computer sciences and mathematics forum

MDPI

Proceeding Paper

On Estimation of the Remainder Term in New Asymptotic Expansions in the Central Limit Theorem †

Vitaly Sobolev [1,*] and Sergey Ramodanov [2]

1 Department of Probability Theory and Applied Mathematics, Moscow Technical University of Communications and Informatics, 123423 Moscow, Russia
2 Mechanical Engineering Research Institute RAS, 101990 Moscow, Russia; ramodanov@mail.ru
* Correspondence: v.n.sobolev@mtuci.ru

† Presented at the 1st International Online Conference on Mathematics and Applications, 1–15 May 2023; Available online: https://iocma2023.sciforum.net/.

Abstract: We offer a new asymptotic expansion with an explicit remainder estimate in the central limit theorem. The results obtained are essentially based on new forms of asymptotic expansions in the central limit theorem. We also present a more accurate estimation of the CLT-expansions remainder which is rigorously proved and backed up numerically. It is shown that our approach can be used for further refinement of allied asymptotic expansions.

Keywords: central limit theorem; asymptotic expansion; random variables; characteristic function; accuracy of approximation; approximation exactness; estimates for the exactness of approximation; Senatov moments; distribution of probabilities

Citation: Sobolev, V.; Ramodanov, S. On Estimation of the Remainder Term in New Asymptotic Expansions in the Central Limit Theorem. *Comput. Sci. Math. Forum* **2023**, *7*, 60. https://doi.org/10.3390/IOCMA2023-14423

Academic Editor: Antonio Di Crescenzo

Published: 28 April 2023

1. Introduction

The central limit theorem (CLT) states that under fairly broad conditions the sum of independent (or weakly dependent) identically distributed (i.i.d.) random variables is approximately normally distributed. In this paper, we deal with CLT for i.i.d. variables $X_1, X_2, \ldots$ with zero mean and unit variance each. Let X_1 follow a distribution P with cumulative distribution function $F(x)$ and characteristic function $f(t)$.

We assume that X_1 has a finite absolute moment of order $m+2$ and for some $\nu > 0$ the function $|f(t)|^\nu$ is integrable on $\mathbb{R}$. Since for any $T > 0$ the integral

$$\int_T^{+\infty} |f(t)|^\nu dt < +\infty$$

converges and

$$\alpha(T) = \sup\{|f(t)| : t \geqslant T\} < 1$$

(see [1], p. 43), it follows that for $n \geqslant \nu$ the distribution P_n of the rescaled sum $(X_1 + \ldots + X_n)n^{-\frac{1}{2}}$ has a continuous density $p_n(x)$. In it is its turn, this implies that for this distribution P_n (with cumulative distribution function $F_n(x)$) the CLT holds [1], $p_n(x) \to \varphi(x)$ as $n \to +\infty$ for any $x \in \mathbb{R}$; here $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the density of the standard normal distribution $\Phi(x)$. The density $p_n(x)$ can be represented as the inverse Fourier transform [1] (pp. 42, 147) of the characteristic function f^n of the convolution of n copies of the original distribution.

Here the question naturally arises concerning the accuracy of the CLT-approximation. One of the main results on the subject is the Berry–Esseen theorem (see, for example, [2]). Nevertheless, this theorem is too general (and not attempting to take any specific properties of the original distribution into play) and, therefore, provides rather crude estimates of the convergence rate [3].

One way to improve the accuracy of CLT-approximations is to use asymptotic expansions. Until quite recently, most of such expansions only gave an estimate of the order of approximation at best and thus were of little use for, say, numeric computations.

Yu.V. Prohorov wrote [4] (p. 7) that "it was P.L. Chebyshev's idea to explore the asymptotic behavior of the difference $F_n(x) - \Phi(x)$ and it was them to give a formal expansion of the difference". A number of such expansions under different restrictions on the original distribution were obtained later by H. Cramer [5] and investigated by C.-G. Esseen [6]. Moreover, H. Cramer [7] claimed that the featured series expansions were introduced by F. Edgeworth [8]. For example, if there exists an integer $m \geqslant 1$, such that $\mathrm{M}|X_1|^{m+2} < \infty$ and Cramer's (C) condition $\lim\limits_{|t|\to\infty} \sup|f(t)| < 1$ is fulfilled, then

$$F_n(x) = \Phi(x) + \sum_{k=1}^{m} \frac{P_k(-\Phi)}{n^{k/2}} + O\left(\frac{1}{n^{m/2}}\right), \, n \to \infty \,,$$

where $P_k(-\Phi) = L_{3k-1}(x)\varphi(x)$ and $L_{3k-1}(x)$ is a polynomial of degree $3k-1$ in x. Explicit formulas for $P_k(-\Phi)$ in terms of semi-invariants were obtained by V.V. Petrov [9] in 1962.

Seeking more efficient forms of such expansions V.V. Senatov took what can be justly characterized a revolutionary step; he offered expansions that allowed for explicit estimation of the remainder (not just barely indicating big-Oh approximation order). At the moment only one type of CLT-asymptotic expansions, namely the Gram–Charlier expansion, was widely known, that is,

$$p_n(x) = \varphi(x) + \sum_{k=3,\, k\neq 3m-4}^{3m-3} \theta_k(P_n) H_k(x)\varphi(x) + O\left(\tfrac{1}{n^{m/2}}\right), \, n \to \infty \,,$$

where $H_k(x) = (-1)^k \frac{\varphi^{(k)}(x)}{\varphi(x)}$ are Chebyshev–Hermite polynomials and

$$\theta_k(P_n) = \frac{1}{k!} \int\limits_{-\infty}^{+\infty} H_k(x) p_n(x) dx \,, \, k \geqslant 0$$

are normalized moments [10].

