Journal of Mathematical Sciences, Vol. 165, No. 1, 2010

ON CONCENTRATION OF MEASURE ON THE CUBE

S. G. Bobkov
University of Minnesota
127 Vincent Hall, Minneapolis, MN, 55455, USA
bobko001@umn.edu
UDC 517.9

Concentration property of the uniform distribution on the cube is considered for the class of permutation invariant sets. Bibliography: 10 *titles.*

1. Introduction

Denote by $\mathbf{P}$ the uniform probability distribution (i.e., the restricted Lebesgue measure) on the n-dimensional cube $[0,1]^n$. A well-known dimension free concentration property of this measure indicates that, for any measurable set $A \subset \mathbf{R}^n$ with $\mathbf{P}(A) \geqslant 1/2$,

$$\mathbf{P}(A + rB_2) \geqslant 1 - e^{-cr^2}, \qquad r > 0. \tag{1.1}$$

Here, B_2 stands for the Euclidean ball in $\mathbf{R}^n$ with unit radius and center at the origin, so that the Minkowski sum $A + rB_2$ represents an open r-neighborhood of A with respect to the Euclidean distance, and $c > 0$ is a numerical constant. For example, the following argument, proposed in [**1**], may be used to reach the inequality (1.1). Consider the map

$$T(x_1, \ldots, x_n) = (\Phi(x_1), \ldots, \Phi(x_n))$$

from $\mathbf{R}^n$ onto $(0,1)^n$, where Φ is the standard normal distribution function on the line. It pushes forward the canonical Gaussian measure γ_n on $\mathbf{R}^n$ into $\mathbf{P}$, i.e., $\mathbf{P}(A) = \gamma_n(T^{-1}(A))$, and has a Lipschitz constant

$$\|T\|_{\rm Lip} = \|\Phi\|_{\rm Lip} = \frac{1}{\sqrt{2\pi}}.$$

Therefore, starting with the Gaussian isoperimetric inequality

$$\gamma_n(A + rB_2) \geqslant \Phi\left(\Phi^{-1}(\gamma_n(A)) + r\right),$$

which holds for all measurable sets $A \subset \mathbf{R}^n$ and, in the case $\gamma_n(A) \geqslant 1/2$, implies

$$\gamma_n(A + rB_2) \geqslant 1 - e^{-r^2/2},$$

we arrive at (1.1) with $c = \pi$. Many other approaches to this concentration result are also available; see [**2**] for the modern exposition of the concentration of measure phenomenon.

On a functional language, (1.1) is equivalent to saying that, for any $f : \mathbf{R}^n \to \mathbf{R}$ with

$$\|f\|_{\rm Lip} \leqslant 1$$

Translated from *Problems in Mathematical Analysis* **44**, January 2010, pp. 55–64.

and $\mathbf{P}$-mean

$$\mathbf{E}f = \int_{[0,1]^n} f(x)\,dx,$$

we have

$$\mathbf{P}\{|f - \mathbf{E}f| \geqslant r\} \leqslant 2\,e^{-cr^2}, \qquad h > 0 \tag{1.2}$$

(for a certain constant $c > 0$). The right-hand side of (1.2) is asymptotically sharp in the class of all Lipschitz functions. However, in many interesting cases, the estimate (1.2) appears to be rather rough. For example, the values of the Lipschitz function $f(x) = \max\{x_1, \dots, x_n\}$ are concentrated around the point 1 and, moreover, the concentration is getting stronger for growing n. More generally, when one might want to improve (1.2), one considers the problem of the deviations of U-statistics

$$f(x) = \sum U(x_{i_1}, \dots, x_{i_k}),$$

where U is a "nice" function of $k \leqslant n$ variables and where summation is performed over all collections of integers $1 \leqslant i_1 < \dots < i_k \leqslant n$.

These examples inspire to consider the concentration property of the uniform distribution, as well as of other product measures, in the class of permutation invariant subsets of the space. Let us say that a set $A \subset \mathbf{R}^n$ is *symmetric under permutations of coordinates* (or just *permutation invariant*) if with every point $x = (x_1, \dots, x_n)$ it contains all points of the form $x = (x_{\pi(1)}, \dots, x_{\pi(n)})$, where π is an arbitrary permutation of $\{1, \dots, n\}$. For such sets (1.1) may considerably be sharpened.

