LARGE DEVIATIONS FOR PARTIAL SUMS U-PROCESSES

PETER EICHELSBACHER, MATTHIAS LÖWE

Universität Bielefeld

Dedicated to Hans Zessin on the occasion of his 50th birthday

ABSTRACT. The large deviations principle (LDP) is known to hold for U-statistics of real valued kernel functions of i. i. d. random variables, where moreover appropriate exponential tail conditions are assumed to hold. We prove that these conditions suffice for the large deviations principle to carry over to the partial sums U-processes corresponding to the statistics.

§1 INTRODUCTION

For a sequence of $\mathbb{R}^d$-valued i. i. d. random variables X_i with a finite moment generating function Borovkov [2, 3], Mogulskii [16] and Varadhan [20] investigated the large deviation behaviour of the partial sums process

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

for different scalings. Their results have been extended in various ways, to functionals of regular finite state Markov chains by Mogulskii [15], to non identically distributed random variables by Schuette [19] and recently by Dembo and Zajic [4] to a level II version of the partial sums process, i.e. for the partial sums empirical process

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

and for the Banach space valued partial sums process $S_n(\cdot)$, where the X_i fulfill Markov dependence or mixing conditions. Their results can be summerized by

1991 *Mathematics Subject Classification.* 60 F 10.
Key words and phrases. Large deviations, partial sums, U-processes.

saying that the conditions implying an LDP for fixed t also suffice to yield the LDP for the corresponding partial sums process.

In this paper we consider the natural next step when going from linear statistics to higher order statistics, namely the partial sums U-processes, i.e.

$$U_n(t) = \frac{1}{\binom{n}{m}} \sum_{C_m^{[nt]}} h(X_{i_1}, \ldots, X_{i_m}), \quad t \in [0,1]. \tag{1.2}$$

Here the X_i are i. i. d. random variables with values in a Polish space S, 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 $\{(i_1, \ldots, i_m) \colon 1 \le i_1 < \cdots < i_m \le k\}$. We show an LDP for $U_n(\cdot)$ in $L_\infty([0,1])$, the space of almost surely bounded, measurable $\mathbb{R}^d$-valued functions on $[0,1]$, where $[0,1]$ is equipped with Lebesgue measure unless explicitly stated otherwise. The LDP for ordinary U-statistics of degree $m \in \mathbb{N}$,

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

with h and C_m^n as above, is proved for $\mathbb{R}^d$-valued h and Polish space valued X_i by the authors [10] in the i. i. d. case. An extended version is proved in [9], where the function h satisfies the condition

$$\int_{S^m} \exp(\Theta \|h(x_1, \ldots, x_m)\|)\, d\mu^m(x_1, \ldots, x_m) < \infty,$$

for each $\Theta > 0$, where μ denotes the common law of X_i. We will show, that this condition suffices for the LDP to carry over to the partial sums U-processes defined as in (1.2).

As a payoff for additional efforts due to the more complex situation we obtain an extension of an application given in [4]. More precisely, we prove an LDP for a cost-sampled empirical distribution (see Corollary 3.8).

While U-statistics were introduced by Hoeffding [11] 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 Loynes [12], Miller and Sen [14], Ronzhin [18] and Mandelbaum and Taqqu [13]. They mainly proved invariance principles and the weak convergence of appropriately normalized sample path U-processes to Brownian motion, functionals of Brownian motion, or limit processes expressible as multiple Wiener integrals. Notice, that a series of papers deal with U-processes indexed by a class of measurable functions, see for example Arcones and Giné [1]. Different to their notion of U-processes we will use the definition (1.2). Another statistics which can be thought of as a biased version of U_n is the so called V-statistics or von-Mises statistics, given by

$$V_n = \frac{1}{n^m} \sum_{D_m^n} h(X_{i_1}, \dots, X_{i_m}),$$

where $D_m^n := \{(i_1, \dots, i_m) : 1 \le i_1, \dots, i_m \le n\}$ and h is as above. Due to the symmetry of h, U_n and V_n mainly differ by the diagonal terms, where at least two of the indices $i_1, \dots, i_m$ are equal. The partial sums V-process for which we also show an LDP is given by

$$V_n(t) = \frac{1}{n^m} \sum_{D_m^{[nt]}} h(X_{i_1}, \dots, X_{i_m}), \quad t \in [0,1]. \tag{1.3}$$

A main aspect of the proof is to consider a nice metric generating the weak topology on the set of all positive signed measures on the state space of the sequence $\{X_n, n \in \mathbb{N}\}$, given by a comparison of integrals over all bounded Lipshitz functions. This metric allows an application of the contraction principle as given in Dembo and Zeitouni [5, Theorem 4.2.23] and Pérez–Abreu and Tudor [17, Theorem 2.1], by an approximation of the kernel by a bounded Lipshitz function via the theorems of Lusin and of Kirszbraun and McShane [7].

This paper is divided into four sections. In Section 2 we present the main results. In Section 3 we collect some preliminaries. Section 4 is devoted to the proofs of the theorems.

§2 Basics and statement of the results

Let $\{X_n, n \in \mathbb{N}\}$ be a sequence of i. i. d. random variables on a probability space $(\Omega, \mathcal{A}, P)$ with Polish state space S and common law μ. Denote by $\mathcal{S}$ the Borel σ-Algebra in S.

By definition a family of probability measures $\{P_\varepsilon, \varepsilon > 0\}$ on a topological space

(X,t) satisfies a large deviation principle (LDP) on X with good rate function $I\colon X \to [0,\infty]$ if for all $A \in \mathcal{B}(X)$, the Borel σ-field on X,

$$-\inf_{x\in \mathrm{int}_t A} I(x) \le \liminf_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(A) \le \limsup_{\varepsilon\to 0} \varepsilon \log P_\varepsilon(A) \le -\inf_{x\in \mathrm{cl}_t A} I(x), \quad (2.1)$$

where $\mathrm{int}_t A$ denotes the interior of A with respect to the topology t on X and $\mathrm{cl}_t A$ the closure. By definition a good rate function is t-lower semi-continuous and has t-compact level sets $\{I(x) \le L\}$, $L \in [0,\infty)$. ε^{-1} is called the speed. We say that a family of random variables satisfies the LDP when the family of measures induced by these variables satisfies the LDP.

We denote by $\mathcal{M}(S)$, $\mathcal{M}^+(S)$ and $\mathcal{M}^m(S)$, respectively, the sets of Borel measures on S which are signed, positive and positive having total mass m, respectively. These spaces are equipped with the topology of weak convergence.

