ESTIMATES OF TRANSPORT DISTANCE
IN THE CENTRAL LIMIT THEOREM

ANDTEI YU. ZAITSEV

1. INTRODUCTION

Sums of independent terms first appeared in probability theory when considering binomial distributions within the Bernoulli scheme. The law of large numbers and the Moivre-Laplace central limit theorem were derived. It was observed that binomial distributions are not only well approximated by the normal law, but also that the tail decay of binomial distributions is similar to that of normal distributions. As a natural extension of the class of binomial distributions, we can consider the class of distributions of sums of independent (generally, non-identically distributed) random variables bounded in absolute value by the same constant. A significant number of works are devoted to estimating the tails of such distributions, see, for example, [2, 3, 4, 15, 19, 21], In this paper we will discuss not only the tail decay, but also the estimation of the approximation accuracy in the one-dimensional central limit theorem.

Let $X, X_1, \ldots, X_n$ be d-dimensional independent random vectors bounded with probability one. For simplicity, we will assume that they have zero mean values:

$$\mathbf{P}\{\|X_j\| \leq \tau\} = 1, \quad \mathbf{E}\, X_j = 0, \quad j = 1, \ldots, n. \tag{1}$$

We will be interested in the behavior of the distribution of the sum $S = X_1 + \cdots + X_n$ depending on the limiting value $\tau > 0$.

From the non-uniform Bikelis estimate [7] in the one-dimensional central limit theorem it follows that in the one-dimensional case

$$W_1(F, \Phi) \leq c\tau. \tag{2}$$

with an absolute constant c, where W_1 is the Kantorovich–Rubinstein–Wasserstein transport distance (see review articles [9, 10]), $F = \mathcal{L}(S)$ is the distribution of the sum S, and $\Phi = \Phi_F$ is the corresponding normal distribution, with the same zero mean and the same variance as those of the distribution F. When proving inequality (2), it should be taken into account that, according to [33],

$$W_1(F, \Phi) = \int |F(x) - \Phi(x)|\, dx, \tag{3}$$

where $F(\,\cdot\,)$ and $\Phi(\,\cdot\,)$ are the corresponding distribution functions. In addition, $\mathbf{E}\,|X_j|^3 \leq \tau\, \mathbf{E}\, X_j^2$.

The main result of the paper is significantly stronger and more precise. It is claimed that

$$W(F, \Phi) = \inf_\pi \int \exp\{|x - y|/c\tau\}\, d\pi(x, y) \leq c, \tag{4}$$

where the infimum is taken over all two-dimensional probability distributions π with marginal distributions F and Φ. The result is also generalized to distributions with sufficiently slowly growing cumulants from the class $\mathcal{A}_1(\tau)$ introduced in the author's paper [36]. In special cases, we obtain some results of Rio [23]. The possibility of generalizing the result to the multidimensional case is discussed.

Following Rio [23], we define the Wasserstein distance associated with the Orlicz function ψ:

$$W_\psi(G, H) = \inf\left\{a > 0 : \inf_\pi \int \psi(|x - y|/a)\, d\pi(x, y) \leq 1\right\}, \tag{5}$$

where the second infimum is taken over all two-dimensional probability distributions π with marginal distributions G and H.

Inequality (4) can be rewritten as

$$W_\psi(F, \Phi) \leq c\tau, \tag{6}$$

with the Orlicz function $\psi(x) = \exp\{|x|\} - 1$. Inequality (2) can also be written in the form (6), but for the Orlicz function $\psi(x) = |x|$. Inequality (6) is also valid for the Orlicz function $\psi(x) = |x|^p$, $p \geq 1$. In this case, the statement is easily deduced from (6) and turns into the estimate

$$W_p(F, \Phi) \leq c(p)\, \tau, \tag{7}$$

where $W_P(\,\cdot\,, \,\cdot\,)$ is the standard Wasserstein p-distance. We took into account that $|x|^p \leq c(p) \exp\{|x|\}$.

The class of distributions of sums $S = X_1 + \cdots + X_n$ satisfying conditions (1) can be considered as a natural generalization of the class of binomial distributions, which historically turned out to be the first distributions of sums of independent terms studied in probability theory. Bernstein [6] found less restrictive conditions (see definition (20) for $d = 1$), under which the tails of the distributions of sums admit estimates similar to those for the tails of binomial distributions. Under the conditions of Bernstein's inequality, the distributions of the terms have finite exponential moments, that is, the Cramér conditions are satisfied, under which the theorems on large deviations of the distributions of sums of independent terms are valid. As is well known, the coefficients of the so-called Cramér–Petrov series arising in the formulations are determined from the cumulants of the distributions of the sums, see [27, Lemma 1.4]. This motivated Statulevičius [31] to a further expansion of the class of distributions for which the results on large deviations are valid. He introduced classes of distributions that are no longer necessarily representable as distributions of sums of a large number of independent terms, but whose cumulants behave similarly to the cumulants of such sums (see (14)). In this paper, we prove inequality (6) not only for distributions of sums $S = X_1 + \cdots + X_n$ satisfying conditions (1), but also for distributions from the class $\mathcal{A}_1(\tau)$, equivalent to the class of one-dimensional distributions considered by Statulevičius.

Let $\mathcal{A}_d(\tau)$, $\tau \geq 0$, $d \in \mathbf{N}$, be the class of d-dimensional distributions introduced in an author's paper [36]. The class $\mathcal{A}_d(\tau)$ (with a fixed $\tau \geq 0$) consists of d-dimensional distributions F, for which the function

$$\varphi(z) = \varphi(F, z) = \log \int_{\mathbf{R}^d} e^{\langle z, x \rangle} F\{dx\} \quad (\varphi(0) = 0) \tag{8}$$

is defined and analytical for $\|z\| \tau < 1$, $z \in \mathbf{C}^d$, and

$$\left| d_u d_v^2 \varphi(z) \right| \leq \|u\| \tau \left\langle \mathbb{D} v, v \right\rangle, \tag{9}$$

for all $u, v \in \mathbf{R}^d$ and $\|z\| \tau < 1$, where $\mathbb{D} = \operatorname{cov} F$ is the covariance operator of distribution F, and $d_u \varphi$ is the derivative of function φ in direction u.

Let's introduce the necessary notation. Below, the symbols $c, c_1, c_2, c_3, \ldots$ will be used for absolute positive constants. Note that c can be different in different (or even in the same) formulas. We will write $A \ll B$ if $A \leq cB$. We will also use the notation $A \asymp B$ if $A \ll B$ and $B \ll A$. If the corresponding constant depends on, say, r, we will write $c(r)$, $A \ll_r B$, and $A \asymp_r B$. By $\widehat{F}(t) = \int e^{itx} F\{dx\}$, $t \in \mathbf{R}$, we denote the characteristic function of the univariate distribution F.

The main result of this paper is contained in the following Theorems 1 and 2. They deal with the proximity of univariate distributions.

Theorem 1. *Let $F = \mathcal{L}(\xi) \in \mathcal{A}_1(\tau)$, $\tau > 0$, $\mathbf{E}\,\xi = 0$. Then there exists an absolute constant c_1 such that*

$$W(F, \Phi) = \inf_\pi \int \exp\{|x - y|/c_1 \tau\} \, d\pi(x, y) \leq c_1, \tag{10}$$

where $\Phi = \Phi_F$ is the corresponding normal distribution, and the infimum is taken over all two-dimensional probability distributions π with marginal distributions F and Φ.

