ON THE DEPENDENCE OF THE BERRY–ESSEEN BOUND ON DIMENSION

V. Bentkus[1]

June 2001

Abstract. Let X be a random vector with values in $\mathbb{R}^d$. Assume that X has mean zero and identity covariance. Write $\beta = \mathbf{E}\,|X|^3$ and $a_d = \sup_{\varepsilon>0} \sup_{A\in\mathcal{C}} \Phi((\partial A)^\varepsilon)/(2\varepsilon)$, where $\mathcal{C}$ is the class of convex subsets of $\mathbb{R}^d$, the measure Φ is the standard d-dimensional normal distribution, and $(\partial A)^\varepsilon$ is ε-neighborhood of the boundary ∂A of the set A. For $\Delta_n = \sup\limits_{A\in\mathcal{C}} |\mathbb{P}\{S_n \in A\} - \Phi(A)|$, we prove a Berry–Esseen bound $\Delta_n \le 100 b_d \beta/\sqrt{n}$ with $b_d = \max\{1,\, a_d\}$, where S_n is a normalized sum of n independent copies of X. This result reduces estimation of Δ_n to estimation of the constants a_d. The best known to the date estimate of a_d is $a_d \le 4d^{1/4}$, which yields $\Delta_n \le 400 d^{1/4}\beta/\sqrt{n}$.

1. Introduction and results

We provide a Berry–Esseen type bound with apparently optimal dependence on dimension. In essence, the question is reduced to a question about optimal dependence on dimension of standard normal d-dimensional measures of convex shells. Our proof seems to be less technically involved compared to the classical proofs using characteristic functions (see Bhattacharya and Rao (1986). It seems to be a difficult problem to prove the result using the Fourier method.

Let $\mathbb{R}^d$ be a real Euclidean d-dimensional space of vectors $x = (x_1, \dots, x_d)$ with the norm $|x|^2 = x_1^2 + \dots + x_d^2$ and the scalar product $\langle x, x\rangle = |x|^2$.

Let $X, X_1, \dots, X_n$ be i.i.d. random vectors with common mean $\mathbf{E}\,X_j = 0$ and the identity covariance, so that $\mathbf{E}\,\langle X, x\rangle^2 = |x|^2$. Write $\beta = \mathbf{E}\,|X|^3$. By $Y, Y_1, \dots, Y_n$ we denote i.i.d. standard normal random vectors, that is Gaussian random vectors with mean zero and identity covariance operator. We assume throughout that X's, Y's and

---

[1]Vilnius Institute of Mathematics and Informatics, Lithuania. Research supported by the Max Planck Institute for Mathematics, Bonn.

2000 *Mathematics Subject Classification.* 60F05.

Key words and phrases. multidimensional, Berry–Esseen bound, dependence on dimension.

other random vectors and variables are independent if the contrary is not clear from the context.

Let $\mathcal{A}$ be a class of (measurable) subsets $A \subset \mathbb{R}^d$ (below we assume that all sets and functions are measurable). We shall obtain upper bounds for the quantity

$$\Delta_n \equiv \Delta_n(\mathcal{A}) = \sup_{A \in \mathcal{A}} |\mathbb{P}\{S_n \in A\} - \mathbb{P}\{Y \in A\}|,$$

where $S_n = (X_1 + \cdots + X_n)/\sqrt{n}$. The bounds have the shape

$$\Delta_n \leq c_d(\mathcal{A})\beta/\sqrt{n}, \qquad \beta = \mathbf{E}\,|X|^3, \tag{1.1}$$

with some constant $c_d(\mathcal{A})$ which depends only on the dimension and the class $\mathcal{A}$. For example, for the class $\mathcal{B}$ of (Euclidean) balls we have $\Delta_n(\mathcal{B}) \leq c\beta/\sqrt{n}$, where c is an absolute constant. For the class $\mathcal{C}$ of convex subsets of $\mathbb{R}^d$, it holds $\Delta_n(\mathcal{C}) \leq 400 d^{1/4}\beta/\sqrt{n}$. Whether the dependence on d in the last bound is optimal, remains an open question. One can show (see Nagaev (1976)) that for the class of convex sets $\Delta_n(\mathcal{C}) \geq c_0\beta/\sqrt{n}$, where $c_0 > 0$ is an absolute positive constant. Hence, the upper and lower bounds for $\Delta_n(\mathcal{C})$ differ by the factor $d^{1/4}$. We conjecture that an optimal bound for $\Delta_n(\mathcal{C})$ has to depend on d as $\ln^\delta d$ with some $\delta > 0$. In view of our result (see Theorem 1.1 below), to prove this it suffices to improve the bound (1.4) for standard Gaussian measures of convex shells, which seems to be a difficult problem.

We need some definitions and conditions related to $\mathcal{A}$ and the standard normal distribution $\Phi(A) = \mathbb{P}\{Y \in A\}$.

For a class, say $\mathcal{A}$, of subsets $A \subset \mathbb{R}^d$, consider the following conditions:

i) $\mathcal{A}$ is invariant under rescaling, that is, $aA \in \mathcal{A}$, if $A \in \mathcal{A}$ and a number $a \in \mathbb{R}$ is positive;

ii) $\mathcal{A}$ is shift invariant, that is, $x + A \in \mathcal{A}$, if $A \in \mathcal{A}$ and $x \in \mathbb{R}^d$;

iii) $\mathcal{A}$ is invariant under taking of ε-neighborhoods, for $\varepsilon > 0$. More precisely, A^ε, $A^{-\varepsilon} \in \mathcal{A}$ if $A \in \mathcal{A}$. Here

$$A^\varepsilon = \{x \in \mathbb{R}^d : \ \rho_A(x) \leq \varepsilon\} \qquad \text{and} \qquad A^{-\varepsilon} = \{x \in A : \ B_\varepsilon(x) \subset A\},$$

where $\rho_A(x) = \inf_{y \in A} |x - y|$ is the distance between $A \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$, and $B_\varepsilon(x) = \{y \in \mathbb{R}^d : \ |x - y| \leq \varepsilon\}$.

For example, the class $\mathcal{C}$ of convex subsets of $\mathbb{R}^d$ satisfies the conditions i)–iii). Another example provides the class $\mathcal{B}$ of Euclidean balls $B_r(x)$, $r \geq 0$, $x \in \mathbb{R}^d$. The class of rectangles does not satisfy the condition iii) since an ε-neighborhood of a rectangle is not a rectangle.

We consider the following condition on a class $\mathcal{A}$ of subsets of $\mathbb{R}^d$ and the standard normal distribution $\Phi(A) = \mathbb{P}\{Y \in A\}$: there exist constants, say $a_d = a_d(\mathcal{A})$, depending only on d and $\mathcal{A}$ such that

$$\mathbb{P}\{Y \in A^\varepsilon \setminus A\} \leq a_d\varepsilon, \quad \mathbb{P}\{Y \in A \setminus A^{-\varepsilon}\} \leq a_d\varepsilon, \qquad \text{for all } A \in \mathcal{A} \text{ and } \varepsilon > 0. \tag{1.2}$$

For the class $\mathcal{B}$ of balls we have $\sup_{d\geq 1} a_d(\mathcal{B}) < \infty$. Ball (1993) has proved that

$$e^{-1}\sqrt{\ln d} \leq \sup_{A\in\mathcal{C}} \int_{\partial A} p(x)\,ds \leq 4d^{1/4}, \tag{1.3}$$

where $p(x)$ is the standard normal d-dimensional density and ds is the surface measure on the boundary ∂A of A. Using (1.3) one can show that for the class of convex sets

$$e^{-1}\sqrt{\ln d} \leq a_d(\mathcal{C}) \leq 4d^{1/4}. \tag{1.4}$$

To find precise asymptotics of a_d as $d \to \infty$ remains an open problem; in view of the next Theorem 1.1, a solution of this problem leads to improvements of bounds of type (1.1).