Recall the definition of the relative entropy $H(\nu\,|\,\mu)$ of $\nu \in \mathcal{M}_1(S)$ with respect to $\mu \in \mathcal{M}_1(S)$:

$$H(\nu\,|\,\mu) := \begin{cases} \displaystyle\int_S f \log f \, d\mu, & \text{if } \nu \ll \mu \text{ and } f = \dfrac{d\nu}{d\mu} \\ +\infty & \text{otherwise.} \end{cases}$$

Let $h\colon S^m \to \mathbb{R}^d$ be a measurable and symmetric function and define the corresponding partial sums U-process $U_n(\cdot)$ and the partial sums V-process $V_n(\cdot)$ as in (1.2) and (1.3). Denote by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^d$. We first prove the LDP for $\{U_n(\cdot), n \in \mathbb{N}\}$ assuming

$$\int_{S^m} \exp(\theta \|h(x)\|)\, \mu^m(dx) < \infty \quad \text{for all } \theta > 0, \quad (2.2)$$

where μ^m denotes the m-fold product measure. Let $\mathbb{P}_r$ denote the set of all decompositions of the set $\{1,\dots,m\}$ into r subsets $P^1,\dots,P^r$ ($r \in \{1,\dots,m\}$). Set $\mathcal{H}_r$ to be the set of all versions of h depending exactly on r variables only, i.e. $h \in \mathcal{H}_r$ iff

$$h_r(y_1,\dots,y_r) = h(x_1,\dots,x_m)$$

such that $x_j = y_i$ if and only if $j \in P^i$ for a partition in $\mathbb{P}_r$ and $1 \le j \le m$, $1 \le i \le r$. We interpret a function in $\mathcal{H}_r$ as a function on S^r rather than a function on S^m. Proving the LDP for $\{V_n(\cdot), n \in \mathbb{N}\}$ we assume

$$\int_{S^r} \exp(\theta \|g(x_1,\dots,x_r)\|)\, \mu^r(dx_1,\dots,dx_r) < \infty \quad \text{for all } \theta > 0 \quad (2.3)$$

and for each $r \in \{1, \ldots, m\}$ and every $g \in \mathcal{H}_r$.

To formulate our main results we need some more definitions. The topological dual of $\mathcal{M}(S)$ is identified with the set $C_b(S)$ of continuous and bounded mappings from S to $\mathbb{R}$ via $\langle f, \mu \rangle = \int_S f d\mu$. $\mathcal{M}^+(S)$ becomes a Polish spaces when equipped with the metric topology induced by

$$\beta(\mu, \nu) := \sup\{|\int_S f d\mu - \int_S f d\nu| : f \in C_b(S), ||f||_\infty + ||f||_L \leq 1\},$$

where $|| \cdot ||_\infty$ denotes the supremum norm and

$$||f||_L := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}$$

and d denotes the metric on S (see [4, Lemma A.1.]).
Let us denote by $D([0,1], (\mathcal{M}^+(S), \beta))$ the space of all maps on $[0,1]$ with values in $(\mathcal{M}^+(S), \beta)$ continuous from the right and having left limits equipped with the uniform metric topology induced by

$$d_\infty(y(\cdot), z(\cdot)) := \sup_{t \in [0,1]} \beta(y(t), z(t)).$$

Then by Theorem 1 in [4] the partial sums processes $\{L_n(\cdot)\}_{n \in \mathbb{N}}$ satisfy the LDP in $D([0,1], (\mathcal{M}^+(S), \beta))$ with the convex rate function

$$I(\nu(\cdot)) = \int_0^1 H(\dot{\nu}|\mu) dt \quad \text{if } \nu \in \mathcal{AC}_0,$$

and $I(\nu(\cdot)) = \infty$ otherwise. Here $\mathcal{AC}_0$ is defined to be the set of all maps $\nu \colon [0,1] \to \mathcal{M}^+(S)$ that are absolutely continuous with respect to the variation norm $|| \cdot ||_{\text{var}}$, satisfy $\nu(t) - \nu(s) \in \mathcal{M}^{t-s}(S)$ for all $t \geq s$, while $\nu(0) = 0$, 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 $f \in C_b(S)$).

For $m \geq 2$ we denote by $D([0,1], (\mathcal{M}^+(S^m), \beta^m))$ the space of all $(\mathcal{M}^+(S^m), \beta^m)$-valued cadlag-functions. S^m as well as $\mathcal{M}^+(S^m)$ are Polish spaces, since S was assumed to be Polish. We take β^m to be the metric on $\mathcal{M}^+(S^m)$ defined analogously to β. Again we equip the space $D([0,1], (\mathcal{M}^+(S^m), \beta^m))$ with the uniform metric topology induced by $d^m_\infty(y(\cdot), z(\cdot)) := \sup_{t \in [0,1]} \beta^m(y(t), z(t))$. Furthermore we denote by $\mathcal{AC}_0([0,1], (\mathbb{R}^d, ||\cdot||))$ the space of all absolutely continuous $\mathbb{R}^d$-valued functions ϕ on $[0,1]$ with $\phi(0) = 0$.

Now our main result can be stated as follows:

Theorem 2.4. *Assuming* (2.2) *holds, the sequence of partial sums U-processes* $\{U_n(\cdot), n \in \mathbb{N}\}$ *satisfies the LDP in* $L_\infty([0,1])$ *with the good rate function*

$$I_\infty(\phi) = \inf\{ \int_0^1 H(\dot\nu|\mu)dt,\, \nu \in \mathcal{A}C_0 \cap K_\infty \text{ and } \int_{S^m} h\, d\nu^m(\cdot) = \phi(\cdot) \} \qquad (2.5)$$

and speed n, if $\phi \in \mathcal{A}C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$ *and* $I_\infty(\phi) = \infty$ *otherwise. Here*

$$K_\infty := \bigcup_{L \geq 0} \{\nu(\cdot) : I(\nu(\cdot)) \leq L\}.$$

Assuming (2.3) *holds, the sequence of partial sums V-processes* $\{V_n(\cdot), n \in \mathbb{N}\}$ *satisfies the LDP in* $L_\infty([0,1])$ *with the same rate function* $I_\infty(\cdot)$ *and speed n.*

Of course we wish to extend Theorem 2.4 to the laws of

$$V_\varepsilon(t) = \varepsilon^m \sum_{D_m^{[t/\varepsilon]}} h(X_{i_1}, \ldots, X_{i_m}), \quad t \in [0,1],$$

and

$$U_\varepsilon(t) = \frac{1}{\binom{1/\varepsilon}{m}} \sum_{C_m^{[t/\varepsilon]}} h(X_{i_1}, \ldots, X_{i_m}), \quad t \in [0,1],$$

for $\varepsilon > 0$. The precise statement is given as follows:

Theorem 2.6. *Assuming* (2.2) *holds, the family* $\{U_\varepsilon(\cdot), \varepsilon > 0\}$ *satisfies the LDP in* $L_\infty([0,1])$ *with the good rate function* $I_\infty(\cdot)$ *given by* (2.5) *and speed* ε^{-1}. *Assuming* (2.3) *holds, the family* $\{V_\varepsilon(\cdot), \varepsilon > 0\}$ *satisfies the LDP in* $L_\infty([0,1])$ *with the same good rate* $I_\infty(\cdot)$ *and speed* ε^{-1}.

As a Corollary we finally obtain the result of Varadhan [20] and Mogulskii [16] by choosing $m = 1$:

Corollary 2.7. *If* $\Lambda(\lambda) := \log \int_S \exp(\langle \lambda, x\rangle)\, \mu(dx) < \infty$ *for all* $\lambda \in \mathbb{R}^d$, *the sequence* $\{\mu_n, n \in \mathbb{N}\}$ *of laws of* $Z_n(t) = 1/n \sum_{i=1}^{[nt]} X_i\,, 0 \leq t \leq 1$, *satisfies the LDP in* $L_\infty([0,1])$ *with the good rate function*

$$J(\phi) = \int_0^1 \Lambda^*(\dot\phi(t))\, dt \quad \text{if} \quad \phi \in \mathcal{A}C([0,1], (\mathbb{R}^d, ||\cdot||)),\, \phi(0) = 0,$$

$J(\phi) = \infty$ *otherwise, where* $\dot\phi$ *denotes the derivative of* ϕ *with respect to* t *and* $\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}^d}\{\langle \lambda, x\rangle - \Lambda(\lambda)\}$ *denotes the Fenchel-Legendre transform of* $\Lambda(\cdot)$. *Under the same condition, the family* $\{\nu_\varepsilon, \varepsilon > 0\}$ *of laws of* $Y_\varepsilon(t) := \varepsilon \sum_{i=1}^{[t/\varepsilon]} X_i$, *satisfies the LDP with the good rate function* $J(\cdot)$.