Theorem 2. *Let $F = \mathcal{L}(\xi) \in \mathcal{A}_1(\tau)$, $\tau > 0$, $\mathbf{E}\,\xi = 0$. Then there exists an absolute constant c_2 such that*

$$W_\psi(F, \Phi) \leq c_2 \tau, \tag{11}$$

with the Orlicz function $\psi(x) = \exp\{|x|\} - 1$, where $\Phi = \Phi_F$.

Theorems 1 and 2 are equivalent. If $c_1 \leq 2$, then Theorem 1 implies that

$$\inf_\pi \int \exp\{|x - y|/c_1 \tau\} \, d\pi(x, y) \leq 2. \tag{12}$$

If $c_1 > 2$, then we can choose c_3 so that $c_1^{c_3} = 2$, and $c_3 < 1$. Then, by Lyapunov's inequality for moments,

$$\inf_\pi \int \exp\{c_3 |x - y|/c_1 \tau\} \, d\pi(x, y) \leq \inf_\pi \left(\int \exp\{|x - y|/c_1 \tau\} \, d\pi(x, y) \right)^{c_3} \leq c_1^{c_3} = 2. \tag{13}$$

Now inequalities (12), (13) imply the statement of Theorem 2. It is also obvious that Theorem 2 implies the statement of Theorem 1.

2. Properties of classes $\mathcal{A}_d(\tau)$

Let's consider the elementary properties of classes $\mathcal{A}_d(\tau)$ (see [36, 38, 40, 41, 42]). It is easy to see that if $\tau_1 < \tau_2$, then $\mathcal{A}_d(\tau_1) \subset \mathcal{A}_d(\tau_2)$. Furthermore, the class $\mathcal{A}_d(\tau)$ is closed under the convolution operation: if $F_1, F_2 \in \mathcal{A}_d(\tau)$, then $F_1 F_2 = F_1 * F_2 \in \mathcal{A}_d(\tau)$. From here on, products and powers of measures are understood in the sense of convolution.

Let $\tau \geq 0$, $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $y \in \mathbf{R}^m$, and $\mathbb{A} : \mathbf{R}^d \to \mathbf{R}^m$ is a linear operator. Then

$$\mathcal{L}(\mathbb{A}\,\xi + y) \in \mathcal{A}_m\left(\|\mathbb{A}\|\,\tau\right), \quad \text{where} \quad \|\mathbb{A}\| = \sup_{x \in \mathbf{R}^d, \|x\| \leq 1} \|\mathbb{A}\,x\|.$$

In particular, for any $a \in \mathbf{R}$

$$\mathcal{L}(a\,\xi) \in \mathcal{A}_d\left(|a|\,\tau\right).$$

The classes $\mathcal{A}_d(\tau)$ are closely related to other naturally defined classes of multivariate distributions. From the definition of $\mathcal{A}_d(\tau)$ it follows that if $\mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, then the vector ξ has finite exponential moments $\mathbf{E}\,e^{\langle h, \xi \rangle} < \infty$, for $h \in \mathbf{R}^d$, $\|h\|\,\tau < 1$. This leads to an exponential decay of the distribution tails.

The condition $\mathcal{L}(\xi) \in \mathcal{A}_1(\tau)$ is equivalent to the condition of Statulevičius [31], see also [13, 26, 27], on the growth rate of the cumulants γ_m of a random variable ξ:

$$|\gamma_m| \leq \frac{1}{2}\,m!\,\tau^{m-2}\gamma_2, \qquad m = 3, 4, \ldots. \tag{14}$$

This equivalence means that if one of these conditions is satisfied with the parameter τ, then the second is valid with the parameter $c\tau$, where c is some positive absolute constant. Note, however, that the condition $\mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$ differs significantly from other multivariate analogues of the Statulevičius condition considered by Rudzkis [24] and Saulis [25]. The review article [27] and the monograph [26] contain a large number of examples of distributions satisfying conditions (14) and which are not distributions of sums of a large number of independent terms. Note also that Statulevičius considered more general conditions under which exponential moments are not necessarily finite.

Another class of distributions, denoted by $\widetilde{\mathcal{A}}_d(\tau)$, $\tau \geq 0$, was mentioned in the paper [42]. It is defined similarly to $\mathcal{A}_d(\tau)$ with replacing (9) with

$$\left|\,d_v^2\,\varphi(z)\right| \leq 2\left\langle \mathbb{D}\,v, v \right\rangle \tag{15}$$

for all $v \in \mathbf{R}^d$ and $\|z\|\,\tau < 1$. That the classes $\widetilde{\mathcal{A}}_d(\tau)$ and $\mathcal{A}_d(\tau)$ are also equivalent is easily verified using Cauchy inequalities. The definition of the classes $\widetilde{\mathcal{A}}_d(\tau)$ in some sense looks even more natural than the definition of the classes $\mathcal{A}_d(\tau)$. The constant 2 in (15) can be replaced by any other constant C, $1 < C < \infty$, bounded away from 1 and from infinity. The result will also be equivalent classes.

Clearly, the class $\mathcal{A}_d(0)$ coincides with the class of all d-dimensional Gaussian distributions. The following inequality (16) was proved in the author's paper [36] and can be considered as an estimate of the stability of this characterization:

$$\text{if } F \in \mathcal{A}_d(\tau), \text{ then } \pi\left(F,\,\Phi_F\right) \leq c\,d^2\tau\,\log^*(\tau^{-1}), \tag{16}$$

where $\pi(\cdot,\cdot)$ is the Prokhorov distance, defined in [22], and Φ_F denotes the Gaussian distribution whose mean and covariance operator are the same as those of F. Here $\log^* b = \max\{1, \log b\}$ for $b > 0$, and $\log$ is used to denote the natural logarithm. Note that Theorems 1 and 2 of this paper can also be viewed as stability estimates for the above-mentioned characterization of Gaussian distributions in transport metrics, and so far in the one-dimensional case.

The Prokhorov distance between distributions F, G can be determined by the formula

$$\pi(F, G) = \inf\{\lambda : \pi(F, G, \lambda) \le \lambda\},$$

where

$$\pi(F, G, \lambda) = \sup_Y \max\left\{F\{Y\} - G\{Y^\lambda\}, G\{Y\} - F\{Y^\lambda\}\right\}, \quad \lambda > 0,$$

and $Y^\lambda = \{y \in \mathbf{R}^d : \inf_{x \in Y} \|x - y\| < \lambda\}$ is the λ-neighborhood of a Borel set Y (see [11, 12]).

In the author's paper [36] it was also established that

$$\text{if } F \in \mathcal{A}_d(\tau), \text{ then } \pi(F, \Phi_F, \lambda) \le c\, d^2 \exp\left\{-\frac{\lambda}{c\, d^2 \tau}\right\}, \quad \lambda > 0. \tag{17}$$

What is important here is that inequality (17) is proved for all $\tau > 0$ and for an arbitrary covariance operator $\operatorname{cov} F$.

