Some Aspects in Large Deviations

Peter Eichelsbacher
Sonderforschungsbereich 343
Diskrete Strukturen der Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld, Germany

Abstract

Let V be a topological space and $\mathcal{B}(V)$ the Borel σ–field on V. A sequence of probability measures $(P_n)_n$ satisfies a large deviation principle (LDP) with rate function $I(\cdot)$ if $P_n(A)$ can be approximated by $\exp(-n \inf_{x \in A} I(x))$ for appropriate events A in $\mathcal{B}(V)$. In the proof of the LDP for m–variate von Mises–statistics and U–statistics combinatorial factorization arguments can be dropped. In general, the governing rate function is not convex. Nevertheless, we are able to give a representation analogously to the convex conjugate concept. Moreover, we analyse a convex minorant of the rate function.

Short running title: Large deviations

1 Introduction

Let V be a topological space and $\mathcal{B}(V)$ the Borel σ–field on V. A sequence of probability measures $(P_n)_n$ satisfies a large deviation principle (LDP) with rate function $I(\cdot)$ if $P_n(A)$ can be approximated by $\exp(-n \inf_{x\in A} I(x))$ for appropriate events A in $\mathcal{B}(V)$. The following definition makes this more precise:

A sequence of probability measures $(P_n)_n$ on $(V, \mathcal{B}(V))$ obeys a large deviation principle (LDP) with speed a_n and rate function I if there exists a sequence $(a_n)_n$ of positive numbers tending to infinity and a function $I : V \to [0, \infty]$ such that the following conditions hold:

1. $I(x)$ is a lower semicontinuous function. $I \not\equiv 0$.
2. $\{x \ : \ I(x) \leq C\} \subseteq V$ is compact for all $C < +\infty$ (compact level sets).
3. For every open set $O \subseteq V$
$$\liminf_{n\to\infty} \frac{1}{a_n} \log P_n(O) \geq - \inf_{x\in O} I(x). \tag{1}$$
4. For every closed set $A \subseteq V$
$$\limsup_{n\to\infty} \frac{1}{a_n} \log P_n(A) \leq - \inf_{x\in A} I(x). \tag{2}$$

A sequence $(X_n)_n$ of random variables with values in V is said to obey a LDP with rate function I satisfying 1. and 2. if the corresponding distributions satisfy 3. and 4.. In this case we will shortly write
$$P(X_n \sim a) \approx e^{-a_n\, I(a)}.$$

Throughout this paper $(X, \mathcal{B}(X), \pi)$ denotes a probability space and, except otherwise specified, $(\Omega, \mathcal{F}, P)$ is its $\mathbb{N}$-fold product. Define $(X_n)_{n\geq 1}$ to be a sequence of i.i.d. random variables with values in X and distribution (law) π, i.e. for every i $P(X_i \in A) = \pi(A)$. Let $h : X^m \to \mathbb{R}^d$ be a measurable and (without loss of generality) symmetric function. The function
$$U_n := U_n(h, X_1, \dots, X_n) := \frac{1}{\binom{n}{m}} \sum_{1\leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}), \quad m \leq n \tag{3}$$
is called U-statistic of degree m with kernel h.
$$V_n := V_n(h, X_1, \dots, X_n) := \frac{1}{n^m} \sum_{1\leq i_1, \dots, i_m \leq n} h(X_{i_1}, \dots, X_{i_m}) \tag{4}$$

is known as the V-statistic or von Mises-statistic (of degree m with kernel h).

After their development by von Mises [18] and Hoeffding [14], U-statistics as well as von Mises-statistics appeared naturally as a kind of generalized mean value in the theory of statistics (where U_n can be considered as the unbiased variant of V_n).

The LDP for such statistics is proved in [10]. Having in mind a technique in [8] we decomposed U_n into an average of dependent averages of i.i.d. random variables using a factorization result for hypergraphs due to Baranyai. We simplify this argument in section 2.

Section 3 represents a convex upper bound for large deviations. It should be remarked that an earlier paper of de Acosta [5] contains a related aspect.

In section 4 we show that the free energy function of U_n

$$C(t) = \lim_{n\to\infty} \frac{1}{n} \log E(\exp(n\langle t, U_n\rangle))$$

exists and is finite, where $\langle\cdot,\cdot\rangle$ is used to denote the Euclidean inner product in $\mathbb{R}^d$. Its Legendre–transform

$$\sup_{t\in\mathbb{R}^d}\{\langle x, t\rangle - C(t)\}$$

is a convex minorant of the LDP rate function. Moreover we present an example where the rate is not convex.

The contents of section 5 may be viewed as a contribution to large deviation theory, especially to the representation of the rate function analogously to the convex conjugate concept. It relies on the asymptotic value method of Bryc [3] and a paper of Dinwoodie [7]. We can apply our identification to the U–statistic case.

2 A Large deviation theorem

First we will use the repesentation of a U–statistic as an average of dependent averages of i.i.d. random variables due to Hoeffding [15] to get an upper bound for the moment generating function of a U–statistic. This representation replaces the Baranyai factorization in our proof of the LDP [10].

Lemma 2.1 *(Hoeffding 1963) Consider a U–statistic with kernel h, taking values in an arbitrary vector space. Let $k := \left[\frac{n}{m}\right]$ be the greatest integer $\leq \frac{n}{m}$ and*

$$B(x_1, \dots, x_n) := \frac{1}{k}\left[h(x_1, \dots, x_m) + h(x_{m+1}, \dots, x_{2m}) + \dots + h(x_{km-m+1}, \dots, x_{km})\right].$$

Then

$$U_n = \frac{1}{n!}\sum_{per} B(X_{i_1}, \dots, X_{i_n}),$$

where $\sum\limits_{per}$ *denotes the summation over all* $n!$ *permutations of* $\{i_1, \ldots, i_n\}$.

