J. Appl. Prob. **55,** 1060–1077 (2018)
*doi:*10.1017/jpr.2018.71
© *Applied Probability Trust* 2018

A CONDITIONAL LIMIT THEOREM FOR HIGH-DIMENSIONAL ℓ^p-SPHERES

STEVEN S. KIM * ** AND
KAVITA RAMANAN,* *** *Brown University*

Abstract

The study of high-dimensional distributions is of interest in probability theory, statistics, and asymptotic convex geometry, where the object of interest is the uniform distribution on a convex set in high dimensions. The ℓ^p-spaces and norms are of particular interest in this setting. In this paper we establish a limit theorem for distributions on ℓ^p-spheres, conditioned on a rare event, in a high-dimensional geometric setting. As part of our proof, we establish a certain large deviation principle that is also relevant to the study of the tail behavior of random projections of ℓ^p-balls in a high-dimensional Euclidean space.

Keywords: Conditional limit theorem; ℓ^p-sphere; large deviation; Gibbs conditioning principle; cone measure; surface measure; relative entropy; Wasserstein topology

2010 Mathematics Subject Classification: Primary 60F10; 52A20
Secondary 52A23; 60D05

1. Introduction

The classical Poincaré–Maxwell–Borel lemma (see, e.g. [15, Section 6.1] or [33]) states that for any fixed number of coordinates of a random vector drawn uniformly from a high-dimensional Euclidean sphere, the joint distribution of the coordinates is approximately standard Gaussian. An analogous description (recalled precisely in Proposition 1) holds in the setting of ℓ^p-spheres for general $p \in [1, \infty]$, which yields an asymptotic probabilistic characterization of the geometry of an ℓ^p-sphere.

In this paper we establish a *conditional* analog of the aforementioned *unconditional* result. Roughly stated, we prove the following asymptotic probabilistic description of the geometry of an ℓ^p-sphere under an ℓ^q-norm constraint. For $1 \le q < p < \infty$, in high dimensions, a random point on the ℓ^p-sphere of $\mathbb{R}^n$ conditioned on having a sufficiently small ℓ^q-norm is close (in the sense of distribution) to a random point drawn from an appropriately scaled ℓ^q-sphere of $\mathbb{R}^n$. That is, conditioning on a small ℓ^q-norm induces a *probabilistic* change that admits a *geometric* interpretation.

As part of our proof, we establish a large deviation principle (LDP) for the empirical measure of the coordinates of a random vector distributed according to the uniform (cone or surface) measure on a suitably scaled ℓ^p-sphere in $\mathbb{R}^n$, for $p \in [1, \infty]$, with respect to the q-Wasserstein topology for every $q < p$. This LDP is also of relevance to the study of the tail behavior of random projections of ℓ^p-balls.

The remainder of the paper is organized as follows. In Section 2 we precisely state our main result (Theorem 1), a conditional limit theorem for ℓ^p-spheres. In Section 3 we establish an

Received 7 February 2017; revision received 8 October 2018.
* Postal address: Brown University, 182 George Street, Box F, Providence, RI 02912, USA.
** Email address: steven_kim@alumni.brown.edu
*** Email address: kavita_rananan@brown.edu

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

LDP for the sequence of empirical measures of the coordinates of the random vector $n^{1/p}X^{(n,p)}$ distributed uniformly on the sphere $n^{1/p}\mathsf{S}_p^{n-1}$. We emphasize that the aforementioned LDP holds in the q-Wasserstein topology for every $q < p$ (which is stronger than the usual weak topology), and as summarized in Remark 1, we crucially utilize this fact in order to prove Proposition 5, as well as to establish a variational formula in a related paper; see [19, Theorem 2.7]. In Section 4 we apply the LDPs of Section 3 and recall the 'Gibbs conditioning principle' in order to establish the proof of Theorem 1.

2. Statement of main result

To precisely state the unconditional limit theorem for ℓ^p-spheres, for $n \in \mathbb{N}$, let $\|\cdot\|_{n,p}$ denote the ℓ^p-norm on $\mathbb{R}^n$, and write $\mathsf{S}_p^{n-1} := \{a \in \mathbb{R}^n : \|a\|_{n,p} = 1\}$ for the unit ℓ^p-sphere in $\mathbb{R}^n$. There are two natural notions of 'uniform' measure on the sphere S_p^{n-1}: the *cone measure* $\gamma_{n,p}$ and the *surface measure* $\sigma_{n,p}$, defined in Definitions 5 and 6, respectively. Assume that all random variables we define are supported on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1. For $n \in \mathbb{N}$, $p \in [1, \infty]$, we write $X^{(n,p)} = (X_1^{(n,p)}, \ldots, X_n^{(n,p)})$ for an n-dimensional random vector in S_p^{n-1}, distributed according to either the surface measure $\sigma_{n,p}$ or the cone measure $\gamma_{n,p}$ (that is, all of our results hold under either distribution for $X^{(n,p)}$, unless explicitly specified otherwise).

Let μ_p denote the *generalized Gaussian* distribution with location 0, shape p, and scale $p^{1/p}$. That is, for $p \in [1, \infty)$,

$$\mu_p(\mathrm{d}y) := \frac{1}{2p^{1/p}\Gamma(1 + 1/p)} \mathrm{e}^{-|y|^p/p}\,\mathrm{d}y, \qquad y \in \mathbb{R}. \tag{1}$$

For $p = \infty$, define

$$\mu_\infty(\mathrm{d}y) = \tfrac{1}{2}\mathbf{1}_{[-1,1]}(y)\,\mathrm{d}y, \qquad y \in \mathbb{R}.$$

The aforementioned 'ℓ^p version' of the Poincaré–Maxwell–Borel lemma is stated as follows. Let '$\xrightarrow{\mathrm{w}}$' denote weak convergence of probability measures, that is, given probability measures $\mu, \mu_n, n \in \mathbb{N}$, on $\mathbb{R}$, the notation $\mu_n \xrightarrow{\mathrm{w}} \mu$ implies that $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$ for every bounded continuous function on $\mathbb{R}$.

Proposition 1. *Fix $p \in [1, \infty]$ and $k \in \mathbb{N}$. As $n \to \infty$,*

$$\mathrm{law}[n^{1/p}(X_1^{(n,p)}, \ldots, X_k^{(n,p)})] \xrightarrow{\mathrm{w}} \mu_p^{\otimes k}.$$

This result is due to [30, Theorems 4.1 and 4.4] in the case of the cone measure, and [27] in the case of the surface measure. In addition, Naor and Romik [29, Theorems 3 and 4] offer a simplification of the proof for both the cone and surface measures.

Note that the scaling $n^{1/p}$ is natural from a geometric perspective due to the following reasoning: the typical coordinate of the random vector $X^{(n,p)}$ (uniformly distributed on the unit ℓ^p-sphere of dimension n) has length $n^{-1/p}$ (since the n-fold sum of pth powers of the coordinates is constrained to sum to 1). Therefore, in order for the joint law of the coordinates of $X^{(n,p)}$ to be nontrivial, one must scale the random vector $X^{(n,p)}$ appropriately; that is, by a factor of $n^{1/p}$.

Our main result (Theorem 1) yields a conditional analog of Proposition 1. To state this precisely, we first set some notation. For $\beta > 0$ and $q \in [1, \infty]$, let

$$\lambda_{n,q}^{(\beta,k)} := \mathrm{law}[\beta^{1/q} n^{1/q}(X_1^{(n,q)}, \ldots, X_k^{(n,q)})]. \tag{2}$$

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

In addition, let $\widehat{\lambda}_{n,p|q}^{(\beta,k)}$ denote the conditional law, that is,

$$\widehat{\lambda}_{n,p|q}^{(\beta,k)} := \mathbb{P}\left(n^{1/p}(X_1^{(n,p)}, \ldots, X_k^{(n,p)}) \in \cdot \;\middle|\; \frac{1}{n}\|n^{1/p} X^{(n,p)}\|_{n,q}^q \le \beta \right). \tag{3}$$

For $1 \le q < p < \infty$, define

$$\beta_{p,q} := \frac{1}{q}\left(\frac{\Gamma(1/q)}{\Gamma((p+1)/q)} \right)^{q/p}. \tag{4}$$

Finally, let ρ be a metric that metrizes the topology of weak convergence of probability measures (e.g. Lévy–Prohorov, bounded Lipschitz); see [7, Appendix III].

