Ann. Inst. H. Poincaré, Probabilités et Statistiques 37, 2 (2001) 245–273

S0246-0203(00)01063-3/FLA

MODERATE DEVIATIONS FOR FUNCTIONAL U-PROCESSES

Peter EICHELSBACHER

Fakultät für Mathematik, Universität Bielefeld, Universitätstrasse 1, D-33501 Bielefeld, Germany

Received in 26 November 1997, revised 15 May 2000

ABSTRACT. – The moderate deviations principle is shown for the partial sums processes built of U-empirical measures and of U-statistics. It is proved that in the non-degenerate case the conditions for the fixed time principles suffice for the moderate deviations principle to carry over to the corresponding partial sums processes. Given a uniformly bounded VC subgraph class of functions, we obtain corresponding moderate deviations for time dependent U-processes. We use decoupling techniques and apply an improved version of a Bernstein-type inequality for degenerate U-statistics. Moreover, we prove and use a Lévy-type maximal inequality for U-statistics. © 2001 Éditions scientifiques et médicales Elsevier SAS

Keywords: Moderate deviations; Partial sums; U-processes; VC-classes; Decoupling inequality; Maximal inequality for U-statistics; Bernstein-type inequality

Math. Subj. Class. (1991): 60F05 (primary); 60E15; 60F10; 60G17

RÉSUMÉ. – Nous établissons le principe de déviations modérées pour les processus de sommes partielles construits à partir des U-mesures empiriques et des U-statistiques. Nous montrons que dans le cas non-dégénéré les conditions pour les résultats à temps fixe suffisent aussi pour le cas des processus de sommes partielles. Etant donné une classe de fonctions de Vapnic–Čhervonenkis uniformément bornée, nous obtenons les déviations modérées correspondantes pour les U-processus dépendant du temps. Nous utilisons des techniques de découplage et appliquons une version améliorée d'une inégalité de type Bernstein pour les U-statistiques dégénérées. De plus nous démontrons et utilisons une ingégalité maximale de Lévy pour les U-statistiques. © 2001 Éditions scientifiques et médicales Elsevier SAS

1. Introduction and statement of the results

For a sequence of $\mathbb{R}^d$-valued i.i.d. random variables X_i with a finite moment generating function Borovkov and Mogulskii [5] investigated the moderate deviation

E-mail address: peter@mathematik.uni-bielefeld.de (P. Eichelsbacher).

behaviour of the polygonal approximation of the partial sums process

$$S_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} X_i, \quad t \in [0,1]. \tag{1.1}$$

In an appropriate topology, Dembo and Zajic [8] proved a moderate deviations principle for the partial sums empirical process

$$L_n(t) = \frac{1}{n} \sum_{i=1}^{[nt]} \delta_{X_i}, \quad t \in [0,1].$$

Given a class of bounded functions $\mathcal{F}$, Dembo and Zajic [8] considered the moderate deviations principle for functional empirical processes

$$L_n(t,f) = \frac{1}{n} \sum_{i=1}^{[nt]} f(X_i), \quad t \in [0,1],\ f \in \mathcal{F}.$$

The aim of this paper is to extend the moderate deviations principle when passing from linear statistics to higher order statistics.

Let us recall the definition of the large deviations principle (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 $\{x\colon I(x) \leqslant \alpha\}$ are compact for all $\alpha < \infty$ and for all $\Gamma \in \mathcal{B}$ the lower bound

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

and the upper bound

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

hold. Here $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively. We say that a sequence of random variables satisfies the LDP 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 satisfying

$$\lim_{n\to\infty} \frac{b_n}{n} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{b_n^2} = 0. \tag{1.2}$$

If $\mathcal{X}$ is a topological vector space then a sequence of random variables $\{Z_n, n \in \mathbb{N}\}$ shall satisfy the moderate deviations principle (MDP) with speed $\frac{n}{b_n^2}$ and with good rate function $I(\cdot)$, if the sequence $\{\frac{n}{b_n} Z_n, n \in \mathbb{N}\}$ satisfies the LDP in $\mathcal{X}$ with the good rate function $I(\cdot)$ and with speed $\frac{n}{b_n^2}$.

Denote by $L_\infty([0,1], \mathbb{R}^d)$ the space of (equivalence classes modulo equality a.e. of) bounded measurable functions on $[0,1]$, equipped with the uniform topology. Consider the polygonal approximation of $S_n(\cdot)$, that is

$$\widetilde{S}_n(t) := S_n(t) + \left(t - \frac{[nt]}{n}\right) X_{[nt]+1}. \tag{1.3}$$

Note that $\widetilde{S}_n(\cdot)$ is continuous and carries the same information as $S_n(\cdot)$. The MDP for $\{\widetilde{S}_n(\cdot), n \in \mathbb{N}\}$ was established in $L_\infty([0,1],\mathbb{R}^d)$: By $\mathcal{AC}_0([0,1],\mathbb{R}^d)$ denote the subspace of absolutely continuous functions ϕ on $[0,1]$ with $\phi(0) = 0$. Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of i.i.d. random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^d$, common law μ and $\mathbb{E}(X_1) = 0$. Borovkov and Mogulskii [5] considered (essentially) the MDP in the i.i.d. case under the condition that $\mathbb{E}\exp(\langle\theta, X_1\rangle) < \infty$ in some ball centered at the origin. $\langle\cdot,\cdot\rangle$ denotes the Euclidean scalar product in $\mathbb{R}^d$ and $\|\cdot\|$ a norm in $\mathbb{R}^d$. The sequence $\{\frac{n}{b_n}\widetilde{S}_n(\cdot), n \in \mathbb{N}\}$ satisfies the LDP in $L_\infty([0,1],\mathbb{R}^d)$, equipped with the uniform topology, with good rate function

$$I_\infty(\phi) = \int_0^1 \Lambda^*(\dot{\phi})\,dt, \tag{1.4}$$

if $\phi \in \mathcal{AC}_0([0,1],\mathbb{R}^d)$ and $I_\infty(\phi) = \infty$ otherwise (see also [22, Theorem 1] and [4, Theorem 3.1]). Here Λ^* denotes the convex dual of $\Lambda(\theta) = \mathbb{E}(\langle\theta, X_1\rangle^2)/2$, that is

$$\Lambda^*(x) := \sup_{\theta\in\mathbb{R}^d}\{\langle\theta, x\rangle - \Lambda(\theta)\}.$$

1.1. Moderate deviations for partial sums U-statistics

We will consider the MDP for different partial sums processes connected with U-statistics. Recall that

$$U_n^m(h) := \frac{1}{\binom{n}{m}} \sum_{C_m^n} h(X_{i_1},\dots,X_{i_m}) = \frac{1}{n_{(m)}} \sum_{I(m,n)} h(X_{i_1},\dots,X_{i_m}),$$

is called a U-statistic. Here the X_i are i.i.d. random variables and h is a measurable, symmetric, $\mathbb{R}^d$-valued function, called kernel function, where symmetric means that h is invariant under all permutations of its arguments. C_m^k with $k, m \in \mathbb{N}$ denotes the set $\{(i_1,\dots,i_m)\colon 1 \leqslant i_1 < \cdots < i_m \leqslant k\}$, $n_{(m)} \equiv \prod_{k=0}^{m-1}(n-k)$ and $I(m,n) \subset \{1,\dots,n\}^m$ contains all m-tuples with pairwise different components.

While U-statistics were introduced by Hoeffding [18] as a generalization of the empirical mean to the case of multivariate functions, later different types of stochastic processes related with U-statistics have been studied, see for example Miller and Sen [21], Hall [17] and Mandelbaum and Taqqu [20]. There the authors basically proved invariance principles and the weak convergence of appropriately normalized partial sums U-statistics to Brownian motion, functionals of Brownian motion, or limit processes expressible as multiple Wiener integrals.

To be able to formulate our results we will next describe the Hoeffding decomposition. The operator $\pi_{k,m}^\mu = \pi_{k,m}$, $k = 0, 1, \dots, m$, acts on $\mu^{\otimes m}$-integrable symmetric functions

$h : S^m \to \mathbb{R}$ as follows:

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

where

$$\nu_1 \otimes \cdots \otimes \nu_m h = \int \cdots \int h(x_1, \ldots, x_m)\, d\nu_1(x_1) \cdots d\nu_m(x_m)$$

and

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

A function h is called μ-*canonical* or *completely degenerate* if $\int h(x_1, \ldots, x_m)\, d\mu(x_i) = 0$ for all $1 \leqslant i \leqslant m$. Note that $\pi_{k,m} h$ is a μ-canonical function of k variables. If $\pi_{1,m} h \neq 0$, then h is called *non-degenerate*. 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

$$U_n^m(h) = \sum_{k=0}^{m} \binom{m}{k} U_n^k(\pi_{k,m} h) \tag{1.5}$$

(cf. (3.5.1) in [15]). Here $U_n^0(\pi_{0,m} h) = \mu^{\otimes m} h$. Consider the *partial sums U-statistics*, that is

$$U_n^m(t, h) := \frac{1}{\binom{n}{m}} \sum_{C_m^{[nt]}} h(X_{i_1}, \ldots, X_{i_m}), \quad t \in [0, 1].$$

We will prove the MDP for the process

$$W_n^m(t, h) = \sum_{k=0}^{m} \binom{m}{k} U_n^k(t, \pi_{k,m} h), \quad t \in [0, 1], \tag{1.6}$$

as well as for $U_n^m(t, h)$, $t \in [0, 1]$, in the non-degenerate case in the uniform topology. In Hall [17] and Mandelbaum and Taqqu [20] functional limit theorems for the process $W_n^m(\cdot, h)$ are discussed. Remark that using (1.5) the process $U_n^m(t, h)$ gets the representation

$$U_n^m(t, h) = \sum_{k=0}^{m} \binom{m}{k} \frac{\binom{[nt]}{m}\binom{n}{k}}{\binom{[nt]}{k}\binom{n}{m}} U_n^k(t, \pi_{k,m} h).$$

The additional factors $\frac{\binom{[nt]}{m}\binom{n}{k}}{\binom{[nt]}{k}\binom{n}{m}}$ influence the behaviour of this process and we will observe another rate function.

To be more precise, we will prove a MDP result for the polygonal approximation process, that is

$$\widetilde{U}_n^m(t, h) := U_n^m(t, h) + \big(nt - [nt]\big) \frac{1}{\binom{n}{m}} \sum_{C_{m-1}^{[nt]}} h(X_{i_1}, \ldots, X_{i_{m-1}}, X_{[nt]+1}),$$

$$t \in [0, 1].$$

Denote by $\widetilde{W}_n^m(t,h)$ the process given by (1.6) where $U_n^k(t,\pi_{k,m}h)$ is replaced by $\widetilde{U}_n^k(t,\pi_{k,m}h)$ for every $k \in \{0,\dots,m\}$. Proving the MDP for $\{\widetilde{W}_n^m(\cdot,h), n \in \mathbb{N}\}$, we will prove that the process $\{\widetilde{U}_n^1(\cdot,\pi_{1,m}h),\ n \in \mathbb{N}\}$ satisfies a MDP and that the sequences $\{\widetilde{U}_n^k(\cdot,\pi_{k,m}h),\ n \in \mathbb{N}\}$ do not contribute to the moderate upper and lower bounds for $2 \leqslant k \leqslant m$. To this end we will use an improved Bernstein-type inequality (Lemma 2.1); for bounded kernel functions h this type of inequality was proved in [1] (see also [15, Theorem 4.1.12]). Moreover we will prove and apply a Lévy-type maximal inequality for U-statistics (Lemmas 2.5 and 2.8). The MDP for $\{\widetilde{U}_n^m(\cdot,h), n \in \mathbb{N}\}$ can be deduced from the MDP for $\{\widetilde{W}_n^m(\cdot,h), n \in \mathbb{N}\}$ using the contraction principle and the concept of exponential equivalence (see [9, Theorem 4.2.1 and Theorem 4.2.13]). Consider

Condition 1.7 (*Weak Cramér condition*). – For each $2 \leqslant k \leqslant m$ there exists a $\delta_k > 0$ such that

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

and there exists a $\delta > 0$ such that

$$\int_S \exp(\delta \|\pi_{1,m}h\|)\, d\mu < \infty. \tag{1.8}$$

Here $\|\cdot\|$ denotes the standard Euclidian norm on $\mathbb{R}^d$.

