

Some Approximation Problems in Statistics and Probability

Yuri V. Prokhorov and Vladimir V. Ulyanov

Dedicated to Friedrich Götze on the occasion of his sixtieth birthday

Abstract We review the results about the accuracy of approximations for distributions of functionals of sums of independent random elements with values in a Hilbert space. Mainly we consider recent results for quadratic and almost quadratic forms motivated by asymptotic problems in mathematical statistics. Some of the results are optimal and could not be further improved without additional conditions.

Keywords Accuracy of approximations • Eigenvalues of covariance operator • Hilbert space • Power divergence family of statistics • Quadratic forms of random elements • Short asymptotic expansions

2000 *Mathematics Subject Classification*. Primary 62E20, 62H10; Secondary 52A20

Y.V. Prokhorov
Steklov Mathematical Institute, Russian Academy of Sciences, Moscow 119991, Russia
e-mail: tvp@mi.ras.ru

V.V. Ulyanov (✉)
Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics,
Moscow State University, Moscow 119991, Russia
e-mail: vulyan@gmail.com

1 Quadratic Forms

All three gems in probability theory—the law of large numbers, the central limit theorem and the law of the iterated logarithm—concern the asymptotic behavior of the sums of random variables. It would be natural to extend the results to functionals of the sums, in particular to quadratic forms. Moreover, in mathematical statistics there are numerous asymptotic problems which can be formulated in terms of quadratic or almost quadratic forms. In this article we review the corresponding results with rates of convergence. Some of these results are optimal and could not be further improved without additional conditions. The review does not pretend to completely illuminate the present state of the area under consideration. It reflects mainly the authors interests.

Let X, X_1, X_2, \dots be independent identically distributed random elements with values in a real separable Hilbert space H . The dimension of H , say $\dim(H)$, could be either infinite or finite. Let (x, y) for $x, y \in H$ denote the inner product in H and put $|x| = (x, x)^{1/2}$. We assume that $\mathbf{E}|X_1|^2 < \infty$ and denote by V the covariance operator of X_1 :

$$(Vx, y) = \mathbf{E}(X_1 - \mathbf{E}X_1, x)(X_1 - \mathbf{E}X_1, y).$$

Let $\sigma_1^2 \geq \sigma_2^2 \geq \dots$ be the eigenvalues of V and let e_1, e_2, \dots be the corresponding eigenvectors which we assume to be orthonormal.

For any integer $k > 0$ we put

$$c_k(V) = \prod_1^k \sigma_i^{-1}, \quad \bar{c}_k(V) = \left(\prod_1^k \sigma_i^{-1}\right)^{(k-1)/k}. \tag{1}$$

In what follows we use c and $c(\cdot)$, with or without indices, to denote the absolute constants and the constants depending on parameters in brackets. Except for $c_i(V)$ and $\bar{c}_i(V)$ the same symbol may be used for various constants.

We define

$$S_n = n^{-1/2} \sigma^{-1} \sum_{i=1}^n (X_i - \mathbf{E}X_i),$$

where $\sigma^2 = \mathbf{E}|X_1 - \mathbf{E}X_1|^2$. Without loss of generality we may assume that $\mathbf{E}X_1 = 0$ and $\mathbf{E}|X_1|^2 = 1$. The general case can be reduced to this one considering $(X_i - \mathbf{E}X_i)/\sigma$ instead of $X_i, i = 1, 2, \dots$. Let Y be H -valued Gaussian $(0, V)$ random element. We denote the distributions of S_n and Y by P_n and Q respectively.

The central limit theorem asserts that

$$P_n(B) - Q(B) \rightarrow 0$$

for any Borel set B in H provided $Q(\partial B) = 0$, where ∂B is the boundary of B . The estimate of the rate of convergence in the central limit theorem is an estimate of the quantity $\sup_{\mathcal{A}} |P_n(A) - Q(A)|$ for various classes \mathcal{A} of measurable sets A .

The most famous is the Berry-Esseen bound (see [5, 9]) when $H = \mathbf{R}$, i.e. $\dim(H) = 1$, and $\mathcal{A} = \mathcal{A}_1 = \{(-\infty, x), x \in \mathbb{R}\}$:

$$\sup_{\mathcal{A}_1} |P_n(A) - Q(A)| \leq c \frac{\mathbf{E}|X_1|^3}{\sqrt{n}}. \tag{2}$$

The bound is optimal with respect to dependence on n and moments of X_1 . The lower bound for the constant c in (2) is known (see [11]):

$$c \geq \frac{3 + \sqrt{10}}{6\sqrt{2\pi}} = 0.40\dots$$

The present upper bounds for $c : c \leq 0.47\dots$ (see [41, 43]) still differ from the lower bound slightly.

In the multidimensional case when $H = \mathbf{R}^d$, i.e. $\dim(H) = d > 1$, it is possible to extend the class \mathcal{A} to the class of all convex Borel sets in H and to get a bound (see e.g. [2, 33])

$$\sup_{\mathcal{A}} |P_n(A) - Q(A)| \leq c \sqrt{d} \frac{\mathbf{E}|X_1|^3}{\sigma_d^3 \sqrt{n}}.$$

If we consider an infinite dimensional space H and take \mathcal{A} as the class of all half-spaces in H then one can show (see e.g. pp. 69–70 in [34]) that there exists a distribution of X_1 such that

$$\sup_{\mathcal{A}} |P_n(A) - Q(A)| \geq 1/2. \tag{3}$$

Therefore, in the infinite dimensional case we can construct upper bound for $\sup_{\mathcal{A}} |P_n(A) - Q(A)|$ provided that \mathcal{A} is a relatively narrow class, e.g. the class of all balls $B(a, x) = \{y : y \in H \text{ and } |y - a|^2 \leq x\}$ with fixed center a or the class of all balls with fixed bounded radius \sqrt{x} . However, the good news are that the numerous asymptotic problems in statistics can be reformulated in terms of these or similar classes (see e.g. Sect. 2).

