REFINED LARGE DEVIATIONS FOR VON MISES–STATISTICS

PETER EICHELSBACHER

ABSTRACT. We give sufficient conditions for the large deviations principle of real valued von Mises–statistics improving previous results. As a consequence we obtain sufficient conditions for the large deviations principle for Banach-space valued U-statistics improving previous results as well. The proofs are based on large deviations results for stochastic processes due to Arcones and a spectral decomposition of the kernel function of the von Mises–statistic and the U-statistic, respectively.

1. INTRODUCTION

Let $(B, \|\cdot\|)$ be a real separable Banach space with a norm $\|\cdot\|$ and let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables taking values in a measurable space $(S, \mathcal{S})$. For simplicity we assume that they are defined on the product space $(\Omega, \mathcal{A}, P) = (S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$, where μ is the distribution of X_1. Denote by $\mathcal{M}_1(S)$ the space of all probability measures on $(S, \mathcal{S})$. Consider for $n \geq m$, $n, m \in \mathbb{N}$,

$$U_n(h) := U_n^m(h, \mu) := \frac{1}{\binom{n}{m}} \sum_{1\leq i_1<\cdots<i_m\leq n} h(X_{i_1}, \ldots, X_{i_m}),$$

where $h : S^m \to B$ is a Bochner integrable symmetric function. $U_n(h)$ is called a B-valued *U-statistic* with *kernel function* h and *degree* m. Moreover we consider the corresponding *von Mises–statistic*

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

In this paper we analyze the large deviations principle for these 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 $I : \mathcal{X} \to [0, \infty]$ if the level sets $\{x : I(x) \leq \alpha\}$ are compact for all $\alpha < \infty$ and for all $\Gamma \in \mathcal{B}$ the lower

1991 *Mathematics Subject Classification.* 60F10, 60G05, 62G20.
Key words and phrases. large deviations, U-statistics, von Mises–statistics, stochastic processes.

bound

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

and the upper bound

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

hold, where $\mathrm{int}(\Gamma)$ and $\mathrm{cl}(\Gamma)$ denote the interior and closure of Γ, respectively. We say that a sequence of random variables satisfies the LDP when the sequence of measures induced by these variables satisfies the LDP.

The LDP of $\{U_n(h)\}_{n\in\mathbb{N}}$ have been studied already. In Eichelsbacher (1997) the real-valued case $B = \mathbb{R}^d$ was considered: If h satisfies the *strong Cramér condition*, that is

$$\int_{S^m} \exp\bigl(\theta \|h\|\bigr)\, d\mu^{\otimes m} < \infty \quad \text{for all} \quad \theta > 0,$$

then the sequence $\{U_n(h)\}_{n\in\mathbb{N}}$ satisfies the LDP with rate function

$$I(x) = \inf\bigl\{H(\varrho|\mu) : \varrho \in K_\infty, F(\varrho) = x\bigr\}, \quad x \in \mathbb{R}^d.$$

Here $H(\varrho|\mu)$ is the relative entropy of $\varrho \in \mathcal{M}_1(S)$ with respect to $\mu \in \mathcal{M}_1(S)$ defined by

$$H(\varrho|\mu) = \begin{cases} \int_S f \log f \, d\mu, & \text{if } \varrho \ll \mu \text{ and } f = \frac{d\varrho}{d\mu}, \\ \infty & \text{otherwise,} \end{cases}$$

and $K_\infty = \bigcup_{L\geq 0} K_L$ with $K_L = \{\varrho \in \mathcal{M}_1(S) : H(\varrho|\mu) \leq L\}$ and $F : \mathcal{M}_1(S) \cap K_\infty \to \mathbb{R}^d$ is defined by

$$F(\varrho) = \int_{S^m} h \, d\, \varrho^{\otimes m},$$

(note that the map is well-defined on K_∞). If in addition h satisfies the condition that there exists at least one $\alpha_h > 0$ such that

$$\int_{S^m} \exp\bigl(\alpha_h \|h \circ \pi_\tau\|\bigr)\, d\mu^{\otimes m} < \infty \tag{1.1}$$

for every map $\tau : \{1, \ldots, m\} \to \{1, \ldots, m\}$, where $\pi_\tau : S^m \to S^m$ is defined by $\pi_\tau(s) = (s_{\tau(1)}, \ldots, s_{\tau(m)})$ for every $s = (s_1, \ldots, s_m) \in S^m$, then $\{V_n(h)\}_{n\in\mathbb{N}}$ satisfies the LDP with the same rate function I. This result is given in Eichelsbacher and Schmock (2002).

