given in (Eichelsbacher and Schmock 1998). Using Newton's formula we obtain for every $h\in\pi_{a,m}\mathcal{H}$ that

$$\begin{aligned}\Big(\frac{n}{b_n}\Big)^a\big(U_n^a(f^{\otimes a}) - V_n^a(f^{\otimes a})\big) &= \Big(\frac{n}{b_n}\Big)^a\Big(1-\frac{n_{(a)}}{n^a}\Big)U_n^a(f^{\otimes a})\\ &\quad + R_a\Big(\frac{n}{b_n}U_n^1(f), \frac{n}{b_n^2}U_n^1(f^2),\ldots,\frac{n}{b_n^a}U_n^1(f^a)\Big).\end{aligned} \tag{6.5}$$

Here $R_a(p_1,\ldots,p_a)$ is a polynomial and every monomial of $R_a(p_1,\ldots,p_a)$ is of the form const. $p_1^{k_1}\ldots p_a^{k_a}$ with const. $\in\mathbb{Z}$ and $k_1,\ldots,k_a\in\mathbb{N}_0$ satisfying $\sum_{j=1}^a jk_j = a$ and $k_1\leq a-1$ (see Giné and de la Peña (1998, Section 4.2)). Since $f^{\otimes a}$ is bounded and completely μ-degenerate and $1-n_{(a)}/n^a\to 0$ as $n\to\infty$, it follows from the Bernstein-type inequality in Proposition 3.1 that we can neglect the first term on the right-hand side of (6.5). Every monomial of the last term of (6.5) contains at least one factor of the form $nb_n^{-k}U_n^1(f^k)$ with $k\in\{2,3,\ldots,a\}$. Since f is bounded, $\lim_{n\to\infty} b_n = \infty$ and $\lim_{n\to\infty} n/b_n^2 = 0$ by (1.1), it follows that there exists a constant $C>0$ such that, for all $n\geq a$,

$$\Big|R_a\Big(\frac{n}{b_n}U_n^1(f),\ldots,\frac{n}{b_n^a}U_n^1(f^a)\Big)\Big| \leq C\sum_{j=0}^{a-1}\frac{n}{b_n^2}\Big(\frac{n}{b_n}\big|U_n^1(f)\big|\Big)^j. \tag{6.6}$$

By applying the Bernstein-type inequality in Proposition 3.1 to every term on the right-hand of (6.6), we see that we can neglect the last term in (6.5). This proves the Theorem. ☐

References

Arcones, M. A. and E. Giné (1993). Limit theorems for U-processes. *Ann. Probab. 21*, 1494–1542.

Arcones, M. A. and E. Giné (1994). U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Process. Appl. 52*, 17–38.

Arcones, M. A. and E. Giné (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Stochastic Process. Appl. 58*, 217–245.

Borell, C. (1979). On the integrability of Banach space valued walsh polynomials. In *Séminaire de Probabilités XIII*, Volume 721 of *Lecture Notes in Mathematics*, pp. 1–3. Springer-Verlag, New York.

Borovkov, A. and A. Mogulskii (1980). Probabilities of large deviations in topological spaces II. *Siberian Math. J. 21*, 653–664.

de Acosta, A. (1992). Moderate deviations and associated Laplace approximations for sums of independent random vectors. *Trans. Amer. Math. Soc. 329*(1), 357–375.

De Acosta, A. (1994). Projective systems in large deviation theory II: some applications. In J. Hoffman-Jørgensen, J. Kuelbs, and M. Marcus (Eds.), *Probability in Banach Spaces, 9*, Volume 35 of *Progress in Probability*, pp. 241–250. Birkhäuser, Boston.

de la Peña, V. (1992). Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab. 20*, 1877–1892.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications*. New York: Springer.

Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*. Braunschweig: Vieweg.

Dudley, R. M. (1984). A course on empirical processes. In A. Dold and B. Eckmann (Eds.), *École d'Eté de Probabilités de Saint-Flour XII. Lecture Notes in Math. 1097*, pp. 1–142. Springer, New York.

Dunford, N. and J. Schwartz (1967). *Linear Operators, Part I*. New York: Interscience Publishers, Inc.

Eichelsbacher, P. (1998). Moderate and large deviations for U-processes. *Stochastic Process. Appl. 74*(2), 273–296.

Eichelsbacher, P. and U. Schmock (1998). Large and rank-dependent moderate deviations for U-empirical measures in strong topologies. preprint.

Giné, E. and V. de la Peña (1998). *Decoupling: from dependence to independence*. Springer-Verlag.

Hoeffding, W. (1948). A class of statistics with asyptotically normal distribution. *Ann. Mat. Statist. 19*, 293–325.

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

Hoffmann-Jørgensen, J. (1991). *Stochastic Processes on Polish Spaces*. Aarhus, Denmark: Various Publication Series 39, Aarhus Universitet.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces*. Berlin: Springer-Verlag.

Lee, A. (1990). *U-Statistics: Theory and Practice*. New York: Marcel Dekker.

Liu, R. V. (1990). On a notion of data depth based on random simplices. *Ann. Statist. 18*, 405–414.

Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *Ann. Statist. 15*, 780–799.

Nolan, D. and D. Pollard (1988). Functional limit theorems for U-processes. *Ann. Probab. 16*, 1291–1298.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Spinger.

Talagrand, M. (1994). Sharper bounds for gaussian and empirical processes. *Ann. Probab. 22*(1), 28–76.

Wu, L. (1994). Large deviations, moderate deviations and LIL for empirical processes. *Ann. Probab. 22*(1), 17–27.

Wu, L. (1995). Moderate deviations of dependent random variables related to CLT. *Ann. Probab. 23*(1), 420–445.

(Peter Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Postfach 10-0131, D-33501 Bielefeld, Germany

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