Proof: Denote by $\sum\limits_{\text{bin}}$ the summation over all $\binom{n}{m}$ combinations $\{i_1, \ldots, i_m\}$ from $\{1, \ldots, n\}$, then

$$k \sum_{\text{per}} B(X_{i_1}, \ldots, X_{i_n}) = k\, m!\, (n-m)! \sum_{\text{bin}} h(X_{i_1}, \ldots, X_{i_m}) = k\, m!\, (n-m)! \binom{n}{m} U_n.$$

□

We can use this representation as follows:

Lemma 2.2 *Let* U_n *be a* $\mathbb{R}^d$*–valued U–statistic and denote by* $||\cdot||$ *the norm on* $\mathbb{R}^d$*. Let* $M_t(h) := E(\exp(t||h(x_1, \ldots, x_m)||))$*, where* E *denotes the expected value with respect to the probability* P*, then*

$$M_t(U_n) := E(\exp(t||U_n||)) \leq (M_{\frac{t}{k}}(h))^k$$

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

Proof:

$$\exp(t||U_n||) = \exp(t||(n!)^{-1} \sum_{\text{per}} B(X_{i_1}, \ldots, X_{i_n})||) \leq (n!)^{-1} \sum_{\text{per}} \exp(t||B(X_{i_1}, \ldots, X_{i_n})||)$$

by Jensen's inequality. Since $B(\cdot)$ is an average of i.i.d. random variables, we get

$$M_t(U_n) \leq E(\exp(t||B(X_{i_1}, \ldots, X_{i_n})||)) = E(\prod_{i=1}^{k} \exp(\frac{t}{k}||h(X_1, \ldots, X_m)||)) = (M_{\frac{t}{k}}(h))^k.$$

□

$\langle\cdot,\cdot\rangle$ is used to denote the Euclidean inner product. With just the same arguments it follows

Corollary 2.3 *Let* U_n *be a* $\mathbb{R}^d$*–valued U–statistic. Then*

$$E(\exp(\langle t, U_n\rangle) \leq \left(E(\exp(\frac{1}{k}\langle t, h\rangle))\right)^k.$$

Now we give a sketch of a new proof of the large deviation theorem for exponentially bounded U–statistics and von Mises–statistics (see also [10]). Many technical arguments are skipped.

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

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

such that $y_j = x_i$ 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 always interprete a function in $\mathcal{H}_r$ as function on X^r rather than a function on X^m.

Theorem 2.4 *Let (X,π) be a Polish probability space and $h : X^m \to \mathbb{R}^d$ be a measurable and symmetric function. $\mathcal{M}^1(X)$ is used to denote the space of all probability measures on X. Assume that for each $r \in \{1,\ldots,m\}$ and for every $g \in \mathcal{H}_r$*

$$\int_{X^r} \exp(\Theta||g(x_1,\ldots,x_r)||)d\pi^r(x_1,\ldots,x_r) < \infty \quad \text{for all } \Theta > 0, \tag{5}$$

where π^r denotes the r-fold product measure with marginal distributions π.
Then, if V_n resp. U_n denotes the associated V-statistics resp. U-statistics with kernel h, for all $a \in \mathbb{R}^d$ we have

$$P(V_n \sim a) \approx e^{-nI(a)} \tag{6}$$

as well as

$$P(U_n \sim a) \approx e^{-nI(a)}. \tag{7}$$

Here

$$I(a) := \inf_{\substack{\varrho\in\mathcal{M}^1(X)\\ F(\varrho)=a}} H(\varrho|\pi), \tag{8}$$

where $H(\varrho|\pi)$ is the entropy function defined in (10) and $F : \mathcal{M}^1(X) \to \mathbb{R}^d$ is given by

$$F(\rho) := \int_{X^m} h(x_1,\ldots,x_m)d\rho^m(x_1,\ldots,x_m). \tag{9}$$

Proof: We will emphazise the new part of the proof and hence will skip technical details. The first step remains unchanged: as a consequence of the comparison principle ([1], Proposition 2), U_n and V_n have identical large deviation behaviour under condition (5). If L_n denotes the empirical measure of the i.i.d. random variables $X_1,\ldots,X_n$,

$$L_n(\omega,B) := \frac{1}{n}\sum_{j=1}^n \delta_{X_j(\omega)}(B)$$

where δ denotes the Dirac–measure, $B \in \mathcal{B}(X)$, then $F(L_n) = V_n$. We will apply a generalized version of the so called contraction principle (see [6], lemma 2.1.4) to Sanov's theorem (see [19]; [6], Theorem 3.2.17) which is the large deviation principle of the distribution of L_n, more precisely: if we endow the space of all probability measures $\mathcal{M}^1(X)$ on X with the weak topology and if $\pi \in \mathcal{M}^1(X)$ is the law of X_1; $P := \bigotimes_{\mathbb{N}} \pi$, then

$(P \circ L_n^{-1})_n$ (sequence in $\mathcal{M}^1(\mathcal{M}^1(X))$) satisfies a LDP with $a_n = n$ and rate $H(\varrho|\pi)$ (the relative entropy) defined by

$$H(\varrho|\pi) := \begin{cases} \int_X f \ \log \ f d\pi & \text{if } \varrho \ll \pi \quad \text{and} \quad f = \frac{d\varrho}{d\pi} \\ \infty & \text{otherwise.} \end{cases} \tag{10}$$

It is a simple exercise that LDP is invariant under a continuous mapping. The main problem is that F is only continuous for bounded and continuous kernels using the fact that $\varrho \longrightarrow \varrho^r$, $\varrho \in \mathcal{M}^1(X)$, is continuous in the weak topology. Thus we truncated the kernel h by $L > 0, L \in \mathbb{R}$ and took a continuous version h_L^c according to Lusin's theorem. Now we had to prove that the continuous mappings

$$F_L(\rho) := \int_{X^m} h_L^c(x_1, \dots, x_m) d\rho^m(x_1, \dots, x_m). \tag{11}$$

converge to the contraction function F in the sense that the two conditions in the contraction principle are satisfied. Checking the first condition that

$$\lim_{L\to\infty} \sup_{\rho\in N} || \int (h - h_L^c) d\rho^m || = 0.$$

for any level set N of $H(\cdot|\pi)$, we adopt the arguments given in [10], which are a direct consequence of an early result of [8]; Lemma 5.1. The remain part is to show that for every $\varepsilon > 0$

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P(||V_n(h - h_L^c)|| > \varepsilon) = -\infty.$$

or equivalently

$$\lim_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log P(||U_n(h - h_L^c)|| > \varepsilon) = -\infty. \tag{12}$$

This is a consequence of our Lemma 2.2. From the exponential boundedness of $h(\cdot)$ (5) it follows that for any non negative $t \in \mathbb{R}$:

$$\begin{aligned} P(||U_n(h - h_L)|| > \varepsilon) &\le \exp(-t\varepsilon) E(\exp(t||U_n(h - h_L)||)) \\ &= \exp(-t\varepsilon) M_t(U_n(||h - h_L||)) \\ &\le \exp(-t\varepsilon)(M_{\frac{t}{k}}(h - h_L))^k. \end{aligned} \tag{13}$$

where $k = \left[\frac{n}{m}\right]$. Therefore substituting t by tk yields

$$P(||U_n(h - h_L)|| > \varepsilon) \le \exp(-tk\varepsilon)(M_t(h - h_L))^k,$$

thus

$$\frac{1}{n} \log P(||U_n(h - h_L)|| > \varepsilon) \le \frac{-tk\varepsilon}{n} + \frac{k}{n} \log M_t(h - h_L),$$

and we conclude

$$\limsup_{n\to\infty}\frac{1}{n}\log P(||U_n(h-h_L)||>\varepsilon)\le\frac{-t\varepsilon}{m}+\frac{1}{m}\log M_t(h-h_L^c).$$