Define

$$\Lambda_m^*(x) := \sup_{\theta \in \mathbb{R}^d} \bigl\{\langle\theta, x\rangle - \Lambda_m(\theta)\bigr\}, \tag{1.9}$$

with

$$\Lambda_m(\theta) = \mathbb{E}\bigl(\langle\theta, \pi_{1,m}h(X_1)\rangle^2\bigr)/2.$$

THEOREM 1.10 (Moderate deviations of partial sums U-statistics). – *Assume that Condition* 1.7 *is satisfied for a symmetric kernel function h, then*

(a) *the sequence $\{\widetilde{W}_n^m(\cdot,h) - \mathbb{E}(h), n \in \mathbb{N}\}$ satisfies the MDP in $L_\infty([0,1],\mathbb{R}^d)$ with good rate function*

$$I_W^m(\phi) := \int_0^1 \Lambda_m^*\bigl(\dot{\phi}/m\bigr)\, dt,$$

if $\phi \in \mathcal{AC}_0([0,1],\mathbb{R}^d)$ and $I_W^m(\phi) = \infty$ otherwise. The speed is n/b_n^2;

(b) *the sequence $\{\widetilde{U}_n^m(\cdot,h) - \mathbb{E}(h), n \in \mathbb{N}\}$ satisfies the MDP in $L_\infty([0,1],\mathbb{R}^d)$ with good rate function*

$$I_U^m(\phi) := \int_0^1 \Lambda_m^*\bigl(\dot{\phi}/\bigl(mt^{m-1}\bigr)\bigr)\, dt,$$

if $\phi \in \mathcal{AC}_0([0,1],\mathbb{R}^d)$ and the integral exists and $I_U^m(\phi) = \infty$ otherwise. The speed is n/b_n^2.

Remark 1.11. –

(a) The weak Cramér conditions in 1.7 are equivalent to the following conditions: Assume that for each $2 \leqslant k \leqslant m$

$$\mathbb{E}\|\pi_{k,m}h\|^{2l} \leqslant \frac{1}{2} l! H^{l-2} \mathbb{E}\|\pi_{k,m}h\|^4, \quad l = 2, 3, \ldots,$$

and

$$\mathbb{E}\|\pi_{1,m}h\|^{l} \leqslant \frac{1}{2} l! H^{l-2} \mathbb{E}\|\pi_{1,m}h\|^2$$

(see for example [29, Remark 3.6.1]).

(b) We can easily adapt the result to a time interval $[0, T]$. Applying Theorem 4.6.1 of [9] for $T \in \mathbb{N}$ yields a MDP for $\{\widetilde{U}_n^m(\cdot, h) - \mathbb{E}(h), n \in \mathbb{N}\}$ and $\{\widetilde{W}_n^m(\cdot, h) - \mathbb{E}(h), n \in \mathbb{N}\}$, respectively, in $\mathcal{AC}_0(\mathbb{R}_+, \mathbb{R}^d)$ equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$.

In the following examples we discuss the MDP for the sequence $\{\widetilde{W}_n^m(\cdot, h) - \mathbb{E}(h), n \in \mathbb{N}\}$.

Example 1.12. – Consider the *sample variance* U_n^{var}, which is a U-statistic of degree 2 with kernel function $h(x, y) = \frac{1}{2}(x - y)^2$. A simple calculation shows that

$$\pi_{1,2}h(x) = \frac{1}{2}\big((x - \mathbb{E}(X_1))^2 - \mathrm{Var}(X_1)\big),$$

where $\mathrm{Var}(X_i)$ denotes the variance of X_i under μ. It is well known, that we are in the non-degenerate regime, if the distribution μ satisfies the condition $\mathbb{E}(X_1 - \mathbb{E}(X_1))^4 > \mathrm{Var}(X_1)^2$. If μ satisfies Condition 1.7, the rate function can be calculated as follows: we obtain for the sample variance that $\Lambda_2(\theta) = \frac{\theta^2}{8}(\mathbb{E}(X_1 - \mathbb{E}(X_1))^4 - (\mathrm{Var}(X_1))^2) =: \frac{\theta^2}{8} c(\mu)$. Hence the rate function is

$$I_{\mathrm{var}}(\phi) = \frac{1}{2c(\mu)} \int_0^\infty (\dot{\phi})^2 \, dt, \quad \phi \in \mathcal{AC}_0\big(\mathbb{R}_+, \mathbb{R}^d\big).$$

Consider the coin tossing with $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$ and $0 < p < 1$, $p \neq 1/2$. We are in the non-degenerate case and the corresponding rate function is

$$I_{\mathrm{var}}^{\mathrm{bernoulli}}(\phi) = \frac{1}{2(1-p)(p - 4p^2 + 4p^3)} \int_0^\infty (\dot{\phi})^2 \, dt.$$

Example 1.13. – Note that for the kernel function $h(x, y) = xy$ (*sample second moment*), we obtain $\pi_{1,2}h(x) = \mathbb{E}(X_1)(x - \mathbb{E}(X_1))$. Therefore

$$\Lambda_2(\theta) = \frac{1}{2}(\mathbb{E}X_1)^2 \mathrm{Var}(X_1)\theta^2 =: \frac{1}{2}\theta^2 c(\mu).$$

The U-statistic is non-degenerate if μ fulfills $(\mathbb{E}X_1)^2\mathrm{Var}(X_1) \neq 0$. In this case the corresponding rate function is

$$I_{\mathrm{sec}}(\phi) = \frac{1}{8c(\mu)} \int_0^\infty (\dot{\phi})^2 \, dt, \quad \phi \in \mathcal{AC}_0(\mathbb{R}_+, \mathbb{R}^d).$$

In the case of Bernoulli random variables we obtain the rate $\frac{1}{8p^3(1-p)} \int_0^\infty (\dot{\phi})^2 \, dt$ for every $0 < p < 1$.

Example 1.14. – In the case of the *Wilcoxon one sample statistic* the kernel is given by $h(x, y) = 1_{\{x+y \geqslant 0\}}(x, y)$. If $F(\cdot)$ denotes the distribution function of X_1, we obtain $\pi_{1,2}h(x) = 1 - F(-x) - \mathbb{P}(X_1 + X_2 \geqslant 0)$ and

$$\pi_{2,2}h(x, y) = h(x, y) - \big(1 - F(-x)\big) - \big(1 - F(-y)\big) + \mathbb{P}\big(X_1 + X_2 \geqslant 0\big).$$

We restrict ourselves to the class of distribution functions $F(\cdot)$ which are continuous and symmetric in 0. Then $\mathbb{E}(h) = \mathbb{P}(X_1 \geqslant X_2) = \frac{1}{2}$. Moreover a simple calculation shows that

$$\begin{aligned}\Lambda_2(\theta) &= \frac{1}{2}\theta^2 \mathbb{E}\left(\big(1 - F(-x)\big)^2 - \big(1 - F(-x)\big) + \frac{1}{4}\right)\\ &= \frac{1}{2}\theta^2 \left(\int_0^1 y^2 \, dy - \int_0^1 y \, dy + \frac{1}{4}\right) = \frac{\theta^2}{24}.\end{aligned}$$

Conditions 1.7 are fulfilled for every μ since h is bounded. Thus we get for every continuous distribution function $F(\cdot)$, symmetric in 0, a MDP for the sequence $\{\widetilde{W}_n^2(\cdot, 1_{\{x+y \geqslant 0\}}) - 1/2, n \in \mathbb{N}\}$ with good rate function

$$I_{\mathrm{wil}}(\phi) = \frac{3}{2} \int_0^\infty (\dot{\phi})^2 \, dt.$$

Hence the Wilcoxon one sample statistic is asymptotically distribution free. The same is true for the *Wilcoxon signed rank statistic*. The MDP can be deduced from the MDP of the one sample statistic: consider the test problem specified by the hypothesis $H_0 := \{F\colon\ F \text{ continuous and symmetric in } 0\}$ against all other symmetric, continuous distribution functions. This test can be performed using the Wilcoxon signed rank statistic W: denote R_i^+ the rank of $|X_i|$ among all $|X_1|, \ldots, |X_n|$. W is defined as $W = 1/2 \sum_{i=1}^n R_i^+(1 + \mathrm{sign}(X_i))$, and can be written as a sum of two U-statistics:

$$W = nU_n^1(h_1) + \binom{n}{2} U_n^2(1_{\{x+y \geqslant 0\}})$$

with $h_1(x) := 1_{\{x>0\}}(x)$. Consider $\widetilde{W}(\cdot) := n\widetilde{W}_n^1(\cdot, h_1) + \binom{n}{2}\widetilde{W}_n^2(\cdot, 1_{\{x+y\geqslant 0\}})$. Checking the MDP for the process $(\widetilde{W}(\cdot) - \mathbb{E}(W))/\binom{n}{2}$ we will prove that for every $\delta > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\sup_{t\in[0,1]} \left\| \frac{\sum_{i=1}^{[nt]} h_1(X_i)}{n-1} - \mathbb{E}(h_1) \right\| \geqslant b_n\delta \right) = -\infty$$

holds (see [9, Theorem 4.2.13]). This example already shows the impact of Bernstein inequalities and the Lévy-type maximal inequality: applying Proposition 2.14 (which is [15, Theorem 1.1.5]) and inequality (3.10) (which is [1, Proposition 2.3(d)]) for $r = 1$ we obtain

$$\mathbb{P}\left(\sup_{t\in[0,1]} \left\| \frac{\sum_{i=1}^{[nt]} h_1(X_i)}{n-1} - \mathbb{E}(h_1) \right\| \geqslant b_n\delta \right) \leqslant c_1 \exp\left(-c_2 \frac{b_n^2\delta^2(n-1)^2}{n} \right),$$

with some constants c_1 and c_2. Hence the MDP follows with rate $I_{\text{wil}}(\cdot)$.

The following corollary which follows directly from Theorem 1.10 does not seem to exist in the literature.

COROLLARY 1.15. – *If Condition* 1.7 *is satisfied for the symmetric kernel function* h*, then* $\{U_n^m(h) - \mathbb{E}(h), n \in \mathbb{N}\}$ *satisfies the MDP in* $\mathbb{R}^d$ *with the good rate function* $\Lambda_m^*(\cdot/m)$ (*see* (1.9)) *and with speed* n/b_n^2.

A further example demonstrating the usefulness of Theorem 1.10 is the following. Consider the function $g : C([0,1], \mathbb{R}^d) \to \mathbb{R}^d$ ($C([0,1], \mathbb{R}^d)$ denotes the space of continuous functions on $[0,1]$ with values in $\mathbb{R}^d$) defined by $g(x) := \sup_{t\in[0,1]} x(t)$. The function g is continuous and we can deduce a MDP for the sequence $\{\sup_{t\in[0,1]}(\widetilde{W}_n^m(t,h) - \mathbb{E}(h)), n \in \mathbb{N}\}$ with rate function

$$J(y) = \inf\left\{ I_W^m(\phi) \colon \sup_{t\in[0,1]} \phi(t) = y \right\}, \quad y \in \mathbb{R}^d,$$

whenever $\{\widetilde{W}_n^m(\cdot, h) - \mathbb{E}(h),\ n \in \mathbb{N}\}$ satisfies the MDP with rate function $I_W^m(\cdot)$. By the convexity of Λ_m^* we obtain $J(y) = \Lambda_m^*(y/m)$.

The large deviation principle (LDP) for $\{U_n^m(h), n \in \mathbb{N}\}$ is proved in [11].

1.2. Moderate deviations for partial sums U-empirical measures

In order to formulate the result on the empirical measure level we need some more notations. Let (S, d) be a Polish space with metric d. Fix $m \in \mathbb{N}$. Denote by $B(S^m, \mathbb{R}^d)$ the set of bounded Borel measurable functions, by $B'(S^m, \mathbb{R}^d)$ the algebraic dual of $B(S^m, \mathbb{R}^d)$ and by $C_b(S^m, \mathbb{R}^d)$ the set of bounded continuous functions.

Denote by $\mathcal{M}(S^m)$ and $\mathcal{M}_t(S^m)$, respectively, the set of Borel measures on the Polish space S^m which are signed and positive having total measure t, respectively. Unless explicitly stated otherwise, these spaces are equipped with the τ-topology. Let $D_1[\mathbb{R}^d]$ denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ equipped with the topology of uniform convergence on compact subsets of $\mathbb{R}_+$ and the corresponding Borel σ-field.

Let $D_1[\mathcal{M}(S^m)]$ denote the space of càdlàg functions from $\mathbb{R}_+$ to $\mathcal{M}(S^m)$ equipped with the weakest topology such that the maps $y(\cdot) \mapsto \langle \varphi, y(\cdot)\rangle : D_1[\mathcal{M}(S^m)] \mapsto D_1[\mathbb{R}]$ are continuous and the smallest σ-field $\mathcal{B}$ such that these maps are measurable. Here $\varphi \in B(S^m, \mathbb{R})$, and $\langle \varphi, y(\cdot)\rangle = \int \varphi\, dy(\cdot)$.

Next let $D_2[[0,T], \mathbb{R}^d]$, $T > 0$, denote the Banach space of càdlàg functions $f : [0,T] \to \mathbb{R}^d$ with finite norm $\sup_{t\in[0,T]} \|f(t)\|/(t+1)$. Let $D_2^{\mathrm{proj}}[\mathbb{R}^d]$ denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ equipped with the projective limit topology of the system $(D_2[[0,T], \mathbb{R}^d], T \in \mathbb{R}_+)$ and let $D_2[\mathbb{R}^d]$ be the Banach space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ with finite norm $\sup_{t\in\mathbb{R}_+} \|f(t)\|/(t+1)^m$. All spaces are equipped with the corresponding Borel σ-field.

