On asymptotics for power divergence family of statistics

Vladimir V. Ulyanov

(Received Xxx 00, 0000)

Abstract. We review the results about asymptotic properties for distributions of statistics from power divergence family of statistics. This family includes such famous goodness-of-fit statistics as Karl Pearson chi-squared test, Freeman-Tukey statistic and log-likelihood ratio statistic. We consider recent results, where for the first time the order of approximation was obtained in the form $O(n^{-c})$ with some positive constant c.

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 the 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. The review does not pretend to completely illuminate the present state of the area under consideration. It reflects mainly the author's interests and results.

We consider accuracy of approximations for 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 study of asymptotic behavior of 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 $H_0\colon \pi=\boldsymbol{p}$. Since the sum of n_i equals n, we can express this multinomial distribution in terms of a vector $\boldsymbol{Y}=(Y_1,\dots,Y_{k-1})$ and define its covariance matrix Ω. It is known that so defined Ω equals

The author is partly supported by Higher School of Economics grant, No. 12-05-0052.

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

Key words and phrases. Accuracy of approximations, goodness-of-fit statistics, power divergence family of statistics.

$(\delta_i^j\, p_i - p_i p_j) \in \mathbf{R}^{(k-1)\times(k-1)}$. The main object of the current study is the power divergence family of goodness-of-fit test statistics:

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

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

These statistics were first introduced in Cressie and Read (1984) and Read (1984). 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

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

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

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

where n_j are non-negative integers.

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

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

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

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

As we see the statistics $T_\lambda(\boldsymbol{X})$ is "close" to quadratic form

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

considered in Section 1.

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

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

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

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

It is a known fact that the distributions of all statistics in the family converge to chi-squared distribution with $k-1$ degrees of freedom (see e.g. Cressie and Read (1984), 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 Yarnold (1972) it was obtained an asymptotic expansion, which in Siotani and Fujikoshi (1984) was converted to

$$\Pr(\boldsymbol{X} \in B) = J_1 + J_2 + O(n^{-1}). \tag{2}$$

with

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

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

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

$$J_2 = -\frac{1}{\sqrt{n}} \sum\nolimits_{l=1}^{r} n^{-(r-l)/2} \sum\nolimits_{x_{l+1} \in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r} \left[\int \cdots \int_{B_l} [S_1(\sqrt{n} x_l + n p_l) \phi(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)}\, dx_1, \cdots, dx_{l-1} \right]; \tag{3}$$

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

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

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

In Yarnold (1972) it was showed 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 Yarnold (1972) it was also examined that expansion for the most known power divergence statistic, which is the chi-squared statistic. Put $B^\lambda = \{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$. It is easy to show that B^1 is an ellipsoid, which is a particular case of a bounded extended convex set. J. Yarnold managed to simplify the item (3) in this simple case and converted the expansion (2) to

$$\begin{aligned} \Pr(\boldsymbol{X} \in B^1) \quad &= \quad G_r(c) + (N^1 - n^{r/2} V^1) e^{-c/2} \Big/ \left((2\pi n)^r \prod\nolimits_{j=1}^{k} p_j \right)^{1/2} \\ &\quad + O(n^{-1}), \end{aligned} \tag{4}$$

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 (1945), he obtained an estimate of the second item in (4) in the form $O(n^{-(k-1)/k})$. If we estimate second term in (4) taking the result from Götze (2004) instead of Esseen's one from Esseen (1945) we get (see Götze and Ulyanov (2003)) in the case of Karl Pearson chi-squared statistics, i.e. when $\lambda = 1$, that for $r \geq 5$

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

In Shiotani and Fujikoshi (1984) it was showed that, when $\lambda = 0, \lambda = -1/2$, we have

$$J_1 = G_r(c) + O(n^{-1})$$

$$J_2 = (N^\lambda - n^{r/2}V^\lambda)\, e^{-c/2} \Big/ \left((2\pi n)^r \prod_{j=1}^k p_j\right)^{1/2} + o(1), \tag{5}$$

$$V^\lambda = V^1 + O(n^{-1}).$$

These results were expanded by T. Read to the case $\lambda \in \mathbf{R}$. In particular Theorem 3.1 in Read (1984) implies

$$\Pr\left(T_\lambda < c\right) = \Pr\left(\chi_r^2 < c\right) + J_2 + O\left(n^{-1}\right). \tag{6}$$

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

It is worth mentioning that in Siotani and Fujikoshi (1984) and in Read (1984) there is no estimate for the residual in (5). Consequently, it is impossible to construct estimates of the rate of convergence of statistics T_λ to the limiting distribution, grounded on the simple representation for J_2 initially suggested by J. Yarnold.

In Ulyanov and Zubov (2009) and in Asylbekov, Zubov and Ulyanov (2011) the rate of convergence in (20) was obtained for any power divergence statistic. Then we construct an estimate for J_2 based on the fundamental number theory results of Hlawka (1950) and Huxley (1993) about approximation of number of integer points in convex sets (more general than ellipsoids) by Lebesgue measure of the set.