Since $t>0$ has been arbitrary and $\varepsilon>0$ we have

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

We make use of the triangle inequality

$$\begin{aligned}&P(||U_n(h-h_L^c)||>\varepsilon)\le\\&P(U_n(||h-h_L||)>\frac{\varepsilon}{2})+P(U_n(||h_L-h_L^c||)>\frac{\varepsilon}{2})\end{aligned}$$

and treat the second term analogously (see [10] for details). □

Remarks 2.5 *1. Although the τ–topology on the space $\mathcal{M}^1(X)$ of all probability measures on X makes integration with respect to bounded measurable functions continuous, it is unknown whether the map $\varrho\longrightarrow\varrho^r$ for $\varrho\in\mathcal{M}^1(X)$, $r\in\mathbb{N}$, is continuous, if $\mathcal{M}^1(X)$ is endowed with the τ–topology. So we cannot use the fact that the LDP for the distribution of L_n is proved in the τ–topology by Groeneboom et.al. ([13]). Therefore Lusin's theorem is needed.*

2. Generalized statistics as an extension of U–statistics resp. V–statistics to the case of several samples are defined in a straightforward way: consider l independent collections of independent observations with l different distribution functions. The extension of the LDP to generalized statistics works dropping the factorization theorem, too. The reason is that the representation of such a statistic as an average of dependent averages of i.i.d. random variables (as in Lemma 2.1) holds true. For simplicity look at a U–statistic of degree (m_1,m_2) with kernel h, given by

$$U_{n_1,n_2}:=\frac{1}{\binom{n_1}{m_1}\binom{n_2}{m_2}}\sum_c h(X_{i_1},\dots,X_{i_{m_1}},Y_{j_1},\dots,Y_{j_{m_2}})$$

where the summation $\sum_c$ extends over all $1\le i_1<\cdots<i_{m_1}\le n_1$ and $1\le j_1<\cdots<j_{m_2}\le n_2$; $(X_i)_i,(Y_i)_i$ are i.i.d. samples with values in Z_1 resp. Z_2,distributed according to μ_1 resp. μ_2 and $h:Z_1^{m_1}\times Z_2^{m_2}\longrightarrow\mathbb{R}^d$ is a measurable, symmetric function. An upper bound for the moment generating function can be proved as in Lemma 2.2:

Lemma 2.6 *Consider a U–statistic with degree (m_1,m_2) and let $k:=\min\{\left[\frac{n_1}{m_1}\right],\left[\frac{n_2}{m_2}\right]\}$, then*

$$M_t(U_{n_1,n_2})\le(M_{\frac{t}{k}}(h))^k.$$

Proof: Let

$$C(X_1,\dots,X_{n_1},Y_1,\dots,Y_{n_2}) :=$$
$$\frac{1}{k}[h(X_1,\dots,X_{m_1},Y_1,\dots,Y_{m_2}) + h(X_{m_1+1},\dots,X_{2m_1},Y_{m_2+1},\dots,Y_{2m_2})$$
$$+\dots + h(X_{km_1-m_1+1},\dots,Y_{km_2})].$$

Then

$$U_{n_1,n_2} = \frac{1}{n_1!n_2!}\sum_{\text{per}} C(X_{i_1},\dots,X_{i_{n_1}},Y_{j_1},\dots,Y_{j_{n_2}});$$

$\sum_{\text{per}}$ denotes the sum over all $n_1!n_2!$ permutations. The term on the right is an average of i.i.d. random variables, so the inequality follows as in Lemma 2.2. □

3 A convex upper bound

Under the assumption of exponential boundedness we apply the techniques of section 2 to derive a result on an upper bound for the probability of large deviations, where the bound is a convex conjugate of a measurable function, more precisely the so called Cramér functional defined as follows:
Given a function $f:\mathbb{R}^d \longrightarrow \mathbb{R}$, its convex conjugate – also called the Legendre transform– is defined by

$$\Lambda_f(x) := \sup_{\lambda\in\mathbb{R}^d}\{\langle\lambda,x\rangle - f(\lambda)\}$$

and the Cramér functional Λ_f of f is defined by

$$\Lambda_f(B) := \inf_{x\in B}\Lambda_f(x); \qquad B\in\mathcal{B}(\mathbb{R}^d),$$

where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ–algebra on $\mathbb{R}^d$.

Theorem 3.1 *Let U_n be a $\mathbb{R}^d$–valued U–statistic. Assume that for all $\Theta > 0$*

$$\int_{X^m}\exp(\Theta||h(x_1,\dots,x_m)||)d\pi^m(x_1,\dots,x_m) < \infty. \tag{14}$$

Then for every closed set $A\subset\mathbb{R}^d$

$$\limsup_{n\to\infty}\frac{1}{n}\log P(U_n\in A) \le -\Lambda_C(A), \tag{15}$$

where $\Lambda_C(\cdot)$ denotes the Cramér functional of

$$C(\lambda) := C_h(\lambda) := \frac{1}{m}\log E(\exp(m\langle\lambda, h(X_1,\dots,X_m)\rangle)).$$

Remark 3.2 *The LDP for $(P \circ U_n^{-1})_n$ was an application of the i.i.d. LDP for the empirical law. Therefore the deviating behaviour of finite–dimensional vectors is described in terms of a variational formula in infinite dimensions. Here we have a finite–dimensional variational formula. But later we will see that the rate cannot be expressed as a Legendre transform in general.*

The proof of the theorem relies on techniques which are step by step known in the Cramér–Chernoff theory for i.i.d.–means. First we collect some elementary properties of $C(\cdot)$ and $\Lambda_C(\cdot)$.

Lemma 3.3 *For every kernel function $h : X^m \longrightarrow \mathbb{R}^d$ $\Lambda_C(\cdot) \geq 0$, moreover*

1. *$C(\cdot)$ and $\Lambda_C(\cdot)$ are convex and lower–semicontinuous.*
2. *$\Lambda_C(\Theta) = 0$ with $\Theta := E(h) < \infty$.*
3. *Λ_C has compact level sets.*
4. *Let $d = 1$, $\Lambda_C(\cdot)$ is non decreasing on $[\Theta, \infty)$ and non increasing on $(-\infty, \Theta]$. In addition $\Lambda_C(x) = \sup_{\lambda \geq 0}\{\lambda\, x - C(\lambda)\}$ for $x \geq \Theta$ and $\Lambda_C(x) = \sup_{\lambda \leq 0}\{\lambda\, x - C(\lambda)\}$ for $x \leq \Theta$.*

Proof: 1. Convexity of C follows from Hölder's inequality, lower semicontinuity by Fatou's lemma. As the pointwise supremum of linear functions, Λ_C is necessarily lower semi–continuous and convex. Since $\langle \lambda, x\rangle - C(\lambda) = 0$ for $\lambda = 0 (\in \mathbb{R}^d)$ and every $x \in \mathbb{R}^d$, $\Lambda_C \geq 0$.

