A NOTE ABOUT THE UNIFORM DISTRIBUTION ON THE INTERSECTION OF A SIMPLEX AND A SPHERE

SOURAV CHATTERJEE

ABSTRACT. Uniform probability distributions on ℓ_p balls and spheres have been studied extensively and are known to behave like product measures in high dimensions. In this note we consider the uniform distribution on the intersection of a simplex and a sphere. Certain new and interesting features, such as phase transitions and localization phenomena emerge.

1. Introduction

Take a real number number $b > 1$ and a positive integer n, and consider the set

$$\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : \textstyle\sum_1^n |x_i| = n,\ \sum_1^n x_i^2 = nb\}.$$

This is the intersection of an ℓ_1^n sphere and an ℓ_2^n sphere in $\mathbb{R}^n$. By sign symmetry, to study the above object it suffices to study

$$K := \{x = (x_1, \ldots, x_n) \in \mathbb{R}_+^n : \textstyle\sum_1^n x_i = n,\ \sum_1^n x_i^2 = nb\}, \tag{1}$$

where $\mathbb{R}_+$ denote the set of all positive real numbers. Consider the uniform distribution on this set, defined as the limit of normalized Lebesgue measures on thin shells around this set as the thickness of the shells tend to zero.

Note that b has to range between 1 and n for K to be non-empty. We are mainly interested in b fixed and $n \to \infty$. Let $X = (X_1, \ldots, X_n)$ be a random vector following the uniform distribution on K. Let us omit the trivial case $b = 1$, when all coordinates are exactly equal to 1. The first theorem covers the range $1 < b \le 2$.

Theorem 1.1. *Suppose $1 < b \le 2$. Then there exist unique $r, s \in \mathbb{R}$ such that the probability density proportional to $\exp(-rx^2 - sx)$ on $[0, \infty)$ has first moment 1 and second moment b. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables following this density. The following hold:*

- (a) *For any fixed k, the random vector $(X_1, \ldots, X_k)$ converges in law to $(Z_1, \ldots, Z_k)$ as $n \to \infty$.*
- (b) *All joint moments of $(X_1, \ldots, X_k)$ converge to the corresponding moments of $(Z_1, \ldots, Z_k)$.*

2000 *Mathematics Subject Classification.* 35Q55, 82B26, 82B10, 60F10.

Key words and phrases. Simplex, sphere, ℓ_p^n ball, large deviation.

Sourav Chatterjee's research was partially supported by NSF grants DMS-0707054 and DMS-1005312, and a Sloan Research Fellowship.

(c) *If $b < 2$, there is a constant C, possibly depending on b, such that*

$$\lim_{n\to\infty} \mathbb{P}\big(\max_{1\le i\le n} X_i > C\sqrt{\log n}\big) = 0.$$

(d) *When $b = 2$, part (c) holds but with $\log n$ instead of $\sqrt{\log n}$.*

Note that in general s can be negative. When $b = 2$, r and s turn out to be 0 and 1; in other words $Z_1 \sim Exp(1)$ when $b = 2$.

The next theorem describes the situation when $b > 2$. An interesting localization phenomenon occurs in this regime.

Theorem 1.2. *Suppose $b > 2$. Let $Z_1, Z_2, \ldots$ be i.i.d. $Exp(1)$ random variables. Then the following hold.*

(a) *For any fixed k, the random vector $(X_1, \ldots, X_k)$ converges in law to $(Z_1, \ldots, Z_k)$ as $n \to \infty$.*
(b) *Convergence of moments does not happen, because $\mathbb{E}(X_1^2) = b$ for any n and $\mathbb{E}(Z_1^2) = 2$.*
(c) *Let $M = \max_{1\le i\le n} X_i$. Then*

$$\frac{M^2}{(b-2)n} \to 1 \text{ in probability.}$$

Consequently, the sum of squares of all other coordinates is roughly $2n$ with high probability.

(d) *Let M_2 be the value of the second largest coordinate. Then*

$$\frac{M_2^2}{n} \to 0 \text{ in probability.}$$

The final theorem in this note provides error bounds for the distributional convergence results in Theorems 1.1 and 1.2. The bounds may not be sharp.

Theorem 1.3. *In the setting of Theorem 1.1,*

$$\sup_{t_1,\ldots,t_k} \big|\mathbb{P}(X_1 \le t_1, \ldots, X_k \le t_k) - \mathbb{P}(Z_1 \le t_1, \ldots, Z_k \le t_k)\big| \le Ck\sqrt{\frac{\log n}{n}},$$

where C is a constant that depends only on b. In Theorem 1.2, the bound on the right hand side becomes $Ckn^{-1/4}$.

If the condition $\sum x_i^2 = nb$ is dropped from the definition of K, the result is a scaled version of the standard $(n-1)$-simplex in $\mathbb{R}^n$. It is a classical result in probability that the coordinates of a point chosen uniformly from this body behave like independent standard Exponential random variables in the large n limit.

On the other hand, if the condition $\sum x_i = n$ is dropped, then K is just a sphere of radius $\sqrt{nb}$. Drawing uniformly from the surface of a sphere results in a vector with approximately independent Gaussian coordinates.

A unified treatment of results of the above type was done by Diaconis and Freedman [4], which is the basic reference for the literature in this area till 1987.

In recent times, attention has shifted to the study of the ℓ_p^n balls and spheres, that is, sets where $\sum |x_i|^p$ is bounded by or equal to a constant. The distribution of low dimensional projections for ℓ_p^n balls was obtained by Naor and Romik [6], who showed that the coordinates behave like i.i.d. random variables with density proportional to $e^{-|x|^p}$. An extensive investigation of the probabilistic structure of ℓ_p^n balls was done by Barthe et. al. [1].

The volumes of intersections of ℓ_p^n balls (not spheres) have been previously investigated, in response to a question raised by Vitali Milman, in a series of papers by Schechtman and Zinn [11], Schechtman and Schmuckenschläger [10] and Schmuckenschläger [12, 13]. They do not, however, study the behavior of uniformly chosen random points from these sets.

A second motivation for studying the uniform distribution on K comes from some recent studies in the statistical mechanics of the discrete nonlinear Schrödinger equation. For more on this connection, see the recent paper [3], which is actually a follow-up of an earlier draft of this article. For the connections with the physics literature, see [9] and references therein.

2. Preliminaries

For a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, define

$$\mu(x) := \frac{1}{n}\sum_{i=1}^n x_i, \quad \mu_2(x) := \frac{1}{n}\sum_{i=1}^n x_i^2 = \frac{\|x\|^2}{n}.$$

Also define

$$\sigma(x) := \sqrt{\mu_2(x) - \mu^2(x)}, \quad m(x) := \min_{1\le i\le n} x_i.$$

Fix $b > 1$ as in Section 1 and let $b' = \sqrt{b-1}$. Let $\mathbb{R}_+$ be the set of positive real numbers. Note that according to the definition (1),

$$\begin{aligned} K &= \{x \in \mathbb{R}_+^n : \mu(x) = 1,\ \mu_2(x) = b\} \\ &= \{x \in \mathbb{R}_+^n : \mu(x) = 1,\ \sigma(x) = b'\}. \end{aligned}$$

The notation introduced above will be used without explicit reference in the rest of the manuscript.

Recall that the uniform distribution on the unit sphere in any dimension is equivalently defined as the unique probability measure that is invariant under rotations (i.e. the action of orthogonal matrices). For each $a, d \in \mathbb{R}$, $c > 0$, define

$$\begin{aligned} S(a, c) &:= \{x \in \mathbb{R}^n : \mu(x) = a,\ \sigma(x) = c\}, \\ S(a, c, d) &:= \{x \in S(a, c) : m(x) > d\}. \end{aligned}$$

Note that $S(0, 1)$ is a sphere in the $n-1$ dimensional hyperplane $\{x : \mu(x) = 0\}$, centered at the origin. This hyperplane can be obtained as the image of $\mathbb{R}^{n-1} = \{x \in \mathbb{R}^n : x_n = 0\}$ under any rotation in $\mathbb{R}^n$ that takes the point $(0, 0, \dots, 0, \sqrt{n})$ to $(1, 1, \dots, 1) =: \mathbf{1}$.