Our result is the following theorem.

Theorem 1.1. *Let a class $\mathcal{A}$ of convex sets satisfy the conditions* i)–iii)*. Furthermore, assume that $\mathcal{A}$ and the standard normal distribution satisfy the condition* (1.2)*. Then*

$$\Delta_n \leq 100 b_d \beta/\sqrt{n}, \qquad\qquad b_d = \max\{1,\, a_d\}.$$

In particular, for the class of convex sets we have $\Delta_n \leq 400 d^{1/4}\beta/\sqrt{n}$. In the case of the class of all Euclidean balls the bound improves to $\Delta_n \leq c\beta/\sqrt{n}$, where c is an absolute constant.

The constants in Theorem 1.1 can be considerably improved, especially for larger d. Our aim was just to provide explicit constants keeping the proof and notation as simple as possible. The conditions i)–iii) on the class $\mathcal{A}$ can be relaxed using a little more refined techniques.

The question of the dependence of the Berry–Esseen bound on dimension has drawn attention of many authors, let us mention here only Nagaev (1976), Senatov (1980), Sazonov (1981), Götze (1991), where the bounds depend on the dimension as d. Our approach is related to that of Bentkus (1986a), where the bound depends on d as $\sqrt{d}$.

The proof of Theorem 1.1 is based on induction in n and Taylor expansions. Usually such proofs are very robust against extensions to general spaces and abstract settings, and can be used as well to prove large deviation theorems (see Bentkus (1986b)). Direct inductional proofs is a commonly used tool to obtain convergence rates in the CLT, see, for example, books Sazonov (1981), Paulauskas and Račkauskas (1989), Senatov (1998). A feature of our proof is that we apply a special smoothing with differentiable functions approximating indicator functions of convex sets. The smoothness properties of these functions rely on nice differentiability properties of Euclidean distances. In the context of limit theorems smoothness of Euclidean distances were earlier applied by Senatov (1984) to estimate Prokhorov's metric.

2. Auxiliary Lemmas

We provide the proof of Theorem 1.1 a little later since first we would like to formulate and prove several auxiliary Lemmas.

Results like the next (very simple technically) Lemma 2.1 usually are called smoothing inequalities or lemmas. Smoothing inequalities allow to replace probabilities like $\mathbb{P}\{S_n \in A\} = \mathbf{E}\,\mathbb{I}\{S_n \in A\}$, which are expectations of non-differentiable indicator functions $x \mapsto \mathbb{I}\{x \in A\}$, by expectations of smooth functions, say φ. Such a replacement allows to apply the differential calculus. In Lemma 2.1 we use smooth functions $x \mapsto \varphi(x)$ which approximate indicator functions. Another often used way to make smoothing is via convolutions with certain measures and Fourier transforms (or characteristic functions), that is, via the family of exponential functions $\varphi_t(x) = \exp\{i\langle t, x\rangle\}$, $t \in \mathbb{R}^d$, see Bhattacharya and Rao (1986).

Lemma 2.1. *Let $\varepsilon > 0$ be a positive number. Let a set $A \subset \mathbb{R}^d$ and functions $0 \le \varphi_1, \varphi_2 \le 1$ satisfy*

$$\varphi_1(x) = 1, \text{ for } x \in A, \quad \varphi_1(x) = 0, \text{ for } x \notin A^\varepsilon,$$

$$\varphi_2(x) = 1, \text{ for } x \in A^{-\varepsilon}, \quad \varphi_2(x) = 0, \text{ for } x \notin A.$$

Then, for arbitrary r. v. X and Y, the difference $\gamma(A) = \mathbb{P}\{X \in A\} - \mathbb{P}\{Y \in A\}$ satisfies

$$|\gamma(A)| \le \max_{k=1,2} |\gamma(\varphi_k)| + \max\Big\{\mathbb{P}\{Y \in A^\varepsilon \setminus A\},\ \mathbb{P}\{Y \in A \setminus A^{-\varepsilon}\}\Big\}, \tag{2.1}$$

where $\gamma(\varphi_k) = \mathbf{E}\,\varphi_k(X) - \mathbf{E}\,\varphi_k(Y)$.

Let a class $\mathcal{A}$ of sets satisfy the condition iii). *Assume that we have a family, say $\{\varphi_{\varepsilon,A} :\ A \in \mathcal{A}\}$, of functions $\varphi_{\varepsilon,A} : \mathbb{R}^d \to \mathbb{R}$ such that $0 \le \varphi_{\varepsilon,A} \le 1$ and $\varphi_{\varepsilon,A}(x) = 1$ for $x \in A$, $\varphi_{\varepsilon,A}(x) = 0$, for $x \notin A^\varepsilon$. Then*

$$\sup_{A\in\mathcal{A}} |\gamma(A)| \le \sup_{A\in\mathcal{A}} |\gamma(\varphi_{\varepsilon,A})| + \max\Big\{\sup_{A\in\mathcal{A}} \mathbb{P}\{Y \in A^\varepsilon \setminus A\},\ \sup_{A\in\mathcal{A}} \mathbb{P}\{Y \in A \setminus A^{-\varepsilon}\}\Big\} \tag{2.2}$$

with $\gamma(\varphi_{\varepsilon,A}) = \mathbf{E}\,\varphi_{\varepsilon,A}(X) - \mathbf{E}\,\varphi_{\varepsilon,A}(Y)$.

Proof. We prove only (2.2) since the proof of (2.1) is simpler. We consider first the case $\gamma(A) \ge 0$. Writing for brevity $\varphi = \varphi_{\varepsilon,A}$, we have

$$\begin{aligned} |\gamma(A)| = \gamma(A) &\le \mathbf{E}\,\varphi(X) - \mathbf{E}\,\varphi(Y) + \mathbf{E}\,\varphi(Y) - \mathbb{P}\{Y \in A\} \\ &\le |\,\mathbf{E}\,\varphi(X) - \mathbf{E}\,\varphi(Y)| + \mathbb{P}\{Y \in A^\varepsilon \setminus A\} \end{aligned} \tag{2.3}$$

since the function φ satisfies $0 \le \varphi \le 1$ and equals 1 on the set A. Taking $\sup\limits_{A\in\mathcal{A}}$ in (2.3) leads to (2.2).

Now consider the case $\gamma(A) < 0$. We have to estimate $-\gamma(A) = \mathbb{P}\{Y \in A\} - \mathbb{P}\{X \in A\}$. The proof depends on whether $A^{-\varepsilon} = \varnothing$ or $A^{-\varepsilon} \neq \varnothing$.

If $A^{-\varepsilon} = \varnothing$ then $\mathbb{P}\{Y \in A^{-\varepsilon}\} = 0$ and

$$-\gamma(A) \le \mathbb{P}\{Y \in A\} = \mathbb{P}\{Y \in A\} - \mathbb{P}\{Y \in A^{-\varepsilon}\} \le \mathbb{P}\{Y \in A \setminus A^{-\varepsilon}\}. \tag{2.4}$$

Taking $\sup\limits_{A\in\mathcal{A}}$ in (2.4) we get (2.2).

