Moderate deviations for degenerate $U$-processes

Peter Eichelsbacher

Fakultät für Mathematik, Universität Bielefeld, Postfach 10-0131, D-33501 Bielefeld, Germany

Received 5 February 1999; received in revised form 22 November 1999

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.

MSC: 60F10 (primary); 60B12; 60G17

Keywords: Rank-dependent moderate deviations; $U$-statistics; $U$-processes; VC classes; Bernstein-type inequality; Metric entropy

1. Introduction

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

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

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

$$U_n^m(h, \mu) := \int_{S^n} h \, dL_n^m = \frac{1}{n(m)} \sum_{(i_1, \ldots, i_m) \in I_{n,m}} h(X_{i_1}, \ldots, X_{i_m}), \quad h \in \mathcal{H}.$$
For a fixed $h$, the expression $U_m^n(h, \mu)$ is called a $U$-statistic of order $m$ with kernel $h$ based on the probability measure $\mu$. $U$-processes appear in statistics frequently as unbiased estimators of the functional $\{\mu^{\otimes m} : h \in \mathcal{H}\}$. For instance, Liu’s simplicial depth process (Liu, 1990) is a $U$-process. Nolan 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 (CLT) 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 $L_n(h) = (1/n) \sum_{i=1}^n h(X_i)$ with $h$ varying in a class of uniformly bounded functions. In a recent paper Arcones (1999) proved necessary and sufficient conditions for the large deviation estimations of empirical processes for unbounded classes. 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) for $U$-processes 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 $\sigma$-field $\mathcal{B}$ is said to satisfy the LDP with speed $a_n$ and good rate function $I(\cdot)$ if the level sets $\{x : I(x) < \alpha\}$ are compact for all $\alpha < \infty$ and for all $\Gamma \in \mathcal{B}$ the lower bound

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

and the upper bound

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

hold, where $\text{int}(\Gamma)$ and $\text{cl}(\Gamma)$ denote the interior and closure of $\Gamma$, 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

$$\lim_{n \to \infty} \frac{b_n}{n} = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{n}{b_n^2} = 0. \quad (1.1)$$

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 $n/b_n^2$ and with good rate function $I(\cdot)$ in case the sequence $\{\Phi(n/b_n)Z_n\}_{n \in \mathbb{N}}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$ and with speed $n/b_n^2$. 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 Hoefding’s decomposition of a $U$-statistic we state it here together with some notation. The operator $\pi_{k,m}^\mu = \pi_{k,m}$ (some times called Hoefding projections) acts on $\mu\otimes^m$-integrable symmetric functions $h : S^m \to \mathbb{R}$ (symmetric in the sense that for all $x_1, \ldots, x_m \in S$ and all permutations $s$ of $\{1, \ldots, m\}$ $h(x_1, \ldots, x_m) = h(x_{s1}, \ldots, x_{sm})$) as follows:

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

where $v_1 \otimes \cdots \otimes v_m h = \int \cdots \int h(x_1, \ldots, x_m) \, dv_1(x_1) \cdots dv_m(x_m)$ and

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

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

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

for all $x_1, \ldots, x_{r-1} \in S$, whereas $\int h(x_1, \ldots, x_m) \, d\mu^{(m-r)}(x_{r+1}, \ldots, x_m)$ 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 $\mu$-canonical if all the functions $h \in \mathcal{H}$ are $\mu$-canonical. With this notation we can decompose a $U$-statistic of rank $r$ into a sum of $\mu$-canonical $U$-statistics of different orders. For all $\mu \otimes^m$-integrable symmetric functions $h : S^m \to \mathbb{R}$ the following equation holds:

$$U^m_n (h, \mu) = \sum_{k=0}^m \binom{m}{k} U^k_n (\pi_{k,m} h, \mu). \quad (1.2)$$

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

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

Hoefding’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 $r \in \{1, \ldots, m\}$ by

$$M^{m,r}_n (h, \mu) = \left( \frac{n}{b_n} \right)^r \sum_{k=r}^m \binom{m}{k} U^k_n (\pi_{k,m} h, \mu).$$

The moderate $U$-process of rank $r$ indexed by $\mathcal{H}$ is defined by $\{M^{m,r}_n (h, \mu), h \in \mathcal{H}\}$. The goal of this paper is to establish sufficient conditions on the class $\mathcal{H}$ of functions to get LDP for $\{M^{m,r}_n (h, \mu), h \in \mathcal{H}\}$ 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 $\|H\|_{\infty} := \sup_{h \in \mathcal{H}} |H(h)|$. This is, in general, a non-separable Banach space if $\mathcal{H}$ is infinite. The aim is to prove a LDP for $\{M^{m,r}_n (h, \mu), h \in \mathcal{H}\}$ 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 $\{n^{m/2}U_n^m(h, \mu)\}$ converges in distribution if and only if $Eh^2$ 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 $G_\mu$, the centered Gaussian process indexed by $L^2(S, \mu)$ with covariance $E(G_\mu(f)G_\mu(g)) = \mu(fg) - (\mu f)(\mu g)$, $f, g \in L^2(S, \mu)$. 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

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

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 $r \in \{1, \ldots, m\}$. 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 $r \leq m$

To the best of our knowledge, the present situation of the MDP problem for $U$-statistics and $U$-empirical measures is as follows. Let $\mathcal{H}(S^m)$, $m \in \mathbb{N}$, denote the set of all signed measures on $(S^m, \mathcal{G}^\otimes m)$ 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 $M_n^{m,1} : \Omega \to \mathcal{H}(S^m)$ by

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

(2.1)
for every \( n \geq m \). 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 \( \tau \) be the coarsest topology on \( \mathcal{M}(S^m) \) such that \( \mathcal{M}(S^m) \ni \varphi \mapsto \int_{S^m} \varphi \, d\nu \) is continuous for every \( \varphi \in B(S^m) \). The large deviations of \( \{M_n\}_{n \in \mathbb{N}} \) on the scale \( \{h^2_n/n\}_{n \in \mathbb{N}} \) have been studied by Borovkov and Mogulskii (1980, Section 3) in the \( \tau \)-topology on the space \( \mathcal{M}(S) \) when \( S \) is a Hausdorff topological space with Borel \( \sigma \)-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 \( \{h^2_n/n\}_{n \in \mathbb{N}} \) when \( (S, \mathcal{S}) \) is a general measurable space. Eichelsbacher and Schmock (1998) proved large deviations of the moderate \( U \)-empirical measures \( \{M_n^{m,1}\}_{n \geq m} \) in stronger topologies, generated by a collection \( \Phi \) of functions \( \varphi : S^m \rightarrow \mathbb{R} \) 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 \( \mathcal{S} \neq \{\emptyset, S\} \) and \( m \geq 2 \), we are not able to separate the zero measure from the measure \( \delta_{x_{m+1}} \otimes \delta_{y_{m+1}} \) with \( x \in A \in \mathcal{S} \) and \( y \in S \setminus A \) for example; hence the topology \( \tau \) would lose 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 \( \mu \)-integrating \( \varphi(s_1, \ldots, s_m) \) with respect to every \( s_i \) with \( i \in \{1, \ldots, m\} \setminus A \). By convention, \( \varphi_{\emptyset} = \int_{S^m} \varphi \, d\mu^{\otimes m} \in \mathbb{R} \). Furthermore, define \( \hat{\varphi}_A \in L_1(\mu^{\otimes |A|}, \mathbb{R}) \) by

\[
\hat{\varphi}_A = \sum_{B \subset A} (-1)^{|A \setminus B|} \varphi_B \quad \text{for every non-empty } A \subset \{1, \ldots, m\}, \text{ and let } \hat{\varphi}_\emptyset = \varphi_\emptyset.
\]