Consider the map $\phi : \mathbb{R}^n \to S(0,1)$ defined as $\phi(x) := 0$ if $x = \alpha\mathbf{1}$ for some scalar α, and

$$\phi(x) := \frac{1}{\sigma(x)}(x - \mu(x)\mathbf{1}) \quad \text{otherwise.} \tag{2}$$

Let Z be an n-dimensional standard Gaussian random vector. Let A be an orthogonal matrix satisfying $A\mathbf{1} = \mathbf{1}$. (This is the set of all rotations preserving $S(0,1)$.) Then AZ is again standard Gaussian. Note that

$$\mu(AZ) = \frac{1}{n}\mathbf{1}^T AZ = \frac{1}{n}\mathbf{1}^T Z = \mu(Z).$$

Since $\|AZ\| = \|Z\|$, this implies that $\sigma(AZ) = \sigma(Z)$. Since $\sigma(Z) > 0$ almost surely, the above steps can be combined to give

$$A\phi(Z) = \frac{1}{\sigma(AZ)}(AZ - \mu(AZ)) = \phi(AZ) \stackrel{d}{=} \phi(Z). \tag{3}$$

Since this holds for every rotation A of the sphere $S(0,1)$, $\phi(Z)$ is uniformly distributed on $S(0,1)$.

Now suppose $S(0,1,d) \neq \emptyset$. Then there exists $x \in S(0,1)$ such that $m(x) > d$. Since $m(x) = m(\phi(x))$ for this x and $m \circ \phi$ is a continuous map in a neighborhood of $S(0,1)$, there exists a ball B of positive radius centered at x such that $m(\phi(y)) > d$ for all $y \in B$. Since $\mathbb{P}(Z \in B) > 0$ this shows that $\mathbb{P}(m(\phi(Z)) > d) > 0$, and hence the uniform distribution on $S(0,1)$ puts positive mass on $S(0,1,d)$. Therefore the uniform distribution on $S(0,1,d)$ is simply the restriction of the uniform distribution on $S(0,1)$ to this set. Since

$$K = b' S(0,1,-1/b') + \mathbf{1}, \tag{4}$$

this gives an alternative characterization of the uniform distribution on K that will be convenient for our purposes.

3. From thin sets to thick sets

In this section, we show how to deduce results about K from a slight 'positively tilted' thickening of K, that we call K^ϵ. For any $\epsilon > 0$, let

$$K^\epsilon := \{x \in \mathbb{R}^n_+ : \epsilon < \mu(x) - 1 < 2\epsilon,\ \epsilon < \mu_2(x) - b < b\epsilon\}.$$

Clearly, K^ϵ has nonzero volume whenever it is non-empty, and therefore the uniform distribution on K^ϵ is naturally defined as restriction of the Lebesgue measure, normalized to have mass 1.

Define a map $\psi : \mathbb{R}^n \to \mathbb{R}^n$ as

$$\psi(x) = b'\phi(x) + \mathbf{1}, \tag{5}$$

where ϕ is the map defined in (2) and $b' = \sqrt{b-1}$.

Proposition 3.1. *Let K, K^ϵ and ψ be defined as above, and suppose K^ϵ is non-empty. Let X be a random vector that is uniformly distributed on K,* *and let X^ϵ be a random vector that is uniformly distributed on K^ϵ. Then for any $f : \mathbb{R}_+^n \to [0, \infty)$ and any $\epsilon \in (0, c(b))$,*

$$\mathbb{E}f(X) \le \frac{\mathbb{E}f(\psi(X^\epsilon))}{\mathbb{P}(m(X^\epsilon) > C(b)\epsilon)},$$

where $c(b)(< 1/2)$ and $C(b)$ are positive constants that depend only on the value of b. The right hand side is interpreted as infinity if the denominator is zero.

Proof. Let ϵ be so small that $\epsilon < 1/2$ and $7\epsilon < b - 1$. Define

$$\hat{K}^\epsilon := \{x \in \mathbb{R}^n : \epsilon < \mu(x) - 1 < 2\epsilon,\ \epsilon < \mu_2(x) - b < b\epsilon\},$$

so that $K^\epsilon = \{x \in \hat{K}^\epsilon : m(x) > 0\}$. Note that

$$\begin{aligned} \sigma(x) &= (\mu_2(x) - \mu^2(x))^{1/2} \\ &< (b + b\epsilon - (1+\epsilon)^2)^{1/2} \\ &\le ((b-1)(1+\epsilon))^{1/2} \le b'(1+\epsilon), \end{aligned} \tag{6}$$

and

$$\begin{aligned} \sigma(x) &> (b + \epsilon - (1+2\epsilon)^2)^{1/2} \\ &= (b - 1 - 3\epsilon - 4\epsilon^2)^{1/2} \\ &> ((b-1)(1 - 7\epsilon/(b-1)))^{1/2} > b'(1 - 7\epsilon/(b-1)). \end{aligned} \tag{7}$$

The last inequality shows that, in particular, $\sigma(x) > 0$ and hence x cannot belong to the diagonal line. Let l be the linear transformation

$$l(x) := b'x + \mathbf{1}.$$

Let $d := 1/b'$, so that $l^{-1}(x) = d(x - \mathbf{1})$. As pointed out before in (4), $S(0, 1, -d) = l^{-1}(K)$. Define

$$\hat{S}^\epsilon := l^{-1}(\hat{K}^\epsilon), \quad S^\epsilon := \{x \in \hat{S}^\epsilon : m(x) > -d\} = l^{-1}(K^\epsilon).$$

Let Y^ϵ be uniformly distributed on $\hat{S}^\epsilon$. Since $\hat{K}^\epsilon$ does not intersect the diagonal line, it follows that $\hat{S}^\epsilon$ does not intersect the diagonal line either. We claim that $\phi(Y^\epsilon)$ is uniformly distributed on $S(0, 1)$. To see this, let A be an orthogonal matrix such that $A\mathbf{1} = \mathbf{1}$. As argued to derive (3), we see that

$$A\phi(Y^\epsilon) = \phi(AY^\epsilon) \quad \text{a.s.} \tag{8}$$

Again, as argued before, $\mu(Ax) = \mu(x)$ and $\mu_2(Ax) = \mu_2(x)$ for any x outside the diagonal line, and therefore, A maps $\hat{K}^\epsilon$ onto itself. By the property that $A\mathbf{1} = \mathbf{1}$ it follows that A and l^{-1} commute, and thus A maps $\hat{S}^\epsilon$ onto itself. Since A is a linear map, this shows that AY^ϵ is uniformly distributed on $\hat{S}^\epsilon$. Combined with (8), this proves the claim that $\phi(Y^\epsilon)$ is uniformly distributed on $S(0, 1)$.

Now, clearly,

$$m(\phi(Y^\epsilon)) = \frac{1}{\sigma(Y^\epsilon)}(m(Y^\epsilon) - \mu(Y^\epsilon)).$$

Thus, $m(\phi(Y^\epsilon)) > -d$ if and only if

$$m(Y^\epsilon) > \mu(Y^\epsilon) - d\sigma(Y^\epsilon).$$

Therefore, if U is distributed uniformly on $S(0,1,-d)$, then for any measurable $h : \mathbb{R}^n \to [0,\infty)$,

$$\begin{aligned} \mathbb{E}h(U) &= \mathbb{E}(h(\phi(Y^\epsilon)) \mid m(\phi(Y^\epsilon)) > -d) \\ &= \mathbb{E}(h(\phi(Y^\epsilon)) \mid m(Y^\epsilon) > \mu(Y^\epsilon) - d\sigma(Y^\epsilon)). \end{aligned} \tag{9}$$

Now take any $y \in \hat{S}^\epsilon$ and let $x = l(y)$. Then $x \in \hat{K}^\epsilon$, and therefore by (6),

$$\begin{aligned} \mu(y) - d\sigma(y) &= d(\mu(x) - 1) - d^2\sigma(x) \\ &> d\epsilon - d(1+\epsilon) = -d. \end{aligned}$$

Thus, the event $m(Y^\epsilon) > \mu(Y^\epsilon) - d\sigma(Y^\epsilon)$ implies $m(Y^\epsilon) > -d$. Let Z^ϵ be uniformly distributed on S^ϵ. Then the law of Z^ϵ is the same as that of Y^ϵ conditioned on the event $m(Y^\epsilon) > -d$. Combined with the previous step and (9), we get

$$\begin{aligned} \mathbb{E}h(U) &= \mathbb{E}(h(\phi(Y^\epsilon)) \mid m(Y^\epsilon) > \mu(Y^\epsilon) - d\sigma(Y^\epsilon)) \\ &= \mathbb{E}(h(\phi(Y^\epsilon)) \mid m(Y^\epsilon) > \mu(Y^\epsilon) - d\sigma(Y^\epsilon),\ m(Y^\epsilon) > -d) \\ &= \mathbb{E}(h(\phi(Z^\epsilon)) \mid m(Z^\epsilon) > \mu(Z^\epsilon) - d\sigma(Z^\epsilon)). \end{aligned}$$

Since h is a non-negative function, this implies that

$$\mathbb{E}h(U) \le \frac{\mathbb{E}h(\phi(Z^\epsilon))}{\mathbb{P}(m(Z^\epsilon) > \mu(Z^\epsilon) - d\sigma(Z^\epsilon))}.$$