In the case of an arbitrary real separable Banach space valued kernel h, Eichelsbacher and Schmock (2002) proved the LDP for U-statistics if h satisfies the strong Cramér condition, and for von Mises–statistics if in addition h satisfies condition (1.1). As one can see from the definition of the rate function I, the proofs in the papers just mentioned apply a generalized contraction principle, see Theorem 4.2.23 in Dembo and Zeitouni (1998) and use a LDP for products of empirical measures in a topology which make integration with respect to certain unbounded functions a continuous operation.

This approach gives rise to the question if $F(\cdot)$ is continuous on the level sets K_L, $L \geq 0$, of the relative entropy (continuity on the level sets is the minimal condition needed to apply the generalized contraction principle). But we obtain from Theorem 3 in Schied (1998) that continuity of $F(\cdot)$ on the level sets of $H(\cdot|\mu)$ is *equivalent* to the strong Cramér condition holding for h. This result shows that in case a weaker assumption than the strong Cramér condition is assumed the LDP for U-statistics cannot be deduced by known contraction techniques from Sanov's theorem, even in the real-valued case. This seems to be surprisingly.
In the case $m = 1$ and $h(x) = x$ it is well known that the classical Cramér theorem states that the *weak Cramér condition*

$$\int_S \exp\big(\delta\|x\|\big)\, d\mu < \infty \quad \text{for some} \quad \delta > 0$$

is a necessary and sufficient condition to have a LDP with a good rate function.
Therefore natural questions are: Do we get a LDP for von Mises–statistics and U-statistics, respectively, under weaker assumptions? Which method could replace the contraction technique?
In this paper we improve previous LDP-results for real-valued von Mises-statistics. Surprisingly the result looks like an exercise. Applying a spectral decomposition of the kernel function h as well as a new LDP-development for stochastic processes due to Arcones (2001), we give new sufficient conditions for the LDP. Moreover we will weaken the assumptions for the LDP of U-statistics. If the diagonal terms fulfill the weak Cramér condition, we can give improved sufficient conditions for the LDP of U-statistics. This is presented in Section 2.
In Section 3 we improve the LDP result for a Banach space valued kernel h, given in Eichelsbacher and Schmock (2002). For the proof we apply results from Arcones (2001) as well as an inequality due to Hoeffding. (Hoeffding 1963).

2. The real valued case

Along the lines of Arcones (1992) and Arcones (2001) we obtain new sufficient conditions for the LDP of a real valued von Mises–statistic in the case $m = 2$. We state this result and after-wards give some comments for the formulation for an arbitrary $m \geq 2$. The result relies on a LDP of sums of Banach space valued i.i.d. random variables. Let us shortly summarize the state of the art. If $(B, \|\cdot\|)$ is a real separable Banach space and if $\{X_i\}$ is a sequence of B-valued i.i.d. random variables, Sethuraman (1964) and Donsker and Varadhan (1976) showed that if $E(\exp(\lambda\|X_1\|)) < \infty$ for each $\lambda > 0$ then the LDP holds for $\big\{\frac{1}{n}\sum_{i=1}^n X_i\big\}$ with speed $1/n$ and with rate function

$$I(x) = \sup\{f(x) - \log\big(E\big(\exp(f(X_1))\big)\big) : f \in B^*\},$$

where B^* is the dual of B. Arcones (2001, Corollary 2.8) proved that the LDP holds for

$\{\frac{1}{n}\sum_{i=1}^n X_i\}$ with speed $1/n$ and a good rate function *if and only if* there exists a $\lambda > 0$ such that

$$E[\exp(\lambda \|X\|)] < \infty,$$

and for each $\lambda > 0$ there exists a $\eta > 0$ such that $E[\exp(\lambda Y^\eta)] < \infty$, where

$$Y^\eta := \sup\{ \quad |f_1(X_1) - f_2(X_1)| : f_1, f_2 \in B_1^*, \\ E\big[|f_1(X_1) - f_2(X_1)|\big] \le \eta\}, \tag{2.1}$$

where B_1^* is the unit ball of B^*. In Arcones (2001, Theorem 2.7) we find some formula for the rate function in this case, but it seems not to be obvious that the formula is the same as in the case where the X_i satisfies the strong Cramér condition. The sequence $\{1/n \sum_{i=1}^n X_i\}$ satisfies the LDP in B with speed $1/n$ if and only if

$$\left\{ \frac{1}{n} \sum_{i=1}^n f(X_i) : f \in B_1^* \right\}$$

satisfies the LDP in $l_\infty(B_1^*)$ with speed $1/n$. Here $l_\infty(B_1^*)$ denotes the set of bounded functions in B_1^* with the norm $\|x\| := \sup_{t \in B_1^*} |x(t)|$, $x \in l_\infty(B_1^*)$. Arcones (2001) gave a formula for the rate function on $l_\infty(B_1^*)$: for $x \in l_\infty(B_1^*)$ he obtained