§3 Preliminaries

We will use the following observation, due to Hoeffding [11, Section 5].

Lemma 3.1. *For $s > 0$*

$$\mathbb{E}_P\big(\exp(s\|U_n\|)\big) \leq \mathbb{E}_P\left(\exp\left(\frac{s}{k}\|h\|\right)\right)^k$$

where $k := [\frac{n}{m}]$.

The process

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

is the polygonal approximation of $U_n(t)$. Recall that the maps $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ and $\{U_n(\cdot), n \in \mathbb{N}\}$ from $(\Omega, \mathcal{S}^{\mathbb{N}}, P)$ into $L_\infty\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ are called exponentially equivalent if for each $\delta > 0$

$$\limsup_{n\to\infty}\frac{1}{n}\log P\big(\sup_{t\in[0,1]}\|\widetilde{U}_n(t) - U_n(t)\| > \delta\big) = -\infty,$$

(see [5, Section 4.2.2]). Remark that $\{\omega : \sup_{t\in[0,1]}\|\widetilde{U}_n(t) - U_n(t)\| > \delta\} \in \mathcal{A}$ for each $\delta > 0$. In the following we always assume that condition (2.2) in the partial sums U-process case and condition (2.3) in the partial sums V-process case holds.

Lemma 3.2. *$\{U_n(\cdot), n \in \mathbb{N}\}$, $\{V_n(\cdot), n \in \mathbb{N}\}$ and $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ are exponentially equivalent in $L_\infty\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$.*

Proof. Note that

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

By the exponential Chebychev-Markov inequality

$$\begin{aligned}
&P\big(\sup_{t\in[0,1]}\|U_n(t) - \widetilde{U}_n(t)\| > \delta\big)\\
&\qquad\leq \sum_{j=1}^{n+1} P\left(\frac{1}{\binom{n}{m-1}}\sum_{C_{m-1}^{j-1}}\|h(X_{i_1},\ldots,X_{i_{m-1}},X_j)\| > \frac{n-m+1}{m}\delta\right)\\
&\qquad\leq (n+1)^m P\left(\|h(X_1,\ldots,X_m)\| > \frac{n-m+1}{m}\delta\right)\\
&\qquad\leq (n+1)^m \exp\left(-\Theta\big(\frac{n-m+1}{m}\big)\delta\right)\mathbb{E}_P\big[\exp(\Theta\|h\|)\big]
\end{aligned}$$

for every $\Theta > 0$. Hence letting $n \to \infty$ and then $\Theta \to \infty$ we get the result on U_n and $\widetilde{U}_n$.

Note that

$$\begin{aligned}
&\|V_n(t) - U_n(t)\| \\
&\quad \le \left| 1 - \frac{n(n-1)\cdots(n-m+1)}{n^m} \right| \|U_n(t)\| + \frac{1}{n^m} \sum_r \|h(X_{i_1}, \ldots, X_{i_m})\|,
\end{aligned}$$

where the sum $\sum_r$ consists of all realizations of h with at least two of $i_1, \ldots, i_m$ being identical and all indices being $\le [nt]$.

Since $n^m - n(n-1)\cdots(n-m+1) \le C n^{m-1}$ with a constant C depending on m only we can write

$$\begin{aligned}
&P\big(\|V_n(t) - U_n(t)\| > \delta\big) \\
&\quad \le P\Big(\|U_n(t)\| > \frac{n\delta}{2C}\Big) + P\Big(\frac{1}{Cn^{m-1}} \Big\| \sum_r h \Big\| > \frac{n\delta}{2C}\Big).
\end{aligned} \tag{3.3}$$

The sum in $U_n(t)$ contains at most $\binom{n}{m}$ summands – for $t = 1$ – and thus we can estimate

$$\begin{aligned}
P\Big(\sup_{t\in[0,1]} \|U_n(t)\| > \frac{n\delta}{2C}\Big) &\le n^m P\Big(\|h\| > \frac{n\delta}{2C}\Big) \\
&\le n^m \exp\Big(-\Theta n \frac{\delta}{2C}\Big) \mathbb{E}_P\big[\exp(\Theta \|h\|)\big].
\end{aligned}$$

The second term on the right in (3.3) also can be regarded as an average with at most $n^m - n(n-1)\cdots(n-m+1)$ summands. Thus the same argument yields

$$\limsup_{n\to\infty} \frac{1}{n} \log P\Big(\frac{1}{Cn^{m-1}} \Big\| \sum_r h \Big\| \ge \frac{n\delta}{2C}\Big) = -\infty$$

for any $\delta > 0$ by assumption (2.3). So by Lemma 1.2.15 in [5] the proof is done. $\square$

Remark 3.4. Since under (2.3) the process $\{U_n(\cdot), n \in \mathbb{N}\}$, $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ and $\{V_n(\cdot), n \in \mathbb{N}\}$ have identical large deviation behaviour, we can concentrate on $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ in the rest of the paper.

We will use the following easy observation (see [9, Corollary 3.3]).

Lemma 3.5. *Let S be Polish and d_v the metric induced by the total variation norm $\|\cdot\|_v$ on $\mathcal{M}_1(S)$. Then a LDP for a sequence of probability measures on $\mathcal{M}_1(S)$ with respect to the weak topology is transfered to every d_v-exponentially equivalent sequence of probability measures on $\mathcal{M}_1(S)$.*

Denote by

$$L_n^{\otimes m}(\cdot) := \frac{1}{n^m} \sum_{D_m^{[n\cdot]}} \delta_{(X_{i_1},\dots,X_{i_m})},$$

and by

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

where $n_{(m)} := \prod_{k=0}^{m-1}(n-k)$ and where $I_{m,n} \subset \{1,\dots,n\}^m$ contains all m-tuples with pairwise different components. Finally let $\widetilde{L}_n^m(\cdot)$ be the polygonal approximation of $L_n^m(\cdot)$, defined by

$$\widetilde{L}_n^m(t) := L_n^m(t) + (nt-[nt])\frac{1}{\binom{n}{m}} \sum_{C_{m-1}^{[nt]}} \delta_{(X_{i_1},\dots,X_{i_{m-1}},X_{[nt]+1})}.$$

Lemma 3.6. *The sequences $\{L_n^{\otimes m}(\cdot)\}_{n\in\mathbb{N}}$ and $\{\widetilde{L}_n^m(\cdot)\}_{n\in\mathbb{N}}$ satisfy the LDP in $D([0,1],(\mathcal{M}^+(S^m),\beta^m))$ with the good rate function $I^m(\nu(\cdot)) = I(\nu_1(\cdot))$ if $\nu_1^m(\cdot) = \nu(\cdot)$ and $\nu_1(\cdot) \in \mathcal{AC}_0$ and $I^m(\nu(\cdot)) = \infty$ otherwise.*