By the Strassen–Dudley theorem (see Dudley [14]) and according to inequality (17), for any distribution $F \in \mathcal{A}_d(\tau)$ and any $\lambda > 0$, one can construct random vectors ξ and η on the same probability space with $\mathcal{L}(\xi) = F$ and $\mathcal{L}(\eta) = \Phi_F$, so that

$$\mathbf{P}\{\|\xi - \eta\| > \lambda\} = \pi(F, \Phi_F, \lambda) \le c\, d^2 \exp\left\{-\frac{\lambda}{c\, d^2 \tau}\right\}. \tag{18}$$

We emphasize that the Strassen–Dudley theorem guarantees the existence of a construction with equality in (18) only for a fixed λ. An example showing impossibility of a construction with equality in (18) for all λ simultaneously can be found in the survey [11]. The Strassen–Dudley theorem enables us to automatically derive statements of the type (18) from estimates for $\pi(F, G, \lambda)$. Strassen's original proof [32] was non-constructive. Dudley [14] gave a complicated constructive proof based on combinatorial ideas. Finally, Schay [30] found a short proof relying on the duality theorem.

If equality (18) were proven for all $\lambda > 0$ simultaneously on the same probability space, then the assertion of Theorem 1 would automatically follow from it, for any dimension d, $1 \le d < \infty$. Therefore, inequality (17) gives grounds to expect a possibility to generalize Theorems 1 and 2 to the multidimensional case.

If F is an infinitely divisible distribution with spectral measure concentrated on the ball $\{x \in \mathbf{R}^d : \|x\| \le \tau\}$, then $F \in \mathcal{A}_d(c\tau)$, where c is some positive absolute constant. In the paper [36], one can find less restrictive conditions on spectral measure, ensuring that an infinitely divisible distribution belongs to $\mathcal{A}_d(c\tau)$.

In particular, for the Poisson distribution Π_λ with parameter $\lambda > 0$, the inclusion $\Pi_\lambda \in \mathcal{A}_1(c)$ holds. It follows from Theorem 2 that

$$\sup_\lambda W_\psi(\Pi_\lambda, \Phi_{\Pi_\lambda}) \leq c, \quad \text{for } \psi(x) = \exp\{|x|\} - 1. \tag{19}$$

This is the statement of Corollary 2.2 of Rio [23]. But Theorem 2 contains a more general assertion. In (19) we can replace the set of Poisson distributions by the set of all infinitely divisible distributions with the Lévy–Khinchin spectral measures concentrated on the interval $[-1, 1]$.

Distributions from the classes $\mathcal{A}_d(\tau)$ are directly used in the formulations of the author's results [40]–[42] on estimating the accuracy of strong Gaussian approximations for sums of independent random vectors in the most important case, when the summands have finite exponential moments (see also the review article [43]). Multivariate analogues of the one-dimensional results of Sakhanenko [28], who generalized and significantly refined the results of Komlós, Major, and Tusnády [18], for the case of non-identically distributed random variables, were obtained. Sakhanenko considered the following classes of univariate distributions:

$$\mathcal{S}_1(\tau) = \left\{ \mathcal{L}(\xi) : \; \mathbf{E}\,\xi = 0, \; \mathbf{E}\,|\xi|^3 \exp\left\{|\xi|/\tau\right\} \leq \tau\,\mathbf{E}\,|\xi|^2 \right\}, \quad \tau > 0.$$

In the author's preprint [35] it was noted that the classes $\mathcal{S}_1(\tau)$ are equivalent to the classes of distributions $\mathcal{B}_1(\tau)$ satisfying the conditions of the Bernstein inequality (see definition (20)), in the sense that if one of the conditions for membership in a class is satisfied with a parameter τ, then the second is true with parameter $c\tau$, where c is some positive absolute constant. Sakhanenko's results [28] were formulated as estimates of exponential moments of maximal deviation of sums of independent random variables with distributions in $\mathcal{S}_1(\tau)$, constructed on a probability space, from corresponding sums of independent normally distributed terms. The form of Sakhanenko's estimates is almost the same as in (10), so that they imply an analogue of inequality (10) with right-hand side replaced by $c(1 + \sigma/\tau)$ for the distributions of sums of independent random variables with distributions from $\mathcal{S}_1(\tau)$. Here, σ^2 denotes the variance of the sum under consideration. For convolutions of distributions from $\mathcal{S}_1(\tau)$, some estimates of moments of exponential type in the central limit theorem are contained in Sakhanenko [29].

In the author's papers [35] and [37], inequalities (16) and (17) (with d^2 replaced by $d^{5/2}$) were proved for convolutions of distributions from the class $\mathcal{B}_d(\tau)$, where $\tau > 0$ and

$$\begin{aligned}
\mathcal{B}_d(\tau) \;=\; & \left\{ F = \mathcal{L}(\xi) : \mathbf{E}\,\xi = 0, \; \left| \mathbf{E}\,\langle \xi, v \rangle^2 \langle \xi, u \rangle^{m-2} \right| \right. \tag{20} \\
& \left. \leq \; \frac{1}{2} m!\,\tau^{m-2}\, \|u\|^{m-2}\,\mathbf{E}\,\langle \xi, v \rangle^2 \quad \text{for all } u, v \in \mathbf{R}^d, \; m = 3, 4, \dots \right\},
\end{aligned}$$

satisfying multidimensional analogues of the Bernstein inequality conditions. Sakhanenko's condition $\mathcal{L}(\xi) \in \mathcal{S}_1(\tau)$ is equivalent to the condition $\mathcal{L}(\xi) \in \mathcal{B}_1(\tau)$. Note that if $F\left\{\left\{x \in \mathbf{R}^d : \|x\| \leq \tau\right\}\right\} = 1$, $\mathbf{E}\,\xi = 0$, then $F \in \mathcal{B}_d(\tau)$.

Let us formulate relations between classes $\mathcal{A}_d(\tau)$ and $\mathcal{B}_d(\tau)$. Let σ_F^2 denote the maximal eigenvalue of the covariance operator of distribution F. Then

a) If $F = \mathcal{L}(\xi) \in \mathcal{B}_d(\tau)$, then $\sigma_F^2 \leq 12\,\tau^2$, $\mathbf{E}\,\xi = 0$ and $F \in \mathcal{A}_d(c\tau)$.

b) If $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $\sigma_F^2 \leq \tau^2$ and $\mathbf{E}\,\xi = 0$, then $F \in \mathcal{B}_d(c\tau)$.

In particular, the distribution of the sum $S = X_1 + \cdots + X_n$ under conditions (1) belongs to the class $\mathcal{A}_d(c\tau)$.

Thus, roughly speaking, $\mathcal{B}_d(\tau)$ forms a subclass of distributions $F = \mathcal{L}(\xi - \mathbf{E}\,\xi)$ such that $\mathcal{L}(\xi) \in \mathcal{A}_d(c\tau)$, and $\sigma_F^2 \leq 12\,\tau^2$. Inequalities (16) and (17) in this case indicate only that both distributions being compared are close to degenerate law E concentrated at the origin. If $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$ and σ_F^2 is significantly greater than τ^2, then $\mathcal{L}(\xi/\sigma_F) \in \mathcal{A}_d(\tau/\sigma_F)$ and inequalities (16) and (17) reflect the proximity of distribution F to the corresponding Gaussian law.