(2.2)

for every non-empty \( A \subset \{1, \ldots, m\} \), and let \( \varphi_A = \varphi_{\emptyset} \). According to the inclusion–exclusion principle or the Möbius inversion formula,

\[
\varphi(s_1, \ldots, s_m) = \sum_{A \subset \{1, \ldots, m\}} \hat{\varphi}_A ((s_i)_{i \in A})
\]

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

\[
\int_{S^n} \varphi \, d\mu^{\otimes n} = \hat{\varphi}_{\emptyset} + \sum_{a=1}^m \int_{S^n} \hat{\varphi}_a \, d\mu^{\otimes n}
\]

(2.3)

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

\[
\hat{\varphi}_a = \sum_{A \subset \{1, \ldots, m\}, |A|=a} \hat{\varphi}_A.
\]

(2.4)

Note that every \( \hat{\varphi}_A \) with non-empty \( A \subset \{1, \ldots, m\} \) is completely \( \mu \)-degenerate. Moreover, decomposition (2.3) is exactly (1.2) for every symmetric \( \varphi \) (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 (De la Peña and Giné, 1999, 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
\( \tau \) of \( \{1, \ldots, m\} \) such that \( \tau(i_j) = j \) for every \( j \in \{1, \ldots, m\} \). Using \( \tau \), define the mapping \( \gamma_{A,m} : \mathcal{M}(S^n) \to \mathcal{M}(S^m) \) by \( \gamma_{A,m}(v) = (v \otimes \mu^{m-a})\pi^{-1}_{\tau} \) for all \( v \in \mathcal{M}(S^n) \), where \( \pi_{\tau} \) is defined by \( \pi_{\tau}(s) = (s_{\tau(1)}, \ldots, s_{\tau(m)}) \) for every \( s = (s_1, \ldots, s_m) \in S^m \). The marginal measure of \( \gamma_{A,m}(v) \), corresponding to the ordered indices contained in \( A \), is then given by \( v \), all other one-component marginals equal \( \mu \). Related to (2.2), define
\[
\tilde{\gamma}_{A,m}(v) = (-1)^{|A|+m} + \sum_{B \subset A, B \neq \emptyset} (-1)^{|A|+|B|} \gamma_{B,m}(v_{A,B}),
\]
where \( v_{A,B} \) denotes the marginal \( v_{i_1, \ldots, i_j} \) of \( v \in \mathcal{M}(S^n) \) when \( B = \{i_1, \ldots, i_j\} \) 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
\[
M_{n,m}^{r} = \left( \frac{n}{b_n} \right)^{r} \left( L_m^m - \mu^{m} - \sum_{A \subset \{1, \ldots, m\}} \sum_{1 \leq |A| \leq r-1} \tilde{\gamma}_{A,m}(L_n^{|A|}) \right).
\]
If \( r = 1 \), then (2.5) reduces to (2.1). Using (2.2)–(2.4), it follows from these definitions that, for every \( \varphi \in L_1(\mu^{m}, \mathbb{P}) \),
\[
\int_{S^m} \varphi \, dM_{n,m}^{r} = \left( \frac{n}{b_n} \right)^{r} \sum_{a=r}^{m} \int_{S^n} \hat{\varphi}_a \, dL_n^a \ \mathbb{P}\text{-a.s.,}
\]
which means that \( M_{n,m}^{r} \) extracts from \( \varphi \) 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
\[
I_{m,r}(v) = \frac{1}{2} \int_{S^m} \left( \frac{d\tilde{v}}{d\mu} \right)^2 \, d\mu
\]
if there exists a \( \tilde{v} \in \mathcal{M}(S) \) satisfying \( \tilde{v}(S) = 0 \) and \( \tilde{v} \ll \mu \) such that
\[
v = \sum_{A \subset \{1, \ldots, m\}} \gamma_{A,m}(\tilde{v}^{\otimes r})
\]
and we define \( I_{m,r}(v) = \infty \) otherwise. In the case \( r = 1 \), Eq. (2.7) reduces to \( v = \sum_{i=1}^{m} \mu^{i-1} \otimes \tilde{v} \otimes \mu^{m-i} \) and every one-component marginal of \( v \) is equal to \( \tilde{v} \). 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 \( \tau \)-topology and rate function \( I_{m,r}(\cdot) \).

Remark 2.8. We apply the results of Eichelsbacher and Schmock (1998) for real valued functions \( h \) in the \( \tau \)-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 takes values in a separable real Banach space \((E, \| \cdot \|_E)\) with Borel \( \sigma \)-algebra \( \mathcal{E} \) and are \( \mathcal{F} \otimes \mathcal{E} \)-measurable. On \( \mathcal{B}(S^m) \) consider the \( \tau(E) \)-topology which makes the maps \( \mathcal{B}(S^m) \ni \varphi \mapsto \int_{S^m} \varphi \, d\nu \) continuous for all \( \varphi \in \mathcal{B}(S^m, E) \), where \( \int \) denotes the Bochner integral. Let \( \{ b_n \}_{n \in \mathbb{N}} \) satisfy (1.1).

Eichelsbacher and Schmoekl (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 \( n/b_n^p \) 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, Vapnik–Cervonenkis 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. Note that decomposition (2.3) coincides with the Hoeffding decomposition (1.2) when we start with a symmetric \( \varphi \). Since the \( \varphi_i \) 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 behavior 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 \((S, \mathcal{F})\) is a separable measurable space and the class \( \mathcal{H} \) is image admissibleSuslin (see Section 3.5 in De la Peña and Giné, 1999). 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 \( \sup_{h \in \mathcal{H}} \left| \sum e_i \cdots e_m h(X_{i_1}, \ldots, X_{i_m}) \right| \), as well as other supers that appear in this paper, now are measurable, where \( e_i, i \in \mathbb{N} \), are i.i.d. Rademacher i.e. \( \mathbb{P}(e_i = 1) = \mathbb{P}(e_i = -1) = \frac{1}{2} \). To ensure measurability, we will assume without further mention that random variables \( e_i, e_i^{(j)}, X_i, X_i^{(j)}, j \leq m, i \in \mathbb{N} \), are the coordinate functions of an infinite product probability measure. The \( e \) variables are Rademacher variables and the \( X \)'s all have law \( \mu \).

