Journal of Mathematical Sciences, Vol. 167, No. 4, 2010

ON A THEOREM OF V. N. SUDAKOV ON TYPICAL DISTRIBUTIONS

S. G. Bobkov* UDC 519

The rate of approximation with respect to the Kantorovich–Rubinstein distance is considered in a theorem of V. N. Sudakov on typical distributions. Bibliography: 18 titles.

Let $X = (X_1, \dots, X_n)$ be a random vector in $\mathbf{R}^n$, $n \geq 2$. We consider linear functionals

$$S_\theta = \theta_1 X_1 + \cdots + \theta_n X_n,$$

whose coefficients satisfy the condition $\theta_1^2 + \cdots + \theta_n^2 = 1$, and the corresponding distribution functions

$$F_\theta(x) = \mathbf{P}\{S_\theta \leq x\}, \quad x \in \mathbf{R}.$$

The above coefficients can be coordinates of arbitrary vectors $\theta = (\theta_1, \dots, \theta_n)$ on the unit sphere S^{n-1}.

In 1978, V. N. Sudakov made the following remarkable observation (see [15]): If the second order absolute moments of the random variables S_θ are bounded uniformly in θ and the dimension n is large, then the F_θ concentrate at a "typical" distribution F. One may take as F the average

$$F(x) = \mathbf{E}_\theta\, F_\theta(x) = \int\limits_{S^{n-1}} F_\theta(x)\, d\sigma_{n-1}(\theta)$$

with respect to the uniform distribution σ_{n-1} on the unit sphere and estimate the closeness of distributions on the real line using, for example, the average metric (Kantorovich–Rubinstein metric),

$$\kappa(F_\theta, F) = \int\limits_{-\infty}^{+\infty} |F_\theta(x) - F(x)|\, dx.$$

In our notation, the Sudakov theorem can be stated as follows.

Theorem 1. *For any $\delta > 0$ there exists a natural number n_δ having the following property. Assume that a random vector X in the Euclidean space $\mathbf{R}^n$ of dimension $n \geq n_\delta$ satisfies the condition*

$$\mathbf{E}\,|S_\theta|^2 \leq \lambda^2 \quad \text{for any} \quad \theta \in S^{n-1}$$

with some $\lambda \geq 0$. Then there exists a measurable subset $\Theta \subset S^{n-1}$ of measure $\sigma_{n-1}(\Theta) \geq 1 - \delta$ such that

$$\kappa(F_\theta, F) \leq \lambda\delta$$

for any element of this set.

The typical distribution F can be characterized as the distribution of the random variable $\xi\|X\|$, where $\|X\| = (X_1^2 + \cdots + X_n^2)^{1/2}$ is the Euclidean length of the initial random vector, and the variable ξ is independent of X and equidistributed with the first coordinate of a unit vector on the sphere with respect to the measure σ_{n-1}. It is well known that the distribution of a random variable $\xi\sqrt{n}$ for large n is close to the standard normal one. Thus, the variables $\xi\|X\|$ and $Z\|X\|/\sqrt{n}$ are almost identically distributed, where Z is a standard normal random variable independent of X. Hence (as was emphasized in [15]), one may take as a typical distribution a certain mixture of normal laws on the line with zero mean.

If in this case the random variable $\|X\|/\sqrt{n}$ is almost constant (i.e., the sequence X_i^2 satisfies a kind of the weak law of large numbers), then it is possible to conclude that "almost" all linear functionals S_θ are distributed

*School of Mathematics, University of Minnesota, Minneapolis, USA, e-mail: bobko001@umn.edu.

Translated from *Zapiski Nauchnykh Seminarov POMI*, Vol. 368, 2009, pp. 59–74. Original article submitted September 20, 2009.

almost normally. One may treat such a conclusion as a variant of the central limit theorem without any essential assumptions on the character of dependence of the initial random variables.

At present, many papers are devoted to the problem of typical distributions, in particular, to refinements and generalizations of Theorem 1 as well as to its variations (see, for example, [1–3, 6, 8, 12–14, 16–17]). Some of these papers are related to complicated problems of convex geometry. We do not give here a survey of the results obtained; let us only make the following remark.

Similarly to the case of problems related to the rate of convergence in the central limit theorem, one of the most interesting topics is the problem on the concentration rate of F_θ around the typical distribution F with respect to various probability metrics. For example, one can characterize the concentration rate by the variable

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) = \int\limits_{S^{n-1}} \kappa(F_\theta, F)\, d\sigma_{n-1}(\theta)$$

or by similar average distances for Lévy or Kantorovich metrics. The problem concerning the smallest distance, $\inf_\theta \kappa(F_\theta, F)$, is of interest as well.

In this note, we prove the following statement.

