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 Polish space valued i. i. d. random variables, where moreover appropriate exponential tail conditions are assumed to hold. We prove that similar 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 and with a law of X_i having a finite moment generating function, Varadhan [19] proved a LDP for the partial sums process

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

Mogulskii [16] extended the result to other scalings and also accomodated moment generating functions, which are finite only in a neighbourhood of the origin. In [15] the case of X_i which are real valued functions of a regular finite state Markov chain is treated. Schuette [18] analysed real valued, independent but not necessarily identical distributed X_i under more restrictive conditions on the corresponding moment generating functions. Moreover Dembo and Zajic [4] proved an extension of Varadhan's result to cover the LDP for

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

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

or for the Banach space valued partial sums process $S_n(\cdot)$, where the X_i fulfill Markov dependence or mixing conditions. They observed that the special conditions, which guarantees the LDP to hold for the empirical measures of Polish space valued random variables and for the empirical means of Banach space valued random variables, suffice for the LDP to carry over to the corresponding partial sums processes.

In this paper we investigate the LDP for partial sums U-processes, which means partial sums of the following form:

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

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 $\{(i_1, \ldots, i_m) : 1 \leq i_1 < \cdots < i_m \leq n\}$. To be more precise, we get an LDP 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 U-statistics of degree $m \in \mathbb{N}$, which means statistics of the form

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

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 [9] in the i. i. d. case. An extended version is proved in [8], 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 prove, that similar conditions suffice for the LDP to carry over to the partial sums U-processes defined as in (1.2).

While U-statistics were introduced by Hoeffding [10] 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 [11], Miller and Sen [14] and Mandelbaum and Taqqu [13]. They mainly proved invariance principles and the weak convergence of appropriate normalized "U-processe" to Brownian motion or limit processes expressable 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

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

where $D_m^n := \{(i_1, \ldots, i_m) : 1 \leq i_1, \ldots, i_m \leq n\}$ 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 $i_1, \ldots, i_m$ are equal. The partial sums V-process is now given by

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

Remark that $U_n(\cdot)$ and $V_n(\cdot)$ cannot be expressed as a contraction mapping of a "process level" empirical measure. Nevertheless, the contraction principle, i. e. a LDP is preserved under sufficiently continuous mappings, plays an important role for our result. Some well established tools of proving an LDP for partial sums will enter. The polygonalized versions of $U_n(\cdot)$ and $V_n(\cdot)$ respectively are used as well as the so called exponential equivalence of U_n and V_n, that is these statistics do not differ in their large deviations behaviour.

Moreover we prove that the LDP for $U_n(\cdot)$ follows as soon as it is established that on the one hand for every fixed $0 < t_1 < \cdots < t_k \leq 1$, $k \in \mathbb{N}$, the sequence

$$\big(L_n(t_1), L_n(t_2) - L_n(t_1), \ldots, L_n(t_k) - L_n(t_{k-1})\big)$$

(see (1.1)) satisfies the LDP in the product space $\big(\mathcal{M}^+(S)\big)^k$, where $\mathcal{M}^+(S)$ denotes the space of positive measures on S, equipped with the topology of weak convergence, and on the other hand the underlying sequence $\{X_n, n \in \mathbb{N}\}$ satisfies an exponential tail condition (see (2.2) below), so that we can apply a generalized version of the contraction principle (cf. Dembo and Zeitouni [5, Theorem 4.2.23] and Pérez–Abreu and Tudor [17, Theorem 2.1]). We need additional moment assumtions on the diagonal terms to establish the same outline for the partial sums V-processes.

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,

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

where $\operatorname{int}_t A$ denotes the interior of A with respect to the topology t on X and $\operatorname{cl}_t A$ the closure. By definition a good rate function is t-lower semicontinuous 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

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

and

$$\int_{S^{m-1}} \exp(\theta \| h(x_1, \dots, x_{m-1}, y) \|) \mu^{m-1} d(x_1, \dots, x_{m-1}) < \infty \quad \text{for all } \theta > 0$$

and all y. 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

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

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

Let Γ denote the space of all maps $\phi \colon [0,1] \to \mathbb{R}^d$ for which $\phi(0) = 0$, equipped with the topology of pointwise convergence on $[0,1]$. Let $C_0\big([0,1], (\mathbb{R}^d, \| \cdot \|)\big)$ denote the subspace of Γ consisting of all continuous elements. Let $\| \cdot \|_\infty$ denote the supremum norm on $L_\infty([0,1])$. Fix $k \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_k \le 1$, set $\varrho_0 = 0$ and define

$$H_k(\varrho \,|\, \mu) := \sum_{i=1}^{k} (t_i - t_{i-1}) H \left(\frac{\varrho_i - \varrho_{i-1}}{t_i - t_{i-1}} \middle| \mu \right)$$

for $\varrho := (\varrho_1, \dots, \varrho_k) \in \big(\mathcal{M}^+(S)\big)^k$. Since by definition $H(\nu \,|\, \mu) = +\infty$ for $\nu \notin \mathcal{M}_1(S)$ we may concentrate on those sequences of measures $\varrho = (\varrho_1, \dots, \varrho_k)$ with $\varrho_j(S) = t_j$, $1 \le j \le k$.