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 \( k \leq m \). 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 De la Peña and Giné (1999, 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–Cervonenkis subgraph class of functions. To state the results we have to introduce some additional notation. Given a pseudo-metric
The covering number \( N(\varepsilon,T,d) \) is defined as

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

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 \( \varkappa \) be a real positive number. Then \( \mathcal{H} \) is a Vapnik–Cervonenkis (VC) class of functions for \( H \) and \( \varkappa \) or \( \mathcal{H} \) is VC \((H;\varkappa)\) if there exists a Lebesgue integrable function \( \lambda : [0,\infty) \to [0,\infty) \) such that

\[
(\log N_2(\tau(\mu(H^2))^{1/2},\mathcal{H},\mu))^{\varkappa/2} \leq \lambda(\tau), \quad \tau \in \mathbb{R}_+.
\]  

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

(a) VC classes of sets (Dudley, 1984) are VC \((1,\varkappa)\) for all \( \varkappa > 0 \). 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 \( h \in \mathcal{H} : \{(x,t) : 0 \leq t \leq h(x_1,\ldots,x_m) \text{ or } h(x_1,\ldots,x_m) \leq t \leq 0\} \). 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 \((H;\varkappa)\) for all \( \varkappa \) provided \( \mathcal{H} \) admits an everywhere finite measurable envelope \( H \) (see Pollard, 1984). Note, that if \( \mathcal{C} \) is a VC class of sets and \( q \) a real function on \( \mathcal{C} \), then the class \( \{C/q(C) : C \in \mathcal{C}\} \) 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 \((H,\varkappa)\) for all \( \varkappa > 0 \) (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 \( \varkappa \) such that, for each probability measure \( \mu \) with \( \mu(H^2) < \infty \),

\[
N_2(\varepsilon,\mathcal{H},\mu) \leq A(\mu(H^2)^{1/2}/\varepsilon)^{\varkappa/2},
\]

where \( \varkappa \) denotes the index of the class of subgraphs (see also De la Peña and Giné, 1999, Theorem 5.1.15).

2.3. The main results

To state our main results we define a (rate) function \( J_{m,r}^\mathcal{H} : l_\infty(\mathcal{H}) \to \mathbb{R} \) by

\[
J_{m,r}^\mathcal{H}(H) := \inf\{I_{m,r}(v) : v \in \mathcal{H}(S^m) \text{ and } v(\mathcal{H}) = H, \quad H \in l_\infty(\mathcal{H})\}, \quad (2.11)
\]

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

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

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 $\varepsilon_H > 0$ such that

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

Let $K(J_{m,r}^H, l) = \{H \in l_\infty(\mathcal{H}): J_{m,r}^H(H) \leq l\}$. 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 $r \in \{1, \ldots, m\}$: If $\pi_{k,m} \mathcal{H}$ has an envelope $\pi_{k,m}H$ which satisfies Cramér’s condition 2.12 for every $k \in \{r, r + 1, \ldots, m\}$, then:

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

(b) $\liminf_{n \to \infty} \frac{n}{b_n} \log P(M_n^{m,r}(\mathcal{H}) \in B) \geq -J_{m,r}^H (\text{int}_{l_\infty(\mathcal{H})}(B))$

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

(c) $\limsup_{n \to \infty} \frac{n}{b_n} \log P(M_n^{m,r}(\mathcal{H}) \in B) \leq -J_{m,r}^H (\text{cl}_{l_\infty(\mathcal{H})}(B))$

for every measurable $B$, where $\text{cl}_{l_\infty(\mathcal{H})}(B)$ 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 $r \in \{1, \ldots, m\}$:

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

(b) $\liminf_{n \to \infty} \frac{n}{b_n} \log P\left( \binom{n}{r}^{\otimes m} \mathcal{U}_n^{r}(\pi_{r,m} \mathcal{H}, \mu) \in B \right) \geq -J_{m,r}^H (\text{int}_{l_\infty(\mathcal{H})}(B))$

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

(c) $\limsup_{n \to \infty} \frac{n}{b_n} \log P\left( \binom{n}{r}^{\otimes m} \mathcal{U}_n^{r}(\pi_{r,m} \mathcal{H}, \mu) \in B \right) \leq -J_{m,r}^H (\text{cl}_{l_\infty(\mathcal{H})}(B))$

for every measurable $B$, where $\text{cl}_{l_\infty(\mathcal{H})}(B)$ 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 $r \in \{1, \ldots, m\}$.

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 De la Peña and Giné (1999).

3. Bernstein-type inequalities

First we will state the Bernstein-type inequality for $U$-processes proved in Arcones and Giné (1994) (we state the formulation of De la Peña and Gine, (1999, Theorem 5.14)). We say that the class $\mathcal{H}$ is uniformly bounded if there is $M < \infty$ such that $\|h\|_{\infty} \leq M$ 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 $0 \leq h \leq 1$. 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 $i$ for $(i_1, \ldots, i_m)$, $\tilde{\varepsilon}_i$ for $\tilde{\varepsilon}_{i_1} \cdots \tilde{\varepsilon}_{i_m}$, $\tilde{\varepsilon}_m^\text{dec}$ for $\tilde{\varepsilon}_{i_1} \cdots \tilde{\varepsilon}_{i_m}$, $h(X_i)$ for $h(X_{i_1}, \ldots, X_{i_m})$, and $h(X_i^\text{dec})$ for $h(X_{i_1}^1, \ldots, X_{i_m}^m)$ (“dec” standing for “decoupled”).

Proposition 3.1 (Arcones and 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 $k \in \{1, \ldots, m\}$ there exist constants $c_k$ and $d_k$ such that, for all $\mu$ on $(S, \mathcal{S})$, $t > 0$, and $n \geq m$,

$$
\mathbb{P}(\|n^{1/2}U_n^k(\pi_{k,m}, h, \mu)\|_{\mathcal{H}} \geq t) \leq c_k \exp(-d_k t^{2/k}).
$$

(3.2)

For a proof see Arcones and Giné (1994, Theorem 3.2) or De la Peña and Giné, (1999, Theorem 5.14).

Analyzing the proof of Proposition 3.1 as well as the proof of Arcones and Giné (1993, Proposition 2.3(c)) (see also De la Peña and Giné, 1999, 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 $\mu$-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 \( \mu \) on \((S, \mathcal{F})\), \( t > 0 \), \( \lambda > 0 \) and \( n \geq m \)

\[
\mathbb{P}(\|n^{-m/2}U_n^m(h, \mu)\|_{\mathcal{H}} \geq t) \leq c \exp(-\lambda \frac{t}{2}M) \mathbb{E}(\exp(c' \lambda^2 M^2)) .
\]  

(3.4)

Here \( M \) is defined by

\[
M = \mathbb{E}_x[n^{-m/2} \sum_{I_{a,n}} e_i \cdots e_n h(X_{i_1}, \ldots, X_{i_n})]_{\mathcal{H}} .
\]

(3.5)

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

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

where \( M' \) is given by

\[
M' = \mathbb{E}_x[n^{-m/2} \sum_{I_{a,n}} e_i \cdots e_n h(X_{i_1}, \ldots, X_{i_n})]_{\mathcal{H}}^2 .
\]

(3.7)

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

\[
\left\{ X_t = \sum_{(i_1, \ldots, i_m) \in I_{a,n}} a_{i_1, \ldots, i_m}(t) e_i \cdots e_n \right\}_{t \in T}
\]

with the functions \( a_{i_1, \ldots, i_n}(t) \in \mathbb{R} \), symmetric in the indices, and \( \{e_i\} \) a Rademacher sequence. Assume that \( \|X\| := \sup_{t \in T} |X_t| < \infty \) a.s. It is well known (see e.g. De la Peña and Giné, 1999, 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

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

(3.8)

where \( \sigma := (\mathbb{E}\|X\|^2)^{1/2} \). Moreover, we apply a metric entropy bound for sub-Gaussian processes (see e.g. Proposition 2.6 in Arcones and Giné, 1993 and Corollary 5.1.8 in De la Peña and Giné, 1999):

\[
\mathbb{E} \sup_{t \in T} |X_t| \leq K_m \int_0^D (\log N(\epsilon, T, \varrho))^{m/2} d\epsilon
\]

(3.9)

and

\[
\mathbb{E} \sup_{x, t \in T, \varrho \leq \varrho(t)} |X_t - X_{ts}| \leq K_m \int_{0}^{\delta} (\log N(\epsilon, T, \varrho))^{m/2} d\epsilon
\]

(3.10)

for all \( 0 < \delta \leq D \), where \( \varrho(s, \varrho) = (\mathbb{E}|X_t - X_{ts}|^2)^{1/2} \), \( D \) is the diameter of \( T \) for the pseudodistance \( \varrho \) and \( K_m \) is a universal constant.