Theorem 1.1. *For any measurable set $A \subset \mathbf{R}^n$, symmetric under permutations of coordinates and with $\mathbf{P}(A) \geqslant 1/2$,*

$$\mathbf{P}(A + rB_\infty) \geqslant 1 - e^{-cnr^2}, \qquad r > 0, \tag{1.3}$$

where $c > 0$ is an absolute constant.

In the sequel, B_p denotes the unit ℓ^p-ball for the norm

$$\|x\|_p = (|x_1|^p + \dots + |x_n|^p)^{1/p}$$

(where $x = (x_1, \dots, x_n) \in \mathbf{R}^n$, $1 \leqslant p \leqslant \infty$). In particular, $B_\infty = (-1,1)^n$ is the open n-dimensional cube in $\mathbf{R}^n$ with side $(-1,1)$.

A functional formulation of Theorem 1.1 is as follows:

Theorem 1.2. *Let f be a function on $[0,1]^n$, symmetric under permutations of coordinates and such that, for all $x, y \in [0,1]^n$,*

$$|f(x) - f(y)| \leqslant \|x - y\|_\infty. \tag{1.4}$$

Then, for all $r > 0$ with some absolute constant $c > 0$,

$$\mathbf{P}\{|f - \mathbf{E}f| > r\} \leqslant 2\,e^{-cnr^2}. \tag{1.5}$$

Thus, in the class of permutation invariant subsets of $[0,1]^n$ there is a stronger concentration of Gaussian type. To compare (1.1) and (1.3), let us rewrite the latter as

$$\mathbf{P}\left(A + \frac{r}{\sqrt{n}}\,B_\infty\right) \geqslant 1 - e^{-cr^2}$$

and note that

$$A + \frac{r}{\sqrt{n}}\, B_\infty \subset A + rB_2.$$

Similarly, on functions the condition (1.4) is telling us that f is differentiable almost everywhere (a.e.), and its gradient satisfies

$$\|\nabla f\|_1 = \sum_{i=1}^n \left| \frac{\partial f(x)}{\partial x_i} \right| \leqslant 1 \quad \text{a.e.}$$

Since

$$\|\nabla f\|_1 \leqslant \sqrt{n}\, \|\nabla f\|_2,$$

(1.5) implies the Gaussian deviation inequality (1.2).

Theorems 1.1–1.2 can be reduced to an appropriate concentration problem about the uniform distribution μ_n on the simplex

$$\Delta_n = \{y \in \mathbf{R}^n_+ : y_1 + \dots + y_n \leqslant 1\}$$

with respect to the ℓ^1-norm. What is needed for (1.3) to hold is the following assertion.

Theorem 1.3. *For any measurable set $A \subset \Delta_n$ with $\mu_n(A) \geqslant 1/2$, with some absolute constant $c > 0$,*

$$\mu_n(A + rB_1) \geqslant 1 - e^{-cnr^2}, \qquad r > 0. \tag{1.6}$$

This concentration inequality is known, although it was stated for the ℓ^1-ball B_1 in place of Δ_n (which is, in fact, equivalent to the case of the simplex). More precisely, if ν_n is the normalized Lebesgue measure on B_1 and a set $A \subset B_1$ has ν_n-measure at least 1/2, it was first shown by Arias-de-Reyna and Villa [**3**] that

$$\nu_n(A + rB_1) \geqslant 1 - n\, e^{-cnr^2}.$$

(Actually, they established this inequality for all ℓ_p-balls with $1 \leqslant p \leqslant 2$.) Afterwards, Schechtman [**4**] removed the unnecessary factor n. In both cases, the proof is essentially based on Talagrand's isoperimetric theorem for the product exponential measure [**5**]. Although this deep theorem is now standard and there are several different approaches to it (cf. [**6, 7**]), we present a direct inductive proof of the concentration inequality (1.6).

2. Reduction to the Simplex

Given a permutation π of $\{1, \dots, n\}$, introduce

$$\Omega(\pi) = \left\{x \in [0,1]^n : 0 \leqslant x_{\pi(1)} \leqslant \ldots \leqslant x_{\pi(n)} \leqslant 1\right\}.$$

Let Ω_n correspond to the identity permutation $\pi(i) = i$. We thus obtain a modulo 0 partition of the cube into $n!$ subsets. In particular, for any measurable set A in $[0,1]^n$

$$\mathbf{P}(A) = \frac{1}{n!} \sum_\pi \mathbf{P}_\pi(A) \tag{2.1}$$

and

$$\mathbf{P}(A + rB_\infty) = \frac{1}{n!} \sum_\pi \mathbf{P}_\pi(A + rB_\infty),$$

where $\mathbf{P}_\pi$ denotes the uniform distribution on Ω_π. If A is symmetric under permutations of coordinates, the terms in each sum coincide, so

$$\mathbf{P}(A) = \mathbf{P}_n(A), \qquad \mathbf{P}(A + rB_\infty) = \mathbf{P}_n(A + rB_\infty),$$

where $\mathbf{P}_n$ temporarily stands for the uniform distribution on Ω_n.