Proof. In view of [4, Theorem 1], all we have to check is that $\nu(\cdot) \mapsto \nu^m(\cdot)$ maps $D([0,1],(\mathcal{M}^+(S),\beta))$ into $D([0,1],(\mathcal{M}^+(S^m),\beta^m))$ and is continuous with respect to the topologies of uniform convergence.
Now for $\nu(\cdot)$ and $\mu(\cdot) \in D([0,1],(\mathcal{M}^+(S),\beta))$ and $t\in[0,1]$ we get

$$\int_{S^m} h(x_1,\dots,x_m)d(\nu^m(t)-\mu^m(t)) =$$
$$\int_{S^m} h(x_1,\dots,x_m)d(\sum_{i=1}^m \nu(t)^{i-1}\,(\nu(t)-\mu(t))\,\mu(t)^{m-i})$$

for every integrable h. If $h \in BL_1(S^m) := \{f : S^m \to \mathbb{R}^d : ||f||_\infty + ||f||_L \le 1\}$ we get for each i

$$\|\int_{S^m} h(x_1,\dots,x_m)d(\nu(t)^{i-1}\,(\nu(t)-\mu(t))\,\mu(t)^{m-i})\| = \|\int_S \widetilde{h}(x_i)d((\nu(t)-\mu(t)))\|$$

with $\widetilde{h}(x_i) := \int_{S^{m-1}} h(x_1, \ldots, x_{i-1}, x_i, x_{i+1} \ldots, x_m) d(\nu(t)^{i-1} \mu(t)^{m-i})$.
Since $\widetilde{h}(x_i) \in BL_1(S)$ we get for each $t \in [0,1]$

$$\| \int_{S^m} h(x_1, \ldots, x_m) d(\nu^m(t) - \mu^m(t)) \| < m\varepsilon$$

if $\beta(\nu(t), \mu(t)) < \varepsilon$ and therefore the lemma is proved for $L_n^{\otimes m}(\cdot)$. Remark that $\beta(\nu(t), \nu(s)) \to 0$ (for $t \to s$) implies $\beta^m(\nu^m(t), \nu^m(s)) \to 0$.
Now for each $t \in [0,1]$

$$\begin{aligned} ||L_n^{\otimes m}(t) - \widetilde{L}_n^m(t)||_{\text{var}} &\leq ||L_n^{\otimes m}(t) - L_n^m(t)||_{\text{var}} + ||L_n^m(t) - \widetilde{L}_n^m(t)||_{\text{var}} \\ &\leq \frac{C(m)}{n}. \end{aligned}$$

Thus

$$\sup_{t \in [0,1]} \beta^m(L_n^{\otimes m}(t), \widetilde{L}_n^m(t)) \leq \sup_{t \in [0,1]} ||L_n^{\otimes m}(t) - \widetilde{L}_n^m(t)||_{\text{var}} \leq \frac{C(m)}{n}$$

and we obtain the result for $\widetilde{L}_n^m(\cdot)$ by Lemma 3.5 and the observation that

$$\limsup_{n \to \infty} \frac{1}{n} \log P(\sup_{t \in [0,1]} ||L_n^{\otimes m}(t) - \widetilde{L}_n^m(t)||_{\text{var}} \geq \varepsilon) = -\infty$$

for each $\varepsilon > 0$. $\square$

Lemma 3.7. *Let $K_\infty^m := \bigcup_{L \geq 0} \{\nu(\cdot) : I^m(\nu(\cdot)) \leq L\}$. If $\nu(\cdot) \in K_\infty^m \cap \mathcal{A}C_0$ then $\int h d\nu(\cdot) \in \mathcal{A}C_0([0,1], (\mathbb{R}^d, || \cdot ||))$ for all $h : S^m \to \mathbb{R}^d$, satisfying condition (2.2). Thus $I_\infty(\phi) < \infty$ implies $\phi \in \mathcal{A}C_0([0,1], (\mathbb{R}^d, || \cdot ||))$. Let $\mathcal{C}_0([0,1], (\mathcal{M}^+(S^m), \beta^m))$ denote the space of all continuous maps equipped with the uniform metric topology induced by d_∞^m. If $\nu(\cdot) \in K_\infty^m \cap \mathcal{C}_0([0,1], (\mathcal{M}^+(S^m), \beta^m))$, then $\int h d\nu(\cdot) \in \mathcal{C}_0([0,1], (\mathbb{R}^d, || \cdot ||))$ for all $h : S^m \to \mathbb{R}^d$, satisfying condition (2.2).*

The proof of Lemma 3.7 is given after the proof of Theorem 2.4.

Note that $\{\nu(\cdot) : I^m(\nu(\cdot)) < \infty\} \subset C_0([0,1], (\mathcal{M}^+(S^m), \beta^m))$, the space of all continuous $(\mathcal{M}^+(S^m), \beta^m)$-valued functions, starting in the origin. Since $P(\widetilde{L}_n^m(\cdot) \in C_0([0,1], (\mathcal{M}^+(S^m), \beta^m))) = 1$, the LDP for $\{\widetilde{L}_n^m(\cdot)\}_{n \in \mathbb{N}}$ holds in the space $C_0([0,1], (\mathcal{M}^+(S^m), \beta^m))$ by [5, Lemma 4.1.5].

A simple consequence of Lemma 3.6 is an improvement of a result of Dembo and Zajic [4]. Given a cost function $f : S^m \to \mathbb{R}$ which is assumed to be bounded and Lipschitz-continuous. Define

$$k^*(L_n^{\otimes m}) := \min\{\, k : \sum_{i_1,\dots,i_m=1}^{k} f(X_{i_1},\dots,X_{i_m}) \geq n^m \,\}$$

and $t^*(\nu) := \inf\{\, t : \langle f, \nu(t)\rangle > 1 \,\}$ for $\nu \in D([0,1],(\mathcal{M}^+(S^m),\beta^m))$. Moreover let

$$\Lambda^m(f) := \lim_{n\to\infty} \frac{1}{n} \log E[\exp(nU_n(f))]$$

(which exits due to our choice of f). We get

Corollary 3.8. *If* $\inf_{\gamma<0} \Lambda^m(\gamma f) < 0$ *then* $\{\frac{1}{t^*(L_n^m)} L_n^m(t^*(L_n^m))\}_{n\in\mathbb{N}}$ *satisfies a LDP on* $(\mathcal{M}^+(S^m),\beta^m)$ *with the good rate function*

$$I_1(\nu) := \frac{1}{\int f d\nu} \Lambda^*(\nu_1),$$

if $\int f d\nu > 0$ *and* $\nu_1^m = \nu$ *and* $I_1(\nu) = \infty$ *otherwise. Here* Λ^* *is defined as in Corollary 2.7. In particular* $\{\frac{1}{k^*}\sum_{i_1,\dots,i_m=1}^{k^*} \delta_{(X_{i_1},\dots,X_{i_m})}\}_{n\in\mathbb{N}}$ *satisfies a LDP with the same rate function* I_1.

Proof. Adopt step by step Proposition 3 in [4], using Lemma 3.6.