Let $D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ and $D_\beta[\mathbb{R}^d]$, respectively, denote the space of càdlàg functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ when in the definition of the norms $t+1$ is replaced by $\beta(t) = \sqrt{2(t \vee 1)\log\log(t \vee 3)}$. Moreover define $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ as the space of càdlàg functions $y : \mathbb{R}_+ \to \mathcal{M}(S^m)$ such that $\langle \varphi, y(\cdot)\rangle \in D_2^{\mathrm{proj}}[\mathbb{R}^d]$ for every $\varphi \in B(S^m, \mathbb{R}^d)$, equipped with the weakest topology such that the maps

$$y(\cdot) \mapsto \langle \varphi, y(\cdot)\rangle : D_2^{\mathrm{proj}}\bigl[\mathcal{M}\bigl(S^m\bigr)\bigr] \mapsto D_2^{\mathrm{proj}}\bigl[\mathbb{R}^d\bigr]$$

are continuous and the smallest σ-field, such that these maps are measurable. The space $D_2[\mathcal{M}(S^m)]$ and $D_2[B'(S^m, \mathbb{R}^d)]$ are defined in a similar way.

For an i.i.d. sequence $\{X_n, n \in \mathbb{N}\}$ with state space S, Dembo and Zajic proved in [8, Theorem 1] the MDP for $\{L_n(\cdot), n \in \mathbb{N}\}$ in $D_1[\mathcal{M}(S)]$ as well as in $D_2[\mathcal{M}(S)]$ with a convex good rate function

$$J_\infty\bigl(\nu(\cdot)\bigr) = \int_0^\infty J(\dot{\nu} \mid \mu)\, dt,$$

if $\nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S)]$ and $J_\infty(\nu(\cdot)) = +\infty$ otherwise. Here the functional $J(\cdot \mid \mu) : \mathcal{M}(S) \to [0, \infty]$ is defined by

$$J(\nu \mid \mu) = \frac{1}{2} \int_S \biggl(\frac{d\nu}{d\mu}\biggr)^2 d\mu$$

if $\nu(S) = 0$, $\nu \ll \mu$ and $J(\nu \mid \mu) = \infty$ otherwise (J is also called the Fisher information). Using the convexity of $\mathbb{R} \ni x \mapsto x^2$, it follows that $J(\cdot \mid \mu)$ is convex. $\mathcal{AC}_0[\mathcal{M}(S)]$ is defined to be the set of all maps $\nu : \mathbb{R}_+ \to \mathcal{M}(S)$ with $\nu(0) = 0$ which are absolutely continuous with respect to the variation norm $\|\cdot\|_{\mathrm{var}}$ and possess a weak derivative for almost all t. The latter means that for almost every t the expression $\langle f, \nu(t+h) - \nu(t)\rangle / h$ converges to a limit $\langle f, \dot{\nu}(t)\rangle$ for every $f \in C_b(S, \mathbb{R})$. In [6] the moderate deviation behaviour of $L_n(\cdot)$ had already been considered in a weaker form.

Consider the measures $L_n^m : \Omega \to \mathcal{M}_1(S^m)$ with $n \geqslant m$, defined by

$$L_n^m = \frac{1}{n_{(m)}} \sum_{(i_1,\dots,i_m)\in I(m,n)} \delta_{(X_{i_1},\dots,X_{i_m})}. \tag{1.16}$$

Due to their application, we call the measures $\{L_n^m\}_{n\geqslant m}$ the U-empirical measures of order m. We define the function $J_m(\cdot \mid \mu) : \mathcal{M}(S^m) \to [0,\infty]$ by

$$J_m(\nu \mid \mu) = \frac{1}{2}\int_S \left(\frac{d\nu_1}{d\mu}\right)^2 d\mu$$

if $\nu_1(S) = 0$, $\nu_1 \ll \mu$ and $\nu = \sum_{i=1}^m \mu^{\otimes i-1} \otimes \nu_1 \otimes \mu^{\otimes m-i}$, and we define $J_m(\nu \mid \mu) = \infty$ otherwise. The convexity of $\mathbb{R} \ni x \mapsto x^2$ implies that $J_m(\cdot \mid \mu)$ is convex, too. In [14, Theorem 1.24] the MDP for $\{L_n^m, n \in \mathbb{N}\}$ (with rate function $J_m(\cdot \mid \mu)$) is proved for an arbitrary measurable space $(S^m, \mathcal{S}^{\otimes m})$ on a suitable subset of all signed measures on $(S^m, \mathcal{S}^{\otimes m})$, endowed with a topology stronger than the usual τ-topology which makes maps $\nu \mapsto \int_{S^m} \varphi\, d\nu$ continuous even for certain unbounded φ taking values in a Banach space.

For technical reasons we have to consider functions $\varphi : S^m \to \mathbb{R}^d$ which are *not symmetric*. Otherwise, for $\mathcal{S} \neq \{\emptyset, S\}$ and $m \geqslant 2$ we would not be able to separate the zero measure from the measure $(\delta_x^{\otimes m-1} \otimes \delta_y) - (\delta_y \otimes \delta_x^{m-1})$ with $x \in A \in \mathcal{S}$ and $y \in S \setminus A$ for example, hence the τ-topology would loose the Hausdorff property. Given φ and a nonempty subset A of $\{1,\dots,m\}$, define $\varphi_A \in L_1(\mu^{\otimes |A|})$ by μ-integrating $\varphi(s_1,\dots,s_m)$ with respect to every s_i with $i \in \{1,\dots,m\} \setminus A$. By convention, $\varphi_\emptyset \equiv \int_{S^m} \varphi\, d\mu^{\otimes m} \in \mathbb{R}^d$. Furthermore, define $\tilde{\varphi}_A \in L_1(\mu^{\otimes [t]|A|})$ by

$$\tilde{\varphi}_A\big((s_i)_{i\in A}\big) = \sum_{B\subset A} (-1)^{|A\setminus B|} \varphi_B\big((s_i)_{i\in B}\big), \tag{1.17}$$

for every nonempty $A \subset \{1,\dots,m\}$, and let $\tilde{\varphi}_\emptyset = \varphi_\emptyset$. According to the inclusion–exclusion principle or the Möbius inversion formula,

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

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

$$\int_{S^m} \varphi\, dL_n^m = \tilde{\varphi}_0 + \sum_{c=1}^m \int_{S^c} \tilde{\varphi}_c\, dL_n^c \tag{1.18}$$

$\mathbb{P}$-almost surely, where,

$$\tilde{\varphi}_c \equiv \sum_{\substack{A\subset\{1,\dots,m\}\\ |A|=c}} \tilde{\varphi}_A \tag{1.19}$$

for every $c \in \{0,1,\dots,m\}$. Note that every $\tilde{\varphi}_A$ with nonempty $A \subset \{1,\dots,m\}$ is completely μ-degenerate. For a symmetric φ, formula (1.18) is closely related to the Hoeffding decomposition of the corresponding U-statistic ($\tilde{\varphi}_c = \binom{m}{c}\pi_{c,m}(\varphi)$). Define $L_n^m(t)$ for $t \in [0,1]$ as in (1.16) replacing $I(m,n)$ by $I(m,[nt])$. We will prove a MDP for the process $(L_n^m - \mu^{\otimes m})(\cdot)$ as well as for $(M_n^m - \mu^{\otimes m})(\cdot)$ defined to be the process such that for every $n \geqslant m$ and every $\varphi : S^m \to \mathbb{R}^d$

$$\int_{S^m} \varphi \, dM_n^m(t) = \int_{S^m} \varphi \, d\mu^{\otimes m} + \sum_{c=1}^{m} \int_{S^c} \tilde{\varphi}_c \, dL_n^c(t) \tag{1.20}$$

holds.

THEOREM 1.21 (Moderate deviations of partial sums U-empirical measures). – *The sequence $\{(M_n^m - \mu^{\otimes m})(\cdot), n \in \mathbb{N}\}$ satisfies the MDP in $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ and in $D_2[\mathcal{M}(S^m)]$, respectively, with good rate function*

$$J_M^m\big(\nu(\cdot)\big) = \int_0^\infty J_m(\dot{\nu} \mid \mu) \, dt$$

for $\nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S^m)]$ and $J_\infty^m(\nu(\cdot)) = +\infty$ otherwise. The speed is n/b_n^2. The sequence $\{(L_n^m - \mu^{\otimes m})(\cdot), n \in \mathbb{N}\}$ satisfies the MDP in $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ with the same speed and good rate function

$$J_L^m\big(\nu(\cdot)\big) = \int_0^\infty J_m\big(\dot{\nu}/\big(t^{m-1}\big) \mid \mu\big) \, dt$$

if $\nu(\cdot) \in \mathcal{AC}_0[\mathcal{M}(S^m)]$ and the integral exists and $J_L^m(\nu(\cdot)) = \infty$ otherwise.

Remark 1.22. – Another representation for the rate function $I_W^m(\cdot)$ in Theorem 1.10 is:

$$I_W^m(\phi) = \inf\bigg\{ \int_0^1 J_m(\dot{\nu} \mid \mu) \, dt, \ \nu \in \mathcal{AC}_0\big[\mathcal{M}\big(S^m\big)\big] \cap K_\infty \\ \text{and} \int_{S^m} h \, d\nu(\cdot) = \phi(\cdot) \bigg\}, \tag{1.23}$$

where $K_\infty := \bigcup_{L \geqslant 0} \{\nu(\cdot)\colon \int_0^1 J_m(\dot{\nu} \mid \mu) \, dt \leqslant L\}$. Therefore Theorem 1.10 can be derived via the contraction principle [9, Section 4.2] from Theorem 1.21. A sketch of the proof of (1.23) is given after the proof of Theorem 1.21.

For the proof of Theorem 1.21 we establish the moderate principle for $\{\widetilde{W}_n^m(\cdot, h) - \mathbb{E}(h), n \in \mathbb{N}\}$ when h is a bounded but possibly asymmetric kernel function. The necessity of this step has been explained above. To this end we apply moment inequalities for U-statistics which can be deduced from decoupling and hyper-contractive methods (cf. [7, Sections 2.5–2.7]). Moreover, we use the MDP for L_n^m, established in [14, Theorem 1.24], and apply the Lévy-type maximal inequality for U-statistics (Lemma 2.8) and a Bernstein-type inequality for bounded kernel functions due to Arcones and Giné. The representation of the rate function is deduced from results of Dembo and Zajic [8] combined with an alternative representation of $J_m(\cdot \mid \mu)$.

1.3. Moderate deviations for functional U-processes

Let $\mathcal{H} \subset B(S^m, \mathbb{R})$ be a class of functions such that $0 \leqslant h \leqslant 1$ for all $h \in \mathcal{H}$. The *U-process* of order m indexed by $\mathcal{H}$ is defined as $\{U_n^m(h), h \in \mathcal{H}\}$. U-processes appear in statistics for example as unbiased estimators of the functional $\{\mu^{\otimes m} h\colon\ h \in \mathcal{H}\}$. Arcones and Giné developed the central limit theorem for U-processes in [1]. An overview over the theory of U-processes is [15]. Define the pseudo-metrics $d_2(g,h) = \left(\int_{S^m} (g-h)^2 \, d\mu^{\otimes m}\right)^{1/2}$. Let $l_\infty(\mathcal{H})$ be the Banach space of all bounded real functions on $\mathcal{H}$ with the supremum norm $\|H\|_{\mathcal{H}} = \sup_{h \in \mathcal{H}} |H(h)|$. Let $D_\beta[l_\infty(\mathcal{H})]$ denote the Banach space of càdlàg functions $H : \mathbb{R}_+ \to l_\infty(\mathcal{H})$ such that

$$\|H\|_{\mathcal{H},\beta} := \sup_{t \in \mathbb{R}^+} \frac{\|H_t\|_{\mathcal{H}}}{\beta(t)^m} < \infty,$$

equipped with the Borel σ-field. Every signed measure $\nu \in \mathcal{M}(S^m)$ of finite variation corresponds to an element $E_\nu \in l_\infty(\mathcal{H})$ such that $E_\nu(h) = \int h \, d\nu$ for all $h \in \mathcal{H}$. We regard the random variables $E_{(n/b_n)(M_n^m - \mu^{\otimes m})(\cdot)}$ as elements of $D_\beta[l_\infty(\mathcal{H})]$. The sequence $\{E_{(L_n^m - \mu^{\otimes m})(\cdot)},\ n \in \mathbb{N}\}$ is called a *functional U-process*. Throughout this paper we assume that the class $\mathcal{H}$ is countable.

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

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

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 on the entropy. A class of real functions $\mathcal{H}$ is a Vapnic–Čhervonenkis (VC for short) subgraph class if the subgraphs of the functions in the class form a VC class of sets (subgraph of $h \in \mathcal{H}$: $\{(x,t) \in S^m \times \mathbb{R}\colon\ 0 \leqslant t \leqslant h(x_1, \ldots, x_m) \text{ or } h(x_1, \ldots, x_m) \leqslant t \leqslant 0\}$). For a definition of a VC class see for example [10]. 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 $\{1_C/q(C)\colon\ C \in \mathcal{C}\}$ 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 [25, 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 $\mu^{\otimes m}(H^2) < \infty$,

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

Consider the following type of class $\mathcal{H}$:

Condition 1.24. – 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 $N_2(\varepsilon, \mathcal{H}, \mu) \leqslant (A/\varepsilon)^v$ for all probability measures μ.