Similarly to (2.1), given a (measurable, bounded) function f on $[0,1]^n$, we have

$$\int f\, d\mathbf{P} = \frac{1}{n!} \sum_\pi \int f\, d\mathbf{P}_\pi.$$

If f is invariant under permutations of coordinates, the $\mathbf{P}$-distribution of f coincides with its $\mathbf{P}_n$-distribution, and the same can be said about $\|\nabla f\|$ with respect to any permutation invariant norm on $\mathbf{R}^n$.

Now, let us consider the measure $\mathbf{P}_n$. Its supporting set Ω_n may be identified with the simplex

$$\Delta_n = \{y \in \mathbf{R}_+^n : y_1 + \cdots + y_n \leqslant 1\}$$

via the map

$$T : x \to y = (x_1, x_1 - x_2, \dots, x_n - x_{n-1}).$$

This map pushes forward $\mathbf{P}_n$ into the uniform measure μ_n on Δ_n. Hence for any measurable set $A \subset \mathbf{R}^n$ we have

$$\mathbf{P}_n(A) = \mu_n(T(A))$$

and, in view of the linearity of T,

$$\mathbf{P}_n(A + rB_\infty) = \mu_n(T(A) + rT(B_\infty)).$$

Note that $T(B_\infty)$ contains the ℓ^1-ball B_1: starting with a point $y = (y_1, \dots, y_n)$ such that $|y_1| + \cdots + |y_n| \leqslant 1$, the point $T^{-1}(y)$ has coordinates $(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_n)$, so it belongs to B_∞. Consequently,

$$\mathbf{P}_n(A + rB_\infty) \geqslant \mu_n(T(A) + rB_1).$$

By the same reasons, starting with a locally Lipschitz function f on Ω_n, the function $y \to f(T^{-1}(y)) = f(y_1, y_1 + y_2, \dots, y_1 + \cdots + y_n)$ is defined on Δ_n and satisfies

$$\|\nabla f(T^{-1}(y))\|_\infty \leqslant \|(\nabla f)(T^{-1})\|_1.$$

This can be summarized in the following.

Lemma 2.1. *Given numbers $p \in (0,1)$, $r > 0$ and a function R on $[p,1]$, any concentration property on the simplex of the form*

$$\mu_n(A + rB_1) \geqslant R(\mu_n(A)),$$

holding for all measurable sets $A \subset \Delta_n$ with $\mu_n(A) \geqslant p$, implies a similar property on the cube

$$\mathbf{P}(A + rB_\infty) \geqslant R(\mathbf{P}(A))$$

in the class of all measurable sets $A \subset [0,1]^n$, symmetric under permutations of coordinates and such that $\mathbf{P}(A) \geqslant p$.

Let us also state a functional form of Lemma 2.1.

Lemma 2.2. *Let $\alpha = \alpha(r)$ be defined on $(0, +\infty)$. Assume that for every function f on Δ_n such that $|f(x) - f(y) \leqslant \|x - y\|_1$, whenever $x, y \in \Delta_n$, we have*

$$\mu_n\left\{\left|f - \int f\, d\mu_n\right| \geqslant r\right\} \leqslant \alpha(r), \quad r > 0.$$

Then, for every function f on $[0,1]^n$, invariant under permutations of coordinates and such that

$$|f(x) - f(y) \leqslant \|x - y\|_\infty,$$

whenever $x, y \in [0,1]^n$, we have

$$\mathbf{P}\left\{\left|f - \int f\, d\mathbf{P}\right| \geqslant r\right\} \leqslant \alpha(r), \quad h > 0.$$

Remark 2.3. As the above discussion shows, instead of the permutation symmetry in the conclusions in Lemma 2.2 and therefore in Theorem 1.2, it suffices to require that $\mathbf{P}_\pi$-means of f on Ω_π are equal to each other. Similarly, in Lemma 2.1 and Theorem 1.1 one may only require that intersections $A \cap \Omega_\pi$ have equal volumes.

