MODERATE AND LARGE DEVIATIONS FOR U-PROCESSES

PETER EICHELSBACHER

Abstract. Sufficient conditions for the moderate and large deviation principle for U-processes are given. For the large deviation result the conditions are in terms of "blockwise" empirical process conditions. On the moderate scaling the case of U-processes indexed by a uniformly bounded VC subgraph class of functions is considered. The proofs are based on an isoperimetric inequality for empirical processes due to Talagrand, a truncation method based on an isoperimetric inequality by Ledoux, the existence of almost regular partitions of complete hypergraphs due to Baranyai and a Bernstein-type inequality for U-processes due to Arcones and Giné.

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 $n_{(m)} := \prod_{k=0}^{m-1}(n-k)$, $I_{m,n} \subset \{1,\dots,n\}^m$ 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 $h :\ S^m \to \mathbb{R}$. The U-process of order m indexed by $\mathcal{H}$ is defined, for every $n \in \mathbb{N}$, 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, $U_n^m(h,\mu)$ is called a U-statistic of order m with kernel h based on the probability mearsure μ. U-processes appear in statistics for example as unbiased estimators of the functional $\{\mu^{\otimes m}h :\ h \in \mathcal{H}\}$. 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 (1991) developed the theory for an arbitrary m. An overview over the theory of empirical processes and U-processes is Giné (1996). Wu (1994) proved necessary and sufficient conditions for the large deviation and moderate deviation

1991 *Mathematics Subject Classification.* 60F05, 60F10, 62G20.

Key words and phrases. Large deviations, moderate deviations, U-statistics, U-processes, isoperimetric inequality, VC classes, Bernstein's inequality.

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 bounded functions. The principles are proved for laws in the Banach space of bounded functionals on the class $\mathcal{H}$.

In this paper we study the large and moderate deviation principles (LDP, MDP) for U-processes under certain conditions on $\mathcal{H}$.

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

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

and the upper bound

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

hold, where $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively. We shall consider situations where the speed in the LDP is given by $a_n = 1/n$. We say that a sequence of random variables satisfies the LDP when 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 $a\,x \in \mathcal{X}$ for all $a > 0$ and $x \in \mathcal{X}$ then a sequence of random variables Z_n shall satisfy the MDP with speed $\frac{n}{b_n^2}$ and with good rate function $I(\cdot)$, if the sequence $\frac{n}{b_n} Z_n$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$ and with speed $\frac{n}{b_n^2}$.

As the best of our knowledge, the present situation of the LDP and MDP problem for U-statistics and U-empirical measures is as follows. Let $\mathcal{M}_1(S^m)$ denote the space of Borel probability measures on S^m. The LDP for the sequence of laws of L_n^m has been studied in Eichelsbacher and Schmock (1996). 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. The LDP for the sequence of laws of L_n^m holds on $\mathcal{M}_1(S^m)$ equipped with the $\tau_1(E)$-topology which makes the maps $\mathcal{M}_1(S^m) \ni \nu \mapsto \int_{S^m} \varphi \, d\nu$ continuous for all $\varphi \in B(S^m, E)$, where $\int$ denotes the Bochner integral. The speed is $1/n$ and the rate function is given by

$$J_m(\nu) = \begin{cases} \frac{1}{m} H_m(\nu \,|\, \mu^{\otimes m}), & \text{if } \nu = \nu_1^{\otimes m}, \\ \infty & \text{otherwise}, \end{cases} \tag{1.2}$$

where ν_1 denotes the first marginal of ν, and

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

(called relative entropy or Kullback-Leibler information). The rate function J_m is non-convex for all $m \geq 2$ (cf. Theorem 1.7, Eichelsbacher and Schmock (1996)). The m-fold products $L_n^{\otimes m} : \Omega \to \mathcal{M}_1(S^m)$ of the empirical measures, i.e.

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

satisfies the LDP with the same rate function (Theorem 1.10, Eichelsbacher and Schmock (1996)).

The results in the case $m = 1$ are an improvement of the general version of Sanov's Theorem, which was proved by de Acosta (1994a) for the general case of an arbitrary measurable state space S and in the $\tau_1(\mathbb{R})$-topology setting. Notice that in Eichelsbacher and Schmock (1996), the LDP result is proved for an even finer topology. This topology is very useful for several applications in statistical mechanics, for example the Gibbs conditioning principle for interacting ensembles of particles and special mean field models.

Moreover in Eichelsbacher and Schmock (1996) we proved a LDP for the moderate U-empirical measure $M_n^m : \Omega \to \mathcal{M}(S^m)$ defined for $n \geq m$ by

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

where $\mathcal{M}(S^m)$ denotes the set of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$ with finite total variation, endowed with the $\tau(E)$-topology, defined analogously to $\tau_1(E)$, and where $\{b_n\}_{n \in \mathbb{N}} \subset (0, \infty)$ satisfies (1.1). Therefore by definition we obtained a MDP for $(L_n^m - \mu^{\otimes m})$. The speed is $\frac{b_n^2}{n}$ and the rate function has the form

$$I_m(\nu) = \frac{1}{2} \int_S \left(\frac{d\nu_1}{d\mu} \right)^2 d\mu \tag{1.3}$$

if $\nu(S^m) = 0$, $\nu_1 \ll \mu$ and $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$, and we define $I_m(\nu) = \infty$ otherwise. I_m is convex. To be more precise we proved the following result. Denote by $\mathbb{P} := \mu^{\otimes \mathbb{N}}$. 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 $\{M_n^m\}_{n \in \mathbb{N}}$ with speed $\frac{n}{b_n^2}$ and rate $I_m(\cdot)$. If the Banach space E is of type 2, then we get the upper bound with the same speed and rate. Notice that in Eichelsbacher and Schmock (1996, Theorem 1.17), the MDP result is proved in an even finer topology under an additional growth restriction on $\varphi \in \Phi$ in terms of the norms of the corresponding completely μ-degenerate functions of the Hoeffding decomposition. The MDP result in the case $m = 1$ is an improvement over de Acosta (1994b, Theorem 3.1).

An obvious application of the contraction principle (cf. Dembo and Zeitouni (1993, 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 and a MDP for $U_n^m(h, \mu)$ and for $(M_n^m)(h) = \int h \, dM_n^m$ with corresponding speeds and rate functions. Without loss of generality we will assume that the h_i are symmetric in their arguments. Thus we obtain a LDP for a finite number of Banach space valued U-statistics and moderate U-statistics (cf. Eichelsbacher and Schmock (1996, Theorem 1.21, 1.23)). But in nonparametric statistics we need some kind of uniform estimates of $(L_n^m - \mu^{\otimes m})(h)$ over a not necessarily finite class of functions $\mathcal{H}$.

In this paper we also consider another class of processes. A V-statistic of order m with kernel h is defined as

$$V_n^m(h,\mu) = \int_{S^m} h\, dL_n^{\otimes m}.$$

A V-process based on μ and indexed by $\mathcal{H}$ is the set $\{V_n^m(h,\mu) : h \in \mathcal{H}\}$.

Denote by $\nu(h) := \int h\, d\nu$ for every $\nu \in \mathcal{M}(S^m)$ and $h : S^m \to \mathbb{R}$. Throughout the paper we assume that the class $\mathcal{H}$ is countable or that the processes $\{(L_n^m - \mu^{\otimes m})(h), h \in \mathcal{H}\}$ (or $\{(M_n^m(h), h \in \mathcal{H}\}$) and $\{(L_n^{\otimes m} - \mu^{\otimes m})(h), h \in \mathcal{H}\}$ (or $\{M_n^{\otimes m}(h), h \in \mathcal{H}\}$), respectively, are separable to avoid measurability problems.

We have to introduce some more notations. Given a class of functions $\mathcal{H} \subset B(S^m, \mathbb{R})$, 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 nonseparable Banach space if $\mathcal{H}$ is infinite. Now every signed measure $\nu \in \mathcal{M}(S^m)$ of finite variation corresponds to an element $\nu(\mathcal{H})$ in the space $l_\infty(\mathcal{H})$ given by

$$\nu(\mathcal{H})(h) = \nu(h) = \int h\, d\nu$$

for all $h \in \mathcal{H}$. The aim of this paper is to prove a LDP for $(L_n^m - \mu^{\otimes m})(\mathcal{H})$ and $(M_n^m)(\mathcal{H})$ and for $(L_n^{\otimes m} - \mu^{\otimes m})(\mathcal{H})$ and $(M_n^{\otimes m})(\mathcal{H})$ in $l_\infty(\mathcal{H})$. The *nonseparability* of $l_\infty(\mathcal{H})$ is the reason why such results do not follow easily via the contraction principle from the results of Eichelsbacher and Schmock (1996).

The content of the different sections is as follows. In Section 2 we consider a LDP and a MDP for U-processes for uniformly bounded classes $\mathcal{H}$, which satisfy "blockwise" empirical process conditions analogously to the conditions obtained in (Wu 1994). Therefore we will rewrite the U-empirical measure L_n^m as a dependent mean of i.i.d. means, using the existence of almost regular partitions of complete hypergraphs due to Baranyai (1975). The i.i.d. mean are built by "blocks" $(X_{(i-1)m+1}, \ldots, X_{(i-1)m+m})$. The advantage of the Baranyai representation is that there are much less dependent means in comparison to the well known Hoeffding formula in (Hoeffding 1963). As in Wu (1994) we prove a LDP and a MDP for certain uniformly bounded classes $\mathcal{H}$ using some well established techniques in the theory of large deviations (see Dembo and Zeitouni (1993), Wu (1994)) and an isoperimetric inequality of Talagrand (1994). Furthermore we extend the results to unbounded classes using well established truncation techniques and a truncation method based on an isoperimetric inequality due to Ledoux (1992). Serfling and Wang (1997) kindly sent me a preprint about large deviations, proving an LDP result for U-processes for Polish space valued random variables. Our result is more general and our methods are quite different from theirs.