Theorem 1. (Conditional limit theorem.) *Assume that* $1 \le q < p < \infty$ *and* $\beta \le \beta_{p,q}$. *Then, for fixed* $k \in \mathbb{N}$,

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} \rho\big(\widehat{\lambda}_{n,p|q}^{(\beta+\varepsilon,k)}, \lambda_{n,q}^{(\beta,k)}\big) = 0.$$

We defer the proof of Theorem 1 to Section 4. It is worth emphasizing that a suitable choice of the conditioning event, as in (3), is required to obtain a meaningful conditional limit theorem that has a geometric interpretation. As discussed above, the typical coordinate of the random vector $X^{(n,p)}$ has length $n^{-1/p}$, and so $(1/n)\|n^{1/p} X^{(n,p)}\|_{n,q}^q$ is typically an order 1 quantity. Thus, for all sufficiently small β, the event $(1/n)\|n^{1/p} X^{(n,p)}\|_{n,q}^q \le \beta$ would be expected to be a rare event, in that its probability decays to 0. In Theorem 1 we show that for the subset of β that lies in the interval $(0, \beta_{p,q}]$, the conditioning event

$$A_{n,\beta} := \left\{ a \in \mathbb{R}^n : \frac{1}{n}\|a\|_{n,q}^q \le \beta \right\} \tag{5}$$

provides a geometric example of a *rare event* in high dimensions for which a geometric conditional limit theorem can be established. A basic intuition from the theory of large deviations suggests that for large n, given a rare event such as $A_{n,\beta+\varepsilon}$, the conditional law will be close to a 'tilted' measure under which the event $A_{n,\beta+\varepsilon}$ is, in fact, typical. The goal of finding a measure under which the event $A_{n,\beta+\varepsilon}$ is typical motivates the scaling $\beta^{1/q}$ and the random variable $X^{(n,q)}$ in (2). That is, it follows from the definition of $\sigma_{n,q}$ and $\gamma_{n,q}$ that for all $n \in \mathbb{N}$ and $\varepsilon > 0$, $\mathbb{P}(\beta^{1/q} n^{1/q} X^{(n,q)} \in A_{n,\beta+\varepsilon}) = 1$. Although the probability of the conditioned event $\mathbb{P}(n^{1/p} X^{(n,p)} \in A_{n,\beta+\varepsilon})$ can decay exponentially to 0 even for some $\beta > \beta_{p,q}$, as explained in Remarks 6–8, the restriction $\beta \le \beta_{p,q}$ allows for the conditioned event to admit the geometric interpretation, of converging to the cone (or surface) measure on a scaled ℓ_q-sphere in $\mathbb{R}^n$.

These general ideas of conditioning and tilting have been realized to great effect in statistical mechanics, where large deviations theory can describe the most probable state of a system of particles under an energy constraint; see, e.g. the surveys [6] and [17]. Of particular utility is the so-called Gibbs conditioning principle, which transforms an LDP for empirical measures into a statement about the most probable behavior of the underlying sequence of random variables, conditional on a rare event. A central motivation of this work is to employ such a conditional probabilistic perspective in a high-dimensional geometric setting by investigating how LDPs can inform the analysis of 'geometric' rare events in high dimensions, such as $A_{n,\beta}$ of (5).

More generally, we continue a body of research that lies at the intersection of large deviations theory and high-dimensional geometry. Properties of finite-dimensional ℓ^p-spaces have been

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

well studied for many decades since they play an important role in the geometry of Banach spaces. Recent years have seen a renewed interest in their study in high dimensions, which falls under the rubric of asymptotic geometric analysis, and is often facilitated by a probabilistic perspective; see, e.g. [3] and [32]. While central limit theorems for projections of ℓ^p-balls (and more general convex sets) were established in several previous works (see, e.g. [23]), a new large deviations perspective was adopted in [18] and [19]. Specifically, Gantert *et al.* [18] and[19] showed that the sequence of one-dimensional scalar projections of a random vector drawn uniformly from an ℓ^p-ball in $\mathbb{R}^n$ satisfies (both quenched and annealed) LDPs when $p \in [1, \infty]$. Subsequently, an LDP for the norms of k_n-dimensional projections of these random vectors, when k_n grows with n, was established in [1], and an LDP for the empirical measure of the coordinates of such k_n-dimensional projections was established in [21] and [22]. The latter result holds, in fact, for a more general class of random vectors, which includes vectors uniformly distributed in a large class of sequences of Orlicz balls and certain Gibbs distributions. Subsequent to the present work, a multivariate central limit theorem for ℓ^q-norms of random vectors from ℓ_p-balls, and some related large deviation results were also established in [20].

3. Large deviations results

In this section we state and prove an LDP for the following empirical measure: for $n \in \mathbb{N}$, $p \in [1, \infty]$, define

$$\mathcal{L}_{n,p} := \frac{1}{n} \sum_{i=1}^{n} \delta_{n^{1/p} X_i^{(n,p)}}. \tag{6}$$

In Section 3.1 we review some basic elements of large deviation theory. In Section 3.2 (respectively, Section 3.3) we prove the LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ in the Wasserstein topology, assuming $X^{(n,p)}$ is distributed according to the cone measure (respectively, the surface measure). This Sanov-type LDP complements the existing Glivenko–Cantelli-type law of large numbers and Donsker-type central limit theorem for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ under the cone measure [32, Theorem 1].

3.1. Background on large deviations

For a broad review of large deviations, we refer the reader to [14]. In particular, we recall the following definition.

Definition 2. Let $\mathcal{X}$ be a topological space equipped with its Borel σ-algebra and let $\mathcal{P}(\mathcal{X})$ denote the space of probability measures on $\mathcal{X}$. A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to satisfy an LDP with a *rate function* $I\colon \mathcal{X} \to [0, \infty]$ if $I(\cdot)$ is lower semicontinuous, and for all Borel-measurable sets $\Gamma \subset \mathcal{X}$,

$$-\inf_{x\in\Gamma^\circ} I(x) \le \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(\Gamma^\circ) \le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(\bar{\Gamma}) \le -\inf_{x\in\bar{\Gamma}} I(x),$$

where Γ° and $\bar{\Gamma}$ denote the interior and closure of Γ, respectively. Furthermore, I is said to be a *good rate function* if it has compact level sets. Similarly, we say that the sequence of $\mathcal{X}$-valued random variables $(\xi_n)_{n\in\mathbb{N}}$ satisfies an LDP if the sequence of laws $(\mathbb{P} \circ \xi_n^{-1})_{n\in\mathbb{N}}$ satisfies an LDP.

In the remaining sections, we will also make occasional reference to the following related notion.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Definition 3. A sequence of probability measures $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is said to be *exponentially tight* if, for every $\alpha < \infty$, there exists a compact set $K_\alpha \subset \mathcal{X}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_\alpha^c) < -\alpha.$$

There are two particularly well-known examples of LDPs. First, *Cramér's theorem* (see, e.g. [14, Section 2.2.2]) establishes an LDP for the sum of independent and identically distributed (i.i.d.) random variables, with the rate function shown to be the so-called Legendre transform of the log moment-generating function of the underlying distribution. Second, *Sanov's theorem* (see, e.g. [14, Section 6.2]) establishes an LDP for the empirical measure of i.i.d. random variables, with the rate function shown to be the so-called relative entropy.

For $\nu, \mu \in \mathcal{P}(\mathcal{X})$, recall that the *relative entropy* of ν with respect to μ is defined as

$$H(\nu \,\|\, \mu) := \begin{cases} \displaystyle\int_{\mathbb{R}} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\nu & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise,} \end{cases} \tag{7}$$

where $\nu \ll \mu$ denotes that ν is absolutely continuous with respect to μ. For $q \in [1, \infty)$, define the qth absolute moment map as

$$m_q(\nu) := \int_{\mathbb{R}} |x|^q \nu(\mathrm{d}x), \qquad \nu \in \mathcal{P}(\mathbb{R}). \tag{8}$$

For $q = \infty$, let

$$m_\infty(\nu) := \inf\{a > 0 \colon \nu([-a, a]^c) = 0\}, \qquad \nu \in \mathcal{P}(\mathbb{R}).$$

For $p \in [1, \infty]$, define $H_p \colon \mathcal{P}(\mathbb{R}) \to [0, \infty]$ to be a modification of the relative entropy with respect to μ_p, perturbed by some pth moment penalty,

$$H_p(\nu) := \begin{cases} H(\nu \,\|\, \mu_p) + \dfrac{1}{p}(1 - m_p(\nu)) & \text{if } m_p(\nu) \le 1, \\ +\infty & \text{otherwise,} \end{cases} \tag{9}$$

where we take the convention $1/\infty = 0$. We will show in Propositions 2 and 4 that $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP with rate function H_p.

Our LDP holds with respect to the so-called q-Wasserstein topologies on $\mathcal{P}_q(\mathbb{R})$, introduced below.