§4 Proofs of the Theorems

Proof of Theorem 2.4. Observe that for the function F defined by $F(\varrho(\cdot)) = \int h d\varrho(\cdot)$ we get $F(\widetilde{L}_n^m(\cdot)) = \widetilde{U}_n(\cdot)$. Therefore we would like to apply the contraction principle (cf. [5, Theorem 4.2.23] and [17, Theorem 2.1]). To do so we have to check that F is "sufficiently continuous", and therefore first of all we have to describe the images under F. For arbitrary kernel functions h the integral $\int h d\varrho(t)$ is generally not defined for all $\varrho(t) \in \mathcal{M}^+(S^m)$. On the other hand if $\varrho(\cdot) \in K_\infty^m$ we get $I^m(\varrho(\cdot)) = I(\nu(\cdot)) \leq K$ for some $K \geq 0$ and $\nu^m(\cdot) = \varrho(\cdot)$. Since by Lemma 4(b) in [4]

$$I(\nu(\cdot)) = \sup_{0=t_0<t_1<\dots<t_k\leq 1, k\in\mathbb{N}} \sum_{i=1}^{k} (t_i - t_{i-1}) H(\frac{\nu(t_i)-\nu(t_{i-1})}{t_i - t_{i-1}}|\mu), \tag{4.1}$$

the convexity of $H(\cdot|\mu)$ yields

$$t_i H(\frac{\nu(t_i)}{t_i}|\mu) = \frac{t_i}{m} H(\frac{\nu^m(t_i)}{t_i^m}|\mu^m) \leq \sum_{j=1}^{i} (t_j - t_{j-1})\, H(\frac{\nu(t_j) - \nu(t_{j-1})}{t_j - t_{j-1}}|\mu) \leq K$$

for each $t_i \in (0,1]$, $i \in \mathbb{N}$. Using the well-known inequality $xy \leq e^x + y \log y$ for $x, y \geq 0$ we get

$$t \int ||h|| \frac{d\rho(t)}{t} \leq t \int \exp(||h||) d\mu^m + t\, H(\frac{\nu^m(t)}{t}|\mu^m)$$

and hence $\int ||h|| d\rho(t) < \infty$ for all $t \in [0,1]$ by the moment condition (2.2), whenever $I^m(\varrho(\cdot)) \leq K$.
Therefore the contraction map

$$F : \mathcal{C}_0([0,1], (\mathcal{M}^+(S^m), \beta^m)) \cap K_\infty^m \to \mathcal{C}_0([0,1], (\mathbb{R}^d, ||\cdot||))$$

is well defined on $K_\infty^m \cap \mathcal{C}_0([0,1], (\mathcal{M}^+(S^m), \beta^m))$ by an application of Lemma 3.7. Notice that by Theorem 1 in [4], the LDP for $\{L_n(\cdot)\}_{n\in\mathbb{N}}$ is established in the space $(\mathcal{C}_0([0,1], (\mathcal{M}^+(S), \beta)), d_\infty)$. By Lemma 3.6 and the remark following Lemma 3.7, we get the LDP for $\{\widetilde{L}_n^m(\cdot)\}_{n\in\mathbb{N}}$ in $C_0([0,1], (\mathcal{M}^+(S^m), \beta^m)) \cap K_\infty^m$.

If h is a continuous and bounded kernel function with $||h||_{BL} := ||h||_\infty + ||h||_L \leq C$ for any finite constant C we get the LDP for $\{U_n(\cdot)\}_{n\in\mathbb{N}}$ and $\{V_n(\cdot)\}_{n\in\mathbb{N}}$ as a consequence of Lemma 3.2 and Lemma 3.6 using the contraction principle (cf. Theorem 4.2.1 in [5]) and Lemma 4.1.5 in [5].

The main step in the proof will be an application of an extended version of the contraction principle (cf. [5, Theorem 4.2.23] and [17, Theorem 2.1]). So let h be an arbitrary, measurable, and symmetric function satisfying condition (2.2). For $0 < L < \infty$ let

$$h_L(x_1, \ldots, x_m) := 1_{\{||h|| \leq L\}} h(x_1, \ldots, x_m)$$

be h truncated at height L. According to Lusin's theorem there exists a compact set $C \subseteq S^m$ with measure $\mu^m(C) \geq 1 - e^{-L^2}$ such that $h_L|_C$ is continuous. Due to the Stone-Weierstrass theorem we then can find a function $g_L \in BL(C)$ such that

$$||g_L - h_L|_C|| \leq \frac{1}{L}$$

and $\sup_{x\in C} g_L(x) \leq \sup_{x\in C} h_L|_C(x) \leq L$. By the extension theorem of Kirszbraun and McShane ([7, Theorem 7.3]) g_L admits an extension $f_L \in BL(S^m)$ with $||f_L||_{BL} = ||g_L||_{BL}$. Now we define $F_L : \mathcal{C}_0([0,1],(\mathcal{M}^+(S^m),\beta^m)) \cap K^m_\infty \to \mathcal{C}_0([0,1],(\mathbb{R}^d,||\cdot||))$ by

$$F_L(\varrho(\cdot)) = \int f_L d\varrho(\cdot)$$

for $L > 0$. Note that F_L is well defined by the same arguments as above. Moreover F_L is continuous with respect to the topologies of uniform convergence by definition of β^m.

We have to check that

$$\limsup_{L\to\infty}\{\sup_{t\in[0,1]} ||F(\varrho(t)) - F_L(\varrho(t))|| : I^m(\varrho(\cdot)) \leq K\} = 0 \tag{4.2}$$

for all $K > 0$. For each $t \in (0,1]$ we get $t\,H(\frac{\varrho(t)}{t}|\mu^m) \leq K$ whenever $I^m(\varrho(\cdot)) \leq K$. Therefore with $\varrho(0) = 0$ we obtain $\{\varrho(\cdot) : I^m(\varrho(\cdot)) \leq K\} \subseteq \{\varrho(\cdot) : tH(\frac{\varrho(t)}{t}|\mu^m) \leq K, \forall t \in [0,1]\} =: N_K$. The triangle inequality yields $||h - f_L|| \leq ||h - h_L|| + ||h_L - f_L||$. A simple adaption of the proof of Lemma 5.1 in [8] shows that

$$\limsup_{L\to\infty} \sup_{\varrho(\cdot)\in N_K} \int ||h - h_L||\, d\varrho(t) = 0$$

uniformly in t and for all $K > 0$. So (4.2) follows if we can prove

$$\limsup_{L\to\infty} \sup_{\varrho(\cdot)\in N_K} \int ||f_L - h_L||\, d\varrho(t) = 0 \tag{4.3}$$

uniformly in t for all $K > 0$. To this end put $||h_L - f_L|| =: f_L^{(1)} + f_L^{(2)}$ with $f_L^{(1)} \leq 1/L$ and $\mu^m(f_L^{(2)} \neq 0) \leq e^{-L^2}$ and $f_L^{(1)}, f_L^{(2)} \geq 0$.