(If the denominator is zero we interpret the right hand side as infinity.) However, for any $y = l(x) \in \hat{S}^\epsilon$, (7) gives

$$\begin{aligned} \mu(y) - d\sigma(y) &= d(\mu(x) - 1) - d^2\sigma(x) \\ &\le 2d\epsilon - d(1 - 7d^2\epsilon) = -d + (2d + 7d^3)\epsilon. \end{aligned}$$

Thus,

$$\mathbb{E}h(U) \le \frac{\mathbb{E}h(\phi(Z^\epsilon))}{\mathbb{P}(m(Z^\epsilon) > -d + (2d+7d^3)\epsilon)}. \tag{10}$$

Since X has the same law as $l(U)$, we get

$$\mathbb{E}f(X) = \mathbb{E}(f \circ l(U)) \le \frac{\mathbb{E}(f \circ l \circ \phi(Z^\epsilon))}{\mathbb{P}(m(Z^\epsilon) > -d + (2d+7d^3)\epsilon)}.$$

Since $\psi = l \circ \phi$ and $\phi = \phi \circ l$, this gives

$$\mathbb{E}f(X) \le \frac{\mathbb{E}(f \circ \psi(l(Z^\epsilon)))}{\mathbb{P}(m(Z^\epsilon) > -d + (2d+7d^3)\epsilon)}.$$

Again, since l is a linear bijection between S^ϵ and K^ϵ, $l(Z^\epsilon)$ is uniformly distributed on K^ϵ. Thus,

$$\mathbb{E}(f \circ \psi(l(Z^\epsilon))) = \mathbb{E}(f \circ \psi(X^\epsilon)).$$

Finally, note that $m(l(Z^\epsilon)) = d^{-1}m(Z^\epsilon) + 1$, and hence

$$\mathbb{P}(m(Z^\epsilon) > -d + (2d + 7d^3)\epsilon) = \mathbb{P}(m(X^\epsilon) > (2 + 7d^2)\epsilon).$$

This completes the proof. □

Proposition 3.2. *Suppose K^ϵ is non-empty. Let $c(b)$ and $C(b)$ be as in Proposition 3.1, and suppose $\epsilon \in (0, c(b))$. Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is a function such that there is a constant L, such that for all $x, y \in \mathbb{R}^n$,*

$$|g(x) - g(y)| \le L \max_{1\le i\le n} |x_i - y_i|.$$

Then for any $a, t \in \mathbb{R}$,

$$\mathbb{P}(|g(X) - a| > t) \le \frac{\mathbb{P}(|g(X^\epsilon) - a| > t - C_3(b)L\epsilon n)}{\mathbb{P}(m(X^\epsilon) > C(b)\epsilon)},$$

where $C_3(b)$ is another constant depending only on b.

Proof. Take any $x \in K^\epsilon$ and let $y = \psi(x)$. Then for any i, we can use the definition of K^ϵ and the inequalities (6) and (7) to conclude that

$$\begin{aligned}|x_i - y_i| &= \left|\left(1 - \frac{b'}{\sigma(x)}\right)x_i + \frac{b'\mu(x)}{\sigma(x)} - 1\right| \\ &\le \frac{|(b' - \sigma(x))x_i|}{\sigma(x)} + \frac{b'|\mu(x) - 1| + |b' - \sigma(x)|}{\sigma(x)} \\ &\le C_2(b)\epsilon(1 + x_i),\end{aligned}$$

where $C_2(b)$ is a constant depending only on b. Since $c(b) < 1/2$, we have

$$x_i \le \sum_j x_j \le n(1 + 2\epsilon) \le 2n.$$

Thus, taking $C_3(b) = 3C_2(b)$, we have

$$\max_i |x_i - y_i| \le C_2(b)\epsilon(2n + 1) \le C_3(b)\epsilon n.$$

Therefore for any $x \in K^\epsilon$,

$$|g(x) - g(\psi(x))| \le C_3(b)L\epsilon n.$$

In particular, the event $|g(\psi(x)) - a| > t$ implies

$$|g(x) - a| > t - C_3(b)L\epsilon n.$$

Taking $f(x) := 1_{\{|g(x)-a|>t\}}$, we get by Proposition 3.1 that

$$\begin{aligned}\mathbb{P}(|g(X) - a| > t) &\le \frac{\mathbb{P}(|g(\psi(X^\epsilon)) - a| > t)}{\mathbb{P}(m(X^\epsilon) > C(b)\epsilon)} \\ &\le \frac{\mathbb{P}(|g(X^\epsilon) - a| > t - C_3(b)L\epsilon n)}{\mathbb{P}(m(X^\epsilon) > C(b)\epsilon)}.\end{aligned}$$

This completes the proof. □

4. From thick sets to conditional distributions

In this section, we show that the uniform distribution on K^ϵ be can approximated by the distribution of a random vector with independent coordinates conditioned to be in K^ϵ.

For each $(r,s) \in (\mathbb{R}_+ \times \mathbb{R}) \cup (\{0\} \times \mathbb{R}_+)$, let $G_{r,s}$ be the probability distribution on $\mathbb{R}_+$ with probability density proportional to $\exp(-rx^2 - sx)$ on $(0,\infty)$. Note that if $(r,s) \notin (\mathbb{R}_+ \times \mathbb{R}) \cup (\{0\} \times \mathbb{R}_+)$, $\exp(-rx^2 - sx)$ is not integrable on $(0,\infty)$. Henceforth, whenever we say 'for any r,s', we will mean 'for any (r,s) in this admissible region'.

In the following $G_{r,s}^{\otimes n}$ will denote the n-fold product of $G_{r,s}$ as a probability measure on $\mathbb{R}^n$.

Lemma 4.1. *Let $c(b)$, $C(b)$ and $C_3(b)$ be as in Proposition 3.2. Take any $\epsilon \in (0, c(b))$ such that K^ϵ is non-empty. Suppose $Y \sim G_{r,s}^{\otimes n}$ for some r,s. Then for any function $f : K^\epsilon \to [0,\infty)$ we have*

$$e^{-B\epsilon n}\mathbb{E}f(X^\epsilon) \le \mathbb{E}(f(Y) \mid Y \in K^\epsilon) \le e^{B\epsilon n}\mathbb{E}f(X^\epsilon),$$

where $B = 2br + 4|s|$.

Proof. Recall that for $x \in K^\epsilon$,

$$|\mu(x) - 1| \le 2\epsilon, \quad |\mu_2(x) - b| \le b\epsilon.$$

Therefore, if we set $B = 2br + 4|s|$, it follows that

$$\begin{aligned}\mathbb{E}(f(Y) \mid Y \in K^\epsilon) &= \frac{\int_{K^\epsilon} f(x) e^{-rn\mu_2(x) - sn\mu(x)} dx}{\int_{K^\epsilon} e^{-rn\mu_2(x) - sn\mu(x)} dx} \\ &\ge e^{-B\epsilon n} \frac{\int_{K^\epsilon} f(x) e^{-rnb - sn} dx}{\int_{K^\epsilon} e^{-rnb - sn} dx} \\ &= e^{-B\epsilon n} \frac{\int_{K^\epsilon} f(x) dx}{\int_{K^\epsilon} dx} = e^{-B\epsilon n}\mathbb{E}f(X^\epsilon).\end{aligned}$$

Similarly, we get the other bound. □

Proposition 4.2. *Let $c(b)$, $C(b)$ and $C_3(b)$ be as in Proposition 3.2. Take $\epsilon \in (0, c(b))$ such that K^ϵ is non-empty. Suppose g is a function as in Proposition 3.2, and $Y \sim G_{r,s}^{\otimes n}$ for some r,s. Then for any $a, t \in \mathbb{R}$, we have*

$$\mathbb{P}(|g(X) - a| > t) \le e^{2B\epsilon n} \frac{\mathbb{P}(|g(Y) - a| > t - C_3(b)L\epsilon n,\ Y \in K^\epsilon)}{\mathbb{P}(m(Y) > C(b)\epsilon,\ Y \in K^\epsilon)},$$

where $B = 2(3b^2|r| + 2|s|)$.