Theorem 2. *Assume that a random vector X in $\mathbf{R}^n$ satisfies the following condition:*

$$\mathbf{E}\, e^{|S_\theta|/\lambda} \le 2 \quad \textit{for any} \quad \theta \in S^{n-1}$$

with some $\lambda > 0$. Then

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) \le C\lambda\, \frac{\log n}{\sqrt{n}},$$

where C is an absolute constant.

Applying the Lipschitz continuity of the function $\theta \to \kappa(F_\theta, F)$, we can sharpen the statement of Theorem 2 in terms of large deviations.

Corollary 1. *Under the conditions of Theorem* 2*,*

$$\sigma_{n-1}\Big\{\kappa(F_\theta, F) \ge C\lambda\, \frac{\log n}{\sqrt{n}} + \lambda t\Big\} \le e^{-(n-1)\, t^2/2}$$

for all $t > 0$.

As we show below, weaker moment assumptions on linear functionals S_θ also lead to definite (weaker) asymptotic estimates for the variables $\mathbf{E}_\theta\, \kappa(F_\theta, F)$.

On the other hand, it is possible to omit the logarithmic term in the statement of Theorem 2 under stronger assumptions. Indeed, assume that X has a logarithmically concave density and all the linear functionals S_θ satisfy the conditions

$$\mathbf{E} S_\theta = 0 \quad \text{and} \quad \mathbf{E}|S_\theta|^2 = 1,$$

which is equivalent to the following requirements: $\mathbf{E}X_i = 0$, $\mathbf{E}X_i^2 = 1$, and $\mathbf{E}X_iX_j = 0$ $(i \ne j)$. Note that in this case, the condition $\mathbf{E}e^{|S_\theta|/C} \le 2$ is satisfied automatically due to a theorem of C. Borell on logarithmibally concave measures (see [5, Lemma 3.1]). In addition, it is shown in [2, Proposition 3.1] that functions of the form

$$u_x(\theta) = F_\theta(x) - F(x), \quad \theta \in S^{n-1},$$

are Lipschitz continuous on the sphere, and their Lipschitz seminorms (in the sense of the distance induced from $\mathbf{R}^n$) satisfy the estimate

$$\|u_x\|_{\mathrm{Lip}} \le Ce^{-c|x|}, \quad x \in \mathbf{R}.$$

The property of concentration of the uniform distribution on the sphere implies that

$$\int |u_x(\theta)|\, d\sigma_{n-1}(\theta) \le \frac{\|u_x\|_{\mathrm{Lip}}}{\sqrt{n}} \le \frac{Ce^{-c|x|}}{\sqrt{n}}.$$

Integrating this inequality in x, we get the following inequality, which correlates with known results for independent random variables:

$$\int_{S^{n-1}} \kappa(F_\theta, F)\, d\sigma_{n-1}(\theta) \le \frac{C}{\sqrt{n}}.$$

In this case, similarly to Corollary 1,

$$\sigma_{n-1}\left\{\sqrt{n}\ \kappa(F_\theta, F) \ge t\right\} \le 2e^{-ct^2/2}$$

for all $t > 0$ (we denote by C and c various absolute positive constants which may differ in different positions).

The above reasoning is essentially based on the property of logarithmic concavity of the distribution of the initial random vector. To include the general case under moment assumptions (possibly, with weaker estimates for the concentration rate), we need the following known lemma (one can prove this statement applying the isoperimetric theorem and logarithmic Sobolev inequality on the sphere).

Lemma 1. *If $u : \mathbf{R}^n \to \mathbf{R}$ is a continuously differentiable function with zero mean with respect to the measure σ_{n-1}, then*

$$\int |u|\, d\sigma_{n-1} \le \frac{\pi}{\sqrt{2n}} \int |\nabla u|\, d\sigma_{n-1}.$$

In addition, if $\|u\|_{\mathrm{Lip}} \le 1$, then

$$\sigma_{n-1}\{u \ge t\} \le e^{-(n-1)\,t^2/2}$$

for all $t > 0$,

At the end of this note, we give several comments on the asymptotic behavior of the best constant in the first inequality (with respect to dimension) and on the concentration property on the sphere.

Let us explain the application of Lemma 1. To simplify notation, we write $\mathbf{E}_\theta\, u(\theta)$ instead of the integral $\int u\, d\sigma_{n-1}$.