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

In Ulyanov and Zubov (2009) it is shown that $B^\lambda = \{\boldsymbol{x} \mid T_\lambda(\boldsymbol{x}) < c\}$ is a bounded extended-convex (strictly convex) set. As it has been already mentioned, in accordance with the results of Yarnold (1972)

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

For the specific case of $r = 2$ this estimate has been considerably refined in Asylbekov, Zubov and Ulyanov (2011):

$$J_2 = O\left(n^{-50/73}(\log n)^{315/146}\right), \quad r = 2.$$

In Asylbekov, Zubov and Ulyanov (2011) it was used the following theorem from Huxley (1993):

THEOREM 1. *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*

$$N = AM^2 + O\left(IM^K(\log M)^\Lambda\right)$$
$$K = \frac{46}{73},\ \Lambda = \frac{315}{146},$$

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

In Ulyanov and Zubov (2009) the results from Asylbekov, Zubov and Ulyanov (2011) 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 applicability the number theory results to B^λ. In Ulyanov and Zubov (2009) we used the following result from Hlawka (1950)

THEOREM 2. *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 of this set is an $(m-1)$-dimensional surface of class $\mathbf{C}^\infty$, 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 $\mathbf{C}^\infty$. 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 curve C, but not on the parameters M or A.

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

The above theorem is applicable in Ulyanov and Zubov (2009) 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 $B^\lambda(n)$ converging in some sense to the limiting set B^1 when $n \to \infty$. 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

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

where C_0 is an absolute constant, we are able to apply Theorem 2 without a change of the overall order of the error with respect to n.

In Ulyanov and Zubov (2009) we prove the following estimate of J_2 in the space of any fixed dimension $r \geqslant 3$.

Theorem 3. *For the term J_2 from decomposition* (6) *the following estimate holds*

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

The Theorem implies that for the statistic $T_\lambda(\boldsymbol{X})$ defined by formula (1) it holds that

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

References

[1] Zh. A. Asylbekov, V. N. Zubov and V. V. Ulyanov, On approximating some statistics of goodness-of-fit tests in the case of three-dimensional discrete data. Siberian Mathematical Journal. 2011, Volume 52, Number 4, Pages 571-584.

[2] Bentkus, V.: On dependence of Berry–Esseen bounds on dimensionality. Lithuanian Math. J. **26**, 205–210 (1986)

[3] Bentkus, V., Götze, F.: Uniform rates of convergence in the CLT for quadratic forms in multidimensional spaces. Probab. Theory Relat. Fields **109**, 367–416 (1997)

[4] Bentkus, V.; Go"tze, F. Optimal bounds in non-Gaussian limit theorems for UU-statistics. Ann. Probab. 27 (1999), no. 1, 454521.

[5] Berry, A.C.: The accuracy of the Gaussian approximation to the sum of independent variates. Trans. Amer. Math. Soc. **49**, 122-136 (1941).

[6] Bhattacharya, R.N. and Ranga Rao, R. Normal approximation and asymptotic expansions. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1986. xiv+291 pp. ISBN: 0-89874-690-6.

[7] Bogatyrev, S.A., Götze, F., Ulyanov, V.V.: 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, Journal of the Royal Statistical Society, Series B, **46** (1984), 440–464.

[9] Esseen, C.G.: On the Liapounoff limit of error in the theory of probability. Ark. Mat. Astr. Fys. **28A**, no. 9, 19 pp. (1942).

[10] Esseen, C.G.: Fourier analysis of distribution functions. Acta Math. **77**, 1–125 (1945)

[11] Esseen, C.G.: A moment inequality with an application to the central limit theorem. Skand. Aktuarietidskr. **39**, 160-170 (1956).

[12] Götze, F.: Asymptotic expansion for bivariate von Mises functionals. Z. Wahrsch. Verw. Gebiete **50**, 333–355 (1979)

[13] Götze, F.: Expansions for von Mises functionals. Z. Wahrsch. Verw. Gebiete **65**, 599–625 (1984)

[14] Götze, F.; Prokhorov, Yu. V.; Ulyanov, V. V. Estimates for the characteristic functions of polynomials of asymptotically normal random variables. (Russian) Uspekhi Mat. Nauk 51 (1996), no. 2(308), 3–26; translation in Russian Math. Surveys 51 (1996), no. 2, 181204

[15] Götze, F.; Prokhorov, Yu. V.; Ulyanov, V. V. On the smooth behavior of probability distributions under polynomial mappings. (Russian) Teor. Veroyatnost. i Primenen. 42 (1997), no. 1, 51–62; translation in Theory Probab. Appl. 42 (1997), no. 1, 2838 (1998).