2. By Jensen's inequality:

$$\frac{1}{m} \log E(\exp(m\langle \lambda, h(X_1, \ldots, X_m)\rangle)) \geq \frac{1}{m} m E(\langle \lambda, h\rangle) = \langle \lambda, \Theta\rangle \tag{16}$$

for all $\lambda \in \mathbb{R}^d$. Thus $\langle \lambda, \Theta\rangle - C(\lambda) \leq 0$ for all $\lambda \in \mathbb{R}^d$ and therefore $\Lambda_C(\Theta) \leq 0$. Since $\Lambda_C \geq 0 \Longrightarrow \Lambda_C(\Theta) = 0$.

3. Consider $K_L := \{x \in \mathbb{R}^d : \Lambda_C(x) \leq L\}$. Since Λ_C is lower–semicontinuous, K_L is closed and for $y \in K_L$

$$\langle z, y\rangle \leq C(z) + \Lambda_C(y) \leq C(z) + L$$

for any $z \in \mathbb{R}^d$, thus for $R > 0$ there exists a suitable A such that

$$\sup_{||z|| \leq R} \langle z, y\rangle \leq \sup_{||z|| \leq R} C(z) + L \leq A < \infty.$$

Hence K_L is bounded.

4. If $d=1$, the monotonicity behaviour is obviously true. Moreover for $x \geq \Theta$ and $\lambda \leq 0$ by (16) $\lambda x - C(\lambda) \leq \lambda x - \lambda\Theta = \lambda(x-\Theta) \leq 0$ and it follows $0 \leq \Lambda_C(x) = \sup_{\lambda \geq 0}\{\lambda x - C(\lambda)\}$. The case $x \leq \Theta$ is done analogously. □

Proof: (of Theorem 3.1) Step 1: Let $d=1$ and $k = \left[\frac{n}{m}\right]$. For $\lambda \geq 0$ we apply Lemma 2.2:

$$P(U_n \in [a,\infty)) \leq \exp(-\lambda a) E(\exp(\lambda U_n)) \leq \exp(-\lambda a)(M_{\frac{\lambda}{k}}(h))^k.$$

Therefore substituting λ by $n\lambda$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \log P(U_n \in [a,\infty)) \leq -\lambda a + \frac{1}{m} \log M_{m\lambda}(h).$$

Now suppose that $a \geq \Theta$, then as a consequence of Lemma 3.3 (4) it follows

$$\limsup_{n\to\infty} \frac{1}{n} \log P(U_n \in [a,\infty)) \leq -\Lambda_C(a) = -\Lambda_C([a,\infty)).$$

Clearly, a similar calculation yields

$$\limsup_{n\to\infty} \frac{1}{n} \log P(U_n \in (-\infty,a]) \leq -\Lambda_C(a)$$

for $a \leq \Theta$. Let $A \subseteq \mathbb{R}$ be closed and $A \cap [\Theta,\infty) \neq \emptyset$ and $A \cap (-\infty,\Theta] \neq \emptyset$, let $i_+ := \inf\{x \geq \Theta : x \in A\}$ and $i_- := \sup\{x \leq \Theta : x \in A\}$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log P(U_n \in A) \leq \lim_{n\to\infty} \frac{1}{n} \log\left(P(U_n \in (-\infty,i_-]) + P(U_n \in [i_+,\infty))\right) \leq -\inf_{x\in A} \Lambda_C(x).$$

Step 2: Now we will prove the statement for vector–valued kernel functions: if $0 < \Lambda_C(A) < \infty$, we can cover A by finitely many closed half–spaces:

$$A \subseteq \bigcup_{i=1}^{k} H(a_i, \Lambda_C(A) - \varepsilon)$$

for any $0 < \varepsilon < \Lambda_C(A)$, where $a_1,\ldots,a_k$ are nonzero points in $\mathbb{R}^d$ and

$$H(a,b) := \{x \in \mathbb{R}^d : \langle a,x\rangle - C(a) \geq b\},$$

$a \in \mathbb{R}^d, b \in \mathbb{R}$.

This is an immediate consequence of Lemma VII 4.1 [12]. Then Chebyshev's inequality implies

$$\begin{aligned} P(U_n \in A) &\leq \sum_{i=1}^{k} P(U_n \in H(a_i, \Lambda_C(A) - \varepsilon)) \\ &= \sum_{i=1}^{k} P(\langle a_i, U_n\rangle \geq C(a_i) + \Lambda_C(A) - \varepsilon) \\ &\leq \sum_{i=1}^{k} \exp(-n(C(a_i) + \Lambda_C(A) - \varepsilon)) E(\exp(n\langle a_i, U_n\rangle)). \qquad (17) \end{aligned}$$

Again applying Corollary 2.3 we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log E(\exp(n\langle a_i, U_n\rangle)) \le \frac{1}{m} \log E(\exp(m\langle a_i, h\rangle)) = C(a_i).$$

Using $\log(a+b) \le \log 2 + \max(\log a, \log b)$ it follows

$$\limsup_{n\to\infty} \frac{1}{n} \log P(U_n \in A) \le -\Lambda_C(A) + \varepsilon,$$

since $0 < \varepsilon < \Lambda(A)$ is arbitrary this yields our result. If $\Lambda(A) = \infty$ Lemma VII 4.1 ([12]) yields a finite covering of A which makes the calculation (17) work. □

U–statistics and V–statistics associated with h are closely related: if condition (5) holds, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log P(||U_n - V_n|| > \varepsilon) = -\infty$$

(see [10]). Now for each $\varepsilon > 0$

$$\frac{1}{n} \log P(V_n \in A) \le \frac{1}{n} \log 2 + \max\left(\frac{1}{n} \log P(U_n \in A^\varepsilon), \frac{1}{n} \log P(||U_n - V_n|| > \varepsilon)\right)$$

where $A^\varepsilon := \{x \in \mathbb{R}^d : d(x, A) < \varepsilon\}, d(x, y) = ||x - y||$ and $d(x, A) := \inf\{d(x, y), y \in A\}$. We obtain as a consequence of (15):

$$\limsup_{n\to\infty} \frac{1}{n} \log P(V_n \in A) \le -\lim_{\varepsilon\searrow 0} \inf_{x\in\overline{A^\varepsilon}} \Lambda_C(x) = -\Lambda_C(A).$$

Using Lemma 3.3, the last equality follows by

Lemma 3.4 *(2.1.2 [6])*
Let $F : \mathbb{R}^d \to [0, \infty]$ be lower–semicontinuous with compact level sets. Then for each closed set $A \subseteq \mathbb{R}^d$:

$$\inf_{x\in A} F(x) = \lim_{\varepsilon\searrow 0} \inf_{x\in A^\varepsilon} F(x).$$

Thus we have proved:

Corollary 3.5 *Let V_n be a $\mathbb{R}^d$–valued V–statistic. Assume condition (5). Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(V_n \in A) \le -\Lambda_C(A)$$

for every closed set $A \subseteq \mathbb{R}^d$.