Put for any $a \in H$

$$F(x) = P_n(B(a, x)), \quad F_0(x) = Q(B(a, x)), \quad \delta_n(a) = \sup_x |F(x) - F_0(x)|.$$

According to (3) it is impossible to prove upper bound for $\sup_a \delta_n(a)$ which tends to 0 as $|a| \rightarrow \infty$. The upper bound for $\delta_n(a)$ should depend on a and becomes in general bad as $|a|$ grows.

The history of constructing bounds for $\delta_n(a)$ in the infinite dimensional case can be divided roughly into three phases: proving bounds with optimal

- Dependence on n ;
- Moment conditions;
- Dependence on the eigenvalues of V .

The first phase started in the middle of 1960s in the twentieth century with bounds of logarithmic order for $\delta_n(a)$ (see [27]) and ended with the result:

$$\delta_n(a) = \mathcal{O}(n^{-1/2}),$$

due to Götze [12], which was based on a Weyl type symmetrization inequality (see Lemma 3.37 (i) in [12]):

Let X, Y, Z be the independent random elements in H . Then

$$\mathbf{E} \exp\{i \tau |X + Y + Z|^2\} \leq (\mathbf{E} \exp\{2i \tau (\widetilde{X}, \widetilde{Y})\})^{1/4}, \tag{4}$$

where \widetilde{X} is the symmetrization of X , i.e $\widetilde{X} = X - X'$ with independent and identically distributed X and X' . The main point of the inequality is that it enables us to reduce the initial problem with non-linear dependence on X in power of \exp to linear one. The inequality since then has been successfully applied and developed by a number of the authors.

The second phase of the history finished with a paper by Yurinskii [48] who proved

$$\delta_n(a) \leq \frac{c(V)}{\sqrt{n}} (1 + |a|^3) \mathbf{E}|X_1|^3,$$

where $c(V)$ denotes a constant depending on V only. The Yurinskii result has the optimal dependence on n under minimal moment condition but dependence of $c(V)$ on characteristics of the operator V was still unsatisfactory.

At the end of the third phase it was proved (see [28, 36, 39])

$$\delta_n(a) \leq \frac{c c_6(V)}{\sqrt{n}} (1 + |a|^3) \mathbf{E}|X_1|^3, \tag{5}$$

where $c_6(V)$ is defined in (1). It is known (see Example 3 in [38]) that for any $c_0 > 0$ and for any given eigenvalues $\sigma_1^2, \dots, \sigma_6^2 > 0$ of a covariance operator V there exist a vector $a \in H = \mathbf{R}^7$, $|a| > c_0$, and a sequence X_1, X_2, \dots of i.i.d. random elements in $H = \mathbf{R}^7$ with zero mean and covariance operator V such that

$$\liminf_{n \rightarrow \infty} \sqrt{n} \delta_n(a) \geq c c_6(V) (1 + |a|^3) \mathbf{E}|X_1|^3. \tag{6}$$

Due to (6) the bound (5) is the best possible in case of the finite third moment of $|X_1|$. For further refinements see e.g. [40]. For the results for the case of non-identically distributed random elements in H see [44].

At the same time better approximations for $F(x)$ are available when we use for approximation an additional term, say $F_1(x)$, of its asymptotic expansion. This term $F_1(x)$ is defined as the unique function satisfying $F_1(-\infty) = 0$ with the Fourier-Stieltjes transform equal to

$$\begin{aligned} \hat{F}_1(t) = & -\frac{2t^2}{3\sqrt{n}} \mathbf{E} e\{t|Y - a|^2\} (3(X, Y - a)|X|^2 \\ & + 2it(X, Y - a)^3). \end{aligned} \tag{7}$$

Here and in the following X and Y are independent and we write $e\{x\} = \exp\{ix\}$.

In case $\dim(H) < \infty$ the term $F_1(x)$ can be defined in terms of the density function of the normal distribution (see [6]). Let φ denote the standard normal density in \mathbf{R}^d . Then the density function $p(y)$ of the normal distribution Q is defined by $p(y) = \varphi(V^{-1/2}y)/\sqrt{\det V}$, $y \in \mathbf{R}^d$. We have

$$F_1(x) = \frac{1}{6\sqrt{n}} \chi(A_x), \quad A_x = \{u \in \mathbf{R}^d : |u - a|^2 \leq x\}$$

with the signed measure

$$\chi(A) = \int_A \mathbf{E} p'''(y) X^3 dy \quad \text{for the Borel sets } A \subset \mathbf{R}^d$$

and

$$p'''(y) u^3 = p(y)(3(V^{-1}u, u)(V^{-1}y, u) - (V^{-1}y, u)^3)$$

is the third Frechet derivative of p in the direction u .