3. Distribution of the Last Coordinate

Thus, Lemmas 2.1 and 2.2 reduce the study of distributions of permutation invariant functionals on the cube to the study of distributions of general functionals on the probability space (Δ_n, μ_n). As the first step, it is useful to look at the last coordinate

$$\eta_n(x) = x_n, \qquad x = (x_1, \dots, x_n),$$

and see how it is concentrated around its μ_n-mean. Denote by F_n the distribution function of η_n. It should be clear that

$$F_n(t) = 1 - (1-t)^n, \quad 0 \leqslant t \leqslant 1,$$

and, by a direct computation,

$$\mathbf{E}\eta_n = \frac{1}{n+1}, \qquad \mathrm{Var}(\eta_n) = \frac{n}{(n+1)^2(n+2)}.$$

The variance is of order $\dfrac{1}{n^2}$. We will need a refinement of this property in terms of the tails of Lipschitz functions of η_n.

Lemma 3.1. *For any function $g : (0,1) \to \mathbf{R}$ with $\|g\|_{\mathrm{Lip}} \leqslant \sigma$ and $\mathbf{E}g(\eta_n) = 0$*

$$\mathbf{E}e^{\lambda g(\eta_n)} \leqslant e^{4\lambda^2\sigma^2/n^2}, \quad \text{for} \quad |\lambda|\,\sigma \leqslant \frac{n}{2}. \tag{3.1}$$

Proof. By the homogeneity of (3.1) with respect to g, we may assume that $\sigma = 1$.

First, it is useful to represent the measure F_n as a Lipschitz transform of the canonical two-sided exponential measure ν on the real line with density

$$p(t) = \frac{1}{2}\, e^{-|t|}, \quad t \in \mathbf{R}.$$

Let

$$F(t) = \int_{-\infty}^{t} p(s)\, ds$$

denote the corresponding distribution function, which has the inverse function $F^{-1} : (0,1) \to \mathbf{R}$. It is easy to check that

$$I(t) \equiv p(F^{-1}(t)) = \min\{t, 1-t\}, \qquad 0 < t < 1.$$

Similarly, introduce

$$I_n(t) \equiv p_n(F_n^{-1}(t)) = n(1-t)^{(n-1)/n}, \qquad 0 < t < 1,$$

where $p_n = F_n'$ is the density of η_n. Consider the increasing map $T_n : \mathbf{R} \to (0,1)$, which pushes forward ν to the measure F_n, i.e., $T_n(t) = F_n^{-1}(F(t))$. Then $T_n'(F^{-1}(t)) = I(t)/I_n(t)$, so T_n has a Lipschitz seminorm

$$\|T_n\|_{\mathrm{Lip}} = \sup_{0<t<1} \frac{I(t)}{I_n(t)} = \frac{I(\frac{1}{2})}{I_n(\frac{1}{2})} = \frac{2^{1/n}}{n}.$$

Hence the superposition $h = g(T_n)$ has a Lipschitz seminorm at most $\dfrac{2^{1/n}}{n}$.

Now, we use the following property of the measure ν (cf. [**8**]), which is an equivalent reformulation of the one-dimensional isoperimetric inequality

$$\nu(A + (-r,r)) \geqslant F(F^{-1}(\nu(A)) + r), \quad r > 0.$$

Namely, for any function $h : \mathbf{R} \to \mathbf{R}$ with $\|h\|_{\mathrm{Lip}} \leqslant b$, there is a nondecreasing function $\widetilde{h} : \mathbf{R} \to \mathbf{R}$, which has the same distribution under ν as h and such that

$$\|\widetilde{h}\|_{\mathrm{Lip}} \leqslant b.$$

On the other hand (cf. [**9**, Lemma 2]), given $\lambda \in \mathbf{R}$ and a random variable η with finite first moment, the quantity

$$\mathbf{E} \exp\{\lambda(h(\eta) - \mathbf{E}\, h(\eta))\}$$

is well defined for all nondecreasing $h : \mathbf{R} \to \mathbf{R}$ with $\|h\|_{\mathrm{Lip}} \leqslant b$ and is maximized in this class at the linear function $h(t) = bt$. Assume that η is distributed according to ν, so that we may put

$$\eta_n = T_n(\eta).$$

Applying this extremal property to the function $\widetilde{h}$, which corresponds to $h = g(T_n)$ and has a Lipschitz seminorm at most $\dfrac{2^{1/n}}{n}$, we conclude that, whenever $|\lambda| < n/2^{1/n}$,

$$\mathbf{E}\, e^{\lambda g(\eta_n)} = \mathbf{E}\, e^{\lambda \widetilde{h}(\eta)} \leqslant \mathbf{E}\, e^{2^{1/n}\lambda\eta/n} = \frac{1}{1 - (\frac{2^{1/n}\lambda}{n})^2}.$$