Let $\tau \geq 0$, $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $\|h\|\tau < 1$, $h \in \mathbf{R}^d$. We define the distribution $\overline{F} = \overline{F}(h)$ by relation

$$\overline{F}\{dx\} = \left(\mathbf{E}\,e^{\langle h, \xi \rangle}\right)^{-1} e^{\langle h, x \rangle} F\{dx\}.$$

We denote by $\overline{\xi} = \overline{\xi}(h)$ a random vector with distribution $\mathcal{L}(\overline{\xi}(h)) = \overline{F}(h)$. Distributions $\overline{F}(h)$ are sometimes called Cramér transforms (or Esscher transforms, see [8]). In the proofs in [36], [37], [40]–[42], distributions $\overline{F}(h)$ are used to estimate probabilities of large deviations, corresponding to the conditional densities. Another important property of classes $\mathcal{A}_d(\tau)$ is that $\overline{F}(h) \in \mathcal{A}_d(2\tau)$ for $\|h\|\,\tau \leq 1/2$, see item b) of Lemma 1. This makes it possible to systematically apply the results obtained for the original distributions to their Cramér transforms and thereby refine the estimates.

Kolmogorov [16] posed the problem of estimating the accuracy of infinitely divisible approximation of distributions of sums of independent random variables whose distributions are concentrated on short intervals of small length $\tau \leq 1/2$ up to a small probability p. In the particular case when $p = 0$, we are talking about approximating the distributions of sums $S = X_1 + \cdots + X_n$ for $d = 1$ and under conditions (1). In this case, Kolmogorov [16, 17] obtained the estimate

$$L(F, \Phi_F) \ll \tau^{1/2} \log^{1/4}(1/\tau), \tag{21}$$

where $L(\,\cdot\,, \,\cdot\,)$ is the Lévy distance. From the above it follows that $F \in \mathcal{A}_1(c\tau)$ and inequalities (16) and (17) can be viewed as generalizations, and refinements of inequality (21). Note that the formulations in [16, 17] differ from (21). To deduce this inequality from them, an elementary additional analysis is required.

Let $X, X_1, \ldots, X_n$ be independent identically distributed random variables such that

$$\mathbf{E}\,X = 0 \quad \text{and} \quad \mathbf{E}\,\exp\{t|X|\} < \infty \quad \text{for some} \quad t > 0. \tag{22}$$

Then it is easy to verify that there exists $c(F)$ such that $F = \mathcal{L}(X) \in \mathcal{A}_1(c(F))$, and distribution of the normalized sum $F_n = \mathcal{L}\big((X_1 + \cdots + X_n)/\sqrt{n}\big)$ belongs to $\mathcal{A}_1(c(F)/\sqrt{n})$. Applying Theorem 2, we obtain that

$$W_\psi(F_n, \Phi_{F_n}) \ll_F 1/\sqrt{n}, \tag{23}$$

with Orlicz function $\psi(x) = \exp\{|x|\} - 1$. This is the statement of Theorem 2.1 of Rio [23]. We emphasize once again that in the main results of this article we consider univariate distributions that satisfy condition (9) on the Laplace transforms and, generally speaking, cannot be represented as convolutions of a large number of identical distributions. In this case, condition (9) turns into

$$|\varphi'''(z)| \leq \tau\sigma^2 \quad \text{for } |z|\tau \leq 1. \tag{24}$$

From the above it follows that the assertions of Theorems 1 and 2 are also valid for convolutions of univariate distributions concentrated on the interval $[-\tau, \tau]$ or satisfying the conditions of Bernstein's inequality, as well as for infinitely divisible distributions with their Lévy–Khinchin spectral measures concentrated on the same interval, and for distributions satisfying the Statulevičius conditions (14). In terms of content and methods of proof, they can be viewed as simply and clearly formulated statements from large deviation theory.

3. Proof of Theorem 1

We will need the following Lemma 1 about the properties of the Cramér transform (it is contained in [36, Lemmas 2.1, 3.1]).

Lemma 1. *Let* $\tau \geq 0$, $\quad F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $\mathbf{E}\,\xi = 0$, $h \in \mathbf{R}^d$, $\|h\|\tau < 1$, $\quad \mathbb{D} = \operatorname{cov}F$, $\mathbb{D}(h) = \operatorname{cov}\overline{F}(h)$. *Denote by* σ^2 *the minimal eigenvalue of* $\mathbb{D}$. *Then*
a) for any $u, v \in \mathbf{R}^d$ *the following relations hold:*

$$\langle \mathbb{D}(h)\,u,\,u\rangle = \langle \mathbb{D}\,u,\,u\rangle\big(1 + \theta\|h\|\tau\big), \tag{25}$$

$$\log \mathbf{E}\,e^{i\langle h, \xi\rangle} = -\frac{1}{2}\,\langle \mathbb{D}\,h,\,h\rangle\Big(1 + \frac{1}{3}\,\theta\|h\|\tau\Big), \tag{26}$$

$$\log \mathbf{E}\,e^{\langle h, \xi\rangle} = \frac{1}{2}\,\langle \mathbb{D}\,h,\,h\rangle\Big(1 + \frac{1}{3}\,\theta\|h\|\tau\Big) \tag{27}$$

(here and below θ *symbolizes various quantities not exceeding one in absolute value:* $|\theta| \leq 1$*);*
b) If $\|h\|\tau \leq 1/2$, *then* $\overline{F}(h) \in \mathcal{A}_d(2\tau)$;
c) For $x \in \Pi = \big\{x \in \mathbf{R}^d : 4.8\,\tau\sigma^{-1}\,\big\|\mathbb{D}^{-1/2}x\big\| \leq 1\big\}$ *there exists a parameter* $h = h(x) \in \mathbf{R}^d$ *such that*

$$\mathbf{E}\,\overline{\xi}(h) = x, \tag{28}$$

$$\|h\|\tau \leq 1/2, \tag{29}$$

$$\sigma\,\|h\| \leq \big\|\mathbb{D}^{1/2}h\big\| \leq 2.4\,\big\|\mathbb{D}^{-1/2}x\big\|, \tag{30}$$

$$\big\|\mathbb{D}^{1/2}h - \mathbb{D}^{-1/2}x\big\| \leq 2.88\,\theta\tau\sigma^{-1}\,\big\|\mathbb{D}^{-1/2}x\big\|^2, \tag{31}$$

$$\mathbf{E}\exp\big\{\langle h, \xi\rangle - \langle h, x\rangle\big\} = \exp\Big\{-\frac{1}{2}\,\big\|\mathbb{D}^{-1/2}x\big\|^2 + 10.08\,\theta\tau\sigma^{-1}\,\big\|\mathbb{D}^{-1/2}x\big\|^3\Big\}. \tag{32}$$

Below

$$\Xi(x) = e^{x^2/2} \int_x^\infty e^{-y^2/2}\,dy, \quad x > 0, \tag{33}$$

is the Mills ratio. We will need the following lemma (see [1, Lemma 1.2 of Chapter VI]).