If $A^{-\varepsilon} \neq \varnothing$, we can in essence repeat the proof of the case $\gamma(A) \ge 0$. Writing for brevity $\varphi = \varphi_{\varepsilon, A^{-\varepsilon}}$, we have

$$\begin{aligned} -\gamma(A) &\le \mathbb{P}\{Y \in A\} - \mathbf{E}\,\varphi(Y) + \mathbf{E}\,\varphi(Y) - \mathbf{E}\,\varphi(X) \\ &\le \mathbb{P}\{Y \in A \setminus A^{-\varepsilon}\} + |\,\mathbf{E}\,\varphi(X) - \mathbf{E}\,\varphi(Y)| \end{aligned} \tag{2.5}$$

since the function φ satisfies $0 \le \varphi \le 1$ and equals 1 on the set $A^{-\varepsilon}$. Noting that by the condition iii) $A^{-\varepsilon} \in \mathcal{A}$ and taking in (2.5) $\sup\limits_{A^{-\varepsilon}} \le \sup\limits_{A}$, we obtain (2.2). □

Lemma 2.2. *Let $A \subset \mathbb{R}^d$ be a closed convex set. Then the distance function*

$$\rho(x) = \min\{|x-y| :\ y \in A\}$$

satisfies $|\rho'(x)| \le 1$, for $x \in A^c = \mathbb{R}^d \setminus A$. Furthermore, its derivative satisfies the following Lipschitz type condition

$$|\rho'(x) - \rho'(y)| \le |x-y|/r, \qquad \text{for all } x, y \in A^c, \tag{2.6}$$

with $r = \min\{\rho(x),\, \rho(y)\}$.

Proof. Without loss of generality we assume that $r = \rho(x) \ge s = \rho(y)$.

For $x \in A^c$, by x_0 we denote a unique point $x_0 \in A$ closest to x. Consider the unit vector $e_1 = (x-x_0)/|x-x_0|$ and the ray $R(x_0) = \{x_0 + te_1 : t > 0\}$. Then $\rho'(z) = e_1$, for all $z \in R_1$. Let e_2 and $R(y_0)$ be a unit vector and ray corresponding to y_0. It is clear that $x = x_0 + re_1 \in \mathbb{R}(x_0)$ and $y = y_0 + se_2 \in \mathbb{R}(y_0)$. Introduce the hyper-plane P_1 (resp. P_2) which is a hyper-plane orthogonal to the ray R_1 (resp. R_2) such that $x_0 \in P_1$ (resp. $y_0 \in P_2$). Analytically P_1 is given by $P_1 = \{x_0 + u \in \mathbb{R}^d :\ \langle u, e_1\rangle = 0\}$. It is clear that the set A is a subset of the half-space $P_1^- = \{x_0 + u \in \mathbb{R}^d :\ \langle u, e_1\rangle \le 0\}$.

We consider separately the following to cases:

i) $P_1 \cap P_2 \neq \varnothing$;

ii) $P_1 \cap P_2 = \varnothing$.

i) There exists $z \in P_1 \cap P_2$. Consider the rays

$$R_3 = \{z + te_1 : t > 0\} \qquad \text{and} \qquad R_4 = \{z + te_2 : t > 0\}$$

and points $\overline{x} = z + re_1 \in R_3$ and $\overline{y} = z + se_2 \in R_4$. It is clear that

$$|\overline{x} - \overline{y}| \le |x - y|. \tag{2.7}$$

Indeed, $A \subset P_1^- \cap P_2^-$ since A is a convex set. The angle between planes P_1 and P_2 is bounded from above by π since A is convex. Therefore, if we have two rays, say $R(u)$ and $R(v)$, then the distance between points $u + re_1 \in R(u)$ and $v + se_2 \in R(v)$ becomes minimal when $u = v \in P_1 \cap P_2$, which proves (2.7) (this fact is obvious geometrically and it is not difficult to provide an analytical proof of it). The inequality (2.7) and $\rho'(x) = e_1$, $\rho'(y) = e_2$ show that (2.6) is implied by the following inequality

$$|e_1 - e_2| \le |\overline{x} - \overline{y}|/r = |e_1 - se_2/r|. \tag{2.8}$$

The inequality (2.8) is just a simple fact from the planar geometry. Nonetheless, let us provide an analytical proof of it. Taking the square, (2.8) reduces to

$$2 - 2\big\langle e_1, e_2\big\rangle \le 1 + \frac{s^2}{r^2} - 2\,\frac{s}{r}\,\big\langle e_1, e_2\big\rangle. \tag{2.9}$$

Note that $0 \le \big\langle e_1, e_2\big\rangle \le 1$ since the angle between planes P_1 and P_2 is bounded from above by π. Hence, for $s \ge r$, the right hand side of (2.9) attains its minimal value $2 - 2\big\langle e_1, e_2\big\rangle$ when $s = r$, which proves (2.9) and (2.6) in the case i).

ii) Now the planes P_1 and P_2 are parallel and $e_2 = -e_1$. Therefore $\rho'(x) = e_1$ and $\rho'(y) = -e_1$. The inequality (2.6) takes the form $2 \le |x_0 - y_0 + (r+s)e_1|/r$. It is clear geometrically that $|x_0 - y_0 + (r+s)e_1|$ assumes its minimal value when $P_1 = P_2$, $x_0 = y_0$ and $s = r$. This value is 2, which proves ii) and the lemma. □

Lemma 2.3. *Let a set $A \subset \mathbb{R}^d$ be convex. Then for any $\varepsilon > 0$ there exist a function φ (which depends only on ε and A) such that*

$$\varphi(x) = 1, \quad \textit{for } x \in A, \qquad \varphi(x) = 0, \quad \textit{for } x \in \mathbb{R}^d \setminus A^\varepsilon, \qquad 0 \le \varphi \le 1, \tag{2.10}$$

and

$$|\varphi'(x)| \le \frac{2}{\varepsilon}, \qquad |\varphi'(x) - \varphi'(y)| \le \frac{8|x-y|}{\varepsilon^2}. \tag{2.11}$$

Furthermore, we can choose φ to have the form $\varphi(x) = \psi(\rho(x)/\varepsilon)$, where $\rho(x)$ is the distance between x and A, and $\psi : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable non-negative non-increasing function such that $\int_{\mathbb{R}} |\psi'(t)|\, dt = 1$.

Proof. Without loss of generality we assume that the set A is closed.

Let $\psi : \mathbb{R} \to [0,1]$ be a function such that $\psi''(x) = 0$, for $x \le 0$ or $x \ge 1$, and $\psi''(x) = -4$, for $0 < x \le 1/2$, $\psi''(x) = 4$, for $1/2 < x < 1$. We define

$$\psi'(x) = \int_{-\infty}^{x} \psi''(t)dt \qquad \text{and} \qquad \psi(x) = 1 + \int_{-\infty}^{x} \psi'(t)dt.$$

The function ψ is non-increasing and has the following properties: $\psi(x) = 1$, for $x \le 0$, $\psi(x) = 0$, for $x \ge 1$, $0 \le \psi \le 1$,

$$|\psi'(t)| \le 4|t|, \qquad |\psi'(t)| \le 2, \qquad |\psi'(t) - \psi'(s)| \le 4|t - s|. \tag{2.12}$$

We define $\varphi(x) = \psi(\rho(x)/\varepsilon)$. The function φ clearly satisfies the relations (2.10).

Using Lemma 2.2 and the properties of ψ, we have

$$|\varphi'(x)| = \varepsilon^{-1}|\psi'(\rho(x)/\varepsilon)\rho'(x)| \le 2/\varepsilon,$$

which proves the first inequality in (2.11).

Let us prove the second inequality in (2.11). Let $x, y \in \mathbb{R}^d$.