In Section 3 we consider a MDP for non-degenerate U-processes indexed by uniformly bounded Vapnik-Červonenkis classes. Since we are using Hoeffding's decomposition of a U-statistic let us state it here together with some notations. The operator $\pi^{\mu}_{k,m} = \pi_{k,m}$ acts on $\mu^{\otimes m}$ -integrable symmetric functions $h:\, S^m \to \mathbb{R}$ 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

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

A function h is called μ-canonical or completely degenerate if $\int h(x_1, \dots, x_m) d\mu(x_i) = 0$ for all $1 \le i \le m$. Note that $\pi_{k,m} h$ is a μ-canonical function of k variables. 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 into a sum of μ-canonical U-statistics of different orders. For all $\mu^{\otimes m}$-integrable functions $h:\, S^m \to \mathbb{R}$ the following relation holds true

$$U_n^m(h, \mu) = \sum_{k=0}^{m} \binom{m}{k} U_n^k(\pi_{k,m} h, \mu). \tag{1.4}$$

Note that the centered empirical process is a canonical U-process of order 1. To prove the MDP in the non-degenerate case for $\{U_n^m(h, \mu) :\, h \in \mathcal{H}\}$, we have to prove that $\{U_n^1(\pi_{1,m} h, \mu) :\, h \in \mathcal{H}\}$ satisfies a MDP and that the processes $\{\frac{n}{b_n} U_n^k(\pi_{k,m} h, \mu) :\, h \in \mathcal{H}\} \to 0$ in probability exponentially fast with speed $\frac{n}{b_n^2}$ for $2 \le k \le m$. The first condition is equivalent to the class $\{\pi_{1,m} h :\, h \in \mathcal{H}\}$ satisfying the conditions of Wu (1994, Theorem 2), which are necessary and sufficient for the MDP. Therefore only the second condition must be dealt with. We will use a Bernstein-type inequality due to Arcones and Giné (1994) and to Arcones (1995).

In Section 4 our results are applied to obtain LDP and MDP results for empirical functions of U-statistics structure, for Liu's simplicial depth process and for classes of functions which are uniformly Hölder.

2. Moderate and large deviations via blocking

We only consider classes of symmetric functions. Define for each kernel function $h:\, S^m \to \mathbb{R}$ in a class $\mathcal{H}$ the function $\tilde{h}:\, S^m \to \mathbb{R}$ by

$$\tilde{h}(x_1, \dots, x_n) = \frac{1}{k} \sum_{j=0}^{k-1} h(x_{jm+1}, \dots, x_{jm+m}), \qquad k = \left\lfloor \frac{n}{m} \right\rfloor.$$

Then, by Hoeffdings formula (Borovskich and Koroljuk 1994, p. 34)

$$L_n^m(h) = \frac{1}{n!} \sum_{\pi \in P_n} \tilde{h}(X_{\pi(1)}, \dots, X_{\pi(n)}), \tag{2.1}$$

where P_n denotes the set of all permutations of $\{1,\dots,n\}$. Moreover let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables defined on $(S^{\mathbb{N}},\mathcal{S}^{\mathbb{N}},\mathbb{P})$ with $\mathbb{P}(\varepsilon_i=\pm 1)=1/2$, independent of the underlying sequence $\{X_i\}_{i\in\mathbb{N}}$. Introduce the following quantities:

$$H(n) := E\|\sum_{i=1}^{\lfloor n/m\rfloor} \varepsilon_i\, h(Y_i)\|_{\mathcal{H}}, \tag{2.2}$$

where

$$Y_i := (X_{(i-1)m+1},\dots,X_{(i-1)m+m}),$$

and

$$H(n,\eta) := E\|\sum_{i=1}^{\lfloor n/m\rfloor} \varepsilon_i\, h(Y_i)\|_{\mathcal{H}_\eta},$$

where

$$\mathcal{H}_\eta := \{f-g:\ f,g\in\mathcal{H} \quad \text{and} \quad d_2(f,g) := \|f-g\|_{L_2(\mu^{\otimes m})}\le \eta\}.$$

2.1. **The case of uniformly bounded $\mathcal{H}$.** We assume that

$$0\le h\le 1 \ \text{ for all } \quad h \quad \text{in} \quad \mathcal{H}. \tag{2.3}$$

We will establish the following theorems.

Theorem 2.4 (Moderate deviations for U-processes and V-processes)**.**
Assume (2.3) and let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence satisfying

$$\lim_{n\to\infty}\frac{b_n}{n}=0 \quad \textit{and} \quad \lim_{n\to\infty}\frac{n\log n}{b_n^2}=0.$$

If $(\mathcal{H},d_2)$ is totally bounded and

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

then the laws of $(L_n^m-\mu^{\otimes m})(\mathcal{H})$ and of $(L_n^{\otimes m}-\mu^{\otimes m})(\mathcal{H})$, respectively, satisfy the MDP on the space $l_\infty(\mathcal{H})$ with speed n/b_n^2 and with the good rate function

$$I_{\mathcal{H}}(H)=\inf\{I_m(\nu):\nu\in\mathcal{M}(S^m) \quad \textit{and} \quad \nu(\mathcal{H})=H\},\quad H\in l_\infty(\mathcal{H}), \tag{2.5}$$

where $I_m(\cdot)$ is given by (1.3).

Theorem 2.6 (Large deviations for U-processes and V-processes)**.**
Under the assumption (2.3) we get: If $(\mathcal{H},d_2)$ is totally bounded and

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{H(n,\eta)}{n}=0, \tag{2.7}$$

then the laws of $L_n^m(\mathcal{H})$ and of $L_n^{\otimes m}(\mathcal{H})$, respectively, satisfy the LDP on the space $l_\infty(\mathcal{H})$ with speed $1/n$ and with the good rate function

$$J_{\mathcal{H}}(H)=\inf\{J_m(\nu):\nu\in\mathcal{M}_1(S^m) \quad \textit{and} \quad \nu(\mathcal{H})=H\},\quad H\in l_\infty(\mathcal{H}), \tag{2.8}$$

where $J_m(\cdot)$ is given by (1.2).

Of course, an important question is, which classes of functions $\mathcal{F}$ satisfy the assumptions of our theorems. We will discuss this and some examples in Section 4. In the case $m = 1$ our results are parts of Theorem 1 and Theorem 2 in Wu (1994).

We will use the following isoperimetric inequality for empirical processes established by Talagrand (1994, Theorem 3.5):

Lemma 2.9 (Talagrand). *Let $m = 1$ and assume (2.3). Set $\sigma := \sup_{f\in\mathcal{H}}(\int(f-\mu(f))^2\, d\mu)^{1/2}$ and $S(n) := n\sigma^2 + H(n)$, where $H(n)$ is given in (2.2). Then for some universal constant K we have for all $t \geq K\, H(n)$:*

$$P\biggl(||\sum_{i=1}^n f(X_i) - n\mu(f)||_{\mathcal{H}} \geq t\biggr) \leq \exp(-\varphi_{K,S(n)}(t)),$$

where the function $\varphi_{K,S(n)}(t)$ for $t \geq 0$ is given by

$$\varphi_{K,S(n)}(t) = \frac{t^2}{K^2 S(n)}, \quad \textit{if} \quad t \leq KS(n),$$

$$\varphi_{K,S(n)}(t) = \frac{t}{K}\biggl(\log\frac{e\,t}{K\,S(n)}\biggr)^{1/2}, \quad \textit{if} \quad t \geq KS(n).$$

We also use the existence of almost regular partitions of completely hypergraphs due to (Baranyai 1975, Theorem 1). Let X be a finite set with $|X| = n$ (the set of vertices) and $\mathcal{E}$ the set of all m-element subsets of X (the edges). Denote by K_n^m this hypergraph which is called the complete m-uniform hypergraph of n vertices. The valency of a vertex x is defined by

$$v_{(X,\mathcal{E})}(x) = \sum_{i: x\in E_i} 1,$$

where $\mathcal{E} = \{E_1, \dots, E_{\binom{n}{m}}\}$. A subset $\mathcal{G} \subset \mathcal{E}$ is a 1-factor if $v_{(X,\mathcal{G})}(x) = 1$ for every $x \in X$. The hypergraph $(X,\mathcal{E})$ is called *1-factorizable* if there are disjoint 1-factors $\mathcal{G}_j$ such that $\mathcal{E} = \cup_j \mathcal{G}_j$. Note that the following result had been an open problem for more than 100 years.

Lemma 2.10 (Baranyai). *If $m|n$ then K_n^m is 1-factorizable.*

Proof of Theorem 2.4*:* Notice that for finite $\mathcal{H}$ the result is a direct consequence of Theorem 1.17 in Eichelsbacher and Schmock (1996). 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 $\varphi : S^m \to E$, where E is a separable real Banach space of type 2 and each φ satisfies the condition

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

for every $\alpha > 0$. We obtain the result for all $\{b_n\}_{n\in\mathbb{N}}$ satisfying (1.1). In Eichelsbacher and Schmock (1996) other conditions for φ related to Ledoux (1992) are considered.

The level sets $K(I_{\mathcal{H}}, L) := \{H \in l_\infty(\mathcal{H}) : I_{\mathcal{H}}(H) \leq L\}, L \geq 0$, are compact in $l_\infty(\mathcal{H})$ by the total boundedness of $(\mathcal{H}, d_2)$ and the theorem of Arcelà-Ascoli (cf. Dunford and Schwartz (1967, Theorem 5, Sec. IV.6)). Therefore we only have to check that a level set is a bounded and closed subset of the set of all bounded and continuous functionals on $(\mathcal{H}, d_2)$ which is easy to see and omitted.

