

LIMIT THEOREMS FOR RANDOM POINTS IN A SIMPLEX

Abstract

In this work the ℓ_q -norms of points chosen uniformly at random in a centered regular simplex in high dimensions are studied. Berry–Esseen bounds in the regime $1 \le q < \infty$ are derived and complemented by a non-central limit theorem together with moderate and large deviations in the case where $q = \infty$. An application to the intersection volume of a regular simplex with an ℓ_p^n -ball is also carried out.

Keywords: Berry–Esseen bound; central limit theorem; high dimensions; moderate deviations; ℓ_q -norm; large deviations; regular simplex

2020 Mathematics Subject Classification: Primary 60F05; 60F10 Secondary 52A23; 60D05

1. Introduction and main results

One of the central aspects of high-dimensional probability theory is the study of random geometric quantities and the phenomena that occur as the dimension of the ambient space tends to infinity. The field is intimately connected to geometric functional analysis as well as convex and discrete geometry, and has attracted considerable attention in the last decade. This is in part because of numerous applications that can be found in the statistics and machine learning literature related to high-dimensional data, e.g. in the form of dimensionality reduction in information retrieval, clustering, principal component regression, community detection, or covariance estimation; see [28] as well as the references provided therein. A famous example of a high-dimensional limit theorem is the Maxwell–Poincaré–Borel lemma (see, e.g., [8]) stating that, for fixed $k \in \mathbb{N}$, the distribution of the first k coordinates of a point chosen uniformly at random from the *n*-dimensional Euclidean ball or sphere of radius one converges weakly to a k-dimensional Gaussian distribution as the space dimension n tends to infinity.

Received 11 May 2020; revision received 29 August 2021.

^{*} Postal address: Faculty of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany.

^{**} Email address: anastas.baci@rub.de

^{***} Postal address: Faculty of Mathematics, University of Münster, 48149 Münster, Germany. Email: zakhar.kabluchko@uni-muenster.de

^{****} Postal address: Institute of Mathematics and Scientific Computing, University of Graz, 8010 Graz, Austria.

^{*****} Postal address: Faculty of Computer Science and Faculty of Computer Science and Mathematics, University of Passau, 94032 Passau, Germany.

^{******} Email address: joscha.prochno@uni-graz.at

^{*******} Email address: mathias.sonnleitner@uni-graz.at

^{*******} Email address: christoph.thaele@rub.de

[©] The Author(s), 2022. Published by Cambridge University Press on behalf of Applied Probability Trust.

Today, there is a vast literature on high-dimensional central limit theorems, which describe the Gaussian fluctuations for various random geometric quantities in different contexts, for instance the famous central limit theorem for convex bodies [20], a central limit theorem for the volume of random projections of the cube [24], or a central limit theorem for the q-norm (Euclidean norm) of (projections of) points chosen randomly from ℓ_p^n -balls [2, 15, 17], to mention just a few. Other limit theorems, such as moderate deviations principles and large deviations principles, have only been studied in high-dimensional probability related to the geometry of convex bodies since their introduction in [1, 11, 15, 17]. In fact, those kinds of limit theorems are more sensitive to the randomness involved and display a non-universal behavior in their speed and/or rate function. The latter fact makes the subject particularly interesting as, contrary to a central limit theorem which implies the somewhat negative result that fluctuations do not provide much information because of universality, the moderate and large deviations limit theorems are distribution dependent and encode subtle geometric information about the underlying structure. In the past three years a number of interesting results in this direction have been obtained, and we refer the reader to [10, 16, 18, 19, 21]. A link between the study of moderate and large deviations and the famous Kannan-Lovász-Simonovits conjecture has recently been discovered [3].

In this work we study limit theorems for (suitably normalized) ℓ_q -norms of points chosen uniformly at random in a centered and regular simplex. When $1 \le q < \infty$, we provide a Berry-Esseen-type rate of convergence to a standard Gaussian distribution. For the case where $q = +\infty$, we complement this result with a non-central limit theorem, establishing the weak convergence to a Gumbel distribution, and provide both moderate and large deviations principles. Moreover, as an application of our central limit theorem, we present a version of the Schechtman–Schmuckenschläger result [25] and compute the volume of the intersection of an ℓ_p -ball with a regular simplex. We note that the method of proof in the Berry–Esseen-type central limit theorem differs from the previously mentioned ones in the sense that here we use a connection to the asymptotic theory of sums of random spacings and do not rely on a Schechtman–Zinn-type probabilistic representation.

In order to be more precise, let $n \in \mathbb{N}$ and consider the (n-1)-dimensional simplex $\Delta_{n-1} := \{x \in \mathbb{R}^n : x_i \ge 0, \sum_{i=1}^n x_i = 1\} = \operatorname{conv}\{e_1, \ldots, e_n\}$, where $\operatorname{conv}(A)$ denotes the convex hull of a set *A* and e_1, \ldots, e_n represent the unit vectors of the standard orthonormal basis of \mathbb{R}^n . Consider a sequence $(E_i)_{i \in \mathbb{N}}$ of independent random variables with an exponential distribution of mean 1 and, for each $n \in \mathbb{N}$, let $S_n := \sum_{i=1}^n E_i$ denote the *n*th partial sum. We study the sequence of random vectors

$$Z_n := \left(\frac{E_1}{S_n} - \frac{1}{n}, \dots, \frac{E_n}{S_n} - \frac{1}{n}\right) \in \Delta_{n-1} - \frac{1}{n}(e_1 + \dots + e_n), \qquad n \in \mathbb{N}.$$

In fact, uniformly distributed points in the standard centered simplex in \mathbb{R}^n , $\Delta_{n-1} - \frac{1}{n}(e_1 + \cdots + e_n)$, have the same distribution as Z_n (see, e.g., [22]). A different method to generate uniform random vectors in the simplex is by letting U_1, \ldots, U_{n-1} be independent and identically distributed random variables with a uniform distribution on [0,1] and considering $G_{n,i} := U_{(i)} - U_{(i-1)}$, $i = 1, \ldots, n$, where $U_{(i)}$ is the *i*th order statistic of U_1, \ldots, U_{n-1} , with the convention that $U_{(0)} := 0$, $U_{(n)} := 1$. Then the vector $G_n := (G_{n,1}, \ldots, G_{n,n})$ is uniformly distributed in Δ_{n-1} and $Z_n \stackrel{d}{=} G_n - \frac{1}{n}(e_1 + \cdots + e_n)$. A proof of this fact can be found, for instance, in [5, Section 6.4].

The first main result of this paper is the following Berry-Esseen-type theorem for the ℓ_q -norm $\|\cdot\|_q$ $(1 \le q < \infty)$ of uniform random points in $\Delta_{n-1} - \frac{1}{n}(e_1 + \cdots + e_n)$. Define

Limit theorems for random points in a simplex

 $\mu_q := \mathbb{E}\left[|E_1 - 1|^q\right] \text{ and } \sigma_q^2 := q^{-2}\mu_q^{-2} \left(\mu_{2q} - \left(q^2 + 2q + 2\right)\mu_q^2 + 2(q+1)\mu_q - 1\right). \text{ For example, } \mu_1 = 2e^{-1}, \sigma_1 = 2e - 5, \text{ and } \mu_2 = \sigma_2 = 1.$

Theorem 1.1 Let $1 \le q < \infty$. There exists some constant $c_q > 0$ depending only on q such that, for all $n \in \mathbb{N}$,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left[\sqrt{n} \left(\frac{n^{1-1/q} \|Z_n\|_q \mu_q^{-1/q} - 1}{\sigma_q} \right) \le x \right] - \mathbb{P}[N \le x] \right| \le c_q \frac{\log n}{\sqrt{n}},$$

where $N \sim \mathcal{N}(0, 1)$ is a standard Gaussian random variable.

In the proof of this result we use a connection to the asymptotic theory of sums of random spacings. This allows us to use a Berry–Esseen result of Mirakhmedov [23], who improved a theorem due to Does and Klaassen [9].

As a direct corollary of Theorem 1.1, we obtain the following central limit theorem, which is the analogue for the regular simplex of the corresponding central limit theorems in [15, 17] for ℓ_p^n -balls. It will find an application in Section 4.

Corollary 1.1 *For all* $1 \le q < \infty$,

$$\sqrt{n}\left(n^{1-1/q} \|Z_n\|_q \, \mu_q^{-1/q} - 1\right) \xrightarrow[n \to \infty]{d} Z \sim \mathcal{N}(0, \, \sigma_q^2).$$

When the parameter q satisfies $q = \infty$, we cannot expect convergence in distribution of $||Z_n||_{\infty}$ to a Gaussian random variable. However, we establish a non-central limit theorem with a double exponential (also known as Gumbel) distribution in the limit. The result in this case reads as follows.