We relate to a continuously differentiable function $f : \mathbf{R} \to \mathbf{R}$ with bounded derivative f' the function

$$u(\theta) = \int_{-\infty}^{+\infty} f\, dF_\theta - \int_{-\infty}^{+\infty} f\, dF = \mathbf{E} f(S_\theta) - \mathbf{E} f(S),$$

where S is an imaginary random variable with distribution function F. Then $\mathbf{E}_\theta\, u = 0$, and the definition of the random variables S_θ implies that

$$\langle \nabla u(\theta), \theta' \rangle = \mathbf{E}\, S_{\theta'}\, f'(S_\theta)$$

for any unit vectors θ and θ'. If a moment assumption of the form

$$(\mathbf{E}\,|S_\theta|^p)^{1/p} \le M_p \quad \text{for any} \quad \theta \in S^{n-1}$$

with a fixed $p > 1$ and a finite constant M_p is satisfied, then we deduce from the Hölder inequality that

$$|\langle \nabla u(\theta), \theta' \rangle| \ \le\ M_p\, (\mathbf{E}\,|f'(S_\theta)|^q)^{1/q}, \quad q = \frac{p}{p-1}.$$

Since θ' is arbitrary, we conclude that

$$|\nabla u(\theta)| \ \le\ M_p\, (\mathbf{E}\,|f'(S_\theta)|^q)^{1/q}.$$

We integrate this inequality over the sphere and apply Lemma 1 (and the Markov inequality) to get the estimate

$$\mathbf{E}_\theta \left| \int_{-\infty}^{+\infty} f\, d(F_\theta - F) \right| \le \frac{3M_p}{\sqrt{n}} \left(\int_{-\infty}^{+\infty} |f'|^q\, dF \right)^{1/q}.$$

Now we integrate by parts to rewrite the first integral in terms of the derivative f'; as a result, we see that

$$\mathbf{E}_\theta \left| \int_{-\infty}^{+\infty} f'(x)\, (F_\theta(x) - F(x))\, dx \right| \le \frac{3M_p}{\sqrt{n}} \left(\int_{-\infty}^{+\infty} |f'|^q\, dF \right)^{1/q}.$$

In this formula, f' is an arbitrary function from the class of bounded continuous functions on the real line. Approximating any bounded (Lebesgue) measurable function by functions of the above-mentioned class, we get the following statement (which is an auxiliary step in the proof of Theorem 2).

Lemma 2. *Assume that a random vector X in $\mathbf{R}^n$ satisfies the following condition:*

$$(\mathbf{E}\,|S_\theta|^p)^{1/p} \leq M_p \quad \textit{for any} \quad \theta \in S^{n-1}$$

with a constant M_p for a fixed $p > 1$. Then, for any bounded measurable function f on the real line,

$$\mathbf{E}_\theta \bigg| \int_{-\infty}^{+\infty} f(x)\,(F_\theta(x) - F(x))\,dx \bigg| \leq \frac{3M_p}{\sqrt{n}} \bigg(\int_{-\infty}^{+\infty} |f(x)|^q \, dF(x) \bigg)^{1/q},$$

where $q = \frac{p}{p-1}$ is the adjoint degree.

In particular, we get the estimate

$$\mathbf{E}_\theta \bigg| \int_a^b (F_\theta(x) - F(x))\,dx \bigg| \leq \frac{3M_p}{\sqrt{n}}\,(F(b) - F(a))^{1/q}$$

for any $a < b$.

The expression on the left in the above inequality looks similar to the average Kantorovich–Rubinstein distance $\mathbf{E}_\theta\,\kappa(F_\theta, F)$; in fact, this expression would become the distance if we could insert the sign of absolute value under the integral sign. The following observation is useful for reaching this goal (but with some losses).

Lemma 3. *Let F and G be arbitrary distribution functions. Then*

$$\int_a^b |F(x) - G(x)|\,dx \leq \sum_{k=1}^N \bigg| \int_{a_{k-1}}^{a_k} (F(x) - G(x))\,dx \bigg| + \frac{2(b-a)}{N}$$

for any $a < b$ and a natural number N, where $a_k = a + (b-a)\frac{k}{N}$.

Proof. Denote by I the family of indices $k = 1, \dots, N$ such that the function $\varphi(x) = F(x) - G(x)$ does not change sign in the kth interval $\Delta_k = (a_{k-1}, a_k)$. The remaining indices form a set $J \subset \{1, \dots, N\}$. If $k \in I$, then

$$\int_{\Delta_k} |F(x) - G(x)|\,dx = \bigg| \int_{\Delta_k} (F(x) - G(x))\,dx \bigg|.$$

If $k \in J$, then it is clear that

$$\sup_{x \in \Delta_k} |\varphi(x)| \leq \sup_{x,y \in \Delta_k} (\varphi(x) - \varphi(y)) \leq F(\Delta_k) + G(\Delta_k),$$

where F and G are considered as probability measures in the last step. In this case,

$$\int_{\Delta_k} |F(x) - G(x)|\,dx \leq (F(\Delta_k) + G(\Delta_k))|\Delta_k|, \quad |\Delta_k| = \frac{b-a}{N}.$$