Proof. By Lemma 4.1 we see that

$$\begin{aligned}&\mathbb{P}(|g(X^\epsilon) - a| > t - C_3(b)L\epsilon n) \\ &\le e^{B\epsilon n}\mathbb{P}(|g(Y^{r,s}) - a| > t - C_3(b)L\epsilon n \mid Y \in K^\epsilon)\end{aligned}$$

and

$$\mathbb{P}(m(X^\epsilon) > C(b)\epsilon) \geq e^{-B\epsilon n}\mathbb{P}(m(Y) > C(b)\epsilon \mid Y \in K^\epsilon).$$

Using these bounds in Proposition 3.2, we get

$$\begin{aligned}\mathbb{P}(|g(X) - a| > t) &\leq e^{2B\epsilon n}\frac{\mathbb{P}(|g(Y) - a| > t - C_3(b)L\epsilon n \mid Y \in K^\epsilon)}{\mathbb{P}(m(Y) > C(b)\epsilon \mid Y \in K^\epsilon)}\\ &= e^{2B\epsilon n}\frac{\mathbb{P}(|g(Y) - a| > t - C_3(b)L\epsilon n,\ Y \in K^\epsilon)}{\mathbb{P}(m(Y^{r,s}) > C(b)\epsilon,\ Y \in K^\epsilon)}.\end{aligned}$$

This completes the proof. □

5. A local limit theorem

In this section we derive some basic properties of the probability distribution $G_{r,s}$ defined in the previous section. Fix r, s, and let $Y_1, Y_2, \ldots$ *i.i.d.* $\sim G_{r,s}$. Suppose $\mathbb{E}(Y_1) = 1$ and let $\beta := \mathbb{E}(Y_1^2)$.

Lemma 5.1. *The pair $(Y_1 + Y_2 + Y_3,\ Y_1^2 + Y_2^2 + Y_3^2)$ has a bounded density in $\mathbb{R}^2$.*

Proof. In this proof, C will denote any positive constant that may depend on b, r or s, but no other parameters. The value of C may change from line to line.

Fix any $u, v \in \mathbb{R}$ such that $u > 0$ and $v > u^2/3$ (the density of $(Y_1 + Y_2 + Y_3, Y_1^2 + Y_2^2 + Y_3^2)$ is zero outside this region), and $\delta \in (0, 1)$. Let

$$E_\delta := \{|Y_1 + Y_2 + Y_3 - u| < \delta,\ |Y_1^2 + Y_2^2 + Y_3^2 - v| < \delta\}.$$

Let A be an orthogonal transformation of $\mathbb{R}^3$ that takes the vector $(1, 1, 1)$ to $(0, 0, \sqrt{3})$. Let $Y = (Y_1, Y_2, Y_3)$ and $Z := AY$. Then

$$Z_3 = \frac{1}{\sqrt{3}}(Y_1 + Y_2 + Y_3), \quad Z_1^2 + Z_2^2 + Z_3^2 = Y_1^2 + Y_2^2 + Y_3^2.$$

Thus, if E_δ happens, then we have

$$|\sqrt{3}Z_3 - u| < \delta, \quad |Z_1^2 + Z_2^2 + Z_3^2 - v| < \delta.$$

From the first inequality, we get

$$|3Z_3^2 - u^2| \leq (\sqrt{3}Z_3 - u)^2 + 2u|\sqrt{3}Z_3 - u| \leq C(1 + u)\delta,$$

Thus, under E_δ,

$$|Z_1^2 + Z_2^2 - (v - u^2/3)| \leq C(1 + u)\delta.$$

Now, the density of Y_1 at y is bounded by $Ce^{-y/C}$ on $[0, \infty)$. Thus, the density of Y at (y_1, y_2, y_3) is bounded by $Ce^{-(y_1+y_2+y_3)/C}$. Since Z is a linear transform of Y and $Y_1 + Y_2 + Y_3 = \sqrt{3}Z_3$, it follows that the density of Z at a point (z_1, z_2, z_3) is bounded by $Ce^{-z_3/C}$. (Note that $z_3 > 0$ on the support of Z.) Combining this fact with the above deduction, we get

$$\begin{aligned}\mathbb{P}(E_\delta) \leq Ce^{-u/C}\mathrm{Vol}(\{z \in \mathbb{R}^3 : |\sqrt{3}z_3 - u| < \delta,&\\ |z_1^2 + z_2^2 - (v - u^2/3)| \leq C(1 + u)\delta\}).&\end{aligned}$$

Now

$$\begin{aligned}
&\mathrm{Vol}(\{z \in \mathbb{R}^3 : |\sqrt{3}z_3 - u| < \delta,\ |z_1^2 + z_2^2 - (v - u^2/3)| \le C(1+u)\delta\}) \\
&= \mathrm{Length}(\{z \in \mathbb{R} : |\sqrt{3}z - u| < \delta\}) \\
&\qquad \times \mathrm{Area}(\{(z_1, z_2) : |z_1^2 + z_2^2 - (v - u^2/3)| \le C(1+u)\delta\}) \\
&\le C\delta \mathrm{Area}(\{(z_1, z_2) : |z_1^2 + z_2^2 - (v - u^2/3)| \le C(1+u)\delta\}).
\end{aligned}$$

For any $a > 0, z_1, z_2 \in \mathbb{R}$,

$$\left|\sqrt{z_1^2 + z_2^2} - \sqrt{a}\right| = \frac{|z_1^2 + z_2^2 - a|}{\sqrt{z_1^2 + z_2^2} + \sqrt{a}} \le \frac{|z_1^2 + z_2^2 - a|}{\sqrt{a}}.$$

Thus, for any $a > c > 0$,

$$\begin{aligned}
&\mathrm{Area}(\{(z_1, z_2) : |z_1^2 + z_2^2 - a| \le c\}) \\
&\le \mathrm{Area}(\{(z_1, z_2) : \left|\sqrt{z_1^2 + z_2^2} - \sqrt{a}\right| \le c/\sqrt{a}\}) \\
&= \pi\left(\left(\sqrt{a} + \frac{c}{\sqrt{a}}\right)^2 - \left(\sqrt{a} - \frac{c}{\sqrt{a}}\right)^2\right) = 4\pi c.
\end{aligned}$$

Again, if $c \ge a > 0$,

$$\begin{aligned}
&\mathrm{Area}(\{(z_1, z_2) : |z_1^2 + z_2^2 - a| \le c\}) \\
&\le \mathrm{Area}(\{(z_1, z_2) : \left|z_1^2 + z_2^2\right| \le 2c\}) = 2\pi c.
\end{aligned}$$

Thus, for any $\delta < 1$,

$$\mathrm{Area}(\{(z_1, z_2) : |z_1^2 + z_2^2 - (v - u^2/3)| \le C(1+u)\delta\}) \le C(1+u)\delta,$$

and consequently,

$$\mathbb{P}(E_\delta) \le Ce^{-u/C}(1+u)\delta^2 \le C\delta^2.$$

From here, it is easy to argue that the distribution of $(Y_1 + Y_2 + Y_3, Y_1^2 + Y_2^2 + Y_3^2)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. Moreover, by Theorem 7.14 of [8] and the last inequality, the Radon-Nikodym derivative is uniformly bounded. □

Lemma 5.2. *The sequence* $V_n := n^{-1/2}(\sum_1^n (Y_i - 1),\ \sum_1^n (Y_i^2 - \beta))$ *satisfies a uniform local limit theorem, meaning that there is a non-degenerate Gaussian density* ρ *on* $\mathbb{R}^2$ *such that if* ρ_n *is the probability density of* V_n*, then*

$$\lim_{n\to\infty} \sup_{(x,y)\in\mathbb{R}^2} |\rho_n(x,y) - \rho(x,y)| = 0.$$

Proof. The result follows directly from Lemma 5.1 and the classical uniform local limit theorem, e.g. Theorem 19.1 in [2]. The non-degeneracy holds because the covariance matrix of (Y_1, Y_1^2) is obviously non-singular. □

Lemma 5.3. *Suppose for each n we have real numbers $a_n \le b_n$, $a'_n \le b'_n$ such that there exist $x_0, y_0 \in \mathbb{R}$, with*

$$\lim_{n\to\infty} \sqrt{n}(a_n - 1) = \lim_{n\to\infty} \sqrt{n}(b_n - 1) = x_0,$$
$$\lim_{n\to\infty} \sqrt{n}(a'_n - \beta) = \lim_{n\to\infty} \sqrt{n}(b'_n - \beta) = y_0.$$

Then

$$\lim_{n\to\infty} \frac{\mathbb{P}(a_n \le \frac{1}{n}\sum_{i=1}^n Y_i \le b_n,\ a'_n \le \frac{1}{n}\sum_{i=1}^n Y_i^2 \le b'_n)}{n(b_n - a_n)(b'_n - a'_n)} = \rho(x_0, y_0),$$

where ρ is as in Lemma 5.2.