Along the lines of the proof of Theorem 1 and 2 in Wu (1994), we will check, that the MDP is established if we can check

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}(||M_n^m(\cdot)||_{\mathcal{H}_\eta} \geq \delta) = -\infty \tag{2.11}$$

for all $\delta > 0$. It is quite standard (for example see Theorem 1.7, Eichelsbacher and Schmock (1996), and its proof) to deduce the upper and lower bound from (2.11).

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

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

is a subset of $l_\infty(\mathcal{H}) \setminus C$. Since the level set $K(I_{\mathcal{H}}, r)$ is compact and $C \cap K(I_{\mathcal{H}}, r) = \emptyset$ there exists a finite subset N of $K(I_{\mathcal{H}}, r)$ such that $U := \bigcup_{F \in N} U(F, \delta_F)$ covers $K(I_{\mathcal{H}}, r)$. We arrive at

$$\mathbb{P}((M_n^m)(\mathcal{H}) \in C) \leq \mathbb{P}\big((M_n^m)(\mathcal{H}) \in l_\infty(\mathcal{H}) \setminus \bigcup_{F \in N} U(F, 2\delta_F)\big).$$

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 $f \in \mathcal{H}$ there exists a $g \in \mathcal{H}^\eta$ such that $d_2(f, g) < \eta$. Denote by $I_\eta(\cdot)$ the rate function corresponding to $\mathcal{H}^\eta$ defined by (2.5) and by $p_\eta(F)$ the restriction of $F \in l_\infty(\mathcal{H})$ to the net $\mathcal{H}^\eta$. Now for every $\delta > 0$ one can find a $\eta' > 0$ such that if $d_2(f, g) < \eta'$ we get $|F(f) - F(g)| < \delta$ for each $F \in N$, since on the compact level set the F are continuous. Thus for every $\delta > 0$ there is a η with $0 < \eta < \eta'$ such that for all $F \in l_\infty(\mathcal{H})$, if

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

and

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

then $F \in \bigcup_{F \in N} U(F, 2\delta_F)$. We obtain

$$\mathbb{P}((M_n^m)(\mathcal{H}) \in C) \leq \\ \mathbb{P}\big((M_n^m)(\mathcal{H}^\eta) \in l_\infty(\mathcal{H}) \setminus \bigcup_{F \in N} U(p_\eta(F), \delta_F)\big) + \mathbb{P}\big(||(M_n^m)(\cdot)||_{\mathcal{H}_\eta} \geq \delta\big). \tag{2.12}$$

By Theorem 1.17 in Eichelsbacher and Schmock (1996) we have

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\big((M_n^m)(\mathcal{H}^\eta) \in l_\infty(\mathcal{H}) \setminus \bigcup_{F\in N} U(p_\eta(F), \delta_F)\big) \le$$
$$- \inf\{I_\eta(p_\eta(F)) : \, p_\eta(F) \in l_\infty(\mathcal{H}) \setminus \bigcup_{F\in N} U(p_\eta(F), \delta_F)\} \le -r.$$

Using (2.11) for the second term in (2.12) 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 $I_{\mathcal{H}}(O) < \infty$. 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 is $\eta > 0$ such that

$$\mathbb{P}((M_n^m)(\mathcal{H}^\eta) \in U(p_\eta(G), \delta_G)) \le \mathbb{P}(||M_n^m(\cdot)||_{\mathcal{H}_\eta} \ge \delta_G) + \mathbb{P}((M_n^m)(\mathcal{H}) \in O).$$

We apply the lower bound of Theorem 1.17 Eichelsbacher and Schmock (1996) to the finite subnet $\mathcal{H}^\eta$:

$$-I_\eta(G) \le -I_\eta(U(p_\eta(G), \delta_G))) \le \liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}((M_n^m)(\mathcal{H}^\eta) \in U(p_\eta(G), \delta_G)).$$

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 (1993, Lemma 4.1.6 (a)) that

$$\lim_{\eta\to 0} I_\eta(G) = \lim_{\eta\to 0} \inf\{I_m(\nu) : \, \nu \in \mathcal{M}_1(S^m), \nu(\mathcal{H}^\eta) = G\} = I_{\mathcal{H}}(G),$$

thus together with (2.11) we get the lower bound.

Proof of (2.11):
To prove the key estimation (2.11), we will apply the result of Talagrand (Lemma 2.9) combined with the factorization result of Baranyai (Lemma 2.10). Therefore consider $n \in \mathbb{N}$ which is divisible by m. P_n denotes the group of all permutations of $\{1, \dots, n\}$. For $\sigma \in P_n$ define

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

This map is measurable and $\mathbb{P}L_{m,n,\sigma}^{-1}$ does not depend on σ, because $\{X_i\}_{i\in\mathbb{N}}$ are independent and identically distributed. We write $L_{m,n}$ for $L_{m,n,\sigma}$, when σ is the identity on $\{1, \dots, n\}$. With Lemma 2.10 there exists a subset P'_n of P_n consisting of $m!\binom{n-1}{m-1}$ permutations such that for every m-tuple $(i_1, \dots, i_m)$ of distinct elements of $\{1, \dots, n\}$ there exists exactly one pair $(\sigma, j) \in (P'_n \times \{0, 1, \dots, n/m - 1\})$ such that $(i_1, \dots, i_m) = (\sigma(jm+1), \dots, \sigma(jm+m))$. (There are $\binom{n}{m}$ subsets $\{i_1, \dots, i_m\}$ of $\{1, \dots, n\}$ with m elements. The problem in wether they can be divided into

$\binom{n}{m}\frac{m}{n}$ disjoint classes such that each class is a partition of $\{1,\dots,n\}$.) Now the U-empirical measure can be rewritten as

$$L_n^m = \frac{1}{m!\binom{n-1}{m-1}} \sum_{\sigma \in P_n'} L_{m,n,\sigma}. \tag{2.13}$$

We set $\mathcal{H}_\eta' := \{(h+1)/2, h \in \mathcal{H}_\eta\}$. Of course, this set satisfies the assumption of Lemma 2.9. We define

$$H'(n,\eta) := E|| \sum_{i=1}^{\lfloor n/m \rfloor} \varepsilon_i \, h(Y_i) ||_{\mathcal{H}_\eta'}$$

and obtain

$$|2H'(n,\eta) - H(n,\eta)| \le E\Big|\sum_{j=1}^n \varepsilon_j\Big| \le \sqrt{n}. \tag{2.14}$$

Analogously to the proof in Wu (1994) we obtain for η sufficiently small and n large enough $KH'(n,\eta) \le b_n \delta/2$ using the assumption $\lim_{\eta\to 0}\lim_{n\to\infty} \frac{H(n,\eta)}{b_n} = 0$. K is the universal constant in Lemma 2.9. We arrive at

$$\mathbb{P}(||M_n^m(\cdot)||_{\mathcal{H}_\eta} \ge \delta) = \mathbb{P}(||M_n^m(\cdot)||_{\mathcal{H}_\eta'} \ge \delta/2).$$

Since L_n^m is an average of the form (2.13), we arrive at (applying Lemma 2.9):

$$\begin{aligned}\mathbb{P}(||M_n^m(\cdot)||_{\mathcal{H}_\eta'} \ge \delta/2) &\le m!\binom{n-1}{m-1}\mathbb{P}(||\frac{n}{b_n}L_{m,n}(\cdot)||_{\mathcal{H}_\eta'} + \eta \ge \delta/2) \\ &\le m!\binom{n-1}{m-1}\exp(-\varphi_{K,S(n)}(b_n(\delta/2-\eta)),\end{aligned}$$

where $S(n) = [n/m](\sigma^2)' + H'(n,\eta)$ and

$$(\sigma^2)' := \sup_{h\in\mathcal{H}_\eta'} \int ((h-\mu^{\otimes m}(h))^2 d\mu^{\otimes m})^{1/2}.$$

If η is small enough, the function $\varphi_{K,S(n)}(b_n(\delta/2-\eta))$ is given by the expression

$$\frac{(b_n(\delta/2-\eta))^2}{K^2 S(n)}.$$

Remark that by the definition of $\mathcal{H}_\eta'$ we get $(\sigma^2)' = const.\eta^2$. Using $S(n) = [n/m]const.\eta^2 + H'(n,\eta)$ the assumption on $H(n,\eta)$ and (2.14), we obtain for η small enough

$$\lim_{\eta\to 0}\lim_{n\to\infty} \frac{n}{b_n^2}\varphi_{K,S(n)}(b_n(\delta/2-\eta)) = \infty.$$

Moreover

$$\limsup_{n\to\infty} \frac{n}{b_n^2}\log m!\binom{n-1}{m-1} = 0$$

by the assumption for $\{b_n\}_{n\in\mathbb{N}}$. Thus (2.11) follows for every n which is divisible by m. Note that the assumption of uniform boundedness of the class $\mathcal{H}$ yields

$$||L_n^{\otimes m}(\cdot) - L_n^m(\cdot)||_{\mathcal{H}} \le 1 - \binom{n}{m}\frac{m!}{n^m} \le \frac{n^m-(n-m)^m}{n^m} \le \frac{m^2}{n}, \tag{2.15}$$

and therefore

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(\frac{n}{b_n} ||L_n^{\otimes m}(\cdot) - L_n^m(\cdot)||_{\mathcal{H}} \geq \delta) = -\infty$$

for each $\delta > 0$. Thus in this case the theorem follows by Dembo and Zeitouni (1993, Theorem 4.2.13).