Theorem 1.2 We have

$$n\|Z_n\|_{\infty} - (\log n - 1) \xrightarrow{\mathrm{d}}_{n \to \infty} G,$$

where *G* has a standard Gumbel distribution, i.e. $\mathbb{P}[G \le x] = \exp(-e^{-x})$ for $x \in \mathbb{R}$.

The next result describes the upper and lower deviations on a moderate scale, which lies between the Gaussian fluctuations of a central limit theorem and the large deviations which occur on the scale of a law of large numbers. For a formal definition of a large deviations principle (LDP), see Section 2.2.

Theorem 1.3 Let $(s_n)_{n \in \mathbb{N}}$ be a positive sequence with $s_n \to \infty$ and $s_n / \log n \to 0$. Then, the sequence $\left(\frac{\log n}{s_n}\left(\frac{n}{\log n} \|Z_n\|_{\infty} - 1\right)\right)_{n \in \mathbb{N}}$ satisfies an LDP with speed s_n and rate function

$$\mathbb{I}(z) := \begin{cases} z, & z \ge 0, \\ +\infty, & z < 0. \end{cases}$$

As a last result, we establish the following large deviations principle for the ℓ_{∞} -norm, which we compare in Section 4 to the LDP for random points in the *n*-dimensional crosspolytope, i.e. the ℓ_1 -ball in \mathbb{R}^n .

Theorem 1.4 The sequence $\left(\frac{n}{\log n} \|Z_n\|_{\infty}\right)_{n \in \mathbb{N}}$ satisfies an LDP with speed $s_n = \log n$ and rate function

$$\mathbb{I}(z) := \begin{cases} z - 1, & z \ge 1, \\ +\infty, & z < 1. \end{cases}$$

The rest of the paper is organized as follows. In Section 2 we collect some background material on large deviations and introduce the notation we use throughout the paper. Section 3 is then devoted to the proofs of Theorems 1.1, 1.2, 1.3, and 1.4. In Section 4 we compare the LDP for the simplex with the one for ℓ_p^n -balls, in particular the one for the crosspolytope, and present an application of the central limit theorem to high-dimensional intersection volumes.

2. Preliminaries

2.1. General notation

Let $n \in \mathbb{N}$. Given $1 \le q \le \infty$ and a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we write

$$\|x\|_{q} = \begin{cases} \left(\sum_{i=1}^{n} |x_{i}|^{q}\right)^{1/q}, & q < \infty, \\ \max_{1 \le i \le n} |x_{i}|, & q = \infty. \end{cases}$$

We will assume that all random quantities are defined on a common probability space $(\Omega, \Sigma, \mathbb{P})$ and we write $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$ for the probability of an event and the expectation of an (integrable) random variable, respectively. For a sequence of independent and identically distributed (i.i.d.) random vectors $(X_i)_{i \in \mathbb{N}}$ we denote by $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ the empirical average. Throughout, E, E_1, E_2, \ldots will be independent exponential random variables having rate 1 and $\bar{E}_n = \frac{1}{n} \sum_{i=1}^n E_i$. Note that $\mathbb{P}[\bar{E}_n = 0] = 0$ and thus we can ignore this event in our analysis. By $\mathcal{N}(\mu, \Sigma)$ we denote the (multivariate) Gaussian distribution with mean μ and covariance matrix Σ . If a random variable N is distributed according to $\mathcal{N}(\mu, \Sigma)$, we write $N \sim \mathcal{N}(\mu, \Sigma)$. With $\stackrel{d}{\rightarrow}$ and $\stackrel{\mathbb{P}}{\rightarrow}$ we indicate convergence in distribution and in probability, respectively. We say that a sequence of real-valued random variables $(X_n)_{n \in \mathbb{N}}$ satisfies a central limit theorem (CLT) if there exists a sequence $(a_n)_{n \in \mathbb{N}}$ of real numbers such that $\sqrt{n}(a_nX_n - 1) \stackrel{d}{\rightarrow} N \sim \mathcal{N}(0, 1)$ as $n \to \infty$. For further background material on asymptotic probability theory consult, for example, [4].

2.2. Large and moderate deviations

In the following, we recall facts from the theory of large deviations as developed, for example, in [6].

A sequence $(X_n)_{n \in \mathbb{N}}$ of real-valued random variables is said to satisfy an LDP with speed $(s_n)_{n \in \mathbb{N}} \subset (0, \infty)$ and rate function $\mathbb{I} : \mathbb{R} \to [0, \infty]$ if \mathbb{I} is lower semi-continuous, has compact level sets $\{z \in \mathbb{R} : \mathbb{I}(z) \le \alpha\}$, $\alpha \in \mathbb{R}$, and if, for all Borel sets $A \subset \mathbb{R}$,

$$-\inf_{z\in A^{\circ}}\mathbb{I}(z)\leq \liminf_{n\to\infty}s_n^{-1}\log \mathbb{P}\left[X_n\in A\right]\leq \limsup_{n\to\infty}s_n^{-1}\log \mathbb{P}\left[X_n\in A\right]\leq -\inf_{z\in\bar{A}}\mathbb{I}(z).$$

Here, A° denotes the interior and \overline{A} the closure of A. For the empirical average of i.i.d. (real-valued) random variables an LDP holds according to Cramér's theorem, which we state next.

Lemma 2.1 (Cramérs theorem [6, Theorem 2.2.3]) Let X, X_1, X_2, \ldots be i.i.d. real-valued random variables. Assume that the origin is an interior point of the domain of the cumulant-generating function $\Lambda(u) = \log \mathbb{E}[\exp(uX)]$. Then the sequence of partial sums $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, $n \in \mathbb{N}$, satisfies an LDP on \mathbb{R} with speed n and rate function Λ^* , where $\Lambda^*(z) = \sup_{u \in \mathbb{R}} (uz - \Lambda(u))$ for all $z \in \mathbb{R}$.

It is often useful to transfer an LDP for a sequence of random variables to another such sequence when they do not differ too much from each other. The following lemma provides a condition (called exponential equivalence) under which such an attempt is possible.

Lemma 2.2 (Exponential equivalence[6, Theorem 4.2.13]) Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be two sequences of real-valued random variables, and assume that $(X_n)_{n\in\mathbb{N}}$ satisfies an LDP with speed s_n and rate function \mathbb{I} . If $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are exponentially equivalent at speed s_n , i.e. we have, for any $\delta > 0$, $\limsup_{n\to\infty} s_n^{-1} \log \mathbb{P}[|X_n - Y_n| > \delta] = -\infty$, then $(Y_n)_{n\in\mathbb{N}}$ satisfies an LDP with the same speed and rate function as $(X_n)_{n\in\mathbb{N}}$.

We shall need the following result for moderate deviations of the empirical average of i.i.d. random variables.

Lemma 2.3 (Moderate deviations [6, Theorem 3.7.1]) Let X, X_1, X_2, \ldots be *i.i.d. real-valued* random variables with $\mathbb{E}[X] = \mu$ and $\operatorname{Var} X = \sigma^2 > 0$. Assume that the origin is an interior point of the domain of the cumulant-generating function $\Lambda(u) = \log \mathbb{E} \exp(uX)$. Fix a sequence $(a_n)_{n \in \mathbb{N}}$ with $a_n \to 0$ and $na_n \to \infty$. Then, the sequence $\sqrt{na_n}(\bar{X}_n - \mu)$ satisfies an LDP on \mathbb{R} with speed a_n^{-1} and rate function $\mathbb{I}(z) = z^2/2\sigma^2$.

3. The proofs

We now present the proofs of Theorems 1.1, 1.2, 1.3, and 1.4, and start with the Berry–Esseen-type central limit theorem followed by the non-central limit theorem together with the moderate and large deviations principles when $q = \infty$.

3.1. Proof of the Berry–Esseen CLT

The general philosophy of the proof is similar to that in [14]. However, as already explained, we shall use a connection to the asymptotic theory of sums of spacings. The following lemma is a version of [2, Lemma 4.1] (see also [14, Lemma 2.8]).

Lemma 3.1 For any real-valued random variables X, Y and any $\varepsilon > 0$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}[Y \le x] - \mathbb{P}[N \le x] \right| \le \sup_{x \in \mathbb{R}} \left| \mathbb{P}[X \le x] - \mathbb{P}[N \le x] \right| + \mathbb{P}[|X - Y| > \varepsilon] + \frac{\varepsilon}{\sqrt{2\pi}},$$

where N is a standard Gaussian random variable.