$$I(x) = \sup\{I_{f_1,\dots,f_m}(x(f_1),\dots,x(f_m)) : f_1,\dots,f_m \in B_1^*, m \ge 1\},$$

where

$$I_{f_1,\dots,f_m}(u_1,\dots,u_m) \quad = \sup\Big\{ \textstyle\sum_{j=1}^m \lambda_j u_j - \\ \log\Big(E\big(\exp(\textstyle\sum_{j=1}^m \lambda_j f_j(X_1))\big)\Big) : \lambda_1,\dots,\lambda_m \in \mathbb{R}\Big\}.$$

For notational reasons let us consider only the case $m = 2$ in detail. Using the notation $\pi_1 h(x) := E_Y h(x,Y) - Eh(X,Y)$ and $\pi_2 h(x,y) := h(x,y) - E_Y h(x,Y) - E_X h(X,y) + Eh(X,Y)$, where E_X (resp. E_Y) indicates expectations with respect to X (resp. Y) only, we recall *Hoeffding's decomposition*

$$\sum_{i,j=1}^n h(X_i,X_j) = n^2 Eh(X,Y) + 2n \sum_{i=1}^n \pi_1 h(X_i) + \sum_{i,j=1}^n \pi_2 h(X_i,X_j)$$

as well as

$$\textstyle\sum_{1\le i<j\le n} h(X_i,X_j) \quad = \sum_{1\le i<j\le n} \pi_2 h(X_i,X_j) + \\ (n-1)\sum_{i=1}^n \pi_1 h(X_i) + \binom{n}{2} Eh(X,Y).$$

The identities hold a.s. for $\mu^{\otimes 2}$. The kernel $\pi_2 h$ is *canonical (degenerate)* for the law of X meaning that $E_X \pi_2 h(X,Y) = E_Y \pi_2 h(X,Y) = 0$ a.s. Note that $\pi_1 h(X)$ is centered. If $\pi_2 h \in L_2(\mu^{\otimes 2})$, then there exists an orthonormal sequence $\{\psi_r\}_{r=1}^\infty$ of centered square integrable functions and a sequence of real constants $\{c_r\}_{r=1}^{\infty}$ such that

$$\pi_2 h(x,y) = \sum_{r=1}^{\infty} c_r\, \psi_r(x)\, \psi_r(y)$$

with convergence in the sense of $L_2(\mu^{\otimes 2})$. Note that the sequence of eigenvalues $\{c_r\}_{r=1}^{\infty}$ is square-summable, see (Lee 1990, Section 3.2.2, Theorem 1). We take $c_0 := 1$ and $\psi_0 := 2\,\pi_1 h - Eh(X,Y)$. Then

$$V_n(h) = \frac{1}{n}\sum_{i=1}^{n}\psi_0(X_i) + \sum_{r=1}^{\infty} c_r \left(\frac{1}{n}\sum_{i=1}^{n}\psi_r(X_i)\right)^2 \tag{2.2}$$

and

$$U_n(h) = \frac{1}{n}\sum_{i=1}^{n}\psi_0(X_i) + \sum_{r=1}^{\infty} c_r \left(\frac{1}{n}\sum_{i=1}^{n}\psi_r(X_i)\right)^2 - \sum_{r=1}^{\infty} c_r \frac{1}{n^2}\sum_{i=1}^{n}\psi_r(X_i)^2. \tag{2.3}$$

In what follows we will assume that h has a representation such that $V_n(h)$ and $U_n(h)$, respectively, can be represented as in (2.2) and (2.3), respectively. Denote by

$$l_2 = \left\{ x = (x_p)_{p=0}^{\infty},\ x_p \in \mathbb{R},\ \sum_{p=0}^{\infty} x_p^2 < \infty \right\}$$

and let l_2^* be the dual space, which is l_2. Let us denote by $\|\cdot\|_2$ the norm in l_2. Now we consider the following conditions:

Condition 1: There exists a $\lambda > 0$ such that

$$E\left[\exp\bigl(\lambda(\sum_{r=0}^{\infty} |c_r| (\psi_r(X_1))^2)^{1/2}\bigr)\right] < \infty.$$

Condition 2: For each $\lambda > 0$ there exists a $\eta > 0$ such that

$$E\bigl[\exp(\lambda F^{\eta})\bigr] < \infty$$

with

$$F^{\eta} := \sup_{f_1, f_2 \in l_2 : \|f_i\|_2 \le 1, d(f_1,f_2) \le \eta} \left|(f_1 - f_2)\bigl(\bigl(|c_r|^{1/2}\psi_r(X_i)\bigr)_{r=0}^{\infty}\bigr)\right|$$

and $d(f_1,f_2) := E\bigl|\sum_{r=0}^{\infty}(f_1 - f_2)(e_r)|c_r|^{1/2}\psi_r(X_i)\bigr|$.
Here $e_r = (0,\ldots,0,1,0,\ldots) \in l_2$. Now we obtain the following results:

Theorem 2.4. *(LDP for von Mises–statistics) The sequence $\{V_n(h)\}_{n\in\mathbb{N}}$ satisfies a LDP with a good rate function $I(\cdot)$ if $h : S^2 \to \mathbb{R}$ satisfies Condition 1 and Condition 2. The rate function has the form*

$$I(x) = \inf\bigl\{J(y) : y \in l_2, y_0 + \sum_{r=1}^{\infty}\mathrm{sign}(c_r) y_r^2 = x\bigr\}, \tag{2.5}$$

where for every $y \in l_2$, $J(y)$ is the rate function of the sequence

$$\Big(|c_r|^{1/2}\frac{1}{n}\sum_{i=1}^n \psi_r(X_i)\Big)_{r=0}^\infty \in l_2.$$

Remark 2.6. If the strong Cramér condition is fulfilled for h, that is for every $\lambda > 0$

$$E\Big[\exp\big(\lambda\big(\sum_{r=0}^\infty |c_r|\psi_r^2(X_i)\big)^{1/2}\big)\Big] < \infty.$$

then one obtains the well known representation

$$J(y) := \sup_{\xi\in l_2}\Big\{\sum_{r=0}^\infty \xi_r\, y_r - \log E\exp\sum_{r=0}^\infty \xi_r c_r \psi_r(X_1)\Big\},$$

see the discussion in Arcones (1992).

Theorem 2.7. *(LDP for U-statistics) If h satisfies Condition 1 and Condition 2 and* (1.1)*, then the sequence $\{U_n(h)\}_{n\in\mathbb{N}}$ satisfies the LDP with good rate function $I(\cdot)$ given by* (2.5).

Our theorems are improvements of all LDP results for U-statistics and von Mises–statistics, that is Theorem 1 in Arcones (1992), Theorem 2 and Corollary 3 in Eichelsbacher (1997) and Theorems (1.7), (1.10) and (1.13) in Eichelsbacher and Schmock (2002).

Remark 2.8. If h satisfies Condition 1, its diagonal does not satisfy the weak Cramér condition (1.1) in any case. We will give a counterexample. Let Y_k be i.i.d. random variables with $N(0,k^2)$-distribution. Independently of this sequence let Z be integer-valued with $P(Z=k)=ce^{-k^2}$, where c is chosen such that the distribution of Z will become a probability distribution. Let X be the l_2-valued random variable with components $X_k = Y_k$, if $Z=k$, and 0 otherwise. Let $c_r = 1$ and $\psi_r(X) = X_r$, thus we consider the case $S = l_2$. Now we obtain that

$$\sum_{r=0}^\infty \psi_r(X)\,\psi_r(X) = Y_Z^2,$$

and therefore we get

$$E\big[\exp\big(\lambda\big(\sum_{r=0}^\infty \psi_r(X)\,\psi_r(X)\big)^{1/2}\big)\big] = \sum_{r=0}^\infty E\big[e^{\lambda Y_k}, Z=k\big].$$

After a little calculation we obtain

$$\sum_{r=0}^\infty E\big[e^{\lambda Y_k}, Z=k\big] = c\sum_{r=0}^\infty e^{\frac12\lambda^2k^2-k^2} \le c\sum_{r=0}^\infty e^{-\frac12 k^2} < \infty$$

for $\lambda \leq 1$. But for any given $\lambda > 0$ we can take m such that $\lambda > \frac{1}{2}m^2$ and find

$$E\big[e^{\lambda Y_Z^2}\big] \geq E\big[e^{\lambda Y_m^2}, Z = m\big] = \infty.$$

Thus we have found an example, where Condition 1 is fulfilled, but not Condition (1.1).

Proof of Theorem 2.4: By the assumptions of the Theorem we obtain that

$$X(n) := \left(|c_r|^{1/2}\frac{1}{n}\sum_{i=1}^{n}\psi_r(X_i)\right)_{r=0}^{\infty} \in l_2.$$

Since l_2 is a real separable Banach space, we obtain by Arcones (2001) result that the sequence $\{X(n)\}_{n\in\mathbb{N}}$ satisfies a LDP *if and only if* Condition 1 and Condition 2 are fulfilled. Given the continuous function $F_h(y) : l_2 \to \mathbb{R}$, defined by

$$F_h(y) := y_0 + \sum_{r=1}^{\infty} \operatorname{sign}(c_r) y_r^2,$$

we get the representation

$$V_n(h) = F_h(X(n)).$$

Using the contraction principle (Theorem 4.2.1, Dembo and Zeitouni (1998)), the theorem follows immediately. Q.e.d.