Proof. Let ρ_n be as in Lemma 5.2. Let

$$u_n := \sqrt{n}(a_n - 1), \quad v_n := \sqrt{n}(b_n - 1),$$
$$u'_n := \sqrt{n}(a'_n - \beta), \quad v'_n := \sqrt{n}(b'_n - \beta).$$

Then

$$\mathbb{P}(a_n \le \tfrac{1}{n}\textstyle\sum_{i=1}^n Y_i \le b_n,\ a'_n \le \tfrac{1}{n}\sum_{i=1}^n Y_i^2 \le b'_n)$$
$$= \int_{u_n}^{v_n} \int_{u'_n}^{v'_n} \rho_n(x, y)dydx.$$

Let

$$\delta_n := \sup_{(x,y)\in\mathbb{R}^2} |\rho_n(x, y) - \rho(x, y)|.$$

and

$$\tau_n := \sup_{u_n \le x \le v_n,\ u'_n \le y \le v'_n} |\rho(x, y) - \rho(x_0, y_0)|.$$

Then $\delta_n \to 0$ by Lemma 5.2, and $\tau_n \to 0$ due to continuity of ρ. Finally, observe that

$$\left|\int_{u_n}^{v_n} \int_{u'_n}^{v'_n} \rho(x, y)dydx - n(b_n - a_n)(b'_n - a'_n)\rho(x_0, y_0)\right|$$
$$= \left|\int_{u_n}^{v_n} \int_{u'_n}^{v'_n} (\rho(x, y) - \rho(x_0, y_0))dydx\right| \le \tau_n n(b_n - a_n)(b'_n - a'_n),$$

and

$$\left|\int_{u_n}^{v_n} \int_{u'_n}^{v'_n} (\rho_n(x, y) - \rho(x, y))dydx\right| \le \delta_n n(b_n - a_n)(b'_n - a'_n),$$

This completes the proof. □

6. Proof of Theorem 1.1

In this section we consider the situation $1 < b \le 2$. First, we need to show the existence of r, s such that the probability distribution $G_{r,s}$ has first moment 1 and second moment b. The uniqueness of r, s will follow automatically from the distributional convergence result for X_1.

Proposition 6.1. *If $1 < b \le 2$, there exist $r, s \in \mathbb{R}$ such that the probability distribution $G_{r,s}$ defined in Section 4 has mean 1 and second moment b.*

Proof. Clearly, if $W \sim G_{r,s}$ then for any $\alpha > 0$, $\alpha W \sim G_{r',s'}$ for some other r', s'. Thus, it suffices to show that for any $b \in (1, 2]$, there exists r, s such if $W \sim G_{r,s}$, then

$$\theta(r,s) := \frac{\mathbb{E}(W^2)}{(\mathbb{E}(W))^2} = b.$$

It is easy to see that θ is a continuous function of r, s. Since $G_{0,1}$ is just the $Exp(1)$ distribution, $\theta(0,1) = 2$. For each $r > 0$, let $W_r \sim G_{r,1}$. Let $Z_r := \sqrt{r} W_r$. Then the density of Z_r on $[0, \infty)$ is proportional to $\exp(-z^2 - z/\sqrt{r})$. It is easy to argue from here that as $r \to \infty$, Z_r converges in law to Z, which has density proportional to $\exp(-z^2)$. Moreover, the moments of Z_r converge to those of Z.

Thus, by the intermediate value theorem for continuous functions, we see that as r ranges between 0 and ∞, $\theta(r, 1)$ takes all values between $\theta_0 := \mathbb{E}(Z^2)/(\mathbb{E}(Z))^2$ and 2. (It is easily verified that $1 < \theta_0 \le 2$.) Next, for $0 \le u < 1$, let V_u follow the density

$$\rho(v) \propto \exp\biggl(-\frac{(v-u)^2}{1-u}\biggr), \quad v \ge 0.$$

In other words, $V_u \sim G_{1/(1-u), -2u/(1-u)}$. Note that V_0 has the same distribution as Z, and as $u \to 1$, the law of V_u tends to the point mass at 1. Convergence of moments is again easy to prove. Therefore, again, by the intermediate value theorem we see that $\theta(1/(1-u), -2u(1-u))$ ranges over all values between θ_0 and 1 as u varies between 0 and 1. This completes the proof. □

Proof of Theorem 1.1, part (a). Choose r, s such that $G_{r,s}$ has first moment 1 and second moment b. In this proof, C will always denote any positive constant that may depend only on b, r or s and no other parameter. (A priori, we do not yet know that r, s are uniquely determined by b, so we treat them as independent parameters.) Let $c(b)$, $C(b)$ and $C_3(b)$ be as in Proposition 4.2.

Fix $\epsilon = n^{-10}$. It will be evident from the proof that the exponent 10 is not of any consequence; any sufficiently large exponent would do. By Lemma 5.3, we see that for sufficiently large n,

$$C^{-1} n\epsilon^2 \le \mathbb{P}(Y \in K^\epsilon) \le Cn\epsilon^2. \tag{11}$$

(Note that this proves, in particular, that K^ϵ is non-empty.) Now, if $Y_1 \le C(b)\epsilon$ and $Y \in K^\epsilon$, then

$$\frac{1}{n}\sum_{i=2}^n Y_i = \mu(Y) - \frac{Y_1}{n} \in (1+\epsilon - n^{-1}C(b)\epsilon,\ 1+2\epsilon), \tag{12}$$

and

$$\frac{1}{n}\sum_{i=2}^n Y_i^2 = \mu_2(Y) - \frac{Y_1^2}{n} \in (b+\epsilon - n^{-1}C(b)^2\epsilon^2, b+\epsilon b). \tag{13}$$

Let E be the event that the two events (12) and (13) happen. By Lemma 5.3, we see that

$$\mathbb{P}(E) \le Cn\epsilon^2.$$

Moreover, the event E is independent of the event $\{Y_1 \le C(b)\epsilon\}$. Thus,

$$\begin{aligned}\mathbb{P}(Y_1 \le C(b)\epsilon,\ Y \in K^\epsilon) &\le \mathbb{P}(\{Y_1 \le C(b)\epsilon\} \cap E)\\ &= \mathbb{P}(Y_1 \le C(b)\epsilon)\mathbb{P}(E) \le Cn\epsilon^3.\end{aligned}$$

Combining with (11), and observing that $n^2\epsilon^3 \ll n\epsilon^2$, we get that for sufficiently large n,

$$\begin{aligned}&\mathbb{P}(m(Y) > C(b)\epsilon,\ Y \in K^\epsilon)\\ &\ge \mathbb{P}(Y \in K^\epsilon) - n\mathbb{P}(Y_1 \le C(b)\epsilon,\ Y \in K^\epsilon)\\ &\ge C^{-1}n\epsilon^2 - Cn^2\epsilon^3 \ge C^{-1}n\epsilon^2.\end{aligned} \tag{14}$$

Next, let $h : \mathbb{R} \to \mathbb{R}$ be a function satisfying $|h(x)| \le 1$ and $|h(x) - h(y)| \le L|x-y|$ for all $x, y \in \mathbb{R}$, where L is some positive constant. Define $g : \mathbb{R}^n \to \mathbb{R}$ as

$$g(x) := \frac{1}{n}\sum_{i=1}^n h(x_i).$$

Note that

$$|g(x) - g(y)| \le L \max_i |x_i - y_i|.$$

Let $a := \mathbb{E}h(Y_1)$. Then by Hoeffding's tail inequality for sums of independent bounded random variables [5], we have that for any $t > 0$,

$$\mathbb{P}(|g(Y) - a| > t) \le 2e^{-nt^2/2}.$$

Therefore it follows from Proposition 4.2 and (14) that for all $t > C_3(b)L\epsilon n$,

$$\mathbb{P}(|g(X) - a| > t) \le Cn^{-1}\epsilon^{-2}e^{-n(t-C_3(b)L\epsilon n)^2/2}. \tag{15}$$

The tail bound decays rapidly in the regime $t > C(n^{-1}\log n)^{1/2}$; from this it is easy to deduce that

$$\mathbb{E}|g(X) - a| \le C\sqrt{\frac{\log n}{n}}.$$

By Jensen's inequality and symmetry, we have

$$\mathbb{E}|g(X) - a| \ge |\mathbb{E}h(X_1) - a| = |\mathbb{E}h(X_1) - \mathbb{E}h(Y_1)|.$$

This shows the convergence of in law for X_1. To show joint convergence for $X_1, \ldots, X_k$, we proceed as follows. Instead of a single function h, consider k functions $h_1, \ldots, h_k$, each satisfying $|h_i(x)| \le 1$ and $|h_i(x) - h_i(y)| \le L|x-y|$. Define $g_1, \ldots, g_k$ and $a_1, \ldots, a_k$ accordingly. Then $|g_i| \le 1$, and therefore by a simple telescoping argument

$$\mathbb{E}\biggl|\prod_{i=1}^k g_i(X) - \prod_{i=1}^k a_i\biggr| \le k \max_i \mathbb{E}|g_i(X) - a_i| \le Ck\sqrt{\frac{\log n}{n}}.$$