Combining both estimates, we conclude that the integral $\int_a^b |F(x) - G(x)|\,dx$ does not exceed

$$\sum_{k \in I} \bigg| \int_{\Delta_k} (F(x) - G(x))\,dx \bigg| + \sum_{k \in J} (F(\Delta_k) + G(\Delta_k))|\Delta_k|$$
$$\leq \sum_{k=1}^N \bigg| \int_{\Delta_k} (F(x) - G(x))\,dx \bigg| + \frac{b-a}{N} \sum_{k=1}^N (F(\Delta_k) + G(\Delta_k)).$$

The lemma is proved.

We can apply Lemma 3 under the conditions of Lemma 2 in estimation of the the Kantorovich–Rubinstein distance. Without loss of generality, we assume that $M_p = 1$; this is possible since $\mathbf{E}_\theta\, \kappa(F_\theta, F)$ is homogeneous in X. Applying Lemma 2 with the same intervals $\Delta_k = (a_{k-1}, a_k)$, we see that

$$\mathbf{E}_\theta \Big| \int\limits_{\Delta_k} (F_\theta(x) - F(x))\, dx \Big| \le \frac{3}{\sqrt{n}}\, F(\Delta_k)^{1/q};$$

hence,

$$\mathbf{E}_\theta \int\limits_a^b |F_\theta(x) - F(x)|\, dx \le \frac{2(b-a)}{N} + \frac{3}{\sqrt{n}} \sum_{k=1}^N F(\Delta_k)^{1/q}$$

by Lemma 3. By the Hölder inequality, the last sum does not exceed

$$N^{1/p} \Big(\sum_{k=1}^N F(\Delta_k) \Big)^{1/q} \le N^{1/p},$$

and we arriwe at the estimate

$$\mathbf{E}_\theta \int\limits_a^b |F_\theta(x) - F(x)|\, dx \le \frac{2(b-a)}{N} + \frac{3N^{1/p}}{\sqrt{n}}.$$

In particular, if $a = -b$, $b > 0$, then

$$\mathbf{E}_\theta \int\limits_{-b}^b |F_\theta(x) - F(x)|\, dx \le \frac{4b}{N} + \frac{3N^{1/p}}{\sqrt{n}}.$$

To extend the integral to the whole real line, we apply the assumption that $M_p = 1$; by the Chebyshev inequality, this assumption implies that $F_\theta\{x : |x| \ge t\} \le t^{-p}$ for $t > 0$. A similar estimate for F is obtained after averaging in θ. Hence,

$$\int\limits_{\{|x| \ge b\}} |F_\theta(x) - F(x)|\, dx \le \int\limits_{\{|x| \ge b\}} |x|^{-p}\, dx = \frac{2b^{1-p}}{p-1}.$$

Combining both estimates, we conclude that

$$\mathbf{E}_\theta \int\limits_{-\infty}^{+\infty} |F_\theta(x) - F(x)|\, dx \le \frac{2b^{1-p}}{p-1} + \frac{4b}{N} + \frac{3N^{1/p}}{\sqrt{n}}$$

for any real b and natural N. It is easily seen that the right-hand side attains its minimum for $b = (\frac{N}{2})^{1/p}$, and

$$\frac{2b^{1-p}}{p-1} + \frac{4b}{N} = \frac{4N^{(1-p)/p}}{2^{1/p}\,(p-1)} + \frac{4N^{(1-p)/p}}{2^{1/p}} = q\, 2^{1+1/q}\, N^{-1/q}$$

in this case.

Thus,

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) \le q\, 2^{1+1/q}\, N^{-1/q} + \frac{3N^{1/p}}{\sqrt{n}}.$$

Now we have to optimize the right-hand side of the obtained inequality over all natural n. Consider the following function of a real variable $x > 0$:

$$\psi(x) = \alpha x^{-1/q} + \beta x^{1/p}, \quad \alpha = q\, 2^{1+1/q}, \quad \beta = \frac{3}{\sqrt{n}}.$$

The function ψ attains its minimum at

$$x_0 = \frac{\alpha}{\beta}\,\frac{p}{q} = \frac{1}{3}\,2^{1+1/q}\,p\sqrt{n}.$$

Note that $x_0 > \frac{2}{3}\sqrt{n} > \frac{2}{3}$. Hence, if $N = [x_0] + 1$, then $x_0 < N \le 2x_0$. For this value,

$$\psi(N) \le \alpha\, x_0^{-1/q} + \beta\,(2x_0)^{1/p} = C_p\, n^{-1/(2q)},$$

where

$$C_p = 2^{\frac{1}{p}(1+\frac{1}{q})}\left[q\left(\frac{3}{p}\right)^{1/q} + 3\left(\frac{2p}{3}\right)^{1/p}\right].$$