Notice that uniformly bounded VC subgraph classes $\mathcal{H}$ of symmetric functions satisfy Condition 1.24.

THEOREM 1.25 (Moderate deviations for functional U-processes). – *Assume that $\mathcal{H}$ is a class of symmetric functions satisfying Condition* 1.24*. Then the sequence $\{E_{(M_n^m-\mu^{\otimes m})(\cdot)}, n \in \mathbb{N}\}$ satisfies the MDP in $D_\beta[l_\infty(\mathcal{H})]$ with speed n/b_n^2 and with good rate function*

$$J^m_{\mathcal{H},\mathcal{M}}\big(H(\cdot)\big) = \inf\left\{ \int_0^\infty J_m(\dot{\nu} \mid \mu)\, dt \colon\ \nu(\cdot) \in \mathcal{AC}_0\big[\mathcal{M}(S^m)\big] \cap K_\infty \text{ and } E_{\nu(\cdot)} = H(\cdot)\right\}.$$

Example 1.26. – There are several examples of classes $\mathcal{H}$ in the literature satisfying the assumptions of Theorem 1.25. We only mention, that the simplicial depth process, empirical distribution functions with the structure of a U-statistic and a class of uniform Hölder functions can be treated. A lot of these examples can be found in [1].

For the proof of Theorem 1.25 we use the MDP for L_n^m [14, Theorem 1.24] and apply the Lévy-type maximal inequality for U-processes (Corollary 2.15) and the Bernstein-type inequality in Proposition 2.24 as essential ingredients. Moreover, we apply the MDP results of [27] as well as Talagrand's isoperimetric inequalities for empirical processes [26]. Again the representation of the rate function is deduced from results of Dembo and Zajic [8] combined with an alternative representation of $J_m(\cdot \mid \mu)$.

2. Decoupling inequalities and consequences

One key for the proofs is a Bernstein type inequality for μ-degenerate U-statistics. For bounded $\mathbb{R}$-valued kernel functions, the proof is given in [1, Proposition 2.3] and in a more detailed version in [15, Theorem 4.1.12]. For bounded kernel functions with values in a Banach space of type 2 a Bernstein type inequality is presented in [7, Theorem 8.15, Corollary 8.1.5]. In [14] the following Bernstein type inequality for unbounded kernel functions with squared norm satisfying the weak Cramér condition is proved.

LEMMA 2.1 (Bernstein-type inequality). – *Consider a symmetric and completely μ-degenerate kernel function $\varphi : S^r \to \mathbb{R}^d$. Assume that there exists an $\alpha > 0$ such that*

$$a \equiv \int_{S^r} \exp\big(\alpha \|\varphi\|^2\big)\, d\mu^{\otimes r} < \infty. \tag{2.2}$$

Define $\sigma^2 = \mathbb{E}[\|\varphi(X_1, \ldots, X_r)\|^2]$. If $\mathbb{E}[\varphi(X_1, \ldots, X_r)] = 0$, then there exist constants c_1, c_2, c_3, depending on r only, and a constant H depending on a and α only such that

$$\mathbb{P}\big(\big\|n^{r/2} U_n^r(\varphi)\big\| \geqslant x\big) \leqslant c_1 \exp\left(-\frac{c_2 x^{2/r}}{\sigma^{2/r} + (c_3 H x^{1/r} n^{-1/2})^{2/(r+1)}}\right) \tag{2.3}$$

for all $x > 0$ and all $n \geqslant r$. Let $\{X_n^k, n \in \mathbb{N}\}$, $k = 1, \ldots, r$, be r independent copies of $\{X_n,\ n \in \mathbb{N}\}$. Then the same inequality (2.3) *holds for the decoupled U-statistics, that is for*

$$\frac{1}{n_{(r)}} \sum_{I(r,n)} \varphi\big(X^1_{i_1}, \ldots, X^r_{i_r}\big).$$

We sketch the proof. Using Chebychev's inequality we obtain for every $t>0$

$$\mathbb{P}\big(\big\|n^{r/2}U_n^r(\varphi)\big\|\geqslant x\big)\leqslant \exp\big(-tx^{2/(r+1)}\big)$$
$$\times\,\mathbb{E}\Bigg[\exp\Bigg(\big(r^r r!\big)^{2/(r+1)}t\Bigg\|n^{-r/2}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}\varphi(X_{i_1},\ldots,X_{i_r})\Bigg\|^{2/(r+1)}\Bigg)\Bigg],$$

since $\binom{n}{r}^{-1}\leqslant n^{-r}r^r r!$. Now we apply Borell's inequality [7, (2.6.5)], the decoupling technique [7, Theorem 2.5.4], a symmetrization lemma [7, Lemma 2.4.5] as well as Hoeffding's formula [7, (1.1.16)]. We get as in the proof of [7, Theorem 8.1.2]:

$$\mathbb{E}\Bigg[\exp\Bigg(\big(r^r r!\big)^{2/(r+1)}t\Bigg\|n^{-r/2}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}\varphi(X_{i_1},\ldots,X_{i_r})\Bigg\|^{2/(r+1)}\Bigg)\Bigg]$$
$$\leqslant c_1\exp\big(c_2\sigma^2t^{r+1}\big)$$
$$\times\,\mathbb{E}\Bigg[\exp\Bigg(2rc_2t^{r+1}n^{-1}\sum_{i=1}^{\lfloor n/r\rfloor}\big(\|\varphi(X_{(i-1)r+1},\ldots,X_{ir})\|^2-\sigma^2\big)\Bigg)\Bigg].\qquad(2.4)$$

Since $a<\infty$ by (2.2), we obtain for each $l\geqslant 2$ the inequality $\int_{S^r}\|\varphi\|^{2l}\,d\mu^{\otimes r}\leqslant\frac{1}{2}l!H^{l-2}\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r}$ with $H\equiv\sup_{l\geqslant 3}(2a)^{1/(l-2)}(\alpha^l\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r})^{-1/(l-2)}<\infty$, cf. [29, Remark 3.6.1]. Thus we can apply [28, Theorem 2.1], especially formula [28, (2.7)]: We obtain

$$\mathbb{E}\Bigg[\exp\Bigg(h\sum_{i=1}^{\lfloor n/r\rfloor}\big(\|\varphi(X_{(i-1)r+1},\ldots,X_{ir})\|^2-\sigma^2\big)\Bigg)\Bigg]\leqslant\exp\bigg(\frac{h^2B_n^2}{2(1-hH)}\bigg)$$

for every $h\in[0,1/(H))$, where $B_n^2=\lfloor n/r\rfloor\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r}$. Considering $h=2rc_2t^{r+1}/n$ we obtain

$$\mathbb{E}\Bigg[\exp\Bigg(\big(r^r r!\big)^{2/(r+1)}t\Bigg\|n^{-r/2}\sum_{1\leqslant i_1<\cdots<i_r\leqslant n}\varphi(X_{i_1},\ldots,X_{i_r})\Bigg\|^{2/(r+1)}\Bigg)\Bigg]$$
$$\leqslant c_1\exp\Bigg(c_2t^{r+1}\sigma^2+\frac{2rc_2^2t^{2(r+1)}\int_{S^r}\|\varphi\|^4\,d\mu^{\otimes r}}{n-2rc_2t^{r+1}H}\Bigg)$$

for all $t^{r+1}\leqslant n(2rc_2H)^{-1}$. Notice that the r-dependent constants change from step to step. Thus we get estimates similar to [1, (2.7), (2.8)] and the result follows by adopting the calculations of the proof of [1, Proposition 2.3(c), p. 1503]. Let $m\in\mathbb{N}$ and $\{X_i^k\colon i\in\mathbb{N}\}_{k=1}^m$ be i.i.d. copies of $\{X_i\colon i\in\mathbb{N}\}$. Regarding notation, we write $\mathbf{i}$ for $(i_1,\ldots,i_m)$, $h(\mathbf{X_i})$ for $h(X_{i_1},\ldots,X_{i_m})$, and $h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})$ for $h(X_{i_1}^1,\ldots,X_{i_m}^m)$ ("dec" standing for "decoupled").

Another key is a Lévy-type maximal inequality for the tails of U-statistics. Although the decoupling method is used to prove some kind of maximal inequalities several times in the literature (see for example [3, Lemma 2.4], [16, Lemma 3.3]), the following result does not seem to exist. The proof was suggested by E. Giné in a private communication.

LEMMA 2.5 (Lévy-type maximal inequality, $m=2$). – *Let $\{X_n,n\in\mathbb{N}\}$ be a sequence of independent identically distributed random variables with values in a* *measurable space* $(S, \mathcal{S})$ *and suppose that the kernel function* $h : S^2 \to \mathbb{R}$ *is symmetric and measurable. Let* $\{X_n^k, n \in \mathbb{N}\}$, $k = 1, 2$, *be two independent copies of* $\{X_n, n \in \mathbb{N}\}$. *Then there exist universal constants* e_1, e_2 *and* f_1, f_2 (*finite and positive*) *such that*

$$\mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{I(2,k)} h(X_i, X_j) \right\| > t\right) \leqslant e_1 \mathbb{P}\left(\left\| \sum_{I(2,n)} h(X_i, X_j) \right\| > e_2 t\right) + f_1 \mathbb{P}\left(\left\| \sum_{i=1}^n h\left(X_i^1, X_i^2\right) \right\| > f_2 t\right). \tag{2.6}$$

Due to the influence of the diagonals, results for U-statistics of arbitrary degree m are more complicated to formulate. Let us define $\sum_{(c)}^* h(X_{i_1}^1, \ldots, X_{i_m}^m)$ for $c = 1, \ldots, m$, where $\sum_{(c)}^*$ is taken over all m-tuples $(i_1, \ldots, i_m)$ formed from the set $\{1, \ldots, c\}$ having exactly c indices distinct. The sum $\sum_{(c)}^*$ contains exactly $c! S_m^{(c)}$ summands, where the quantities $S_m^{(c)}$ are Stirling numbers of the second kind. We obtain

$$\begin{aligned}\sum_{i_1, \ldots, i_m = 1}^n h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) &= \sum_{c=1}^m \sum_{\substack{\mathbf{i}: c \text{ among } i_1, \ldots, i_m \\ \text{distinct}}} h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) \\ &= \sum_{c=1}^m \sum_{1 \leqslant i_1 < \cdots < i_c \leqslant n} \sum_{(c)}^* h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right).\end{aligned} \tag{2.7}$$

For simplicity let us denote by $\left(h_{c*}(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\right)$ the family of all elements of the sum $\sum_{(c)}^*$. For example if $m = 3$, we obtain $\sum_{(1)}^* h(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}) = h(X_{i_1}^1, X_{i_1}^2, X_{i_1}^3)$,

$$\begin{aligned}\sum_{(2)}^* h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) = {} & h\left(X_{i_1}^1, X_{i_1}^2, X_{i_2}^3\right) + h\left(X_{i_1}^1, X_{i_2}^2, X_{i_1}^3\right) + h\left(X_{i_2}^1, X_{i_1}^2, X_{i_1}^3\right) \\ & + h\left(X_{i_1}^1, X_{i_2}^2, X_{i_2}^3\right) + h\left(X_{i_2}^1, X_{i_1}^2, X_{i_2}^3\right) + h\left(X_{i_2}^1, X_{i_2}^2, X_{i_1}^3\right)\end{aligned}$$

and

$$\begin{aligned}\sum_{(3)}^* h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) = {} & h\left(X_{i_1}^1, X_{i_2}^2, X_{i_3}^3\right) + h\left(X_{i_1}^1, X_{i_3}^2, X_{i_2}^3\right) + h\left(X_{i_2}^1, X_{i_1}^2, X_{i_3}^3\right) \\ & + h\left(X_{i_2}^1, X_{i_3}^2, X_{i_1}^3\right) + h\left(X_{i_3}^1, X_{i_1}^2, X_{i_2}^3\right) + h\left(X_{i_3}^1, X_{i_2}^2, X_{i_1}^3\right).\end{aligned}$$

LEMMA 2.8 (Lévy-type maximal inequality). – *Let* $\{X_n, n \in \mathbb{N}\}$ *be a sequence of independent identically distributed random variables with values in a measurable space* $(S, \mathcal{S})$ *and suppose that the kernel function* h *is a symmetric measurable* $\mathbb{R}^d$*-valued function. Let* $\{X_n^k, n \in \mathbb{N}\}$, $k = 1, \ldots, m$, *be* m *independent copies of* $\{X_n, n \in \mathbb{N}\}$. *Then there exist universal constants* e_c^m *and* f_c^m, $c = 1, \ldots, m$ (*finite, positive and depending on* m *only*) *such that*

$$\mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{I(m,k)} h(X_{i_1}, \ldots, X_{i_m}) \right\| > t\right)$$

$$\leqslant e_m^m \mathbb{P}\left(\left\| \sum_{I(m,n)} h(\mathbf{X_i}) \right\| > f_m^m t\right) + \sum_{c=1}^{m-1} \sum_{(c)}^{*} e_c^m \mathbb{P}\left(\left\| \sum_{I(c,n)} h_{c*}(\mathbf{X_i^{\mathrm{dec}}}) \right\| > f_c^m t\right). \quad (2.9)$$

Remark 2.10. – Lemmas 2.5 and 2.8 are true for any symmetric kernel h with values in any separable Banach space.