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*

$$I_\infty(\phi) = \sup_{0 \le t_1 < \cdots < t_k \le 1} \inf \Big\{ H_k(\varrho \,|\, \mu) \colon \int h \, d\varrho_j^m = \phi(t_j), \ 1 \le j \le k, \ \varrho \in K_\infty(t) \Big\} \tag{2.5}$$

and speed n, where $\phi \in C_0\big([0,1], (\mathbb{R}^d, \| \cdot \|)\big)$ and where for $t = (t_1, \dots, t_k)$

$$K_\infty(t) := \bigcup_{L \ge 0} \Big\{ \varrho \in \big(\mathcal{M}^+(S)\big)^k \colon H\left(\frac{\varrho_i}{t_i} \middle| \mu \right) \le L, \ 1 \le i \le k \Big\}$$

is the "skeleton of $H_k(\cdot\,|\,\mu)$". Set $I_\infty(\phi)=\infty$ otherwise.
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 good rate $I_\infty(\cdot)$ and speed ε^{-1}.*

Finally we get in the case $m=1$ the result of Varadhan [19] and Mogulskii [16]:

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 $Z_n(t)=1/n\sum_{i=1}^{[nt]}X_i\,, 0\le t\le 1$, satisfies the LDP in $L_\infty([0,1])$ with the good rate function*

$$I(\phi)=\int_0^1\Lambda^*(\dot\phi(t))\,dt\quad \textit{if}\quad \phi\in\mathcal{AC},\ \phi(0)=0,$$

$I(\phi)=\infty$ otherwise, where $\mathcal{AC}$ denotes the space of absolutely continuous functions, $\dot\phi$ 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 familiy $\{\nu_\varepsilon,\varepsilon>0\}$ of laws of $Y_\varepsilon(t):=\varepsilon\sum_{i=1}^{[t/\varepsilon]}X_i\,, 0\le t\le 1$, satisfies the LDP with the good rate function $I(\cdot)$.

§3 Preliminaries

We will use the following observation, due to Hoeffding [10, Section 5].

Lemma 3.1. *For $s>0$*

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

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 and is used in the proof of Theorem 2.4. The proof is based on the approach to Mogulskii's result taken in Dembo and Zeitouni [5, Section 5.1] and in Dembo and Zajic [4]. The following two Lemmas are standard in the sense of this approach. 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 $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log P\big(\|\widetilde{U}_n - U_n\|_\infty > \delta\big) = -\infty.$$

Remark that $\{\omega\colon \|\widetilde{U}_n - U_n\|_\infty > \delta\} \in \mathcal{A}$ 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

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

Remark that the right hand side is nearly a U-statistics of degree $m-1$. Thus by the exponential Chebychev-Markov inequality and the assumption on the moment generating function we have

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

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

$$\begin{aligned} &\|V_n(t) - U_n(t)\| \\ &\quad \le \left| 1 - \frac{n(n-1)\cdots(n-m+1)}{n^m} \right| \|U_n(t)\| + \frac{1}{n^m} \sum_r \|h(X_{i_1}, \dots, X_{i_m})\|, \end{aligned} \tag{3.3}$$

where the sum $\sum_r$ consists of all realizations of h with at least two of $i_1, \dots, i_m$ 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

$$\begin{aligned} &P\big(\|V_n(t) - U_n(t)\| > \delta\big) \\ &\quad \le P\bigg(\|U_n(t)\| > \frac{n\delta}{2C}\bigg) + P\bigg(\frac{1}{Cn^{m-1}} \Big\| \sum_r h \Big\| > \frac{n\delta}{2C}\bigg). \end{aligned} \tag{3.4}$$

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

$$\begin{aligned} P\bigg(\|U_n\|_\infty > \frac{n\delta}{2C}\bigg) &\le n^m P\bigg(\|h\| > \frac{n\delta}{2C}\bigg) \\ &\le n^m \exp\bigg(-\Theta n \frac{\delta}{2C}\bigg) \mathbb{E}_P\big[\exp(\Theta\|h\|)\big]. \end{aligned}$$

The second term on the right in (3.4) 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

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

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

Remark 3.5. Since under (2.3) the process $\{U_n(\cdot), n \in \mathbb{N}\}$ and $\{V_n(\cdot), n \in \mathbb{N}\}$ have identical large deviation behaviour, we can concentrate on $\{U_n(\cdot), n \in \mathbb{N}\}$ in the rest of the paper.