To replace this expression by a simpler one, we apply the obvious estimates

$$\frac{1}{p}\left(1+\frac{1}{q}\right) = \frac{1}{p}\left(2-\frac{1}{p}\right) \le 1 \quad \text{and} \quad 3\left(\frac{2p}{3}\right)^{1/p} \le 3\,e^{2/(3e)} < 4.$$

It is easily seen that $(\frac{3}{p})^{1/q} < 2$. Hence, $C_p < 2\,(2q+4) < 12\,q$, and

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) \le 12\,q\,n^{-1/(2q)}.$$

As a result, we get the following statement.

Theorem 3. *Assume that a random vector X in $\mathbf{R}^n$ satisfies the following condition:*

$$(\mathbf{E}\,|S_\theta|^p)^{1/p} \le M_p \quad \textit{for any} \quad \theta \in S^{n-1}$$

with a constant M_p for a fixed $p > 1$. Then

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) \le 12\, M_p\, \frac{p}{p-1}\, n^{-\frac{p-1}{2p}}.$$

For example, if $p = 2$, then $\mathbf{E}_\theta\, \kappa(F_\theta, F) \le 24\, M_2\, n^{-1/4}$. As a corollary, we see that

$$\sigma_{n-1}\left\{\theta \in S^{n-1} : \kappa(F_\theta, F) \ge M_2\delta\right\} \le \frac{24\, n^{-1/4}}{\delta}$$

for any $\delta > 0$. The expression on the right does not exceed δ whenever $n \ge 24^4/\delta^8$.

Corollary 2. *One can take $n_\delta = [\,24^4\,\delta^{-8}] + 1$ in Theorem 1.*

It remains to make the last step in the proof of Theorem 2.

Proof of Theorem 2. We may set $\lambda = 1$. In this case, the elementary inequality $t^p \le p^p e^{-p}\, e^t$, which is valid for all $t \ge 0$ and $p > 1$, implies that

$$\mathbf{E}\,|S_\theta|^p \le p^p\, e^{-p}\, \mathbf{E}\, e^{|S_\theta|} \le 2\, p^p\, e^{-p},$$

so that $M_p < 2p/e$. By Theorem 3,

$$\mathbf{E}_\theta\, \kappa(F_\theta, F) \le 12\,\frac{2p}{e}\,\frac{p}{p-1}\, n^{\frac{1}{2p}}\, \frac{1}{\sqrt{n}}.$$

Take $p = 1 + \frac{1}{2}\log n$ $(n \ge 2)$. Then $n^{\frac{1}{2p}} < e$, and it is clear that the coefficient of $\frac{1}{\sqrt{n}}$ is bounded by $C\log n$ with an absolute constant C. Theorem 2 is proved.

Proof of Corollary 1. Similarly to the case of Theorem 2, we take $\lambda = 1$. It is enough to show that the function

$$u(\theta) = \kappa(F_\theta, F) = \int_{-\infty}^{+\infty} |F_\theta(x) - F(x)|\, dx$$

has a bounded Lipschitz seminorm on the unit sphere. By the Kantorovich–Rubinstein theorem (see, for example, [7, p. 330]), the following representation is valid:

$$u(\theta) = \sup \left[\int_{-\infty}^{+\infty} f \, dF_\theta - \int_{-\infty}^{+\infty} f \, dF \right] = \sup \left[\mathbf{E}\, f(S_\theta) - \mathbf{E}\, f(S) \right],$$

where S is a random variable with distribution F, and the supremum is taken over the set of all functions $f : \mathbf{R} \to \mathbf{R}$ with Lipschitz seminorm $\|f\|_{\mathrm{Lip}} \le 1$. It follows that

$$u(\theta) - u(\theta') \le \sup \left[\mathbf{E}\, f(S_\theta) - \mathbf{E}\, f(S_{\theta'}) \right]$$

for any pair θ, θ' of unit vectors. Note that

$$|\mathbf{E}\, f(S_\theta) - \mathbf{E}\, f(S_{\theta'})| \le \mathbf{E}\, |S_\theta - S_{\theta'}| = \mathbf{E}\, |\langle X, \theta - \theta' \rangle| \le M_1 \, \|\theta - \theta'\|,$$

where $M_1 = \sup_{\theta \in S^{n-1}} \mathbf{E}\, |S_\theta|$. Hence, $\|u\|_{\mathrm{Lip}} \le M_1$. By the condition of our theorem, $\mathbf{E}\, e^{|S_\theta|} \le 2$; hence, $\mathbf{E}\, |S_\theta| \le e^{-1} \, \mathbf{E}\, e^{|S_\theta|} < 1$. Thus, $\|u\|_{\mathrm{Lip}} < 1$.