Introduce the error

$$\Delta_n(a) = \sup_x |F(x) - F_0(x) - F_1(x)|.$$

Note, that $\hat{F}_1(t) = 0$ and hence $F_1(x) = 0$ when $a = 0$ or X has a symmetric distribution, i.e. when X and $-X$ are identically distributed. Therefore, we get

$$\Delta_n(0) = \delta_n(0).$$

Similar to the developments of the bounds for $\delta_n(a)$ the first task consisted in deriving the bounds for $\Delta_n(a)$ with the optimal dependence on n . Starting with a seminal paper by Esseen [10] for the finite dimensional spaces $H = \mathbf{R}^d$, $d < \infty$, who proved

$$\Delta_n(0) = \mathcal{O}(n^{-d/(d+1)}), \tag{8}$$

a comparable bound

$$\Delta_n(0) = \mathcal{O}(n^{-\gamma})$$

with $\gamma = 1 - \varepsilon$ for any $\varepsilon > 0$ was finally proved in [12, 13], based on the Weyl type inequalities mentioned above. Further refinements and generalizations in the case $a \neq 0$ and $\gamma < 1$ are due to Nagaev and Chebotarev [29], Sazonov et al. [35].

Note however, that the results in the infinite dimensional case did not even yield (8) as corollary when $\sigma_{d+1} = 0$, i.e. $\dim(H) = d$. Only 50 years after Esseen's result the optimal bounds (in n) were finally established in [3]

$$\Delta_n(0) \leq \frac{c(9, V)}{n} \mathbf{E}|X_1|^4, \tag{9}$$

$$\Delta_n(a) \leq \frac{c(13, V)}{n} (1 + |a|^6) \mathbf{E}|X_1|^4, \tag{10}$$

where $c(i, V) \leq \exp\{c\sigma_i^{-2}\}$, $i = 9, 13$, and in the case of the bound (9) it was additionally assumed that the distribution of X_1 is symmetric. In order to derive these bound new techniques were developed, in particular the so-called multiplicative inequality for the characteristic functions (see Lemma 3.2, Theorem 10.1 and formulas (10.7)–(10.8) in [4]):

Let $\varphi(t)$, $t \geq 0$, denote a continuous function such that $0 \leq \varphi \leq 1$. Assume that

$$\varphi(t) \varphi(t + \tau) \leq \theta \mathcal{M}^d(\tau, N)$$

for all $t \geq 0$ and $\tau > 0$ with some $\theta \geq 1$ independent of t and τ , where

$$\mathcal{M}(t, n) = 1/\sqrt{|t|n} + \sqrt{|t|} \text{ for } |t| > 0.$$

Then for any $0 < B \leq 1$ and $N \geq 1$

$$\int_{B/\sqrt{N}}^1 \frac{\varphi(t)}{t} dt \leq c(s) \theta (N^{-1} + (B\sqrt{N})^{-d/2}) \text{ for } d > 8. \tag{11}$$

The previous Weyl type inequality (4) gave the bounds for the integrals

$$\int_{D(n, \gamma)} \frac{\mathbf{E}e\{t|S_n - a|^2\}}{|t|} dt$$

for the areas $D(n, \gamma) = \{t : n^{1/2} < |t| \leq n^\gamma\}$ with $\gamma < 1$ only, while (11) enables to extend the areas of integration up to $\gamma = 1$.

The bounds (9) and (10) are optimal with respect to the dependence on n [14] and on the moments. The bound (9) improves as well Esseen’s result (8) for the Euclidean spaces \mathbf{R}^d with $d > 8$. However, the dependence on covariance operator V in (9), (10) could be improved. Nagaev and Chebotarev [30] considered the case $a = 0$ and got a bound of type (9) replacing $c(9, V)$ by the following function $c(V)$:

$$c(V) = c (\bar{c}_{13}(V) + (c_9(V))^{4/9} \sigma_9^{-6}),$$

where $\bar{c}_{13}(V)$ and $c_9(V)$ are defined by (1). The general case $a \neq 0$ was considered in [31] (see their Theorem 1.2). The Nagaev and Chebotarev results improve the dependence on the eigenvalues of V (compared to (10)) but still require that $\sigma_{13} > 0$ instead of the weaker condition $\sigma_9 > 0$ in (9). However, it follows from Lemma 2.6 in [17] that for any given eigenvalues $\sigma_1^2, \dots, \sigma_{12}^2 > 0$ of a covariance operator V

there exist $a \in H = \mathbf{R}^{13}$, $|a| > 1$, and a sequence X_1, X_2, \dots of i.i.d. random elements in $H = \mathbf{R}^{13}$ with zero mean and covariance operator V such that

$$\liminf_{n \rightarrow \infty} n \Delta_n(a) \geq c c_{12}(V) (1 + |a|^6) \mathbf{E}|X_1|^4. \tag{12}$$

The bound with dependence on 12 largest eigenvalues of the operator V was obtained only in [46] (for the first version see Corollary 1.3 in [17]). Moreover, in [46] the dependence on the eigenvalues is given in the bound in the explicit form which coincides with the form given by the lower bound (12):

Theorem 1.1. *There exists an absolute constant c such that for any $a \in H$*

$$\begin{aligned} \Delta_n(a) \leq \frac{c}{n} \cdot c_{12}(V) \cdot (\mathbf{E}|X_1|^4 + \mathbf{E}(X_1, a)^4) \\ \times (1 + (Va, a)), \end{aligned} \tag{13}$$

where $c_{12}(V)$ is defined in (1).