Proof of Theorem 2.7: The only thing we have to prove is that Condition (1.1) for h implies that the sequences $\{U_n(h)\}_{n\in\mathbb{N}}$ and $\{V_n(h)\}_{n\in\mathbb{N}}$ have identical large deviations behavior, that is, for every $\delta > 0$ we obtain

$$\limsup_{n\to\infty} \frac{1}{n}\log P\Big(\|U_n(h) - V_n(h)\| \geq \delta\Big) = -\infty. \tag{2.9}$$

Notice that Condition (1.1) implies that

$$\int_{S^2} \exp\big(\lambda\|h(x,x)\|\big)\, d\mu < \infty \quad \text{for some} \quad \lambda > 0.$$

This condition suffices to get (2.9). This is proved more generally for kernel functions with $m \in \mathbb{N}$ variables in Eichelsbacher and Schmock (2002), proof of Theorem (1.10), applying Baranyai's result on the factorization of complete uniform hypergraphs. Q.e.d.

Example 2.10. Let us 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

$$\pi_1 h(x) = \frac{1}{2}\big((x - E(X))^2 - \operatorname{Var}(X)\big),$$

where $\operatorname{Var}(X)$ denotes the variance of X under μ. Moreover $\pi_2 h(x, y) = -(x - E(X))(y - E(X))$. Thus $\psi_0(x) = \big((x - E(X))^2 - \operatorname{Var}(X)\big) - Eh(X, Y)$, $c_1 = -1$

and $\psi_1(x) = x - E(X_1)$. $\psi_i = 0$ for $i \geq 2$. Let us assume that X_i is $N(0,1)$ distributed. Then the strong Cramér condition does not hold for the variance-kernel. But it is easily seen that Condition 1 is fulfilled. Remark that Condition 2 is fulfilled since every $f \in l_2$ is a linear and continuous function and

$$f(x) \leq \|f(x)\| \leq \beta \|x\|$$

for some $\beta > 0$. Using the decomposition of h we obtain Condition 2.

The theorems in this section can be formulated and proved for von Mises–statistics and U-statistics, respectively, of arbitrary degree $m \geq 2$. Therefore we need to describe the spectral decomposition of a kernel function $h : S^m \to \mathbb{R}$ and its Hoeffding-decomposition. Denote $\theta := Eh(X_1, \dots, X_m)$ and define the *canonical* functions $h_c : S^c \to \mathbb{R}$ by the formula

$$h_c(x_1, \dots, x_c) := \int_{S^c} h(y_1, \dots, y_m) \prod_{s=1}^{c} (\delta_{x_s} - \mu(dy_s)) \prod_{s=c+1}^{m} \mu(dy_s).$$

According to the canonical Hoeffding decomposition we have the representation

$$V_n(h) = \theta + \sum_{c=1}^{m} \binom{m}{c} V_n(h_c).$$

We only state the spectral decomposition in the case $m = 3$. It is straightforward to adapt the technique for an arbitrary degree m, but we omit this. We already have seen that we can find a sequence $\{\psi_{r_1}^{(2)}\}_{r_1=0}^{\infty}$ of centered square integrable functions and a sequence of constants $c_{r_1}^{(2)}$ such that

$$h_2(x, y) = \sum_{r_1=1}^{\infty} c_{r_1}^{(2)} \psi_{r_1}^{(2)}(x) \psi_{r_1}^{(2)}(y).$$

Now we proceed iteratively to write for $h_3(x_1, x_2, x_3)$ an expansion in L_2 like

$$h_3(x_1, x_2, x_3) = \sum_{r_1, r_2=1}^{\infty} c_{r_1}^{(3)} c_{r_1,r_2}^{(3)} \psi_{r_1}^{(3)}(x_1) \psi_{r_1,r_2}^{(3)}(x_2) \psi_{r_1,r_2}^{(3)}(x_3),$$

where the functions are centered square integrable. The third term in the Hoeffding decomposition can be represented as

$$\sum_{r_1, r_2=1}^{\infty} \operatorname{sign}(c_{r_1,r_2}^{(3)}) \, \Psi_{n,r_1}^{(3)} \left(\Psi_{n,r_1,r_2}^{(3)}\right)^2$$

with

$$\Psi_{n,r_1}^{(3)} = c_{r_1}^{(3)} \frac{1}{n} \sum_{j=1}^{n} \psi_{r_1}^{(3)}(X_j)$$

and

$$\Psi_{n,r_1,r_2}^{(3)} = |c_{r_1,r_2}^{(3)}|^{1/2} \frac{1}{n} \sum_{j=1}^{n} \psi_{r_1,r_2}^{(3)}(X_j).$$