It remains to apply the second inequality of Lemma 1 to the function $u - \mathbf{E}_\theta u$ and to use the estimate of $\mathbf{E}_\theta u$ according to Theorem 2.

Remarks. (1) If we assume in Theorem 2 that the linear functionals are sub-Gaussian uniformly in θ, i.e., $\mathbf{E}\, e^{S_\theta^2/\lambda^2} \le 2$, then the statement can be sharpened up to the estimate

$$\mathbf{E}_\theta \, \kappa(F_\theta, F) \le C\lambda \, \frac{\sqrt{\log n}}{\sqrt{n}}.$$

The proof is the same, only Theorem 3 must be applied with parameter p of order $\sqrt{\log n}$.

(2) It follows from Theorem 3 that the condition $(\mathbf{E}\, |S_\theta|^2)^{1/2} \le \lambda$ in Theorem 1 can be weakened up to the condition $(\mathbf{E}\, |S_\theta|^p)^{1/p} \le \lambda$ for any fixed $p > 1$. At the same time, it is not clear whether it is enough to assume that the first absolute moments are bounded, i.e., that $\mathbf{E}\, |S_\theta| \le \lambda$. The main argument opposing the statement of Theorem 1 in such a general case is the following fact: The family of all probability distributions on the real line with a bounded absolute moment fails to be compact in the Kantorovich–Rubinstein metric (while it remains compact in the topology of weak convergence). Thus, it is natural to consider the problem on the rate of concentration around a typical distribution in different standard metrics which are responsible for weak convergence. For example, the following statement holds.

Theorem 4. *Assume that a random vector X in $\mathbf{R}^n$ satisfies the condition $\mathbf{E}\, |S_\theta| \le \lambda$ for any $\theta \in S^{n-1}$ with a constant λ. Then*

$$\mathbf{E}_\theta \, L(F_\theta, F) \le C_\lambda \, \frac{\log n}{n^{1/4}},$$

where C_λ depends on λ.

In this theorem, $L(F_\theta, F)$ is the Lévy distance, which is defined as the smallest value $h \in [0, 1]$ such that

$$F(x - h) - h \le F_\theta(x) \le F(x + h) + h$$

for all $x \in \mathbf{R}$.

Proof. We apply the Zolotarev inequality (see [18]):

$$L(F_\theta, F) \le \frac{1}{\pi} \int_0^T \frac{|f_\theta(t) - f(t)|}{t} \, dt + 2e \, \frac{\log T}{T}, \quad T > 1.3,$$

which relates the Lévy distance between distributions F_θ and F to their characteristic functions

$$f_\theta(t) = \mathbf{E} e^{itS_\theta} \quad \text{and} \quad f(t) = \mathbf{E} e^{itS} = \mathbf{E}_\theta f_\theta(t).$$

Here S is a random variable distributed according to F (such a relation holds for any pair of probability distributions on the line).

For any fixed $t > 0$ the (complex-valued) function

$$u_t(\theta) = f_\theta(t) - f(t) = \mathbf{E}\,(e^{itS_\theta} - e^{itS}), \quad \theta \in \mathbf{R}^n,$$

is defined and continuously differentiable on the whole space, and $\langle \nabla u_t(\theta), \theta' \rangle = it\,\mathbf{E}\,S_{\theta'}\,e^{itS_\theta}$ for any unit vectors θ and θ'. Hence,

$$|\langle \nabla u_t(\theta), \theta' \rangle| \le t\,\mathbf{E}\,|S_{\theta'}| \le \lambda t.$$

Since θ' is arbitrary, $|\nabla u_t(\theta)| \le \lambda t$ on the unit sphere. The isoperimetric theorem on the sphere implies that

$$\mathbf{E}_\theta\,|f_\theta(t) - f(t)| = \mathbf{E}_\theta |u_t| \le \frac{2\|u_t\|_{\mathrm{Lip}}}{\sqrt{n}} \le \frac{2\lambda t}{\sqrt{n}}.$$

Hence, averaging the Zolotarev inequality in θ, we conclude that

$$\mathbf{E}_\theta\, L(F_\theta, F) \le \frac{2\lambda T}{\pi\sqrt{n}} + 2e\,\frac{\log T}{T}$$

for all $T > 1.3$. It remains to optimize the right-hand side in T (it is enough to take T of order $n^{1/4}$).