Definition 4. Let $q \in [1, \infty]$ and

$$\mathcal{P}_q(\mathbb{R}) := \{\mu \in \mathcal{P}(\mathbb{R}) \colon m_q(\mu) < \infty\}.$$

We equip this space with the *q-Wasserstein* topology: a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_q(\mathbb{R})$ is said to converge to $\mu \in \mathcal{P}_q(\mathbb{R})$ (with respect to the q-Wasserstein topology) if μ_n converges weakly to μ (denoted by $\mu_n \xrightarrow{\mathrm{w}} \mu$) and $m_q(\mu_n) \to m_q(\mu)$ as $n \to \infty$.

Remark 1. For an extensive review of the Wasserstein topology and, in particular, to see that the definition given above is equivalent to alternative definitions of the q-Wasserstein topology, see [37, Section 6]. We consider the q-Wasserstein topology on probability measures (instead

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

of, e.g. the weak topology or the τ-topology) because we must consider a topology that is weak enough to allow an LDP to hold, but at the same time strong enough to allow certain moments to be continuous functionals of measures. In particular, we require a topology strong enough such that the moment map m_q is continuous for $q < p$, which we use in the proofs of both the Gibbs conditioning result of Proposition 5 and the variational formula of [19] that relates 'quenched' and 'annealed' LDP rate functions for random projections of ℓ^p-balls.

3.2. LDP under the cone measure

In this section we establish an LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ assuming $X^{(n,p)}$ is distributed according to the cone measure on S_p^{n-1} defined below.

Definition 5. Let $\mathrm{vol}_n(\cdot)$ denote the Lebesgue measure on $\mathbb{R}^n$, and let $\gamma_{n,p}$ denote the *cone measure* on S_p^{n-1}, that is,

$$\gamma_{n,p}(S) := \frac{\mathrm{vol}_n(\{cx : x \in S,\ c \in [0,1]\})}{\mathrm{vol}_n(\mathbb{B}_{n,p})} \quad \text{for } S \text{ a Borel-measurable subset of } \mathsf{S}_p^{n-1}.$$

Proposition 2. (LDP under the cone measure.) *Let $p \in [1,\infty]$ and assume that $X^{(n,p)} \sim \gamma_{n,p}$. For $q < p$, the sequence of empirical measures $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ of (6) satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ with the strictly convex good rate function H_p of (9).*

The proof of Proposition 2 is given at the end of this subsection.

Remark 2. The primary obstacle in the proof of Proposition 2 (which differentiates it from the classical i.i.d. setting of Sanov's theorem) is that the coordinates $X_i^{(n,p)}$, $i = 1,\dots,n$, are *dependent* by virtue of the fact that the vector $X^{(n,p)}$ is constrained to lie in the sphere S_p^{n-1}. Although LDPs for empirical measures of random vectors with other dependency structures have been established (such as in [14, Section 6.3–6.6] for Markov chains or stationary sequences satisfying certain mixing conditions, or [5] for empirical spectral measures of random matrices), none of these apply in our setting.

Remark 3. It is natural to try to establish an LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ by exploiting the fact that the sequence $(X^{(1,p)}, X^{(2,p)}, \dots)$ can be viewed as an infinite triangular array, where each row $X^{(n,p)}$, for $n \in \mathbb{N}$, is an exchangeable vector. However, such an attempt fails. Indeed, in [36, p. 653], which contains general large deviations results for the empirical measure of a row-wise exchangeable triangular array, it was stated that: 'Even in the simple case of binary valued finite exchangeable random variables there is no general result for the LD behavior of [the row-wise empirical measure].' The aforementioned quote is in reference to the result [36, Theorem 2]. That is, an exchangeable structure on its own is not sufficient for a general LDP result (and, in particular, not sufficient for an LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$). Instead, we take a different route, which relies on the following preliminary results.

The analysis of the LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ is facilitated by the following probabilistic representation for cone measures.

Lemma 1. ([30, Section 3] and [31, Lemma 1].) *Fix $n \in \mathbb{N}$ and $p \in [1,\infty]$. Suppose that $X^{(n,p)} \sim \gamma_{n,p}$, and let $Y^{(n,p)} \sim \mu_p^{\otimes n}$. Then*

$$X^{(n,p)} \overset{\mathrm{D}}{=} \frac{Y^{(n,p)}}{\|Y^{(n,p)}\|_{n,p}}.$$

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Some extensions of Lemma 1 can also be found in [3]. The preceding representation is used in the proof of the following result, which generalizes a result from [4, Theorem 6.6] beyond the special case where $p = 2$ (for which the cone and surface measures coincide, and the surface measure can be expressed in terms of i.i.d. Gaussians). Specifically, the preceding representation of the cone measure in terms of product measures allows us to use the approach of [4, Theorem 6.6]; namely, we establish an LDP for sums of i.i.d. random variables in the space $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$. We then derive a more interpretable representation of the rate function via the Donsker–Varadhan formula for relative entropy (see Lemma 2).

Proposition 3. *Let $p \in [1, \infty]$ and assume that $X^{(n,p)} \sim \gamma_{n,p}$. The sequence $(\mathcal{L}_{n,p})_{n\in N}$ of (6) satisfies an LDP in $\mathcal{P}(\mathbb{R})$ equipped with the weak topology, with the good rate function H_p of (9).*

To prove Proposition 3, we appeal to the following representation of relative entropy; recall the definition of relative entropy given in (7).

Lemma 2. (Donsker–Varadhan; see, e.g. [16, Lemma 1.4.3(a)].) *Let $C_b(\mathcal{X})$ denote the space of bounded continuous functions from $\mathcal{X}$ into $\mathbb{R}$. Then, for all $\nu, \mu \in \mathcal{P}(\mathcal{X})$, we have*

$$H(\nu \,\|\, \mu) = \sup_{f \in C_b(\mathbb{R})} \left\{ \int_{\mathcal{X}} f \, \mathrm{d}\nu - \log \int_{\mathcal{X}} \mathrm{e}^f \, \mathrm{d}\mu \right\}.$$

Proof of Proposition 3. Let $Y^{(n,p)} \sim \mu_p^{\otimes n}$ be as in Lemma 1, and let $L^Y_{n,p}$ denote the empirical measure of $Y^{(n,p)}$, that is

$$L^Y_{n,p} := \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i^{(n,p)}}.$$

Since $(L^Y_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ by Sanov's theorem (see, e.g. [14, Section 6.2]), $(L^Y_{n,p})_{n\in\mathbb{N}}$ is also exponentially tight (recall Definition 3, and see, e.g. [26, Lemma 2.6]). Furthermore, due to the finite exponential moment $\mathbb{E}[\exp(t\|Y^{(n,p)}\|^p_{n,p})] < \infty$ for $|t| < 1/p$, for m_p as in (8), the sequence $(m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$ satisfies an LDP in $\mathbb{R}$ by Cramér's theorem (see, e.g. [14, Section 2.2.2]), and, hence, $(m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$ is also exponentially tight. As a consequence, the joint sequence $(L^Y_{n,p}, m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$ is also exponentially tight with respect to the product topology on $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$.

We claim (and prove below) that the joint sequence $(L^Y_{n,p}, m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$ with the convex good rate function J defined as follows: for $\lambda \in \mathcal{P}(\mathbb{R})$ and $c \in \mathbb{R}_+$, let

$$J(\lambda, c) := \sup_{f \in C_b(\mathbb{R}),\, t\in\mathbb{R}} \left\{ \int_{\mathbb{R}} f(y)\lambda(\mathrm{d}y) + tc - \log \int_{\mathbb{R}} \mathrm{e}^{f(y)+t|y|^p} \mu_p(\mathrm{d}y) \right\}, \tag{10}$$

where $C_b(\mathbb{R})$ denotes the space of all bounded continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Due to Cramér's theorem for general Polish spaces (see, e.g. [14, Theorem 6.1.3]), the sequence $(L^Y_{n,p}, m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$ satisfies a *weak* LDP in the space $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$ with rate function J of (10); that is, the LDP upper bound in Definition 2 holds only for compact sets (see [14, p. 7] for a precise definition of a weak LDP). The full LDP with good rate function J then follows as a consequence of the weak LDP combined with the previously established exponential tightness (see, e.g. [14, Lemma 1.2.18]).