For general $n \in \mathbb{N}$ satisfying $n \geq m$, we define $L_n^m = L_{m[n/m]}^m$. Since $||L_n^{\otimes m}(\cdot) - L_{m[n/m]}^{\otimes m}(\cdot)||_{\mathcal{H}} \leq 1 - \frac{(n-m)^m}{n^m} \leq \frac{m^2}{n}$, it follows with (2.15) that $||L_n^{\otimes m}(\cdot) - L_n^m(\cdot)||_{\mathcal{H}} \leq 2m^2/(n-m+1)$. Using Dembo and Zeitouni (1993, Theorem 4.2.13) once more we get the result. $\square$

Proof of Theorem 2.6*:* Again notice that for finite $\mathcal{H}$ the result is a direct consequence of Theorem 1.7 in Eichelsbacher and Schmock (1996). Since this result is true even in a finer topology, we obtain the statement for any finite collection $\mathcal{H}$ of $\mathcal{S}^{\otimes m}$-$\mathcal{E}$-measurable functions $\varphi : S^m \to E$, where E is an arbitrary real Banach space and φ satisfies the condition

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

for each $\alpha > 0$.

The level sets $K(J_{\mathcal{H}}, L) := \{H \in l_\infty(\mathcal{H}) : J_{\mathcal{H}}(H) \leq L\}, L \geq 0$, are compact in $l_\infty(\mathcal{H})$ by the same argument as in the proof of Theorem 2.6.
The LDP is established if we can check

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(||L_n^m(\cdot) - \mu^{\otimes m}(\cdot)||_{\mathcal{H}_\eta} \geq \delta) = -\infty \tag{2.16}$$

for all $\delta > 0$. Now it is clear how to deduce the upper and lower bound from (2.16). We omit the proof.

Proof of (2.16):
To prove (2.16), we will again apply Lemma 2.9 combined with Lemma 2.10. Consider $n \in \mathbb{N}$ which is divisible by m. The proof follows the lines of the proof of Theorem 2 in Wu (1994). Consider the class $\mathcal{H}'_\eta$. For η sufficiently small and n large enough we obtain $KH'(n, \eta) \leq n\delta/2$ using the assumption $\lim_{\eta\to 0} \lim_{n\to\infty} \frac{H(n,\eta)}{n} = 0$ and (2.14). K is the universal constant in Lemma 2.9. Again we arrive at

$$\mathbb{P}(||L_n^m(\cdot) - \mu^{\otimes m}(\cdot)||_{\mathcal{H}_\eta} \geq \delta) = \mathbb{P}(||L_n^m(\cdot) - \mu^{\otimes m}(\cdot)||_{\mathcal{H}'_\eta} \geq \delta/2).$$

The representation (2.13) of L_n^m and Lemma 2.9 yield:

$$\begin{aligned}\mathbb{P}(||L_n^m(\cdot) - \mu^{\otimes m}||_{\mathcal{H}'_\eta} \geq \delta/2) &\leq m! \binom{n-1}{m-1} \mathbb{P}(||L_{m,n}(\cdot)||_{\mathcal{H}'_\eta} + \eta \geq \delta/2) \\ &\leq m! \binom{n-1}{m-1} \exp(-\varphi_{K,S(n)}(n(\delta/2 - \eta)),\end{aligned}$$

where $S(n) = [n/m](\sigma^2)' + H'(n, \eta)$. If η is small enough, the function $\varphi_{K,S(n)}(b_n(\delta/2 - \eta))$ is now given by the expression

$$\frac{n(\delta/2-\eta)}{K}\left(\log\frac{en(\delta/2-\eta)}{KS(n)}\right)^{1/2}.$$

Using $S(n) = [n/m] const.\eta^2 + H'(n, \eta)$, the assumption on $H(n, \eta)$ and (2.14), we obtain for η small enough

$$\lim_{\eta\to 0}\lim_{n\to\infty}\frac{1}{n}\varphi_{K,S(n)}(b_n(\delta/2-\eta)) = \infty.$$

Thus (2.16) follows. For general $n \in \mathbb{N}$ satisfying $n \geq m$ we can adapt the arguments of the proof of Theorem 2.6. □

2.2. **Extension to unbounded functions.** Consider a class $\mathcal{H}$ of real measurable functions on S^m with $||h(x_1, \dots, x_m)|| < \infty$ for every $(x_1, \dots, x_m) \in S^m$. Thus the maps $\mathcal{H} \ni h \mapsto h(X_{i_1}, \dots, X_{i_m})$ define random elements in the space $l_\infty(\mathcal{H})$. Via a well established truncation method we get the following extension of Theorem 2.6:

Theorem 2.17 (Large deviations for U-processes of oder m). *Assume that $\mathcal{H}$ is a class of functions in $L_2(S^m, \mu^{\otimes m})$ such that*

$$\int_{S^m}\exp(\alpha\sup_{h\in\mathcal{H}}||h||)\,d\mu^{\otimes m} < \infty \tag{2.18}$$

for every $\alpha > 0$. If $(\mathcal{H}, d_2)$ is totally bounded and

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

then the laws of $L_n^m(\mathcal{H})$ satisfy the LDP on $l_\infty(\mathcal{H})$ with speed $1/n$ and with the rate function $J_{\mathcal{H}}(\cdot)$ given by (2.8).

Proof. Consider the class $\mathcal{H}_N := \{h_N := (h \vee (-N)) \wedge N : \ h \in \mathcal{H}\}$. We claim that the class $\mathcal{H}_N$ satisfies the assumptions of Theorem 2.6. Since $d_2(f_N, g_N) \leq d_2(f, g)$, $\mathcal{H}_N$ is totally bounded in $L_2(S^m, \mu^{\otimes m})$. If $H_N(n, \eta)$ denotes the quantity $H(n, \eta)$ for the class $\mathcal{H}_N$ we have to check

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{H_N(n,\eta)}{n} = 0, \tag{2.19}$$

which is a little involved. For details see the proofs of Theorem 1 and 4 in Wu (1994). The arguments can be applied for the "blocking" mean $L_{m,n}(\mathcal{H})$, introduced in the proof of Theorem 2.6. With Wu (1994, Theorem 1) we obtain that $(L_{m,n} - \mu^{\otimes m})(\mathcal{H}) \to 0$ in probability in $l_\infty(\mathcal{H})$. Via the steps of the proof of Wu (1994, Theorem 4), we get (2.19). Thus Theorem 2.6 can be applied to $\mathcal{H}_N$ for every $N \in \mathbb{N}$. By Dembo and Zeitouni (1993, Theorem 4.2.13) we only have to check

$$\lim_{N\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sup_{h\in\mathcal{H}}||L_n^m(h-h_N) - \mu^{\otimes m}(h-h_N)|| \geq \delta) = -\infty$$

for all $\delta > 0$. Therefore we can use Hoeffdings formula (2.1) and (Hoeffding 1963, (5.2)) (see also Eichelsbacher (1996)) to obtain

$$\mathbb{P}(\sup_{h\in\mathcal{H}} ||L_n^m(h-h_N) - \mu^{\otimes m}(h-h_N)|| \geq \delta)$$
$$\leq \exp(-[n/m]\delta\alpha)\left(\int_{S^m} \exp(\alpha \sup_{h\in\mathcal{H}} ||h-h_N-\mu^{\otimes m}(h-h_N)||)\right)^{[n/m]},$$

which gives us the result for $L_n^m(\mathcal{H})$ using the dominated convergence theorem and the moment condition (2.18). □

Notice that for a seperable Banach space the condition (2.18) is the same as that of the LDP result for Banach space valued U-statistics in Eichelsbacher and Schmock (1996).

Corollary 2.20 (Large deviations for V-processes of oder m)**.** *Assume that $\mathcal{H}$ is a class of functions satisfying the assumptions of Theorem 2.17. Assume moreover that there exists at least one $\alpha_{\mathcal{H}} > 0$ such that*

$$\int_{S^m} \exp(\alpha_{\mathcal{H}} \sup_{h\in\mathcal{H}} ||h\circ\pi_\tau||)\, d\mu^{\otimes} < \infty \tag{2.21}$$

for every map $\tau : \{1,\dots,m\} \to \{1,\dots,m\}$, where $\pi_\tau : S^m \to S$ is defined by $\pi_\tau(s) = (s_{\tau(1)},\dots,s_{\tau(m)})$ for every $s = (s_1,\dots,s_m) \in S^m$. Then the laws of $L_n^{\otimes m}(\mathcal{H})$ satisfy the LDP on $l_\infty(\mathcal{H})$ with speed $1/n$ and with the rate function $J_{\mathcal{H}}(\cdot)$ given by (2.8).