Now we can put these sample-means in a vector

$$Y(n) = \left(c^{(3)}_{r_1} \frac{1}{n} \sum_{j=1}^{n} \psi^{(3)}_{r_1}(X_j), |c^{(3)}_{r_1,r_2}|^{1/2} \frac{1}{n} \sum_{j=1}^{n} \psi^{(3)}_{r_1,r_2}(X_j) \right)$$

and proceed as in the case $m = 2$ representing any event like $\{V_n(h) \in A\}$ in the form $\{Y(n) \in F_h(A)\}$ with an appropriate function F_h.
Since (2.9) is true for any degree $m \in \mathbb{N}$, we obtain Theorems 2.4 and 2.7 for an arbitrary degree $m \in \mathbb{N}$ using the spectral decomposition just described.

3. The Banach space case

Let $(B, \|\cdot\|)$ be an arbitrary real separable Banach space and let B^* be the dual space. For B-valued U-statistics we obtain the following result:

Theorem 3.1. *(LDP for Banach space valued statistics) Given a measurable, symmetric kernel function $h : S^2 \to B$, assume that for any $f \in B^*$ the real valued function $f \circ h$ satisfies Condition 1 and Condition 2 and Condition* (1.1) *(considering a decomposition for $f \circ h$ as in Section 2 for every $f \in B^*$). Moreover assume that for any $\lambda > 0$ there exists a $\eta > 0$ such that*

$$E[\exp(\lambda Y^\eta)] < \infty, \tag{3.2}$$

where we define

$$Y^\eta := \sup\big\{ |f_1(h(X,Y)) - f_2(h(X,Y))| : f_1, f_2 \in B^*, \|f_i\| \le 1, \\ E|f_1(h(X,Y)) - f_2(h(X,Y))| \le \eta \big\}.$$

Then the sequence $\{U_n(h)\}_{n\in\mathbb{N}}$ satisfies the LDP with a good rate function, that is the sequence $\{U_n(f \circ h), f \in T\}_{n\in\mathbb{N}}$ satisfies the LDP in $l_\infty(T)$ with a good rate function, with $T := \{f \in B^* : \|f\| \le 1\}$.

For the proof of this Theorem we will use the following result of Arcones (2001), Theorem 2.1: Let $\{X_n(t) : t \in T\}$ be a sequence of stochastic processes, where T is an index set. Let $\{\varepsilon_n\}$ be a sequence of positive numbers that converges to zero. Let $I : l_\infty(T) \to [0, \infty]$ and let $I_{t_1,\dots,t_k} : \mathbb{R}^k \to [0, \infty]$ be a function for each $t_1, \dots, t_k \in T$. Let d be a pseudo-metric in T. Consider the following conditions:
(1): (T, d) is totally bounded.
(2): For each $t_1, \dots, t_k \in T$, the vector $(X_n(t_1), \dots, X_n(t_k))$ satisfies the LDP with speed ε_n and good rate function $I_{t_1,\dots,t_k}$.
(3): For each $\tau > 0$

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \varepsilon_n \log P^* \Big\{ \sup_{d(s,t)\le\eta} |X_n(t) - X_n(s)| \ge \tau \Big\} = -\infty.$$

Then for each $0 < \alpha < \infty$ the set $\{z \in l_\infty(T) : I(z) \le \alpha\}$ is a compact set in $l_\infty(T)$, where

$$I(z) = \sup\big\{ I_{t_1,\dots,t_k}(z(t_1), \dots, z(t_k)) : t_1, \dots, t_k \in T, k \ge 1 \big\}.$$

Moreover one gets the upper and lower bound in the LDP with respect to outer and inner probabilities (because of the lack of measurability): for each $A \subset l_\infty(T)$,

$$\begin{aligned} -\inf_{z \in \mathrm{int}(A)} I(z) &\leq \liminf_{n\to\infty} \varepsilon_n \log P_*\{X_n \in A\} \\ &\leq \limsup_{n\to\infty} \varepsilon_n \log P^*\{X_n \in A\} \leq -\inf_{z\in \mathrm{cl}(A)} I(z). \end{aligned}$$

Proof of Theorem 3.1: We will check conditions (1)-(3) of the general LDP approach for stochastic processes due to Arcones (2001): The pseudo-metric on $T = \{f \in B^* : \|f\| \leq 1\}$ is chosen to be the L_1-pseudo-metric $d(f_1, f_2) := E\big(|f_1(h)(X,Y) - f_2(h)(X,Y)|\big)$. Condition 1 implies that $E\big(\|h(X,Y)\|\big) < \infty$. Given $\varepsilon > 0$ there exists a simple (measurable, finitely valued) symmetric kernel $\tilde{h}$ such that $E\big(\|h(X,Y) - \tilde{h}(X,Y)\|\big) < \varepsilon$. It follows that $\{f(\tilde{h}(X,Y)) : f \in B^*, \|f\| \leq 1\}$ is totally bounded in L_1 and therefore $\{f(h(X,Y)) : f \in B^*, \|f\| \leq 1\}$ is totally bounded in L_1.