At last we prove a general result for Banach–valued kernel functions $h : X^m \longrightarrow B$, B a separable real Banach space. Denote by B' the dual space of B and $\langle\cdot,\cdot\rangle$ the dual relation. Then we define analogously the Cramér–functional as

$$\Lambda_C(A) := \inf_{x\in A} \sup_{\lambda\in B'} \{\langle\lambda, x\rangle - C(\lambda)\},$$

where $C(\lambda) = \frac{1}{m} \log E(\exp(m\langle\lambda, h(X_1, \ldots, X_m)\rangle))$.

Theorem 3.6 *Let h be B–valued, measurable and symmetric. Assume the exponential boundedness condition (14) for h. If for every $L > 0$ there exists a compact set $K_L \subset B$ and an $\hat{n} \in \mathbb{N}$ such that*

$$P(U_n \in K_L^c) \leq \exp(-n\,L) \quad \textit{for all} \quad n \geq \hat{n},$$

where $K_L^c := B \backslash K_L$ ($(P \circ U_n^{-1})_n$ is exponentially tight), then

$$\limsup_{n \to \infty} \frac{1}{n} \log P(U_n \in A) \leq -\Lambda_C(A)$$

for every closed subset $A \subset B$.

Proof: As a consequence of Corollary 2.3

$$E(\exp(n\langle\lambda, U_n\rangle)) \leq \left(E(\exp(\frac{n}{k}\langle\lambda, h\rangle))\right)^k, \tag{18}$$

hence we are in the situation of Theorem 2.1 of de Acosta ([5]). He has proved that the upper bounds for large deviations of dependent random vectors depend only on a limiting inequality for the "free–energy" term (18). □

Remarks 3.7 *1. $E(\exp(<\lambda, h>)) < \infty$ is a well known conclusion from the exponential boundedness condition (14).*

2. Theorem 3.1 is included in Theorem 3.6, but our direct proof does not use any abstract elements of large deviations theory as the exponential tightness concept. Nevertheless, if a sequence P_n of probability measures on a Polish space satisfies a LDP, then $(P_n)_n$ is exponentially tight by Lemma 2.6 of Lynch and Sethuraman ([17]). Thus with Theorem 2.4, especially under the stronger condition (5), the conditions of Theorem 3.6 are fulfilled. But in the $\mathbb{R}^d$–case, we do not need to go this detour. Independently of this we can prove the exponantial tightness directly:

Lemma 3.8 *Let U_n be a $\mathbb{R}^d$–valued U-statistic. Then $(P \circ U_n^{-1})_n$ is exponentially tight.*

Proof: Let K_L be the set $\{x \in \mathbb{R}^d : ||x|| \leq L\}$, K_L is compact and

$$\begin{aligned} P(U_n \notin K_L) &= P(||U_n|| > L) \\ &\leq P\left(\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \cdots < i_m \leq n} ||h(X_{i_1}, \ldots, X_{i_m})|| > L\right) \\ &\leq \exp(-n\Lambda_C(L) + \varepsilon) \end{aligned}$$

for n sufficiently large, where $\Lambda_C(\cdot)$ denotes the Cramér–functional on $\mathbb{R}$. The last inequality follows immediatly by Theorem 3.1. Now

$$\limsup_{x\to\infty}\frac{\Lambda_C(x)}{x}\geq\limsup_{x\to\infty}\frac{\lambda x-C(\lambda)}{x}=\lambda$$

for $\lambda>0$ if $C(\lambda)<\infty$ and hence $\lim_{x\to\infty}\Lambda_C(x)=\infty$ and the lemma is proved. □

4 Free energy function of U_n

In this section we gather from the techniques of section 2 the existence of the so called free energy function of U_n (or pressure) defined by

$$C(t):=\lim_{n\to\infty}C_n(t),\quad t\in\mathbb{R}^d$$

where $C_n(t)$ is defined by

$$C_n(t):=\frac{1}{n}\log E(\exp(n\langle t,U_n\rangle)).$$

One of the main applications of large deviation property is to calculate the asymptotics of certain integrals. Varadhan's theorem on the asymptotics of integrals (Laplace method) gives the leading order asymptotic behaviour of a class of integrals. For every bounded and continuous function $F:\mathbb{R}^d\to\mathbb{R}$ the variation of $F(x)-I(x)$ is finite and

$$\lim_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{R}^d}\exp(nF(x))dP\circ U_n^{-1}=\sup_{x\in\mathbb{R}^d}\{F(x)-I(x)\}.$$

Dinwoodie ([7]) represented a generalized version where F is continuous and bounded from above. But this is a direct consequence of the following two standard results (see [6], Lemma 2.1.7 and 2.1.8).

Lemma 4.1 *Suppose that $(P_n)_n$ satisfies a LDP with rate function I. If the function $F:V\to\mathbb{R}$ is lower–semicontinuous, then*

$$\liminf_{n\to\infty}\frac{1}{n}\log\int_V\exp(nF(x))dP_n(x)\geq\sup_{x\in V}\{F(x)-I(x)\};\quad F(x)\wedge I(x)<\infty\}.$$

Lemma 4.2 *Assume that $(P_n)_n$ satisfies a LDP with rate function I. If $F:V\to\mathbb{R}$ is an upper–semicontinuous function which satisfies*

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\left(\int_{x:F(x)\geq L}\exp(nF(x))dP_n\right)=-\infty\tag{19}$$

then

$$\limsup_{n\to\infty}\frac{1}{n}\log\int_V\exp(nF(x))dP_n\leq\sup_{x\in V}\{F(x)-I(x)\}.$$

Theorem 4.3 *Let U_n be a U–statistic with exponentially bounded kernel $h : X^m \to \mathbb{R}^d$ (condition (5)), then $C(t)$ exists and is finite for all $t \in \mathbb{R}^d$. Furthermore*

$$C(t) = \sup_{x \in \mathbb{R}^d} \{\langle t, x\rangle - I(x)\}, \tag{20}$$

where $I(\cdot)$ is the rate of U_n.