We need a convex modification \( \Psi \) of \( \exp(x^2) \), \( \alpha \leq 1 \), \( \lambda > 0 \). Note that \( \exp(x^2) \) is only convex in the range \( x^2 \geq (1 - \alpha)/\alpha \). Replace \( y = \exp(x^2) \) by \( y = \exp((1 - \alpha)/\alpha) \) for \( 0 < x^2 \leq (1 - \alpha)/\alpha \). Hence, \( \Psi \) satisfies

\[
\exp(|x|^\alpha) \leq \Psi(|x|) \leq a_\alpha \Psi(|x|)
\]

(3.11)

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 \( \Psi \). We can apply the decoupling inequality (De la Peña, 1992) and (3.11) to obtain
\[
\mathbb{E} \left[ \exp \left( \lambda \left\| n^{-m/2} \sum_{l_{n,s}} h(X_l) \right\|^{1/m} \right) \right] \leq c \mathbb{E} \left[ \Psi \left( c' \lambda^m \left\| n^{-m/2} \sum_{l_{n,s}} h(X_{l_{n,s}}^{\text{dec}}) \right\| \right) \right].
\]

Applying symmetrization (sometimes called a randomization inequality, contained in De la Peña, 1992, see also De la Peña and Gine, 1999, Theorem 3.5.3, which uses the completely \( \mu \)-degeneracy of \( h \), the decoupling inequality (De la Peña and Gine, 1999 Theorem 3.5.3) as well as (3.11), it follows that
\[
\mathbb{E} \left[ \Psi \left( c' \lambda^m \left\| n^{-m/2} \sum_{l_{n,s}} h(X_{l_{n,s}}^{\text{dec}}) \right\| \right) \right] \leq c \mathbb{E} \left[ \exp \left( c' \lambda^m \left\| n^{-m/2} \sum_{l_{n,s}} v_i h(X_i) \right\|^{1/m} \right) \right].
\]

Applying (3.8) with \( t = \lambda \) and \( \alpha = 1/m \), we obtain
\[
\mathbb{E} \left[ \exp \left( \lambda \left\| n^{-m/2} \sum_{l_{n,s}} h(X_l) \right\|^{1/m} \right) \right] \leq c \mathbb{E} \left[ \exp \left( c' \lambda^2 \left( \mathbb{E}_v \left\| n^{-m/2} \sum_{l_{n,s}} v_i h(X_i) \right\|^{2/m} \right) \right) \right],
\]

letting \( \mathbb{E}_v \) denote integration with respect to the Rademacher variables only. With Borell’s inequality for a Rademacher chaos process (see Borell, 1979), which is
\[
(E \|X\|^p)^{1/p} \leq ((p - 1)/(q - 1))^{m/2}(E \|X\|^q)^{1/q} \quad \text{for all } 1 < q < p < \infty,
\]

it follows that \( (E \|X\|^2)^{1/2} \leq 2^{3m/2}E \|X\| \): as in Littlewood’s proof of Khinchin’s inequality (see De la Peña and Gine, 1999, Theorem 1.3.2) for \( q = 1 \) and \( p = 2 \) it follows using Hölder’s inequality:
\[
E \|X\|^2 = E(E \|X\|^{1/2} \|X\|^{3/2}) \leq (E \|X\|^{1/2})(E \|X\|^{3/2})^{1/2}.
\]

Using Borell’s inequality it follows
\[
E \|X\|^2 \leq (E \|X\|)^{1/2} 2^{3m/4} (E \|X\|^{2})^{1/4}.
\]

This gives \( (E \|X\|^2)^{1/2} \leq 2^{3m/2}E \|X\| \). Eq. (3.4) follows now by Markov’s inequality applied to the function \( e^{2\|\cdot\|_\Psi^{2(m+1)}} \).

Proving (3.6) we take \( t = \lambda \) and \( \alpha = 2/(m + 2) \) in (3.8) and apply the decoupling and symmetrization inequality to the corresponding convex modification, to conclude
\[
\mathbb{E} \left[ \exp \left( \lambda \left\| n^{-m/2} \sum_{l_{n,s}} h(X_l) \right\|^{2/(m+1)} \right) \right] \leq c \mathbb{E} \left[ \exp \left( c' \lambda^{m+1} n^{-m} \left( \mathbb{E}_v \left\| n^{-m/2} \sum_{l_{n,s}} v_i h(X_i) \right\|^{2} \right) \right) \right].
\]

Eq. (3.6) follows by Markov’s inequality applied to the function \( e^{2\|\cdot\|_\Psi^{2(m+1)}} \).
Remark 3.12. An immediate consequence of the proof of Lemma 3.3 is that we can replace $M$ and $M'$, respectively, by
\[
M = \mathbb{E} \left\| n^{-m/2} \sum_{I_{n,m}} \epsilon_{I_1} \cdots \epsilon_{I_m} h(X_{I_1}, \ldots, X_{I_m}) \right\|_{\mathcal{F}}
\]
and
\[
M' = \mathbb{E} \left\| n^{-m/2} \sum_{I_{n,m}} \epsilon_{I_1} \cdots \epsilon_{I_m} h(X_{I_1}, \ldots, X_{I_m}) \right\|^2_{\mathcal{F}}.
\]

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

Remark 3.13. If $\mathcal{H}$ is VC($1,m$), then there are constants $c$ and $c'$ such that for all $t > 0$, $\lambda > 0$, and $n \geq m$ we have
\[
\mathbb{P}(\|n^{k/2} U_k(\sigma_{k,m} h, \mu)\|_{\mathcal{F}} \geq t) \leq c \exp(-\lambda t^{1/2}) \mathbb{E}(\exp(c' t^2 (M'')^{2/k})).
\]

Here $M''$ is given by the following quantity. Consider the Rademacher process
\[
M(h) = n^{-k/2} \sum_{I_{n,m}} \epsilon_i \sigma_{k,m} h(X_{I_i}, \ldots, X_{I_k}), \quad h \in \mathcal{H}
\]
and the $L_2$ pseudodistance associated to the Rademacher process $M$, the square root of
\[
\mathbb{E}(M(f) - M(g))^2 = \mu_{n,k}(f, g) = n^{-k} \sum_{I_{n,m}} (\sigma_{k,m}(f - g))^2(X_{I_1}, \ldots, X_{I_k})
\]
and let $D_2 := \sup_{f,g \in \mathcal{H}} \mu_{n,k}(f, g)$. Then using (3.9) we obtain
\[
M'' = \int_0^D (\log N_2(\tau, \mathcal{H}, \mu_{n,k}))^{3/2} d\tau.
\]

By De la Peña and Giné (1999, Lemma 5.3.5) the covering number of $(\mathcal{H}, \mu_{n,k})$ is a finite product of covering numbers of $\mathcal{H}$ of $L_2$ distances of probability measures on $S^m$ and therefore, $\mathcal{H}$ being VC($1,m$), $M''$ is uniformly bounded in $n$. Proposition 3.1 follows by Markov’s inequality choosing $\lambda = t^{2/k} / (2c')$.