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

We now derive an alternative form of the rate function J defined in (10). Note that for $t \geq 1/p$, we have $\log \int_{\mathbb{R}} \mathrm{e}^{f(y)+t|y|^p} \mu_p(\mathrm{d}y) = +\infty$. Define $Z_{p,t} \coloneqq \int_{\mathbb{R}} \mathrm{e}^{t|y|^p} \mu_p(\mathrm{d}y)$ and note that for $t < 1/p$, $Z_{p,t}$ is finite due to the definition of μ_p from (1), and serves as a normalizing constant that makes the measure

$$\tilde{\mu}_{p,t}(\mathrm{d}y) \coloneqq \frac{1}{Z_{p,t}} \mathrm{e}^{t|y|^p} \mu_p(\mathrm{d}y)$$

a probability measure. Using the fact that $Z_{p,t}$ is infinite for $t > 1/p$ and the definition of relative entropy in (7), we can then rewrite the functional $J(\lambda, c)$ defined in (10) as follows: for $(\lambda, c) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$,

$$\begin{aligned} J(\lambda, c) &= \sup_{t<1/p,\, f \in C_{\mathrm{b}}(\mathbb{R})} \left\{ \int_{\mathbb{R}} f(y)\lambda(\mathrm{d}y) + tc - \log Z_{p,t} - \log \int_{\mathbb{R}} \mathrm{e}^{f(y)} \tilde{\mu}_{p,t}(\mathrm{d}y) \right\} \\ &= \sup_{t<1/p} \left\{ tc - \log Z_{p,t} + \sup_{f \in C_{\mathrm{b}}(\mathbb{R})} \left\{ \int_{\mathbb{R}} f(y)\lambda(\mathrm{d}y) - \log \int_{\mathbb{R}} \mathrm{e}^{f(y)} \tilde{\mu}_{p,t}(\mathrm{d}y) \right\} \right\}. \end{aligned}$$

Next, applying the Donsker–Varadhan variational formula for relative entropy (as stated in Lemma 2) to the space $\mathcal{X} = \mathbb{R}$, the bounded continuous function f and the probability measure $\tilde{\mu}_{p,t}$, we find that, for $(\lambda, c) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$,

$$\begin{aligned} J(\lambda, c) &= \sup_{t<1/p} \{tc + H(\lambda \,\|\, \tilde{\mu}_{p,t}) - \log Z_{p,t}\} \\ &= H(\lambda \,\|\, \mu_p) + \sup_{t<1/p} \left\{ tc + \int_{\mathbb{R}} \log \frac{\mathrm{d}\mu_p}{\mathrm{d}\tilde{\mu}_{p,t}} \lambda(\mathrm{d}y) - \log Z_{p,t} \right\} \\ &= H(\lambda \,\|\, \mu_p) + \sup_{t<1/p} \left\{ tc - \int_{\mathbb{R}} t|y|^p \lambda(\mathrm{d}y) + \log Z_{p,t} - \log Z_{p,t} \right\} \\ &= H(\lambda \,\|\, \mu_p) + \sup_{t<1/p} \{tc - t m_p(\lambda)\}, \end{aligned}$$

where we used the definition (7) of relative entropy, and the decomposition $\mathrm{d}\lambda/\mathrm{d}\mu_{p,t} = (\mathrm{d}\lambda/\mathrm{d}\mu)(\mathrm{d}\mu/\mathrm{d}\mu_{p,t})$ in the second equality above. Thus, we have

$$J(\lambda, c) = \begin{cases} H(\lambda \,\|\, \mu_p) + \dfrac{1}{p}(c - m_p(\lambda)) & \text{if } m_p(\lambda) \leq c, \\ +\infty & \text{otherwise.} \end{cases} \tag{11}$$

Next, for $(\lambda, c) \in \mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$, define $G_p(\lambda, c) \coloneqq \lambda(\cdot \times c^{1/p})$. An elementary consequence of Slutsky's theorem is that the map $G_p \colon \mathcal{P}(\mathbb{R}) \times \mathbb{R}_+ \mapsto \mathcal{P}(\mathbb{R})$ is continuous (with respect to the weak topology on $\mathcal{P}(\mathbb{R})$ and the induced product topology on $\mathcal{P}(\mathbb{R}) \times \mathbb{R}_+$). Using (6) and Lemma 1, $\mathcal{L}_{n,p}$ is equal in distribution to $G_p(L^Y_{n,p}, m_p(L^Y_{n,p}))$, so the contraction principle (see, e.g. [14, Theorem 4.2.1]) and the expression for J in (11) yield an LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ with the good rate function

$$\begin{aligned} J_p(\nu) &\coloneqq \inf\{J(\lambda, c) \colon \lambda \in \mathcal{P}(\mathbb{R}), c \in \mathbb{R}_+, G_p(\lambda, c) = \nu\} \\ &= \begin{cases} \inf\limits_{c \geq 0} \left\{ H(\nu(\cdot \times c^{-1/p}) \,\|\, \mu_p) + \dfrac{1}{p}(c - c m_p(\nu)) \right\} & \text{if } m_p(\nu) \leq 1, \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

If $m_p(\nu) \le 1$ then using the definition of μ_p in (1) to calculate $\mathrm{d}\mu_p/\mathrm{d}\mu_p(\cdot \times c^{1/p})$, and the definition of relative entropy in (7), we have

$$\begin{aligned} J_p(\nu) &= \inf_{c\ge 0}\left\{H(\nu \,\|\, \mu_p(\cdot \times c^{1/p})) + \frac{c}{p}(1 - m_p(\nu))\right\} \\ &= H(\nu \,\|\, \mu_p) + \inf_{c\ge 0}\left\{-\frac{1}{p}\log c - \frac{1-c}{p}m_p(\nu) + \frac{c}{p}(1 - m_p(\nu))\right\} \\ &= H_p(\nu). \end{aligned}$$

□

Remark 4. Note that the above proof of Proposition 3 relies on an LDP for the joint sequence $(L^Y_{n,p}, m_p(L^Y_{n,p}))_{n\in\mathbb{N}}$. An alternative approach to the one presented would be to appeal to results on the 'partial LDP' established for self-normalized processes; see [12, Theorem 1.1].

Remark 5. Note that Proposition 3 cannot be obtained via a direct application of the contraction principle to the map $\nu \mapsto G_p(\nu, m_p(\nu))$. Indeed, there does not appear to be a standard topology on $\mathcal{P}(\mathbb{R})$ such that both the sequence $(L^Y_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP with good rate function *and* the map $m_p(\cdot)$ is continuous with respect to that topology. In particular, $m_p(\cdot)$ is continuous with respect to the r-Wasserstein topology if and only if $r \ge p$; on the other hand, the authors of [25] and [38] showed that the sequence $(L^Y_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP with good rate function with respect to the Wasserstein-r-topology if and only if $r < p$.

Lemma 3. *For $p \in [1,\infty]$, let $K_p := \{\nu \in \mathcal{P}(\mathbb{R}) : m_p(\nu) \le 1\}$. Then, for all $q < p$, the set $K_p \subset \mathcal{P}_q(\mathbb{R})$ is compact in $\mathcal{P}_q(\mathbb{R})$. In addition, K_p is convex and nonempty.*

Proof. The convexity and nonemptiness of K_p are elementary. As for compactness, we first prove that K_p is weakly compact in $\mathcal{P}(\mathbb{R})$. For $\nu \in K_p$ and for all $M > 0$, Chebyshev's inequality yields $\nu([-M, M]^{\mathrm{c}}) \le m_p(\nu)/M^p \le 1/M^p$, so K_p is tight, and by Prokhorov's theorem, precompact. Note that K_p is weakly closed in $\mathcal{P}(\mathbb{R})$ since it is a level set of the lower semicontinuous map m_p. Therefore, K_p is weakly compact.

To verify Wasserstein compactness, it suffices to show that the set of probability measures in K_p have uniformly integrable qth moments; see [2, Proposition 7.1.5]. This latter condition follows from the de la Vallée-Poussin criterion (see, e.g. [11, Theorem II.22]) since, for $g(x) := |x|^{p/q}$ (which satisfies the superlinear growth condition $\lim_{|t|\to\infty} g(t)/t = \infty$), we have

$$\sup_{\nu\in K_p}\int_{\mathbb{R}} g(|x|^q)\nu(\mathrm{d}x) = \sup_{\nu\in K_p} m_p(\nu) \le 1 < \infty.$$

□

Proof of Proposition 2. Fix $p \in [1,\infty]$ and $q < p$. In view of the fact that $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}(\mathbb{R})$ with respect to the weak topology due to Proposition 3, in order to establish the LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ in $\mathcal{P}_q(\mathbb{R})$ with respect to the q-Wasserstein topology, it suffices to show exponential tightness of $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ in $\mathcal{P}_q(\mathbb{R})$ (see [14, Corollary 4.2.6]). Let K_p be the compact set defined in Lemma 3. Note that $m_p(\mathcal{L}_{n,p}) = 1$ almost surely, so $\mathbb{P}(\mathcal{L}_{n,p} \in K_p^{\mathrm{c}}) = 0$ and, hence, $\log \mathbb{P}(\mathcal{L}_{n,p} \in K_p^{\mathrm{c}}) = -\infty$ for all $n \in \mathbb{N}$, implying exponential tightness of $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$.