Proof. Notice that for every $n \geq m$ and every $h \in \mathcal{H}$ we can define a function h_n such that $L_n^{\otimes m}(h) = L_n^m(h_n)$. One possibility is to define, for all $x = (x_1,\dots,x_m) \in S^m$ and $h \in \mathcal{H}$,

$$h_n(x) = \sum_{j=1}^m \frac{n_{(j)}}{n^m} \sum_{\tau\in\mathcal{T}_j} h(x_{\tau(1)},\dots,x_{\tau(m)}), \tag{2.22}$$

where $\mathcal{T}_j$ denotes the set of all surjectiv maps $\tau : \{1,\dots,m\} \to \{1,\dots,j\}$ with $\tau(1) = 1$ and $\tau(k) \leq 1 + \max\{\tau(1),\dots,\tau(k-1)\}$ for all $k \in \{2,\dots,m\}$. (cf. Eichelsbacher and Schmock (1996, (2.24))). We only have to check

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sup_{h\in\mathcal{H}} ||L_n^m(h_n) - L_n^m(h)|| \geq \delta) = -\infty$$

for every $\delta > 0$. Define $\alpha = (m^{m-1}+m^2)^{-1}\alpha_{\mathcal{H}}$ with $\alpha_{\mathcal{H}}$ as in (2.21). By using the exponential Chebychev inequality and independence,

$$\begin{aligned}
&\mathbb{P}(\sup_{h\in\mathcal{H}} ||L_n^m(h_n) - L_n^m(h)|| \geq \delta)\\
&\quad\leq e^{-\alpha\delta n^2} E[\exp(\alpha n^2 \sup_{h\in\mathcal{H}} ||L_n^m(h_n) - L_n^m(h)||)]\\
&\quad\leq e^{-\alpha\delta n^2}\left(\int_{S^m} \exp(\alpha m n \sup_{h\in\mathcal{H}} ||h-h_n||)\, d\mu^{\otimes m}\right)^{\lfloor n/m\rfloor}.
\end{aligned} \tag{2.23}$$

Note that $1 - n_{(m)}/n^m \leq m^2/n$. Using (2.22), it follows that

$$n\|h(x) - h_n(x)\| \leq m^2\|h(x)\| + \sum_{j=1}^{m-1} \sum_{\tau \in \mathcal{T}_j} \|h \circ \pi_\tau(x)\|$$

for all $x \in S^m$. Using this estimate, $|\bigcup_{j=1}^{m-1} \mathcal{T}_j| \leq m^{m-1}$, Hölder's inequality and (2.21), it follows that the expectation in (2.23) is bounded by a constant which does not depend on n. Hence we are done (cf. Eichelsbacher and Schmock (1996, Proof of Theorem 1.10(b))). $\square$

As in Wu (1994), for the moderate estimations we can use a truncation method established by Ledoux (1992, Thèorèm) based on isoperimetric inequality (we state the result for the Banach space $l_\infty(\mathcal{H}_\eta)$).

Lemma 2.24 (Ledoux). *Let $Y_i = (X_{(i-1)m+1}, \dots, X_{(i-1)m+m})$ and let $\{b_n\}_{n\in\mathbb{N}}$ satisfies (1.1) and the auxiliary condition*

$$b_{nk} \leq A k^{-\delta} b_n \tag{2.25}$$

for some $A \geq 1$ and some $0 < \delta < 1$ (b_n cannot be too near n). If $\frac{n}{b_n^2}(L_{m,n} - \mu^{\otimes m})(\mathcal{H}) \to 0$ in probability in $l_\infty(\mathcal{H})$ and if there exists a constant $M > 0$ such that for all $u > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log\bigl(n\mathbb{P}(||\delta_{Y_i}(\cdot)||_{\mathcal{H}_\eta} > u)\bigr) \leq -\frac{u^2}{M},$$

then there exists some constant C depending only on M and A and δ such that

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\bigl(||\frac{n}{b_n} L_{m,n}(\cdot)||_{\mathcal{H}_\eta} \geq \varepsilon\bigr) \leq -\frac{\varepsilon}{C\sigma^2}, \tag{2.26}$$

where σ^2 is defined as in the proof of Theorem 2.6.

Theorem 2.27 (Moderate deviations for U-processes of oder m)**.** *Assume that $\mathcal{H}$ is a class of functions in $L_2(S^m, \mu^{\otimes m})$ such that there exists a $M > 0$ such that for all $u > 0$*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log(n\mathbb{P}(||\delta_{Y_i}(\cdot)||_{\mathcal{H}_\eta} > u)) \leq -\frac{u^2}{M}.$$

If $\mathcal{H}$ is totally bounded and

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

and $\{b_n\}_{n\in\mathbb{N}}$ satisfies (1.1) and (2.25), then the laws of $(L_n^m - \mu^{\otimes m})(\mathcal{H})$ satisfy the MDP on $l_\infty(\mathcal{H})$ with speed n/b_n^2 and with the rate function $I_{\mathcal{H}}(\cdot)$ given by (2.5).

Proof. Let us only mention what has to be changed in Wu (1994, Theorem 5). The MDP is established if we can check (2.16). To do so, we use again the representation of L_n^m due to Baranyai. The only change from the proof of Theorem 2.4 is now to establish

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}(||\frac{n}{b_n} L_{m,n}(\cdot)||_{\mathcal{H}'_\eta} + \eta \geq \delta/2) = -\infty.$$

without using Lemma 2.9. But this estimation follows from (2.26) using the fact that σ^2 is smaller than η^2. We can apply the Lemma, since our assumptions on $\mathcal{H}$ imply via the proof of Theorem 2 in Wu (1994), that $\frac{n}{b_n^2}(L_{m,n} - \mu^{\otimes m})(\mathcal{H}) \to 0$ in probability in $l_\infty(\mathcal{H})$. $\square$

3. Moderate deviation result for VC classes of functions

Here we will prove MDPs for non-degenerate U-processes indexed by a countable class $\mathcal{H}$ which satisfies some conditions which will be given in terms of metric entropy. Especially we consider the case when $\mathcal{H}$ is a uniformly bounded Vapnik-Červonenkis subgraph class of functions. As mentioned in the introduction, we will apply the results of Wu (1994) to the linear part of the Hoeffding decomposition and will handle the rest by means of different types of Bernstein-type inequalities for uniformly bounded μ-canonical processes proved in Arcones and Giné (1994) and Arcones (1995). The conditions on $\mathcal{H}$ seem to be more natural than the conditions via the blocking approach. Nevertheless let us mention that this approach is not applicable for the large deviation scale.

To state the results we have to introduce some more notations. Given a pseudometric space (T, d), the ε-covering number $N(\varepsilon, T, d)$ is defined as

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

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 us recall the definition of a Vapnik-Čhervonenkis (VC for short) subgraph class of functions. Let S be a set, $\mathcal{C}$ a class of subsets of S and A be a subset of S with cardinality k. $\mathcal{C}$ shatters A if each subset of A is the trace of an element of $\mathcal{C}$, i.e. $\#(\mathcal{C} \cap A) = 2^k$ where $\mathcal{C} \cap A = \{C \cap A : C \in \mathcal{C}\}$. $\mathcal{C}$ is called a VC class if there exists an integer $k \geq 1$ such that no subset A of S of cardinality k is shattered by $\mathcal{C}$, i.e. for every A in S with $\#(A) = k$ we have $\#(\mathcal{C} \cap A) < 2^k$. Examples in $\mathbb{R}^d$ include classes of all rectangles, all ellipoids, and all polyhedra of at most l sides (for any fixed l). If $\mathcal{G}$ is a finite-dimensional vector space of functions on S, then $\text{pos}(\mathcal{G}) := \{\{s : g(s) > 0\} : g \in \mathcal{G}\}$ is a VC class. 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 $h \in \mathcal{H}$: $\{(x,t) \in S^m \times \mathbb{R} : 0 \leq t \leq h(x_1, \ldots, x_m) \text{ or } h(x_1, \ldots, x_m) \leq t \leq 0\}$). Any finite-dimensional vector space of functions (e.g., polynomials of bounded degree on $\mathbb{R}^d$) is a VC subgraph class. Notice moreover, 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.

The envelope H of $\mathcal{H}$ is defined as $\sup_{h \in \mathcal{H}} |h|$. It is well known (Pollard (1984, Proposition II 2.5)) that if $\mathcal{H}$ is a VC subgraph class then there are finite constants A and v such that, for each probability measure μ with $\mu^{\otimes m}(H^2) < \infty$,

$$N_2(\varepsilon, \mathcal{H}, \mu) \leq A(\mu^{\otimes m}(H^2)^{1/2}/\varepsilon)^v,$$

where H denotes the envelope of $\mathcal{H}$. Consider the following different typ of classes $\mathcal{H}$:

Condition 3.1. *Let $\mathcal{H}$ be a measurable class of functions $h: S^m \to \mathbb{R}$ satisfying:*

(a) *$\mathcal{H}$ is uniformly bounded.*
(b) *There are $A > 0$ and $v < \infty$ such that $N_2(\varepsilon, \mathcal{H}, \mu) \leq (A/\varepsilon)^v$ for all probability measures μ.*

Let $\{X_i^{(k)} : i \in \mathbb{N}\}_{k=1}^m$ be i.i.d. copies of $\{X_i : i \in \mathbb{N}\}$ and $\{\varepsilon_i^{(k)} : i \in \mathbb{N}\}_{k=1}^m$ be independent copies of $\{\varepsilon_i : i \in \mathbb{N}\}$ (defined as in Section 2), independent of $\{X_i^{(k)} : i \in \mathbb{N}\}_{k=1}^m$.

Condition 3.2. *Let $\mathcal{H}$ be a measurable class of functions $h: S^m \to \mathbb{R}$ satisfying:*

(a) *$\mathcal{H}$ is uniformly bounded.*
(b) *There is a finite constant c such that*

$$E\big(||n^{1-m} \sum_{I_{m,n}} \varepsilon_{i_1}^{(1)} \varepsilon_{i_2}^{(2)} h(x_{i_1}^{(1)}, \dots, x_{i_m}^{(m)})||_{\mathcal{H}}\big) \leq c. \tag{3.3}$$

Notice that uniformly bounded VC subgraph classes $\mathcal{H}$ satisfy Condition 3.1 as well as Condition 3.2 (for Condition 3.2, see Arcones and Giné (1991, Proposition 2.6)). Condition 3.2 is of interest if one uses the decoupling inequality by de la Peña (1992) together with symmetrization techniques (cf. Ledoux and Talagrand (1991, Lemma 6.3)) to get a bound for $E(\exp|tU_n^m(h)|)$ in terms of the left hand side of (3.3). For details see Arcones (1995).