Clearly

$$\limsup_{L\to\infty} \sup_{\varrho(\cdot)\in N_K} \int f_L^{(1)} d\varrho(t) = 0$$

uniformly in t and on each level set of $H(\cdot\,|\,\mu)$. With $x\,y \leq e^{\sigma\, x} + \frac{y}{\sigma}\log\frac{y}{\sigma}$ for all $x, y \geq 0$ and $\sigma > 0$ we obtain for $\sigma \geq 1$ as in the proof of Lemma 5.1 in [8]

$$\begin{aligned} t\int f_L^{(2)} d\frac{\varrho(t)}{t} &= t\int_{\{f_L^{(2)}\neq 0\}} f_L^{(2)} \frac{d\big(\frac{\varrho(t)}{t}\big)}{d\mu^m} d\mu^m \\ &\leq \int_{\{f_L^{(2)}\neq 0\}} \exp\big(\sigma\, f_L^{(2)}\big) d\mu^m + \frac{1}{\sigma}\int_{S^m} t\left|\log\frac{d\big(\frac{\varrho(t)}{t}\big)}{d\mu^m}\right| d\Big(\frac{\varrho(t)}{t}\Big) \qquad (4.4) \\ &\leq \exp(L(\sigma - L)) + \frac{1}{\sigma}\int_{S^m} t\left|\log\frac{d\big(\frac{\varrho(t)}{t}\big)}{d\mu^m}\right| d\Big(\frac{\varrho(t)}{t}\Big). \end{aligned}$$

Now the first term in (4.4) converges to zero for every fixed σ by letting $L \to \infty$ (independent of t and N_K, of course). Since the second term on the right in (4.4) converges uniformly in t on N_K for each K we prove (4.3) and therefore (4.2) by first letting $L \to \infty$ and then $\sigma \to \infty$.

The last step in applying the contraction principle is to verify the exponential equivalence between $F(\widetilde{L}_n^m(\cdot)) = \widetilde{U}_n(\cdot)$ and $F_L(\widetilde{L}_n^m(\cdot)) = \widetilde{U}_n(f_L)(\cdot)$, where $\widetilde{U}_n(f_L)(\cdot)$ denotes the polygonal approximation of the partial sums U-process with kernel function f_L. We have to check

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(\sup_{t\in[0,1]} ||\widetilde{U}_n(t) - \widetilde{U}_n(f_L)(t)|| \geq \varepsilon \big) = -\infty \tag{4.5}$$

for every $\varepsilon > 0$. By definition the values of $U_n(\cdot)$ and $U_n(f_L)(\cdot)$ are constant on each interval $(i/n, (i+1)/n]$ for $i \in \{0, \dots, n-1\}$ and $U_n(t) = U_n(i/n)$ for all $t \in (i/n, (i+1)/n]$. On the other side in each intervall $(i/n, (i+1)/n]$ the oscillation $||\widetilde{U}_n(t) - \widetilde{U}_n(f_L)(t)||$ is by definition dominated by $\sup_{l\in\{i,i+1\}} ||U_n(t_l) - U_n(f_L)(t_l)||$, where $t_l = l/n$. Thus

$$P\big(\sup_{t\in[0,1]} ||\widetilde{U}_n(t) - \widetilde{U}_n(f_L)(t)|| \geq \varepsilon \big) \leq \sum_{i=0}^{n-1} P\big(||U_n(t_i) - U_n(f_L)(t_i)|| \geq \varepsilon \big).$$

Hence by the Chebyshev-Markov inequality we obtain for any positive Θ

$$\begin{aligned} P\big(\|U_n(t_i) - U_n(h_L)(t_i)\| > \frac{\varepsilon}{2} \big) &\leq \exp\left(-\frac{\Theta\varepsilon}{2} \right) \mathbb{E}_P \left(\exp(\Theta \|U_n(h - h_L)(t_i)\|) \right) \\ &\leq \exp\left(-\frac{\Theta\varepsilon}{2} \right) \mathbb{E}_P \left(\exp\Big(\frac{\Theta}{\binom{i}{m}} \sum_{C_m^i} \|h - h_L\| \Big) \right). \end{aligned}$$

Applying Lemma 3.1 to the right hand side we get, substituting Θ by $\Theta \frac{i}{m}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P\big(\|U_n(h - h_L)(t_i)\| > \frac{\varepsilon}{2} \big) \leq -\frac{\Theta}{m} \frac{\varepsilon}{2} + \frac{1}{m} \log \mathbb{E}_P \big(\exp(\Theta \|h - h_L\|) \big).$$

By the arbitrariness of Θ we conclude

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(\|U_n(h - h_L)(t_i)\| > \frac{\varepsilon}{2} \big) = -\infty.$$

Recall $\|h_L - f_L\| = f_L^{(1)} + f_L^{(2)}$. If we take L large such that $\frac{1}{L} < \frac{\varepsilon}{4}$ we get $P\big(\|U_n(f_L^{(1)})(t_i)\| > \frac{\varepsilon}{4} \big) = 0$ and thus it remains to show that

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(\|U_n(f_L^{(2)})(t_i)\| > \frac{\varepsilon}{4} \big) = -\infty$$

for every positive ε. Again applying Chebyshev-Markov inequality and Lemma 3.1 we get

$$\limsup_{n\to\infty}\frac{1}{n}\log P\big(\|U_n(f_L^{(2)})(t_i)\| > \frac{\varepsilon}{4}\big) \le \frac{1}{m}\bigg(-\frac{\Theta\varepsilon}{4} + \log\big(1+\exp(L(\Theta - L))\big)\bigg)$$

for every $\Theta, \varepsilon > 0$. So letting $L \to \infty$ and then $\Theta \to \infty$ proves (4.5). Thus we get the LDP for $\{U_n(\cdot)\}_{n\in\mathbb{N}}$ in $L_\infty([0,1])$ by Lemma 4.1.5 in [5]. The exponential equivalence to $\{U_n(\cdot)\}_{n\in\mathbb{N}}$ and $\{V_n(\cdot)\}_{n\in\mathbb{N}}$ in $L_\infty([0,1],(\mathbb{R}^d,||\cdot||))$ established in Lemma 3.2 implies the result for $\{V_n(\cdot)\}_{n\in\mathbb{N}}$ (cf. [5, Theorem 4.2.13]). □

Proof of Lemma 3.7. Since we are only interested in the case $\nu(\cdot) \in K_\infty^m$, by the same arguments as in the proof of Theorem 2.4 the integrals $\int h d\nu(t)$ are defined for all $t \in [0,1]$. Now fix $\nu(\cdot) \in \mathcal{AC}_0$: By definition this means that for each $\varepsilon > 0$ there is a $\delta > 0$ such that for any finite sequences $0 \le s_1 < t_1 \le s_2 < t_2 \le \cdots \le s_k < t_k \le 1$

$$\sum_{l=1}^k (t_l - s_l) \le \delta \quad \text{implies} \quad \sum_{l=1}^k ||\nu(t_l) - \nu(s_l)||_{\text{var}} \le \varepsilon.$$