The strict convexity of H_p follows from the strict convexity of the relative entropy $H(\cdot \,\|\, \mu_p)$ and the linearity of the pth moment m_p. □

3.3. LDP under the surface measure

In this section we establish an LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ assuming $X^{(n,p)}$ is distributed according to the surface measure on S_p^{n-1}, which is defined below.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Definition 6. Let $\text{area}_{n,p}(\cdot)$ denote the (unnormalized) surface area measure on S_p^{n-1}, and let $\sigma_{n,p}$ denote the (normalized) *surface measure* on S_p^{n-1},

$$\sigma_{n,p}(S) := \frac{\text{area}_{n,p}(S)}{\text{area}_{n,p}(\mathsf{S}_p^{n-1})} \quad \text{for } S \text{ a Borel-measurable subset of } \mathsf{S}_p^{n-1}.$$

Proposition 4. (LDP under the surface measure.) *Let $p \in [1,\infty]$ and assume that $X^{(n,p)} \sim \sigma_{n,p}$. For $q < p$, the sequence of empirical measures $(\mathcal{L}_{n,p})_{n\in N}$ of (6) satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ with the strictly convex good rate function H_p of (9).*

The proof of Proposition 4 is deferred to the end of this subsection, after we first establish some preliminary lemmas.

Note that the cone measure coincides with the surface measure on S_p^{n-1} if and only if $p = 1, 2, \infty$. See [28, Section 3] and [30, Section 3] for more extensive discussions. More generally, we have the following relationship between the cone and surface measures.

Lemma 4. ([29, Lemma 2].) *Let $p \in [1,\infty)$. Then*

$$\frac{\mathrm{d}\sigma_{n,p}}{\mathrm{d}\gamma_{n,p}}(x) = C_{n,p}\left(\sum_{i=1}^n |x_i|^{2p-2}\right)^{1/2}, \qquad x \in \mathsf{S}_p^{n-1},$$

where $C_{n,p}$ is a normalizing constant.

Next, we state a general result about LDPs for two sequences of measures that satisfy a particular absolute continuity relation.

Lemma 5. *Let $\mathcal{X}$ be a Polish space, and for $n \in \mathbb{N}$, let $\psi_n \colon \mathcal{X} \to \mathbb{R}$. Suppose that there exists a sequence of finite constants $(M_n)_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty} M_n/n = 0$ such that*

$$|\psi_n(\lambda)| \le M_n, \qquad \mu_n\text{-a.e. } \lambda \in \mathcal{X} \quad \text{for all } n \in \mathbb{N}, \tag{12}$$

where we abbreviate almost every to a.e. Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ be a sequence of probability measures that satisfies an LDP with a good rate function $I(\cdot)$. Define

$$\nu_n(\mathrm{d}\lambda) := \frac{1}{\int_{\mathcal{X}} \exp(\psi_n(\lambda))\mu_n(\mathrm{d}\lambda)} \mathrm{e}^{\psi_n(\lambda)}\mu_n(\mathrm{d}\lambda). \tag{13}$$

Then $(\nu_n)_{n\in\mathbb{N}}$ satisfies an LDP with the same good rate function $I(\cdot)$.

Proof. For $\phi \colon \mathcal{X} \to \mathbb{R}$ continuous and bounded, define

$$\Lambda_\phi := \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)}\mu_n(\mathrm{d}\lambda),$$

where the limit exists due to the LDP for $(\mu_n)_{n\in\mathbb{N}}$ and Varadhan's Lemma (see, e.g. [14, Theorem 4.3.1]). The definition of ν_n in (13), the common bound on ψ_n for all $n \in \mathbb{N}$ given in (12), and the assumption that $M_n/n \to 0$ imply that

$$\liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)}\nu_n(\mathrm{d}\lambda) \ge \liminf_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)-M_n}\mu_n(\mathrm{d}\lambda) = \Lambda_\phi - \lim_{n\to\infty} \frac{M_n}{n} = \Lambda_\phi,$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)}\nu_n(\mathrm{d}\lambda) \le \limsup_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)+M_n}\mu_n(\mathrm{d}\lambda) = \Lambda_\phi + \lim_{n\to\infty} \frac{M_n}{n} = \Lambda_\phi.$$

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Thus, we have shown that

$$\Lambda_\phi = \tilde{\Lambda}_\phi := \lim_{n\to\infty} \frac{1}{n} \log \int_{\mathcal{X}} \mathrm{e}^{n\phi(\lambda)} \nu_n(\mathrm{d}\lambda).$$

Note that $(\mu_n)_{n\in\mathbb{N}}$ is exponentially tight since it satisfies an LDP with good rate function and $\mathcal{X}$ is Polish; see [26, Lemma 2.6]. We claim that $(\nu_n)_{n\in\mathbb{N}}$ is also exponentially tight. Let $L < \infty$, and let $K_L \subset \mathcal{X}$ be a compact set such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_L^{\mathrm{c}}) < -L. \tag{14}$$

Then, given $L < \infty$, note that

$$\begin{aligned} \log \nu_n(K_L^{\mathrm{c}}) &= \log \int_{K_L^{\mathrm{c}}} \mathrm{e}^{\psi_n(\lambda)} \mu_n(\mathrm{d}\lambda) - \log \int_{\mathcal{X}} \mathrm{e}^{\psi_n(\lambda)} \mu_n(\mathrm{d}\lambda) \\ &\le M_n + \log \mu_n(K_L^{\mathrm{c}}) + M_n - \log \mu_n(\mathcal{X}). \end{aligned}$$

Taking the limit supremum as $n \to \infty$, using (14), and applying the fact that $M_n/n \to 0$, we find that

$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n(K_L^{\mathrm{c}}) \le \limsup_{n\to\infty} \frac{1}{n} \log \mu_n(K_L^{\mathrm{c}}) < -L.$$

Since $\Lambda_\phi = \tilde{\Lambda}_\phi$, and both sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ are exponentially tight, Bryc's inverse lemma (see, e.g. [14, Theorem 4.4.2]) implies that the two sequences satisfy the same LDP with the good rate function $I(\lambda) = \sup_{\phi\in C_{\mathrm{b}}(\mathcal{X})}\{\phi(\lambda) - \Lambda_\phi\}$. □

We apply the preceding lemma to the absolute continuity relation of Lemma 4 to prove the LDP under the surface measure.

Proof of Proposition 4. For $p = 1, 2, \infty$, we have $\sigma_{n,p} = \gamma_{n,p}$, and so the result trivially follows from Proposition 2. Therefore, we restrict to $p \in (1, 2)$ or $p \in (2, \infty)$. For fixed p in this range and $q < p$, define the map $\mathcal{L}_n \colon \mathsf{S}_p^{n-1} \to \mathcal{P}_q(\mathbb{R})$ that takes a vector $a \in \mathsf{S}_p^{n-1}$ to the measure $(1/n)\sum_{i=1}^n \delta_{n^{1/p}a_i} \in \mathcal{P}_q(\mathbb{R})$, and note that $\mathcal{L}_n$ maps S_p^{n-1} onto the set

$$A_n := \left\{ \frac{1}{n} \sum_{i=1}^n \delta_{n^{1/p}a_i} : a \in \mathsf{S}_p^{n-1} \right\} \subset \mathcal{P}_q(\mathbb{R}).$$

We then apply Lemma 5 to the following setting:

- $\mathcal{X} = \mathcal{P}_q(\mathbb{R})$;
- $\mu_n = \gamma_{n,p} \circ \mathcal{L}_n^{-1}$ and $\nu_n = \sigma_{n,p} \circ \mathcal{L}_n^{-1}$ on A_n, and $\mu_n(A_n^{\mathrm{c}}) = \nu_n(A_n^{\mathrm{c}}) = 0$;
- $\psi_n = \psi = \frac{1}{2} \log m_{2p-2}$;
- $M_n := |\frac{1}{2} - (1/p)| \log n$;

and note that clearly $M_n/n \to 0$. It remains to show that our setup satisfies the remaining hypotheses of Lemma 5. We know from Proposition 2 that $(\mu_n)_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

(for $q < p$), with good rate function H_p. Further, using Lemma 4, we can write the Radon–Nikodym derivative of ν_n with respect to μ_n for $\lambda \in A_n$ (which is a set of full μ_n measure since $\mu_n(A_n) = 1$) as

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}(\lambda) = \frac{\mathrm{d}\sigma_{n,p}}{\mathrm{d}\gamma_{n,p}}(\mathcal{L}_n^{-1}\lambda) = C_{n,p}(n^{2/p}m_{2p-2}(\lambda))^{1/2} = n^{1/p}C_{n,p}\mathrm{e}^{\psi(\lambda)},$$

from which (13) is verified.