One can now apply a simple bound

$$\frac{1}{1-s} \leqslant e^{2s},$$

holding in $0 \leqslant s \leqslant \frac{1}{2}$, which implies

$$\mathbf{E}\, e^{\lambda g(\eta_n)} \leqslant e^{2^{1+2/n}\,\lambda^2/n^2},$$

provided that

$$\Big(\frac{2^{1/n}\lambda}{n}\Big)^2 \leqslant \frac{1}{2},$$

i.e.,

$$|\lambda| \leqslant n\, 2^{-(\frac{1}{2}+\frac{1}{n})}.$$

If $n \geqslant 2$, this range covers the interval $|\lambda| \leqslant n/2$, and we obtain (3.1).

Remark 3.2. The statement of Lemma 3.1 can be sharpened for $n = 1$. In this case, the random variable η_1 is uniformly distributed on (0,1), and, by a similar argument, for all $\lambda \in \mathbf{R}$

$$\mathbf{E}e^{\lambda g(\eta_1)} \leqslant \mathbf{E}e^{\lambda\sigma(\eta_1 - \frac{1}{2})} = \frac{\operatorname{sh}(\lambda\sigma/2)}{\lambda\sigma/2} \leqslant e^{\lambda^2\sigma^2/8}.$$

The next observation, obtained by a direct application of the Fubini theorem, shows how the measure F_n is related to the uniform distribution μ_n on the simplex Δ_n.

Lemma 3.3. *Let ξ_n denote a random vector in $\mathbf{R}^n$ distributed according to μ_n. Then for all $n \geqslant 2$ the random vectors*

$$((1-\eta_n)\xi_{n-1}, \eta_n) \quad \text{and} \quad \xi_n$$

are equidistributed provided that ξ_{n-1} and η_n are independent.

4. Deviations of Lipschitz Functions on the Simplex

Now, we are prepared to study the deviation problem on the simplex and prove a functional variant of Theorem 1.3. Recall that $\Delta_n = \{x \in \mathbf{R}^n_+ : x_1 + \dots + x_n \leqslant 1\}$ is equipped with the uniform measure μ_n.

Theorem 4.1. *For every function f on Δ_n such that*

$$|f(x) - f(y)| \leqslant \|x - y\|_1,$$

whenever $x, y \in \Delta_n$, we have

$$\mu_n\left\{\left|f - \int f\,d\mu_n\right| \geqslant r\right\} \leqslant 2\,e^{-cnr^2}, \quad r > 0. \tag{4.1}$$

where $c > 0$ is a numerical constant.

One may take $c = 1/144$, for example.

This inequality may equivalently be rewritten in terms of the Orlicz norm $\|\cdot\|_{\psi_2}$ generated by the Young function $\psi_2(t) = e^{t^2} - 1$ as

$$\left\|f - \int f\,d\mu_n\right\|_{\psi_2} \leqslant \frac{C}{\sqrt{n}}$$

with some numerical constant C. In particular, the μ_n-variances of such functions are bounded by C^2/n (for some different constant).

Proof of Theorem 4.1. One may assume that f is smooth and is such that

$$\max_{1 \leqslant i \leqslant n} \left|\frac{\partial f(x)}{\partial x_i}\right| \leqslant \sigma \tag{4.2}$$

throughout the interior int (Δ_n) of the simplex (where σ is a positive parameter). Using induction on n, we verify under (4.2) the hypothesis

$$\mathbf{E}\,e^{\lambda f(\xi_n)} \leqslant \exp\{\lambda \mathbf{E} f(\xi_n) + c_n \lambda^2 \sigma^2\} \quad \text{for} \quad |\lambda|\,\sigma \leqslant \lambda_n, \tag{4.3}$$

for suitably chosen sequences c_n and λ_n, where ξ_n is a random vector in $\mathbf{R}^n$ with distribution μ_n.

According to Remark 3.2, one may take $c_1 = 1/8$, which fits any $\lambda_1 > 0$.