Proof. Let $x \in \mathbb{R}$ and $\varepsilon > 0$. Then,

$$\begin{split} \mathbb{P}[Y \le x] - \mathbb{P}[N \le x] \le \mathbb{P}[Y \le x, |X - Y| \le \varepsilon] + \mathbb{P}[|X - Y| > \varepsilon] - \mathbb{P}[N \le x] \\ \le \mathbb{P}[X \le x + \varepsilon] + \mathbb{P}[|X - Y| > \varepsilon] - \mathbb{P}[N \le x] \\ = \mathbb{P}[X \le x + \varepsilon] - \mathbb{P}[N \le x + \varepsilon] + \mathbb{P}[|X - Y| > \varepsilon] \\ + \mathbb{P}[N \le x + \varepsilon] - \mathbb{P}[N \le x]. \end{split}$$

Using that $\mathbb{P}[N \le x + \varepsilon] - \mathbb{P}[N \le x] \le \varepsilon/\sqrt{2\pi}$ for all $x \in \mathbb{R}$, taking absolute values, and forming the supremum completes the proof.

The next lemma shows that, similar to CLTs, Berry–Esseen-type bounds can also be transferred by "nice" functions.

Lemma 3.2 Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of real-valued random variables. Suppose that there exist constants $\mu \in \mathbb{R}$ and $\sigma > 0$ such that the Berry–Esseen-type bound

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left[\sqrt{n} \left(\frac{X_n - \mu}{\sigma} \right) \le x \right] - \mathbb{P}[N \le x] \right| \le a_n$$

holds for some sequence $(a_n)_{n \in \mathbb{N}}$ and all $n \in \mathbb{N}$, where N is a standard Gaussian random variable. If $g : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable at μ with $g'(\mu) > 0$ then, for some constant C > 0 and all $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left[\sqrt{n} \left(\frac{g(X_n) - g(\mu)}{g'(\mu)\sigma} \right) \le x \right] - \mathbb{P}[N \le x] \right| \le C \max\left\{ a_n, \frac{\log n}{\sqrt{n}} \right\}$$

Proof. Set $\xi_n := \sqrt{n}(X_n - \mu)/\sigma$ and $\zeta_n := \sqrt{n}(g(X_n) - g(\mu))/g'(\mu)\sigma$. We use Lemma 3.1 to infer, for each $n \in \mathbb{N}$ and every $\varepsilon > 0$, that

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}[\zeta_n \le x] - \mathbb{P}[N \le x] \right| \le \sup_{x \in \mathbb{R}} \left| \mathbb{P}[\xi_n \le x] - \mathbb{P}[N \le x] \right| + \mathbb{P}[|\xi_n - \zeta_n| > \varepsilon] + \frac{\varepsilon}{\sqrt{2\pi}}.$$
 (3.1)

Fix $n \in \mathbb{N}$. We will estimate $\mathbb{P}[|\xi_n - \zeta_n| > \varepsilon]$ and choose ε suitably.

Making use of the Taylor expansion of g at μ yields that there exists some $\delta > 0$ such that, for all $x \in (\mu - \delta, \mu + \delta)$, $g(x) - g(\mu) = g'(\mu)(x - \mu) + \Phi(x - \mu)$, where $\Phi : \mathbb{R} \to \mathbb{R}$ is a function such that, for all $x \in (\mu - \delta, \mu + \delta)$, $|\Phi(x)| \le M_g |x - \mu|^2$ for $M_g = \sup_{x \in (\mu - \delta, \mu + \delta)} |g''(x)|/2$. Thus, if $|X_n - \mu| < \delta$, $|g(X_n) - g(\mu) - g'(\mu)(X_n - \mu)| \le M_g |X_n - \mu|^2$. We get, after division by $g'(\mu)\sigma$ and multiplication by \sqrt{n} ,

$$|\xi_n-\zeta_n|\leq \frac{M_g}{g'(\mu)\sigma}\sqrt{n}|X_n-\mu|^2.$$

Therefore, for every $n \in \mathbb{N}$,

$$\mathbb{P}[|\xi_n - \zeta_n| > \varepsilon] \le \mathbb{P}\left[\left| \sqrt{n}(X_n - \mu) \right| > \sqrt{\frac{\varepsilon g'(\mu)\sigma\sqrt{n}}{M_g}} \right] + \mathbb{P}\left[\left| \sqrt{n}(X_n - \mu) \right| > \delta\sqrt{n} \right].$$
(3.2)

If $M_g = 0$, the first summand disappears and we can set $\varepsilon = a_n$. By the assumed Berry–Esseen bound and the symmetry of a Gaussian random variable, for each $n \in \mathbb{N}$ and every $x \in \mathbb{R}$,

$$\mathbb{P}\left[\left|\sqrt{n}(X_n-\mu)\right| > \sigma x\right] = \mathbb{P}\left[\sqrt{n}(X_n-\mu) < -\sigma x\right] + \mathbb{P}\left[\sqrt{n}(X_n-\mu) > \sigma x\right]$$
$$\leq 2\left(\mathbb{P}\left[N \ge x\right] + a_n\right).$$

Together with the bound $\mathbb{P}[N \ge x] \le e^{-x^2/2}$, the second summand in (3.2) is

$$\mathbb{P}\left[\left|\sqrt{n}(X_n-\mu)\right| > \delta\sqrt{n}\right] \le 2\left(\exp\left\{-\frac{\delta^2}{2\sigma^2}n\right\} + a_n\right) \le c \max\left\{a_n, \frac{\log n}{\sqrt{n}}\right\}$$

for some c > 0 independent of $n \in \mathbb{N}$. If $M_g > 0$, by setting $\varepsilon = \sigma M_g g'(\mu)^{-1} n^{-1/2} \log n$ for each $n \in \mathbb{N}$, the first summand is

$$\mathbb{P}\left[\left|\sqrt{n}(X_n-\mu)\right| > \sigma\sqrt{\log n}\right] \le 2(n^{-1/2}+a_n) \le c' \max\left\{a_n, \frac{\log n}{\sqrt{n}}\right\}$$

for some constant c' > 0 independent of *n*. This choice of ε yields

$$\mathbb{P}[|\xi_n - \zeta_n| > \sigma M_g g'(\mu)^{-1} n^{-1/2} \log n] \le (c+c') \max\left\{a_n, \frac{\log n}{\sqrt{n}}\right\}$$

Together with inequality (3.1) we have, for all $n \in \mathbb{N}$,

$$\sup_{\mathbf{x}\in\mathbb{R}} \left| \mathbb{P}[\zeta_n \le \mathbf{x}] - \mathbb{P}[N \le \mathbf{x}] \right| \le a_n + (c+c') \max\left\{ a_n, \frac{\log n}{\sqrt{n}} \right\} + \frac{\sigma M_g}{\sqrt{2\pi}g'(\mu)} \frac{\log n}{\sqrt{n}},$$

whereupon choosing C > 0 suitably completes the proof.

Remark 3.1 There exist results in the literature which are similar to Lemma 3.2. For example, [4, Theorem 11.6] deals with the case of X_n being empirical averages and g a function with Hölder-continuous derivative. In this case we have $a_n = n^{-1/2}$ and the guaranteed bound for the modified sequence is of order $\log n/\sqrt{n}$. We do not know if this rate in Lemma 3.2 can be improved in general.

Recall the definition of the spacings in the introduction by $G_{n,i} = U_{(i)} - U_{(i-1)}$, i = 1, ..., n, where $U_{(i)}$ is the *i*th order statistic of $U_1, ..., U_{n-1}$, sampled independently and uniformly from the unit interval, with the convention that $U_{(0)} = 0$, $U_{(n)} = 1$. Also recall that *E* denotes an exponential random variable with rate 1.

We deduce the following theorem from [23] which refined a Berry–Esseen theorem from [9].

Theorem 3.1 Let $G_{n,1}, \ldots, G_{n,n}$ be as above. Suppose $f : \mathbb{R} \to \mathbb{R}$ is measurable with $\mathbb{E}[f(E)^3] < \infty$ and $\sigma^2 := \operatorname{Var} f(E) - \operatorname{Cov}(E, f(E))^2 > 0$. Then there exists a constant C > 0 such that, for all $n \in \mathbb{N}$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^{n}f(nG_{n,i})-\mathbb{E}[f(E)]}{\sigma}\right)\leq x\right]-\mathbb{P}[N\leq x]\right|\leq\frac{C}{\sqrt{n}},$$

where N is a standard Gaussian random variable.

The following result will be used for the proof of Theorem 1.1 and we skip its elementary proof. Again, *E* stands for a standard exponential random variable.