Corollary 3.14. Let $\mathcal{H}$ be a uniformly bounded VC($1,m$) class of measurable functions $h : S^m \rightarrow \mathbb{R}$, symmetric in their entries. Let $\mathcal{G} := \{(f - g)^2 : f, g \in \mathcal{H}\}$. For $h_1, f_1 \in \mathcal{H}$ denote by $\tilde{\sigma}_{k,m}(h_1 - f_1)^2 := (\sigma_{k,m}(h_1 - f_1))^2 - \mu_{n,k}(h_1 - f_1)^2$. Then, for each $k \in \{1, \ldots, m\}$ there exist constants $c_k$ and $d_k$ such that, for all $\mu$ on $(S, \mathcal{S})$, $t > 0$ and $n \geq m$,
\[
\mathbb{P}(\|n^{k/2} U_k(\tilde{\sigma}_{k,m} h, \mu)\|_{\mathcal{G}} \geq t) \leq c_k \exp(-d_k t^{2/k}). \tag{3.15}
\]
Proof. By Lemma 3.3 we have to consider
\[ \mathbb{E} \left\| n^{-k/2} \sum_{l_k \omega} e_l \hat{\pi}_{k,m} h(X_l) \right\|_g. \]
Remark that for \( h_1, h_2, g_1, g_2 \in \mathcal{H} \), using the uniformly boundedness, we have for every \( 1 \leq k \leq m \)
\[ ((\hat{\pi}_{k,m}(h_1 - g_1))^2 - (\hat{\pi}_{k,m}(h_2 - g_2))^2) \leq 4((\hat{\pi}_{k,m}(h_1 - g_1))^2 - (\hat{\pi}_{k,m}(h_2 - g_2))^2). \]
With the definition of \( \mu_k \) as in Remark 3.13 we obtain
\[ \mathbb{E} \left\| n^{-k/2} \sum_{l_k \omega} e_l \hat{\pi}_{k,m} h(X_l) \right\|_g \leq \text{const}(k) \int_0^\infty (\log N(e/8, \mathcal{H}, \mu_k))^k e^k \, de < \infty. \]
Now the proof follows the lines of the proof of Proposition 3.1, see Remark 3.13.

The next inequality is basic in the proof of the rank dependent MDP for degenerate \( U \)-processes over VC-classes. Given a pseudometric \( \epsilon \) on \( \mathcal{H} \) and \( \eta > 0 \), we set
\[ \mathcal{H}'(\eta, \epsilon) := \{ h_2 - h_1: h_1, h_2 \in \mathcal{H}, e(h_1, h_2) \leq \eta \}. \]
In the following, we take the pseudodistance \( \epsilon^2(f, g) := \mu \ominus m(f - g)^2 \) for \( f, g \in \mathcal{H} \).
For fixed \( 1 \leq k \leq m \) define the pseudodistance \( \epsilon^2(\pi_{k,m} f, \pi_{k,m} g) = \mu \ominus m(f - g)^2 \) with \( f, g \in \mathcal{H} \).

**Lemma 3.16.** 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 \( k \in \{1, \ldots, m\} \) there exist constants \( c_k \) and \( d_k \), depending only on \( m \), such that, for all \( \mu \) on \( (S, \mathcal{H}) \), \( t > 0 \) and all \( \eta > 0 \)
\[ \mathbb{P}\left( \| n^{k/2} U_n^k(\pi_{k,m} h, \mu) \|_{\mathcal{H}'(\eta, \epsilon)} \geq t \right) \leq c_k \exp \left( -d_k \left( \frac{t}{\kappa(\eta)} \right)^{2k} \right) + c_k \exp \left( d_k \left( \frac{t^{2k}(K - 2\kappa(\eta)^{2k})}{\kappa(\eta)^{4k}} \right) \right) \mathbb{P}(A_{n,k}(\eta)). \]
(3.17)
where $\kappa(\eta)$ is a function with the property $\lim_{\eta \to 0} \kappa(\eta) = 0$, $K$ is a constant which depends on $\mathcal{H}$ only and the set $A_{n,k}(\eta)$ is defined in (3.18).

For the proof of Lemma 3.16 we combine arguments from the proof of Theorem 5.3.7 (CLT for $U$-processes) in De la Peña and Giné (1998) with arguments from the proof of Proposition 3.1:

**Proof.** For every fixed $k \in \{1, \ldots, m\}$ we apply (3.4) in Lemma 3.3 with $\mathcal{H}$ replaced by $\mathcal{H}(\eta, \varepsilon)$. Hence we have to estimate $E(\exp(\lambda^2(M_{n,k})^2))$ with $M_{n,k}$ defined by

$$M_{n,k} = \frac{1}{n^{k/2}} E_{\varepsilon} \left\| \sum_{I_{n,k}} \varepsilon_i \pi_{k,m} h(X_i) \right\|^2_{\mathcal{H}(\eta, \varepsilon)}.$$  

For each $n \in \mathbb{N}$ the $L_2$ distance associated to the conditional Rademacher chaos process $M_{n,k}$ is the pseudometric $e_{n,k}$ already defined in the proof of Corollary 3.14,

$$e_{n,k}(h_1, h_2) := \frac{1}{n^k} E_{\varepsilon} \left( \sum_{I_{n,k}} \varepsilon_i (\pi_{k,m} h_1 - \pi_{k,m} h_2)(X_i, \ldots, X_i) \right)^2$$

$$= \frac{1}{n^k} \sum_{I_{n,k}} (\pi_{k,m} h_1 - \pi_{k,m} h_2)(X_i, \ldots, X_i).$$

For each $\eta > 0$ and $n \geq k$ let

$$A_{n,k}(\eta) := \{ \omega: ||e_{n,k}(h_1, h_2) - \mathcal{E}^2(\pi_{k,m} h_1, \pi_{k,m} h_2)||_{\mathcal{H} \times \mathcal{H}} \geq \eta^2 \}.$$  

(3.18)

The law of large numbers in De la Peña and Giné (1999, Lemma 5.3.6) (see also Arcones and Giné (1993, Theorem 3.1 and Corollary 3.3)), implies

$$\lim_{n \to \infty} P(A_{n,k}(\eta)) = 0$$

for all $\eta > 0$. Since $\pi_{k,m}$ is a projection in $L_2$, $\{f, g \in \mathcal{H}: e(f, g) < \eta \} \subseteq \{f, g \in \mathcal{H}: \mathcal{E}(\pi_{k,m} (f - g))^2 < \eta^2 \}$ and therefore

$$\{f, g \in \mathcal{H}: e(f, g) < \eta \} \subseteq \{f, g \in \mathcal{H}: e_{n,k}(f, g) < \eta^2 \}$$

on the set $(A_{n,k}(\eta))^c$. The entropy maximal inequality for Rademacher chaos processes (3.10) then gives on $(A_{n,k}(\eta))^c$

$$M_{n,k} \leq K \int_0^\eta (\log N(\varepsilon, \mathcal{H}, e_{n,k}))^{1/2} d\varepsilon$$

(3.19)

and on $A_{n,k}(\eta)$

$$M_{n,k} \leq K \int_0^\infty (\log N(\varepsilon, \mathcal{H}, e_{n,k}))^{1/2} d\varepsilon.$$  

(3.20)