[16] Götze, F.: Lattice point problems and the central limit theorem in Euclidean spaces. Doc. Math. J.DMV, Extra Vol. ICM, III, 245–255 (1998)
[17] Götze, F. and Margulis, G.A.: Distribution of values of quadratic forms at integral points, Preprint http://arxiv.org/abs/1004.5123, 2010.
[18] Götze, F., Ulyanov, V.V.: Uniform approximations in the CLT for balls in Euclidian spaces, Preprint 00-034 SFB 343, Univ.Bielefeld (2000).
[19] Götze, F., Ulyanov, V.V.: Asymptotic disrtribution of χ^2-type statistics, Preprint 03-033, Research group "Spectral analysis, asymptotic distributions and stochastic dynamics", 2003.
[20] F. Götze, Lattice point problems and values of quadratic forms, Inventiones mathematicae, **157** 2004, 195–226.
[21] Götze, F., Zaitsev, A. Yu.:Uniform rates of convergence in the CLT for quadratic forms. Preprint 08119. SFB 701, Univ.Bielefeld (2008).
[22] Götze, F., Zaitsev, A. Yu.: Explicit rates of approximation in the CLT for quadratic forms. http://arxiv.org/pdf/1104.0519.pdf (2011).
[23] E. Hlawka Über integrale auf konvexen körpern I. Mh. Math **54** (1950), 1–36.
[24] M. N. Huxley, Exponential sums and lattice points II. Proceedings of London Mathematical Society, **66** (1993), 279–301.
[25] Kandelaki, N.P.: On limit theorem in Hilbert space. Trudy Vychisl. Centra Akad. Nauk Gruzin. SSR **11**, 46–55 (1965)
[26] Nagaev, S.V.: 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)
[27] Nagaev, S.V., Chebotarev, V.I.: A refinement of the error estimate of the normal approximation in a Hilbert space. Siberian Math. J. **27**, 434–450 (1986)
[28] Nagaev, S.V., Chebotarev, V.I.: On the accuracy of Gaussian approximation in Hilbert space. Acta Applicandae Mathematicae **58**, 189–215 (1999)
[29] T. R. C. Read, Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics., The Annals of Mathematical Statistics, Part A, **36** (1984), 59–69.
[30] Sazonov, V.V.: On the multi-dimensional central limit theorem. Sankhya Ser. A **30**, no.2, 181–204 (1968)
[31] Sazonov, V.V.: Normal approximation – some recent advances. Lecture Notes in Mathematics, 879, Springer-Verlag, Berlin, NY (1981)
[32] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: Normal approximation in a Hilbert space. I, II. Theory Probab. Appl. **33** , 207–227, 473–483 (1988a)
[33] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: A sharp estimate for the accuracy of the normal aapproximation in a Hilbert space. Theory Probab. Appl. **33**, 700–701 (1988b)
[34] Sazonov, V.V., Ulyanov, V.V., Zalesskii, B.A.: A precise estimate of the rate of covergence in the CLT in Hilbert space. Mat.USSR Sbornik **68**, 453–482 (1991)
[35] Senatov, V.V.: Four examples of lower bounds in the multidimensional central limit theorem. Theory Probab.Appl. **30**, 797–805 (1985)
[36] Senatov, V.V.: 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)
[37] Senatov, V.V.: Qualitive effects in the estimates of convergence rate in the central limit theorem in multidimensional spaces. Proceedings of the Steklov Institute of Mathematics. Vol.**215**, Moscow, Nauka (1996)
[38] Shevtsova, I.G.: On the absolute constants in the BerryEsseen type inequalities for identically distributed summands, Preprint http://arxiv.org/pdf/1111.6554.pdf, 2011.
[39] M. Siotani, Y. Fujikoshi, Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics, Hiroshima Mathematical Journal, **14** (1984), 115–124.

[40] Tyurin, I.S.: Sharpening the upper bounds for constants in Lyapunov's theorem. (Russian) Uspekhi Mat. Nauk 65 (2010), no. 3(393), 201–201; translation in Russian Math. Surveys 65 (2010), no. 3, 586–588.

[41] Ulyanov, V.V.: Normal approximation for sums of nonidentically distributed random variables in Hilbert spaces. Acta Sci. Math. (Szeged) **50**, no. 3-4, 411-419, (1986).

[42] Ulyanov, Vladimir V.; Zubov, Vasily N. Refinement on the convergence of one family of goodness-of-fit statistics to chi-squared distribution. Hiroshima Math. J. 39 (2009), no. 1, 133161.

[43] Ulyanov, V.V., Götze, F.: Short asymptotic expansions in the CLT in Euclidian spaces: a sharp estimate for its accuracy. Proceedings 2011 World Congress on Engineering and Technology. Oct.28-Nov.2, 2011. Shanghai, China, IEEE Press, **1**, 260–262, 2011.

[44] J. K. Yarnold, Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. The Annals of Mathematical Statistics, **43** (1972), 1566–1580.

[45] Yurinskii, V.V.: On the accuracy of normal approximation of the probability of hitting a ball. Theory Probab. Appl. **27**, 280–289 (1982)

Vladimir V. Ulyanov
Department of Methods of Collection and Analysis of Sociological Information
Faculty of Sociology
National research university "Higher school of economics",
Moscow, RUSSIA
E-mail: sentea@mail.ru