Lemma 3.3 Let $1 \le q < \infty$. Then $\text{Cov}(E, |E-1|^q) = (q+1)\mathbb{E}[|E-1|^q] - 1$, where $\mathbb{E}[|E-1|^q] = e^{-1}(\Gamma(q+1) + \int_0^1 x^q e^x dx).$

Proof of Theorem 1.1. Using the connection to the spacings $G_{n,1}, \ldots, G_{n,n}$ we have

$$\|Z_n\|_q^q = \sum_{i=1}^n \left|Z_{n,i} - \frac{1}{n}\right|^q \stackrel{\mathrm{d}}{=} \sum_{i=1}^n \left|G_{n,i} - \frac{1}{n}\right|^q = n^{-q} \sum_{i=1}^n |nG_{n,i} - 1|^q.$$

Thus, writing $n^{q-1} ||Z_n||_q^q \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^n f(nG_{n,i})$, we are in the situation of Theorem 3.1 with $f(x) = |x-1|^q$, which is not of the form $x \mapsto ax + b$ for any $a, b \in \mathbb{R}$ and all $x \in \mathbb{R}$. Because of this and the fact that the exponential distribution has finite moments of all orders, the assumptions of Theorem 3.1 are satisfied and there exists a constant $c_q > 0$ depending only on q such that, for all $n \in \mathbb{N}$,

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left[\sqrt{n} \left(\frac{n^{q-1} \|Z_n\|_q^q - \mu_q}{d_q} \right) \le x \right] - \mathbb{P}[N \le x] \right| \le \frac{c_q}{\sqrt{n}},$$

with

$$d_q^2 = \operatorname{Var}|E - 1|^q - \operatorname{Cov}\left(E, |E - 1|^q\right)^2 = \mu_{2q} - \mu_q^2 - \left((q+1)\mu_q - 1\right)^2$$

and $\mu_q = \mathbb{E}[|E-1|^q]$ according to Lemma 3.3. Note that $d_q^2 > 0$ due to the Cauchy–Schwarz inequality and the fact that *E* and $|E-1|^q$ are linearly independent. Applying Lemma 3.2 with $g(x) = x^{1/q}$ and $g'(\mu_q) = q^{-1} \mu_q^{1/q-1}$, and rearranging terms, concludes the proof.

Remark 3.2 Using the so-called delta method, we can alternatively derive the central limit theorem stated as Corollary 1.1 from a CLT for sum-functions of spacings [13].

Remark 3.3 We briefly want to put μ_q in a more accessible form and compare it to the centering constant in the central limit theorem [15, Theorem 1.1], stating for $1 < q \le \infty$ and random vectors Y_n , which are uniformly distributed in the ℓ_1^n -ball, that

$$\sqrt{n} \left(n^{1-1/q} \|Y_n\|_q M_1(q)^{-1/q} - 1 \right) \xrightarrow{d}{n \to \infty} Z \sim \mathcal{N} \left(0, C_1(q, q) \right),$$

where $M_1(q) := \Gamma(q+1)$ and

$$C_1(q, q) = q^{-2} \left(\frac{\Gamma(2q+1)}{\Gamma(q+1)^2} - 1 \right) - 1.$$

As can be seen from Corollary 1.1, the same rate of $n^{1-1/q}$ appears. With repeated partial integration we can derive, for $q \in \mathbb{N}$,

$$\mu_q = \mathbb{E}\left[|E-1|^q\right] = \begin{cases} \mathbb{E}\left[(E-1)^q\right], & q \text{ even,} \\ 2e^{-1}q! - \mathbb{E}\left[(E-1)^q\right], & q \text{ odd,} \end{cases}$$

where $\mathbb{E}[(E-1)^q]$ equals the subfactorial $!q = q! \sum_{i=0}^q (-1)^i / i!$, which is also the nearest integer to $e^{-1}q!$. This is smaller than $M_1(q) = q!$ by roughly a factor of e.

3.2. Proof of the non-central limit theorem

In the following, we give a proof of Theorem 1.2 and analyze the limiting distribution of

$$\|Z_n\|_{\infty} = n \max_{1 \le i \le n} |Z_{n,i}| = n \max_{1 \le i \le n} \left| \frac{E_i}{S_n} - \frac{1}{n} \right|$$

Set $M_n := \max_{1 \le i \le n} E_i$ and let us recall the well-known fact that $M_n - \log n =: G_n \xrightarrow{d}_{n \to \infty} G$, where *G* is standard Gumbel distributed. First, we prove that $||Z_n||_{\infty}$ and

$$T_n := \max_{1 \le i \le n} \left(\frac{E_i}{S_n} - \frac{1}{n}\right) = \frac{M_n}{S_n} - \frac{1}{n}$$

are exponentially equivalent in the following sense.

Lemma 3.4 $\lim_{n\to\infty} n^{-1} \log \mathbb{P}[||Z_n||_{\infty} \neq T_n] < 0.$

Proof. We first prove that everywhere except for $\{S_n = 0\}$ we have the implication $||Z_n||_{\infty} \neq T_n \Longrightarrow M_n/S_n < 2/n$. Note that $||Z_n||_{\infty} \ge T_n$, and if $||Z_n||_{\infty} \neq T_n$ there must be some index $i_0 \in \{1, \ldots, n\}$ such that

$$\left|\frac{E_{i_0}}{S_n}-\frac{1}{n}\right|>T_n=\frac{M_n}{S_n}-\frac{1}{n}.$$

This can only occur if $E_{i_0}/S_n - 1/n$ is negative (otherwise we would have equality), i.e.

$$\frac{1}{n} - \frac{E_{i_0}}{S_n} = \left| \frac{E_{i_0}}{S_n} - \frac{1}{n} \right| > \frac{M_n}{S_n} - \frac{1}{n}$$

Since $E_{i_0}/S_n \ge 0$, this gives the desired implication. Therefore,

$$\mathbb{P}\Big[\|Z_n\|_{\infty} \neq T_n\Big] \le \mathbb{P}\bigg[M_n < 2\frac{S_n}{n}\bigg] \le \mathbb{P}[M_n < 4] + \mathbb{P}\bigg[\frac{S_n}{n} > 2\bigg].$$

By means of the inequality $1 + x \le e^x$, the first summand evaluates to $\mathbb{P}[M_n < 4] = (1 - e^{-4})^n \le e^{-e^{-4}n}$ and, by Cramér's theorem, Lemma 2.1, $\lim_{n\to\infty} n^{-1}\log \mathbb{P}[S_n/n > 2] < 0$. This completes the proof of the lemma.

We return to the proof of Theorem 1.2. The previous lemma implies that $n \|Z_n\|_{\infty} - nT_n \xrightarrow{\mathbb{P}} 0$, and to establish Theorem 1.2 it suffices to prove that $nT_n - (\log n - 1) \xrightarrow{d} G$.

The connection to the spacings $G_n = (G_{n,1}, \ldots, G_{n,n})$ gives $nT_n \stackrel{d}{=} n \max_{1 \le i \le n} G_{n,i} - 1$. A classical result [7, Lemma 2.4] states that $n \max_{1 \le i \le n} G_{n,i} - \log n \stackrel{d}{\xrightarrow{n \to \infty}} G$, which was originally derived by Lévy. Rewriting this for nT_n gives $nT_n - \log n + 1 \stackrel{d}{\xrightarrow{n \to \infty}} G$ and completes the proof.

Remark 3.4 We can read off from Theorem 1.2 that the maximum norm of random points in the shifted simplex converges to zero at the rate $\log n/n$ with fluctuations of order $(\log n)^{-1}$. Uniformly distributed random vectors Y_n in the ℓ_1^n -ball exhibit similar behavior, as can be read off from [15, Theorem 1.1(c)]. Namely, $n ||Y_n||_{\infty} - \log n \xrightarrow[n \to \infty]{d} G$.

3.3. Proof of the moderate deviations principle

We derive the proof of Theorem 1.3 from [7, Lemma 3.2], which we rephrase in our notation.

Lemma 3.5 Let $(a_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers satisfying $a_n \to 0$ and $a_n \log n \to \infty$ as $n \to \infty$. Then

$$\lim_{n \to \infty} n^{a_n} \mathbb{P}\left[\frac{n}{\log n} \frac{M_n}{S_n} - 1 > a_n\right] = 1, \qquad \lim_{n \to \infty} \exp\left(n^{a_n}\right) \mathbb{P}\left[\frac{n}{\log n} \frac{M_n}{S_n} - 1 < -a_n\right] = 1.$$

In order to prove Theorem 1.3, we restate this result in the following form.

Lemma 3.6 Let $(s_n)_{n \in \mathbb{N}}$ be a positive sequence with $s_n \to \infty$ and $s_n / \log n \to 0$ as $n \to \infty$. Then, for any x > 0,

$$\lim_{n \to \infty} \frac{1}{s_n} \log \mathbb{P}\left[\frac{\log n}{s_n} \left(\frac{n}{\log n} \|Z_n\|_{\infty} - 1\right) > x\right] = -x,$$
$$\lim_{n \to \infty} \frac{1}{s_n} \log \mathbb{P}\left[\frac{\log n}{s_n} \left(\frac{n}{\log n} \|Z_n\|_{\infty} - 1\right) < -x\right] = -\infty.$$

Proof. Using the continuity of the logarithm and inserting $T_n = M_n/S_n - 1/n$, we can deduce from Lemma 3.5 that

$$\lim_{n \to \infty} a_n \log n + \log \mathbb{P}\left[\frac{n}{\log n}T_n - 1 > a_n - \frac{1}{\log n}\right] = 0,$$
(3.3)

$$\lim_{n \to \infty} n^{a_n} + \log \mathbb{P}\left[\frac{n}{\log n}T_n - 1 < -a_n - \frac{1}{\log n}\right] = 0$$
(3.4)

for any such sequence $(a_n)_{n \in \mathbb{N}}$.