To prove the lemma we approximate kernel function h by $h_L := 1_{\{||h||\le L\}}h$. Analogously to (4.2) for a given $\varepsilon > 0$ we can choose $L > 0$ such that $||\int h_L d\nu(t) - \int h d\nu(t)|| \le \varepsilon$ uniformly in $t \in [0,1]$. Moreover choose the sequence $0 \le s_1 < t_1 \le s_2 < t_2 \le \cdots \le s_k < t_k \le 1$ such that $\sum_{l=1}^k ||\nu(t_l) - \nu(s_l)||_{\text{var}} \le \frac{\varepsilon}{L}$.
Thus by $||\varrho_1 - \varrho_2||_{var} = \sup\{|\int f d\varrho_1 - \int f d\varrho_2| : ||f|| \le 1\}$ and the triangle inequality we obtain

$$\begin{aligned}\sum_{l=1}^k \Big|\int h d\nu(t_l) - \int h d\nu(s_l)\Big| &\le \sum_{l=1}^k \Big\{\Big|\int h d\nu(t_l) - \int h_L d\nu(t_l)\Big| \\ &+ \Big|\int h_L d\nu(t_l) - \int h_L d\nu(s_l)\Big| + \Big|\int h_L d\nu(s_l) - \int h d\nu(s_l)\Big|\Big\} < 3\varepsilon,\end{aligned}$$

and therefore $\int h d\nu(\cdot) \in \mathcal{AC}_0([0,1],(\mathbb{R}^d,||\cdot||))$. The second statement of the lemma follows by the same steps with just replacing h_L by f_L as defined in the proof of Theorem 2.4. □

Proof of Theorem 2.6. $\{V_\varepsilon(\cdot), \varepsilon > 0\}$ and $\{U_\varepsilon(\cdot), \varepsilon > 0\}$ are exponentially equivalent in $L_\infty\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$. The proof may be carried out in the same way as the proof of 3.2. For sequences $\varepsilon_l \to 0$ such that $\varepsilon_l^{-1} \in \mathbb{N}$ for every $l \in \mathbb{N}$, the theorem follows from Theorem 2.4. For an arbitrary $\varepsilon_l \to 0$ let $n_l := [\varepsilon_l^{-1}]$. By Theorem 2.4 $\{U_{n_l}(\cdot), l \in \mathbb{N}\}$ satisfies the LDP with rate $I_\infty(\cdot)$ and speed n_l^{-1}. To obtain the assertion of the theorem it remains to show that for any $\delta > 0$

$$\limsup_{l\to\infty} \frac{1}{n_l} \log P(\sup_{t\in[0,1]} \|U_{\varepsilon_l}(t) - U_{n_l}(t)\| \geq \delta) = -\infty,$$

since $n_l\, \varepsilon_l \to 1$ and again applying Theorem 4.2.13 in [5]. By definition of $U_\varepsilon(\cdot)$ we get

$$\begin{aligned}
&\|U_{\varepsilon_l}(t) - U_{n_l}(t)\| \\
&\quad \leq \left| 1 - \binom{n_l}{m}\binom{\frac{1}{\varepsilon_l}}{m}^{-1} \right| \|U_{n_l}(t)\| + \binom{\frac{1}{\varepsilon_l}}{m}^{-1} \sum_{C_{m-1}^{[n_l t]}} \|h(X_{i_1}, \dots, X_{i_{m-1}}, X_{[\frac{t}{\varepsilon_l}]})\|,
\end{aligned}$$

since $[\frac{t}{\varepsilon_l}] \in \{[n_l t], [n_l t]+1\}$ and $\binom{n_l}{m}\binom{1/\varepsilon_l}{m}^{-1} \in [1-(m+2)^2\varepsilon_l, 1]$ for l sufficiently large by the Bernoulli inequality, we get

$$\|U_{\varepsilon_l}(t) - U_{n_l}(t)\| \leq \mathrm{const}(m)\, \varepsilon_l \max_{(i_1,\dots,i_m)\in C_m^{n_l}} \|h(X_{i_1}, \dots, X_{i_m})\|.$$

Thus we get

$$\begin{aligned}
&\frac{1}{n_l} \log P\big(\sup_{t\in[0,1]} \|U_{\varepsilon_l}(t) - U_{n_l}(t)\| \geq \delta\big) \\
&\qquad \leq \frac{1}{n_l} \log |C_m^{n_l}| + \frac{1}{n_l} \log P\left(|h(X_1, \dots, X_m)| > \frac{\delta}{\mathrm{const}(m)\varepsilon_l} \right)
\end{aligned}$$

and the result follows by our assumption (2.2) and by $n_l\, \varepsilon_l \to_{l\to\infty} 1$ via the Chebyshev-Markov inequality. □

Proof of Corollary 2.7. We only have to identify the rate function, that means, we have to check

$$\inf\{I(\nu(\cdot)), \nu(\cdot) \in K_\infty \text{ and } \int x d\nu(\cdot) = \phi(\cdot)\} = \int_0^1 \Lambda^*(\dot{\phi}(t))dt \tag{4.6}$$

Assume that the right hand side is finite. Using the identity (4.1) we get that the left hand side is greater than or equal to

$$\inf\{\sum_{i=1}^k (t_i - t_{i-1}) H(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}}|\mu); \int x d\nu(t_i) = \phi(t_i), 1 \leq i \leq k\} \tag{4.7}$$

for arbitrary $0 = t_0 < t_1 < \ldots < t_k \leq 1$. Putting

$$H_k(\nu(t_1, \ldots, t_k) \,|\, \mu) := \sum_{i=1}^{k} (t_i - t_{i-1}) H(\frac{\nu(t_i) - \nu(t_{i-1})}{t_i - t_{i-1}} | \mu)$$

and considering the case $k = 1, t_1 = 1$ we obtain

$$\inf\{H(\nu(1)|\mu); \int x d\nu(1) = \phi(1), \nu(1) \in K_{1,\infty}\} = \Lambda^*(\phi(1)),$$

where

$$K_{1,\infty} = \bigcup_{L \geq 0} \{\varrho \in \mathcal{M}_1(S) : H(\varrho|\mu) \leq L\},$$

as a consequence of Exercise 2.1.20 (i) and Exercise 3.3.12 in Deuschel and Stroock [6]. Since $\mathcal{M}^+(S) \ni \varrho = (\varrho_1, \ldots, \varrho_k) \mapsto (\int x d\varrho_1, \ldots, \int x d\varrho_k)$ is linear and $\Lambda^*(\cdot)$ is convex we get by a simple calculation that

$$H_k(\nu(t_1, \ldots, t_k) \,|\, \mu) = \sum_{i=1}^{k} (t_i - t_{i-1}) \Lambda^*(\frac{\phi(t_i) - \phi(t_{i-1})}{t_i - t_{i-1}}).$$

Since this is true for every $0 = t_0 < t_1 < \ldots < t_k \leq 1$ we get by Lemma 5.1.6 in [5] that the right hand side in (4.6) is smaller than the left hand side. In the case that $\int_0^1 \Lambda^*(\dot{\phi}(t))dt = \infty$ the same arguments as in the proof of Lemma 3.7 show that $\{\nu(\cdot) : \nu(\cdot) \in K_\infty$ and $\int x d\nu(\cdot) = \phi(\cdot)\} = \emptyset$.