Recall that sequence $\{P \circ U_n(t)^{-1}, n \in \mathbb{N}\}$ is called exponentially tight if there exists compact sets K_α, $\alpha \in (0, \infty)$, such that

$$\lim_{\alpha\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(U_n(t) \in K_\alpha^c\big) = -\infty.$$

Lemma 3.6. *$\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ is exponentially tight in $C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$, equipped with the supremum norm topology.*

Proof. For each $t \in [0,1]$ fixed, the sequences $\{U_n(t), n \in \mathbb{N}\}$ and $\{\widetilde{U}_n(t), n \in \mathbb{N}\}$ are exponentially equivalent by the proof of Lemma 3.2. Therefore by the LDP for $\{U_n(t), n \in \mathbb{N}\}$ (cf. [9]) and Dembo and Zeitouni [5, Theorem .2.13], $\{\widetilde{U}_n(t), n \in \mathbb{N}\}$ satisfies the LDP in the Polish space $(\mathbb{R}^d, \|\cdot\|)$ and is therefore exponentially tight by Lynch and Sethuraman [12, Lemma 2.6]. For $f \in C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$ define the modulus of continuity of f by

$$\omega_f(\delta) := \sup_{|t-s|<\delta} \|f(t) - f(s)\| \qquad \text{for } \delta > 0.$$

We will check that there exists $n_0 \in \mathbb{N}$ such that for all $\eta > 0$

$$\lim_{\delta \to 0} \sup_{n \geq n_0} \frac{1}{n} \log P\big(\omega_{\widetilde{U}_n}(\delta) \geq \eta\big) = -\infty. \tag{3.7}$$

Then by Dembo and Zajic [4, Lemma A.2] $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ is exponentially tight in $C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$. By the polygonal structure of $\widetilde{U}_n$ we get

$$\begin{aligned}
P\big(\omega_{\widetilde{U}_n}(\delta) \geq \eta\big) &= P\Big(\bigcup_{j=1}^{[n\delta]} \bigcup_{i=1}^{n-j} \|U_n(\tfrac{i}{n}) - U_n(\tfrac{i+j}{n})\| \geq \eta\Big) \\
&\leq \sum_{i=1}^{n} P\bigg(\frac{1}{\binom{n}{m}} \sum_{C_{m-1}^{i}} \|h(X_{i_1}, \ldots, X_{i_{m-1}}, X_{i+1})\| > \frac{\eta}{[\delta n]}\bigg) \\
&\leq nP\bigg(\frac{n}{\binom{n}{m}} \sum_{C_{m-1}^{n-1}} \|h(X_{i_1}, \ldots, X_{i_{m-1}}, X_n)\| > \frac{\eta}{\delta + \frac{1}{n}}\bigg).
\end{aligned}$$

Now there exists a constant C such that $\frac{Cn}{\binom{n}{m}} \geq \frac{1}{\binom{n}{m-1}}$ and thus

$$P\big(\omega_{\widetilde{U}_n}(\delta) \geq \eta\big) \leq nP\bigg(\frac{1}{\binom{n}{m-1}} \sum_{C_{m-1}^{n-1}} \|h(X_{i_1}, \ldots, X_{i_{m-1}}, X_n)\| > \frac{\eta C}{\delta + \frac{1}{n}}\bigg).$$

We now apply Lemma 3.1 to the $(m-1)$-variate U-statistics case:

$$\limsup_{n \to \infty} \frac{1}{n} \log P\big(\omega_{\widetilde{U}_n}(\delta) \geq \eta\big) \leq -\frac{\eta C}{\delta} + \log \mathbb{E}_P\big[\exp(\eta \|\overline{h}\|)\big]$$

with $\overline{h}(x_1, \ldots, x_{m-1}) := h(x_1, \ldots, x_{m-1}, X_n)$. $\square$

Moreover we use the following easy observation (see [8, Lemma 3.1 and Corollary 3.4]).

Lemma 3.8. *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 transformed to every d_v-exponentially equivalent sequence of probability measures on $\mathcal{M}_1(S)$.*

§4 Proofs of the Theorems

First let us consider a bounded and continuous kernel function h.

Lemma 4.1. *Fix $k \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_k \leq 1$, setting $U_{n,k} = (U_n(t_1), \ldots, U_n(t_k))$. If h is bounded and continuous then $\{U_{n,k}, n \in \mathbb{N}\}$ satisfies the LDP in $(\mathbb{R}^d)^k$ with the good rate function*

$$I_k(z) := \inf\big\{H_k(\varrho\,|\,\mu) \colon \int h\, d\varrho_j^m = z_j \ , \ j = 1, \ldots, k \ , \ \varrho \in K_\infty(t)\big\},$$

where $z = (z_1, \ldots, z_k) \in (\mathbb{R}^d)^k$ and $\varrho = (\varrho_1, \ldots, \varrho_k) \in \big(\mathcal{M}_1(S)\big)^k$.

Proof. First of all notice, that $\int h\, d\varrho_j^m$ is not defined for all $\varrho_j \in \mathcal{M}_{t_j}(S)$. But with $x \cdot y \leq e^x + y \log y$, $x \geq 0$, $y \geq 0$, we get for $\widehat{\varrho}_j := \frac{\varrho_j}{t_j}$

$$\int \|h\| \, \frac{d\widehat{\varrho}_j^m}{d\mu^m}\, d\mu^m \leq \int \exp(\|h\|)\, d\mu^m + H(\widehat{\varrho}_j^m\,|\,\mu^m)$$