We will use the following two typs of Bernstein-type inequalities for U-processes: The first is proved in Arcones and Giné (1994, Theorem 3.2):

Lemma 3.4 (Arcones, Giné). *Let $\mathcal{H}$ be a measurable class of μ-canonical functions $h: S^m \to \mathbb{R}$ satisfying:*

(a) *$\mathcal{H}$ is uniformly bounded by 1.*
(b) *There is a Lebesgue integrable function $\lambda: (0,\infty) \to [0,\infty)$ such that for each probability measure ν*

$$(\log N_2(\varepsilon, \mathcal{H}, \nu))^{m/2} \leq \lambda(\varepsilon), \quad \varepsilon > 0.$$

Then there are constants c and c' such that for all $t > 0$ and $n \geq m$

$$\mathbb{P}\Big(||n^{-m/2} \sum_{(i_1,\dots,i_m) \in I_{m,n}} h(X_{i_1}, \dots, X_{i_m})||_{\mathcal{H}} \geq t\Big) \leq c\, \exp(-c'\, t^{2/m}). \tag{3.5}$$

The next result is proved in Arcones (1995, Proposition 4):

Lemma 3.6 (Arcones). *Let $\mathcal{H}$ be a measurable class of symmetric functions $h : S^m \to \mathbb{R}$ satisfying Condition 3.2 with $\sup_{h\in\mathcal{H}} |h(X)| \le b$. Then for each $t > 0$*

$$\mathbb{P}\Big(||n^{1/2} \sum_{k=2}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge t \Big) \le 2\ \exp\Big(- \frac{tn^{1/2}}{2^{m+5} m^{m+1} b\, c} \Big),$$

To prove the MDP for the U-process, by Hoeffding's decomposition it is enough to show that $\{U_n^m(\pi_{1,m}h, \mu) :\ h \in \mathcal{H}\}$ satisfies the MDP in $l_\infty(\mathcal{H})$ and that the processes $\{\frac{n}{b_n} U_n^k(\pi_{k,m}h, \mu) :\ h \in \mathcal{H}\} \to 0$ in probability exponentially fast with speed $\frac{n}{b_n^2}$ for $2 \le k \le m$. For the second part we will apply the Bernstein inequalities:

Lemma 3.7. *Let $\mathcal{H}$ be a class of symmetric functions satisfying Condition 3.1 or Condition 3.2 and let $\{b_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ satisfy (1.1). In each case we obtain*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\Big(||\frac{n}{b_n} \sum_{k=2}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \delta \Big) = -\infty \tag{3.8}$$

for all $\delta > 0$.

Proof. Assume that $\mathcal{H}$ satisfies Condition 3.1. Note that for any probability measure ν on S^k the quantity $||\pi_{k,m}h||_{L_2(\nu)}$ is dominated by the sum of 2^k L_2-distances. Thus the condition $N_2(\varepsilon, \mathcal{H}, \mu) \le (A/\varepsilon)^v$ (for all probability measures μ) implies that the classes of canonical functions $\{\pi_{k,m}h :\ h \in \mathcal{H}\}$ satisfy the second part of the conditions of Lemma 3.4. With (3.5) we get for each $2 \le k \le m$:

$$\begin{aligned}
&\mathbb{P}\Big(||\frac{n}{b_n} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \delta \Big) = \\
&\quad \mathbb{P}\Big(||\frac{1}{n^{k/2}} \sum_{(i_1,\dots,i_k)\in I_{k,n}} \pi_{k,m}h(X_{i_1}, \dots, X_{i_k})||_{\mathcal{H}} \ge \frac{b_n \delta n_{(k)}}{n^{1+k/2}} \Big) \\
&\qquad\qquad \le c_k \exp\Big(- c_k' \big(\frac{\delta b_n}{n} \frac{n_{(k)}}{n^{k/2}} \big)^{2/k} \Big).
\end{aligned}$$

This implies

$$\frac{n}{b_n^2} \log \mathbb{P}\Big(||\frac{n}{b_n} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \delta \Big) \le \frac{n}{b_n^2} \log c_k - c_k' \delta^{2/k} \big(\frac{n}{b_n} \big)^k,$$

and the right hand side decreases to $-\infty$ by the assumptions for $\{b_n\}_{n\in\mathbb{N}}$. Since

$$\Big\{ ||\frac{n}{b_n} \sum_{k=2}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \delta \Big\} \subset \bigcup_{k=2}^{m} \Big\{ \frac{n}{b_n} || \binom{m}{k} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \frac{\delta}{m} \Big\}$$

Dembo and Zeitouni (1993, Lemma 1.2.15) yields (3.8).
Now assume that $\mathcal{H}$ satisfies Condition 3.2. We can apply Lemma 3.6 and obtain immediately

$$\mathbb{P}\Big(||n^{1/2} \sum_{k=2}^{m} \binom{m}{k} U_n^k(\pi_{k,m}h, \mu)||_{\mathcal{H}} \ge \frac{n^{1/2} b_n \delta}{n} \Big) \le 2\ \exp\Big(- \frac{\delta b_n}{2^{m+5} m^{m+1} bc} \Big),$$

which implies

$$\frac{n}{b_n^2}\log \mathbb{P}\left(||\frac{n}{b_n}\binom{m}{k} U_n^k(\pi_{k,m}h,\mu)||_{\mathcal{H}} \geq \delta \right) \leq \frac{n}{b_n^2}\log 2 - \frac{1}{2^{m+5}m^{m+1}bc}\ \frac{n}{b_n}.$$

and the Lemma is proved. □

Remark that the linear term $U_n^1(\pi_{1,m}h,\mu)$ in the Hoeffding decomposition of $U_n^m(h,\mu)$ is of the form

$$U_n^1(\pi_{1,m},\mu) = m\frac{1}{n}\sum_{i=1}^n \pi_{1,m}h(X_i).$$

To deduce the MDP for non-degenerate U-processes from Lemma 3.7, we only have to take care of the class $\{U_n^1(\pi_{1,m}h,\mu) :\ h \in \mathcal{H}\}$. In Wu (1994, Theorem 2) necessary and suffient conditions for uniformly bounded classes of functions in one variable are obtained. Introduce the following quantity:

$$H_1(n,\eta) := E||\sum_{i=1}^n \varepsilon_i \pi_{1,m}h(X_i)||_{\mathcal{H}_{\eta,1}},$$

where

$$\mathcal{H}_{\eta,1} := \{\pi_{1,m}f - \pi_{1,m}g :\ f,g \in \mathcal{H};\ d_2(\pi_{1,m}f,\pi_{1,m}g) := ||\pi_{1,m}f - \pi_{1,m}g||_{L_2(\mu)} \leq \eta\}.$$

The combination of these conditions with Lemma 3.7 yields the following theorem:

Theorem 3.9 (Moderate deviations for U-processes)**.** *Let $\{b_n\}_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$ which satisfies (1.1). Assume that the class $\mathcal{H}$ satisfies one of*

(a) *Let $\mathcal{H}$ be a class of symmetric functions satisfying Condition 3.1.*
(b) *Assume that $\mathcal{H}$ satisfies Condition 3.2. Moreover assume that $\{p_{1,m}h : h \in \mathcal{H}\}$ is totally bounded with respect to d_2 and satisfies*

$$\lim_{\eta\to 0}\limsup_{n\to\infty} \frac{H_1(n,\eta)}{b_n} = 0. \tag{3.10}$$

Then the laws of $(L_n^m - \mu^{\otimes m})(\mathcal{H})$ satisfy the MDP on the space $l_\infty(\mathcal{H})$ with speed n/b_n^2 and with the good rate function $I_{\mathcal{H}}(\cdot)$ given in (2.5).

Proof. Assume that the $\mathcal{H}$ satisfies the first condition. We can use the proof of Corollary 5.7 and the following remark in Arcones and Giné (1991), using the uniform boundedness of $\mathcal{H}$, to get the following remarkable fact: all projections $\pi_{k,m}\mathcal{H} = \{\pi_{k,m}h :\ h \in \mathcal{H}\}$, $k = 1,\dots,m$, satisfy the Central Limit Theorem in $l_\infty(\mathcal{H})$. Especially this implies using Ledoux and Talagrand (1991, Theorem 14.6) that the class $\{\pi_{1,m}h :\ h \in \mathcal{H}\}$ satisfies the sufficient (and necessary) conditions of Theorem 2 in Wu (1994): the class is totally bounded and (3.10) holds.

Notice that in both cases $\{\pi_{1,m}h :\ h \in \mathcal{H}\}$ is uniformly bounded.

Following the lines of the proof of Theorem 2.4, we have only to check (2.16). By (1.4),

$$||U_n^m(h,\mu)-\mu^{\otimes m}(h)||_{\mathcal{H}_\eta}\leq ||U_n^1(\pi_{1,m}h,\mu)||_{\mathcal{H}_\eta}+\sum_{k=2}^m ||\binom{m}{k}U_n^k(\pi_{k,m}h,\mu)||_{\mathcal{H}_\eta}.$$

Now we use the fact that for linear functionals Φ, we have $||\Phi||_{\mathcal{H}_\eta}\leq 2||\Phi||_{\mathcal{H}}$ for each $\eta>0$. Thus Lemma 3.7 implies that (2.16) is checked if the first term on the right hand side satisfies

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\big(||\frac{n}{b_n}U_n^1(\pi_{1,m}h,\mu)||_{\mathcal{H}_\eta}\geq\delta\big)=-\infty.$$

But by our assumptions this is an immediate consequence of Lemma 2.9, see the proof of Theorem 2 in Wu (1994). □

Notice that in the last theorem the assumptions on the sequence $\{b_n\}_{n\in\mathbb{N}}$ are weaker than in Theorem 2.4. Here the $\log n$ term has been removed.