A decoupling inequality for the tail probabilities of multivariate U-statistics is the key for the proof of Lemmas 2.5 and 2.8. We state here the pertinent result from de la Peña and Montgomery-Smith [24, Theorem 1]:

PROPOSITION 2.11 (Decoupling for tails). – *For natural numbers $n \geqslant m$, let $\{X_n,\ n \in \mathbb{N}\}$ be a sequence of independent random variables with values in a measurable space $(S, \mathcal{S})$, and let $\{X_n^k, n \in \mathbb{N}\}$, $k = 1, \ldots, m$, be m independent copies of this sequence. Let E be a separable Banach space and, for each $(i_1, \ldots, i_m) \in I(n,m)$, let $h_{i_1,\ldots,i_m} : S^m \to E$ be measurable functions. Then there are constants $c_m \in (0, \infty)$, depending on m only, such that for all $t > 0$*

$$\mathbb{P}\left(\left\| \sum_{I(m,n)} h_{i_1,\ldots,i_m}(\mathbf{X_i}) \right\|_E > t\right) \leqslant c_m \mathbb{P}\left(c_m \left\| \sum_{I(m,n)} h_{i_1,\ldots,i_m}(\mathbf{X_i^{\mathrm{dec}}}) \right\|_E > t\right). \quad (2.12)$$

If, moreover, the functions $h_{i_1,\ldots,i_m}$ are symmetric in the sense that, for all $x_1, \ldots, x_m \in S$ and all permutations π of $\{1, \ldots, m\}$,

$$h_{i_1,\ldots,i_m}(x_1, \ldots, x_m) = h_{i_{\pi(1)},\ldots,i_{\pi(m)}}(x_{\pi(1)}, \ldots, x_{\pi(m)}),$$

then the reverse inequality holds true. In particular, there are constants $d_m \in (0, \infty)$ depending on m only, such that for all $t > 0$

$$\mathbb{P}\left(\left\| \sum_{I(m,n)} h_{i_1,\ldots,i_m}(\mathbf{X_i^{\mathrm{dec}}}) \right\|_E > t\right) \leqslant d_m \mathbb{P}\left(d_m \left\| \sum_{I(m,n)} h_{i_1,\ldots,i_m}(\mathbf{X_i}) \right\|_E > t\right). \quad (2.13)$$

Furthermore, for the proof of Lemmas 2.5 and 2.8 we use an extension of the classical Lévy inequality for sums of independent symmetric random vectors to sums of not necessarily symmetric i.i.d. random vectors. The proof has been found in [23, Theorem 1, Corollary 4] (see also [15, Section 2.4]).

PROPOSITION 2.14 (Lévy's inequality for asymmetric random vectors). – *There exist universal constants c_1 and c_2 such that if $\{X_i\}_{i \in \mathbb{N}}$ are i.i.d. E-valued random variables, where $(E, \|\cdot\|_E)$ is a separable Banach space, then, for $1 \leqslant k \leqslant n$,*

$$\mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{i=1}^{k} X_i \right\|_E > t\right) \leqslant c_1 \mathbb{P}\left(\left\| \sum_{k=1}^{n} X_i \right\|_E > c_2 t\right).$$

For example we can choose $c_1 = 9$ and $c_2 = 1/30$.

Propositions 2.11 and 2.14 can be extended to Banach spaces E not necessarily separable, a situation that arises in the context of U-processes. A good reference for details is [15] and references therein. Thus we obtain (details of the proof are omitted):

COROLLARY 2.15. – *There are versions of Proposition* 2.11*, Proposition* 2.14 *and Lemma* 2.8 *for U-processes indexed by arbitrary families of kernels.*

Proof of Lemmas 2.5 and 2.8. – For each $k \in \mathbb{N}$ define functions h_k taking values in $(l_\infty^n, \|\cdot\|_\infty)$, the space $(\mathbb{R}^d)^n$ endowed with the supremum norm, as follows:

$$h_k := (\underbrace{0, \ldots, 0}_{k \text{ times}}, h, h, \ldots, h).$$

Thus it is obvious that

$$\max_{k \leqslant n} \left\| \sum_{I(m,k)} h(X_{i_1}, \ldots, X_{i_m}) \right\| = \left\| \sum_{I(m,n)} h_{i_1 \vee i_2 \vee \cdots \vee i_m}(X_{i_1}, \ldots, X_{i_m}) \right\|_\infty . \tag{2.16}$$

Applying first (2.16), then Proposition 2.11 for $E = l_\infty^n$, and again (2.16) it follows that

$$\begin{aligned} \mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{I(m,k)} h(\mathbf{X_i}) \right\| > t \right) &\leqslant c_m \mathbb{P}\left(c_m \left\| \sum_{I(m,n)} h_{i_1 \vee i_2 \vee \cdots \vee i_m}\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) \right\| > t \right) \\ &= c_m \mathbb{P}\left(c_m \max_{k \leqslant n} \left\| \sum_{I(m,k)} h\left(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\right) \right\| > t \right). \end{aligned} \tag{2.17}$$

For the clarity of exposition we present only the proof of Lemma 2.5. This is notationally much less involved than the general case and already contains the main idea. The general case is proved by iteration. Remark that

$$\begin{aligned} &\mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{I(2,k)} h\left(X_i^1, X_j^2\right) \right\| > t \right) \\ &\quad \leqslant \mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{i,j=1}^{k} h\left(X_i^1, X_j^2\right) \right\| > t/2 \right) + \mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{i=1}^{k} h\left(X_i^1, X_i^2\right) \right\| > t/2 \right). \end{aligned} \tag{2.18}$$

The result can now be obtained by conditionally applying Lévy's inequality (Proposition 2.14) twice: For every $1 \leqslant l \leqslant n$ let s_l be an element in $(l_\infty^n, \|\cdot\|_\infty)$, defined by

$$s_l := \left(\sum_{i=1}^{k} h\left(X_i^1, X_l^2\right) \right)_{k=1}^{n} .$$

We denote by $\mathbb{P}_2$ the conditional probability given $\{X_i^1,\ i \in \mathbb{N}\}$. Now Lévy's inequality applied to the conditionally independent and identically distributed random variables s_l gives

$$\begin{aligned} \mathbb{P}\left(\max_{k \leqslant n} \left\| \sum_{i,j=1}^{k} h\left(X_i^1, X_j^2\right) \right\| > t/2 \right) &\leqslant \mathbb{E}\mathbb{P}_2\left(\max_{k \leqslant n} \left\| \sum_{l=1}^{k} s_l \right\|_\infty > t/2 \right) \\ &\leqslant 9\mathbb{E}\mathbb{P}_2\left(30 \left\| \sum_{l=1}^{n} s_l \right\|_\infty > t/2 \right) = 9\mathbb{P}\left(\max_{k \leqslant n} \left\| 30 \sum_{i=1}^{k} \sum_{j=1}^{n} h\left(X_i^1, X_j^2\right) \right\| > t/2 \right). \end{aligned} \tag{2.19}$$

The inequality follows now by iteration and by (2.13) of Proposition 2.11: the first summand in (2.18) can be bounded by

$$9^2 d_2 \mathbb{P}\left(30^2 d_2 \left\| \sum_{I(2,n)} h(X_i, X_j) \right\| > t/4\right) + 9^2 \mathbb{P}\left(30^2 \left\| \sum_{i=1}^{n} h(X_i^1, X_i^2) \right\| > t/4\right).$$

We apply Proposition 2.14 for the second term in (2.18). Using (2.17) we obtain the asserted inequality with constants $e_1 = c_2 d_2 9^2$ and $e_2 = (4c_2 d_2 30^2)^{-1}$ as well as $f_1 = c_2(9 + 9^2)$ and $f_2 = (4c_2 30^2)^{-1}$. Using (2.7) we obtain the result. To this end we decompose these statistics again and again and conditionally apply Lévy's inequality. The constants e_c^m and f_c^m can be calculated explicitly as a function of the degree m. We omit this calculation, since for our applications it suffices to know, that the constants depend only on m. □

Remark 2.20. – We have to take care of the diagonal terms in the maximal inequality for U-statistics we presented in (2.9). However, since the arguments of these diagonal terms are decoupled, we don't need any additional information about integrability of diagonal terms for our results as we will see in the proof of Theorem 1.10.

Remark 2.21. – Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables defined on the space $(S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}}, \mathbb{P})$ with $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$, independent of the underlying sequence $\{X_i\}_{i\in\mathbb{N}}$. Let $\{\varepsilon_i^k\colon\ i \in \mathbb{N}\}_{k=1,2}$ be independent copies of $\{\varepsilon_i\colon\ i \in \mathbb{N}\}$, independent of $\{X_i^k\colon\ i \in \mathbb{N}\}_{k=1,2}$. For a symmetrized kernel function $\varepsilon_i^1 \varepsilon_j^2 h(X_i^1, X_j^2)$ we can neglect the diagonal terms in the maximal inequality (2.9). This is an immediate consequence of the proof of Lemma 3.3 in [16]. However, inequalities comparing tail probabilities for the random variables $\sum_{I(2,n)} h(X_i, X_j)$ and $\sum_{I(2,n)} \varepsilon_i^1 \varepsilon_j^2 h(X_i, X_j)$ (which are referred to as randomization inequalities) have been obtained only one sided: Theorem 3.5.6, Chapter 3, in [15] yields:

$$\mathbb{P}\left(\left\| \sum_{I(2,n)} \varepsilon_i^1 \varepsilon_j^2 h(X_i^1, X_j^2) \right\| > t\right) \leqslant c\mathbb{P}\left(\left\| \sum_{I(2,n)} h(X_i, X_j) \right\| > ct\right)$$

for some constant c. There is no converse of this inequality in general, even for $m = 1$: see the counterexample in Chapter 3 in [15] (after Theorem 3.5.6). Thus it is not obvious if in general one can neglect diagonal terms.

We will use the following moment inequality for U-statistics established with decoupling and hyper-contractive methods (see for example [7, Theorem 2.7.1, Corollary 2.7.1]).

PROPOSITION 2.22. – *Let E be a Banach space of type p, $1 \leqslant p \leqslant 2$ with norm $\|\cdot\|_E$. Let $\{X_n,\ n \in \mathbb{N}\}$ be a sequence of independent random variables with identical distributions and suppose that the kernel function h is a symmetric E-valued function with rank m and $\mathbb{E}\|h\|_E^q < \infty$, $q \geqslant 1$. Then*

$$\mathbb{E}\left\| \sum_{I(m,n)} h(X_{i_1}, \dots, X_{i_m}) \right\|_E^q$$

$$\leqslant \left(6A^{1/p}m^3\right)^{mq}\left(1 \vee (q-1)^{mq/2}\right)\left(n_{(m)}\right)^{1\vee(q/p)}\mathbb{E}\|h(X_1,\ldots,X_m)\|_E^q, \quad (2.23)$$

where A is the type p constant, depending on E and p only.

Furthermore we will use the following Bernstein-type inequality for U-processes. It is proved with decoupling techniques in [2, Theorem 3.2].