Remark (to Lemma 1)**.** It follows from the general theory of Sobolev type inequalities (see, for example, [10, 4]) that the best constant in the inequality

$$c_n\,\mathbf{E}_\theta\,|u - \mathbf{E}_\theta u| \le \mathbf{E}_\theta |\nabla u|$$

is attained in the asymptotic sense at indicator functions $u = 1_A$ of measurable subsets $A \subset S^{n-1}$. For such functions, the considered integro-differential inequality turns into an inequality of isoperimetric type,

$$\sigma_{n-1}^+(A) \ge 2c_n\,t(1-t), \quad t = \sigma_{n-1}(A),$$

where

$$\sigma_{n-1}^+(A) = \liminf_{\varepsilon \downarrow 0} \left[\frac{1}{\varepsilon}\,\sigma_{n-1}\left\{w \in S^{n-1} \setminus A : \quad \text{there exists} \quad v \in A \quad \text{such that} \quad |v - w| < \varepsilon\right\}\right]$$

denotes the perimeter of A is the sense of the measure σ_{n-1}. Minimizing in A, we get an equivalent relation

$$I(t) \ge 2c_n\,t(1-t), \quad 0 < t < 1,$$

in terms of the isoperimetric function $I(t) = \inf\{\sigma_{n-1}^+(A) : \sigma_{n-1}(A) = t\}$; hence,

$$c_n = \inf_{0<t<1} \frac{I(t)}{2t(1-t)}.$$

This formula remains valid in a rather abstract situation, namely, for arbitrary metric spaces.

In the case of a sphere, it follows from the Lévy–Schmidt isoperimetric theorem that, for any fixed $t \in (0,1)$, the perimeter $\sigma_{n-1}^+(A)$ in the class of sets with fixed measure $\sigma_{n-1}(A) = t$ is minimized at balls of the sphere. Hence, it is possible to get an explicit analytic description of the corresponding isoperimetric function $I(t)$. In addition, it is known that the function $t(1-t)/I(t)$ is concave (see [4, Theorem 1.9]), which immediately implies that $c_n = 2I(1/2)$. We note that $I(1/2)$ equals the perimeter of a half-sphere whose boundary is a sphere of dimension $n-2$ and that the $(n-1)$-dimensional area of the sphere S^{n-1} is given by the formula $s_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$. It follows that

$$c_n = \frac{2s_{n-2}}{s_{n-1}} = \frac{2}{\sqrt{\pi}}\,\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}.$$

In particulat, $c_2 = 2/\pi$. It follows from the Stirling formula that $c_n\sqrt{n} \to \sqrt{2/\pi}$ as $n \to \infty$. Hence, the best (dimensionless) constant in an inequality of the form

$$\mathbf{E}_\theta\, |u - \mathbf{E}_\theta u| \le \frac{C}{\sqrt{n}}\, \mathbf{E}_\theta |\nabla u|$$

is not less than $\sqrt{\frac{\pi}{2}}$. A routine analysis of the formula for c_n shows that the value of $c_n/\sqrt{n}$ is minimal for $n = 2$, which corresponds to the optimal value $C = \frac{\pi}{\sqrt{2}}$.

If we know, in addition, that $|\nabla u(\theta)| \le 1$, i.e., $\|u\|_{\mathrm{Lip}} \le 1$, then

$$\mathbf{E}_\theta\, |u - \mathbf{E}_\theta\, u| \le \frac{C}{\sqrt{n}}$$

by Lemma 1. In this case, however, the constant $C = \frac{\pi}{\sqrt{2}}$ is not optimal, and we can take $C = 1$. Indeed, by the Lévy–Schmidt theorem, the distribution of any Lipschitz continuous function u with respect to the measure σ_{n-1} is the image of the distribution of the function $u_1(\theta) = \theta_1$ with respect to the same measure under some nondecreasing mapping $\Psi : \mathbf{R} \to \mathbf{R}$ with Lipschitz seminorm $\|\Psi\|_{\mathrm{Lip}} \le 1$. It is easy to conclude that, for dispersions in the sence of the measure σ_{n-1},

$$\begin{aligned}\mathrm{Var}\,(u) &= \frac{1}{2}\iint |\Psi(u_1(\theta)) - \Psi(u_1(\theta'))|^2\, d\sigma_{n-1}(\theta)\, d\sigma_{n-1}(\theta') \\ &\le \frac{1}{2}\iint |u_1(\theta) - u_1(\theta')|^2\, d\sigma_{n-1}(\theta)\, d\sigma_{n-1}(\theta') = \mathrm{Var}\,(u_1) = \frac{1}{n}.\end{aligned}$$

Hence, $\mathbf{E}_\theta\, |u - \mathbf{E}_\theta u| \le \frac{1}{\sqrt{n}}$. This inequality was applied in our remark on logarithmically concave probability distributions and in the proof of Theorem 4 (with a doubled constant since the function u is complex-valued).