According to the lower bound (12) the estimate (13) is the best possible in the following sense:

- It is impossible that $\Delta_n(a)$ is of order $\mathcal{O}(n^{-1})$ uniformly for all distributions of X_1 with arbitrary eigenvalues $\sigma_1^2, \sigma_2^2, \dots$;
- The form of the dependence of the right-hand side in (13) on the eigenvalues of V , on n and on $\mathbf{E}|X_1|^4$ coincides with one given in the lower bound.

For earlier versions of this result on the optimality of 12 eigenvalues and a detailed discussion of the connection of the rate problems in the central limit theorem with classical lattice point problems in analytic number theory, see the ICM-1998 Proceedings paper by Götze [14], and also Götze and Ulyanov [17].

Note however, that in the special ‘symmetric’ cases of the distribution of X_1 or of the center, say a , of the ball, the number of the eigenvalues which are necessary for optimal bounds may well decrease below 12. For example, when $\mathbf{E}(X, b)^3 = 0$ for all $b \in H$, by Corollary 2.7 in [17], for any given eigenvalues $\sigma_1^2, \dots, \sigma_8^2 > 0$ of a covariance operator V there exists a center $a \in H = \mathbf{R}^9$, $|a| > 1$, and a sequence X, X_1, X_2, \dots of i.i.d. random elements in $H = \mathbf{R}^9$ with zero mean and the covariance operator V such that

$$\liminf_{n \rightarrow \infty} n \Delta_n(a) \geq c c_8(V) (1 + |a|^4) \mathbf{E}|X_1|^4.$$

Hence, in this case an upper bound of order $\mathcal{O}(n^{-1})$ for $\Delta_n(a)$ has to involve at least the *eight* largest eigenvalues of V .

Furthermore, lower bounds for $n\Delta_n(a)$ in the case $a = 0$ are not available. A conjecture, see [14], said that in that case the five first eigenvalues of V suffice. That conjecture was confirmed in Theorem 1.1 in [19] with result $\Delta_n(0) = \mathcal{O}(n^{-1})$ provided that $\sigma_5 > 0$ only. Note that for some centered ellipsoids in \mathbf{R}^d with $d \geq 5$

the bounds of order $\mathcal{O}(n^{-1})$ were obtained in [18]. Moreover, it was proved recently (see Corollary 2.4 in [20]) that even for $a \neq 0$ we have $\Delta_n(a) = \mathcal{O}(n^{-1})$ when $H = \mathbf{R}^d$, $5 \leq d < \infty$, and the upper bound for $\Delta_n(a)$ is written in the explicit form and depends on the smallest eigenvalue σ_d (see Theorem 1.4 in [21] as well). It is necessary to emphasize that (13) implies $\Delta_n(a) = \mathcal{O}(n^{-1})$ for general *infinite* dimensional space H with dependence on the first twelve eigenvalues of V only.

The proofs of the recent results due to Götze, Ulyanov and Zaitsev are based on the reduction of the original problem to lattice valued random vectors and on the symmetrization techniques developed in a number of papers, see e.g. Götze [12], Yurinskii [48], Sazonov et al. [35–37], Götze and Ulyanov [17], Bogatyrev et al. [7]. In the proofs we use also the new inequalities obtained in Lemma 6.5 in [20] and in [16] (see Lemma 8.2 in [20]). In fact, the bounds in [20] are constructed for more general quadratic forms of the type $(\mathbb{Q}x, x)$ with non-degenerate linear symmetric bounded operator in \mathbf{R}^d .

One of the basic lemma to prove (13) is the following (see Lemma 2.2 in [17]):

Let $T > 0$, $b \in \mathbf{R}^1$, $b \neq 0$, l be an integer, $l \geq 1$, $Y = (Y_1, \dots, Y_{2l})$ be a Gaussian random vector with values in \mathbf{R}^{2l} ; Y_1, \dots, Y_{2l} be independent and $\mathbf{E}Y_i = 0$, $\mathbf{E}Y_i^2 = \sigma_i^2$ for $i = 1, 2, \dots, 2l$; $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_{2l}^2 > 0$ and $a \in \mathbf{R}^{2l}$. Then there exists a positive constant $c = c(l)$ such that

$$\left| \int_{-T}^T s^{l-1} \mathbf{E} \exp\{is|Y + a|^2\} e^{ibs} ds \right| \leq c \prod_{j=1}^{2l} \sigma_j^{-1}.$$

For non-uniform bounds with 12 eigenvalues of covariance operator V see [7].

For estimates for the characteristic functions of polynomials (of order higher than 2) of asymptotically normal random variables see [22], for related results see also [23].

2 Applications in Statistics: Almost Quadratic Forms

In this section we consider the accuracy of approximations for the distributions of sums of independent random elements in $k - 1$ -dimensional Euclidian space. The approximation is considered on the class of sets which are “similar” to ellipsoids. Its appearance is motivated by the study of the asymptotic behavior of the goodness-of-fit test statistics—power divergence family of statistics.