Now, putting $A := \prod a_i$ and using Jensen's inequality, we see that

$$\begin{aligned}\mathbb{E}\biggl|\prod_{i=1}^k g_i(X) - \prod_{i=1}^k a_i\biggr| &\ge \biggl|\frac{1}{n^k}\sum_{1\le i_1,\ldots,i_k\le n} (\mathbb{E}(h_{i_1}(X)h_{i_2}(X)\cdots h_{i_k}(X)) - A)\biggr| \\ &= |\mathbb{E}(h_1(X)\cdots h_k(X)) - A| + O(1/n).\end{aligned}$$

This shows the joint convergence of $X_1, \ldots, X_k$ and completes the proof of part (a). Finally, as we noted before, the distributional convergence automatically proves the uniqueness of r, s. □

Proof of parts (c) *and* (d). Let $g(x) = \max_i x_i$. Then

$$|g(x) - g(y)| \le \max_i |x_i - y_i|.$$

When $b < 2$, we must have $r > 0$. In this situation, it is not difficult to conclude that

$$\mathbb{P}(g(Y) > t) \le ne^{-t^2/C}.$$

Thus by Proposition 4.2 and (14) it follows that for all $t > C_3(b)\epsilon n$,

$$\mathbb{P}(g(X) > t) \le C\epsilon^{-2} e^{-(t - C_3(b)\epsilon n)^2/C}.$$

This shows that there exists a constant C depending only on b such that

$$\lim_{n\to\infty} \mathbb{P}(\max_{1\le i\le n} X_i > C\sqrt{\log n}) = 0.$$

When $b = 2$, we have $r = 0$ and $s = 1$. The argument is exactly the same, except that the tail bound is $e^{-t/C}$ instead of $e^{-t^2/C}$, which gives the $\log n$ instead of $\sqrt{\log n}$. □

Proof of part (b). Let us first prove for $k = 1$. Suppose we want to prove the convergence of the pth moment. Fix n. For $x > 0$, let

$$h(x) := \min\{x^p, (\log n)^{2p}\}.$$

Let us compute a Lipschitz constant for h. If $x > (\log n)^2$ and $y > (\log n)^2$, then $h(x) - h(y) = 0$. If $x \le (\log n)^2$ and $y \le (\log n)^2$, then

$$\begin{aligned}|h(x) - h(y)| &= |x^p - y^p| \\ &= |x - y||x^{p-1} + x^{p-2}y + \cdots + y^{p-1}| \\ &\le r(\log n)^{2p-2}|x - y|.\end{aligned}$$

Finally, if $x \le (\log n)^2$ but $y > (\log n)^2$, the

$$\begin{aligned}|h(x) - h(y)| &= |h(x) - h((\log n)^2)| \\ &\le p(\log n)^{2p-2}|x - (\log n)^2| \le p(\log n)^{2p-2}|x - y|.\end{aligned}$$

Thus, we can take $L = p(\log n)^{2p-2}$ and proceed as in the proof of part (a) to get (15). This proves that $\mathbb{E}\min\{X_1^p, (\log n)^2\} \to \mathbb{E}(Z_1^p)$. Now, from the proof of part (c) and the fact that $0 \le X_1 \le n$, we see that when $1 < b \le 2$,

$$\begin{aligned}|\mathbb{E}\min\{X_1^p, (\log n)^2\} - \mathbb{E}(X_1^p)| &\le n^p \mathbb{P}(X_1 > (\log n)^2) \\ &\le n^p \epsilon^{-2} e^{-(\log n)^2/C} \to 0.\end{aligned}$$

This completes the proof for $k = 1$. For $k > 1$, and a monomial like $x_1^{p_1} \cdots x_k^{p_k}$ we proceed as in part (a) by defining

$$g_i(x) := \frac{1}{n}\sum_{j=1}^n \max\{x_j^{p_i}, (\log n)^{2p_i}\}, \quad i = 1, \ldots, k.$$

Noting that g_i is bounded by $(\log n)^{2p_i}$, the proof can be completed as before. □

7. Proof of Theorem 1.2

In this section we deal with the case $b > 2$. As usual, C will denote any constant that depends only on b. We also set $q := \sqrt{b-2}$, a constant that will occur often.

The proof of the localization draws inspiration from Talagrand's localization theorem for the p-spin Hopfield model (see Section 5.11 of [14]).

Proof of parts (c) *and* (d). Let $Y_1, Y_2, \ldots$ be i.i.d. $Exp(1)$ random variables, and let $Y = (Y_1, \ldots, Y_n)$. Take $\epsilon = n^{-10}$ as before. Let $a := 1/100$. Let $Z_i = Y_i 1_{\{Y_i \le n^a\}}$ and let $v := \mathbb{E}(Z_1^2)$. By Hoeffding's inequality, we have

$$\mathbb{P}\biggl(\biggl|\sum_{i=1}^n (Z_i^2 - v)\biggr| > t\biggr) \le 2e^{-t^2/2n^{1+4a}}.$$

Thus, if we define

$$A := \biggl\{\biggl|\sum_{i=1}^n (Z_i^2 - v)\biggr| > n^{5/6}\biggr\},$$

then

$$\mathbb{P}(A) \le 2\exp\biggl(-\frac{n^{\frac{2}{3}-4a}}{2}\biggr).$$

Next, let B be the event that there is a set $I \subseteq \{1, \ldots, n\}$ of size $k := [n^{(1-a)/2}]$ such that $Y_i > n^a$ for all $i \in I$. Then

$$\mathbb{P}(B) \le \binom{n}{k} e^{-kn^a} \le n^k e^{-kn^a} \le C\exp\biggl(-\frac{n^{(1+a)/2}}{C}\biggr).$$

Let D be the event that $\sum_{i\in I} Y_i > qn^{1/2} + n^{(2-a)/4}$ for some subset $I \subseteq \{1,\ldots,n\}$ of size $< k$. Note that for any j, $\sum_{i=1}^j Y_i$ follows a $Gamma(j,1)$ distribution. Therefore, for any $j \ge 2$, $t > 2$,

$$\begin{aligned}
\mathbb{P}\biggl(\sum_{i=1}^j Y_i > t\biggr) &= \int_t^\infty \frac{x^{j-1}}{(j-1)!}e^{-x}dx \\
&= e^{-t}\int_0^\infty \frac{(x+t)^{j-1}}{(j-1)!}e^{-x}dx \\
&\le e^{-t}\int_0^\infty \frac{2^{j-2}(x^{j-1}+t^{j-1})}{(j-1)!}e^{-x}dx \\
&\le e^{-t}(2^{j-2}+t^{j-1}) \le 2t^{j-1}e^{-t}.
\end{aligned}$$

(Note that the inequality is also true for $j = 1$.) Thus, if n is sufficiently large so that $k < n/2$ and we take $t := qn^{1/2} + n^{(2-a)/4}$, then

$$\mathbb{P}(D) \le \sum_{j=1}^{k-1} \binom{n}{j} 2t^{j-1}e^{-t} \le Ckn^k t^k e^{-t}.$$

Since $k = [n^{(1-a)/2}]$ and $(1-a)/2 < (2-a)/4$, we see that

$$\mathbb{P}(D) \le C\exp\biggl(-qn^{1/2} - \frac{n^{(2-a)/4}}{C}\biggr).$$

Now suppose $A^c \cap B^c \cap D^c \cap \{Y \in K^\epsilon\}$ happens. Let I be the set of i such that $Y_i > n^a$. Since B^c has happened, therefore $|I| < k$. Since D^c has occurred, we must have that

$$\sum_{i\in I} Y_i \le qn^{1/2} + n^{(2-a)/4}. \tag{16}$$

Again, since $Y \in K^\epsilon$, we have

$$\biggl|\sum_{i=1}^n Y_i^2 - bn\biggr| < nb\epsilon = bn^{-9}.$$

But due to A^c,

$$\biggl|\sum_{i\notin I} Y_i^2 - vn\biggr| \le n^{5/6}.$$

Combining the last two inequalities, we get

$$\biggl|\sum_{i\in I} Y_i^2 - (b-v)n\biggr| \le bn^{-9} + n^{5/6} \le Cn^{5/6}.$$