Lemma 2. *Let* $x, \varepsilon > 0$*. Then*

$$0 \leq \Xi(x) - \Xi(x + \varepsilon) \leq \frac{\varepsilon}{x^2}, \tag{34}$$

$$\Xi(x) = \frac{1}{x}\left(1 - \frac{|\theta|}{x^2}\right). \tag{35}$$

Let Φ_σ *be the univariate Gaussian distribution with zero mean and variance* σ^2*. Then*

$$1 - \Phi_\sigma(x + \varepsilon) \leq (1 - \Phi_\sigma(x)) \exp\left\{-\frac{2x\varepsilon + \varepsilon^2}{2\sigma^2}\right\}. \tag{36}$$

Hence, for $x > \varepsilon$

$$1 - \Phi_\sigma(x) \leq (1 - \Phi_\sigma(x - \varepsilon)) \exp\left\{-\frac{2x\varepsilon - \varepsilon^2}{2\sigma^2}\right\}. \tag{37}$$

Let $\rho(F, \Phi) = \sup_x |F(x) - \Phi(x)|$ be the Kolmogorov distance, uniform distance between distribution functions.

Lemma 3. *Under the conditions of Theorem* 1,

$$\rho(F, \Phi) \ll \tau/\sigma, \tag{38}$$

where $\sigma^2 = \operatorname{Var}\xi$ *is the common variance of distributions* F *and* Φ*.*

Proof. Using inequality

$$\left|e^{z_1} - e^{z_2}\right| \leq \left|z_1 - z_2\right| \max\left\{\left|e^{z_1}\right|, \left|e^{z_2}\right|\right\}, \tag{39}$$

valid for $z_1, z_2 \in \mathbb{C}$, and applying inequality (26), we obtain that for $|t|\tau \leq 1$

$$\left|\widehat{F}(t) - \widehat{\Phi}(t)\right| \leq \frac{\tau}{6}\sigma^2|t|^3 \exp\left\{-\frac{1}{3}\sigma^2 t^2\right\}. \tag{40}$$

Therefore, using the standard smoothing inequality (see [1, Theorem 1.2 of Chapter III]), we find that for $T = 1/\tau$

$$\rho(F, \Phi) \ll \int_0^T \left|\frac{\widehat{F}(t) - \widehat{\Phi}(t)}{t}\right| dt + \frac{1}{\sigma T} \ll \frac{\tau}{\sigma}. \tag{41}$$

The following lemma contains an analogue of Bernstein's inequality for distributions from the class $\mathcal{A}_1(\tau)$.

Lemma 4. *Let, under the conditions of Theorem* 1, $\operatorname{Var}\xi = \sigma^2$*. Then*

$$\mathbf{P}\{\xi \geq x\} \leq \max\left\{\exp\left\{-\frac{x^2}{4\,\sigma^2}\right\}, \exp\left\{-\frac{x}{4\,\tau}\right\}\right\}, \qquad x \geq 0. \tag{42}$$

The proof of this lemma almost literally repeats the proof of Bernstein's inequality. Let $0 \le h\tau \le \frac{1}{2}$. By (27),

$$\mathbf{E}\, e^{h\xi} \le \exp\{h^2\sigma^2\}$$

and

$$\mathbf{P}\{\xi \ge x\} \le e^{-hx}\, \mathbf{E}\, e^{h\xi} \le \exp\{h^2\sigma^2 - hx\}.$$

Let's choose the parameter h depending on x. If $0 \le x \le \frac{\sigma^2}{\tau}$, we take $h = \frac{x}{2\sigma^2}$ and obtain the bound

$$\mathbf{P}(S \ge x) \le \exp\left\{-\frac{x^2}{4\,\sigma^2}\right\}. \tag{43}$$

And if $x > \frac{\sigma^2}{\tau}$, take $h = \frac{1}{2\tau}$ and get

$$\mathbf{P}(S \ge x) \le \exp\left\{\frac{\sigma^2}{4\tau^2} - \frac{x}{2\tau}\right\} \le \exp\left\{-\frac{x}{4\,\tau}\right\}. \tag{44}$$

Now the inequality (42) follows from (43) and (44).

Proof of Theorem 1. Without loss of generality, we assume that the distribution function F is infinitely differentiable and strictly increasing. To justify this, it suffices to consider, instead of the distribution F, the convolution of this distribution with a Gaussian distribution having zero mean and positive variance tending to zero, and also to use the standard tool for proving theorems on strong approximation, Lemma A from Berkes and Philipp [5]; see, for example, the proof of Theorem 3.1 in Rio [23]. Thus, under these assumptions, the strictly increasing inverse function $F^{-1}(\,\cdot\,)$ is well defined.

Let a random variable η have the distribution $\mathcal{L}(\eta) = \Phi$. Write $\xi = F^{-1}\big(\Phi(\eta)\big)$. It is clear that $\mathcal{L}(\xi) = F$ and $\eta = \Phi^{-1}\big(\Phi(\eta)\big)$. This means that the random variables ξ and η are defined as the Smirnov transforms of a random variable $\Phi(\eta)$ uniformly distributed on the interval $[0,1]$. This is exactly how random variables with given distributions are constructed in the proof of equality (3) in [33]. Then if the random variable ξ takes some specific value $x \in \mathbf{R}$, the random variable η will take the value $\Phi^{-1}\big(F(x)\big)$.

Further reasoning is carried out under the assumption that $\xi = x$ and $\tau \le c_4\sigma$, where $\sigma^2 = \mathrm{Var}\,\xi$, and the choice of c_4 will be refined during the proof.

First we consider the case when $|x| \le 2\sigma$. Let

$$\phi(u) = \frac{1}{\sqrt{2\pi}\,\sigma}\, \exp\left(-\frac{u^2}{2\,\sigma^2}\right) \tag{45}$$

be the density of distribution Φ. Recall that according to Lemma 3

$$\rho(F, \Phi) \le c_5 \frac{\tau}{\sigma}. \tag{46}$$

Let $|u| \le 2\sigma$, $|y| \le 3\sigma$. It is obvious that then

$$\big|\Phi(u) - \Phi(y)\big| \ge |u-y|\,\phi(3\sigma) = \frac{e^{-9/2}\,|u-y|}{\sqrt{2\pi}\,\sigma} = \frac{c_6\,|u-y|}{\sigma}, \quad c_6 = \frac{e^{-9/2}}{\sqrt{2\pi}}, \tag{47}$$

$$|F(u) - F(y)| \geq \frac{c_6\,|u - y|}{\sigma} - 2c_5\frac{\tau}{\sigma}. \tag{48}$$

Lemma 5. *There exist absolute positive constants c_4, c_7 such that for $\tau \leq c_4\sigma$, $|x| \leq 2\sigma$ the following inequalities hold:*

$$\Phi(x + c_7\tau) \geq F(x), \quad F(x + c_7\tau) \geq \Phi(x). \tag{49}$$

Indeed, according to (46)–(48), for $|x| \leq 2\sigma$

$$\Phi(x + c_7\tau) - F(x) \geq F(x + c_7\tau) - F(x) - c_5\frac{\tau}{\sigma} \geq \frac{c_6\,c_7\tau}{\sigma} - 3c_5\frac{\tau}{\sigma} \geq 0, \tag{50}$$

if we choose $c_7 = 3c_5/c_6$ and if $x + c_7\tau \leq 3\sigma$. The last inequality becomes obvious if $c_7\tau \leq \sigma$. For this, it is sufficient to choose $c_4 \leq c_7^{-1}$. The second inequality in (49) is verified similarly.