Proof: We try to apply Lemma 4.2 for the upper bound. If we can show that

$$\sup_n \left\{ \int \exp(n\alpha\langle t, x\rangle) dP \circ U_n^{-1} \right\}^{\frac{1}{n}} < \infty \tag{21}$$

for some $\alpha \in (1, \infty)$, we will see that the contribution of the integral $\int_{F>L} \exp(nF(x))dP \circ U_n^{-1}$ decreases exponentially fast if the level L grows up to infinity and thus condition (19) in Lemma 4.2 is fulfilled. By Hölder's and Chebyshev's inequality

$$\begin{aligned}
\int_{\langle t,x\rangle \geq L} \exp(n\langle t, x\rangle) dP \circ U_n^{-1} &\leq \left(\int \exp(\alpha n\langle t, x\rangle) dP \circ U_n^{-1} \right)^{\frac{1}{\alpha}} (P(\langle t, U_n\rangle \geq L))^{1-\frac{1}{\alpha}} \\
&\leq \left(\int \exp(\alpha n\langle t, x\rangle) dP \circ U_n^{-1} \right)^{\frac{1}{\alpha}} \left(e^{-\alpha Ln} \int \exp(\alpha n\langle t, x\rangle dP \circ U_n^{-1} \right)^{1-\frac{1}{\alpha}} \\
&= \left(\int \exp(\alpha n\langle t, x\rangle) dP \circ U_n^{-1} \right) \exp[(1-\alpha)Ln],
\end{aligned}$$

so condition (19) holds, when (**??**) can be shown. Now for all $t \in \mathbb{R}^d$:

$$(E(\exp(2n\langle t, U_n\rangle)))^{\frac{1}{n}} \leq (E(\exp(2m\langle t, h\rangle)))^{\frac{1}{m}}$$

by Corollary 2.3. and the Theorem is proved. □

The Legendre transform of the free energy function of U_n is a good candidate for the rate function for $P \circ U_n^{-1}$. If the rate I was convex, I would equal this Legendre transform (see Theorem 2.2.21 [6]) but in the case of U–statistics I usually is not convex:

Example 4.4 $h(x_1, \ldots, x_4) := \prod_{i=1}^4 x_i$ *then the corresponding V–statistic satisfies a LDP (as well as the corresponding U–statistic). The rate function can be easily calculated:*

$$V_n = \frac{1}{n^4} \sum_{i_1, \ldots, i_4 = 1}^{n} X_{i_1} \cdots X_{i_4} = \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^4,$$

hence by contraction principle the rate is

$$I(x) = \inf_{y : F(y) = x} J(y)$$

where $J(\cdot)$ is the rate function of Cramer's theorem given by

$$J(y) := \sup_{\lambda \in \mathbb{R}} \{\lambda y - \log \int \exp(\lambda q) d\pi(q)\},$$

where π denotes the law of X_1 and $F(y) = y^4$. If the X_i are $N(0,1)$–distributed, $J(y) = \frac{y^2}{2}$, hence

$$I(x) = \inf_{y: y^4 = x} \frac{y^2}{2} = \frac{\sqrt{x}}{2}$$

for $x > 0$, which is even concave.

According to the upper bound approach in section 3,

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

is an upper bound which is convex and as a consequence of (20) is majorized by the rate I. Moreover it is optimal in the sense that it is the largest lower–semicontinuous convex function $\leq I$. Hence the example show, that Λ cannot be the rate in general. Summarizing we will prove the following Lemma:

Lemma 4.5 *(optimal convex upper bound) In the situation of Theorem 3.6 the following conclusions hold:*

1. $$\limsup_{n \to \infty} \frac{1}{n} \log P(U_n \in A) \leq - \inf_{x \in A} \Lambda(x)$$

   for every closed set $A \subset B$.

2. *$\Lambda \leq I$, I denotes the rate of U_n.*

3. *If $G : B \to \mathbb{R}$ is an arbitrary lower–semicontinuous convex function satisfying $G \leq I$ than $\Lambda \geq G$. Λ is called a regularization.*

Proof: 1. Repeat the arguments in the proof of Theorem 3.1 and 3.6.
2. By definition of $C(\cdot)$ and $\Lambda(\cdot)$ it follows

$$\Lambda(x) = \sup_{t \in \mathbb{R}^d} \{\langle x, t \rangle - C(t)\} \leq \langle x, t \rangle - \langle x, t \rangle + I(x).$$

3. This is a well known result from convex analysis: $C(\cdot)$ is the Legendre transform of $I(\cdot)$, performing the Legendre transform twice gives $\Lambda(\cdot)$. Thus $\Lambda(\cdot)$ is a regularization of the class of lower–semicontinuous convex functions by Proposition 4.1, [11], Chapter I. $\square$

5 A rate function representation

Although the Legendre transform of the pressure of U_n is not the correct rate function we give a representation of the rate analogously to the convex conjugate concept. We have in mind that a usual approach to large deviation results will consists of two steps: the first being an extential proof the second being an identification of the rate taking into account the fact that a LDP rate is unique (see [6], Lemma 2.1.1).

The concept of an asymptotic value method, due to Bryc [3], contributes to our representation as follows: Bryc proved a new approach to LDP which converses Varadhan's Laplace method. He replaced the main patterns like change of measures by verification of exponential tightness and showing, that the Laplace integral exists for a sufficiently large class of functions, to be more precise:

A sequence $(P_n)_n$ of probability measures on $(X, \mathcal{B}(X))$ admits an asymptotic value over a class $\mathcal{F}$ of measurable functions if

$$L(F) := \lim_{n\to\infty} \frac{1}{n} \log \int_X \exp(nF(x)) dP_n(x)$$

exists and is finite for each function $F \in \mathcal{F}$. Now Bryc proved:

Theorem 5.1 *If $(P_n)_n$ admits an asymtotic value over $C_b(X)$, the Banach space of all bounded continuous functions $F : X \to \mathbb{R}$ with the supremum norm, and is exponentially tight, the LDP holds with rate function*

$$I(x) = \sup\{F(x) - L(F), F \in C_b(X)\}.$$

As a consequence of the Stone–Weierstrass theorem, Bryc also proved a criterion for the LDP, where the functions, which admit an asymptotic value, are in a subset of the space of all continuous and not necessarily bounded functions $X \to \mathbb{R}$:

Theorem 5.2 *Let Y be a locally convex Hausdorff topological linear space, Y^* its conjugate. Suppose $X \subset Y$ is a metric space in the relative topology and $D \subseteq Y^*$ is a dense linear subspace. Define*

$$\mathcal{G} := \{g : g(x) = \min_i\{\lambda_i(x) + c_i\}; c_i \in \mathbb{R}, \lambda_i \in D, 1 \le i \le n, n \in \mathbb{N}\}.$$

Suppose $(P_n)_n$ is exponentially tight and admits an asymptotic value over $\mathcal{G}$. Then $(P_n)_n$ satisfies a LDP with rate function

$$I(x) = \sup\{F(x) - L(F); F \in \mathcal{G}\}.$$

Remark 5.3 *Bryc [4] successfully used the last criterion to prove a LDP for stationary weakly dependent random fields.*

Now we can list simple but useful consequences:

Corollary 5.4 *Let X be Polish. Then for a sequence of probability measures $(P_n)_n$ on X the following statements are equivalent:*

1. *$(P_n)_n$ satisfies a LDP with rate I.*

2. *$(P_n)_n$ admits an asymptotic value over $C_b(X)$ an is exponentially tight.*

In the context of Theorem 5.2 the statements are replaced by

1. *$(P_n)_n$ satisfies a LDP with rate I and*

$$\sup_n \left(\int_X \exp(n\lambda^*(x))dP_n(x) \right)^{\frac{1}{n}} < \infty \quad \forall \quad \lambda^* \in Y^* \tag{22}$$

   holds.