and thus if $\widehat{\varrho}_j \in \{\widehat{\varrho} \in \mathcal{M}_1(S) \colon H(\widehat{\varrho}\,|\,\mu) \leq L\}$ for some L we get $\int \|h\|\, d\widehat{\varrho}_j^m < +\infty$ by the moment condition (2.2). Therefore $\big(\int h\, d\, \varrho_j^m\big)_{1 \leq j \leq k}$ is well defined on $K_\infty(t)$. By Sanov's theorem (see for example Dembo and Zeitouni [5, Theorem 6.2.10]), $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ satisfies the LDP with the good rate function $H(\cdot\,|\,\mu)$. Thus $L_n(t)$ satisfies the LDP with rate function $H(\frac{\cdot}{t}\,|\,\mu)$ and speed nt as an easy consequence of the contraction principle (cf. Dembo and Zeitouni, [5, Theorem 4.2.1]). For $t_j < t_l$, we observe by the same argument, that $L_n(t_l) - L_n(t_j)$ satisfies the LDP with rate function $H\big(\frac{\cdot}{t_l - t_j}\,|\,\mu\big)$ and speed $n(t_l - t_j)$. Since the components of $\widetilde{L}_n := \big(L_n(t_1), L_n(t_2) - L_n(t_1), \ldots, L_n(t_k) - L_n(t_{k-1})\big)$ are independent, $\{\widetilde{L}_n, n \in \mathbb{N}\}$ satisfies the LDP on $\big(\mathcal{M}^+(S)\big)^k$ with rate function

$$\sum_{i=1}^k (t_i - t_{i-1}) H\bigg(\frac{\varrho_i}{t_i - t_{i-1}}\bigg|\mu\bigg)$$

and speed n (cf. Bezuidenhout [2, Proposition 2.17]). Again applying the contraction principle to the continuous map $(z_1, \ldots, z_k) \mapsto (z_1, z_1 + z_2, \ldots, \sum_{i=1}^k z_i)$ on

$\big(\mathcal{M}^+(S)\big)^k$, it follows that $\big(L_n(t_1),\dots,L_n(t_k)\big)$ satisfies the LDP in $\big(\mathcal{M}^+(S)\big)^k$ with the good rate function $H_k(\varrho\,|\,\mu)$. Since S is Polish the topology of weak convergence in $\mathcal{M}^+(S)$ is homeomorphic to the metric topology induced by the Prohorov metric d. Consider on $\big(\mathcal{M}^+(S)\big)^k$ the product topology, which is induced by the metric $\widetilde{d}(\varrho,\nu) := \sup\big\{d(\varrho_i,\nu_i)\colon 1\le i\le k\big\}$ for $\varrho,\nu\in\big(\mathcal{M}^+(S)\big)^k$. Since the map $\big(\mathcal{M}^+(S)\big)^k \ni \varrho=(\varrho_1,\dots,\varrho_k)\mapsto(\varrho_1^m,\dots,\varrho_k^m)\in\big(\mathcal{M}^+(S^m)\big)^k$ is continuous in the weak topological setting, the vector $\big(L_n^m(t_1),\dots,L_n^m(t_k)\big)$ satisfies the LDP with the good rate function $H_k^m(\cdot)\colon \big(\mathcal{M}^+(S^m)\big)^k\to[0,\infty]$ given by

$$H_k^m(\nu) := \begin{cases} H_k(\varrho\,|\,\mu) & (\varrho_1^m,\dots,\varrho_k^m)=\nu;\ \varrho=(\varrho_1,\dots,\varrho_k)\\ +\infty & \text{otherwise.}\end{cases}$$

Define

$$\widetilde{L}_{n,m}(t)=\frac{1}{\binom{n}{m}}\sum_{C_m^{[nt]}} f(\delta_{X_{i_1}},\dots,\delta_{X_{i_m}}),$$

where $f(\delta_{X_{i_1}},\dots,\delta_{X_{i_m}}) := \frac{1}{m!}\sum_p \delta_{X_{i_1}}\otimes\cdots\otimes\delta_{X_{i_m}}$ and p denotes the summation of all $m!$ permutations $(i_1,\dots,i_m)$ of $(1,\dots,m)$. Analogously to (3.3) we get

$$\|L_n^m(t)-\widetilde{L}_{n,m}(t)\|_v \le \frac{C}{n}\|\widetilde{L}_{n,m}(t)\|_v+\frac{C}{n}\le\frac{2C}{n}$$

with a constant C depending on m only. Thus with $\|\varrho-\nu\|_{vp} := \sup\big\{\|\varrho_i-\nu_i\|_v\colon 1\le i\le k\big\}$ for $\varrho,\nu\in\big(\mathcal{M}^+(S)\big)^k$, we get

$$\Big\|\big(L_n^m(t_1),\dots,L_n^m(t_k)\big)-\big(\widetilde{L}_{n,m}(t_1),\dots,\widetilde{L}_{n,m}(t_k)\big)\Big\|_v\le\frac{2C}{n}$$

and thus by Lemma 3.8, applied to $\|\cdot\|_v$ and $\widetilde{d}$, $\big(\widetilde{L}_{n,m}(t_1),\dots,\widetilde{L}_{n,m}(t_k)\big)$ satisfies the LDP on $\big(\mathcal{M}^+(S)\big)^k$ with respect to the weak topology. Now observe that $\big(\int h\,d\widetilde{L}_{n,m}(t_1),\dots,\int h\,d\widetilde{L}_{n,m}(t_k)\big)=\big(U_n(t_1),\dots,U_n(t_k)\big)$, thus for bounded and continuous kernel functions h the assertion of the Lemma follows by the simple form of the contraction principle. Define the contraction map $F\colon\big(\mathcal{M}^+(S^m)\big)^k\to(\mathbb{R}^d)^k$ by $F(\varrho_1,\dots,\varrho_k) := \big(\int h\,d\varrho_1,\dots,\int h\,d\varrho_k\big)$ and observe that F is continous. $\square$