Thus, according to Lemma 5,

$$|\xi - \eta| < c_7\tau, \qquad \text{if} \quad |\xi| \leq 2\sigma. \tag{51}$$

Let $2\sigma \leq x \leq \sigma^2/5\tau$, and the parameter $h = h(x) \in \mathbf{R}$ be chosen in accordance with item c) of the one-dimensional version of Lemma 1 (whose condition $x \in \Pi$ is satisfied) and it is such that $\mathbf{E}\,\bar{\xi}(h) = x$, $\|h\|\tau \leq 1/2$, $\mathcal{L}(\bar{\xi}(h)) = \overline{F} = \overline{F}(h)$,

$$|\sigma h - x/\sigma| \leq 2.88\,\tau\sigma^{-1}x^2\sigma^{-2}. \tag{52}$$

According to relation (25) of Lemma 1,

$$\sigma^2(h) = \operatorname{Var}\bar{\xi}(h) = \sigma^2\big(1 + \theta\|h\|\tau\big). \tag{53}$$

Introduce the distribution $H = \Phi_{\overline{F}}$. Then, in accordance with item b) of Lemma 1, $\overline{F}(h) \in \mathcal{A}_d(2\tau)$ and, by Lemma 3 taking into account (53),

$$\rho(H, \overline{F}) \ll \frac{\tau}{\sigma(h)} \ll \frac{\tau}{\sigma}. \tag{54}$$

Integrating by parts, we obtain

$$1 - F(x) = \mathbf{E}\,e^{h\xi}\int_x^\infty e^{-hy}\overline{F}\{dy\} = \mathbf{E}\,e^{h\xi}\left(\int_x^\infty h\,e^{-hy}\overline{F}(y)\,dy - e^{-hx}\,\overline{F}(x)\right), \tag{55}$$

$$\int_x^\infty e^{-hy}H\{dy\} = \int_x^\infty h\,e^{-hy}H(y)\,dy - e^{-hx}\,H(x). \tag{56}$$

On the other hand, it is easy to verify that

$$\int_x^\infty e^{-hy}H\{dy\} = \frac{1}{\sqrt{2\pi}}e^{-hx}\,\Xi(h\sigma(h)). \tag{57}$$

From (54)–(57) it follows that

$$\left| \int_x^\infty e^{-hy}\overline{F}\{dy\} - \frac{1}{\sqrt{2\pi}}e^{-hx}\,\Xi(h\sigma(h)) \right|$$

$$\leq \left| \int_x^\infty h\,e^{-hy}(\overline{F}(y) - H(y))\,dy - e^{-hx}(\overline{F}(x) - H(x)) \right|$$

$$\leq 2\,e^{-hx}\,\rho(H,\overline{F}) \ll e^{-hx}\,\frac{\tau}{\sigma}. \quad (58)$$

Applying inequality (34) of Lemma 2, we obtain

$$\left| \Xi(h\sigma(h)) - \Xi(h\sigma) \right| \ll \frac{h|\sigma(h) - \sigma|}{h^2\sigma^2} \ll \frac{\tau}{\sigma}. \quad (59)$$

If $x \geq 2\sigma$, then $h\sigma \ll x/\sigma$ and $\Xi(h\sigma) \gg \Xi(x/\sigma) \gg \sigma/x$. Applying inequality (34) again, as well as (52), we obtain

$$\left| \Xi(h\sigma) - \Xi(x/\sigma) \right| \ll \frac{\left| h\sigma - x/\sigma \right|\sigma^2}{x^2} \ll \frac{\tau}{\sigma}. \quad (60)$$

Hence,

$$\left| \int_x^\infty e^{-hy}\overline{F}\{dy\} - \frac{1}{\sqrt{2\pi}}e^{-hx}\,\Xi(x/\sigma) \right| \ll \frac{\tau}{\sigma}\,e^{-hx}. \quad (61)$$

Applying the above inequalities, we obtain that

$$1 - F(x) = \mathbf{E}\,e^{h\xi}\left(\frac{1}{\sqrt{2\pi}}e^{-hx}\,\Xi(x/\sigma) + \int_x^\infty e^{-hy}\overline{F}\{dy\} - \frac{1}{\sqrt{2\pi}}e^{-hx}\,\Xi(x/\sigma) \right)$$

$$= \frac{1}{\sqrt{2\pi}}\mathbf{E}\,e^{h\xi - hx}\,\Xi(x/\sigma)\left(1 + \theta\,c\,\frac{\tau}{\sigma}\,\frac{x}{\sigma} \right)$$

$$= (1 - \Phi(x))\exp\left\{ \theta\,c_8\,\frac{\tau}{\sigma}\,\frac{x^3}{\sigma^3} \right\}$$

$$= \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2\sigma^2}\,\Xi(x/\sigma)\exp\left\{ \theta\,c_8\,\frac{\tau}{\sigma}\,\frac{x^3}{\sigma^3} \right\}. \quad (62)$$

Lemma 6. *There exist absolute positive constants $c_9,\ldots,c_{11}$ such that*

$$1 - \Phi(x + \beta(x)) \leq 1 - F(x) \leq 1 - \Phi(x - \beta(x)) \quad (63)$$

for $\tau/\sigma \leq c_9$, $2\sigma \leq x \leq z = c_{10}\sigma^2/\tau$, where

$$\beta(x) = c_{11}\tau\,x^2\sigma^{-2}. \quad (64)$$

We set $c_{11} = 4c_8$. Then, by choosing a sufficiently small c_{10}, we ensure that the inequality $\beta(x) \leq x$ is satisfied. Now applying inequalities (36) and (37) with $\varepsilon = \beta(x)$, we obtain

$$
\begin{aligned}
1 - F(x) &= (1 - \Phi(x)) \exp\left\{\theta\, c_8\, \frac{\tau}{\sigma}\, \frac{x^3}{\sigma^3}\right\} \\
&\leq (1 - \Phi(x - \beta(x))) \exp\left\{-\frac{(2x - \beta(x))\beta(x)}{2\sigma^2} + c_8\, \frac{\tau}{\sigma}\, \frac{x^3}{\sigma^3}\right\} \leq 1 - \Phi(x - \beta(x)), \quad (65)
\end{aligned}
$$

$$
\begin{aligned}
1 - F(x) &= (1 - \Phi(x)) \exp\left\{\theta\, c_8\, \frac{\tau}{\sigma}\, \frac{x^3}{\sigma^3}\right\} \\
&\geq (1 - \Phi(x + \beta(x))) \exp\left\{\frac{(2x + \beta(x))\beta(x)}{2\sigma^2} - c_8\, \frac{\tau}{\sigma}\, \frac{x^3}{\sigma^3}\right\} \geq 1 - \Phi(x + \beta(x)), \quad (66)
\end{aligned}
$$

completing the proof of the lemma.

Applying Lemma 6, we obtain that

$$
|\xi - \eta| < c_{11}\tau\xi^2/\sigma^2, \qquad \text{if} \quad 2\sigma \leq \xi \leq z = c_{10}\sigma^2/\tau, \quad \tau/\sigma \leq c_9. \qquad (67)
$$

To prove Theorem 1 it suffices to prove that the absolute constant c_{14} can be chosen so large that $\mathbf{E} \exp\left\{|\xi - \eta|/c_{14}\tau\right\} \ll 1$.