2. *$(P_n)_n$ admits an asymptotic value over $\mathcal{G}$ and is exponentially tight.*

Corollary 5.5 *If $(P_n)_n$ satisfies a LDP with rate function I, then*

$$I(x) = \sup\{F(x) - L(F); F \in C_b(X)\}. \tag{23}$$

In the context of Theorem 5.2, if in addition (22) holds, the rate is of the form

$$I(x) = \sup\{F(x) - L(F); F \in \mathcal{G}\}. \tag{24}$$

Proof: (Corollary 5.4): 1. $\Longrightarrow$ 2. is the Laplace method of Varadhan. The exponential tightness follows from lemma 2.6. of Lynch, Sethuraman [17], where X has to be Polish. 2. $\Longrightarrow$ 1. is Theorem 5.1.
In the second context: 1. $\Longrightarrow$ 2.: Let $F \in \mathcal{G}$, we have to prove that $L(F)$ exists and is finite. Let $F(x) = \min_i\{\lambda_i(x) + c_i\}; c_i \in \mathbb{R}; \lambda_i \in D; 1 \leq i \leq n$. Thus for all i

$$\int \exp(nF(x))dP_n(x) \leq \int \exp(n\lambda_i(x) + nc_i)dP_n(x),$$

and with assumption (22)

$$\sup_n \left(\int \exp(n\alpha F(x))dP_n(x) \right)^{\frac{1}{n}} < \infty$$

with $\alpha = 2$ holds. Therefore the calculation in Theorem 4.3 shows that we can apply Lemma 4.2, so $(P_n)_n$ admits the asymptotic value over $\mathcal{G}$.
2. $\Longrightarrow$ 1.: This is Theorem 5.2 and if $(P_n)_n$ admits the asymptotic value over $\mathcal{G}$ this

especially implies that $L(\lambda^*)$ exists for all $\lambda^* \in D$. Moreover, given any $\varepsilon > 0$ there exists n_0 such that for all $n \geq n_0$

$$\left(\int_X \exp(n\lambda^*(x))dP_n(x)\right)^{\frac{1}{n}} \leq \exp(L(\lambda^*) + \varepsilon),$$

thus (22) holds.

(Corollary 5.5:) Combining Theorem 5.1 and Corollary 5.4, resp. Theorem 5.2 and Corollary 5.4. □

Of course, the proof of Corollary 5.5 is in some sense a diversion, because the asymptotic value method is an approach to large deviations, but we start with a LDP. Dinwoodie [7] has proved the identification (23) in Corollary 5.5 directly without using Bryc's theorems. Thus we will do so for the identification (24) which can be applied in the U–statistic case: Let Y, Y^*, X, D and $\mathcal{G}$ be as in Theorem 5.2.

Theorem 5.6 *(Rate representation in LD–theory)*
If $(P_n)_n$ satisfies a LDP with rate function I and if $L(F)$ exists for each $F \in \mathcal{G}$ then

$$I(x) = \sup\{F(x) - L(F); F \in \mathcal{G}\}. \tag{25}$$

Corollary 5.7 *Given a U–statistic U_n with $\mathbb{R}^d$–valued kernel function satisfying the exponential boundedness condition (5). Then the rate function in Theorem 2.4 has a representation as in (25).*

Proof: (of Corollary 5.7) $P \circ U_n^{-1}$ satisfies a LDP with rate I. Let $F \in \mathcal{G}$ with representation $\min_i(\langle \lambda_i, x\rangle + c_i)$, we can calculate for all $i \in \{1, \ldots, n\}$:

$$\begin{aligned} \int \exp(nF(x))dP \circ U_n^{-1}(x) &\leq \int \exp(n\langle \lambda_i, x\rangle + nc_i)dP \circ U_n^{-1} \\ &\leq \exp(nc_i)\left(\int \exp(m\langle \lambda_i, h\rangle)dP\right)^k, \end{aligned} \tag{26}$$

where $k = \left[\frac{n}{m}\right]$ and thus

$$\left(\int \exp(nF(x))dP \circ U_n^{-1}\right)^{\frac{1}{n}} \leq (\exp(c_i))\,(E(\exp(m\langle \lambda_i, h\rangle)))^{\frac{1}{m}}$$

and as in Theorem 4.3 $P \circ U_n^{-1}$ admits the asymptotic value over $\mathcal{G}$. □

Proof: (of Theorem 5.6) Y^* separates points of X by the Hahn–Banach theorem (see e.g. [11], Chapter 1) and since D is dense and linear, it separates points of X, too.

Moreover $\mathcal{G}$ contains the constant and is closed under finite minima, thus the Stone–Weierstrass theorem (see e.g. Schaefer p.243) can be applied.
As a consequence of $L(F) < \infty$ for each $F \in \mathcal{G}$ we have

$$L(F) = \sup_{x \in X}\{F(x) - I(x)\} = -\inf_{x \in X}\{I(x) - F(x)\}.$$

To prove this variational formula, pick $\varepsilon > 0$ then there exists an n_0 such that for all $n \geq n_0$

$$\left(\int_X \exp(nF(x))dP_n(x)\right)^{\frac{1}{n}} \leq \exp(L(F) + \varepsilon),$$

thus by Lemma 4.2 $L(F) = \sup\{F(x) - I(x)\}$. An immediate consequence is that

$$I(x) \geq \sup_{F \in \mathcal{G}}\{F(x) - L(F)\}.$$

To show that $I(x) \leq \sup_{F\in\mathcal{G}}\{F(x) - L(F)\}$, we follow an earlier result of Dinwoodie [7], Theorem 2.1: Without loss of generality we can assume $I(x) > 0$, let $\delta < I(x)$ and choose $\varepsilon > 0$ such that $I(y) > \delta$ for each $y \in B_{x,2\varepsilon} := \{y : d(y,x) \leq 2\varepsilon\}$ (this is possible by the semi–continuity of I). For a fixed $N \geq 0$ let g_N be a continuous bounded function defined on X such that $g_N(y) = 0$ in $B_{x,\varepsilon}$, $g_N(y) = -N$ in $(B^c_{x,2\varepsilon})$ and $-N \leq g_N(y) \leq 0$ for $\varepsilon < d(y,x) < 2\varepsilon$ (see Theorem 1.2 [2]). Denote $\overline{B}$ the closure of $B \subset X$. Then by simple calculation

$$L(g_N) \leq -\min\{I(\overline{B_{x,2\varepsilon}}), N\} : \tag{27}$$

$$\begin{aligned} L(g_N) &= \lim_{n\to\infty} \frac{1}{n} \log\left(\int_{d(x,y)<2\varepsilon} \exp(ng_N)dP_n + \int_{d(x,y)\geq 2\varepsilon} \exp(ng_N)dP_n\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log\left(P_n(\overline{B_{x,2\varepsilon}}) + \exp(-n\,N)\right) \\ &\leq \max\{-I(\overline{B_{x,2\varepsilon}}), -N\} \end{aligned} \tag{28}$$