Lemma 4.2. *Fix $k\in\mathbb{N}$ and $0=t_0<t_1<\cdots<t_k\le 1$. If h satisfies the condition* (2.2)*, then $\{U_{n,k}, n\in\mathbb{N}\}$ satisfies the LDP in $(\mathbb{R}^d)^k$ with the good rate function $I_k(\cdot)$ as in Lemma* 4.1*.*

Proof. For general h fulfilling the assumption (2.2) we apply an extended version of the contraction principle (cf. Dembo and Zeitouni [5, Theorem 4.2.23] and

Pérez–Abreu and Tudor [17, Theorem 2.1]) and Lusin's theorem. For an arbitrary measurable and symmetric h, put for $0 < L < \infty$

$$h_L(x_1, \dots, x_m) := \begin{cases} h(x_1, \dots, x_m) & \text{if } \|h(x_1, \dots, x_m)\| \le L \\ 0 & \text{otherwise.} \end{cases}$$

Due to Lusin's theorem there exists a continuous function $h_L^c \colon S^m \to \mathbb{R}^d$ such that $\|h_L^c\| \le L$ and $\mu^m(\|h_L - h_L^c\| \ge \frac{1}{L}) \le \exp(-L^2)$. Define $F_L \colon \big(\mathcal{M}^+(S^m)\big)^k \to (\mathbb{R}^d)^k$ by $F_L(\varrho) := \big(\int h_L^c \, d\varrho_1, \dots, \int h_L^c \, d\varrho_k\big)$. F_L is continuous and well defined on $K_\infty(t)$. We have to check

$$\limsup_{L \to \infty} \sup \big\{ \|F(\varrho) - F_L(\varrho)\|_{\sup} \colon H_k^m(\varrho) \le C \big\} = 0, \tag{4.3}$$

where $\|x\|_{\sup} := \sup_{i=1,\dots,k} \|x_i\|$, $x = (x_1, \dots, x_k) \in (\mathbb{R}^d)^k$. If $\widehat{\varrho} \in \{H_k^m(\varrho) \le C\}$ we get $H\big(\frac{\varrho_i - \varrho_{i-1}}{t_i - t_{i-1}} \big| \mu\big) \le \widetilde{C}$ for every $1 \le i \le k$, where for each $\widehat{\varrho}$ the ϱ_i's are the one-dimensional marginal distributions of the i's component. By convexity we get

$$H\left(\frac{\varrho_l}{t_l}\middle|\mu\right) = H\left(\sum_{i=1}^{l} \frac{t_i - t_{i-1}}{t_l} \, \frac{\varrho_i - \varrho_{i-1}}{t_i - t_{i-1}}\middle|\mu\right) \le \widetilde{C} \tag{4.4}$$

for every $1 \le l \le k$. Now since $\{H(\cdot \,|\, \mu) \le K\}$ is uniformly absolutly continuous with respect to μ for every $K \le 0$ one gets

$$\limsup_{L \to \infty} \int_{S^m} \|h - h_L\| d\left(\frac{\varrho_l}{t_l}\right)^m = 0$$

uniformly in $\frac{\varrho_l}{t_l}$ with $H\big(\frac{\varrho_l}{t_l} \big| \mu\big) \le \widetilde{C}$ (cf. Donsker and Varadhan [7, Lemma 5.1]) and (4.3) follows if we can prove

$$\limsup_{L \to \infty} \Big\{ \sup_{1 \le i \le k} \int \|h_L - h_L^c\| d\varrho_i^m, \; H_k^m(\varrho) \le C \Big\} = 0. \tag{4.5}$$

To this end put $\|h_L - h_L^c\| = f_L^{(1)} + f_L^{(2)}$ with $f_L^{(1)} \le \frac{1}{L}$ and $\mu^m(f_L^{(2)} \ne 0) \le e^{-L^2}$ and $f_L^{(1)}, f_L^{(2)} \ge 0$.

Clearly

$$\limsup_{L \to \infty} \sup_{1 \le i \le k} \int f_L^{(1)} d\varrho_i^m = 0$$

uniformly on each level set of $H(\cdot\,|\,\mu)$. With $x\,y \le e^{\sigma\,x} + \frac{y}{\sigma}\log\frac{y}{\sigma}$ for all $x, y \ge 0$ and $\sigma > 0$ we obtain for $\sigma \ge 1$

$$\begin{aligned}\int f_L^{(2)} d\varrho_i^m &= \int_{\{f_L^{(2)}\neq 0\}} t_i^m f_L^{(2)} \frac{d\big(\frac{\varrho_i}{t_i}\big)^m}{d\mu^m} d\mu^m \\ &\le \int_{\{f_L^{(2)}\neq 0\}} \exp\big(\sigma\, t_i^m f_L^{(2)}\big) d\mu^m + \frac{1}{\sigma}\int_{S^m} \left|\log \frac{d\big(\frac{\varrho_i}{t_i}\big)^m}{d\mu^m}\right| d\Big(\frac{\varrho_i}{t_i}\Big)^m.\end{aligned} \tag{4.6}$$