To verify the remaining boundedness property of $\psi_n = \psi = \frac{1}{2}\log m_{2p-2}$, as stipulated in (12) of Lemma 5, first consider the $1 < p < 2$ case. To develop an upper bound for $\psi(\lambda) = m_{2p-2}(\lambda)$, note that due to Hölder's inequality, for any $0 < r < s < \infty$ and $\lambda \in \mathcal{P}(\mathbb{R})$, we have $m_r(\lambda) \le [m_s(\lambda)]^{r/s}$. Then $0 < 2p-2 < p$, and so applying the preceding inequality with $r = 2p-2$ and $s = p$, we see that for $\lambda \in A_n$,

$$\begin{aligned}
\mu_n(m_{2p-2}(\lambda) > 1) &\le \mu_n(m_p(\lambda)^{2-(2/p)} > 1) \\
&= \mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n}[n^{1/p}X^{(n,p)}]^p\right)^{2-(2/p)} > 1\right) \\
&= \mathbb{P}(\|X^{(n,p)}\|_{n,p}^{2-(2/p)} > 1) \\
&= 0,
\end{aligned}$$

where the final equality follows since $X^{(n,p)} \in \mathsf{S}_p^{n-1}$, so $\|X^{(n,p)}\|_{n,p}^{2-(2/p)} = 1$, $\mathbb{P}$-a.s. ($\mathbb{P}$-almost surely). On the other hand, to obtain a lower bound for $\psi(\lambda) = m_{2p-2}(\lambda)$, recall that for $0 < r < s < \infty$ and $a \in \mathbb{R}^n$, we have $\|a\|_{n,s} \le \|a\|_{n,r}$. Applying this inequality with $r = 2p-2$ and $s = p$, and recalling the definition of the empirical measure $\mathcal{L}_{n,p}$ from (6), we have $\mathbb{P}$-a.s.,

$$\begin{aligned}
m_{2p-2}(\mathcal{L}_{n,p}) &= \frac{1}{n}\sum_{i=1}^{n}[|n^{1/p}X^{(n,p)}|]^{2p-2} \\
&= n^{1-(2/p)}\|X^{(n,p)}\|_{n,2p-2}^{2p-2} \\
&\ge n^{1-(2/p)}\|X^{(n,p)}\|_{n,p}^{2p-2} \\
&= n^{1-(2/p)},
\end{aligned}$$

where the last equality once again uses the fact that $\mathbb{P}$-a.s. (under the cone measure), and $X^{(n,p)}$ lies on the unit ℓ^p-sphere S_p^{n-1}.

In a similar fashion, for the $2 < p < \infty$ case, we can show that $\mu_n(m_{2p-2}(\lambda) < 1) = 0$ for $\lambda \in A_n$, and $m_{2p-2}(\mathcal{L}_{n,p}) \le n^{1-(2/p)}$, $\mathbb{P}$-a.s.

Since $\mu_n(A_n) = 1$, we have shown that, for μ_n-a.e. $\lambda \in \mathcal{P}(\mathbb{R})$,

$$|\psi(\lambda)| = \left|\frac{1}{2}\log m_{2p-2}(\lambda)\right| \le \left|\frac{1}{2} - \frac{1}{p}\right|\log n = M_n,$$

which concludes the verification of all the hypotheses of Lemma 5. Thus, it follows from that lemma that the sequence of empirical measures $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ under the surface measure $\sigma_{n,p}$ satisfies an LDP with the same good rate function as under the cone measure $\gamma_{n,p}$. □

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

4. Gibbs conditioning

In this section we use the LDPs of Propositions 2 and 4 to prove our main result Proposition 5, in which we state a conditional limit law; that is, the asymptotically most likely distribution of a sequence of random variables, conditional on a large deviation. In the novel setting of ℓ^p-spheres, we are able to lend a geometric meaning to this conditional limit law.

To state our result, we first introduce some additional notation. Let $h(\cdot)$ denote the differential entropy of a probability measure $\nu \in \mathcal{P}(\mathbb{R})$,

$$h(\nu) := \begin{cases} -\displaystyle\int_{\mathbb{R}} \frac{\mathrm{d}\nu}{\mathrm{d}x} \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}x}\right) \mathrm{d}x & \text{if } \nu \ll \text{Lebesgue measure on } \mathbb{R}, \\ -\infty & \text{otherwise.} \end{cases}$$

Differential entropy arises naturally in the analysis of the rate function H_p since for measures $\nu \in \mathcal{P}(\mathbb{R})$ that are absolutely continuous with respect to Lebesgue measure and satisfy $m_p(\nu) \le 1$, we can use the definition of μ_p in (1) to rewrite the function H_p defined in (9) as

$$H_p(\nu) = -h(\nu) + \log\left(2p^{1/p}\Gamma\left(1+\frac{1}{p}\right)\right) + \frac{1}{p}. \tag{15}$$

We obtain the following result as a consequence of the LDP of Theorem 2, which involves a constrained maximum entropy problem. In particular, we crucially use the fact that the LDP for $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ holds with respect to the Wasserstein topology, which is stronger than the usual weak topology.

Proposition 5. *Fix $p \in [1,\infty]$ and suppose that either $X^{(n,p)} \sim \sigma_{n,p}$ or $X^{(n,p)} \sim \gamma_{n,p}$. Fix a closed interval $C = [\alpha,\beta] \subset \mathbb{R}$, and for $\varepsilon > 0$, let $C_\varepsilon := [\alpha-\varepsilon, \beta+\varepsilon]$. Then, for $q < p$, the optimizing measure*

$$\nu_* := \arg\max\{h(\nu) \colon m_p(\nu) \le 1,\ m_q(\nu) \in C\} \tag{16}$$

is well defined (that is, exists and is unique), and

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty} \mathbb{P}(\mathcal{L}_{n,p} \in \cdot \mid m_q(\mathcal{L}_{n,p}) \in C_\varepsilon) = \delta_{\nu_*}. \tag{17}$$

Moreover, for $k \in \mathbb{N}$,

$$\mathbb{P}\left(n^{1/p}(X_1^{(n,p)},\dots,X_k^{(n,p)}) \in \cdot \;\middle|\; \frac{1}{n}\|n^{1/p}X^{(n,p)}\|_{n,q}^q \in C_\varepsilon\right) \xrightarrow{\mathrm{w}} \nu_*^{\otimes k} \quad \text{as } n\to\infty \tag{18}$$

followed by $\varepsilon \to 0$.

The proof of Proposition 5 relies crucially on the so-called Gibbs conditioning principle (see, e.g. [10], [13], and [34]), stated precisely below.

Proposition 6. (Gibbs conditioning principle, [24, Theorem 7.1].) *Let $\mathcal{X}$ be a topological space, and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{X}$-valued random variables that satisfies an LDP with good rate function I. In addition, let $F \subset \mathcal{X}$ be a subset such that*

(i) *$I(F) := \inf_{x\in F} I(x) < \infty$;*

(ii) *F is closed;*

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

(iii) $F = \bigcap_{\varepsilon>0} F_\varepsilon$ *for a family of sets* $(F_\varepsilon)_{\varepsilon>0}$ *such that* $F_\varepsilon \subset \mathcal{X}$ *for all* $\varepsilon > 0$ *and* $\mathbb{P}(\xi_n \in F_\varepsilon) > 0$ *for all* $\varepsilon > 0$ *and* $n \in \mathbb{N}$*;*

(iv) $F \subset (F_\varepsilon)^\circ$ *for all* $\varepsilon > 0$.

Let $\mathcal{M}_F$ *be the set of* $x \in F$ *which minimizes* I*. That is,*

$$\mathcal{M}_F := \{x \in F : I(x) = I(F)\}.$$

Then, for all open $G \subset \mathcal{X}$ *such that* $\mathcal{M}_F \subset G$*, we have*

$$\limsup_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\xi_n \notin G \mid \xi_n \in F_\varepsilon) < 0.$$

As a consequence, if $\mathcal{M}_F = \{\bar{x}\}$ *is a singleton then*

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \mathbb{P}(\xi_n \in \cdot \mid \xi_n \in F_\varepsilon) = \delta_{\bar{x}}. \tag{19}$$

Proof of Proposition 5. Fix $p \in [1, \infty]$ and $q < p$, and recall that due to Propositions 2 and 4, the sequence $(\mathcal{L}_{n,p})_{n\in\mathbb{N}}$ satisfies an LDP in $\mathcal{P}_q(\mathbb{R})$ (equipped with the q-Wasserstein topology) with strictly convex good rate function H_p.