First, we will assume that $\tau/\sigma \leq c_9$. It is clear that

$$
\begin{aligned}
\exp\left\{|\xi - \eta|/c_{14}\tau\right\} &\leq \exp\left\{|\xi - \eta|/c_{14}\tau\right\} \mathbf{1}\left\{|\xi| \leq 2\sigma\right\} \\
&\quad + \exp\left\{|\xi - \eta|/c_{14}\tau\right\} \mathbf{1}\left\{2\sigma \leq |\xi| \leq z\right\} \\
&\quad\quad + \exp\left\{|\eta|/c_{14}\tau + |\xi|/c_{14}\tau\right\} \mathbf{1}\left\{|\xi| \geq z\right\}. \quad (68)
\end{aligned}
$$

According to (51), for $c_{14} > c_7$

$$
\mathbf{E} \exp\left\{|\xi - \eta|/c_{14}\tau\right\} \mathbf{1}\left\{|\xi| \leq 2\sigma\right\} \leq e.
$$

By Lemma 4,

$$
\mathbf{P}\left\{\xi^2 \geq x^2\right\} = \mathbf{P}\left\{|\xi| \geq x\right\} \leq 2 \max\left\{\exp\left\{-\frac{x^2}{4\,\sigma^2}\right\}, \exp\left\{-\frac{x^2}{4\,c_{10}\,\sigma^2}\right\}\right\} \qquad (69)
$$

for $0 \leq x \leq z = c_{10}\,\sigma^2/\tau$. Set

$$
W = c_{11}\xi^2/\sigma^2.
$$

Using (69), we obtain that there exists c_{13} such that

$$
\begin{aligned}
\mathbf{P}\left\{W \geq u\right\} &\leq \mathbf{P}\left\{|\xi| \geq \sigma\, \sqrt{u/c_{11}}\right\} \\
&\leq 2 \exp\left\{-u/c_{13}\right\}
\end{aligned}
$$

for $0 \leq u \leq \gamma = c_{11} z^2 / \sigma^2$. Set $c_{12} = 2 c_{13}$, $v = 1/c_{12}$. Integrating by parts, we obtain

$$
\begin{aligned}
\mathbf{E} \exp\{v\,W\}\, \mathbf{1}\{|\xi| \leq z\} &\leq 1 + \int_0^\gamma v\, e^{v\,u}\, \mathbf{P}\{W \geq u\}\, du \\
&\leq 1 + \frac{2}{c_{12}} \int_0^\gamma e^{u/c_{12}}\, \exp\{-u/c_{13}\}\, du \\
&\leq 1 + \frac{2}{c_{12}} \int_0^\infty e^{-u/c_{12}}\, du \ll 1,
\end{aligned}
$$

By (67), for $c_{14} > c_{12}$

$$
\exp\{|\xi - \eta|/c_{14}\tau\}\, \mathbf{1}\{2\sigma \leq |\xi| \leq z\} \leq \exp\{c_{11}\xi^2/c_{12}\sigma^2\}\, \mathbf{1}\{2\sigma \leq |\xi| \leq z\}.
$$

Therefore,

$$
\mathbf{E} \exp\{|\xi - \eta|/c_{14}\tau\}\, \mathbf{1}\{2\sigma \leq |\xi| \leq z\} \leq \mathbf{E} \exp\{v\,W\}\, \mathbf{1}\{|\xi| \leq z\} \ll 1.
$$

By the Cauchy–Bunyakovsky–Schwartz inequality, for $t \in \mathbf{R}$, $z > 0$ we have

$$
\mathbf{E} \exp\{t\,|\xi|+t\,|\eta|\}\, \mathbf{1}\{|\xi| \geq z\} \leq \left(\mathbf{E} \exp\{2t\,|\xi|\}\, \mathbf{1}\{|\xi| \geq z\} \cdot \mathbf{E} \exp\{2t\,|\eta|\}\, \mathbf{1}\{|\xi| \geq z\} \right)^{1/2},
$$
$$
\tag{70}
$$

and also

$$
\mathbf{E} \exp\{2t\,|\xi|\}\, \mathbf{1}\{|\xi| \geq z\} \leq \left(\mathbf{E} \exp\{4t\,|\xi|\} \cdot \mathbf{P}\{|\xi| \geq z\} \right)^{1/2}. \tag{71}
$$

It is clear that

$$
\exp\{4t\,|\xi|\} \leq \exp\{4t\,\xi\} + \exp\{-4t\,\xi\}. \tag{72}
$$

Applying Lemma 4, we obtain that for $z = c_{10}\,\sigma^2/\tau$

$$
\mathbf{P}\{|\xi| \geq z\} \leqslant 2 \exp\{-c\sigma^2/\tau^2\}. \tag{73}
$$

Let $0 \leq |h|\tau \leqslant \frac{1}{2}$. By (27),

$$
\mathbf{E}\, e^{h\xi} \leqslant \exp\{h^2\sigma^2\}. \tag{74}
$$

Applying (71)–(74) with $h = 4t = \pm 4/c_{14}\tau$, $z = c_{10}\,\sigma^2/\tau$ and choosing a sufficiently large constant c_{14}, we obtain

$$
\mathbf{E} \exp\{2t\,|\xi|\}\, \mathbf{1}\{|\xi| \geq z\} \leq \left(\mathbf{E}\left(\exp\{4t\,\xi\} + \exp\{-4t\,\xi\}\right) \cdot \mathbf{P}\{|\xi| \geq z\} \right)^{1/2} \leq \sqrt{2}. \tag{75}
$$

It is similarly verified that

$$
\mathbf{E} \exp\{2t\,|\eta|\}\, \mathbf{1}\{|\xi| \geq z\} \leq \sqrt{2}. \tag{76}
$$

Hence,

$$
\mathbf{E} \exp\{|\eta|/c_{14}\tau + |\xi|/c_{14}\tau\}\, \mathbf{1}\{|\xi| \geq c_{10}\,\sigma^2/\tau\} \leq 2\sqrt{2}. \tag{77}
$$

Let now $\tau > c_9\sigma$. Then

$$
\mathbf{E} \exp\{|\xi - \eta|/c_{14}\tau\} \leq \left(\mathbf{E} \exp\{2\,|\eta|/c_{14}\tau\} \cdot \mathbf{E} \exp\{2\,|\xi|/c_{14}\tau\} \right)^{1/2}. \tag{78}
$$

$$
\mathbf{E} \exp\{2\,|\xi|/c_{14}\tau\} \leq \mathbf{E} \exp\{2\,\xi/c_{14}\tau\} + \mathbf{E} \exp\{-2\,\xi/c_{14}\tau\}. \tag{79}
$$

Applying (74) with $h = \pm 2/c_{14}\tau$, and choosing a sufficiently large constant c_{14}, we obtain that

$$\mathbf{E} \exp \left\{ |\xi - \eta|/c_{14}\tau \right\} \ll 1, \tag{80}$$

completing the proof of the theorem.

References

[1] Arak, T.V., Zaitsev, A.Yu.: Uniform limit theorems for sums of independent random variables. Trudy MIAN, **174**, 3–214 (1986) (in Russian). English translation in: Proc. Steklov Inst. Math., **174**, 1–222 (1988).

[2] Bennett G., *Probability inequalities for the sum of independent random variables.* — J. Amer. Statist. Assoc. **57** (1962), 33–45.

[3] Bentkus V., *A remark on Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand inequalities.* — Lith. Math. J. **42**, No. 3, 262–269 (2002).