But

$$v = \int_0^{n^a} x^2e^{-x}dx = 2 - \int_{n^a}^\infty x^2e^{-x}dx,$$

and therefore

$$|v-2| \le Ce^{-n^a/C}.$$

Thus, under $A^c \cap B^c \cap D^c \cap \{Y \in K^\epsilon\}$,

$$\left|\sum_{i\in I} Y_i^2 - (b-2)n\right| \le Cn^{5/6}.$$

Let $M^Y := \max_i Y_i$. The above inequality combined with (16) shows that under $A^c \cap B^c \cap C^c \cap \{Y \in K^\epsilon\}$, we have

$$\begin{aligned} q^2 n - Cn^{5/6} &\le \sum_{i\in I} Y_i^2 \\ &\le M^Y \sum_{i\in I} Y_i \le M^Y(qn^{1/2} + n^{(2-a)/4}). \end{aligned}$$

Therefore, since $a/4 < 1/6$,

$$\begin{aligned} M^Y &\ge \frac{q^2 n - Cn^{5/6}}{qn^{1/2} + n^{(2-a)/4}} \\ &= qn^{1/2}\frac{1 - Cn^{-1/6}}{1 + n^{-a/4}} \ge qn^{1/2}(1 - Cn^{-a/4}). \end{aligned}$$

But under D^c we have

$$M^Y \le qn^{1/2} + n^{(2-a)/4}.$$

Thus, under $A^c \cap B^c \cap D^c \cap \{Y \in K^\epsilon\}$, we have

$$|M^Y - qn^{1/2}| \le Cn^{(2-a)/4}.$$

Therefore, from the bounds on $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(D)$ obtained above (and observing that the bound on $\mathbb{P}(D)$ dominates the other two), we get

$$\begin{aligned} &\mathbb{P}(|M^Y - qn^{1/2}| > Cn^{(2-a)/4},\ Y \in K^\epsilon) \\ &\le \mathbb{P}(A \cup B \cup D) \\ &\le C\exp\left(-qn^{1/2} - \frac{n^{(2-a)/4}}{C}\right). \end{aligned} \tag{17}$$

Let M_2^Y be the second largest among the Y_i's. Then either $M_2^Y < n^a$, or under $A^c \cap B^c \cap D^c \cap \{Y \in K^\epsilon\}$,

$$\begin{aligned} M_2^Y &\le \sum_{i\in I} Y_i - M^Y \\ &\le qn^{1/2} + n^{(2-a)/4} - (qn^{1/2} - Cn^{(2-a)/4}) \\ &= Cn^{(2-a)/4}. \end{aligned}$$

Thus, again, we have

$$\begin{aligned} &\mathbb{P}(M_2^Y > Cn^{(2-a)/4},\ Y \in K^\epsilon) \\ &\le C\exp\left(-qn^{1/2} - \frac{n^{(2-a)/4}}{C}\right). \end{aligned} \tag{18}$$

This gives us the bounds on the numerator in Proposition 4.2, except that we have to evaluate the Lipschitz constant L for M and M_2. For a vector x, let $g_1(x)$ and $g_2(x)$ denote the largest and second-largest components of x. Since

$$|g_1(x) - g_1(y)| = |\max_i x_i - \max_i y_i| \le \max_i |x_i - y_i|,$$

it follows that we can take $L = 1$ for g_1. By the same logic,

$$\begin{aligned}|\max_{i<j}(x_i + x_j) - \max_{i<j}(y_i + y_j)| &\le \max_{i<j} |(x_i + x_j) - (y_i + y_j)| \\ &\le 2\max_i |x_i - y_i|.\end{aligned}$$

However, $\max_{i<j}(x_i + x_j) = g_1(x) + g_2(x)$. Thus, we can take $L = 3$ for g_2. Thus by (17), (18) and Proposition 4.2, we have

$$\begin{aligned}&\mathbb{P}(|M - qn^{1/2}| > Cn^{(2-a)/4},\ Y \in K^\epsilon) \\ &\le \frac{C\exp\bigl(-qn^{1/2} - C^{-1}n^{(2-a)/4}\bigr)}{\mathbb{P}(m(Y) > c(b)\epsilon,\ Y \in K^\epsilon)}\end{aligned} \tag{19}$$

and

$$\mathbb{P}(M_2 > Cn^{(2-a)/4},\ Y \in K^\epsilon) \le \frac{C\exp\bigl(-qn^{1/2} - C^{-1}n^{(2-a)/4}\bigr)}{\mathbb{P}(m(Y) > c(b)\epsilon,\ Y \in K^\epsilon)}. \tag{20}$$

Let us now start working on the denominator in the above expressions. Let $\delta := C(b)\epsilon$. Note that

$$\begin{aligned}\mathbb{P}(m(Y) > \delta,\ Y \in K^\epsilon) &= \mathbb{P}(Y \in K^\epsilon \mid m(Y) > \delta)\mathbb{P}(m(Y) > \delta) \\ &= \mathbb{P}(Y \in K^\epsilon \mid m(Y) > \delta)e^{-\delta n}.\end{aligned} \tag{21}$$

Since $Y_1, \ldots, Y_n$ are i.i.d. $Exp(1)$, it follows from the memoryless property of the exponential distribution that the conditional distribution of Y given $m(Y) > \delta$ is the same as the unconditional distribution of $Y + \delta\mathbf{1}$. Thus,

$$\mathbb{P}(Y \in K^\epsilon \mid m(Y) > \delta) = \mathbb{P}(Y + \delta\mathbf{1} \in K^\epsilon). \tag{22}$$

Note that

$$\mu(Y + \delta\mathbf{1}) = \mu(Y) + \delta, \quad \mu_2(Y + \delta\mathbf{1}) = \mu_2(Y) + 2\delta\mu(Y) + \delta^2.$$

Let

$$E := \Bigl\{\bigl|\mu(Y) - \bigl(1 - \delta + \tfrac{3}{2}\epsilon\bigr)\bigr| < \epsilon^2,\ \bigl|\mu_2(Y) - \bigl(b - 2\delta + \tfrac{b+1}{2}\epsilon\bigr)\bigr| < \epsilon^2\Bigr\}.$$

If E happens, then

$$1 + \tfrac{3}{2}\epsilon - \epsilon^2 < \mu(Y) + \delta < 1 + \tfrac{3}{2}\epsilon + \epsilon^2,$$

and thus, if n is sufficiently large (so that $\epsilon^2 = n^{-20} \ll \epsilon$), we have

$$1 + \epsilon < \mu(Y + \delta\mathbf{1}) < 1 + 2\epsilon. \tag{23}$$

Again, under E, we have

$$\begin{aligned}&\left|\mu_2(Y)+2\delta\mu(Y)+\delta^2-(b+\tfrac{b+1}{2}\epsilon)\right|\\&\le\left|\mu_2(Y)-(b-2\delta+\tfrac{b+1}{2}\epsilon)\right|+2\delta|\mu(Y)-1|+\delta^2\\&\le C\epsilon^2.\end{aligned}$$

Thus, if n is sufficiently large, and E happens, then we have

$$\begin{aligned}&b+\epsilon<b+\tfrac{b+1}{2}\epsilon-C\epsilon^2\\&\qquad<\mu_2(Y+\delta\mathbf{1})<b+\tfrac{b+1}{2}\epsilon+C\epsilon^2<b+b\epsilon.\end{aligned}\tag{24}$$

By (23) and (24), we see that E implies $Y+\delta\mathbf{1}\in K^\epsilon$, provided n is large enough. Now let

$$\mu^-(Y):=\frac{1}{n}\sum_{i=2}^n Y_i,\quad \mu_2^-(Y):=\frac{1}{n}\sum_{i=2}^n Y_i^2.$$

Define

$$\begin{aligned}E':=&\left\{\left|\mu^-(Y)-\left(1-\delta+\tfrac{3}{2}\epsilon-qn^{-1/2}\right)\right|<\tfrac{1}{2}\epsilon^2\right\}\\&\cap\left\{\left|\mu_2^-(Y)-\left(2-2\delta+\tfrac{b+1}{2}\epsilon\right)\right|<\tfrac{1}{2}\epsilon^2\right\}\\&\cap\left\{\left|Y_1^2-q^2n\right|<\tfrac{1}{2}\epsilon^2\right\}.\end{aligned}$$

Suppose E' happens. Then

$$\begin{aligned}&\left|\mu_2(Y)-\left(b-2\delta+\tfrac{b+1}{2}\epsilon\right)\right|\\&\le\left|\mu_2^-(Y)-\left(2-2\delta+\tfrac{b+1}{2}\epsilon\right)\right|+\tfrac{1}{n}|Y_1^2-(b-2)n|\\&<\tfrac{1}{2}\epsilon^2+\tfrac{1}{2n}\epsilon^2\le\epsilon^2.\end{aligned}$$