Since $(P_n)_n$ is exponentially tight by Lemma 2.6 [17], we can choose a large compact set $\hat{K}$. Then $K := \hat{K} \cup \{x\}$ is compact, too. By the Stone–Weierstrass theorem (see e.g. [20]) there is a finite collection $\{g_i\}_{1\leq i\leq n}$ of functions in $\mathcal{G}$ such that

$$\sup_{y\in K} |g_N(y) - \max_i g_i(y)| \leq \varepsilon.$$

Moreover, since $g_N(y) \leq 0$, passing to $g_i \wedge 0$ if necessary, we assume $g_i \leq 0$ for all i and we can choose $g_0 = 0$. Without loss of generality let $\max_i g_i(y) = 0$ for all $y \in B_{x,\varepsilon} \cap K$. By Lemma 3.2 [3],

$$L(\max(g_1, \ldots, g_n)) = \max(L(g_1), \ldots, L(g_n)),$$

if $L(g_i)$ exists for some measurable functions $g_i(x), 1 \leq i \leq n$. Define $L_A(F) := \lim_n \frac{1}{n} \log \int_A \exp(nF(x)) dP_n(x)$ for suitable, measurable $F : X \to \mathbb{R}$. Then for each j

$$L(g_N) \geq L_K(g_N) \geq L_K(\max_i g_i(x)) - \varepsilon = \max_i L_K(g_i) - \varepsilon \geq L_K(g_j) - \varepsilon$$

thus using (27) we get

$$\min\{I(\overline{B_{x,2\varepsilon}}), N\} \leq -L_K(g_j) + \varepsilon = -L_K(g_j) + g_j(x) + \varepsilon,$$

if we choose j such that $\max_i g_i(x) = g_j(x)(= 0)$. Now if $I(\overline{B_{x,2\varepsilon}}) = \infty$ we get

$$N \leq g_j(x) - L_K(g_j) + \varepsilon.$$

Since $\max_i L(g_i) \geq L(g_0) = 0$, by Lemma 4.1, [3]:

$$\lim_{n\to\infty} \left(\frac{1}{n} \log \int \exp(n \max_i g_i) dP_n - \frac{1}{n} \log \int_K \exp(n \max_i g_i) dP_n \right) = 0$$

and hence

$$I(x) = \infty = \sup_{g\in\mathcal{G}} \{g(x) - L(g)\}.$$

Otherwise choose $N > I(\overline{B_{x,2\varepsilon}})$ and by an argument of [7] we obtain

$$\delta < I(B_{x,2\varepsilon}) \leq g_j(x) - L_K(g_j) + \varepsilon \leq \sup_{g\in\mathcal{G}} \{g(x) - L(g)\}.$$

Now δ was choosen arbitrarily such that $0 < \delta < I(x) < \infty$. It follows

$$I(x) \leq \sup_{g\in\mathcal{G}} \{g(x) - L(g)\}$$

which proves the result. □

Acknowledgment: I would like to thank Matthias Löwe for helpful comments.

References

[1] Baxter, J.R.; Jain, N.C.: A comparison principle for large deviations; Proceedings of the American Mathematical Society Vol.103, No.4, 1235–1240 (1988)

[2] Billingsley, P.: Convergence of Probability Measures; New York, Wiley & Sons (1968)

[3] Bryc, W.: Large deviations by the asymptotic value method, Diffusion Processes and Related Problems in Analysis 1 (M. Pinsky ed.) Birkhäuser Boston, 447–472 (1990)

[4] Bryc, W.: On the large deviation principle for stationary weakly dependent fields, Ann. Probab. 20, 1004–1030 (1992)

[5] de Acosta, A.: Upper bounds for large deviations of dependent random vectors, Z. Wahrsch. verw. Geb. 69, 551–565 (1985)

[6] Deuschel, J.-D.; Stroock, D.W.: Large Deviations; Academic Press 137, (1989)

[7] Dinwoodie, I.H.: Identifying a large deviation rate function, Ann.Probab. 21 , 216–231 (1993)

[8] Donsker,M.D.; Varadhan,S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time (I,III,IV); Commun.Pure Appl.Math.28, 1-47 (1975); 29, 389-461 (1976); 36, 182-212 (1983)

[9] Eichelsbacher, P.: Varadhans Prinzip großer Abweichungen für spezielle Statistik-Klassen; Dissertation, Universität Bielefeld, (1992)

[10] Eichelsbacher, P.; Löwe, M.: Large deviation principle for m–variate von Mises–statistics and U–statistics, preprint 93–001, Sonderforschungsbereich 343, Diskrete Strukturen in der Mathematik, Universität Bielefeld (1993)

[11] Ekeland, I.; Teman, R.: Convex analysis and variational problems, North Holland, Amsterdam (1979)

[12] Ellis,R.S.: Entropy, large deviations and statistical mechanics; Grundlehren der Mathematischen Wissenschaften vol.271, Springer (1985)

[13] Groeneboom, P.; Oosterhoff, J.; Ruymgaart, F.H.: Large deviation theorems for empirical probability measures; Ann. Probab. 7, 553–586 (1979)

[14] Hoeffding, W.: A class of statistics with asymptotically normal distribution; Ann. of Math. Stat. 19, 293–325 (1948)

[15] Hoeffding, W.: Probability inequalities for sums of bounded random variables; Journal of the American Statistical Association 58, 13–30 (1963)

[16] Löwe, M.: Exponential Inequalities and Principles of Large Deviations for U–Statistics and von Mises–Statistics; Dissertation; Universität Bielefeld, (1992)

[17] Lynch, J.; Sethuraman, J.: Large deviations for processes with independent increments, Ann. Probab. 15, 610–627 (1987)

[18] von Mises, R.: On the asymptotic distribution of differentiable statistical functions; Ann. of Math. Stat. 18, 309–348 (1947)

[19] Sanov, J.N.: On the probability of large deviations of random variables; Selected Translations in Mathematical Statistics and Probability I, 214–244 (1957)

[20] Schaefer, H.: Topological Vector Spaces, Springer–Verlag, New–York (1971)