We will now discuss examples of classes $\mathcal{H}$ which satisfy the conditions of the last Theorem. An important and well studied class is the class of uniformly bounded VC subgraph classes of symmetric functions on S^m with values in $\mathbb{R}$.

Corollary 3.11. *If $\mathcal{H}$ is a uniformly bounded VC subgraph class of symmetric functions, then in both cases of Theorem 3.9 the assumptions on $\mathcal{H}$ are fullfilled.*

Proof. We have already mentioned that a uniformly bounded VC subgraph class of functions satisfies Condition 3.1 as well as Condition 3.2. The proof of Corollary 5.7 in Arcones and Giné (1991) shows that $\{\pi_{1.m}h:\ h\in\mathcal{H}\}$ is a measurable uniformly bounded VC subgraph class. Thus the conditions for $\{\pi_{1,m}h:\ h\in\mathcal{H}\}$ hold. □

Remark 3.12. In our scaling the linear part of the Hoeffding decomposition is the leading one. The form of the rate function emphasizes this fact. Look at the non-degenerate form of the rate function $I_m(\cdot)$ in (1.3), which is the rate for the moderate U-empirical measure. We have checked in Eichelsbacher and Schmock (1996) that $I_m(\nu)$ is finite for all measures $\nu\in\mathcal{M}(S^m)$ which satisfy the conditions: $\nu(S^m)=0,\nu_1\ll\mu$ and

$$\nu=\sum_{i=1}^m\mu^{\otimes i-1}\otimes\nu_1\otimes\mu^{\otimes m-i}.$$

By definition (2.5) of the rate function $I_{\mathcal{H}}(H)$ we have to calculate for a given $H\in l_\infty(\mathcal{H})$ the infimum of $I_m(\nu)$ over all measures $\nu\in\mathcal{M}(S^m)$ with $I_m(\nu)\neq\infty$, such that $\nu(\mathcal{H})=H$. The symmetry of h yields

$$\nu(h)=m\int\cdots\int h(x_1,\dots,x_{m-1},x_m)\,d\mu^{\otimes(m-1)}d\nu_1=m\int\mu^{\otimes(m-1)}hd\nu_1.$$

So actually we operate on the class $\{\pi_{1,m}h:\ h\in\mathcal{H}\}$ and build the integral with respect to a measure on S. This reflects the non-degenerate form of the rate function.

From Theorem 3.9 we can derive a law of the iterated logarithm for non-degenerate U-processes for classes $\mathcal{H}$ satisfying condition 3.1.

Corollary 3.13. *Let $\mathcal{H}$ be a class of symmetric functions satisfying Condition 3.1. Let $b_n = \sqrt{2n \log\log n}$ Then the compact LIL holds for $\mathcal{H}$, e.g. $\{\frac{\sqrt{n}}{b_n}(U_n^m(h,\mu) - \mu^{\otimes} h) : h \in \mathcal{H}\}$ is almost surely compact and its limit set coincide with*

$$K := \{H \in l_\infty(\mathcal{H}) : \ I_{\mathcal{H}}(H) \leq 1/2\}.$$

We only point out that the MDP estimations obtained in Theorem 3.9 give the crucial estimation required by the classical proof of the LIL (cf. Wu (1994) for the case of empirical processes). The result was obtained in Arcones and Giné (1994, Corollary 3.4, 3.5). Arcones (1993) has shown that the uniform boundedness hypothesis for the class $\mathcal{H}$ can be relaxed to finite second moment of its envelop. Remark that the Corollary also implies a law of the iterated logarithm for U-processes indexed by a uniformly bounded VC subgraph class.

4. Examples

First of all we will discuss the conditions of Theorem 2.6 and Theorem 2.4. The sum $\sum_{i=1}^{[n/m]} h(Y_i)$ is a sum over the independent blocks $Y_i = (X_{(i-1)m+1}, \dots, X_{(i-1)m+m})$ with values in S^m, which are identically distributed with law $\mathcal{L}(Y_i) = \mu^{\otimes m}$. So the functions h can be interpreted as functions on $\tilde{S} := S^m$. Now the conditions on $\mathcal{H}$ (as a class of functions on $\tilde{S}$) are very well studied. For example if $\mathcal{H}$ is a Donsker class of functions on $\tilde{S}$ (with values in $\mathbb{R}$), then $(\mathcal{H}, d_2)$ is totally bounded and $\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{H(n,\eta)}{\sqrt{n}} = 0$ (cf. Ledoux and Talagrand (1991, Theorem 14.6)). Therefore there is a variety of well studied classes $\mathcal{H}$ which are possible examples of our theorems. Notice moreover that on the moderate scale we are interested in classes $\mathcal{H}$ which satisfy one of the two conditions 3.1, 3.2, because we get a better result with respect to the conditions for the sequence $\{b_n\}_{n \in \mathbb{N}}$.

Since the LDP estimations offers a precise estimation associated to the the law of large numbers (LLN) and the MDP is often used to give further estimations related to the central limit theorem (CLT) and the law of the iterated logarithm (LIL), we will point out in each of the following examples, whether the class $\mathcal{H}$ satisfies these laws. We will investigate classes which are well known and studied for example in Arcones and Giné (1991) and Arcones (1995).

4.1. **Simplicial depth process.** Let $S = \mathbb{R}^d$ and μ be a probability measure on $\mathbb{R}$. Liu introduced the following objects (Liu 1990): The depth $D(x) := D(x, \mu)$ of a point $x \in \mathbb{R}^d$ is defined as the probability that the simplex whose vertices are $d+1$ independent obversations from μ contains x, that is,

$$D(x) = \mu^{d+1}\{x \in S(x_1, \dots, x_{d+1})\},$$

where $S(x_1, \dots, x_{d+1})$ is the closed simplex of $\mathbb{R}^d$ determined by the points $x_1, \dots, x_{d+1}$. For an i.i.d. sequence $\{X_i\}_{i\in\mathbb{N}}$ of random vectors in $\mathbb{R}^d$ with common law μ the n-th empirical simplicial depth process is

$$D_n(x) = U_n^m(C_x), \quad \text{where} \quad C_x = I_{\{(x_1,\dots,x_{d+1})\,:\, x\in S(x_1,\dots,x_{d+1})\}}.$$

For each x, C_x is a function defined on $(\mathbb{R}^d)^{d+1}$ or more secifically, on simplices of $\mathbb{R}^d$. It is well known that $\mathcal{H} := \{C_x : x \in \mathbb{R}^d\}$ is a collection of indicator functions of a measurable VC class of sets (cf. Arcones and Giné (1991, Corollary 6.7)). An immediate consequence is, that the simplicial depth process satisfies both the LLN and the CLT (cf. Arcones and Giné (1991, Corollary 6.8)) as well as the LIL (cf. Theorem 2.2 in Arcones (1993)). Since $\mathcal{H}$ is a uniformly bounded VC subgraph class of functions the MDP holds for this U-process with the rate function (2.5). Notice that we can apply Theorem 3.9 and thus we obtain the result for all $\{b_n\}_{n\in\mathbb{N}}$ satisfying (1.1). The MDP result includes the law of the iterated logarithm. Since $\mathcal{H}$ is a uniformly bounded VC class of functions, it is totally bounded and (2.7) holds. Thus the LDP holds for the simplicial depth process. We get the interesting implication that $D_n(x)$ converges exponentially (uniformly in x) to $D(x)$. The argument is standard in the theory of large deviations: let K be a nonempty closed set in $l_\infty(\mathcal{H})$ not containing a minimum point of $J_{\mathcal{H}}(\cdot)$. Then the upper bound of the LDP result implies that

$$\mathbb{P}\big(U_n^m(\mathcal{H}) \in K\big) \le \exp\big(-nJ_{\mathcal{H}}(K)/2\,\big)$$

for all sufficiently large n. If $J_{\mathcal{H}}(K) > 0$, there is a number $N = N(K)$ such that $\mathbb{P}(U_n^m(\mathcal{H}) \in K) \le e^{-nN}$ for all sufficiently large n. Consider the set $\tilde{K} = K \cap \{H \in l_\infty(\mathcal{H}) : J_{\mathcal{H}}(H) \le J_{\mathcal{H}}(K) + 1\}$. This set is nonempty and is compact. Therefore $I_{\mathcal{H}}(K) = I_{\mathcal{H}}(\tilde{K})$. Since a lower semi-continuous function attains its infimum over a nonempty compact set, there exists a $H \in \tilde{K}$ such that $I_{\mathcal{H}}(H) = I_{\mathcal{H}}(K)$. If $I_{\mathcal{H}}(K) = 0$ then H is a minimum point of $I_{\mathcal{H}}(\cdot)$ and thus we are done. Now choose K such that $\{U_n^m(\mathcal{H}) \in K\} = \{\sup_{x\in\mathbb{R}^d} |D_n(x) - D(x)| \ge \varepsilon\}$, then the Borel-Cantelli lemma implies the strong law of large numbers and we obtain that $D_n(x)$ converges (uniformly in x) exponentially to $D(x)$.