If both $x, y \in A$, then $\varphi'(x) = \varphi'(y) = 0$ and the second inequality in (2.11) is obviously fulfilled. To check that $\varphi'(x) = 0$, for $x \in A$, we consider two cases: i) x is in the interrior $A \setminus \partial A$ of A; ii) $x \in \partial A$. In the case i) $\varphi'(x) = 0$ since φ is a constant function in the open set $A \setminus \partial A$. In the case ii), using $|\psi'(t)| \le 4|t|$, we have

$$|\varphi(x + h) - \varphi(x)| \le \int\limits_0^{\rho(x+h)} |\psi'(t)|\, dt \le 4\rho^2(x + h) \le 4|h|^2,$$

which implies $\varphi'(x) = 0$.

Now let both $x, y \in A^c$. We can assume that $r = \rho(x) \le s = \rho(y)$. Lemma 2.2 and properties of ψ yield

$$\begin{aligned}
|\varphi'(x) - \varphi'(y)| &= \varepsilon^{-1}|\psi'(r/\varepsilon)\rho'(x) - \psi'(s/\varepsilon)\rho'(y)| \\
&\le \varepsilon^{-1}|\psi'(r/\varepsilon)|\,|\rho'(x) - \rho'(y)| + \varepsilon^{-1}|\psi'(r/\varepsilon) - \psi'(s/\varepsilon)|\,|\rho'(y)| \\
&\le \frac{4r}{\varepsilon^2}\,\frac{|x - y|}{r} + \frac{4}{\varepsilon^2}\,|r - s| \le \frac{8}{\varepsilon^2}\,|x - y|,
\end{aligned}$$

which proves the second inequality in (2.11).

It remains to consider the case $x \in A$ and $y \in A^c$. Now $\varphi'(x) = 0$ and the second inequality in (2.11) takes the form $|\varphi'(y)| \le 8|x - y|/\varepsilon^2$. Write $s = \rho(y)$. Using $|\psi'(t)| \le 4|t|$, we have

$$|\varphi'(y)| = \varepsilon^{-1}\,|\psi'(s/\varepsilon)| \le \frac{4s}{\varepsilon^2} \le \frac{4|x - y|}{\varepsilon^2}$$

since $s = \rho(y) \le |x - y|$. This completes the proof of (2.11) and of the lemma. □

The next two simple technical Lemmas will be often used in the forthcoming proof of Theorem 1.1. To prove the bound $\Delta_n \le cb_d\beta/\sqrt{n}$ with an inexplicit absolute constant c, all we need is that expectations and integrals in Lemmas 2.4–2.5 are bounded from above by $c_0\beta$ (with an absolute constant c_0), which is almost obvious.

Lemma 2.4. *Write* $\sigma = 2/\sqrt{2\pi}$. *We have*

$$1 \le d^{-3/2}\beta, \quad \mathbf{E}\,|X| \le \beta, \quad \mathbf{E}\,|Y| \le \sigma\beta \le \beta, \quad \mathbf{E}\,|Y|^3 \le 2\sigma\beta \le 2\beta \tag{2.13}$$

$$\mathbf{E}\,|X|^2|\langle X, Y\rangle| \le \sigma\beta \le \beta, \quad \mathbf{E}\,|Y_1|^2|\langle Y_1, Y\rangle| \le 2\sigma^2\beta \le 2\beta, \tag{2.14}$$

$$\begin{aligned} &\mathbf{E}\,|\langle X, Y\rangle| \le \sigma\beta \le \beta, \quad \mathbf{E}\,|\langle Y_1, Y\rangle| \le \sigma^2\beta \le \beta, \\ &\mathbf{E}\,|\langle X, Y\rangle|^3 \le 2\sigma\beta \le 2\beta, \quad \mathbf{E}\,|\langle Y_1, Y\rangle|^3 \le 4\sigma^2\beta \le 3\beta. \end{aligned} \tag{2.15}$$

Proof. To prove that $\beta \ge d^3/2$, it suffices to notice that $d = \mathbf{E}\,|X|^2 \le \beta^{2/3}$. The inequality $\mathbf{E}\,|X| \le \beta$ is implied by an application of Hölder inequality and $\beta \ge 1$.

To prove other inequalities, let us consider first the case $d = 1$. Then $\langle X, Y\rangle = XY$, the random variable Y is a standard normal variable, and (2.13)–(2.15) are implied by an application of $\mathbf{E}\,|Y| = \sigma$ and $\mathbf{E}\,|Y|^3 = 2\sigma$.

Henceforth we assume that $d \ge 2$. Let η, η_j be independent real standard normal variables. Using that coordinates η_i of $Y = (\eta_1, \dots, \eta_d)$ are independent and identically distributed, $\mathbf{E}\,\eta^4 = 3$ and $\mathbf{E}\,\eta^2 = 1$, we have

$$\mathbf{E}\,|Y|^4 = d\,\mathbf{E}\,\eta^4 + (d^2 - d)(\,\mathbf{E}\,\eta^2)^2 = d^2 + 2d.$$

Therefore, for $d \ge 2$, we have

$$\mathbf{E}\,|Y|^3 \le (\,\mathbf{E}\,|Y|^2)^{1/2}(\,\mathbf{E}\,|Y|^4)^{1/2} = d^{3/2}\sqrt{1 + 2/d} \le \sqrt{2}\beta \le 2\sigma\beta.$$

Using polar coordinates and integrating by parts we have

$$\mathbf{E}\,|Y|^3 = s_{d-1}\int_0^\infty r^{d+2}\exp\{-r^2/2\}dr = (d+1)s_{d-1}\int_0^\infty r^d\{-r^2/2\}dr = (d+1)\,\mathbf{E}\,|Y|,$$

where s_{d-1} is the area of the unit sphere $\{x \in \mathbb{R}^d : |x| = 1\}$. Hence, $\mathbf{E}\,|Y|^3 \ge 2\,\mathbf{E}\,|Y|$, for $d \ge 2$. Combining this inequality with the already proved $\mathbf{E}\,|Y|^3 \le 2\sigma\beta$, we conclude the proof of (2.13).

Let us prove (2.14) and (2.15). Given X, the random variable $\langle X, Y\rangle$ is a mean zero normal random variable with variance $|X|^2$. Therefore, conditioning on X, we have $\mathbf{E}\,|X|^2|\langle X, Y\rangle| = \sigma\,\mathbf{E}\,|X|^3 = \sigma\beta$. The proof of the second inequality in (2.14) is similar. We have $\mathbf{E}\,|\langle X, Y\rangle| = \sigma\,\mathbf{E}\,|X| \le \sigma\beta$ and $\mathbf{E}\,|\langle X, Y\rangle|^3 = 2\sigma\,\mathbf{E}\,|X|^3 = 2\sigma\beta$. The proof of the last inequality in (2.15) is similar. □

Lemma 2.5. *Let*

$$p(y) = (2\pi)^{-d/2} \exp\{-|y|^2/2\}, \qquad y \in \mathbb{R}^d, \tag{2.16}$$

be the standard normal d-dimensional density. Then we have

$$A = \mathbf{E} \int_{\mathbb{R}^d} |p'(y)X|\, dy \le \sigma\beta \le \beta, \qquad B = \mathbf{E} \int_{\mathbb{R}^d} |p'(y)Y|\, dy \le \sigma^2\beta \le \beta, \tag{2.17}$$

and

$$C = \mathbf{E} \int_{\mathbb{R}^d} |p'''(y)X^3|\, dy \le \zeta\beta \le 2\beta, \qquad D = \mathbf{E} \int_{\mathbb{R}^d} |p'''(y)Y^3|\, dy \le 2\sigma\zeta\beta \le 3\beta \tag{2.18}$$

with $\sigma = 2/\sqrt{2\pi}$ *and* $\zeta = 2(1 + 4\,\mathrm{e}^{-3/2})/\sqrt{2\pi}$.