Induction step. Let $n \geqslant 2$. We associate with $f(x)$, $x = (x_1, \dots, x_{n-1}, x_n) \in \Delta_n$, the functions

$$g(u,t) = f((1-t)u, t), \qquad h(u) = \int_0^1 g(u,t)\, dF_n(t) = \mathbf{E} g(u, \eta_n),$$

where $u = (u_1, \dots, u_{n-1}) \in \Delta_{n-1}$, $0 < t < 1$, and where a random variable η_n has the distribution F_n, as in the previous section. Then for each $i = 1, \dots, n-1$

$$\frac{\partial g(u,t)}{\partial u_i} = (1-t)\, \frac{\partial f((1-t)u,t)}{\partial x_i}$$

which, by (4.2), does not exceed $\sigma(1-t)$ in absolute value. Hence

$$\left|\frac{\partial h(u)}{\partial u_i}\right| \leqslant \sigma \int_0^1 (1-t)\, dF_n(t) = \sigma\, \frac{n}{n+1}. \tag{4.4}$$

In addition,

$$\frac{\partial g(u,t)}{\partial t} = -\sum_{i=1}^{n-1} \frac{\partial f((1-t)u,t)}{\partial x_i}\, u_i + \frac{\partial f((1-t)u,t)}{\partial x_n},$$

so again by (4.2),

$$\left|\frac{\partial g(u,t)}{\partial t}\right| \leqslant \sigma\bigg(\sum_{i=1}^{n-1} u_i + 1\bigg) \leqslant 2\sigma.$$

Hence, by Lemma 3.1 applied to the function $t \to g(u,t)$ with 2σ in place of the Lipschitz seminorm, for all $u \in \Delta_{n-1}$

$$\mathbf{E}\, e^{\lambda g(u,\eta_n)} \leqslant \exp\{\lambda \mathbf{E} g(u,\eta_n) + 16\lambda^2\sigma^2/n^2\} = e^{\lambda h(u)}\, e^{16\lambda^2\sigma^2/n^2}, \qquad |\lambda|\,\sigma \leqslant \frac{n}{4}. \tag{4.5}$$

According to (4.4), with respect to the ℓ^1-distance on Δ_{n-1} the function h has a Lipschitz constant, bounded by $\sigma\, \frac{n}{n+1}$. Hence, by the induction hypothesis (4.3) for the dimension $n-1$,

$$\mathbf{E}\, e^{\lambda h(\xi_{n-1})} \leqslant \exp\{\lambda \mathbf{E} h(\xi_{n-1}) + c_{n-1}(\frac{n}{n+1})^2\lambda^2\sigma^2\} \quad \text{for} \quad |\lambda|\,\sigma \leqslant \frac{n+1}{n}\,\lambda_{n-1}, \tag{4.6}$$

where ξ_{n-1} has distribution μ_{n-1} (ξ_{n-1} is supposed to be independent of η_n). Inserting $u = \xi_{n-1}$ in (4.5), integrating over μ_{n-1}, and applying (4.6), we arrive at

$$\mathbf{E}\, e^{\lambda g(\xi_{n-1},\eta_n)} \leqslant \exp\{\lambda \mathbf{E} h(\xi_{n-1})\}\, e^{\left[c_{n-1}(\frac{n}{n+1})^2 + \frac{16}{n^2}\right]\lambda^2\sigma^2}, \tag{4.7}$$

which holds for all λ such that

$$|\lambda|\,\sigma \leqslant \min\left\{\frac{n}{4}, \frac{n+1}{n}\,\lambda_{n-1}\right\}. \tag{4.8}$$

But by Lemma 3.3, the random variables $f(\xi_n)$ and $g(\xi_{n-1},\eta_n)$ are equidistributed. In addition,

$$\mathbf{E} h(\xi_{n-1}) = \mathbf{E} g(\xi_{n-1},\eta_n) = \mathbf{E} f(\xi_n).$$

Therefore, (4.7) and (4.8) imply the desired statement (4.3) for dimension n provided that

$$c_n \geqslant c_{n-1}\Big(\frac{n}{n+1}\Big)^2 + \frac{16}{n^2}, \qquad \lambda_n \leqslant \min\left\{\frac{n}{4}, \frac{n+1}{n}\,\lambda_{n-1}\right\}.$$

For the second inequality we may take $\lambda_n = \alpha(n+1)$ with a constant satisfying

$$\alpha(n+1) \leqslant \frac{n}{4}.$$

Since $n \geqslant 2$, the optimal choice is $\alpha = 1/6$.