Since the second term on the right in (4.6) converges uniformly on each level set of $H(\cdot\,|\,\mu)$, and therefore on each level set of $H(\cdot\,|\,\mu^m)$, to zero as σ tends to infinity (cf. Donsker and Varadhan [7, proof of Lemma 5.1]), by the same argument as above, it works uniformly in $1 \le i \le k$. Since moreover

$$\int_{\{f_L^{(2)}\neq 0\}} \exp\big(\sigma\, t_i^m f_L^{(2)}\big) d\mu^m \le \exp\big(L(t_i^m \sigma - L)\big) \tag{4.7}$$

which is converging to zero with $L \to \infty$ for every fixed σ we arrive at (4.5) and therefore at (4.3). The last step in applying the contraction principle is to verify the exponential equivalence between $U_{n,k}$ with kernel h and $U_{n,k}$ with kernel h_L^c. We have to check

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P\big(\|U_{n,k}(h) - U_{n,k}(h_L^c)\|_{\sup} \ge \varepsilon\big) = -\infty \tag{4.8}$$

for every $\varepsilon > 0$. Note that (4.8) only has to be shown for every fixed $i \in \{1,\dots,k\}$ by the definition of $\|\cdot\|_{\sup}$. Applying the exponential Chebyshev-Markov inequality yields for any positiv $\Theta \in \mathbb{R}$

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

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

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

By the arbitraryness of Θ we conclude

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

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

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

for every positiv ε. Again applying Chebyshev-Markov inequality, Lemma 3.1 and (4.7) we get

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

for every $\Theta, \varepsilon > 0$. So letting $L \to \infty$ and then $\Theta \to \infty$ proves (4.8) and thus the lemma. □

Remark. Since by Lemma 3.2 $U_{n,k}$ and $V_{n,k} := (V_n(t_1), \dots, V_n(t_k))$ are exponentially equivalent in $(\mathbb{R}^d)^k$ under the assumption (2.3), we get the LDP for $\{V_{n,k}, n \in \mathbb{N}\}$ with the same rate.

Lemma 4.9. *If* (2.2) *holds, then* $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ *satisfies the LDP in* Γ *with the good rate function* (2.5)*. If* (2.3) *holds,* $\{\widetilde{V}_n(\cdot), n \in \mathbb{N}\}$ *satisfies the LDP in* Γ *with the same rate function.*

Proof. (see construction in Dembo and Zeitouni [5, Section 5.1]) The key oberservation is that Γ can be regarded as a projective limit. Let J denote the collection of all ordered finite subsets of $(0,1]$ and partially order J by inclusion. If $i, j \in J$ and $i \le j$ define the projection $p_{ij} : (\mathbb{R}^d)^{|j|} \to (\mathbb{R}^d)^{|i|}$ in the natural way. It is easy to check that the projective limit of $\{(\mathbb{R}^d)^{|j|}\}_{j\in J}$ with respect to the projections p_{ij} can be identified with Γ (cf. Dembo and Zeitouni [5, Lemma 5.1.6]). Therefore the statement of the lemma follows by Dawson and Gärtner [3, Theorem 3.3] in conjunction with Lemma 3.2 and Lemma 4.1 and 4.2. □

Next we will show that the right hand side of (2.5) can only be finite for functions in $C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$. The idea behind this is to apply the variational formula for the relative entropy $H(\cdot\,|\,\mu)$.

Lemma 4.10. *If* $I_\infty(\phi) < \infty$ *then* $\phi \in C_0\big([0,1], (\mathbb{R}^d, \|\cdot\|)\big)$*, so the typical realizations of the process are continuous.*

Proof. Suppose ϕ is not continuous in $0 \le t \le 1$. Let us assume that

$$\inf_{t<s\le s_0} |\phi(s) - \phi(t)| \ge \delta$$

for some $\delta > 0$ and some $s_0 \le 1$. The other case $\big(\inf_{s_0 \le s < t} |\phi(t) - \phi(s)| \ge \delta\big)$ is treated analogously. For fixed $s \in (t, s_0]$ consider a $\varrho \in \big(\mathcal{M}^+(S)\big)^2$ with $\int h \, d\varrho_t^m = \phi(t)$ and $\int h \, d\varrho_s^m = \phi(s)$. Then for all $\lambda > 1$ we have

$$\begin{aligned}\lambda\,\delta \le \lambda|\phi(t) - \phi(s)| &= \lambda\bigg|\int h\, d\varrho_t^m - \int h\, d\varrho_s^m\bigg| \\ &\le \lambda \int \|h\|\, d(\varrho_s^m - \varrho_t^m) \\ &= \lambda \int \|h\| d\bigg(\sum_{i=1}^m \varrho_s^{i-1}(\varrho_s - \varrho_t)\varrho_t^{m-i}\bigg).\end{aligned}$$