Proof. We have

$$p'(x)h = -\langle x, h\rangle p(x), \qquad p'''(x)h^3 = (3|h|^2\langle x, h\rangle - \langle x, h\rangle^3)p(x). \tag{2.19}$$

Let us prove (2.17). Using (2.19) we can write $A = \mathbf{E}\,|\langle Y, X\rangle|$. Applying (2.15) we get $A \le \sigma\beta$. Similarly $B \le \sigma^2\beta$.

To prove (2.18) we notice that

$$\mathbf{E}\left|3|h|^2\langle Y, h\rangle - \langle Y, h\rangle^3\right| = |h|^3\,\mathbf{E}\,|3\eta - \eta^3| = 2(1 + 4\,\mathrm{e}^{-3/2})/\sqrt{2\pi},$$

where η is a standard real normal random variable. Applying (2.13), we derive (2.18). □

For a sufficiently smooth function $f : \mathbb{R}^d \to \mathbb{R}$, we shall apply the Taylor formula

$$f(x+h) = f(x) + f'(x)h + \cdots + \frac{1}{s!}\, f^{(s)}(x)h^s + \frac{1}{s!}\,\mathbf{E}\,(1-\tau)^s f^{(s+1)}(x+\tau h)h^{s+1}, \tag{2.20}$$

where τ is a random variable uniformly distributed on the interval $[0, 1]$.

3. Proof of Theorem 1.1

The proof is based on induction in n. We shall prove that

$$\Delta_n \le M b_d \delta, \qquad \delta \overset{\text{def}}{=} \beta/\sqrt{n}, \tag{3.1}$$

holds for all n provided that an absolute constant $M \ge 10$ is sufficiently large (recall that $b_d = \max\{1, a_d\}$).

Since $\beta \ge 1$ and $\Delta_n \le 1$, for $n \le M^2\beta^2$ we have

$$\Delta_n \le 1 \le b_d \sqrt{M^2\beta^2/n} = M b_d \delta. \tag{3.2}$$

Hence, (3.1) holds for $n \le 100 \le M^2\beta^2$. In particular, (3.1) holds for $n = 1$.

Now assume that (3.1) holds for $1, \dots, n-1$. We shall prove that then (3.1) holds for n as well. In the proof we can assume that $n \ge M^2\beta^2 \ge 100$, cf. (3.2).

Let $\varepsilon > 0$ be a positive number (later we shall choose $\varepsilon = a\delta$ with a sufficiently large absolute constant $a > 0$). Applying the smoothing Lemma 2.1, the condition (1.2) and using $a_d \le b_d$, we have

$$\Delta_n \le \sup_{A \in \mathcal{A}} |\,\mathbf{E}\,\varphi(S_n) - \mathbf{E}\,\varphi(Y)| + b_d\varepsilon, \tag{3.3}$$

where a function $\varphi = \varphi_{\varepsilon,A}$ satisfies the conditions of Lemma 2.1. Fix $A \in \mathcal{A}$. By Lemma 2.3 we can choose $\varphi(x) = \psi(\rho(x)/\varepsilon)$, where $\rho(x) \equiv \rho_A(x)$ is the distance between x and A, and φ and ψ satisfy the conditions of Lemma 2.3.

To simplify the notation, we write

$$U = X/\sqrt{n}, \qquad U_k = X_k/\sqrt{n}, \qquad V = Y/\sqrt{n}, \qquad V_k = Y_k/\sqrt{n}.$$

Write

$$W_k = V_1 + \dots + V_{k-1} + U_{k+1} + \dots + U_n, \qquad Z_n = V_1 + \dots + V_n. \tag{3.4}$$

Due to the i.i.d. assumption, S_n is distributed as $W_1 + U$. Moreover, Z_n, $W_n + V$ and Y are i.i.d. standard normal random vectors. We have

$$\begin{aligned} |\,\mathbf{E}\,\varphi(S_n) - \mathbf{E}\,\varphi(Y)| &= |\,\mathbf{E}\,\varphi(W_1+U) - \mathbf{E}\,\varphi(W_n+V)| \\ &\le |\,\mathbf{E}\,\varphi(W_1+U) - \mathbf{E}\,\varphi(W_1+V)| + |\,\mathbf{E}\,\varphi(W_1+V) - \mathbf{E}\,\varphi(W_n+V)| \\ &= |\,\mathbf{E}\,\varphi(W_1+U) - \mathbf{E}\,\varphi(W_1+V)| + |\,\mathbf{E}\,\varphi(W_2+U) - \mathbf{E}\,\varphi(W_n+V)| \end{aligned} \tag{3.5}$$

since by the i.i.d. assumption $\mathbf{E}\,\varphi(W_1+V) = \mathbf{E}\,\varphi(W_2+U)$. Repetitions of arguments used to obtain (3.5), lead to

$$|\,\mathbf{E}\,\varphi(S_n) - \mathbf{E}\,\varphi(Y)| \le \gamma_1 + \dots + \gamma_n, \tag{3.6}$$

where
$$\gamma_k = |\,\mathbf{E}\,\varphi(W_k + U) - \mathbf{E}\,\varphi(W_k + V)|.$$

Introduce
$$\theta^2 = n/(k-1), \qquad \alpha^2 = n/(n-k), \qquad Z = U_{k+1} + \cdots + U_n. \tag{3.7}$$

Then, for $2 \le k \le n$, we can write
$$\gamma_k = |\,\mathbf{E}\,\varphi(Y_1/\theta + U + Z) - \mathbf{E}\,\varphi(Y_1/\theta + V + Z)|. \tag{3.8}$$

Write $n_0 = [n/2]$, where $[z]$ stands for the integer part of z. Below we shall prove that, for any $m = 2, \dots, n_0$,
$$\gamma_1 \le b_d(c_1\delta + c_2 M\delta^2/\varepsilon), \tag{3.9}$$
$$\gamma_2 + \cdots + \gamma_m \le \frac{\beta b_d\sqrt{m-1}}{\varepsilon^2 n}(c_3 M\delta + c_4\varepsilon), \tag{3.10}$$
$$\gamma_{m+1} + \cdots + \gamma_{n_0} \le \frac{c_5\beta b_d M\delta}{\sqrt{m-1}} + c_6 b_d\delta, \tag{3.11}$$
$$\gamma_{n_0+1} + \cdots + \gamma_n \le c_7 b_d\delta \tag{3.12}$$
with some absolute constants $c_1, \dots, c_7$. These constants can be chosen to be
$$\begin{aligned} &c_1 = (2+2\sigma)\sqrt{100/99}, \quad c_2 = 2c_1, \quad c_3 = 2c_4, \quad c_4 = 16\sqrt{2}(\sigma + 2\sigma^2)/3, \\ &c_5 = \zeta(1+2\sigma)\sqrt{2}/3, \quad c_6 = c_5\sqrt{100/98}, \quad c_7 = c_5\sqrt{100/97} \end{aligned} \tag{3.13}$$
with $\sigma = 2/\sqrt{2\pi}$ and $\zeta = 2(1 + 4\,\mathrm{e}^{-3/2})/\sqrt{2\pi}$.