PROPOSITION 2.24 (Bernstein inequality for U-processes). – *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*

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

Then there are constants c and c', depending on m only, such that for all $t > 0$ and $n \geqslant m$

$$\mathbb{P}\left(\left\| n^{-m/2}\sum_{I(m,n)} h(X_{i_1},\ldots,X_{i_m})\right\|_{\mathcal{H}} \geqslant t\right) \leqslant c\exp\left(-c't^{2/m}\right). \quad (2.25)$$

3. Proofs of the moderate deviations results

Proof of Theorem 1.10. – Part (a): We will check that the linear term $k = 1$ in (1.6) is the leading term for the moderate deviations behaviour. Analogously to (1.3), we define a random function $T_n(t)$ for $t \in [0,1]$ by $T_n(0) = 0$ and

$$T_n(t) := U_n^1(t,\pi_{1,m}h) + \left(t - \frac{[nt]}{n}\right)\pi_{1,m}h(X_{[nt]+1}).$$

With the help of condition (1.8) we obtain the MDP for the sample path sequence $\{T_n(\cdot), n \in \mathbb{N}\}$ (as well as for the sequence $\{mT_n(\cdot), n \in \mathbb{N}\}$ via the contraction principle [9, Theorem 4.2.1]) as a consequence of [5]. Notice that $U_n^1(t,\pi_{1,m}h) = \frac{1}{n}\sum_{i=1}^{[nt]}\pi_{1,m}h(X_i)$. We will prove that

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{n}{b_n}\sup_{t\in[0,1]}\left\|\widetilde{W}_n^m(t,h) - \mathbb{E}(h) - mT_n(t)\right\| \geqslant \delta\right) = -\infty \quad (3.1)$$

for all $\delta > 0$. Applying [9, Theorem 4.2.13] we obtain the result. Using (1.6) is suffices to prove that

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{n}{b_n}\max_{i\in\{0,1,\ldots,n\}}\left\|\sum_{k=2}^m\binom{m}{k}U_n^k(i/n,\pi_{k,m}h)\right\| \geqslant \delta\right) = -\infty \quad (3.2)$$

for all $\delta > 0$. By definition the values of $U_n^k(\cdot,\pi_{k,m}h)$ are constant on each interval $[i/n,(i+1)/n)$ for $i \in \{0,\ldots,n-1\}$ and $U_n^k(t,\pi_{k,m}h) = U_n^k(i/n,\pi_{k,m}h)$ for all $t \in [i/n,(i+1)/n)$. Lemma 2.8 yields

$$\begin{aligned}
&\mathbb{P}\left(\max_i \frac{n}{b_n}\|U_n^k(i/n, \pi_{k,m}h)\| \geqslant \delta\right)\\
&\quad\leqslant e_k^k \mathbb{P}\left(\frac{n}{b_n n_{(k)}}\left\|\sum_{I(k,n)}(\pi_{k,m}h)(\mathbf{X_i})\right\| \geqslant f_k^k\delta\right)\\
&\qquad+\sum_{c=1}^{k-1}\sum_{(c)}^{*} e_c^k \mathbb{P}\left(\frac{n}{b_n n_{(k)}}\left\|\sum_{I(c,n)}(\pi_{k,m}h)_{c*}\big(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\big)\right\| \geqslant f_c^k\delta\right).
\end{aligned}\tag{3.3}$$

Lemma 2.1 applied to the first summand yields for each fixed $2 \leqslant k \leqslant m$ and sufficiently large n:

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{n}{b_n}U_n^k(1,\pi_{k,m}h)\right\| \geqslant f_k^k\delta\right) &= \mathbb{P}\left(\|n^{k/2}U_n^k(1,\pi_{k,m}h)\| \geqslant \frac{f_k^k b_n \delta n^{k/2}}{n}\right)\\
&\leqslant c_1 \exp\left(-\frac{c_2\delta^{2/k}b_n^{2/k}n^{1-2/k}}{\sigma^{2/k}+(\delta^{1/k}Hc_3^k)^{2/(k+1)}(b_n/n)^{2/(k(k+1))}}\right),
\end{aligned}\tag{3.4}$$

where the constants c_i depend on k, δ and h only and which might change from step to step. For n sufficiently large this implies

$$\begin{aligned}
&\frac{n}{b_n^2}\log\mathbb{P}\left(\left\|\frac{n}{b_n}U_n^k(1,\pi_{k,m}h)\right\| \geqslant f_k^k\delta\right)\\
&\quad\leqslant \frac{n}{b_n^2}\log c_1 - c_2\left(c_1\left(\frac{n}{b_n}\right)^{2-2/k}\right)\left(\sigma^{2/k}+c_2\left(\frac{b_n}{n}\right)^{2/(k(k+1))}\right)^{-1}.
\end{aligned}\tag{3.5}$$

The right hand side decreases to $-\infty$ as n tends to ∞ by the assumptions on $\{b_n\}_{n\in\mathbb{N}}$. For $c = 1, \ldots, k-1$ and for every $(\pi_{k,m}h)_{c*}$ we obtain that each summand in the double-sum in (3.3) can be bounded by

$$\mathbb{P}\left(\frac{n^{k-(c/2)}}{n_{(k)}}\left\|\sum_{I(c,n)}(\pi_{k,m}h)_{c*}\big(\mathbf{X}_\mathbf{i}^{\mathrm{dec}}\big)\right\| \geqslant \frac{f_c^k\delta b_n n^{k-(c/2)}}{n}\right).\tag{3.6}$$

The assumption that the squared norm of the $\pi_{k,m}h$, $k \geqslant 2$, satisfies the weak Cramér conditions, enables us to apply the Bernstein-type inequality for the functions $(\pi_{k,m}h)_{c*}$ for every $c \in \{1, \ldots, k-1\}$ without any additional assumption for $(\pi_{k,m}h)_{c*}$. This can be done since we obtain in (2.4) in the proof of Lemma 2.1 terms of the form $\|(\pi_{k,m}h)_{c*}(X^1_{(i-1)k+1}, \ldots, X^k_{ik})\|^2$. Hence the application of [28, Theorem 2.1] works with these decoupled entries: $\int_{S^k}\exp(\delta_k\|\pi_{k,m}h\|^2)\,d\mu^{\otimes k} < \infty$ implies

$$\int_{S^k}\exp\big(\delta_k\big\|(\pi_{k,m}h)_{c*}\big(X^1_{i_1}, \ldots, X^k_{i_k}\big)\big\|^2\big)\,d\mathcal{L}\big(X^1_{i_1}\big)\otimes\cdots\otimes d\mathcal{L}\big(X^k_{i_k}\big) < \infty$$

for every $c \in \{1, \ldots, k-1\}$, where $\mathcal{L}(X_1^i)$ denotes the law of X_1^i. Thus (3.2) follows applying Lemma 2.1 in (3.6) and Lemma 1.2.15 in [9].

Part (b): Proving the MDP for $\{\widetilde{U}_n^m(\cdot,h)-\mathbb{E}(h), n\in\mathbb{N}\}$ first we check that

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{n}{b_n}\sup_{t\in[0,1]}\big\|\widetilde{U}_n^m(t,h)-\mathbb{E}(h)-\widehat{T}_n(t)\big\|\geqslant\delta\right)=-\infty \tag{3.7}$$

for all $\delta>0$, where $\widehat{T}_n(t):=mT_n(t)\frac{n\binom{[nt]}{m}}{[nt]\binom{n}{m}}$. Using the representation of $\widetilde{U}_n^m(t,h)$ via Hoeffding's decomposition, this follows from (3.2), since

$$\left\|\sum_{k=2}^m\binom{m}{k}\frac{\binom{[nt]}{m}\binom{n}{k}}{\binom{[nt]}{k}\binom{n}{m}}U_n^k(i/n,\pi_{k,m}h)\right\|\leqslant\sum_{k=2}^m\binom{m}{k}\big\|U_n^k(i/n,\pi_{k,m}h)\big\|.$$

The result follows from the contraction principle (using that multiplying with t^{m-1} is a continuous operation on $L_\infty([0,1],\mathbb{R}^d)$ with respect to the supremum norm) and the (easily proven) fact that

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{n}{b_n}\sup_{t\in[0,1]}\big\|\widehat{T}_n(t)-mT_n(t)t^{m-1}\big\|\geqslant\delta\right)=-\infty$$

for every $\delta>0$. Here we use

$$\left|\frac{n\binom{[nt]}{m}}{[nt]\binom{n}{m}}-t^{m-1}\right|\leqslant\frac{C}{n}\quad\text{for } t\in[0,1], nt\geqslant 1$$

and apply [1, Proposition 2.3(d)]. □

Proof of Theorem 1.21. – We follow the lines of the proof of Theorem 1(b) in [8]. We will first check that for fixed $d\in\mathbb{N}$ and $f\in B(S^m,\mathbb{R}^d)$ the sequence $\{\langle f,(M_n^m-\mu^{\otimes m})(\cdot)\rangle, n\in\mathbb{N}\}$ satisfies a MDP in $D_1[\mathbb{R}^d]$ with a convex good rate function

$$I_f(\phi)=\int_0^\infty\Lambda^*(\dot\phi)\,dt$$

for $\phi\in\mathcal{AC}_0(\mathbb{R}_+,\mathbb{R}^d)$ and $I_f(\phi)=+\infty$ otherwise. Here $\Lambda^*(\theta)=\sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda,\theta\rangle-\frac{1}{2}\int(\langle\lambda,\widetilde{f}_1-\int\widetilde{f}_1\,d\mu\rangle)^2\,d\mu\}$, where $\widetilde{f}_1$ denotes the completely μ-degenerate function for f as in (1.19). To do this we first prove the MDP on the space $D([0,1],\mathbb{R}^d)$, the space of càdlàg functions from $[0,1]$ to $\mathbb{R}^d$, equipped with the topology of uniform convergence: if we can show that the non-linear terms of the decomposition (1.20) do not contribute to the MDP the result follows from [5] (see also [22] and [4]). The polygonal approximation $\widetilde{T}_n(\cdot)$ of $T_n(t):=\langle\widetilde{f}_1,L_n^1(t)\rangle$ satisfies the MDP and the MDP for $\{T_n(\cdot),n\in\mathbb{N}\}$ follows since $\sup_{t\in[0,1]}\|n/b_n(T_n(t)-\widetilde{T}_n(t))\|\leqslant L/b_n$ for some constant $L<\infty$ (cf. [9, Theorem 4.2.13]). Hence the result on $D([0,1],\mathbb{R}^d)$ follows if for every $\eta>0$ and every $a\in\{2,\dots,m\}$

$$\limsup_{n\to\infty}\frac{n}{b_n^2}\log\mathbb{P}\left(\frac{n}{b_n}\sup_{t\in[0,1]}\left\|\int_{S^a}\tilde{f}_a\,dL_n^a(t)\right\|\geqslant\eta\right)=-\infty, \tag{3.8}$$

where $\tilde{f}_a$ denotes the completely μ-degenerate function for f as in (1.19). Define $k_n = b_n^2/n$. Due to (1.2), we may assume that $k_n \geqslant 2$ for all $n \in \mathbb{N}$. Using Chebychev's inequality for the first step and Proposition 2.22 for the second step we obtain the estimate

$$\mathbb{P}\left(\frac{n}{b_n}\left\|\int_{S^a} \tilde{f}_a \, dL_n^a(1)\right\| \geqslant \eta\right) \leqslant \left(\frac{n}{b_n \eta n_{(a)}}\right)^{k_n} \mathbb{E}\left[\left\|\sum_{(i_1,\ldots,i_a)\in I(a,n)} \tilde{f}_a(\mathbf{X}_\mathbf{i})\right\|^{k_n}\right]$$
$$\leqslant \left(\frac{n}{b_n \eta \sqrt{n_{(a)}}}\right)^{k_n} \left(6a^3\sqrt{\mathrm{const}(k_n-1)}\right)^{ak_n} \int_{S^a} \|\tilde{f}_a\|^{k_n} \, d\mu^{\otimes a}.$$

Moreover with the notations of (2.7) we obtain for every $c = 1, \ldots, a-1$ a similar bound for

$$\mathbb{P}\left(\frac{n}{b_n n_{(a)}}\left\|\sum_{I(c,n)} (\tilde{f}_a)_{c*}(\mathbf{X}_\mathbf{i}^{\mathrm{dec}})\right\| \geqslant \eta\right)$$

via Chebychev's inequality and Proposition 2.22. Using Lemma 2.8 this implies (3.8), since f is bounded. The result can easily be adapted to cover the time interval $[0,T]$ instead of $[0,1]$. Next we apply the projective limit approach Theorem 4.6.1 of [9] for $T \in \mathbb{N}$ and the MDP in $D_1[\mathbb{R}^d]$ for the sequence $\{\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle, n \in \mathbb{N}\}$ follows (the representation of the rate function follows as in [9, (5.1.11)]). The MDP for $\{\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle, n \in \mathbb{N}\}$ with the same rate function extends to the space $D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ (and hence also to $D_2^{\mathrm{proj}}[\mathbb{R}^d]$). To this end first note that almost surely $(n/b_n)\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle \in D_\beta[\mathbb{R}^d] \subset D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ (this is the LIL for non-degenerate U-statistics, see for example [2, Corollary 3.5]). Moreover, we will show that $\mathcal{D}_{I_f} := \{\phi\colon I_f(\phi) < \infty\} \subset D_\beta[\mathbb{R}^d] \subset D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$. Let ϕ be chosen such that $I_f(\phi) \leqslant L$, $L > 0$. It follows that $\phi(t) \leqslant \sqrt{LMt}$ for every $t \in \mathbb{R}_+$, where M is a constant such that $\Lambda^*(y) \geqslant \|y\|^2/M$. Thus $\phi \in D_\beta[\mathbb{R}^d]$. Notice that the topology on $D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ is generated by the family of metrics $d_T(y,z) := \sup_{t\in[0,T]} \|y(t) - z(t)\|/(\beta(t))$, $y, z \in D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ with $T > 0$. This family is separating. We apply the concept of exponential equivalence in completely regular topological spaces [13, Theorem 1.6]: By Chebychev's inequality, Lemma 2.8 and Proposition 2.22 we obtain that for every $\eta > 0$ and every $T > 0$

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} \sup_{t\in[0,T]} \left\|\int_{S^a} \tilde{f}_a \, dL_n^a(t)\right\| / \beta(t) \geqslant \eta\right) = -\infty \tag{3.9}$$

using similar arguments as in the proof of (3.8). Applying the contraction principle [9, Theorem 4.2.23], Etemadi's maximal inequality and Hoeffdings's inequality, we get the MDP for $\{\int_S \tilde{f}_1 \, dL_n(\cdot), n \in \mathbb{N}\}$ (even) in $D_\beta^1[\mathbb{R}^d]$ as in the proof of [8, Theorem 1(b)]. Here $D_\beta^1[\mathbb{R}^d]$ denotes the space of all càdlàg functions $f:\mathbb{R}_+ \to \mathbb{R}^d$ such that $\sup_{t\in\mathbb{R}_+} \|f(t)\|/\beta(t) < \infty$. The MDP in $D_\beta^{\mathrm{proj}}[\mathbb{R}^d]$ follows immediately.