We will check condition (2): for each $f_1, \ldots, f_k \in T$, the vector $(V_n(f_1(h)), \ldots, V_n(f_k(h)))$ satisfies the LDP with speed $1/n$ and a good rate function. Notice that for every $k \in \mathbb{N}$ the product space $(l_2)^k$ is a real separable Banach space. Consider for each $f_1 \circ h, \ldots, f_k \circ h$ the same expansion as in Section 2, that is

$$\pi_2 f_l \circ h(x,y) = \sum_{r=1}^\infty c_{r,f_l} \psi_{r,f_l}(x) \psi_{r,f_l}(y),$$

$c_{0,f_l} = 1$, $\psi_{0,f_l} = 2\pi_1 f_l \circ h - E f_l h(X,Y)$, where we skip the dependence of h in the notation. With

$$X^k(n) := \left(\left(|c_{r,f_l}|^{1/2} \frac{1}{n} \sum_{i=1}^n \psi_{r,f_l}(X_i) \right)_{r=0}^\infty \right)_{l=1}^k \in (l_2)^k,$$

we obtain that the sequence $\{X^k(n)\}_{n\in\mathbb{N}}$ satisfies the LDP, if Condition 1 and Condition 2 are fulfilled for every $f_l \circ h$, $l = 1, \ldots, k$. Therefore notice that every continuous, linear map on $(l_2)^k$ is described by an element $z = (z_1, \ldots, z_k) \in (l_2)^k$ via $f(x_1, \ldots, x_k) = \sum_{\nu=1}^k \big(\sum_{l=1}^\infty x_{\nu l}\, z_{\nu l}\big)$ for every $x = (x_1, \ldots, x_k) \in (l_2)^k$ with the notation $x_\nu = (x_{\nu r})_{r=0}^\infty \in l_2$. Applying the result of Arcones we obtain the LDP for the sequence $\{X^k(n)\}_{n\in\mathbb{N}}$ if Condition 1 and Condition 2 are fulfilled for f in the dual of $(l_2)^k$. Using the representation of each such f, both conditions follow by applying Hölder's inequality and Condition 1 and 2, given for each $f_l \circ h$, $l = 1, \ldots, k$.

Let us define the continuous function $F_h^k : (l_2)^k \to \mathbb{R}^k$ by

$$F_h^k(y_1, \ldots, y_k) := \left(y_{l0} + \sum_{r=1}^\infty \mathrm{sign}(c_{r,f_l}) y_{lr}^2 \right)_{l=1}^k$$

with the notation $y_l = (y_{lr})_{r=0}^\infty$, we obtain the representation

$$(V_n(f_1(h)), \ldots, V_n(f_k(h))) = F_h^k(X^k(n)).$$

Using the contraction principle, the LDP for the vector follows. We obtain that the vector $(U_n(f_1(h)), \dots, U_n(f_k(h)))$ satisfies the LDP with the same rate function: for $F_k := (f_1 \circ h, \dots, f_k \circ h)$ we get

$$\int_{S^2} \exp\bigl(\lambda \|F_k(x,x)\|\bigr)\, d\mu^{\otimes 2} < \infty$$

for some $\lambda > 0$, choosing the norm $\|\cdot\|$ on $\mathbb{R}^k$ defined by $\|(x_1, \dots, x_k)\| = |x_1| + \cdots + |x_k|$ and applying Hölder's inequality and Condition (1.1) for every $f_l \circ h$. With the proof of Theorem (1.10) in Eichelsbacher and Schmock (2002) we obtain the LDP for $(U_n(f_1(h)), \dots, U_n(f_k(h)))$ as in the proof of Theorem 2.7.
In the last step we will check that

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log P\Bigl\{ \sup_{d(f_1,f_2) \le \eta} |U_n(f_1(h)) - U_n(f_2(h))| \ge \tau \Bigr\} = -\infty \tag{3.3}$$

for each $\tau > 0$. This is a bit involved.
Set

$$h_\eta := \sup_{d(f_1,f_2) \le \eta} |f_1(h) - f_2(h)|.$$

Note that

$$\sup_{d(f_1,f_2) \le \eta} |U_n(f_1(h)) - U_n(f_2(h))| \le U_n(h_\eta). \tag{3.4}$$