Again, under E',

$$\left|Y_1-qn^{1/2}\right|=\frac{|Y_1^2-q^2n|}{Y_1+qn^{1/2}}\le Cn^{-1/2}\epsilon^2,$$

and therefore, for sufficiently large n,

$$\begin{aligned}&\left|\mu(Y)-\left(1-\delta+\tfrac{3}{2}\epsilon\right)\right|\\&\le\left|\mu^-(Y)-\left(1-\delta+\tfrac{3}{2}\epsilon-qn^{-1/2}\right)\right|+\tfrac{1}{n}\left|Y_1-qn^{1/2}\right|\\&<\tfrac{1}{2}\epsilon^2+Cn^{-3/2}\epsilon^2\le\epsilon^2.\end{aligned}$$

Thus, E' implies E. Since Y_1 is independent of $(Y_2,\dots,Y_n)$ and $\mathbb{E}(Y_i)=1$, $\mathbb{E}(Y_i^2)=2$, we can apply Lemma 5.3 to the pair $(\mu^-(Y),\mu_2^-(Y))$ conclude that

$$\begin{aligned}\mathbb{P}(E')&\ge C^{-1}n\epsilon^4\mathbb{P}(|Y_1^2-q^2n|<\tfrac{1}{2}\epsilon^2)\\&\ge C^{-1}n\epsilon^4\mathbb{P}(|Y_1-qn^{1/2}|<\tfrac{1}{2}qn^{1/2}\,\epsilon^2)\\&\ge C^{-1}n^{3/2}\epsilon^6e^{-qn^{1/2}}.\end{aligned}$$

Therefore by (21) and (22),

$$\begin{aligned}\mathbb{P}(m(Y) > \delta,\ Y \in K^\epsilon) &= \mathbb{P}(Y \in K^\epsilon \mid m(Y) > \delta)e^{-\delta n} \\ &= \mathbb{P}(Y + \delta\mathbf{1} \in K^\epsilon)e^{-\delta n} \\ &\geq \mathbb{P}(E)e^{-\delta n} \geq \mathbb{P}(E')e^{-\delta n} \geq C^{-1}n^{-60}e^{-qn^{1/2}}.\end{aligned} \tag{25}$$

Combining this with (19) and Proposition 4.2 (and the value of L obtained before), we get

$$\mathbb{P}(|M - qn^{1/2}| > Cn^{(2-a)/4}) \leq Ce^{-C^{-1}n^{(2-a)/4}}.$$

Similarly from (20) we get

$$\mathbb{P}(M_2 > Cn^{(2-a)/4}) \leq Ce^{-C^{-1}n^{(2-a)/4}}.$$

This completes the proof of parts (c) and (d). □

Proof of parts (a) *and* (b). Part (b) is obvious by symmetry. So we only have to prove part (a). We proceed exactly as in the proof of part (a) in Theorem 1.1. Let $h : \mathbb{R} \to \mathbb{R}$ be a function satisfying $|h(x)| \leq 1$ and $|h(x) - h(y)| \leq L|x - y|$ for all $x, y \in \mathbb{R}$, where L is some positive constant. Define $g : \mathbb{R}^n \to \mathbb{R}$ as

$$g(x) := \frac{1}{n}\sum_{i=1}^n h(x_i).$$

Setting $a := \mathbb{E}h(Y_1)$ and using Hoeffding's inequality, we get

$$\mathbb{P}(|g(Y) - a| > t) \leq 2e^{-nt^2/2}.$$

However, the lower bound (25) for $\mathbb{P}(m(Y) > C(b)\epsilon,\ Y \in K^\epsilon)$ is different from (14). Using (25) and Proposition (4.2), we get the following analog of (15):

$$\mathbb{P}(|g(X) - a| > t) \leq Ce^{C\sqrt{n}}e^{-n(t - C_3(b)L\epsilon n)^2/2}.$$

The tail bound decays rapidly in the regime $t > Cn^{-1/4}$. This gives

$$\mathbb{E}|g(X) - a| \leq Cn^{-1/4}.$$

As before, by Jensen's inequality we get

$$|\mathbb{E}h(X_1) - a| \leq Cn^{-1/4}.$$

The joint distribution of $(X_1, \ldots, X_k)$ is handled similarly. □

8. Proof of Theorem 1.3

The proof of Theorem 1.3 is basically contained in the earlier proofs. Fix $x_0 > 0$ and $1 < b \le 2$. In the proof of part (a) of Theorem 1.1, if instead of taking a fixed h let us take

$$h_n(x) := \begin{cases} 1 & \text{if } x < x_0, \\ 1 - (x - x_0)n^5 & \text{if } x_0 \le x < x_0 + n^{-5}, \\ 0 & \text{if } x \ge x_0 + n^{-5}, \end{cases}$$

then $|h_n(x)| \le 1$ and $|h_n(x) - h_n(y)| \le n^5|x - y|$ for all x, y. Hereafter we can proceed exactly as in the proof of (15) (taking $L = n^5$) and conclude that

$$|\mathbb{E}h_n(X_1) - \mathbb{E}h_n(Z_1)| \le C\sqrt{\frac{\log n}{n}}. \tag{26}$$

Since Z_1 has a bounded density, this gives

$$\begin{aligned} \mathbb{P}(X_1 \le x_0) &\le \mathbb{E}h_n(X_1) \\ &\le \mathbb{E}h_n(Z_1) + C\sqrt{\frac{\log n}{n}} \\ &\le \mathbb{P}(Z_1 \le x_0) + Cn^{-5} + C\sqrt{\frac{\log n}{n}}. \end{aligned}$$

Next, let us slightly modify the definition of h_n by replacing x_0 with $x_0 - n^{-5}$. Let us call the new function $\tilde{h}_n$. Then (26) holds for $\tilde{h}_n$ too, and hence

$$\begin{aligned} \mathbb{P}(X_1 \le x_0) &\ge \mathbb{E}\tilde{h}_n(X_1) \\ &\ge \mathbb{E}\tilde{h}_n(Z_1) - C\sqrt{\frac{\log n}{n}} \\ &\ge \mathbb{P}(Z_1 \le x_0) - Cn^{-5} - C\sqrt{\frac{\log n}{n}}. \end{aligned}$$

This completes the proof for $k = 1$. The general case is similar, as in the proof of Theorem 1.1. When $b > 2$, the proof is exactly the same, except that the bound in (26) becomes $Cn^{-1/4}$, as in the proof of part (a) of Theorem 1.2.

Acknowledgment. The author thanks Persi Diaconis for bringing this problem to his attention and Julien Barré for posing the problem to Persi, and Assaf Naor for helpful references.

References

[1] Barthe, F., Guédon, O., Mendelson, S. and Naor, A. (2005). A probabilistic approach to the geometry of the ℓ_p^n-ball. *Ann. Probab.* **33** no. 2, 480–513.

[2] Bhattacharya, R. N. and Ranga Rao, R. (1986). *Normal approximation and asymptotic expansions.* Reprint of the 1976 original. Robert E. Krieger Publishing Co., Inc., Melbourne, FL.

[3] Chatterjee, S. and Kirkpatrick, K. (2010). Probabilistic methods for discrete nonlinear Schrödinger equations. *Preprint.* Available at http://arxiv.org/abs/1009.5737

[4] Diaconis, P. and Freedman, D. (1987). A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.* **23** no. 2, suppl., 397–423.

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

[6] Naor, A. and Romik, D. (2003). Projecting the surface measure of the sphere of ℓ_p^n. *Ann. Inst. H. Poincaré Probab. Statist.* **39** no. 2, 241–261.

[7] Rudin, W. (1973). *Functional analysis.* McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg.

[8] Rudin, W. (1987). *Real and complex analysis.* Third edition. McGraw-Hill Book Co., New York.

[9] Rumpf, B. (2004). Simple statistical explanation for the localization of energy in nonlinear lattices with two conserved quantities. *Phys. Rev. E* **69**, 016618.

[10] Schechtman, G. and Schmuckenschläger, M. (1991). Another remark on the volume of the intersection of two L_p^n balls. *Geometric aspects of functional analysis (1989–90)*, 174–178, Lecture Notes in Math., **1469**, Springer, Berlin.

[11] Schechtman, G. and Zinn, J. (1990). On the volume of the intersection of two L_p^n balls. *Proc. Amer. Math. Soc.* **110** no. 1, 217–224.

[12] Schmuckenschläger, M. (1998). Volume of intersections and sections of the unit ball of ℓ_p^n. *Proc. Amer. Math. Soc.* **126** no. 5, 1527–1530.

[13] Schmuckenschläger, M. (2001). CLT and the volume of intersections of ℓ_p^n-balls. *Geom. Dedicata* **85** no. 1-3, 189–195.

[14] Talagrand, M. (2003). *Spin glasses: a challenge for mathematicians. Cavity and mean field models*. Springer-Verlag, Berlin.

Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012