In the following, let x > 0 be arbitrary. First, set $a_n = (s_n/\log n)x + 1/\log n = (s_nx + 1)/\log n$, where $(s_n)_{n \in \mathbb{N}}$ is any sequence with $s_n \to \infty$ and $s_n/\log n \to 0$. This choice of $(a_n)_{n \in \mathbb{N}}$ meets the assumptions of Lemma 3.5. Then, inserting into (3.3) gives

$$\lim_{n \to \infty} s_n x + 1 + \log \mathbb{P}\left[\frac{n}{\log n} T_n - 1 > \frac{s_n}{\log n} x\right] = 0,$$

which, by considering the limit of the sequence divided by s_n , implies that

$$x + \lim_{n \to \infty} \frac{1}{s_n} \log \mathbb{P}\left[\frac{n}{\log n}T_n - 1 > \frac{s_n}{\log n}x\right] = 0.$$

For (3.4) we choose $a_n = (s_n/\log n)x - 1/\log n = (s_nx - 1)/\log n$ for all $n \in \mathbb{N}$ such that $s_nx - 1 > 0$ and set $a_n = 1$ for all other $n \in \mathbb{N}$. Since $s_n \to \infty$, we have $s_nx - 1 > 0$ for all $n \ge n_0$, where $n_0 \in \mathbb{N}$ may depend on x, and only need to set finitely many terms $a_n = 1$. This choice of $(a_n)_{n \in \mathbb{N}}$ satisfies the assumptions of Lemma 3.5. Therefore,

$$\lim_{n\to\infty} n^{(s_nx-1)/\log n} + \log \mathbb{P}\left[\frac{n}{\log n}T_n - 1 < -\frac{s_n}{\log n}x\right] = 0.$$

Proceeding as before,

$$\lim_{n\to\infty}\frac{1}{s_n}n^{(s_nx-1)/\log n}+\frac{1}{s_n}\log \mathbb{P}\left[\frac{n}{\log n}T_n-1<-\frac{s_n}{\log n}x\right]=0.$$

Noting that $n^{(s_n x - 1)/\log n} / s_n = \exp(s_n x - 1) / s_n \to \infty$, we have

$$\lim_{n\to\infty}\frac{1}{s_n}\log \mathbb{P}\left[\frac{n}{\log n}T_n-1<-\frac{s_n}{\log n}x\right]=-\infty.$$

Since $s_n/n \to 0$ as $n \to \infty$, we can apply Lemma 3.4 and rearrange terms to complete the proof of Lemma 3.6.

We now deduce Theorem 1.3 using a standard technique in large deviations theory.

Proof of Theorem 1.3. Let $(s_n)_{n \in \mathbb{N}}$ be an arbitrary positive sequence with $s_n \to \infty$ and $s_n / \log n \to 0$. Theorem 1.3 follows if we can show for arbitrary open $U \subset \mathbb{R}$ and closed $C \subset \mathbb{R}$ the bounds

$$\liminf_{n \to \infty} s_n^{-1} \log \mathbb{P} \left[X_n \in U \right] \ge -\inf_{z \in U} \mathbb{I}(z),$$

$$\limsup_{n \to \infty} s_n^{-1} \log \mathbb{P} \left[X_n \in C \right] \le -\inf_{z \in C} \mathbb{I}(z),$$
(3.5)

where we used the notation $X_n := \frac{\log n}{s_n} \left(\frac{n}{\log n} \|Z_n\|_{\infty} - 1 \right)$. Recall that

$$\mathbb{I}(z) := \begin{cases} z, & z \ge 0, \\ +\infty, & z < 0. \end{cases}$$

If $A \subset \mathbb{R}$, we use the notation $A_- := A \cap (-\infty, 0]$ and $A_+ := A \cap [0, \infty)$ as well as $a_- := \sup A_-$ and $a_+ = \inf A_+$ such that $\inf_{z \in A} \mathbb{I}(z) = a_+$.

We first prove the upper bound in (3.5) and choose a closed set $C \subset \mathbb{R}$. If *C* is empty, the upper bound is trivial as the probability of X_n being in an empty set is zero. On the other hand, if $0 \in C$, the infimum is zero and the upper bound is satisfied due to $\mathbb{P}[X_n \in C] \leq 1$. Therefore, assume without loss of generality that at least one of C_- and C_+ is not empty and that $0 \notin C$. If both are non-empty, then $c_- < 0 < c_+$ and $C = C_- \cup C_+ \subset (-\infty, c_-] \cup [c_+, \infty)$, and $\mathbb{P}[X_n \in C] \leq \mathbb{P}[X_n < c_-] + \mathbb{P}[X_n > c_+]$ for $n \in \mathbb{N}$. This also makes sense if C_- is empty, i.e. $c_- = -\infty$ or if C_+ is empty, i.e. $c_+ = \infty$, if we interpret $\mathbb{P}[X_n < -\infty] = \mathbb{P}[X_n > \infty] = 0$. Because of the monotonicity of the logarithm and [6, Lemma 1.2.15], $\limsup_{n\to\infty} s_n^{-1} \log \mathbb{P}[X_n \in C] \leq \max\{M_-, M_+\}$ with $M_- := \limsup_{n\to\infty} s_n^{-1} \log \mathbb{P}[X_n < c_-]$ and $M_+ := \limsup_{n\to\infty} s_n^{-1} \log \mathbb{P}[X_n > c_+]$. Now the upper bound follows from Lemma 3.6 since $M_- = -\infty$ and $\max\{M_-, M_+\} = M_+ = -c_+ = -\inf_{z \in C} \mathbb{I}(z)$.

We now prove the lower bound and choose an open set $U \subset \mathbb{R}$. If U_+ is empty, the infimum is ∞ and the lower bound is trivially satisfied. Assume therefore, without loss of generality, that U_+ is not empty. Take any $z \in U_+$. We will show that

$$\liminf_{n \to \infty} s_n^{-1} \log \mathbb{P}[X_n \in U] \ge -\mathbb{I}(z).$$
(3.6)

This will conclude the proof of the lower bound in (3.5), since then

$$\liminf_{n\to\infty} s_n^{-1} \log \mathbb{P}[X_n \in U] \ge -\inf_{z\in U_+} \mathbb{I}(z) = -\inf_{z\in U} \mathbb{I}(z).$$

The openness of U yields that, for all $\varepsilon \ge 0$ small enough, $(z + \varepsilon, z + 2\varepsilon] \subset U$. Therefore,

$$\liminf_{n \to \infty} s_n^{-1} \log \mathbb{P} \left[X_n \in U \right] \ge \liminf_{n \to \infty} s_n^{-1} \log \mathbb{P} \left[X_n \in (z + \varepsilon, z + 2\varepsilon) \right]$$
$$= \liminf_{n \to \infty} s_n^{-1} \log \left(\mathbb{P} \left[X_n \ge z + \varepsilon \right] - \mathbb{P} \left[X_n \ge z + 2\varepsilon \right] \right)$$
$$\ge \liminf_{n \to \infty} s_n^{-1} \log \mathbb{P} \left[X_n \ge z + \varepsilon \right]$$
$$+ \liminf_{n \to \infty} s_n^{-1} \log \left(1 - \frac{\mathbb{P} \left[X_n \ge z + 2\varepsilon \right]}{\mathbb{P} \left[X_n \ge z + \varepsilon \right]} \right)$$
(3.7)

by the superadditivity of lim inf. We use Lemma 3.6 to conclude the proof. The first summand in (3.7) satisfies lim $\inf_{n\to\infty} s_n^{-1} \log \mathbb{P}[X_n > z + \varepsilon] = -z - \varepsilon = -\mathbb{I}(z) - \varepsilon$. Let $\delta > 0$ be small. For the second summand in (3.7) we choose $n_0 \in \mathbb{N}$ large enough such that both $\mathbb{P}[X_n > z + 2\varepsilon] \le e^{(-(z+2\varepsilon)+\delta)s_n}$ and $\mathbb{P}[X_n > z + \varepsilon] \ge e^{(-(z+\varepsilon)-\delta)s_n}$ for $n \ge n_0$. This shows that the fraction in the second summand satisfies, for every $n \ge n_0$,

$$0 \leq \frac{\mathbb{P}\left[X_n > z + 2\varepsilon\right]}{\mathbb{P}\left[X_n > z + \varepsilon\right]} \leq e^{(2\delta - \varepsilon)s_n}$$

which tends to zero for $n \to \infty$ if $\delta < \frac{1}{2}\varepsilon$. Thus, the second summand in (3.7) is equal to zero and, after letting $\varepsilon \to 0$, the inequality (3.6) is established for any $z \in U_+$, which proves the lower bound in (3.5) and thus Theorem 1.3.