Consider a vector $(Y_1, \dots, Y_k)^T$ with multinomial distribution $M_k(n, \pi)$, i. e.

$$\Pr(Y_1 = n_1, \dots, Y_k = n_k) = \begin{cases} n! \prod_{j=1}^k (\pi_j^{n_j} / n_j!), & n_j = 0, 1, \dots, n \ (j = 1, \dots, k) \\ & \text{and } \sum_{j=1}^k n_j = n, \\ 0, & \text{otherwise,} \end{cases}$$

where $\pi = (\pi_1, \dots, \pi_k)^T, \pi_j > 0, \sum_{j=1}^k \pi_j = 1$. From this point on, we will assume the validity of the hypothesis $H_0: \pi = \mathbf{p}$. Since the sum of n_i equals n , we can express this multinomial distribution in terms of a vector $\mathbf{Y} = (Y_1, \dots, Y_{k-1})$ and denote its covariance matrix Ω . It is known that so defined Ω equals $(\delta_i^j p_i - p_i p_j) \in \mathbf{R}^{(k-1) \times (k-1)}$. The main object of the current study is the power divergence family of goodness-of-fit test statistics:

$$t_\lambda(\mathbf{Y}) = \frac{2}{\lambda(\lambda + 1)} \sum_{j=1}^k Y_j \left[\left(\frac{Y_j}{np_j} \right)^\lambda - 1 \right], \lambda \in \mathbf{R},$$

When $\lambda = 0, -1$, this notation should be understood as a result of passage to the limit.

These statistics were first introduced in [8] and [32]. Putting $\lambda = 1, \lambda = -1/2$ and $\lambda = 0$ we can obtain the chi-squared statistic, the Freeman-Tukey statistic, and the log-likelihood ratio statistic respectively.

We consider transformation

$$X_j = (Y_j - np_j) / \sqrt{n}, j = 1, \dots, k, r = k - 1, \mathbf{X} = (X_1, \dots, X_r)^T.$$

Herein the vector \mathbf{X} is the vector taking values on the lattice,

$$L = \left\{ \mathbf{x} = (x_1, \dots, x_r)^T; \mathbf{x} = \frac{\mathbf{m} - n\mathbf{p}}{\sqrt{n}}, \mathbf{p} = (p_1, \dots, p_r)^T, \mathbf{m} = (n_1, \dots, n_r)^T \right\},$$

where n_j are non-negative integers.

The statistic $t_\lambda(\mathbf{Y})$ can be expressed as a function of \mathbf{X} in the form

$$T_\lambda(\mathbf{X}) = \frac{2n}{\lambda(\lambda + 1)} \left[\sum_{j=1}^k p_j \left(\left(1 + \frac{X_j}{\sqrt{np_j}} \right)^{\lambda+1} - 1 \right) \right], \tag{14}$$

and then, via the Taylor expansion, transformed to the form

$$T_\lambda(\mathbf{X}) = \sum_{i=1}^k \left(\frac{X_i^2}{p_i} + \frac{(\lambda - 1)X_i^3}{3\sqrt{np_i^2}} + \frac{(\lambda - 1)(\lambda - 2)X_i^4}{12p_i^3 n} + O(n^{-3/2}) \right).$$

As we see the statistics $T_\lambda(\mathbf{X})$ is “close” to quadratic form

$$T_1(\mathbf{X}) = \sum_{i=1}^k \frac{X_i^2}{p_i},$$

considered in Sect. 1.

We call a set $B \subset \mathbf{R}^r$ an extended convex set, if for all $l = 1, \dots, r$ it can be expressed in the form:

$$B = \{ \mathbf{x} = (x_1, \dots, x_r)^T : \lambda_l(x^*) < x_l < \theta_l(x^*) \text{ and } x^* = (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_r)^T \in B_l \},$$

where B_l is some subset of \mathbf{R}^{r-1} and $\lambda_l(x^*), \theta_l(x^*)$ are continuous functions on \mathbf{R}^{r-1} . Additionally, we introduce the following notation

$$[h(\mathbf{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} = h(x_1, \dots, x_{l-1}, \theta_l(x^*), x_{l+1}, \dots, x_r) - h(x_1, \dots, x_{l-1}, \lambda_l(x^*), x_{l+1}, \dots, x_r).$$

It is known that the distributions of all statistics in the family converge to chi-squared distribution with $k - 1$ degrees of freedom (see e.g. [8], p. 443). However, more intriguing is the problem to find the rate of convergence to the limiting distribution.

For any bounded extended convex set B in [47] it was obtained an asymptotic expansion, which in [42] was converted to

$$\Pr(X \in B) = J_1 + J_2 + O(n^{-1}). \tag{15}$$

with

$$J_1 = \int \dots \int_B \phi(\mathbf{x}) \left\{ 1 + \frac{1}{\sqrt{n}} h_1(\mathbf{x}) + \frac{1}{n} h_2(\mathbf{x}) \right\} dx, \text{ where}$$

$$h_1(\mathbf{x}) = -\frac{1}{2} \sum_{j=1}^k \frac{x_j}{p_j} + \frac{1}{6} \sum_{j=1}^k x_j \left(\frac{x_j}{p_j} \right)^2,$$

$$h_2(\mathbf{x}) = \frac{1}{2} h_1(\mathbf{x})^2 + \frac{1}{12} \left(1 - \sum_{j=1}^k \frac{1}{p_j} \right) + \frac{1}{4} \sum_{j=1}^k \left(\frac{x_j}{p_j} \right)^2 - \frac{1}{12} \sum_{j=1}^k x_j \left(\frac{x_j}{p_j} \right)^3;$$

$$J_2 = -\frac{1}{\sqrt{n}} \sum_{l=1}^r n^{-(r-l)/2} \sum_{x_{l+1} \in L_{l+1}} \dots \sum_{x_r \in L_r} \left[\int \dots \int_{B_l} [S_1(\sqrt{n}x_l + np_l)\phi(\mathbf{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} dx_1, \dots, dx_{l-1} \right]; \tag{16}$$

$$L_j = \left\{ \mathbf{x} : x_j = \frac{n_j - np_j}{\sqrt{n}}, n_j \text{ and } p_j \text{ defined as before} \right\};$$

$$S_1(x) = x - [x] - 1/2, [x] \text{ is the integer part of } x;$$

$$\phi(\mathbf{x}) = \frac{1}{(2\pi)^{r/2} |\Omega|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T \Omega^{-1} \mathbf{x}\right).$$

In [47] it was shown that $J_2 = O(n^{-1/2})$.