Hence for at least one i and all $\varepsilon > 0$ we get

$$\begin{aligned}\frac{\lambda\,\delta}{m} - \varepsilon &\le \lambda \int \|h\|\, d(\varrho_s^{i-1}(\varrho_s - \varrho_t)\varrho_t^{m-i}) - \varepsilon \\ &\le \lambda \int \|h\|\, d(\varrho_s^{i-1}(\varrho_s - \varrho_t)\varrho_t^{m-i}) - s^i(s-t)t^{m-i} \log \int \exp(\lambda\|h\|)\, d\pi^m \\ &\le H\bigg(\frac{\varrho_s^{i-1}(\varrho_s - \varrho_t)\varrho_t^{m-i}}{s^i(s-t)t^{m-i}}\bigg|\pi^m\bigg) s^i(s-t)t^{m-i},\end{aligned}$$

where for the second inequality $s - t$ is chosen small enough, and where the last inequality follows from $x\,y \le \exp(\lambda\,x) + \frac{y}{\lambda}\log\frac{y}{\lambda}$, $x, y > 0$. As

$$\begin{aligned}&H\bigg(\frac{\varrho_s^{i-1}(\varrho_s - \varrho_t)\varrho_t^{m-i}}{s^i(s-t)t^{m-i}}\bigg|\pi^m\bigg) \\ &\qquad = (i-1)H\bigg(\frac{\varrho_s}{s}\bigg|\pi\bigg) + H\bigg(\frac{\varrho_s - \varrho_t}{s-t}\bigg|\pi\bigg) + (m-i)H\bigg(\frac{\varrho_t}{t}\bigg|\pi\bigg)\end{aligned}$$

we obtain

$$\sup\bigg\{s\,H\bigg(\frac{\varrho_s}{s}\bigg|\pi\bigg), (s-t)\,H\bigg(\frac{\varrho_s - \varrho_t}{s-t}\bigg|\pi\bigg), t\,H\bigg(\frac{\varrho_t}{t}\bigg|\pi\bigg)\bigg\} = \infty.$$

As this holds true for all ϱ with $\int h\, d\varrho_s^m = \phi(s)$ and $\int h\, d\varrho_t^m = \phi(t)$ we obtain

$$I_\infty(\phi) \ge \sup_{(t,s),(t,0),(s,0)} \inf\bigg\{s\,H\bigg(\frac{\varrho_s}{s}\bigg|\pi\bigg), (s-t)H\bigg(\frac{\varrho_s - \varrho_t}{s-t}\bigg|\pi\bigg), t\,H\bigg(\frac{\varrho_t}{t}\bigg|\pi\bigg)\bigg\} = \infty.$$

$\square$

Remark. Other than in the case $m = 1$ where the rate function is infinite even on the set of all absolutely continuous functions we only can show that the interesting functions are contained in $C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$. Technically this originates in the fact that a sudden change of the realized function may have its origin in the choice of an untypical instantaneous measure as well as in former realizations of the empirical measure. Anyhow we think that this weaker result is probably the price one has to pay for considering the multivariate case.

Proof of the Theorem 2.4. By Lemma 4.9, $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ satisfies the LDP in Γ with rate function $I_\infty(\cdot)$ and consequently, $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ satisfies the LDP in $C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big) \subset \Gamma$, equipped the the relative topology induced by the pointwise convergence topology on Γ, since by definition of $U_n(\cdot)$ and $\widetilde{U}_n(\cdot)$ we get $P\big(\widetilde{U}_n(\cdot) \in C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)\big) = 1$ and by Lemma 4.10 $I_\infty(\phi) = \infty$ for $\phi \notin C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ (cf. Dembo and Zeitouni [5, Lemma 4.1.5]). Since the topology of uniform convergence in $C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ is finer than the pointwise convergence topology, the exponential tightness of $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ (cf. 3.6) allows for the strengthening of the LDP to the supremum norm topology on $C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ by Corollary 4.2.6 in Dembo and Zeitouni [5]. Since $C_0\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ is a closed subset of $L_\infty([0,1])$, the LDP also holds in $L_\infty([0,1])$ by Lemma 4.1.5 in Dembo and Zeitouni [5]. The exponential equivalence of $\{U_n(\cdot), n \in \mathbb{N}\}$ and $\{\widetilde{U}_n(\cdot), n \in \mathbb{N}\}$ in $L_\infty\big([0,1],(\mathbb{R}^d,\|\cdot\|)\big)$ established in Lemma 3.2 implies the result for $\{U_n(\cdot), n \in \mathbb{N}\}$ (cf. Dembo and Zeitouni [5, Theorem 4.2.13]). The result for $\{V_n(\cdot), n \in \mathbb{N}\}$ follows by 3.2 analogously. □

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 $\delta > 0$

$$\limsup_{l\to\infty} \frac{1}{n_l} \log P(\|U_{\varepsilon_l} - U_{n_l}\|_\infty \geq \delta) = -\infty,$$

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

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

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

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

Thus we get

$$\begin{aligned}
&\frac{1}{n_l}\log P\big(\|U_{\varepsilon_l} - U_{n_l}\|_\infty \ge \delta\big)\\
&\qquad \le \frac{1}{n_l}\log|C_m^{n_l}| + \frac{1}{n_l}\log P\bigg(|h(X_1,\dots,X_m)| > \frac{\delta}{\mathrm{const}(m)\varepsilon_l}\bigg)
\end{aligned}$$