Let $U_{n}^r \times \mu_{\otimes^{m-r}}$ denote the random probability measure

$$U_{n}^r \times \mu_{\otimes^{m-r}} = \left( \frac{(n-r)!}{n!} \sum_{I_{n,k}} \delta_{(X_i, \ldots, X_i)} \right) \times \mu_{\otimes^{m-r}}.$$
By Lemma 5.3.5 in De la Peña and Giné (1999) the right-hand side of (3.19) is dominated by a constant times the sum from $r = 0$ to $k$ of

$$
\int_0^{\eta'} \left( \log(N(\varepsilon, \mathcal{H}, ||H_{L}||_{L^2(\mathbb{Z}^\omega \times \mu^{\otimes n-r})})) \right)^{k^2} \, d\varepsilon,
$$

where $\eta'$ is a function of $\eta$ that tends to 0 as $\eta \to 0$. The right-hand side of (3.20) is dominated by a constant times the sum from $r = 0$ to $k$ of

$$
\int_0^{\infty} \left( \log(N(\varepsilon, \mathcal{H}, ||H_{L}||_{L^2(\mathbb{Z}^\omega \times \mu^{\otimes n-r})})) \right)^{k^2} \, d\varepsilon.
$$

Since $\mathcal{H}$ is a VC($1, m$) class we obtain that there exist a function $\kappa(\eta)$ with $\lim_{\eta \to 0} \kappa(\eta) = 0$ and a constant $K$ such that

$$
\mathbb{E}(\exp(c'\lambda^2 M'_{n,k})) \leq \exp(c'\lambda^2 \kappa(\eta)^{2/k}) + \mathbb{P}(A_{n,k}(\eta)) \exp(c'\lambda^2 K).
$$

Using the fact that $\kappa(\eta) \leq K$, we choose in (3.4) $\lambda = t^{1/k}/(2c'(\kappa(\eta))^{2/k})$ and obtain the result. \(\square\)

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

**Lemma 3.21.** Let $\mathcal{H}$ be a measurable class of $\mu$-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

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

(3.22)

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

$$
M' \leq c n^{-m} \left( \mathbb{E} \left\| \sum_{i=0}^{n} \hat{e}_{i, dec} \hat{h}(X_{i, dec}) \right\|_{\mathcal{H}} \right)^2.
$$

Now since $\mathcal{H}$ is assumed to be VC($H, m$), we can apply Lemma 2.2 in Arcones and Giné (1995):

$$
\mathbb{E} \left\| \sum_{i=0}^{n} \hat{e}_{i, dec} \hat{h}(X_{i, dec}) \right\|_{\mathcal{H}} \leq c \left( \sum_{i=0}^{n} H^2(X_{i, dec}) \right)^{1/2}.
$$

We conclude that

$$
\mathbb{E}(\exp(c'\lambda^{m+1} M')) \leq \mathbb{E} \left( \exp \left( c' \lambda^{m+1} \frac{1}{n^m} \sum_{i=0}^{n} H^2(X_{i, \ldots, X_{im}}) \right) \right).
$$
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 De la Peña and Giné, 1999); 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. 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 $e$.

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

$$
\limsup_{n \to \infty} \frac{n}{b_n^r} \log \mathbb{P} \left( \left\| \frac{M_{n,r}^m(h)}{b_n^r} - \left( \frac{n}{b_n^r} \right)^r \right\|_{\mathcal{H}} \geq \delta \right) = -\infty \quad (4.1)
$$

(see Dembo and Zeitouni, 1998, Theorem 4.2.13, the concept of exponential equivalence). We can apply Proposition 3.1. With (3.2) in Proposition 3.1 we get for each $r + 1 \leq k \leq m$:

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

This implies

$$
\frac{n}{b_n^r} \log \mathbb{P} \left( \left\| \left( \frac{n}{b_n^r} \right)^r U_n^k(\pi_{k,m}, h, \mu) \right\|_{\mathcal{H}} \geq \delta \right) \leq \frac{n}{b_n^r} \log c_k - c_k' \delta^{2-k} \frac{n^{2-(2r-k)}}{b_n^{2-(2r-k)}}.
$$
Since \( k \in \{r+1, r+2, \ldots, m\} \), the right-hand side decreases to \(-\infty\) by the assumptions for \( \{b_n\}_{n \in \mathbb{N}} \). Since
\[
\left\{ \left\| \left( \frac{n}{b_n} \right)^r \sum_{k=r+1}^m \binom{m}{k} U_n^k(\pi_{b_n \mu}, \mu) \right\|_{\mathcal{F}} \geq \delta \right\}
\]
\[
\subset \bigcup_{k=r+1}^m \left\{ \left( \frac{n}{b_n} \right)^r \left\| \binom{m}{k} U_n^k(\pi_{b_n \mu}, \mu) \right\|_{\mathcal{F}} \geq \frac{\delta}{m-r} \right\},
\]
Lemma 1.2.15 in Dembo and Zeitouni (1998) yields (4.1). Hence what follows is the proof of Theorem 2.14.

**Step 2:** Note 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 \( \mathcal{F}\otimes m;\mathcal{E} \)-measurable functions \( h : S^m \to E \), where \( E \) is a separable real Banach space of type 2 and each \( h \) satisfies the condition
\[
\int_{S^m} \exp(x \| \pi_{r,m} h \|_E) \, d\mu^\otimes r < \infty
\]
for every \( x > 0 \) as well as there exist \( x_k > 0 \) for \( k \in \{r+1, r+2, \ldots, m\} \) such that
\[
\int_{S^m} \exp(x_k \| \pi_{b_n \mu} \|_{L_k}^2) \, d\mu^\otimes k < \infty
\]
for each such \( k \).

**Step 3:** The level sets \( K(J_{m,r}^\mathcal{H}, L) := \{ H \in l_\infty(\mathcal{H}) : J_{m,r}^\mathcal{H}(H) \leq L \}, L \geq 0 \), are compact in \( l_\infty(\mathcal{H}) \) by Arzelà–Ascoli (Dunford and Schwartz, 1967, Theorem 5, Section IV.6). Therefore, we have to check that each level set is a bounded subset of \( C_b(\mathcal{H}, e) \), the set of all bounded and continuous functionals on \( (\mathcal{H}, e) \) and that \( (\mathcal{H}, e) \) is totally bounded (which holds, since \( \mathcal{H} \) is a VC(1, m)-class). For every \( H \) with \( J_m^\mathcal{H}(H) < \infty \) 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 \( (\mathcal{H}, e) \) and \( K(J_{m,r}^\mathcal{H}, L) \) is a subset of this space. Using the uniform boundedness of \( \mathcal{H} \) we obtain \( \| H \|_{\mathcal{H}} \leq 1 \), i.e. \( K(J_{m,r}^\mathcal{H}, L) \) is a bounded subset of \( C_b(\mathcal{H}, e) \). Therefore, \( K(J_{m,r}^\mathcal{H}, L) \) is compact in \( l_\infty(\mathcal{H}) \) for every \( L > 0 \).