The isoperimetric theorem allows one to estimate the σ_{n-1}-measure for large deviations as well. In particular, under the same assumption that $\|u\|_{\mathrm{Lip}} \le 1$, we conclude that

$$\sigma_{n-1}\{u - m(u) \ge t\} \le \sigma_{n-1}\{u_1 \ge t\}$$

for all $t > 0$, where $m(u)$ is the median (or one of medians) for u in the sense of σ_{n-1}. The right-hand side of the above inequality is estimated by the expression $e^{-(n-1)\,t^2/2}$, similarly to Lemma 1. However, if we want to pass from a median to mathematical expectation without losses, it is better to apply the logarithmic Sobolev inequality,

$$\mathbf{E}_\theta\, u^2 \log u^2 - \mathbf{E}_\theta\, u^2\, \log\, \mathbf{E}_\theta\, u^2 \le 2C_n\, \mathbf{E}_\theta |\nabla u|^2,$$

where it is assumed that u is an arbitrary continuously differentiable function. It is well known (see, for example, [9]) that, for functions with $\|u\|_{\mathrm{Lip}} \le 1$, such an integro-differential inequality guarantees that tails of the corresponding distributions are sub-Gaussian; namely, the estimate

$$\sigma_{n-1}\{u - \mathbf{E}_\theta\, u \ge t\} \le e^{-t^2/2C_n}, \quad t > 0,$$

holds. Finally, it is shown in [11] that the best constant in the logarithmic Sobolev inequality for the measure σ_{n-1} is $C_n = 1/(n-1)$.

This research was supported by the NSF (project DMS-0706866).

Transalted by S. Yu. Pilyugin.

REFERENCES

1. M. Antilla, K. Ball, and I. Perissinaki, "The central limit problem for convex bodies," *Trans. Amer. Math. Soc.*, **355**, 4723–4735 (2003).
2. S. G. Bobkov, "On concentration of distributions of random weighted sums," *Ann. Probab.*, **31**, 195–215 (2003).

3. S. G. Bobkov, “Concentration of distributions of the weighted sums with Bernoullian coefficients,” *Lect. Notes Math.*, **1807**, 27–36 (2003).
4. S. G. Bobkov and C. Houdré, “Some connections between isoperimetric and Sobolev-type inequalities,” *Mem. Amer. Math. Soc.*, **129** (1997).
5. C. Borell, “Convex measures on locally convex spaces,” *Ark. Mat.*, **12**, 239–252 (1974).
6. P. Diaconis and D. Freedman. “Asymptotics of graphical projection pursuit,” *Ann. Stat.*, **12**, 793–815 (1984).
7. R. M. Dudley, *Real Analysis and Probability.* The Wadsworth & Brooks/Cole Mathematics Series, Pacific Grove, CA (1989).
8. B. Klartag, “A central limit theorem for convex sets,” *Invent. Math.*, **168**, 91–131 (2007).
9. M. Ledoux, “Concentration of measure and logarithmic Sobolev inequalities,” *Lect. Notes Math.*, **1709**, 120–216 (1999).
10. V. G. Mazya, *S. L. Sobolev Spaces* [in Russian], Leningr. Univ. (1985).
11. C. E. Mueller and F. B. Weissler, “Hypercontractivity for the heat semigroup for ultraspherical polynomials and on the n-sphere,” *J. Funct. Anal.*, **48**, 252–283 (1982).
12. S. V. Nagayev, “On distributions of linear functionals in finite-dimensional spaces of high dimension,” *Dokl. Akad. Nauk SSSR*, **263**, 295–296 (1982).
13. S. Sodin, “Tail-sensitive Gaussian asymptotics for marginals of concentrated measures in high dimension,” *Lect. Notes Math.*, **1910**, 271–295 (2007).
14. A. V. Sudakov and V. N. Sudakov, “Empiric distributions and uniform in dimension estimates of dispersion in the problem on typical distributions,” *Zap. Nauchn. Semin. POMI*, **260**, 258–262 (1999).
15. V. N. Sudakov, “Typical distributions of linear functionals on spaces of high dimension,” *Dokl. Akad. Nauk SSSR*, **243**, 1402–1405 (1978).
16. V. N. Sudakov, “On distributions of randomly chosen random values,” *Zap. Nauchn. Semin. POMI*, **165**, 153–160 (1994).
17. H. von Weizsäcker, “Sudakov’s typical marginals, random linear functionals and a conditional central limit theorem,” *Probab. Theory Rel. Fields*, **107**, 313–324 (1997).
18. V. M. Zolotarev, “Estimates of closeness of distributions in the Lévy metric,” *Trudy Mat. Inst. AN SSSR*, **112**, 224–231 (1971).