Hence by assumption (3.2), for any $\lambda > 0$,

$$\Psi_{h_\eta}(\lambda) := E \exp\bigl(\lambda\, h_\eta(X_1, X_2)\bigr) = E \exp\bigl(\lambda\, Y^\eta\bigr) < \infty,$$

for all $\eta > 0$ sufficiently small. Moreover, with $k := \lfloor n/2 \rfloor$, we obtain that

$$E \exp\bigl(\lambda\, k\, U_n(h_\eta)\bigr) \le \Psi^k_{h_\eta}(\lambda)$$

(Hoeffdings's inequality, see for example Lemma C, page 200, in Serfling (1980)).
Further, since

$$h_\eta(X_1, X_2) \to_P 0, \;\text{ as } \eta \to 0,$$

for any choice of $\tau > 0$, we have for all $\eta > 0$ sufficiently small both

$$\theta(\eta) := E\bigl(h_\eta(X_1, X_2)\bigr) < \tau/2$$

and

$$\begin{aligned} \Psi_{h_\eta}(\lambda)\, e^{-\lambda \theta(\eta)} &= E \exp\bigl(\lambda\bigl(h_\eta(X_1,X_2) - \theta(\eta)\bigr)\bigr) \\ &\le 1 + \frac{\lambda^2}{2} E\Bigl(\bigl(h_\eta(X_1,X_2) - \theta(\eta)\bigr)^2 \exp\bigl(\lambda |h_\eta(X_1,X_2) - \theta(\eta)|\bigr)\Bigr) \\ &\le 1 + \frac{\lambda \tau}{4}. \end{aligned}$$

Thus

$$\begin{aligned} P\Bigl\{U_n(h_\eta) \ge \tau\Bigr\} &\le P\Bigl\{U_n(h_\eta) - \theta(\eta) \ge \frac{\tau}{2}\Bigr\} \\ &\le e^{-k\lambda\tau/2}\, E \exp\Bigl(k\, \lambda \bigl(U_n(h_\eta) - \theta(\eta)\bigr)\Bigr) \end{aligned}$$

$$\leq e^{-k\,\lambda\,\tau/2}\Big(\Psi_{h_\eta}(\lambda)\,e^{-\lambda\theta(\eta)}\Big)^k \leq e^{-k\lambda\tau/2}\Big(1+\frac{\lambda\tau}{4}\Big)^k \leq e^{-k\lambda\tau/4}.$$

From this bound we immediately get that for each choice of $\tau > 0$ and any $\lambda > 0$

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log P\Big\{U_n(h_\eta) > \tau\Big\} \leq -\frac{\lambda\tau}{8},$$

which by the arbitrary choice of λ and the inequality in (3.4) implies (3.3). Q.e.d.

Remark 3.5. Theorem 3.1 holds for any degree $m \geq 2$. This follows using the spectral decomposition introduced in Section 2 as well as the facts that (2.9) is proved for arbitrary degree in Eichelsbacher and Schmock (2002) and that the decomposition arguments using Baranyai's result is applicable for any degree $m \geq 2$, see Eichelsbacher and Schmock (2002).

Acknowledgment. I would like to thank Prof. D. Mason for fruitful discussions concerning an easier proof of (3.3).

References

Arcones, M. (2001). The large deviation principle for empirical processes. To appear in the proceedings in "High Dimensional Probability III".

Arcones, M. A. (1992). Large deviations for U-statistics. *J. Multivariate Anal. 42*(2), 299–301.

Dembo, A. and O. Zeitouni (1998). *Large Deviations Techniques and Applications*. New York: Springer.

Donsker, M. and S. Varadhan (1976). Asymptotic evaluation of certain Markov process expectations for large time III. *Comm. Pure Appl. Math. 29*, 389–461.

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

Eichelsbacher, P. and U. Schmock (2002). Large deviations for products of empirical measures in strong topologies and applications. *Ann. Inst. H. Poincaré. Probab. Statist. 38*(5), 779–797.

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

Lee, A. (1990). *U-Statistics: Theory and Practice*. New York: Marcel Dekker.

Schied, A. (1998). Cramer's condition and Sanov's theorem. *Statist. Probab. Lett. 39*(1), 55–60.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

Sethuraman, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist. 35*, 1304–1316.

(P. Eichelsbacher) Fakultät für Mathematik, Ruhr–Universität Bochum, NA 3/68, D-44780 Bochum, Germany, e-mail: peter.eichelsbacher@ruhr-uni-bochum.de