The bounds (3.9)–(3.12) imply the desired bound for Δ_n. Indeed, summing the inequalities (3.9)–(3.12), using (3.3) and setting $\varepsilon = a\delta$ with some $1 \le a \le M$ to be chosen later, we have (below $c_8 = c_8(c_1, \dots, c_7)$ and $c_9 = c_9(c_8)$ are sufficiently large absolute constants)
$$\begin{aligned} \Delta_n &\le b_d\delta\Big(a + c_8 + c_8\frac{\sqrt{m-1}}{a\beta} + c_8 M\Big(\frac{1}{a} + \frac{\sqrt{m-1}}{a^2\beta} + \frac{\beta}{\sqrt{m-1}}\Big)\Big) \\ &\le b_d\delta\Big(a + c_9 M\Big(\frac{1}{a} + \frac{\sqrt{m-1}}{a^2\beta} + \frac{\beta}{\sqrt{m-1}}\Big)\Big) \end{aligned}$$

using $\dfrac{M}{a} \ge 1$. Therefore, the inequality $\Delta_n \le M b_d\delta$ reduces to checking that
$$a + c_9 M\Big(\frac{1}{a} + \frac{\sqrt{m-1}}{a^2\beta} + \frac{\beta}{\sqrt{m-1}}\Big) \le M. \tag{3.14}$$

We choose $m-1$ such that $a^2\beta^2 \le m-1 \le 2a^2\beta^2$. Then (3.14) is implied by

$$a + c_{10}M/a \le M, \tag{3.15}$$

where $c_{10} \ge 1$ is an absolute constant. Let us take $a = 2c_{10}$ and $M = 10c_{10}$. Then (3.15) holds. Notice, that we have to check the condition $1 \le m \le n/2$. Since $m-1 \le 2a^2\beta^2$, it suffices to check that $3a^2\beta^2 \le n/4$ or, equivalently, that $12c_{10}^2\beta^2 \le n$. This inequality holds since we assume $n \ge M^2\beta^2 = 100c_{10}^2\beta^2$. A little more careful choice of the parameters using numerical values (3.13) of constants shows that it suffices to take $M \le 100$, which proves the theorem.

It remains to prove (3.9)–(3.12). We shall prove (3.9)–(3.12) with

$$c_1 \le 8, \quad c_2 \le 16, \quad c_3 \le 192, \quad c_4 \le 48, \quad c_5 \le 20, \quad c_6 \le 10, \quad c_7 \le 20.$$

Better constants (3.13) can be obtained using refined versions of inequalities of Lemmas 2.4–2.5, applying $n \ge 100$ to estimate better θ and α, and using $E\tau = 1/2$ and $\mathbf{E}\,\tau^2 = 1/3$. In order to keep the notation as simple as possible, we omit related details.

Let us prove (3.9). A little later we prove that

$$\gamma_1 \le \frac{4\beta}{\varepsilon\sqrt{n}}(2\Delta_{n-1} + \alpha b_d \varepsilon), \qquad \alpha^2 = n/(n-1). \tag{3.16}$$

Using the induction assumption $\Delta_{n-1} \le Mb_d\beta/\sqrt{(n-1)} = Mb_d\delta\alpha$ and estimating $\alpha \le 2$, for $n \ge 2$, the inequality (3.16) yields (3.9).

Let us prove the inequality (3.16). Using the Taylor formula (2.20) with $s=0$, noting that $\varphi'(x) = \varphi'(x)\mathbb{I}\{A^\varepsilon \setminus A\}$ since φ is a constant function outside the set $A^\varepsilon \setminus A^{-\varepsilon}$, and estimating the derivative φ' by (2.11) of Lemma 2.3, we have

$$\gamma_1 = |\,\mathbf{E}\,\varphi'(W_1+\tau U)U - \mathbf{E}\,\varphi'(W_1+\tau V)V| \le 2\,\mathbf{E}\,I(U)/\varepsilon + 2\,\mathbf{E}\,I(V)/\varepsilon \tag{3.17}$$

with

$$I(U) = |U|\mathbb{I}\{W_1 + \tau U \in (A^\varepsilon \setminus A\}, \qquad I(V) = |V|\mathbb{I}\{W_1 + \tau V \in (A^\varepsilon \setminus A\}.$$

We shall prove that

$$\mathbf{E}\,I(U) \le \mathbf{E}\,|U|(2\Delta_{n-1} + \alpha b_d\varepsilon) \qquad \mathbf{E}\,I(V) \le \mathbf{E}\,|V|(2\Delta_{n-1} + \alpha b_d\varepsilon). \tag{3.18}$$

By (2.13) of Lemma 2.4, we have $\mathbf{E}\,|V|$, $\mathbf{E}\,|U| \le \beta/\sqrt{n}$, and (3.17)–(3.18) imply (3.16).

Let us prove (3.18). We shall prove only the bound for $\mathbf{E}\, I(U)$ since a proof of the bound for $\mathbf{E}\, I(V)$ is similar. Conditioning on τ and U, we have

$$\begin{aligned}\mathbf{E}\, I(U) &= \mathbf{E}\, |U| \mathbb{P}\big\{\, W_1 \in A^\varepsilon \setminus A - \tau U \,\big|\, \tau, U \big\} \le \mathbf{E}\, |U| \sup_{z\in\mathbb{R}^d} \mathbb{P}\big\{\, W_1 \in A^\varepsilon \setminus A + z \,\big\} \\ &\le \mathbf{E}\, |U| \sup_{A\in\mathcal{A}} \mathbb{P}\big\{\, W_1 \in A^\varepsilon \setminus A \,\big\} = \mathbf{E}\, |U| \sup_{A\in\mathcal{A}} \mathbb{P}\big\{\, S_{n-1}/\alpha \in A^\varepsilon \setminus A \,\big\} \\ &= \mathbf{E}\, |U| \sup_{A\in\mathcal{A}} \mathbb{P}\big\{\, S_{n-1} \in A^{\alpha\varepsilon} \setminus A \,\big\}. \end{aligned} \tag{3.19}$$

We estimate $\mathbb{P}\big\{\, S_{n-1} \in A^{\alpha\varepsilon} \setminus A \,\big\} = \mathbb{P}\big\{\, S_{n-1} \in A^{\alpha\varepsilon} \,\big\} - \mathbb{P}\big\{\, S_{n-1} \in A \,\big\}$ applying the induction assumption. Using in addition the condition (1.2) to estimate $\mathbb{P}\big\{\, Y \in A^{\alpha\varepsilon} \setminus A \,\big\}$, we get

$$\mathbb{P}\big\{\, S_{n-1} \in A^{\alpha\varepsilon} \setminus A \,\big\} \le 2\Delta_{n-1} + \mathbb{P}\big\{\, Y \in A^{\alpha\varepsilon} \setminus A \,\big\} \le 2\Delta_{n-1} + \alpha b_d \varepsilon. \tag{3.20}$$

Combining (3.19) and (3.20), we obtain (3.18), which concludes the proof of (3.16) and of (3.9).

Let us prove (3.10). We shall show that (now $\alpha^2 = n/(n-k)$)

$$\gamma_k \le \frac{24\beta}{\varepsilon^2 n\sqrt{k-1}} \big(2\Delta_{n-k} + b_d \alpha\varepsilon \big), \qquad \text{for } k = 2, \dots, n-1. \tag{3.21}$$

This inequality implies (3.10). Indeed, using the induction assumption we can estimate $\Delta_{n-k} \le M b_d \alpha\delta$. Furthermore, $\alpha \le 2$ since $1 \le k \le n_0 = [n/2]$, and

$$\sum_{k=2}^{m} \frac{1}{\sqrt{k-1}} \le 1 + \int_1^m \frac{dt}{\sqrt{t}} = 2\sqrt{m} - 1 \le 2\sqrt{m-1}, \qquad \text{for } m \ge 2.$$

We note that the Lipschitz condition (2.11) implies

$$|\varphi'(x) - \varphi'(y)| \le \frac{8|x-y|}{\varepsilon^2} \Big(\mathbb{I}\{x \in A^\varepsilon \setminus A\} + \mathbb{I}\{y \in A^\varepsilon \setminus A\} \Big). \tag{3.22}$$

Indeed, if at least one of x and y belong to $A^\varepsilon \setminus A$, then $\mathbb{I}\{x \in A^\varepsilon \setminus A\} + \mathbb{I}\{x \in A^\varepsilon \setminus A\} \ge 1$, and (3.22) becomes (2.11). If neither x nor y belongs to $A^\varepsilon \setminus A$, then $\varphi'(x) = \varphi'(y) = 0$ since φ is a constant function outside the set $A^\varepsilon \setminus A$, and (3.22) is just the identity $0 = 0$.