**Step 4:** We show, that the MDP is established once we have
\[
\lim_{n \to 0} \log \mathbb{P} \left( \left\| \left( \frac{n}{b_n} \right)^r \sum_{k=r+1}^m \binom{m}{k} U_n^k(\pi_{b_n \mu}, \mu) \right\|_{\mathcal{F}} \geq \delta \right) = -\infty \tag{4.2}
\]
for all \( \delta > 0 \). Similar to the proof of Eichelsbacher (1998, Theorems 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 \( J_{m,r}^\mathcal{H}(B) = \inf_{H \in B} J_{m,r}^\mathcal{H}(H) \). Let \( C \) be an arbitrary closed set in \( l_\infty(\mathcal{H}) \). It suffices to consider the case \( J_{m,r}^\mathcal{H}(C) > 0 \). Choose \( r \in (0, J_{m,r}^\mathcal{H}(C)) \). For every element \( H \) in \( l_\infty(\mathcal{H}) \) \( C \) 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 \( K(J_{m, r}^F) \) is compact and \( C \cap K(J_{m, r}^F) = \emptyset \), there exists a finite subset \( N \) of \( K(J_{m, r}^F) \) such that \( U := \bigcup_{H \in N} U(H, \delta_H) \) covers \( K(J_{m, r}^F) \). We abbreviate
\[
O_n^{m, r}(h) := \left( \frac{n}{b_n} \right) \binom{m}{r} U_n^r(\pi_r m, h, \mu)
\]
and arrive at
\[
\mathbb{P}((O_n^{m, r})(\mathcal{H}) \in C) \leq \mathbb{P} \left( (O_n^{m, r})(\mathcal{H}) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_H, \delta_H) \right).
\]
For a fixed \( \eta > 0 \) let \( \mathcal{H}^\eta \) denote a finite \( \eta \)-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 \( \epsilon(h, g) < \eta \). 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 \( \epsilon(h, g) < \eta' \) 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 \( \eta \) 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
\[
\sup \{|F(h) - F(g)|; \; h, g \in \mathcal{H} \text{ and } \epsilon(h, g) < \eta\} < \delta,
\]
then \( F \in \bigcup_{H \in N} U(H, 2\delta_H) \). Hence we obtain
\[
\mathbb{P}((O_n^{m, r})(\mathcal{H}) \in C) \leq \mathbb{P} \left( (O_n^{m, r})(\mathcal{H}) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H) \right)
\]
\[
+ \mathbb{P}(\|O_n^{m, r}(h)\|_{\mathcal{H}^\eta} \geq \delta).
\]
By Theorem 1.17 in Eichelsbacher and Schmock (1998) we have
\[
\limsup_{n \to \infty} \frac{\mu}{b_n} \log \mathbb{P} \left( (O_n^{m, r})(\mathcal{H}) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H) \right)
\]
\[
\leq - \inf \left\{ J_\eta(p_\eta(F)); \; p_\eta(F) \in l_\infty(\mathcal{H}) \setminus \bigcup_{H \in N} U(p_\eta(H), \delta_H) \right\} \leq - r.
\]
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 \( J_{m, r}^F(O) < \infty \). There exists a \( \delta_G \) such that \( U(G, \delta_G) \) is contained in \( O \). As in the proof of the upper bound for \( \delta_G \) there exists an \( \eta > 0 \) such that
\[
\mathbb{P}((O_n^{m, r})(\mathcal{H}) \in U(p_\eta(G), \delta_G)) \leq \mathbb{P}(\|O_n^{m, r}(h)\|_{\mathcal{H}^\eta} \geq \delta_G) + \mathbb{P}((O_n^{m, r})(\mathcal{H}) \in O).
\]
We apply the lower bound of Theorem 1.17, Eichelsbacher and Schmock (1998), to the finite subnet $\mathcal{H}$:

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

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

$$\lim_{n \to 0} J_\eta(G) = \liminf_{n \to 0} \{I_{m_n}(v); v \in \mathcal{M}(S^m), v(\mathcal{H}) = G\} = J_{m_n}(G),$$

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 3.16. With $c(m, r) := \binom{m}{r}$ we obtain that for all $n \geq m$

$$\mathbb{P}\left(\left\| \frac{n}{b_n^2} c(m, r) U^n_{\eta}(\pi_{r,m} h, \mu) \right\|_{\mathcal{H}^r(\eta, c)} \geq \delta \right)$$

$$= \mathbb{P}\left(\left\| n^{r/2} U^n_{\eta}(\pi_{r,m} h, \mu) \right\|_{\mathcal{H}^r(\eta, c)} \geq \delta \cdot n^{r/2} b_n^2 \right)$$

$$\leq c \exp\left(-c' \frac{\delta^2 r b_n^2}{nc(m, r)^{2r}(\eta{k})^{2r}}\right) + c \exp\left(c' \frac{\delta^2 r b_n^2 (K - 2\kappa(\eta)^{2r})}{nc(m, r)^{2r}(\eta{k})^{2r}}\right) \mathbb{P}(A_{n, r}(\eta)).$$

(4.4)

We get for the first summand in (4.4) for every $n \geq m$ and $\eta > 0$:

$$\frac{n}{b_n^2} \log \left(c \exp\left(-c' \frac{\delta^2 r b_n^2}{nc(m, r)^{2r}(\eta{k})^{2r}}\right)\right) = \frac{n}{b_n^2} \log c - \frac{c' \delta^2 r}{c(m, r)^{2r}(\eta{k})^{2r}}.$$

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 $\eta \to 0$. For the second summand in (4.4) we obtain for every $n \geq m$ and $\eta > 0$ taking the log function and multiplying with $n/b_n^2$:

$$\frac{n}{b_n^2} \log c + \frac{c' \delta^2 (K - 2\kappa(\eta)^{2r})}{c(m, r)^{2r}(\eta{k})^{2r}} + \frac{n}{b_n^2} \log \mathbb{P}(A_{n, r}(\eta)).$$

By definition of $A_{n, r}(\eta)$ we observe with Corollary 3.14 that

$$\mathbb{P}(A_{n, r}(\eta)) = \mathbb{P}\left(\left\| \frac{1}{n^r} \sum_{l=H} \hat{\pi}_{r,m}(h_1 - h_2)^2 \right\|_{\mathcal{H} \times \mathcal{H}} \geq \eta^2\right)$$

$$\leq \mathbb{P}(\| n^{r/2} U^n_{\eta}(\hat{\pi}_{r,m} h, \mu) \|_{\eta} \geq \eta^2 n^{r/2}) \leq c \exp(-c' \eta^{4r} n)$$

and therefore

$$\frac{n}{b_n^2} \log \mathbb{P}(A_{n, r}(\eta)) \leq \frac{n}{b_n^2} \log c - c' n^2 \eta^{4r},$$

which increases to $-\infty$ when letting $n \to \infty$ using the assumptions for $\{b_n\}_{n \in \mathbb{N}}$. Now (4.2) follows with the principle of the largest term, see Lemma 1.2.15 in Dembo and Zeitouni (1998). 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 $\pi_{k,m} H$ satisfies Condition (2.12) if $r$ is fixed. $\pi_{k,m} \mathcal{H}$ is VC($\pi_{k,m} H, k$) for every $k \in \{1, \ldots, m\}$. This
follows from De la Peña and Giné (1999, Lemma 5.3.5). Theorems 2.13 and 2.14 are equivalent: we prove (4.1) applying Lemma 3.21. We obtain for each $r + 1 \leq k \leq m$:

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

$$\leq c_1 \exp \left( -\frac{\delta^{2/k} b_n^{2r/k} n^{-(2r/k)} 1/\sigma^{2/k}}{c_2 + F_{\pi_{k,m}}(\delta^{1/k} b_n^{r/k}/(n^{r/k}))^{2/(k+1)}} \right).$$

This implies

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

$$\leq n \frac{b_n}{b_n^k} \log c_1 - c_2 \delta^{2/k} b_n^{2r/k} n^{-(2r/k)} \sigma^{2/k}$$

$$\leq c_2 + F_{\pi_{k,m}}(\delta^{1/k} b_n^{r/k}/(n^{r/k}))^{2/(k+1)}.$$
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 $\{M_{n}^{m,1}(\mathcal{H}), n \in \mathbb{N}\}$ satisfies a LDP in $l_\infty(\mathcal{H})$ with speed $n/b_n^2$ and good rate function $J_{m,1}(\cdot)$, defined in (2.11).