In terms of $d_n = (n+1)\, c_n$, the first inequality may be rewritten as

$$d_n \geqslant \left(1 - \frac{1}{n+1}\right) d_{n-1} + \frac{16\,(n+1)}{n^2}.$$

Hence to prove that the optimal value of d_n is bounded by a constant d (which can be performed by induction on n), it suffices to bound by d the quantity $\frac{16\,(n+1)^2}{n^2}$. The worst situation corresponds to $n = 2$, which yields $d = 36$. Therefore, one may take $c_n = \frac{36}{n+1}$.

Restricting ourselves to the case $\sigma = 1$ and $\mathbf{E} f(\xi_n) = 0$, we thus proved the following: If

$$\max_{1 \leqslant i \leqslant n} \left|\frac{\partial f(x)}{\partial x_i}\right| \leqslant 1 \tag{4.9}$$

in the interior int (Δ_n), then

$$\mathbf{E} e^{\lambda f(\xi_n)} \leqslant e^{36\,\lambda^2/(n+1)} \tag{4.10}$$

as soon as

$$|\lambda| \leqslant \frac{n+1}{6}.$$

But this range of λ is large enough in order to extend (4.10) to all λ on the real line. Indeed, for some point $x_0 \in \Delta_n$ necessarily $f(x_0) = 0$ (since f has mean zero). Since $|f(x) - f(y)| \leqslant \|x - y\|_1$, the function f must take values in $[-1, 1]$, so one always has

$$\int e^{\lambda f}\, d\mu_n \leqslant e^{|\lambda|} \leqslant e^{36\,\lambda^2/(n+1)},$$

where the second inequality is valid for $|\lambda| \geqslant \frac{n+1}{36}$. This allows one to cover the remaining values of λ in (4.10).

Finally, starting with (4.10) and applying the Chebyshev inequality, we get for any $r > 0$

$$\mathrm{Prob}\,\{f(\xi_n) \geqslant r\} \leqslant e^{-(n+1)\, r^2/144}, \tag{4.11}$$

and, combining it with a similar inequality for the function $-f$,

$$\mathrm{Prob}\,\{|f(\xi_n)| \geqslant r\} \leqslant 2\, e^{-(n+1)\, r^2/144}.$$

The latter yields the inequality (4.1) with $c = 1/144$.

At this step, the condition (4.9) may slightly be relaxed to the Lipschitz condition

$$|f(x) - f(y)| \leqslant \|x - y\|_1 \quad (x, y \in \Delta_n).$$

Theorem 4.1 is proved. □

5. Gaussian Concentration on the Simplex

The transition from functional inequalities, such as (4.1) in Theorem 4.1, to concentration inequalities, such as (1.6) in Theorem 1.3, is standard. Let us recall the argument. In fact, it is better to start with the one-sided estimate (4.11), which we write here as

$$\mu_n\left\{f - \int f\, d\mu_n \geqslant r\right\} \leqslant e^{-cnr^2}, \quad r > 0. \tag{5.1}$$

It holds with $c = 1/144$ for any function f on Δ_n such that $|f(x) - f(y)| \leqslant \|x - y\|_1$, whenever $x, y \in \Delta_n$.

First, let us remind one general observation. Given a random variable $\xi \geqslant 0$ with a median at zero, i.e., such that

$$\mathrm{Prob}\,\{\xi > 0\} \leqslant \frac{1}{2},$$

we have

$$\mathbf{E}\xi \leqslant \sqrt{\mathrm{Var}(\xi)}. \tag{5.2}$$

Indeed, putting $F = \{\xi > 0\}$, we get, by the Cauchy inequality,

$$(\mathbf{E}\xi)^2 = (\mathbf{E}\,\xi\,1_F)^2 \leqslant \mathbf{E}\xi^2\,\mathrm{Prob}\,(F) \leqslant \frac{1}{2}\,\mathbf{E}\xi^2,$$

which is the same as (5.2).