Let us return to the proof of (3.21). Using the notation (3.7)–(3.8), we can write

$$\gamma_k = |I - J| \qquad I \overset{\text{def}}{=} \mathbf{E}\, \varphi(Y_1/\theta + U + Z), \quad J \overset{\text{def}}{=} \mathbf{E}\, \varphi(Y_1/\theta + V + Z).$$

Let us show that

$$I = I_0 + I_1, \qquad J = I_0 + J_1 \tag{3.23}$$

with $I_0 = \mathbf{E}\,\varphi(Y_1/\theta + Z)$ and

$$I_1 = \theta\,\mathbf{E}\,\tau\big\langle Y_1, U\big\rangle\varphi'(Y_1/\theta + \tau_1\tau U + Z)U,$$
$$J_1 = \theta\,\mathbf{E}\,\tau\big\langle Y_1, V\big\rangle\varphi'(Y_1/\theta + \tau_1\tau V + Z)V, \tag{3.24}$$

where τ and τ_1 are independent random variables uniformly distributed in the interval $[0, 1]$. We have

$$I = \mathbf{E}\int\limits_{\mathbb{R}^d} \varphi(x/\theta + U + Z)p(x)\,dx, \tag{3.25}$$

where p is standard normal density (2.16). Changing the variable x as $x/\theta + U = y$, we can rewrite (3.25) as

$$I = \mathbf{E}\int\limits_{\mathbb{R}^d} \varphi(y + Z)p(\theta y - \theta U)\theta^d\,dy. \tag{3.26}$$

Given U, expanding the density

$$p(\theta y - \theta U) = p(\theta y) - \theta\,\mathbf{E}\,p'(\theta y - \tau\theta U)U = p(\theta y) + \theta\,\mathbf{E}\,\big\langle\theta y - \tau\theta U, U\big\rangle p(\theta y - \theta\tau U)$$

(cf. (2.19)) and changing the variable y in integrals as $\theta y = x$ and $\theta y - \tau\theta U = x$, the relation (3.25) yields

$$I = I_0 + I_1, \quad I_1 \overset{\text{def}}{=} \theta\,\mathbf{E}\,\big\langle Y_1, U\big\rangle\varphi(Y_1/\theta + \tau U + Z). \tag{3.27}$$

Expanding the function φ by (2.20) with $s = 0$ and using $\mathbf{E}\,U = 0$, we have

$$I_1 = \theta\,\mathbf{E}\,\tau\big\langle Y_1, U\big\rangle\varphi'(Y_1/\theta + \tau_1\tau U + Z)U. \tag{3.28}$$

The relations (3.25)–(3.28) yield (3.23) for I. The proof for J is similar, we have just to replace everywhere U by V.

Write

$$I_2 = \theta\,\mathbf{E}\,\tau\big\langle Y_1, U\big\rangle\varphi'(Y_1/\theta + Z)U, \qquad J_2 = \theta\,\mathbf{E}\,\tau\big\langle Y_1, V\big\rangle\varphi'(Y_1/\theta + Z)V$$

and notice that $I_2 = J_2$ since the covariances of U and V are equal. Therefore (3.23) implies

$$\gamma_k = |I - J| = |I_1 - J_1| \le |I_1 - I_2| + |J_1 - J_2|. \tag{3.29}$$

Applying to the function φ' the Lipschitz condition (3.22), writing for a while

$$\xi = \frac{8\theta}{\varepsilon^2}\,\tau_1\tau^2|U|^2\,|\big\langle Y_1, U\big\rangle|$$

and $\mathbb{P}_*$ for the conditional probability given all random variables except Z, we have

$$|I_1 - I_2| \le \mathbf{E}\,\xi\Big(\mathbb{P}_*\{Y_1/\theta + \tau_1\tau U + Z \in A^\varepsilon \setminus A\} + \mathbb{P}_*\{Y_1/\theta + Z \in A^\varepsilon \setminus A\}\Big)$$
$$\le 2\,\mathbf{E}\,\xi \sup_{x\in\mathbb{R}^d} \mathbb{P}\{x + Z \in A^\varepsilon \setminus A\}. \tag{3.30}$$

Note that Z is a normalized sum of $n-k$ independent copies of X. Therefore, using the induction assumption and arguing similarly to (3.19)–(3.3), we get

$$\mathbb{P}\{x+Z\in A^{\varepsilon}\setminus A\}\le 2\Delta_{n-k}+b_d\alpha\varepsilon. \tag{3.31}$$

In order to estimate $\mathbf{E}\,\xi$ we notice that $\tau,\tau_1\le 1$ and $\mathbf{E}\,|U|^2\,|\langle Y_1,U\rangle|\le\beta n^{-3/2}$ (cf. (2.14) of Lemma 2.4). We get

$$\mathbf{E}\,\xi\le 8\theta\beta/(\varepsilon^2 n^{3/2}). \tag{3.32}$$

Collecting the bounds (3.30)–(3.32), we obtain

$$|I_1-I_2|\le 8\theta\beta(2\Delta_{n-k}+b_d\alpha\varepsilon)/(\varepsilon^2 n^{3/2}). \tag{3.33}$$

Similarly to the proof of (3.33) we have

$$|J_1-J_2|\le 16\theta\beta(2\Delta_{n-k}+b_d\alpha\varepsilon)/(\varepsilon^2 n^{3/2}). \tag{3.34}$$

Now (3.29) combined with (3.33) and (3.34) yields (3.21) since $\theta=\sqrt{n/(k-1)}$.

Let us prove (3.11) and (3.12). Below we show that

$$\gamma_k\le 5\beta\Delta_{n-k}(k-1)^{-3/2}+5\beta(n-k)^{-3/2},\qquad \text{for } k=2,\dots,n. \tag{3.35}$$

$$\gamma_k\le 5\beta(k-1)^{-3/2}\qquad \text{for } k=2,\dots,n. \tag{3.36}$$

These bounds imply (3.11) and (3.12). To prove (3.11) we use (3.35) and the induction assumption $\Delta_{n-k}\le 2b_dM\delta$, for $k\le n_0$. Using

$$\sum_{k=m+1}^{\infty}\frac{1}{(k-1)^{3/2}}\le\int_{m-1}^{\infty}\frac{dt}{t^{3/2}}=\frac{2}{(m-1)^{1/2}},$$

estimating $n-k\ge n/2$, for $k\le n_0$, we have

$$\gamma_{m+1}+\dots+\gamma_{n_0}\le\frac{20b_d\beta M\delta}{\sqrt{m-1}}+\frac{10\beta}{\sqrt{n}},$$

which proves (3.11) since $b_d\ge 1$. To prove (3.12), we use (3.36). Summing the bound (3.36) over $k>n_0=[n/2]$ and using

$$\sum_{k=n_0+1}^{\infty}\frac{1}{(k-1)^{3/2}}\le\int_{n_0-1}^{\infty}\frac{dt}{t^{3/2}}=\frac{2}{\sqrt{n_0-1}}\le\frac{4}{\sqrt{n}},$$

we obtain (3.12).