3.4. Proof of the LDP

For the proof of Theorem 1.4 we use a rather general result on large deviations for maxima and minima [12, Proposition 3.1]. Recall that a function $H : \mathbb{R} \to \mathbb{R}$ is said to be regularly varying at ∞ of index $\alpha \in \mathbb{R}$ if, for all u > 0, $\lim_{t\to\infty} [H(ut)/H(t)] = u^{\alpha}$.

Lemma 3.7 Let $X, X_1, X_2, ...$ be i.i.d. real-valued random variables with $\mathbb{P}[X \le x] < 1$ for all x > 0 such that $-\log \mathbb{P}[X > x]$ is regularly varying at ∞ of index $\alpha > 0$ as a function of $x \in \mathbb{R}$. Choose $m_n \in \mathbb{R}$ such that $\mathbb{P}[X > m_n] = \frac{1}{n}$ and set $M_n = \max_{1 \le i \le n} X_i$. Then $\left(\frac{M_n}{m_n}\right)_{n \ge 1}$ satisfies an LDP with speed $s_n = \log n$ and good rate function

$$\mathbb{I}(z) := \begin{cases} z^{\alpha} - 1, & z \ge 1, \\ +\infty, & z < 1. \end{cases}$$

In order to prove Theorem 1.4 we make use of this result, Lemma 2.2, and the next lemma, which states the exponential equivalence of the sequences in question.

Lemma 3.8 Set $M_n = \max_{1 \le i \le n} E_i$. The sequences $\left(\frac{n}{\log n} \|Z_n\|_{\infty}\right)_{n \in \mathbb{N}}$ and $\left(\frac{M_n}{\log n}\right)_{n \in \mathbb{N}}$ are exponentially equivalent at speed log *n*.

Proof. This amounts to showing that, for every $\delta > 0$,

$$\limsup_{n \to \infty} \frac{1}{\log n} \log \mathbb{P}[|n||Z_n||_{\infty} - M_n| > \delta \log n] = -\infty$$

Fix $\delta > 0$. Then, for every $n \in \mathbb{N}$,

$$\mathbb{P}\left[\left|n\|Z_n\|_{\infty}-M_n\right|>\delta\log n\right]\leq \mathbb{P}\left[\left|nT_n-M_n\right|>\delta\log n\right]+\mathbb{P}\left[\|Z_n\|_{\infty}\neq T_n\right].$$

By Lemma 3.4 the second summand satisfies $\frac{1}{\log n} \log \mathbb{P}[\|Z_n\|_{\infty} \neq T_n] \to -\infty$ as $n \to \infty$. Recall the notation $\bar{E}_n = \frac{S_n}{n}$. For the first summand we compute $nT_n - M_n = M_n \bar{E}_n^{-1} - 1 - M_n = -(M_n(\bar{E}_n - 1)\bar{E}_n^{-1} + 1)$. Setting $A_n := M_n(\bar{E}_n - 1)\bar{E}_n^{-1}$, it follows that $\mathbb{P}[|nT_n - M_n| > \delta \log n] = \mathbb{P}[|A_n + 1| > \delta \log n]$. Whenever *n* is large enough $\log n > 2\delta^{-1}$ holds, and thus $\mathbb{P}[|A_n + 1| > \delta \log n] \leq \mathbb{P}[|A_n + 1| > 2] \leq \mathbb{P}[|A_n| > 1]$. Introducing $B_n := n^{-1/4}M_n$ and $C_n := n^{1/4}(\bar{E}_n - 1)\bar{E}_n^{-1}$, such that $A_n = B_nC_n$, gives $\mathbb{P}[|nT_n - M_n| > \delta \log n] \leq \mathbb{P}[|B_n| > 1/2] + \mathbb{P}[|C_n| > 2]$. By the union bound, the first summand satisfies

$$\mathbb{P}[|B_n| > 1/2] = \mathbb{P}\left[M_n > n^{1/4}/2\right] \le n\mathbb{P}\left[E_1 > n^{1/4}/2\right] = n\mathrm{e}^{-n^{1/4}/2}$$

and thus

$$\frac{1}{\log n}\log \mathbb{P}\big[|B_n| > 1/2\big] \le 1 - \frac{n^{1/4}}{2\log n} \xrightarrow{n \to \infty} -\infty$$

The other summand can be estimated by means of $\mathbb{P}[|C_n| > 2] \le \mathbb{P}[\bar{E}_n < 1/2] + \mathbb{P}[n^{1/4}|\bar{E}_n - 1| > 1]$. Cramér's theorem (Lemma 2.1) implies that

$$\frac{1}{\log n}\log \mathbb{P}\left[\bar{E}_n < 1/2\right] \xrightarrow{n \to \infty} -\infty$$

After splitting the second summand into

$$\mathbb{P}\left[n^{1/4}|\bar{E}_n-1|>1\right] = \mathbb{P}\left[n^{1/4}(\bar{E}_n-1)>1\right] + \mathbb{P}\left[n^{1/4}(\bar{E}_n-1)<-1\right],$$

we apply the moderate deviations principle (Lemma 2.3) with $a_n = n^{-1/2}$, $n \in \mathbb{N}$, giving

$$\lim_{n \to \infty} n^{-1/2} \log \mathbb{P}\left[n^{1/4} (\bar{E}_n - 1) > 1\right] = -\frac{1}{2},$$
$$\lim_{n \to \infty} n^{-1/2} \log \mathbb{P}\left[n^{1/4} (\bar{E}_n - 1) < -1\right] = -\frac{1}{2}.$$

Consequently,

$$\frac{1}{\log n}\log \mathbb{P}\left[n^{1/4}|\bar{E}_n-1|>1\right] \xrightarrow{n\to\infty} -\infty,$$

and thus

$$\frac{1}{\log n} \log \mathbb{P}[|C_n| > 2] \xrightarrow{n \to \infty} -\infty,$$

completing the proof.

Remark 3.5 The choice of the sequence $n^{1/4}$ was arbitrary; any sequence growing faster than $\log n$ and slower than $\left(\frac{n}{\log n}\right)^{1/2}$ would have done the job.

Proof of Theorem 1.4. We apply Lemma 3.7 to the case of standard exponential random variables and verify the assumptions. We have $\mathbb{P}[X \le x] = 1 - e^{-x} < 1$ for all $x \in \mathbb{R}$ and $-\log \mathbb{P}[X > x] = x$, which is regularly varying at ∞ of index $\alpha = 1$. We choose $m_n = \log n$ and thus obtain an LDP of speed $\log n$ and rate function as in Lemma 3.7 with $\alpha = 1$. By Lemma 2.2 the just-proven Lemma 3.8 implies Theorem 1.4.

4. Comparison with and application to ℓ_p^n -balls

4.1. An LDP for ℓ_p^n -balls

In [15] LDPs for the ℓ_q -norm of uniformly distributed points in ℓ_p^n -balls were proven for the cases $1 \le p < \infty$, $1 \le q < \infty$ and $p = \infty$, $1 \le q \le \infty$. In order to compare the LDP for the simplex with the one for the crosspolytope (i.e. the unit ball in ℓ_1^n), we complete the picture presented in [15] by deriving, without making the argument more complicated, a slightly more general LDP covering all possible values $1 \le p < \infty$ and put $q = \infty$. Let $1 \le p < \infty$ and $\mathbb{B}_p^n := \{x \in \mathbb{R}^n : \|x\|_p \le 1\}$ be the ℓ_p^n -unit ball. Further, let Z'_n be uniformly distributed in \mathbb{B}_p^n for all $n \in \mathbb{N}$.

Theorem 4.1 Let $1 \le p < \infty$. The sequence $((n/p \log n)^{1/p} ||Z'_n||_{\infty})_{n \in \mathbb{N}}$ satisfies an LDP with speed log *n* and rate function

$$\mathbb{I}(z) := \begin{cases} z^p - 1, & z \ge 1, \\ \infty, & z < 1. \end{cases}$$

In particular, taking p = 1 and comparing the result to Theorem 1.4 we see that the LDPs for the ℓ_{∞} -norm of uniformly distributed random points in a crosspolytope and a regular simplex are identical.