Using elementary transformations it can be easily shown that the determinant of the matrix Ω equals $\prod_{i=1}^k p_i$.

In [47] it was also examined the expansion for the most known power divergence statistic, which is the chi-squared statistic. Put $B^\lambda = \{\mathbf{x} \mid T_\lambda(\mathbf{x}) < c\}$. It is easy to show that B^1 is an ellipsoid, which is a particular case of a bounded extended convex set. Yarnold managed to simplify the item (16) in this simple case and converted the expansion (15) to

$$\begin{aligned} \Pr(X \in B^1) &= G_r(c) + (N^1 - n^{r/2} V^1) e^{-c/2} / \left((2\pi n)^r \prod_{j=1}^k p_j \right)^{1/2} \\ &+ O(n^{-1}), \end{aligned} \tag{17}$$

where $G_r(c)$ is the chi-squared distribution function with r degrees of freedom; N^1 is the number of points of the lattice L in B^1 ; V^1 is the volume of B^1 . Using the result from Esseen [10], Yarnold obtained an estimate of the second item in (17) in the form $O(n^{-(k-1)/k})$. If we estimate the second term in (17) taking the result from Götze [15] instead of Esseen's one from Esseen [10] we get (see [18]) in the case of the Pearson chi-squared statistics, i.e. when $\lambda = 1$, that for $r \geq 5$

$$\Pr(X \in B^1) = G_r(c) + O(n^{-1}).$$

In [42] it was shown that, when $\lambda = 0, \lambda = -1/2$, we have

$$\begin{aligned} J_1 &= G_r(c) + O(n^{-1}) \\ J_2 &= (N^\lambda - n^{r/2} V^\lambda) e^{-c/2} / \left((2\pi n)^r \prod_{j=1}^k p_j \right)^{1/2} + o(1), \\ V^\lambda &= V^1 + O(n^{-1}). \end{aligned} \tag{18}$$

These results were expanded by Read to the case $\lambda \in \mathbf{R}$. In particular, Theorem 3.1 in [32] implies

$$\Pr(T_\lambda < c) = \Pr(\chi_r^2 < c) + J_2 + O(n^{-1}). \tag{19}$$

This reduces the problem to the estimation of the order of J_2 .

It is worth mentioning that in [42] and in [32] there is no estimate for the residual in (18). Consequently, it is impossible to construct estimates of the rate of convergence of the statistics T_λ to the limiting distribution, based on the simple representation for J_2 initially suggested by Yarnold.

In [45] and in [1] the rate of convergence for the residual in (18) was obtained for any power divergence statistic. Then we constructed an estimate for J_2 based on the fundamental number theory results of Hlawka [25] and Huxley [26] about an approximation of a number of the integer points in the convex sets (more general than ellipsoids) by the Lebesgue measure of the set.

Therefore, one of the main point is to investigate the applicability of the aforementioned theorems from number theory to the set B^λ .

In [45] it is shown that $B^\lambda = \{x \mid T_\lambda(x) < c\}$ is a bounded extended convex (strictly convex) set. As it has been already mentioned, in accordance with the results of Yarnold [47]

$$J_2 = O(n^{-1/2}).$$

For the specific case of $r = 2$ this estimate has been considerably refined in [1]:

$$J_2 = O(n^{-3/4+\varepsilon}(\log n)^{315/146}) \tag{20}$$

with $\varepsilon = 3/4 - 50/73 < 0,0651$. As it follows from (18), the rate of convergence of J_2 to 0 cannot be better than the results in the lattice point problem for the ellipsoids in number theory, where for the case $r = 2$ we have the lower bound of the order $O(n^{-3/4} \log \log n)$ (see [24]). Therefore, the relation (20) gives for J_2 the order that is not far from the optimal one.

In [1] it was used the following theorem from Huxley [26]:

Theorem 2.2. *Let D be a two-dimensional convex set with area A , bounded by a simple closed curve C , divided into a finite number of pieces each of those being 3 times continuously differentiable in the following sense. Namely, on each piece C_i the radius of curvature ρ is positive (and not infinite), continuous, and continuously differentiable with respect to the angle of contingence ψ . Then in a set that is obtained from D by translation and linear expansion of order M , the number of integer points equals*

$$N = AM^2 + O(IM^K(\log M)^\Lambda)$$

$$K = \frac{46}{73}, \quad \Lambda = \frac{315}{146},$$

where I is a number depending only on the properties of the curve C , but not on the parameters M or A .