Let us prove (3.35) and (3.36). We have $\varphi(x) = \psi(\rho_A(x)/\varepsilon)$ and $\int_{\mathbb{R}} |\psi'(t)|\, dt = 1$ (cf. Lemma 2.3). Integrating by parts and using the property iii) of the class $\mathcal{A}$, we obtain

$$\begin{aligned}\gamma_k &= \Big| \int_{\mathbb{R}} \psi(t/\varepsilon) d(\mathbb{P}\{\rho_A(W_k+U) \le t\} - \mathbb{P}\{\rho_A(W_k+V) \le t\}) \Big| \\ &\le \int_{\mathbb{R}} |\psi'(t/\varepsilon)|\, |\mathbb{P}\{\rho_A(W_k+U) \le t\} - \mathbb{P}\{\rho_A(W_k+V) \le t\})| dt/\varepsilon \\ &\le \sup_{B\in\mathcal{A}} \gamma_k(B)\end{aligned}$$

with

$$\gamma_k(B) = |P\{W_k + U \in B\} - \mathbb{P}\{W_k + V \in B\})|.$$

Hence, it suffices to prove that instead of γ_k the quantity $\gamma_k(B)$ is bounded as γ_k in (3.35) and (3.36). Using the notation (3.7)–(3.8), we have

$$\gamma_k(B) = |\mathbb{P}\{Y/\theta + U + Z \in B\} - \mathbb{P}\{Y/\theta + V + Z \in B\}|.$$

Changing the variable y as $y/\theta + U = u$, we get

$$\gamma_k(B) = \Big| \mathbf{E} \int_{\mathbb{R}^d} \mathbb{I}\{y/\theta + U + Z \in B\} p(y) dy - \mathbf{E} \int_{\mathbb{R}^d} \mathbb{I}\{\theta y + V + Z \in B\} p(y) dy \Big| = |I_1 - J_1|,$$

where

$$I_1 = \mathbf{E} \int_{\mathbb{R}^d} \mathbb{I}\{u + Z \in B\} p(\theta u - \theta U)\theta^d\, du,$$

and J_1 is defined as I_1 just replacing U by V.

Applying the Taylor expansion (2.20) with $s = 2$ to the density p with $h = \theta U$ and $h = \theta V$, using that U and V have the same means and covariances, changing the variables $\theta u - \tau\theta U = y$ and $\theta u - \tau\theta V = y$, we obtain

$$\gamma_k(B) \le |I_0| + |J_0|$$

with

$$I_0 = \frac{\theta^3}{2}\, \mathbf{E}\, (1-\tau)^2 \int_{\mathbb{R}^d} \mathbb{I}\{y/\theta + \tau U + Z \in B\}\, p'''(y) U^3\, dy \tag{3.37}$$

and J_0 defined as I_0 just replacing U by V.

Let us prove (3.36). Using (3.37) and (2.18) of Lemma 2.4, we have

$$|I_0| \le \theta^3\, \mathbf{E} \int_{\mathbb{R}^d} |p'''(y) U^3|\, dy \le \frac{2\beta\theta^3}{n^{3/2}} = \frac{2\beta}{(k-1)^{3/2}}.$$

Similarly $|J_0| \le 3\beta/(k-1)^{3/2}$. These bounds and $\gamma_k(B) \le |I_0| + |J_0|$ prove (3.36).

Let us return to the proof of (3.35). Write $\alpha^2 = n/(n-k)$. The sum Z is a normalized sum of $n-k$ independent copies of X. Conditioning in (3.37) on all random variables except Z and using the induction assumption, we get

$$|I_0| \le \Delta_{n-k} I_2 + |I_3|, \tag{3.38}$$

where

$$I_2 = \frac{\theta^3}{6}\,\mathbf{E}\int_{\mathbb{R}^d} |p'''(y)U^3|\,dy, \qquad I_3 = \frac{\theta^3}{2}\,\mathbf{E}\,(1-\tau)^2 I_4,$$

$$I_4 \overset{\text{def}}{=} \int_{\mathbb{R}^d} \mathbb{I}\{y/\theta + \tau U + Y_1/\alpha \in B\}\, p'''(y)U^3\,dy.$$

We can represent I_4 as

$$I_4 = \theta^3 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{I}\{x + \tau U + z \in B\}\, p'''(\theta x)U^3 p(\alpha z)\,\theta^d\,dx\alpha^d dz.$$

To estimate I_4, we integrate by parts. Changing the variable x as $x + z = y$ and writing $\partial_h[z]$ for the partial derivative in the direction h with respect to the variable z, we have

$$\begin{aligned} I_4 &= -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{I}\{y + \tau U \in B\}\,\big(\partial_U^3[z]p(\theta y - \theta z)\big)p(\alpha z)\theta^d\,dy\alpha^d dz \\ &= \alpha^3 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{I}\{y + \tau U \in B\}\,p(\theta y - \theta z)p'''(\alpha z)U^3\theta^d\,dy\alpha^d dz. \end{aligned}$$

Therefore

$$|I_4| \le \alpha^3 \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} p(\theta y - \theta z)\,|p'''(\alpha z)U^3|\,\theta^d dy\,\alpha^d dz = \alpha^3 \int_{\mathbb{R}^d} |p'''(z)U^3|\,dz. \tag{3.39}$$

Estimating the last integral in (3.39) by (2.18) of Lemma 2.5 and collecting the bounds in (3.38), we get

$$|I_0| \le 2\beta\big(\Delta_{n-k}\theta^3 n^{-3/2} + \alpha^3 n^{-3/2}\big) = 2\beta\big(\Delta_{n-k}(k-1)^{-3/2} + (n-k)^{-3/2}\big). \tag{3.40}$$

A similar proof shows that $|J_0|$ is bounded as I_0 in (3.40) but with the factor 3 instead of 2. Therefore the inequality $\gamma_k(B) \le |I_0| + |J_0|$ yields (3.35). $\square$

References

Ball, K., *The reverse isoperimetric problem for Gaussian measure*, Discrete Comput. Geom. **10** (1993), no. 4, 411–420.

Bhattacharya, R.N; Rao, R.R., *Normal Approximation and Asymptotic Expansions*, Krieger, Melbourne, 1986.

Bentkus, V., *Dependence of the Berry–Esseen bound on the dimension*, Lithuanian Math. J. **26** (1986a), no. 2, 205–210.

Bentkus, V., *Large deviations in Banach spaces*, Probab. Theory Appl. **31** (1986b), no. 4, 710–716.

Götze, F., *On the rate of convergence in the multivariate CLT*, Ann. Probab. **19** (1991), no. 2, 724–739.

Nagaev, S.V., *An estimate of the remainder term in the multidimensional CLT*, Proc. of the third Japan–USSR Symposium on Probability theory, LNM, Springer, 1976, pp. 419–438.

Paulauskas, V.; Račkauskas, A., *Approximation theory in the CLT. Exact results in Banach spaces*, Kluwer, Dordrecht, 1989.

Sazonov, V., *Normal approximations some recent advances.*, Springer, Berlin–New-York, 1981.

Senatov, V., *Normal approximations: new results, methods and problems*, VSP, Utrecht, 1998.

Senatov, V., *An estimate of the Levy-Prokhorov metric*, Theory Probab. Appl. **29** (1984), no. 1, 109–114.

Senatov, V., *Several estimates of the rate of convergence in the multidimensional CLT*, Dokl. Akad. Nauk SSSR **254** (1980), no. 4, 809–812.

Vidmantas Bentkus
Institute of mathematics and informatics
Akademijos 4
232600 Vilnius
Lithuania
E-mail address: `bentkus@takas.lt`