The proof of Theorem 4.1 is based on the Schechtman–Zinn probabilistic representation, saying that for any $n \in \mathbb{N}$ the distributional identity

$$Z'_n \stackrel{\mathrm{d}}{=} U^{1/n} \frac{Y'_n}{\|Y'_n\|_p}$$

holds, where *U* is uniformly distributed on [0,1] and independent of $Y'_n := (Y_1, \ldots, Y_n)$ which has i.i.d. *p*-generalized Gaussian distributed entries; see [26]. We recall that if a random variable *Y* is distributed according to the *p*-generalized Gaussian distribution its Lebesgue density f_p is given by $c_p e^{-|y|^p/p}$, $x \in \mathbb{R}$, with the normalization constant given by $c_p := (2p^{1/p}\Gamma(1+1/p))^{-1}$.

Next, we define for each $n \in \mathbb{N}$ the number $m_n(p) > 0$ by $\mathbb{P}[|Y| > m_n(p)] = \frac{1}{n}$, where *Y* is a *p*-generalized Gaussian random variable. From lemma 3.7 we can deduce the following result for $M_n(p) := \max_{1 \le i \le n} |Y_i|$.

Lemma 4.1 Let $1 \le p < \infty$. The sequence $(M_n(p)m_n(p)^{-1})_{n \in \mathbb{N}}$ satisfies an LDP with speed log *n* and rate function \mathbb{I} as in Theorem 4.1.

Proof. We check the assumptions of Lemma 3.7, and note that by the symmetry of the generalized Gaussian distribution, $\mathbb{P}[|Y| > x] = 2\mathbb{P}[Y > x] > 0$ for every x > 0. In order to check if $-\log \mathbb{P}[|Y| > x]$ is regularly varying at ∞ , we compute, for every u > 0,

$$\lim_{t \to \infty} \frac{\log \mathbb{P}[|Y| > ut]}{\log \mathbb{P}[|Y| > t]} = \lim_{t \to \infty} \frac{\log 2 + \log \mathbb{P}[Y > ut]}{\log 2 + \log \mathbb{P}[Y > t]} = \lim_{t \to \infty} \frac{(ut)^p}{t^p} = u^p.$$

Here, we used that according to the inequality

$$\frac{x}{x^p + p - 1} e^{-x^p/p} \le \int_x^\infty e^{-y^p/p} \, \mathrm{d}y \le \frac{1}{x^{p-1}} e^{-x^p/p},\tag{4.1}$$

which is valid for all $1 \le p < \infty$ and x > 0, $\log \mathbb{P}[|Y| > x]$ is asymptotically equivalent to $-x^p/p$ as $x \to \infty$. Thus, the assumptions of Lemma 3.7 are satisfied with $\alpha = \alpha_p := p > 0$, and we can complete the proof.

Lemma 4.2 The two sequences $(M_n(p)m_n(p)^{-1})_{n\in\mathbb{N}}$ and $(n^{1/p}\|Z'_n\|_{\infty}m_n(p)^{-1})_{n\in\mathbb{N}}$ are exponentially equivalent at speed log *n* for $1 \le p < \infty$.

Proof. Note that

$$n^{1/p} \|Z'_n\|_{\infty} m_n(p)^{-1} \stackrel{\mathrm{d}}{=} U^{1/n} \frac{M_n(p)m_n(p)^{-1}}{\left(\frac{1}{n} \sum_{i=1}^n |Y_i|^p\right)^{1/p}}$$

with the Y'_i being independent and *p*-generalized Gaussians. Then, the exponential equivalence is proved using Cramér's theorem as in the proof of [15, Theorem 1.3]. It remains to note that, for $0 < \varepsilon \le \delta$, we have, by Lemma 4.1,

$$\limsup_{n \to \infty} \frac{1}{\log n} \log \mathbb{P}\left[M_n(p)m_n(p)^{-1} > \frac{\delta}{\varepsilon}\right] = -\left(\left(\frac{\delta}{\varepsilon}\right)^p - 1\right),$$

$$-\infty \text{ as } \varepsilon \to 0.$$

which tends to $-\infty$ as $\varepsilon \to 0$.

After these preparations, we can now complete the proof of Theorem 4.1.

Proof of Theorem 4.1. By Lemma 4.1 the sequence $(M_n(p)m_n(p)^{-1})_{n\in\mathbb{N}}$ satisfies the required LDP, and so does $(n^{1/p}\|Z'_n\|_{\infty}m_n(p)^{-1})_{n\in\mathbb{N}}$ by Lemma 4.2. Using (4.1), we

can see that $m_n(p)(p \log n)^{-1/p}$ tends to 1 as $n \to \infty$. Consequently, the sequences $(n^{1/p} || Z'_n ||_{\infty} m_n(p)^{-1})_{n \in \mathbb{N}}$ and $((n/p \log n)^{1/p} || Z'_n ||_{\infty})_{n \in \mathbb{N}}$ satisfy the same LDP. \Box

4.2. Intersection with ℓ_p^n -balls

As an application of our central limit theorem, we sketch here a geometric application in the spirit of classical results of [26, 27] on the intersection of two different ℓ_p^n -balls. In our case, we consider the intersection volume of an ℓ_p -ball and a regular simplex. To be precise, we denote by $\mathbb{D}_p^n := \mathbb{B}_p^n / |\mathbb{B}_p^n|_n$ the normalized ℓ_p^n -ball, where $|\cdot|_n$ is a shorthand notation for the *n*-dimensional Lebesgue measure. Similarly, we define the normalized (n-1)-dimensional regular simplex $\Sigma_{n-1} := \tilde{\Delta}_{n-1} / |\tilde{\Delta}_{n-1}|_{n-1}$, where $\tilde{\Delta}_{n-1} := \Delta_{n-1} - \frac{1}{n}(e_1 + \cdots + e_n)$ is the centered (n-1)-dimensional regular simplex (note that $|\tilde{\Delta}_{n-1}|_{n-1} = |\Delta_{n-1}|_{n-1}$).

Theorem 4.2 Let
$$s > 0$$
, $1 \le p < \infty$, and define $A_p := 2e^{-1}\mu_p^{1/p}\Gamma\left(1+\frac{1}{p}\right)(pe)^{1/p}$. Then
$$\lim_{n \to \infty} \left|\Sigma_{n-1} \cap s\mathbb{D}_p^n\right|_{n-1} = \begin{cases} 1, & sA_p^{-1} > 1, \\ \frac{1}{2}, & sA_p^{-1} = 1, \\ 0, & sA_p^{-1} < 1. \end{cases}$$

Proof. We start by introducing some notation. For each $n \in \mathbb{N}$ we put

$$A_{p,n} := \mu_p^{1/p} \frac{n^{1/p} \left| \mathbb{B}_p^n \right|_n^{1/n}}{n |\tilde{\Delta}_{n-1}|_{n-1}^{1/(n-1)}}$$

and let s_n be such that $s_n(A_{p,n}/A_p) = s$. Using Stirling's formula repeatedly, we can verify that $A_{p,n} = A_p(1 + O(n^{-1} \log n))$. Indeed,

$$\begin{split} A_{p,n} &= 2\mu_p^{1/p} \Gamma\left(1 + \frac{1}{p}\right) \left\{ \frac{n^{1/p}}{\left(\sqrt{2\pi n/p}(n/(ep))^{n/p}(1 + O(n^{-1}))\right)^{1/n}} \right\} \\ &\times \left\{ n \left(\frac{\sqrt{n}}{\sqrt{2\pi (n-1)}((n-1)/e)^{n-1}(1 + O(n^{-1}))}\right)^{1/(n-1)} \right\}^{-1} \\ &= 2\mu_p^{1/p} \Gamma\left(1 + \frac{1}{p}\right) \frac{(ep)^{1/p}}{e} \frac{\left(1 + O(n^{-1})\right)^{1/n}}{(n/p)^{1/(2n)}(1 + O(n^{-1}))^{1/n}} = A_p \left(1 + O(n^{-1}\log n)\right), \end{split}$$

where we used that $(n/p)^{1/(2n)} = 1 + O(n^{-1} \log n)$ in the last step.