Now, as was already mentioned, the inequality (4.1) implies that, up to some numerical constant $C > 0$,

$$\mathrm{Var}_{\mu_n}(f) \leqslant \frac{C^2}{n}, \tag{5.3}$$

where Var_{μ_n} denotes variance with respect to the measure μ_n. In particular, this may be applied to the distance functions

$$f_A(x) \;=\; \mathrm{dist}(A, x) \;=\; \inf_{y \in A}\, \|x - y\|_1, \qquad x \in \Delta_n,$$

associated with nonempty subsets A of Δ_n. If, in addition,

$$\mu_n(A) \geqslant \frac{1}{2},$$

we may combine (5.2) and (5.3) to obtain

$$\mathbf{E}_{\mu_n} f_A = \int f_A(x)\,d\mu_n(x) \leqslant \frac{C}{\sqrt{n}}.$$

Finally, for any

$$r \geqslant \frac{2C}{\sqrt{n}}$$

we have

$$\mu_n\{f_A \geqslant r\} \leqslant \mu_n\Big\{f_A - \mathbf{E}_{\mu_n} f_A \geqslant r - \frac{C}{\sqrt{n}}\Big\} \leqslant \mu_n\Big\{f_A - \mathbf{E}_{\mu_n} f_A \geqslant \frac{r}{2}\Big\} \leqslant \; e^{-cnr^2/4}.$$

where we applied (5.1) at the last step. The obtained inequality continues to hold for

$$0 < r < \frac{2C}{\sqrt{n}}$$

with a smaller constant c (if needed) since

$$e^{-cnr^2/4} \geqslant \frac{1}{2} \geqslant \mu_n\{f_A \geqslant r\}.$$

It remains to note that $\{f_A \geqslant r\}$ represents the complement to the set $A + rB_1$ in Δ_n, i.e., we arrived at the concentration inequality of the form

$$1 - \mu_n(A + rB_1) \leqslant e^{-cnr^2}, \quad r > 0. \tag{5.4}$$

Remark 5.1. Let us also mention an isoperimetric result of Sodin [**10**], which provides us with a similar concentration inequality with respect to the Euclidean distance:

$$1 - \mu_n(A + rB_2) \leqslant e^{-cnr}, \quad r > 0. \tag{5.5}$$

Using $B_2 \subset \sqrt{n}\, B_1$ and making the change of variable, one obtains on the basis of (5.5) an exponential bound

$$1 - \mu_n(A + rB_1) \leqslant \ e^{-c\sqrt{n}r},$$

which is however weaker than (5.4).

Acknowledgement. The work is partially supported by the NSF grant.

References

1. B. S. Cirel'son, I. A. Ibragimov, V. N. Sudakov, "Norms of Gaussian sample functions," In: *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)*, pp. 20–41. *Lect. Notes Math.* **550**, Springer, Berlin (1976).
2. M. Ledoux, *The Concentration of Measure Phenomenon* Am. Math. Soc., Providence, RI (2001).
3. J. Arias-de-Reyna, R. Villa, "The uniform concentration of measure phenomenon in ℓ_p^n $(1 \leqslant p \leqslant 2)$," In: *Geometric Aspects of Functional Analysis*, pp. 13–18. *Lect. Notes Math.* **1745**, Springer, Berlin (2000).
4. G. Schechtman, "An editorial comment on the preceding paper: "The uniform concentration of measure phenomenon in ℓ_p^n $(1 \leqslant p \leqslant 2)$" by J. Arias-de-Reyna and R. Villa," In: *Geometric Aspects of Functional Analysis*, pp. 19–20. *Lect. Notes Math.* **1745**, Springer, Berlin (2000).
5. M. Talagrand, "A new isoperimetric inequality and the concentration of measure phenomenon" In: *Geometric Aspects of Functional Analysis (1989–90)*, pp. 94–124. *Lect. Notes Math.* **1469**, Springer, Berlin (1991).
6. B. Maurey, "Some deviation inequalities," *Geom. Funct. Anal.* **1**, No. 2, 188–197 (1991).
7. S. Bobkov, M. Ledoux, "Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution," *Probab. Theory Related Fields* **107**, No. 3, 383–400 (1997).
8. S. G. Bobkov, "Extremal properties of half-spaces for log-concave distributions," *Ann. Probab.* **24**, No. 1, 35–48 (1996).
9. S. G. Bobkov, "A localized proof of the isoperimetric Bakry-Ledoux inequality and some applications" [in Russian], *Teor. Veroyatnost. Primenen.* **47**, No. 2, 340–346 (2002); English transl.: *Theory Probab. Appl.* **47**, No. 2, 308–314 (2003).
10. S. Sodin, "An isoperimetric inequality on the l_p balls," *Ann. Inst. Henri Poincaré Probab. Stat.* **44**, No. 2, 362–373 (2008).

Submitted date: January 7, 2010