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

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

(c) $(\pi_{1,m,H},e_1)$ 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 $\pi_{k,m,H}$ has an envelope $\pi_{k,m,H}$ which satisfies Cramer’s Condition 2.12 for every $k \in \{1,\ldots,m\}$, 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 $\{(n/b_n^2) m(U_{n}^{1}(\pi_{1,m,H})-\mu_{H}^{m}) , n \in \mathbb{N}\}$ satisfies a LDP in $l_\infty(\mathcal{H})$ with speed $n/b_n^2$ and good rate function $J_{m,1}(\cdot)$, 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. Note 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 $U_{n}^{1}(\pi_{r,\hat{h}},\mu)$ is replaced by $V_{n}^{1}(\pi_{r,\hat{h}},\mu)$. 
Remark 6.2. Fix \( r \in \{1, 2\} \). Assume that there exists a \( \alpha_{r,m} > 0 \) such that
\[
\int_S \exp(\alpha_{r,m} |\pi_{r,m} H \circ \pi_r|^2) \, d\mu^{\otimes r} < \infty
\]
for every map \( \tau : \{1, \ldots, r\} \to \{1, \ldots, r\} \), where \( \pi_{r,s} : S^r \to S^r \) is defined by \( \pi_{r,s}(s) = (s_{\tau(1)}, \ldots, s_{\tau(r)}) \) for every \( s = (s_1, \ldots, s_r) \in S^r \). Then for every fixed \( r \in \{1, 2\} \) the three assertions of Theorem 2.14 hold true for a VC\((H, 2)\) class when \( U_n^r(\alpha_{r,2}) \) is replaced by \( V_n^r(\alpha_{r,2}) \). We omit the details.

Proof of Theorem 6.1. It suffices to check that for every \( \delta > 0 \)
\[
\limsup_{n \to \infty} \frac{n}{\log n} \log \mathbb{P} \left( \left\| \frac{n}{b_n} \right\| L_n^{\otimes 2}(\pi_{2,2}) - L_n^2(\pi_{2,2}) \right\| \geq \delta \right) = -\infty.
\]
Using
\[
L_n^{\otimes 2}(\pi_{2,2}) = \frac{1}{n} \sum_{j=1}^n \pi_{2,2}(x_j, x_j) + \frac{n(2)}{m^2} U_n^2(\pi_{2,2}, \mu),
\]
we apply twice Proposition 3.1 and observe
\[
\mathbb{P} \left( \left\| \frac{n}{b_n} \right\| L_n^{\otimes 2}(\pi_{2,2}) - L_n^2(\pi_{2,2}, \mu) \right\| \geq \delta / 2 \) \leq c \exp \left( -d \frac{\delta n b_n^2}{2(n^2 - (n)_2)} \right)
\]
and
\[
\mathbb{P} \left( \left\| \frac{1}{b_n} \sum_{j=1}^n \pi_{2,2}(x_j, x_j) \right\| \geq \delta / 2 \right) \leq c' \exp \left( -d' \frac{\delta^2 b_n^4}{4n} \right),
\]
where \( c, c' \) and \( d,d' \) are some constants. Thus the theorem is proved. \( \square \)

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
\[
\limsup_{n \to \infty} \frac{n}{\log n} \log \mathbb{P} \left( \left\| L_n^{\otimes a}(\pi_{a,m}) - L_n^a(\pi_{a,m}) \right\| \geq \delta \right) = -\infty \quad (6.3)
\]
for every \( \delta > 0 \) and for every \( a \in \{s,s+1, \ldots, m\} \), where \( s \in \mathbb{N} \) is the rank: We use the well known decomposition
\[
L_n^{\otimes a}(\pi_{a,m}) = \frac{1}{n^a} \sum_{j=1}^a j! S_a^{(j)} \binom{n}{j} U_n^j(h_{a,j}, \mu).
\]
Here \( h_{a,j} \) is a function on \( S^j \) defined by
\[
h_{a,j}(x_1, \ldots, x_j) := \frac{1}{j! S_a^{(j)}} \sum_{i,j} \pi_{a,m} h(x_i_1, \ldots, x_{i_j})
\]
and the sum \( \sum_{i,j} \) is taken over all \( a \)-tuples \( (i_1, \ldots, i_a) \) formed from \( \{1, 2, \ldots, j\} \) 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
\[
\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P} \left( \left( \frac{n}{b_n} \right)^a \left\| U_n^a(\pi_{a,m} h, \mu) \right\| \geq \delta \right) = -\infty.
\]
In general, the diagonals \( h_{a,j} \) \((1 \leq j \leq a-1)\) are not \( \mu \)-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 \( h_{a,j} \) \((h_t \text{ is a function in } t \text{ variables and actually depends on } a \text{ and } j)\). We consider with Proposition 3.1 that
\[
\frac{n}{b_n^2} \log \mathbb{P} \left( \left( \frac{n}{b_n} \right)^a \left\| U_n^a(h_t, \mu) \right\| \geq \delta \right) \leq - \delta^2 t^{-2} \frac{(2a)^{t-2}}{n^{t-2}}.
\]
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
\[
\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P} \left( \left( \frac{n}{b_n} \right)^a \left\| U_n^a(f^{\otimes a}, \mu) - V_n^a(f^{\otimes a}, \mu) \right\| \geq \delta \right) = -\infty.
\]

**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
\[
\left( \frac{n}{b_n} \right)^a \left( U_n^a(f^{\otimes a}) - V_n^a(f^{\otimes a}) \right) = \left( \frac{n}{b_n} \right)^a \left( 1 - \frac{n(a)}{a!} \right) U_n^a(f^{\otimes a})
+ R_a \left( \frac{n}{b_n} U_n^1(f), \frac{n}{b_n^2} U_n^1(f^2), \ldots, \frac{n}{b_n^a} U_n^1(f^a) \right).
\]
(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. \( p_1^{k_1} \cdots p_a^{k_a} \) 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 De la Peña and Giné, 1999, Section 4.2). Since \( f^{\otimes a} \) is bounded
and completely \( \mu \)-degenerate and \( 1 - n(a)/a! \to 0 \) 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, \( \lim_{a \to \infty} b_n = \infty \) and \( \lim_{a \to \infty} n/b_n^a = 0 \) by (1.1), it follows that there exists a constant \( C > 0 \) such
that, for all \( n \geq a \),
\[
\left| R_a \left( \frac{n}{b_n} U_n^1(f), \ldots, \frac{n}{b_n^a} U_n^1(f^a) \right) \right| \leq C \sum_{j=0}^{a-1} \frac{n}{b_n} \left( \frac{n}{b_n} |U_n^1(f)| \right)^j.
\]
(6.6)
By applying the Bernstein-type inequality in Proposition 3.1 to every term on the right-hand side of (6.6), we see that we can neglect the last term in (6.5). This proves the theorem.

References