In [45] the results from Asylbekov et al. [1] were generalized to any dimension. The main reason why two cases when $r = 2$ and $r \geq 3$ are considered separately consists in the fact that for $r \geq 3$ it is much more difficult than for $r = 2$ to check

the applicability of the number theory results to B^λ . In [45] we used the following result from Hlawka [25]:

Theorem 2.3. *Let D be a compact convex set in \mathbf{R}^m with the origin as its inner point. We denote the volume of this set by A . Assume that the boundary C of this set is an $(m - 1)$ -dimensional surface of class \mathbf{C}^∞ , the Gaussian curvature being non-zero and finite everywhere on the surface. Also assume that a specially defined “canonical” map from the unit sphere to D is one-to-one and belongs to the class \mathbf{C}^∞ . Then in the set that is obtained from the initial one by translation along an arbitrary vector and by linear expansion with the factor M the number of integer points is*

$$N = AM^m + O\left(IM^{m-2+\frac{2}{m+1}}\right)$$

where the constant I is a number dependent only on the properties of the surface C , but not on the parameters M or A .

Providing that $m = 2$, the statement of Theorem 2.3 is weaker than the result of Huxley.

The above theorem is applicable in [45] with $M = \sqrt{n}$. Therefore, for any fixed λ we have to deal not with a single set, but rather with a sequence of sets $B^\lambda(n)$ which are, however, “close” to the limiting set B^1 for all sufficiently large n (see the representation for $T_\lambda(X)$ after (14)). It is necessary to emphasize that the constant I in our case, generally speaking, is $I(n)$, i.e. it depends on n . Only having ascertained the fulfillment of the inequality

$$|I(n)| \leq C_0,$$

where C_0 is an absolute constant, we are able to apply Theorem 2.3 without a change of the overall order of the error with respect to n .

In [45] we prove the following estimate of J_2 in the space of any fixed dimension $r \geq 3$.

Theorem 2.4. *For the term J_2 from the decomposition (19) the following estimate holds*

$$J_2 = O\left(n^{-r/(r+1)}\right), \quad r \geq 3,$$

The Theorem implies that for the statistics $t_\lambda(Y)$ and $T_\lambda(X)$ (see formula (14)) it holds that

$$\Pr(t_\lambda(Y) < c) = \Pr(T_\lambda(X) < c) = G_r(c) + O\left(n^{-1+\frac{1}{r+1}}\right), \quad r \geq 3.$$

Acknowledgements The authors are partly supported by RFBR grants, No. 11-01-00515 and No. 11-01-12104. The second author is partly supported as well by CRC 701 at Bielefeld University.

References

1. Zh.A. Asylbekov, V.N. Zubov, V.V. Ulyanov, On approximating some statistics of goodness-of-fit tests in the case of three-dimensional discrete data. *Siberian Math. J.* **52**(4), 571–584 (2011)
2. V. Bentkus, On dependence of Berry–Esseen bounds on dimensionality. *Lithuanian Math. J.* **26**, 205–210 (1986)
3. V. Bentkus, F. Götze, Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. *Probab. Theor. Relat. Fields* **109**, 367–416 (1997)
4. V. Bentkus, F. Götze, Optimal bounds in non-Gaussian limit theorems for UU -statistics. *Ann. Probab.* **27**(1), 454–521 (1999)
5. A.C. Berry, The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Am. Math. Soc.* **49**, 122–136 (1941)
6. R.N. Bhattacharya, R. Ranga Rao, *Normal Approximation and Asymptotic Expansions* (Robert E. Krieger Publishing Co., Inc., Melbourne, 1986), pp. xiv+291. ISBN: 0-89874-690-6
7. S.A. Bogatyrev, F. Götze, V.V. Ulyanov, Non-uniform bounds for short asymptotic expansions in the CLT for balls in a Hilbert space. *J. Multivariate Anal.* **97**(9), 2041–2056 (2006)
8. N.A.C. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B*, **46**, 440–464 (1984)
9. C.G. Esseen, On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.* **28A**(9), 19 (1942)
10. C.G. Esseen, Fourier analysis of distribution functions. *Acta Math.* **77**, 1–125 (1945)
11. C.G. Esseen, A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39**, 160–170 (1956)
12. F. Götze, Asymptotic expansion for bivariate von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **50**, 333–355 (1979)
13. F. Götze, Expansions for von Mises functionals. *Z. Wahrsch. Verw. Gebiete* **65**, 599–625 (1984)
14. F. Götze, Lattice point problems and the central limit theorem in Euclidean spaces. *Doc. Math. J. DMV, Extra vol. ICM III*, 245–255 (1998)
15. F. Götze, Lattice point problems and values of quadratic forms. *Inventiones mathematicae* **157**, 195–226 (2004)
16. F. Götze, G.A. Margulis, Distribution of values of quadratic forms at integral points. Preprint. <http://arxiv.org/abs/1004.5123> (2010)
17. F. Götze, V.V. Ulyanov, Uniform approximations in the CLT for balls in Euclidean spaces. Preprint 00-034 SFB 343, Univ.Bielefeld (2000)
18. F. Götze, V.V. Ulyanov, Asymptotic distribution of χ^2 -type statistics. Preprint 03-033, Research group “Spectral analysis, asymptotic distributions and stochastic dynamics” (2003)
19. F. Götze, A.Yu. Zaitsev, Uniform rates of convergence in the CLT for quadratic forms. Preprint 08119, SFB 701, Univ.Bielefeld (2008)
20. F. Götze, A.Yu. Zaitsev, Explicit rates of approximation in the CLT for quadratic forms. <http://arxiv.org/pdf/1104.0519.pdf> (2011)
21. F. Götze, A.Yu. Zaitsev, Uniform rates of approximation by short asymptotic expansions in the CLT for quadratic forms of sums of i.i.d. random vectors. Preprint 09073 SFB 701, Univ. Bielefeld, Bielefeld, (2009); published in *J.Math.Sci. (N.Y.)* **176**(2), 162–189 (2011)
22. F. Götze, Yu.V. Prokhorov, V.V. Ulyanov, Estimates for the characteristic functions of polynomials of asymptotically normal random variables. (Russian) *Uspekhi Mat. Nauk* **51** 2(308), 3–26 (1996); translation in *Russ. Math. Surv.* **51**(2), 181–204 (1996)
23. F. Götze, Yu.V. Prokhorov, V.V. Ulyanov, On the smooth behavior of probability distributions under polynomial mappings. (Russian) *Teor. Veroyatnost. i Primenen.* **42**(1), 51–62 (1997); translation in *Theor. Probab. Appl.* **42**(1), 28–38 (1998)
24. G. Hardy, On Dirichlet’s divisor problem. *Proc. Lond. Math. Soc.* **15**, 1–25 (1916)
25. E. Hlawka, Über integrale auf konvexen körpern I. *Mh. Math.* **54**, 1–36 (1950)