We first show that Proposition 6 applies in the setting of Proposition 5. Fix $C = [\alpha, \beta] \subset \mathbb{R}$ and for $\varepsilon > 0$, let $C_\varepsilon = [\alpha - \varepsilon, \beta + \varepsilon]$. The set $F_\varepsilon := \{\nu \in \mathcal{P}_q(\mathbb{R}) : m_q(\nu) \in C_\varepsilon\}$ is closed due to the continuity of m_q of (8) with respect to the q-Wasserstein topology and, hence, as an intersection of closed sets, the set $F := \bigcap_{\varepsilon>0} F_\varepsilon = \{\nu \in \mathcal{P}(\mathbb{R}) : m_q(\nu) \in C\}$ is also closed. Moreover, for $\varepsilon > 0$, we also have

$$F = m_q^{-1}(C) \subset m_q^{-1}(C_\varepsilon^\circ) \subset [m_q^{-1}(C_\varepsilon)]^\circ = (F_\varepsilon)^\circ,$$

where the second inclusion again uses the continuity of m_q. Next, we show that $\mathcal{M}_F$ is a singleton (that is, ν_* of (16) is well defined). Recall that

$$\mathcal{M}_F := \left\{\nu \in F : H_p(\nu) = \min_{\nu\in F} H_p(\nu)\right\}.$$

Note that F is closed and convex since it is the pre-image of a closed, convex set C under a continuous linear map m_q. Because H_p is lower semicontinuous and has compact level sets, it attains its minimum within F. That the minimum is attained at a unique $\nu_* \in F$ follows from the strict convexity of H_p. The 'maximum entropy' representation for ν_* given in (16) follows from the expression for H_p given in (15). Thus, the limit (17) follows from the limit (19) of Proposition 6 applied in the following setting: the probability measure $\xi_n = \mathcal{L}_{n,p}$; the rate function minimizing set $\mathcal{M}_F$ consists of the singleton $\{\nu_*\}$; and the conditioning event is $F_\varepsilon = \{\nu \in \mathcal{P}_q(\mathbb{R}) : m_q(\nu) \in C_\varepsilon\}$, as defined above.

As for the second result (18), this follows from an elementary 'propagation of chaos' result (see [35, Proposition 2.2]) that, under the assumption of exchangeability, allows us to establish the equivalence of statements such as (17) (regarding convergence in probability of the empirical measure to a deterministic measure) and statements such as (18) (regarding joint convergence in distribution of any fixed k of the underlying random variables). □

As a prerequisite for the proof of Theorem 1, recall the following basic information-theoretic fact.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Lemma 6. *Fix $r_i : \mathbb{R} \to \mathbb{R}$, $\alpha_i \in \mathbb{R}$ for $i = 1, \ldots, m$, and $s_j : \mathbb{R} \to \mathbb{R}$, $\beta_j \in \mathbb{R}$ for $j = 1, \ldots, n$. Consider the following maximization problem:*

$$\begin{aligned} \max_{\nu \in \mathcal{P}(\mathbb{R})} \quad & h(\nu) \\ \text{subject to} \quad & \int_{\mathbb{R}} r_i(x)\nu(\mathrm{d}x) = \alpha_i \quad \text{for } i = 1, \ldots, m, \\ & \int_{\mathbb{R}} s_j(x)\nu(\mathrm{d}x) \le \beta_j \quad \text{for } j = 1, \ldots, n. \end{aligned} \tag{20}$$

Then a probability measure $\nu_* \in \mathcal{P}(\mathbb{R})$ attains the maximum in (20) if and only if it is of the form

$$\nu_*(\mathrm{d}x) = \exp\left(-1 - \kappa_0^* - \sum_{i=1}^m \lambda_i^* r_i(x) - \sum_{j=1}^n \mu_j^* s_j(x)\right) \mathrm{d}x$$

with nonnegative constants κ_0^*, $(\lambda_i^*)_{i=1}^m$, and $(\mu_j^*)_{j=1}^n$ chosen such that ν_* lies in $\mathcal{P}(\mathbb{R})$ and satisfies the constraints in (20). Moreover, if ν_* attains the maximum in (20) then

$$\left(\int_{\mathbb{R}} s_j(x)\nu_*(\mathrm{d}x) - \beta_j\right) \cdot \mu_j^* = 0 \quad \text{for all } j = 1, \ldots, n. \tag{21}$$

The preceding lemma is standard. See [9, Section 12.1] for a slight simplification of (20) with only equality constraints, or [8, Exercise 5.3] for a version of (20) with discrete entropy and both equality and inequality constraints. Claim (21) is the so-called 'complementary slackness' condition (see, e.g. [8, Section 5.5.2]).

As a final preliminary, define the following family of probability measures for $q \in [1, \infty)$, $\beta > 0$:

$$\mu_{q,\beta}(\mathrm{d}x) := \frac{1}{2(\beta q)^{1/q}\Gamma(1 + 1/q)} \mathrm{e}^{-|x|^q/(\beta q)}\, \mathrm{d}x, \qquad x \in \mathbb{R}. \tag{22}$$

Note that $\mu_{p,1}$ corresponds to μ_p of (1).

Proof of Theorem 1. For a random variable $X \sim \mu_q$, note that $\beta^{1/q} X \sim \mu_{q,\beta}$. Due to the unconditional limit theorem (Proposition 1), under either the surface measure or the cone measure, we have $\lambda_{n,q}^{(\beta,k)} \xrightarrow{\mathrm{w}} \mu_{q,\beta}^{\otimes k}$, where $\beta \le \beta_{p,q}$ from (4), and $\mu_{q,\beta}$ is as in (22). Therefore, it suffices to show that under either the surface measure or the cone measure, the conditional distribution $\widehat{\lambda}_{n,p|q}^{(\beta+\varepsilon,k)}$ also converges weakly to the same limit $\mu_{q,\beta}^{\otimes k}$. In view of (18) of Proposition 5, it suffices to show that when $C = [0, \beta]$, the unique maximizer ν_* of (16) satisfies $\nu_* = \mu_{q,\beta}$. Note that due to the maximum entropy calculations of Lemma 6, we have for $\beta > 0$,

$$\mu_{q,\beta} = \arg\max\{h(\nu) : m_q(\nu) \le \beta\}. \tag{23}$$

To show that $\mu_{q,\beta} = \nu_*$, it suffices to show that $m_p(\mu_{q,\beta}) \le 1$. After some elementary calculations, we find that since $\beta \le \beta_{p,q}$, and using the expression for $\beta_{p,q}$ in (4),

$$m_p(\mu_{q,\beta}) = \beta^{p/q} q^{p/q} \frac{\Gamma((p+1)/q)}{\Gamma(1/q)} \le \frac{1}{q^{p/q}} \left(\frac{\Gamma(1/q)}{\Gamma((p+1)/q)}\right) q^{p/q} \frac{\Gamma((p+1)/q)}{\Gamma(1/q)} = 1,$$

completing the proof. □

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

Remark 6. (*'Small' β.*) The constant $\beta_{p,q}$ is chosen such that for $\beta \leq \beta_{p,q}$ and the interval $C = [0, \beta]$, the variational problem of (16) has an explicit solution with a natural geometric interpretation. Note that the explicit solution to the simpler variational problem (23) holds for all $\beta > 0$. However, the original variational problem of (16) involves not only the qth moment constraint $m_q(\nu) \in C$, but also an additional constraint on the pth moment, $m_p(\nu) \leq 1$. To simplify variational problem (16), it suffices to consider values of β small enough such that $m_p(\mu_{q,\beta}) \leq 1$, so that the maximizer of (23) is also the maximizer of (16) when $C = [0, \beta]$. It is easy to see that

$$m_p(\mu_{q,\beta_{p,q}}) = \beta_{p,q}^{p/q} q^{p/q} \frac{\Gamma((p+1)/q)}{\Gamma(1/q)} = 1. \tag{24}$$

That is, $\beta_{p,q}$ is the threshold value such that for $\beta \leq \beta_{p,q}$, we have $m_p(\mu_{q,\beta}) \leq 1$. Note, however, that this discussion of $\beta_{p,q}$ (and, thus, Theorem 1) is valid only for $p \in [1, \infty)$ since for $p = \infty$, we have $m_\infty(\mu_{q,\beta}) = \infty$ for all $q < \infty$ and $\beta > 0$. That is, for $q < \infty$, there is no value of $\beta_{\infty,q} > 0$ such that an analog of (24) can hold for $p = \infty$.