Next, we rewrite $|\Sigma_{n-1} \cap s\mathbb{D}_p^n|_{n-1}$ in probabilistic terms by using the definition of s_n :

$$\begin{split} \left| \Sigma_{n-1} \cap s \mathbb{D}_{p}^{n} \right|_{n-1} \\ &= \left| \left\{ z \in \Sigma_{n-1} : z \in s_{n} \frac{A_{p,n}}{A_{p}} \mathbb{D}_{p}^{n} \right\} \right|_{n-1} \\ &= \left| \left\{ z \in \Sigma_{n-1} : z \in s_{n} n^{\frac{1}{p}-1} \mu_{p}^{1/p} A_{p}^{-1} \frac{|\mathbb{B}_{p}^{n}|_{n}^{1/n}}{|\Sigma_{n-1}|_{n-1}^{1/(n-1)}} \mathbb{D}_{p}^{n} \right\} \right|_{n-1} \\ &= \left| \left\{ z \in |\tilde{\Delta}_{n-1}|_{n-1}^{-1/(n-1)} \tilde{\Delta}_{n-1} : z \in s_{n} n^{\frac{1}{p}-1} \mu_{p}^{1/p} A_{p}^{-1} | \tilde{\Delta}_{n-1} |_{n-1}^{-1/(n-1)} \mathbb{B}_{p}^{n} \right\} \right|_{n-1} \end{split}$$

$$= \frac{\left|\left\{z \in \tilde{\Delta}_{n-1} : z \in s_n n^{\frac{1}{p}-1} \mu_p^{1/p} A_p^{-1} \mathbb{B}_p^n\right\}\right|_{n-1}}{\left|\tilde{\Delta}_{n-1}\right|_{n-1}}$$
$$= \mathbb{P}(\|Z_n\|_p \le s_n n^{\frac{1}{p}-1} \mu_p^{1/p} A_p^{-1}),$$

where Z_n is uniformly distributed in $\tilde{\Delta}_{n-1}$. Now, we observe that

$$\mathbb{P}\Big(\|Z_n\|_p \le s_n n^{\frac{1}{p}-1} \mu_p^{1/p} A_p^{-1}\Big) = \mathbb{P}\Big(\sqrt{n} \Big(n^{1-\frac{1}{p}} \|Z_n\|_p \mu_p^{-1/p} - 1\Big) \le \sqrt{n} \Big(s_n A_p^{-1} - 1\Big)\Big)$$

This allows us to apply the central limit theorem from Corollary 1.1. Indeed, if $sA_p^{-1} > 1$ or $sA_p^{-1} < 1$, then $\sqrt{n}(s_nA_p^{-1} - 1) \to +\infty$ or $-\infty$ respectively as $n \to \infty$. In the equality case $sA_p^{-1} = 1$, since $s_n = s(1 + O(n^{-1}\log n))$ for arbitrary $\varepsilon > 0$, we have $\sqrt{n}(s_nA_p^{-1} - 1) = \sqrt{n}(sA_p^{-1}(1 + O(n^{-1}\log n)) - 1) = O(n^{-1/2}\log n) \to 0$ as $n \to \infty$. The proof is thus complete.

It is interesting to know the value of *s* in the critical case, i.e. $s = A_p$, for different choices of *p*. For example, if p = 1, which corresponds to the intersection of the simplex with the crosspolytope, we have that $s = A_1 = 2\mu_1 = 4/e \approx 1.4715$. For p = 2, which corresponds to the intersection of the simplex with the Euclidean ball, we obtain $s = A_2 = \sqrt{2\pi/e} \approx 1.5203$.

Remark 4.1 As in the case of ℓ_p -balls (see [15, Corollary 2.2]) one can show that for any $r \in (0, 1)$ there exists a sequence $(s_n^{(r)})_{n \in \mathbb{N}}$ for which $|\Sigma_{n-1} \cap s_n^{(r)} \mathbb{D}_p^n|_{n-1} \to r$ as $n \to \infty$.

Acknowledgements

We gratefully acknowledge the suggestions of an anonymous referee which improved the exposition.

Funding information

ZK has been supported by the German Research Foundation under Germany's Excellence Strategy EXC 2044 – 390685587, Mathematics Münster: Dynamics - Geometry - Structure. JP and MS are supported by the Austrian Science Fund (FWF) Project P32405 Asymptotic geometric analysis and applications and MS also received support from Project FF5513-N26, which is part of the Special Research Program *Quasi-Monte Carlo Methods: Theory and Applications*. Most of this work was done while AB was a visiting PhD student at the University of Graz and we thank the RTG 2131 *High-dimensional phenomena in probability – Fluctuations and discontinuity* for the financial support.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- ALONSO-GUTIÉRREZ, D., PROCHNO, J. AND THÄLE, C. (2018). Large deviations for high-dimensional random projections of lⁿ_p-balls. Adv. Appl. Math. 99, 1–35.
- [2] ALONSO-GUTIÉRREZ, D., PROCHNO, J. AND THÄLE, C. (2019). Gaussian fluctuations for high-dimensional random projections of lⁿ_p-balls. *Bernoulli* 25, 3139–3174.

- [3] ALONSO-GUTIÉRREZ, D., PROCHNO, J. AND THÄLE, C. (2021). Large deviations, moderate deviations, and the KLS conjecture. J. Funct. Anal. 280, 108779.
- [4] DasGupta, A. (2008). Asymptotic Theory of Statistics and Probability. Springer, New York.
- [5] DAVID, H. A. AND NAGARAJA, H. N. (2003). Order Statistics, 3rd edn. Wiley-Interscience, Hoboken, NJ.
- [6] DEMBO, A. AND ZEITOUNI, O. (2010). Large Deviations Techniques and Applications, 2nd edn. Springer, Berlin.
- [7] DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. Ann. Prob. 9, 860–867.
- [8] DIACONIS, P. AND FREEDMAN, D. (1987). A dozen de Finetti-style results in search of a theory. Ann. Inst. H. Poincaré Prob. Statist. 23, 397–423.
- [9] DOES, R. J. M. M. AND KLAASSEN, C. A. J. (1984). The Berry-Esseen theorem for functions of uniform spacings. Z. Wahrscheinlichkeitsth. 65, 461–471.
- [10] GANTERT, N., KIM, S. S. AND RAMANAN, K. (2016). Cramér's theorem is atypical. In Advances in the Mathematical Sciences (Assoc. Women Math. Ser. 6), eds G. LETZTER et al. Springer, Cham, pp. 253–270.
- [11] GANTERT, N. Kim, S. S. and Ramanan, K. (2017). Large deviations for random projections of l^p balls. Ann. Prob. 45, 4419–4476.
- [12] GIULIANO, R. AND MACCI, C. (2014). Large deviation principles for sequences of maxima and minima. Commun. Statist. Theory Meth. 43, 1077–1098.
- [13] HOLST, L. (1979). Asymptotic normality of sum-functions of spacings. Ann. Prob. 7, 1066–1072.
- [14] JOHNSTON, S. G. G. AND PROCHNO, J. (2019). Berry–Esseen bounds for random projections of ℓ_p^n -balls. Preprint, arXiv:1911.00695.
- [15] KABLUCHKO, Z., PROCHNO, J. AND THÄLE, C. (2019). High-dimensional limit theorems for random vectors in lⁿ_p-balls. Commun. Contemp. Math. 21, 1750092.
- [16] KABLUCHKO, Z., PROCHNO, J. AND THÄLE, C. (2020). Sanov-type large deviations in Schatten classes. Ann. Inst. Henri Poincaré Prob. Stat. 56, 928–953.
- [17] KABLUCHKO, Z., PROCHNO, J. AND THÄLE, C. (2021). High-dimensional limit theorems for random vectors in lⁿ_p-balls. II. Commun. Contemp. *Math.* 23, 1950073.
- [18] KIM, S. S. AND RAMANAN, K. (2018). A conditional limit theorem for high-dimensional l^p-spheres. J. Appl. Prob. 55, 1060–1077.
- [19] KIM, S. S. AND RAMANAN, K. (2019). An asymptotic thin shell condition and large deviations for multidimensional projections. Preprint, arXiv:1912.13447.
- [20] KLARTAG, B. (2007). A central limit theorem for convex sets. Invent. Math. 168, 91–131.
- [21] LIAO, Y.-T. AND RAMANAN, K. (2020). Geometric sharp large deviations for random projections of ℓ_p^n spheres. Preprint, arXiv:2001.04053.
- [22] MATHAI, A. M. (1999). An Introduction to Geometrical Probability. Gordon and Breach, Amsterdam.
- [23] MIRAKHMEDOV, S. A. (2005). Lower estimation of the remainder term in the CLT for a sum of the functions of k-spacings. *Statist. Prob. Lett.* 73, 411–424.
- [24] PAOURIS, G., PIVOVAROV, P. AND ZINN, J. (2014). A central limit theorem for projections of the cube. Prob. Theory Relat. Fields 159, 701–719.
- [25] SCHECHTMAN, G. AND SCHMUCKENSCHLÄGER, M. (1991). Another remark on the volume of the intersection of two Lⁿ_p balls. In *Geometric Aspects of Functional Analysis (1989–90)* (Lect. Notes Math. 1469), eds J. LINDENSTRAUSS AND V. D. MILMAN. SPRINGER, BERLIN, pp. 174–178.
- [26] SCHECHTMAN, G. AND ZINN, J. (1990). On the volume of the intersection of two L_p^n balls. *Proc. Amer. Math.* Soc. **110**, 217–224.
- [27] SCHMUCKENSCHLÄGER, M. (2001). CLT and the volume of intersections of l_p^n -balls. Geom. Dedicata 85, 189–195.
- [28] VERSHYNIN, R. (2018). High-Dimensional Probability. Cambridge University Press.