26. M.N. Huxley, Exponential sums and lattice points II. Proc. Lond. Math. Soc. **66**, 279–301 (1993)
27. N.P. Kandelaki, On limit theorem in Hilbert space. Trudy Vychisl. Centra Akad. Nauk Gruzin. SSR **11**, 46–55 (1965)
28. S.V. Nagaev, On new approach to study of distribution of a norm of a random element in a Hilbert space. Fifth Vilnius conference on probability theory and mathematical statistics. Abstracts **4**, 77–78 (1989)
29. S.V. Nagaev, V.I. Chebotarev, A refinement of the error estimate of the normal approximation in a Hilbert space. Siberian Math. J. **27**, 434–450 (1986)
30. S.V. Nagaev, V.I. Chebotarev, On the accuracy of Gaussian approximation in Hilbert space. Acta Applicandae Mathematicae **58**, 189–215 (1999)
31. S.V. Nagaev, V.I. Chebotarev, On the accuracy of Gaussian approximation in a Hilbert space. (Russian) Mat. Tr. **7**(1), 91–152 (2004); translated in Siberian Adv. Math. **15**(1), 11–73 (2005)
32. T.R.C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics. Ann. Math. Stat. Part A **36**, 59–69 (1984)
33. V.V. Sazonov, On the multi-dimensional central limit theorem. Sankhya Ser. A **30**(2), 181–204 (1968)
34. V.V. Sazonov, *Normal Approximation – Some Recent Advances*. Lecture Notes in Mathematics, vol. 879 (Springer, Berlin, 1981)
35. V.V. Sazonov, V.V. Ulyanov, B.A. Zalesskii, Normal approximation in a Hilbert space. I, II. Theor. Probab. Appl. **33**, 207–227, 473–483 (1988)
36. V.V. Sazonov, V.V. Ulyanov, B.A. Zalesskii, A sharp estimate for the accuracy of the normal approximation in a Hilbert space. Theor. Probab. Appl. **33**, 700–701 (1988)
37. V.V. Sazonov, V.V. Ulyanov, B.A. Zalesskii, A precise estimate of the rate of convergence in the CLT in Hilbert space. Mat. USSR Sbornik **68**, 453–482 (1991)
38. V.V. Senatov, Four examples of lower bounds in the multidimensional central limit theorem. Theor. Probab. Appl. **30**, 797–805 (1985)
39. V.V. Senatov, On rate of convergence in the central limit theorem in a Hilbert space. Fifth Vilnius conference on probability theory and mathematical statistics. Abstracts **4**, 222 (1989)
40. V.V. Senatov, Qualitive effects in the estimates of convergence rate in the central limit theorem in multidimensional spaces, in *Proceedings of the Steklov Institute of Mathematics*, vol. 215, Moscow, Nauka (1996)
41. I.G. Shevtsova, On the absolute constants in the BerryEsseen type inequalities for identically distributed summands. Preprint <http://arxiv.org/pdf/1111.6554.pdf> (2011)
42. M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics. Hiroshima Math. J. **14**, 115–124 (1984)
43. I.S. Tyurin, Sharpening the upper bounds for constants in Lyapunov’s theorem. (Russian) Uspekhi Mat. Nauk **65** 3(393), 201–201 (2010); translation in Russ. Math. Surv. **65**(3), 586–588 (2010)
44. V.V. Ulyanov, Normal approximation for sums of nonidentically distributed random variables in Hilbert spaces. Acta Sci. Math. (Szeged) **50**(3–4), 411–419 (1986)
45. V.V. Ulyanov, V.N. Zubov, Refinement on the convergence of one family of goodness-of-fit statistics to chi-squared distribution. Hiroshima Math. J. **39**(1), 133–161 (2009)
46. V.V. Ulyanov, F. Götzte, Short asymptotic expansions in the CLT in Euclidian spaces: a sharp estimate for its accuracy. *Proceedings 2011 World Congress on Engineering and Technology*, vol. 1, 28 Oct–2 Nov 2011, Shanghai, China (IEEE, New York, 2011), pp. 260–262
47. J.K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. Ann. Math. Stat. **43**, 1566–1580 (1972)
48. V.V. Yurinskii, On the accuracy of normal approximation of the probability of hitting a ball. Theor. Probab. Appl. **27**, 280–289 (1982)