Remark 7. (*'Large' β.*) Now suppose that $\beta \geq \overline{\beta}_{p,q} := m_q(\mu_p)$. An easy consequence of Lemma 6 is

$$\mu_p = \arg\max\{h(\nu) \colon m_p(\nu) \leq 1\}. \tag{25}$$

For $\beta \geq \overline{\beta}_{p,q}$ and $C = [0, \beta]$, the qth moment constraint of (16) is automatically satisfied by $\nu = \mu_p$, so the maximizer of (25) is also the maximizer of (16). In this case, the 'conditional' limit of Proposition 5 is equivalent to the 'unconditional' limit of Proposition 1. In other words, for sufficiently large β, the ℓ^q-norm conditioning event of (3) is extraneous.

Remark 8. (*'Intermediate' β.*) As for $C = [0, \beta]$ and $\beta_{p,q} < \beta < \overline{\beta}_{p,q}$, this regime is less amenable to an immediate geometric interpretation. Whereas the small β regime of Theorem 1 and Remark 6 allows an ℓ^q-sphere interpretation via the measure $\mu_{q,\beta}$, and the large β regime of Remark 7 allows an ℓ^p-sphere interpretation via the measure μ_p, we have a different behavior in the intermediate β regime. Recall that the maximum entropy considerations of Lemma 6 imply that the unique maximizer of Proposition 5 is of the form

$$\nu_*(\mathrm{d}x) = \exp(-1 - \kappa_0 - \kappa_p |x|^p - \kappa_q |x|^q) \tag{26}$$

with constants $\kappa_0, \kappa_p, \kappa_q$ such that ν_* is a probability measure that satisfies $m_p(\nu_*) \leq 1$ and $m_q(\nu_*) \leq \beta$. To gain insight on these constants, consider the following cases.

- If $\kappa_p = 0$ and $\kappa_q > 0$ then the complementary slackness condition of (21) implies $m_q(\nu_*) = \beta$, so the form (26) implies $\nu_* = \mu_{q,\beta}$. But this is not a feasible solution since $m_p(\mu_{q,\beta}) > 1$ for $\beta > \beta_{p,q}$.
- Similarly, if $\kappa_p > 0$ and $\kappa_q = 0$ then we have $\nu_* = \mu_p$, but this is not a feasible solution since $m_q(\mu_p) = \overline{\beta}_{p,q} > \beta$.
- Finally, if $\kappa_p = 0$ and $\kappa_q = 0$ then ν_* is not a probability measure for any choice of κ_0.

Therefore, it must be the case that $\kappa_p > 0$ *and* $\kappa_q > 0$, which implies that ν_* of (26) is not of the form $\mu_{r,b}$ for any $r \in [1, \infty)$, $b > 0$. Instead, ν_* is of an exponential family that is genuinely different from the generalized normal distributions.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

In this paper we have discussed applications of Proposition 5 only to intervals C of the form $C = [0, \beta]$, and primarily for small β, with some discussion of larger β in Remark 7 and Remark 8. We leave for future work the question of finding other examples of intervals $C \subset \mathbb{R}$ which lead to an explicit and meaningful maximizing measure ν_*, and other geometric conditional laws.

Acknowledgements

SSK was partially supported by a US Department of Defense NDSEG fellowship. KR was partially supported by Army Research Office grant W911NF-12-1-0222 and National Science Foundation grants DMS 1407504 and DMS-1713032. SSK and KR would also like to thank Microsoft Research New England for their hospitality during the Fall of 2014, when some of this work was carried out.

References

[1] Alonso-Gutiérrez, D., Prochno, J. and Thäle, C. (2018). Large deviations for high-dimensional random projections of ℓ_p^n balls. *Adv. Appl. Math.* **99,** 1–35.
[2] Ambrosio, L., Gigli, N. and Savaré, G. (2008). *Gradient flows in metric spaces and in the space of probability measures* (Lectures Math. ETH Zürich), 2nd edn. Birkhäuser, Basel.
[3] Barthe, F., Guédon, O., Mendelson, S. and Naor, A. (2005). A probabilistic approach to the geometry of the ℓ_p^n-ball. *Ann. Prob.* **33,** 480–513.
[4] Ben Arous, G., Dembo, A. and Guionnet, A. (2001). Aging of spherical spin glasses. *Prob. Theory Relat. Fields* **120,** 1–67.
[5] Ben Arous, G. and Guionnet, A. (1997). Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Prob. Theory Relat. Fields* **108,** 517–542.
[6] Bertini, L. *et al.* (2015). Macroscopic fluctuation theory. *Rev. Modern Phys.* **87,** 593–636.
[7] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
[8] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
[9] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*, 2nd edn. John Wiley, Hoboken, NJ.
[10] Csiszár, I. (1984). Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Prob.* **12,** 768–793.
[11] Dellacherie, C. and Meyer, P.-A. (2011). *Probabilities and Potential, C* (North-Holland Math. Studies **151**). North-Holland, Amsterdam.
[12] Dembo, A. and Shao, Q.-M. (1998). Self-normalized large deviations in vector spaces. In *High Dimensional Probability* (Progr. Prob. **43**). Birkhäuser, Basel, pp. 27–32.
[13] Dembo, A. and Zeitouni, O. (1996). Refinements of the Gibbs conditioning principle. *Prob. Theory Relat. Fields* **104,** 1–14.
[14] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications* (Appl. Math. **38**), 2nd edn. Springer, Berlin.
[15] Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Statist.* **23,** 397–423.
[16] Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.
[17] Ellis, R. S. (2006). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, Berlin.
[18] 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**). Springer, Cham, pp. 253–270.
[19] Gantert, N., Kim, S. S. and Ramanan, K. (2017). Large deviations for random projections of ℓ^p balls. *Ann. Prob.* **45,** 4419–4476.
[20] Kabluchko, Z., Prochno, J. and Thäle, C. (2017). High-dimensional limit theorems for random vectors in ℓ_p^n-balls. *Commun. Contemp. Math.* 1750092.
[21] Kim, S. S. (2017). Problems at the interface of probability and convex geometry: Random projections and constrained processes. Doctoral Thesis, Brown University.
[22] Kim, S. S. and Ramanan, K. (2018). Large deviations on the Stiefel manifold. Preprint.
[23] Klartag, B. (2007). A central limit theorem for convex sets. *Invent. Math.* **168,** 91–131.
[24] Léonard, C. (2010). Entropic projections and dominating points. *ESAIM: Prob. Statist.* **14,** 343–381.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press

[25] Léonard, C. and Najim, J. (2002). An extension of Sanov's theorem: application to the Gibbs conditioning principle. *Bernoulli* **8,** 721–743.
[26] Lynch, J. and Sethuraman, J. (1987). Large deviations for processes with independent increments. *Ann. Prob.* **15,** 610–627.
[27] Mogul′skii, A. A. (1991). De Finetti-type results for ℓ_p. *Siberian Math. J.* **32,** 609–616.
[28] Naor, A. (2007). The surface measure and cone measure on the sphere of ℓ_p^n. *Trans. Amer. Math. Soc.* **359,** 1045–1079.
[29] Naor, A. and Romik, D. (2003). Projecting the surface measure of the sphere of ℓ_p^n. *Ann. Inst. H. Poincaré Prob. Statist.* **39,** 241–261.
[30] Rachev, S. T. and Rüschendorf, L. (1991). Approximate independence of distributions on spheres and their stability properties. *Ann. Prob.* **19,** 1311–1337.
[31] 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.
[32] Spruill, M. C. (2007). Asymptotic distribution of coordinates on high dimensional spheres. *Electron. Commun. Prob.* **12,** 234–247.
[33] Stam, A. J. (1982). Limit theorems for uniform distributions on spheres in high-dimensional euclidean spaces. *J. Appl. Prob.* **19,** 221–228.
[34] Stroock, D. W. and Zeitouni, O. (1991). Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In *Random Walks, Brownian Motion, and Interacting Particle Systems* (Progr. Prob. **28**). Birkhäuser, Boston, MA.
[35] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989* (Lecture Notes Math. **1464**). Springer, Berlin, pp. 165–251.
[36] Trashorras, J. (2002). Large deviations for a triangular array of exchangeable random variables. *Ann. Inst. H. Poincaré Prob. Statist.* **38,** 649–680.
[37] Villani, C. (2009). *Optimal Transport: Old and New*, **338**, Springer, Berlin.
[38] Wang, R., Wang, X. and Wu, L. (2010). Sanov's theorem in the Wasserstein distance: a necessary and sufficient condition. *Statist. Prob. Lett.* **80,** 505–512.

https://doi.org/10.1017/jpr.2018.71 Published online by Cambridge University Press