In order to refine the MDP for $\{\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle, n \in \mathbb{N}\}$ to the topology induced by the norm $\|f\|_\beta := \sup_{t\in\mathbb{R}_+} \|f(t)\|/(\beta(t)^m)$ (and hence also to $D_2[\mathbb{R}^d]$), we apply the contraction principle [9, Theorem 4.2.23]. Let F denote a measurable extension to $D_1[\mathbb{R}^d]$ of the identity map on $D_\beta[\mathbb{R}^d]$ and consider its continuous approximations $F_l : D_1[\mathbb{R}^d] \to D_\beta[\mathbb{R}^d]$ such that $F_l(y)(t) = y(t)1_{[0,l)}(t)$. We have already checked that $\mathbb{P}(\frac{n}{b_n}\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle \in D_\beta[\mathbb{R}^d]) = 1$ and since $\phi(t) \leqslant \sqrt{LMt}$ for all ϕ with $I_f(\phi) \leqslant L$, we obtain for every $L < \infty$

$$\lim_{l\to\infty} \sup_{\{\phi:\ I_f(\phi)\leqslant L\}} \|F_l(\phi) - F(\phi)\|_\beta = 0.$$

Thus the MDP extends to the Banach space $D_\beta[\mathbb{R}^d]$ (and hence also to $D_2[\mathbb{R}^d]$) if we can show that for every $\eta > 0$ and for each $a \in \{1, \dots, m\}$

$$\lim_{l\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} \sup_{t\geqslant l} \frac{\|\int_{S^a} \tilde{f}_a \, dL_n^a(t)\|}{\beta(t)^m} \geqslant \eta\right) = -\infty.$$

Notice that for a symmetric and completely μ-degenerate function $g : S^r \to \mathbb{R}^d$ with $\|g\| \leqslant c$ we get by [1, Proposition 2.3(d)]:

$$\mathbb{P}\left(\left\| n^{-r/2} \sum_{(i_1,\dots,i_r)\in I(r,n)} g(\mathbf{X_i})\right\| \geqslant x\right) \leqslant c_1 \exp\bigl(-c_2 (x/c)^{2/r}\bigr) \tag{3.10}$$

for all $x > 0$ and all $n \geqslant r$, where the constants c_i depend only on r. The same inequality holds for $\mathbf{X_i}$ replaced by $\mathbf{X_i^{\mathrm{dec}}}$. Lemma 2.8 and (3.10) yield

$$\begin{aligned}
&\mathbb{P}\left(\frac{n}{b_n} \sup_{t\geqslant l} \frac{\|\int_{S^a} \tilde{f}_a \, dL_n^a(t)\|}{\beta(t)^m} \geqslant \eta\right) \\
&\quad\leqslant \sum_{k=0}^{\infty} \mathbb{P}\left(\max_{nl2^k \leqslant j \leqslant nl2^{k+1}} \left\| \sum_{(i_1,\dots,i_a)\in I(a,j)} \tilde{f}_a(\mathbf{X_i})\right\| \geqslant \frac{\eta n_{(a)} b_n \beta(l2^k)^m}{n}\right) \\
&\quad\leqslant \sum_{k=0}^{\infty} \left(\sum_{c=1}^{a-1} \sum_{(c)}^{*} e_c^a \mathbb{P}\left(\frac{1}{(nl2^{k+1})^{c/2}} \left\| \sum_{\mathbf{i}\in I(c,l2^{k+1})} (\tilde{f}_a)_{c*}(\mathbf{X_i^{\mathrm{dec}}})\right\| \geqslant \frac{\eta f_c^a n_{(a)} b_n \beta(l2^k)^m}{n(nl2^{k+1})^{c/2}}\right)\right. \\
&\qquad \left. + e_a^a \mathbb{P}\left(\frac{1}{(nl2^{k+1})^{a/2}} \left\| \sum_{\mathbf{i}\in I(a,l2^{k+1})} \tilde{f}_a(\mathbf{X_i})\right\| \geqslant \frac{\eta f_a^a n_{(a)} b_n \beta(l2^k)^m}{n(nl2^{k+1})^{a/2}}\right)\right) \\
&\quad\leqslant \sum_{k=0}^{\infty} \sum_{c=1}^{a} \tilde{e}_c^a \exp\left(-\eta^2 \tilde{f}_c^a \frac{(b_n n_{(a)})^{2/c}}{n^{1+2/c}} \log(\log l + k \log 2)\right),
\end{aligned} \tag{3.11}$$

where we have assumed that $\eta < 1$ and therefore $\eta^{2/c} > \eta^2$. Hence

$$\begin{aligned}
&\mathbb{P}\left(\frac{n}{b_n} \sup_{t\geqslant l} \frac{\|\int_{S^a} \tilde{f}_a \, dL_n^a(t)\|}{\beta(t)^m} \geqslant \eta\right) \\
&\quad\leqslant \sum_{k=0}^{\infty} c_1(a) \exp\left(-\eta^2 c_2(a) \frac{(b_n n_{(a)})^{2/c}}{n^{1+2/c}} \log(\log l + k \log 2)\right).
\end{aligned}$$

Now we use that

$$\frac{(b_n n_{(a)})^{2/c}}{n^{1+2/c}} \geqslant \frac{1}{n}\left(\frac{b_n n_{(a)}}{n}\right)^{2/a} = n^{1-1/a}\left(\frac{b_n^2 n_{(a)}^2}{n^{2a+1}}\right)^{1/a} \to \infty,$$

and that $\sum_{k=0}^{\infty} \exp(-C\log(\log l + k\log 2)) < \infty$ if and only if $C > 1$. We obtain

$$\frac{n}{b_n^2}\log \mathbb{P}\left(\frac{n}{b_n}\sup_{t\geqslant l}\frac{\|\int_{S^a}\tilde{f}_a\,dL_n^a(t)\|}{\beta(t)^m} \geqslant \eta\right) \\ \leqslant \frac{n}{b_n^2}\log c_1(a) - \frac{n}{b_n^2}\frac{(b_n n_{(a)})^{2/c}}{n^{1+2/c}}\log\log l + \frac{n}{b_n^2}\,\text{const}\,.$$

For $a=1$, taking $n\to\infty$ and then $l\to\infty$ yields the result. Since for $a\in\{2,\dots,m\}$ the right hand side converges to $-\infty$ for $n\to\infty$, independent of l, we actually have proved that all terms of order $a\geqslant 2$ do not contribute to the MDP in $D_\beta[\mathbb{R}^d]$ and therefore the result in $D_\beta[\mathbb{R}^d]$ is proved.

We equip $D_2[B'(S^m,\mathbb{R})]$, the space of all càdlàg functions from $\mathbb{R}_+$ to $B'(S^m,\mathbb{R})$ with the weakest topology such that the maps $y(\cdot)\mapsto\langle\phi, y(\cdot)\rangle : D_2[B'(S^m,\mathbb{R})]\mapsto D_2[\mathbb{R}]$ are continuous and with the minimal σ-field $\mathcal{B}_2$ for which these maps are measurable. Here $\phi\in B(S^m,\mathbb{R})$. Moreover choose $\mathcal{B}_2$ such that $D_2[\mathcal{M}(S^m)]\in\mathcal{B}_2$. We apply the projective limit approach for càdlàg function spaces, introduced in [8, Theorem A.1]. Hence we obtain the MDP for $\{(M_n^m - \mu^{\otimes})(\cdot), n\in\mathbb{N}\}$ in the space $D_2[B'(S^m,\mathbb{R})]$. The corresponding good rate function has the form

$$J_\infty^m\big(\nu(\cdot)\big) = \sup_{l\in\mathbb{N}, 0<t_1<\cdots<t_l<\infty}\sum_{i=1}^{l}(t_i - t_{i-1})J_m\left(\frac{\nu(t_i)-\nu(t_{i-1})}{t_i - t_{i-1}}\middle|\mu\right). \tag{3.12}$$

Since $\{L_n^m(1), n\in\mathbb{N}\}$ satisfies the MDP in $B'(S^m,\mathbb{R}^d)$ with the convex rate function $J_m(\cdot\mid\mu)$ (cf. [14, Theorem 1.21]), this representation follows from Lemma A.3 in [8]. Moreover since both, $\{(n/b_n)(M_n^m - \mu^{\otimes m})(\cdot), n\in\mathbb{N}\}$ and $\{\nu(\cdot)\colon\ J_\infty^m(\nu(\cdot)) < \infty\}$, are subsets of $D_2[\mathcal{M}(S^m)]$ the MDP also holds in this space by Lemma 4.1.5 in [9]. Next we define

$$K_m(\nu\mid\mu^{\otimes m}) = \frac{1}{2}\int_{S^m}\left(\frac{d\nu}{d\mu^{\otimes m}}\right)^2 d\mu^{\otimes m},$$

if $\nu\ll\mu^{\otimes m}$ and $\nu(S^m)=0$ and $K_m(\nu\mid\mu^{\otimes m})=\infty$ otherwise. Let $\nu\in\mathcal{M}(S^m)$ satisfy $J_m(\nu\mid\mu)<\infty$. Then the representation $\nu=\sum_{i=1}^m \mu^{\otimes i-1}\otimes\widetilde{\nu}\otimes\mu^{\otimes m-i}$ implies that

$$\frac{d\nu}{d\mu^{\otimes m}}(s_1,\dots,s_m) = \sum_{i=1}^m\frac{d\widetilde{\nu}}{d\mu}(s_i)\quad \mu^{\otimes m}\text{-a.s.}$$

Using the definition of $K_m(\cdot\mid\mu^{\otimes m})$, $\widetilde{\nu}(S)=0$ and the definition of $J_m(\cdot\mid\mu)$, it follows that

$$K_m(\nu\mid\mu^{\otimes m}) = mJ_m(\nu\mid\mu).$$

Since $K_m(\nu \mid \mu^{\otimes m}) \geqslant \|\nu\|_{\mathrm{var}}^2/2$ we obtain $J_m(\nu \mid \mu) \geqslant \|\nu\|_{\mathrm{var}}^2/2m$. Using again the convexity of $J_m(\cdot \mid \mu)$, we can apply Lemma A.6 in [8] to get $J_\infty^m(\nu(\cdot)) = J_M^m(\nu(\cdot))$ for all relevant paths. Remark that the result in $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ as well as in $D_1[\mathcal{M}(S^m)]$ follows by the same arguments.

In order to proof the MDP for the sequence $\{(L_n^m - \mu^{\otimes m})(\cdot), n \in \mathbb{N}\}$ we use the fact that the weak topology on $D_2^{\mathrm{proj}}[\mathcal{M}(S^m)]$ is generated by the family of pseudo-metrics

$$d_{f,T}(y,z) := \sup_{t\in[0,T]} \left\| \int_{S^m} f\,dy(t) - \int_{S^m} f\,dz(t) \right\| \Big/ (t+1), \qquad y,z \in D_2^{\mathrm{proj}}\big[\mathcal{M}\big(S^m\big)\big],$$

with $f \in B(S^m, \mathbb{R}^d)$ and $T > 0$. This family is separating. With [13, Theorem 1.6] the proof is an adaption of the proof of part (b) in the proof of Theorem 1.10. We omit the details. □

Remark 3.13. – The calculations in (3.11) show that

$$\lim_{l\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\sup_{t\geqslant l} \left(\frac{n}{b_n}\right)^a \left\| \int_{S^a} \tilde{f}_a\,dL_n^a(t) \right\| \Big/ \big(\beta(t)\big)^a \geqslant \eta \right) = -\infty$$

for all $1 \leqslant a \leqslant m$.