The simplicial median x_0 of μ is defined as the argument of the supremum of $D(x)$ if it exists and is unique, or as any point in $\{\arg\max D(x)\}$, if this set consists of more than one point. The n-th empirical simplicial median is any random variable x_n almost surely in the set $\{\arg\max D_n(x)\}$. So the estimator of x_0 is a maximizer of a U-process. The consistency (cf. Arcones and Giné (1991, Theorem 6.9)) of x_n for x_0 follows from the law of large numbers for the corresponding U-process $D_n(x)$. Another interesting implication of the LDP is the exponentially convergence of x_n to x_0 in the following sense: Let μ be a probability measure on $\mathbb{R}^d$ satisfing: (a) $D(\cdot)$ is uniquely maximized at x_0 and (b) x_n is a sequence of random variables with

$D_n(x_n) = \operatorname{argmax} D_n(x)$, then

$$x_n \to x_0 \quad \text{a.s exponentially fast.}$$

The proof is a simple adaption of the proof of Arcones and Giné (1991, Theorem 6.9), using the exponential decay of the probability for any closed $K \in l_\infty(\mathcal{H})$ not containing a minimum point of $J_{\mathcal{H}}(\cdot)$ as seen above.

4.2. **Empirical distribution function of U-statistic structure.** Let h be a fixed measurable and symmetric function on S^m. Consider the random process defined by $U_n^m(I_{h\le t}, \mu) - \mu^{\otimes}(h \le t)$, $t \in \mathbb{R}$. These processes are studied for example in Helmers, Janssen, and Serfling (1988), Nolan and Pollard (1988) and Arcones (1995). Building the supremum over t we get a Kolmogorov-Smirnov type statistic. The class $\mathcal{H} = \{I_{h\le t} : t \in \mathbb{R}\}$ is Euclidean in the sense of Nolan and Pollard (1987): there exists constants A and v such that $N_1(\varepsilon, \mathcal{H}, \mu) \le A\varepsilon^{-v}$. A direct consequence is $N_2(\varepsilon, \mathcal{H}, \mu) \le (A2^{2v})\varepsilon^{-2v}$. Therefore the class $\mathcal{H}$ satisfies Condition 3.1 and thus we obtain the MDP for the empirical distribution of U-statistic structure. The LLN and the CLT are discussed in Nolan and Pollard (1988, Theorem 5). The LIL is a consequence of Arcones and Giné (1994, Theorem 3.3). Moreover $\mathcal{H}$ fulfills all conditions of Theorem 2.6 as discussed already at the beginning of this section. Thus the LDP holds for this class of processes, too. Moreover it is easy to see that the class of functions $\tilde{\mathcal{H}} = \{I_{h\le r} - I_{h\le s} : r, s \in \mathbb{R}\}$ is a VC subgraph class of functions. Considering the MDP and the LDP bounds we can prove lower and upper bounds for $\mathbb{P}\big(\sup_{a\le r\le s\le b} |U_n^m(I_{h\le s} - I_{h\le r})|\big)$. It is not easy to a get an explicit expression for the rate functions (for a fixed h). Therefore the bounds are not as "precise" as the one in Arcones (1995). Notice, that we get a sharp result in the sense that we get an upper and a lower bound with the same speed and the same rate function.

4.3. **Weighted empirical distributions of U-statistic structure.** In Helmers, Janssen, and Serfling (1988) and Arcones and Giné (1991) weighted discrepancies between $U_n^m(I_{h\le t})$ and $\mu^{\otimes m}(h \le t)$ are studied. Assume that $q : (0,1) \to \mathbb{R}^+$ is continuous, nondecreasing on $(0, \delta]$ and nonincreasing on $[1-\delta, 1)$ for some $0 < \delta < 1/2$, and satisfies $\int_0^1 dt/q(t) \le c$ for some constant c. Then the class

$$\frac{I_{h\le t} - \mu^{\otimes m}(h \le t)}{q(\mu^{\otimes}(h \le t))}$$

is a uniformly bounded VC-subgraph class. If q satisfies $\int_0^1 dt/q(t) < \infty$, the LLN follows as in Arcones and Giné (1991, Example 3.4). Under our more restrictive assumption on q we obtain the LDP and the MDP via the same arguments as in the last example.

4.4. **Uniform Hölder functions.** Let $H: \mathbb{R}^m \to \mathbb{R}$, $H > 0$ and

$$\int_{\mathbb{R}^m} \exp(\lambda H(x)) d\mu^{\otimes m}(x) < \infty$$

for every $\lambda > 0$. Let, for $0 < \alpha \leq 1$ and $c < \infty$ $\mathcal{H} = \{h : \mathbb{R}^m \to \mathbb{R} : |h(x)| \leq H(x)$ for all $x \in \mathbb{R}^m$, $|h(x) - h(y)| \leq c|x-y|^\alpha$ for all $x, y \in \mathbb{R}^m\}$.

First of all we restrict ourselves to the case of a uniformly bounded class, i.e. $|H(x)| \leq K$ for all $x \in \mathbb{R}^m$ and $K \in \mathbb{R}$ fixed. Moreover we will assume that the functions of $\mathcal{H}$ are supported by a fixed compact set. But then, by a result of Kolmogorov (cf. Dudley (1984, Theorem 7.1.1)) the metric entropy of $\mathcal{H}$ with respect to the sup norm is of the order of $\varepsilon^{-m/\alpha}$ which especially implies that there is a Lebesgue integrable function $\lambda : (0, \infty) \to [0, \infty)$ such that for each probability measure ν

$$\big(\log N_2(\varepsilon, \mathcal{H}, \nu)\big)^{m/2} \leq \lambda(\varepsilon), \quad \varepsilon > 0.$$

Hence both conditions of Lemma 3.4 are satisfied. Following the lines of the proof of Lemma 3.7, we can prove the statement of this Lemma. Notice moreover, that we can use the proof of Corollary 5.7 and the following remark in Arcones and Giné (1991) (using the uniform boundedness of $\mathcal{H}$) to see, that all projections $\pi_{k,m}\mathcal{H}$ satisfy the CLT in $l_\infty(\mathcal{H})$. Therefore the proof of Theorem 3.9 can be adapted and the MDP for the class $\mathcal{H}$ is verified.

The observation from Corollary 5.7 in Arcones and Giné (1991) implies the LDP for the U-process with the class $\mathcal{H}$. But we can prove an LDP for the U-process of $\mathcal{H}$ in general via truncation. Let $h_M = h \wedge M \vee (-M)$, $M > 0$. Using Hoeffdings formula (2.1) and formula (5.2) in Hoeffding (1963) we obtain

$$\begin{aligned}
&\mathbb{P}(\sup_{h \in \mathcal{H}} ||L_n^m(h - h_M) - \mu^{\otimes m}(h - h_M)|| \geq \delta) \\
&\qquad \leq \exp(-[n/m]\delta\alpha) \bigg(\int_{S^m} \exp(\alpha |2H|) \bigg)^{[n/m]}.
\end{aligned}$$

The dominated convergence theorem and the moment condition on H yields

$$\lim_{M \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\sup_{h \in \mathcal{H}} ||L_n^m(h - h_M) - \mu^{\otimes m}(h - h_M)|| \geq \delta) = -\infty$$

for all $\delta > 0$. Next we can state the LDP for the class $\mathcal{H}$, where each function $h \in \mathcal{H}$ is supported by the coordinate hypercube R_M of side M centered at 0. Via Hoeffdings formula we can check that the influence of $hI_{R_M^c}$ is exponentially small. We omit the details. Remark that we can prove the results for more general classes of differentiable functions considered in Dudley (1984, Chapter 7.1). There for any $\alpha > 0$ the class of interest is the space of functions which have "bounded derivatives through order α." If β is the greatest integer $< \alpha$, the functions in this class have partial derivatives through order β bounded, and the derivatives of order β satisfy a uniform Hölder condition of order $\alpha - \beta$. Notice that the LLN is proved in Arcones and Giné (1991, Example 3.10).

References

Arcones, M. A. (1993). The law of the iterated logarithm for U-processes. *J. Mult. Anal. 47*, 139–151.

Arcones, M. A. (1995). A Bernstein-type inequality for U-statistics and U-processes. *Statist. Probab. Lett. 22*, 239–247.

Arcones, M. A. and E. Giné (1991). 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.

Baranyai, Z. (1975). On the factorisation of the complete uniform hypergraph. In A. Hajnal, T. Rado, and V. Sós (Eds.), *Infinite and Finite Sets*, pp. 91–108. North-Holland, Amsterdam.

Borovskich, Y. and V. Koroljuk (1994). *Theory of U-Statistics*, Volume 273 of *Mathematics and Its Applications.* Dordrecht: Kluwer Academic Publishers.

de Acosta, A. (1994a). On large deviations of empirical measures in the τ-topology. *J. Appl. Probab. 31A*, 41–47. (special volume).

de Acosta, A. (1994b). 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 (1993). *Large Deviations Techniques and Applications.* Boston, MA: Jones and Bartlett.

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. (1996). Large deviations for products of empirical probability measures in the τ-topology. *J. Theoret. Probab.*. (to appear).

Eichelsbacher, P. and U. Schmock (1996). Large and moderate deviations of products of empirical measures and U-empirical measures in strong topologies. submitted.

Giné, E. (1996). Empirical processes and applications: an overview. *Bernoulli 2*(1), 1–28.

Helmers, R., P. Janssen, and R. Serfling (1988). Glivenko-Cantelli properties of some generalized empirical DF's and strong convergence of generalized L-statistics. *Probab. Theor. Relat. Fields 79*, 75–93.

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

Ledoux, M. (1992). Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. Henri Poincaré 28*(2), 267–280.

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

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.

Serfling, R. and W. Wang (1997). Large deviation results for U- and V- statistics, -empiricals, and -processes. preprint.

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.

(Peter Eichelsbacher) Fakultät für Mathematik, Universität Bielefeld, Universitätsstrasse 1, D-33501 Bielefeld, Germany

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