Showing the other inequality is a little more involved. It suffices the prove that the left hand side of (4.6) is bounded by

$$J_\infty(\phi) := \sup_{0=t_0<t_1<\ldots<t_k\leq 1} \inf\{H_k(\nu(t_1, \ldots, t_k)|\mu)) : \int x d\nu(t_i) = \phi(i), \\ 1 \leq i \leq k, \nu \in K_\infty(t_1, \ldots, t_k)\}$$

where

$$K_\infty(t_1, \ldots, t_k) := \bigcup_{L \geq 0} \{\mu \in (\mathcal{M}^+(S))^k : H(\frac{\nu_i}{t_i}|\mu) \leq L; 1 \leq i \leq k\}.$$

Assume $J_\infty(\phi) < \infty$. Then by definition we can find a sequence $\varrho(t_1, \ldots, t_k) \in K_\infty(t_1, \ldots, t_k) \cap \{\int x d\varrho(t_i) = \phi(i), 1 \leq i \leq k\}$ such that $H_k(\varrho(t_1, \ldots, t_k)|\mu)) \nearrow J_\infty(\phi)$ (to see why, note that $H_k(\varrho(t_1, \ldots, t_k)|\mu))$ is increasing for a nested sequence of $\{(t_1, \ldots, t_k)\}_k)$. We may assume that the sequence $\{(t_1, \ldots, t_k)\}_k$ is increasing (nested) and the sets $\{(t_1, \dots, t_k)\}$ become dense in $[0,1]$. A simple computation gives

$$H_k(\varrho(t_1, \dots, t_k|\mu)) = \int_0^1 H_k(\dot{\tilde{\varrho}}(t_1, \dots, t_k)(s)|\mu)\, ds$$

where $\tilde{\varrho}(t_1, \dots, t_k)(\cdot)$ is the polygonalization of $\varrho(t_1, \dots, t_k)$. Now, since there exists a subsequence ϱ_n of these polygonalized measure valued functions that converges in the topology generated by uniform weak convergence on $[0,1]$ to some measure valued function ϱ, we get

$$J_\infty(\phi) = \liminf_{n\to\infty} \int_0^1 H(\dot{\varrho_n}|\mu) ds \geq \int_0^1 H(\dot{\varrho}|\mu)\, ds$$

and we are done if we can show that ϱ satisfies

$$\int x d\varrho(t) = \phi(t), \quad t \in [0,1].$$

Clearly $\varrho(\cdot) \in K_\infty$. For each $t \in \cup_k \{(t_1, \dots, t_k)\}$ we get by truncating the identity at height L (and denoting this map by x_L) and an application of Lusin's theorem as in the proof of Theorem 2.4 that

$$\begin{aligned} &|\int x d\varrho(t) - \phi(t)| \leq \\ &\quad |\int x d\varrho(t) - \int x_L d\varrho(t)| + |\int x_L d\varrho(t) - \int x_L d\varrho_n(t)| \leq 2\varepsilon \end{aligned}$$

by the weak convergence of the $\varrho_n(t)$ to $\varrho(t)$. For any other $s \in [0,1]$ we obtain for $t \in [0,1]$

$$\begin{aligned} &|\int x d\varrho(s) - \phi(s)| \leq \\ &|\int x d\varrho(s) - \int x_L d\varrho(s)| + |\int x_L d\varrho(s) - \int x_L d\varrho(t)| \\ &+ |\int x_L d\varrho(t) - \int x d\varrho(t)| + |\int x d\varrho(t) - \phi(t)| + |\phi(t) - \phi(s)|. \end{aligned}$$

Therefore, first choosing t in some set $\{(t_1, \dots, t_k)\}$ close enough to s such that the second term on the right hand side is bounded by a given $\varepsilon > 0$ and $|\phi(t) - \phi(s)| < \varepsilon$ (since $\varrho(t)$ weakly converges to $\varrho(s)$ and ϕ is continuous) and then choosing L large enough bounds $|\int x d\varrho(s) - \phi(s)|$ by 6ε and thus finishes the proof. $\square$

Let us conclude with an easy example where the rate function can be explicitly calculated.

Example 4.8. Let the kernel function h be given by $h(x,y) = x\,y$ and assume that μ is an arbitrary probability distribution such that (2.2) is fulfilled. Then – as a simple consequence of the contraction principle and Corollary 2.7 (Mogulskii's theorem) – the partial sums U-process with kernel h fulfills the LDP with speed n and rate function

$$I_\infty(\phi) = \int_0^1 \Lambda^*(\dot{\sqrt{\phi(t)}})\,dt \quad \text{if} \quad \phi \geq 0 \text{ and } \sqrt{\phi(t)} \in \mathcal{AC}([0,1],(\mathbb{R}^d,\|\cdot\|)),$$

and $I_\infty(\phi) = \infty$ otherwise. This can be easily seen by observing that $U_n(t) = (\int x dL_n(t))^2$. This calculation heavily depends on the special kernel function. For arbitrary kernel functions h such a nice representation of the rate function is, of course, out of reach.

Acknowledgment. We thank Tim Zajic for communicating us his observations concerning the representation of the rate function. Some of his ideas have entered into the proof of Corollary 2.7.

References

1. Arcones, M.A. and E. Giné, *Limit theorems for U-processes*, Ann. Probab. **7** (1993), 1494–1542.
2. Borovkov, A. A., *New limit theorems in boundary problems for sums of independent terms*, Sel. Transl. Math. Stat. **5** (1965), 315–372.
3. Borovkov, A. A., *Boundary-value problems for random walks and large deviations in function spaces*, Th. Prob. Appl. **12** (1967), 575–595.
4. Dembo, A. and T. Zajic, *Large deviations: from empirical mean and measure to partial sums process*, Stochastic Process. Appl. **57** (1995), 191–224.
5. Dembo, A. and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett, Boston, MA, 1993.
6. Deuschel, J. D. and D. W. Stroock, *Large Deviations*, Academic Press, Boston, MA, 1989.
8. Donsker, M.D. and S.R.S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time III*, Comm. Pure Appl. Math. **29** (1976), 389–461.
7. Dudley, R. M., *Probabilities and metrics*, Aarhus Universitet, Lecture notes series No.45, Aarhus Universitet, Mathematisk Institut, 1976.
9. Eichelsbacher, P., *Large deviations for products of empirical probability measures in the τ-topology*, to appear in J. Theoret. Probab..

10. Eichelsbacher, P. and M. Löwe, *Large deviation principle for m-variate von Mises-statistics and U-statistics*, J. Theoret. Probab. **8** (1995), 807–824.
11. Hoeffding, W., *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1948), 13–30.
12. Loynes, R.M., *An invariance principle for reversed martingales*, Proc. Amer. Math. Soc. **25** (1970), 56–64.
13. Mandelbaum, A. and M.S. Taqqu, *Invariance principle for symmetric statistics*, Ann. Statist. **12** (1984), 483–496.
14. Miller, R.G. and P.K. Sen, *Weak convergence of U-statistics and von Mises differentiable functions*, Ann. Math. Statist. **43** (1972), 31–41.
15. Mogulskii, A.A., *Large deviations in the space $C(0,1)$ for sums given on a finite Markov chain*, Sibirian Math.J., Eng. Translation **15** (1974), 43–53.
16. Mogulskii, A.A., *Large deviations for trajectories of multidimensional random walks*, Theory Probab. Appl. **21** (1976), 300–315.
17. Pérez–Abreu, V. and C. Tudor, *Large deviations for a class of chaos expansions*, J. Theoret. Probab. **7**, No. 4 (1994), 757–765.
18. Ronzhin, A. F., *A functional limit theorem for homogeneous U-statictics with a degenerate kernel*, Theory Prob. Applns. **30** (1986), 806–810.
19. Schuette, P.H., *Large deviations for trajectories of sums of independent random variables*, J. Theoret. Probab. **7** (1994), 3–45.
20. Varadhan, S.R.S., *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math. **19** (1966), 261–286.

Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D–33501 Bielefeld, Germany