[4] Bentkus V., *On Hoeffding's inequalities.* — Ann. Probab. **32**, No. 2 (2004), 1650–1673.

[5] Berkes I., Philipp W., *Approximation theorems for independent and weakly dependent random vectors.* — Ann. Probab. **7**, (1979), 29–54.

[6] Bernstein S., Probability Theory, Moscow, Gostekhizdat., 1946 (in Russian).

[7] Bikelis A., *Estimates of the remainder term in the central limit theorem.* — Lit. Matem. Sb. **6**, No. 3 (1966), 323–346 (in Russian).

[8] Bobkov S., Götze F., *Esscher transform and the central limit theorem.* — J. Funct. Anal. **289**, No. 5, Article ID 110999, 39 p. (2025).

[9] Bogachev V. I., Kolesnikov A. V., *The Monge–Kantorovich problem: achievements, connections, and perspectives.* — Russian Math. Surv. **67**, No. 5 (2012), 785–890.

[10] Bogachev V. I., *Kantorovich problem of optimal transportation of measures: new directions of research.* — Russian Math. Surv. **77**, No. 5 (2022), 769–817.

[11] Borovkov A. A., Mogul skij A. A., Sakhanenko A. I., *Limit theorems for random processes.* Probability theory – 7, Itogi Nauki Tekh., Ser. Sovrem. Probl. Mat., Fundam. Napravleniya, **82**, Moskva: VINITI, 1995, 5–194 (in Russian).'

[12] Borovkov A. A., Sakhanenko A. I., *On the rate of convergence in invariance principle.* — Lect. Notes Math. **1021** (1981), 59–66.

[13] Döring H., Jansen S., Schubert K., *The method of cumulants for the normal approximation.* — Probab. Surv. **19** (2022), 185–270.

[14] Dudley R. M., *Distances of probability measures and random variables.* — Ann. Math. Statist. **39**, No. 5 (1968), 1563–1572.

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

[16] Kolmogorov A. N., *Two uniform limit theorems for sums of independent random variables.* — Theory Probab. Appl. **1**, No. 4 (1956), 384–394.

[17] Kolmogorov A. N., *Approximation to distributions of sums of independent terms by means of infinitely divisible distributions.* — Trans. Mosc. Math. Soc. **12** (1963), 492–509.

[18] Komlós J., Major P., Tusnády G., *An approximation of partial sums of independent RV's and the sample DF, I; II.* — Z. Wahrscheinlichkeitstheor. verw. Geb., **32** (1975), 111–131; **34** (1976), 34–58.

[19] Nagaev S. V., *Large deviations of sums of independent random variables.* — Ann. Probab. **7** (1979), 745–789.

[20] Petrov V. V., *Limit theorems for the sums of independent random variables*, Moscow, Nauka, 1987 (in Russian).

[21] Pinelis I., *On the Bennett–Hoeffding inequality.* — Ann. Inst. Henri Poincaré, Probab. Statist. **50**, No. 1 (2014), 15–27.

[22] Prokhorov Yu. V., *Convergence of random processes and limit theorems in probability theory.* — Theory Probab. Appl. **1** (1956), 157–214.

[23] Rio E., *Upper bounds for minimal distances in the central limit theorem.* — Ann. Inst. Henri Poincaré, Probab. Statist. **45**, No. 3 (2009), 802–817.

[24] Rudzkis R., *Probabilities of large deviations of random vectors.* — Lith. Math. J. **23** (1983), 113–120.

[25] Saulis L., *Large deviations for random vectors for certain classes of sets. I.,* — Lith. Math. J. **23** (1983), 308–317.

[26] Saulis L., Statulevičius V. A., *Limit theorems on large deviations.* Vilnius, Mokslas, 1989 (in Russian).

[27] Saulis L., Statulevičius V. A., *Limit theorems on large deviations.* — Probability theory – 6, Itogi Nauki Tekh., Ser. Sovrem. Probl. Mat., Fundam. Napravleniya,, **81** (1991), 219–312 (in Russian).

[28] Sakhanenko A. I., *Convergence rate in the invariance principle for non-identically distributed variables with exponential moments.* — Advances in probability theory: Limit theorems for sums of random variables, Transl. Ser. Math. Eng., 2–73, 1986.

[29] Sakhanenko A. I., *Estimates for the accuracy of coupling in the central limit theorem*, Siberian Math. J., **37**, No. 4 (1996), 811–823.

[30] Schay G., *Nearest random variables with given distributions.* — Ann. Probab. **2** (1974), 163–166.

[31] Statulevičius V. A., *On large deviations.* — Z. Wahrscheinlichkeitstheor. verw. Geb. **62** (1966), 133–144.

[32] Strassen V., *The existence of probability measures with given marginals.* — Ann. Math. Statist. **36** (1965), 423–439.

[33] Vallander S. S., *Calculation of the Wasserstein distance between probability distributions on the line.* — Theory Probab. Appl. **18**, No. 4 (1973), 784–786.

[34] Yurinskii V. V., *Exponential inequalities for sums of random vectors.* — J. Multiv. Anal. **6** (1976), 473–499.

[35] Zaitsev A. Yu., *On the Gaussian approximation of convolutions under multidimensional analogues of Bernstein's inequality conditions.* — Preprint P-9-84, Leningrad, LOMI, 1984, 54 pp. (in Russian).

[36] Zaitsev A. Yu., *Estimates for the Lévy–Prokhorov distance in the multivariate central limit theorem for random vectors with finite exponential moments.* — Theory Probab. Appl. **31** (1986), 203–220.

[37] Zaitsev A. Yu., *On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein inequality conditions.* — Probab. Theor. Rel. Fields **74** (1987), 535–566.

[38] Zaitsev A. Yu., *On a connection between two classes of probability distributions.* — In: Rings and modules. Limit theorems of probability theory. **2**, Leningrad, LGU Press, 1988, 153–158.

[39] Zaitsev A. Yu., *Multivariate version of the second Kolmogorov's uniform limit theorem.* — Theory Probab. Appl. **34**, No. 1 (1989), 108–128.

[40] Zaitsev A. Yu., *Estimates for the quantiles of smooth conditional distributions and the multidimensional invariance principle.* — Siberian Math. J. **37**, No. 4 (1996), 706–729.

[41] Zaitsev A. Yu., *Multidimensional version of the results of Komlós, Major, and Tusnády for vectors with finite exponential moments.* — ESAIM: Probab. Statist. **2** (1998), 41–108.

[42] Zaitsev A. Yu., *Multidimensional version of the results of Sakhanenko in the invariance principle for vectors with finite exponential moments, I; II; III.* — Theory Probab. Appl. **45** (2001), 624–641; **46** (2001/02), 490–514; 676–698.

[43] Zaitsev A. Yu., *The accuracy of strong Gaussian approximation for sums of independent random vectors.* — Russian Math. Surv. **68**, No. 4 (2013), 721–761.

Keywords: inequalities, sums of independent random vectors, transport distance estimates, central limit theorem

St. Petersburg Branch of the Steklov Mathematical Institute, Fontanka 27, St. Petersburg 191023, Russia, and St. Petersburg State University,, Universitetskaya Embankment 7/9, St. Petersburg, 199034 Russia

Email address: zaitsev@pdmi.ras.ru