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

Proof of Corollary 2.7. By Theorem 5.1.19 in Dembo and Zeitouni ([5]), $\{\nu_\varepsilon, \varepsilon > 0\}$ and $\{\mu_\varepsilon, \varepsilon > 0\}$ have identical LDP behaviour. Thus we only have to identify the rate function. Considering the case $k = 1, t_1 = 1, m = 1$ and $h(x) = x$ we obtain

$$\inf\{H(\varrho_1\,|\,\mu) : \int x\varrho_1 = \phi(1)\ ,\ \varrho_1 \in K_{1,\infty}\} = \Lambda^*(\phi(1)),$$

where $K_{1,\infty} := \cup_{L\ge 0}\{\varrho \in \mathcal{M}_1(S) : H(\varrho\,|\,\mu) \le L\}$, as a consequence of Exercise 2.1.20(i) and Exercise 3.3.12 of Deuschel and Stroock ([6]). Since $\big(\mathcal{M}^+(S)\big)^k \ni \varrho = (\varrho_1,\dots,\varrho_k) \mapsto \big(\int x\, d\varrho_1, \dots, \int x\, d\varrho_k\big)$ is linear and $\Lambda^*(\cdot)$ is convex, we get by a simple calculation

$$I_\infty(\phi) = \sup_{0\le t_1<\dots<t_k\le 1} \sum_{i=1}^k (t_i - t_{i-1})\,\Lambda^*\bigg(\frac{\phi(t_i)-\phi(t_{i-1})}{t_i - t_{i-1}}\bigg)$$

and by Lemma 5.1.6 and its proof in Dembo and Zeitouni ([5]), $I_\infty(\phi) = I(\phi)$. $\square$

Example. If $\mu = \delta_0(1-p) + \delta_1 p, 0 \le p \le 1$, and $\varrho_i = \delta_0(t_i - q_i) + \delta_1 q_i, 0 \le q_i \le 1, 0 \le t_i \le 1, 1 \le i \le k$, we get

$$H_k(\varrho\,|\,\mu) = \sum_{i=1}^k (t_i - t_{i-1}) I_p\bigg(\frac{q_i - q_{i-1}}{t_i - t_{i-1}}\bigg),$$

where $I_p(x) := x\log\frac{x}{p} + (1-p)\log\frac{(1-x)}{(1-p)}$ for $x \in [0,1]$ and $I_p(x) = \infty$ otherwise (the Cramér rate function for the Bernoulli(p)-distributed case). If $h(x,y) = x\,y$, we get $\int h \, d\varrho_i^2 = q_i^2$ and thus $\int h \, d\varrho_i^2 = \phi(t_i) \Leftrightarrow q_i = \sqrt{\phi(t_i)}$, if $\phi \geq 0$. The rate function governing this LDP is

$$\sup_{0 \leq t_1 < \cdots < t_k \leq 1} \sum_{i=1}^{k} (t_i - t_{i-1}) I_p \left(\frac{\sqrt{\phi(t_i)} - \sqrt{\phi(t_{i-1})}}{t_i - t_{i-1}} \right),$$

and via the same calculation as in the proof of Lemma 5.1.6 in Dembo and Zeitouni ([5]), we get

$$I_\infty(\phi) = \int_0^1 \Lambda^*(\dot{\sqrt{\phi(t)}}) \, dt \quad \text{if} \quad \sqrt{\phi(t)} \in \mathcal{A}C,$$

and $I_\infty(\phi) = \infty$ otherwise. This calculation heavily depends on the special kernel function and the underlying law with only two atoms. For arbitrary kernel functions h and arbitrary laws μ the calculation becomes more difficult.

References

1. Arcones, M.A. and E. Giné, *Limit theorems for U-processes*, Ann. Probab. **7** (1993), 1494–1542.
2. Bezuidenhout, C., *A large deviations principle for small perturbations of random evolution equations*, Ann. Probab. **15** (1987), 1014–1043.
3. Dawson, D. and J. Gärtner, *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions*, Stochastics **20** (1987), 247–308.
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.
7. 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.
8. Eichelsbacher, P., *Large deviations for products of empirical probability measures in the τ-topologie,* preprint 95–28, Sonderforschungsbereich 343, Diskrete Strukturen in der Mathematik, Universität Bielefeld.
9. Eichelsbacher, P. and M. Löwe, *Large deviation principle for m-variate von Mises-statistics and U-statistics*, J. Theoret. Probab. **8** (1995), 807–824.
10. Hoeffding, W., *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1948), 13–30.
11. Loynes, R.M., *An invariance principle for reversed martingales*, Proc. Amer. Math. Soc. **25** (1970), 56–64.
12. Lynch, M. and M. Sethuraman, *Large deviations for processes with independent increments*, Ann. Probab. **15** (1987), 610–627.

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. Schuette, P.H., *Large deviations for trajectories of sums of independent random variables*, J. Theoret. Probab. **7** (1994), 3–45.
19. 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