Proof of Remark 1.22. – We only give a sketch of the proof of (1.23). A similar proof was already given in [12, Corollary 2.7]. We have to check that

$$\inf\left\{ \int_0^1 J_m(\dot{\nu} \mid \mu)\,dt,\ \nu \in \mathcal{AC}_0\big[\mathcal{M}\big(S^m\big)\big] \cap K_\infty \text{ and } \int_{S^m} h\,d\nu(\cdot) = \phi(\cdot) \right\} = I_W^m(\phi). \quad (3.14)$$

Assume that the right hand side is finite. Using $J_\infty^m(\nu(\cdot)) = J_M^m(\nu(\cdot))$ and the identity (3.12) we obtain that the left hand side is greater than or equal to

$$\inf\left\{ \sum_{i=1}^k (t_i - t_{i-1}) J_m\left(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}} \Big| \mu \right);\ \int_{S^m} h\,d\nu(t_i) = \phi(t_i),\ 1 \leqslant i \leqslant k \right\} \qquad (3.15)$$

for arbitrary $0 = t_0 < t_1 < \cdots < t_k \leqslant 1$. Consider the case $k = 1$ and $t_1 = 1$, then we obtain

$$\inf\left\{ J_m(\nu(1) \mid \mu);\ \int_{S^m} h\,d\nu(1) = \phi(1),\ \nu(1) \in K_{1,\infty} \right\} = \Lambda_m^*(\phi(1)/m),$$

where $K_{1,\infty} = \bigcup_{L\geqslant 0}\{\varrho \in \mathcal{M}(S^m)\colon\ J_m(\varrho \mid \mu) \leqslant L\}$ and $\Lambda_m^*(\cdot)$ is the convex dual in the statement of Theorem 1.10. This fixed-time variational identity follows from [14, Lemma 3.19]. A simple argument, using the linearity of $\nu \mapsto \int_{S^m} h\,d\nu$ and the convexity of $J_m(\cdot \mid \mu)$ yields that (3.15) is equal to

$$\sum_{i=1}^{k}(t_i - t_{i-1})\Lambda_m^*\left(\frac{\phi(t_i) - \phi(t_{i-1})}{m(t_i - t_{i-1})}\right).$$

Since this is true for every $(t_0, \ldots, t_k)$ we obtain by Lemma 5.1.6 in [9] that the right hand side in (3.14) is smaller than the left hand side. In the case $\int_0^1 \Lambda^*(\dot{\phi})\,dt = \infty$ it follows easily that $\{\nu(\cdot)\colon \nu(\cdot) \in K_\infty \quad \text{and} \quad \int h\,d\nu = \phi(\cdot)\} = \emptyset$. The other inequality, which is actually more involved, can be proved following step by step the proof of [12, Corollary 2.7]. $\square$

Proof of Theorem 1.25. – First we will check that in our scaling the non-linear terms of the Hoeffding decomposition do not contribute to the MDP. Denote by R_n^m the empirical measure process such that for every $n \geqslant m$ and every $\varphi : S^m \to \mathbb{R}^d$

$$\int_{S^m} \varphi\,dR_n^m(t) = \sum_{c=2}^{m} \int_{S^c} \tilde{\varphi}_c\,dL_n^c(t).$$

Let us define

$$L_n^{m,1} := \sum_{i=1}^{m} \mu^{\otimes i-1} \otimes L_n^1 \otimes \mu^{\otimes m-i}. \tag{3.16}$$

It suffices to prove that every $\eta > 0$

$$\lim_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\big(\|E_{(n/b_n)R_n^m}\|_{\mathcal{H},\beta} > \eta\big) = -\infty. \tag{3.17}$$

Assume that $\mathcal{H}$ satisfies Condition 1.24. Note that for any probability measure ν on S^a the quantity $\|\pi_{a,m}h_1 - \pi_{a,m}h_2\|_{L_2(\nu)}$ is dominated by the sum of 2^a L_2-distances. Thus the condition $N_2(\varepsilon, \mathcal{H}, \mu) \leqslant (A/\varepsilon)^v$ (for all probability measures μ) implies that the classes of canonical functions $\{\pi_{a,m}h\colon h \in \mathcal{H}\}$ satisfy part (b) of the conditions in Proposition 2.24. With (2.25) and Corollary 2.15 (maximal inequality) we get a bound for

$$\mathbb{P}\left(\sup_{t\geqslant l} \frac{n}{b_n}\big\|R_n^m(t)\big\|_{\mathcal{H}}/(\beta(t)^m) \geqslant \eta\right)$$

similar to (3.11) and since the calculations are true for any $l \geqslant 0$, we get (3.17). In the remaining part we prove that the sequence $\{E_{L_n^{m,1}(\cdot)}, n \in \mathbb{N}\}$ satisfies the MDP with the correct rate function. If $\mathcal{H}$ satisfies Condition 1.24 we can use the proof of Corollary 5.7 and the following remark in [1], using the uniform boundedness of $\mathcal{H}$, to get the following remarkable fact: all projections $\pi_{a,m}\mathcal{H} = \{\pi_{a,m}h\colon h \in \mathcal{H}\}$, $a = 1, \ldots, m$, satisfy the Central Limit Theorem in $l_\infty(\mathcal{H})$. Especially this implies using [19, Theorem 14.6] that the class $\{\pi_{1,m}h\colon h \in \mathcal{H}\}$ satisfies the sufficient (and necessary) conditions of Theorem 2 in [27]: the class is totally bounded and $E_{(n/b_n)L_n^{m,1}(\cdot)} \to 0$ in probability in $l_\infty(\mathcal{H})$ for $n \to \infty$. Notice that $\{\pi_{1,m}h\colon h \in \mathcal{H}\}$ is uniformly bounded.

For fixed $k \in \mathbb{N}$ let $\mathcal{F}_k \in \mathcal{H}$ be finite $1/k$-nets of $\mathcal{H}$, i.e. $\sup_{h\in\mathcal{H}} \min_{f\in\mathcal{F}_k} d_2(h, f) \leqslant 1/k$. Define $f_k(h)$ via $d_2(h, f_k(h)) = \min_{f\in\mathcal{F}_k} d_2(h, f)$ with tie-breaking such that $h \mapsto f_k(h)$ is Borel measurable in $(\mathcal{H}, d_2)$. Moreover, denote $E_{k,\nu(\cdot)} = \langle f_k(\cdot), \nu\rangle \in l_\infty(\mathcal{H})$. Adopting the arguments of [8, Theorem 2(b)], which uses the isoperimetric inequality of [26, Theorem 3.5] as well as Etemadi's maximal inequality, we obtain

$$\lim_{k\to\infty} \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\left\| E_{\frac{n}{b_n} L_n^{m,1}} - E_{k,\frac{n}{b_n} L_n^{m,1}} \right\|_{\mathcal{H},\beta} > \eta \right) = -\infty.$$

For the last step of the proof we adapt the proof of [8, Theorem 2(b)] using some properties of the rate function $J_m(\cdot \mid \mu)$. Let $\mathcal{F}_k = \{f_1, \ldots, f_d\}$ for some $d = d_k$ and denote $f = (f_1, \ldots, f_d)$. By Theorem 1.21, applying the contraction principle, we know that $\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle$ satisfies the MDP in $D_2[\mathbb{R}^d]$ with the good rate function $I_f(\phi(\cdot)) = \inf_{\{\nu(\cdot):\ \phi(\cdot)=\langle f,\nu(\cdot)\rangle\}} J_M^m(\nu(\cdot))$. Denote by

$$K_L := \bigl\{\nu(\cdot):\ J_M^m\bigl(\nu(\cdot)\bigr) \leqslant L\bigr\} \subset D_2\bigl[\mathcal{M}\bigl(S^m\bigr)\bigr].$$

With (3.12) it follows that $t J_m\bigl(\frac{\nu(t)}{t}\big|\mu\bigr) \leqslant J_M^m(\nu(t))$ for every $t > 0$. For each $\nu(\cdot) \in K_L$ and every $t > 0$ there exists a density $d\nu(t)/d\mu^{\otimes m}$ such that

$$\frac{1}{2t} \int\limits_{S^m} \left(\frac{d\nu(t)}{d\mu^{\otimes m}}\right)^2 d\mu^{\otimes m} = t K_m\left(\frac{\nu(t)}{t}\middle|\mu^{\otimes m}\right) = tm J_m\left(\frac{\nu(t)}{t}\middle|\mu\right) \leqslant L\, m.$$

Thus $\phi(t) \leqslant \sqrt{2tLm}$ for $I_f(\phi(\cdot)) < \infty$ and is follows that $\{\phi(\cdot):\ I_f(\phi(\cdot)) < \infty\} \subset D_\beta[\mathbb{R}^d]$. We have seen in the proof of Theorem 1.21 that $\langle f, (M_n^m - \mu^{\otimes m})(\cdot)\rangle$ satisfies the MDP in $D_\beta[\mathbb{R}^d]$. The uniqueness of the rate function (cf. [9, Lemma 4.1.4]) and [9, Lemma 4.1.5] imply that the rate function is $I_f(\cdot)$. Identifying $E_{k,\nu(\cdot)}$ with $\langle f, \nu(\cdot)\rangle$ (remember that f depends on k) the MDP in $D_\beta[l_\infty(\mathcal{H})]$ with the good rate

$$J_{\mathcal{H},k}^m\bigl(H(\cdot)\bigr) = \inf\bigl\{J_M^m\bigl(\nu(\cdot)\bigr):\ \nu(\cdot) \in \mathcal{AC}_0\bigl[\mathcal{M}\bigl(S^m\bigr)\bigr] \cap K_\infty \text{ and } E_{k,\nu(\cdot)} = H(\cdot)\bigr\}$$

follows via the contraction principle [9, Theorem 4.2.23]. The mapping $\nu(\cdot) \mapsto E_{k,\nu(\cdot)} : K_L \to D_\beta[l_\infty(\mathcal{H})]$ is continuous for every k. Define

$$\mathcal{H}_\eta := \{h - g:\ h, g \in \mathcal{H} \text{ and } d_2(h, g) \leqslant \eta\}.$$

For $\nu(\cdot) \in K_L$ it follows that

$$\|\nu(t)\|_{\mathcal{H}_{1/k}} \leqslant \sup_{\{h:\ \int_{S^m} h^2 d\mu^{\otimes m} \leqslant 1/k^2, \|h\|\leqslant 1\}} \left\| \int\limits_{S^m} h \frac{d\nu(t)}{d\mu^{\otimes m}}\, d\mu^{\otimes m} \right\| \leqslant c(m) \frac{\sqrt{2Ltm}}{k},$$

where $c(m)$ is a constant, depending only on m. Since $h - f_k(h) \in \mathcal{H}_{1/k}$ for every $h \in \mathcal{H}$ we conclude that $\|E_{\nu(t)} - E_{k,\nu(t)}\|_{\mathcal{H}} \leqslant \|\nu(t)\|_{\mathcal{H}_{1/k}}$ for every $\nu(t) \in \mathcal{M}(S^m)$. Hence we obtain that $\|E_\nu - E_{k,\nu}\|_{\mathcal{H},\beta} \to 0$ as $k \to \infty$, uniformly over K_L. Thus we can apply the contraction principle [9, Theorem 4.2.23] and the result is proved. $\square$

Acknowledgement

I would like to thank Prof. E. Giné for discussions concerning Lemma 2.8 and an anonymous referee for having read the paper with great care and having made a number of very useful remarks which corrected the exposition.

REFERENCES

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

[2] Arcones M.A., Giné E., 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 (1994) 17–38.

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

[4] Baldi P., Large deviations and stochastic homogenization, Ann. Mat. Pura Appl. 151 (1988) 161–177.

[5] Borovkov A.A., Mogulskii A.A., Probabilities of large deviations in topological spaces I, Siberian Math. J. 19 (1978) 697–709.

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

[7] Borovskikh Yu.V., U-Statistics in Banach Spaces, VSP, Utrecht, 1996.

[8] Dembo A., Zajic T., Uniform large and moderate deviations for functional empirical processes, Stochastic Process. Appl. 67 (1997) 195–211.

[9] Dembo A., Zeitouni O., Large Deviations Techniques and Applications, Springer, New York, 1998.

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

[11] Eichelsbacher P., Large deviations for products of empirical probability measures in the τ-topology, J. Theoret. Probab. 10 (4) (1997) 903–920.

[12] Eichelsbacher P., Löwe M., Large deviations for partial sums U-processes, Theory Probab. Appl. 43 (1998) 97–115.

[13] Eichelsbacher P., Schmock U., Exponential approximations in completely regular topological spaces and extensions of Sanov's theorem, Stochastic Process. Appl. 77 (2) (1998) 233–251.

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

[15] Giné E., de la Peña V., Decoupling: From Dependence to Independence, Springer-Verlag, 1999.

[16] Giné E., Zhang C.-H., On integrability in the LIL for degenerate U-statistics, J. Theoret. Probab. 9 (2) (1996) 385–412.

[17] Hall P., On the invariance principle for U-statistics, Stochastic Process. Appl. 9 (1979) 163–174.

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

[19] Ledoux M., Talagrand M., Probability in Banach Spaces, Springer-Verlag, Berlin, 1991.

[20] Mandelbaum A., Taqqu M.S., Invariance principle for symmetric statistics, Ann. Statist. 12 (1984) 483–496.
[21] Miller R.G., Sen P.K., Weak convergence of U-statistics and von Mises differentiable, Ann. Math. Statist. 43 (1972) 31–41.
[22] Mogulskii A.A., Large deviations for trajectories of multidimensional random walks, Theory Probab. Appl. 21 (1976) 300–315.
[23] Montgomery-Smith S.J., Comparison of sums of independent identically distributed random vectors, Probab. Math. Statist. 14 (2) (1994) 281–285.
[24] Montgomery-Smith S.J., de la Peña V., Decoupling inequalities for the tail probabilities of multivariate U-statistics, Ann. Probab. 23 (2) (1995) 806–816.
[25] Pollard D., Convergence of Stochastic Processes, Springer, New York, 1984.
[26] Talagrand M., Sharper bounds for gaussian and empirical processes, Ann. Probab. 22 (1) (1994) 28–76.
[27] Wu L., Large deviations, moderate deviations and LIL for empirical processes, Ann. Probab. 22 (1) (1994) 17–27.
[28] Yurinsky V.V., Exponential inequalities for sums of random vectors, J. Mult. Anal. 6 (4) (1976) 473–499.
[29] Yurinsky V.V., Sums and Gaussian Vectors, Lecture Notes in Mathematics, Vol. 1617, Springer-Verlag, Berlin, 1995.