As mentioned by V.V. Senatov [1] (p. 124), already, in 1920, H. Cramer [7] noticed that this accuracy can be attained with only $m+2$ moments at hand (the Gram–Charlier expansion provides the same accuracy only when $3m-3$ moments are used). That was the reason, in Senatov's opinion, why researchers' primary focus was on the Edgeworth–Cramer (not Gram–Charlier) expansion.

At the same, it turns out that there exists an expansion of the Gram–Charlier type that makes use of only $m+2$ absolute moment of variable X_1. V.V. Senatov to construct this kind of asymptotic expansions proposed together with the moments:

$$\theta_k = \sum_{j=0}^{\left[\frac{k}{2}\right]} (-1)^j \alpha_{k-2j} \alpha_{2j}(\varphi) \,,$$

use the so-called [10] Senatov incomplete moments:

$$\theta_k^{(l)} = \sum_{j=0}^{\left[\frac{l}{2}\right]} (-1)^j \alpha_{k-2j} \alpha_{2j}(\varphi) \,, \, l \leqslant k \,,$$

where

$$\alpha_j = \frac{\mathrm{M}X_1^j}{j!} \,, \quad \beta_j = \frac{\mathrm{M}|X_1|^j}{j!} \,, \quad \alpha_{2l}(\varphi) = \frac{1}{2^l l!} \,.$$

V.V. Senatov came up [11] with what he dubbed a shortened Gram–Charlier expansion:

$$p_n(x) = \varphi(x) + \sum_{k=3}^{m+1} \theta_k(P_n) H_k(x)\varphi(x) + \sum_{k=m+2,\, k\neq 3m-4}^{3m-3} \theta_k^{(m+1)}(P_n) H_k(x)\varphi(x) + R\,.$$

He used arbitrary measures (not necessarily probabilistic measures) in their derivation, which imposes additional limitations on the moments of the original distribution.

In attempts to remove the limitations and make the estimation of the remainder more accurate, V.V. Senatov and V.N. Sobolev [12] suggested a novel form of asymptotic expansion that does not impose any additional restrictions on the momenta (as compared to [1]).

The Gram–Charlier expansion differs architecturally from the Edgeworth–Cramer expansion, the former is in powers of the Chebyshev–Hermite polynomials, while the latter is in powers of n. Expansions of the third type [12,13] are obtained as follows, the terms in such an expansion are to be ordered with respect to the number of factors, which are Senatov's moments.

Before the results of V.V. Senatov and V.N. Sobolev [12,13], only two types of such expansions were widely known, the Gram–Charlier and Edgeworth–Cramer expansions. In the former, the terms are grouped in the order of the Chebyshev–Hermite polynomials, and in the latter, they are grouped in powers of n. In [12], V.V. Senatov and V.N. Sobolev proposed grouping the terms according to the number of factors of Senatov's moments

$$p_n(x) = \varphi(x) + \sum_{s=1}^{m-1} C_n^s \sum_{l=3s}^{m-1+2s} \frac{\Theta_{s,l}}{n^{l/2}} H_l x)\varphi(x) + O\left(\frac{1}{n^{m/2}}\right),\ n\to\infty\,,$$

where

$$\Theta_{s,l} = \sum_{t_1+\ldots+t_s=l} \theta_{t_1}\ldots\theta_{t_s}$$

the summation is carried out over tuples of natural numbers $t_1, \ldots, t_s$, such that $t_j \geqslant 3$, $j = 1, \ldots, m-1$ and $t_1 + \ldots + t_s = l$.

2. Main Result

To improve the accuracy of the expansions from [12] note first that the values

$$\alpha_0 = 1, \quad \alpha_1 = 0, \quad \alpha_2 = \frac{1}{2}, \quad \alpha_{2l}(\varphi) = \frac{1}{2^l l!}$$

are known. Then, we introduce the following non-negative quantities:

$$\|\theta_s\| = \beta_{s+2} + \sum_{j=1}^{\left[\frac{s}{2}\right]-2} \alpha_{2j}(\varphi)\left|\alpha_{s-2j}\right| + \left|\sum_{j=\left[\frac{s}{2}\right]-1}^{\left[\frac{s}{2}\right]} (-1)^j \alpha_{2j}(\varphi)\alpha_{s-2j}\right|,$$

for which the inequalities

$$\|\theta_s\| \leqslant \beta_{s+2} + \sum_{j=1}^{\left[\frac{s}{2}\right]} \alpha_{2j}(\varphi)\left|\alpha_{s-2j}\right|$$

are valid. Therefore, using $\|\theta_s\|$ in the estimates instead of the right side of the last inequality increases accuracy of asymptotic expansions.

Let $\left\|\theta_s^{(l)}\right\|$ be the abbreviated versions of $\|\theta_s\|$: they are calculated by the same formulas in which $\alpha_k = 0$ for $k > l$. There is also an improvement here.

Another improvement consists of preserving the minus signs between the summands when evaluating the estimates. The following quantities naturally arise

$$S_{0,2}(\Theta) = \|\theta_{m+2}\|\,, \qquad S_{0,3}(\Theta) = \left\|\theta_{m+3}^{(m+1)}\right\|$$

which for $l \geqslant 1$ read

$$S_{l,2}(\Theta) = \sum_{k_1,k_2,...,k_l}^{m} \left|\theta_{3+k_1}\theta_{3+k_2}...\theta_{3+k_l}\right| \left\|\theta_{m+2-(l+k_1+k_2+...+k_l)}\right\|,$$

$$S_{l,3}(\Theta) = \sum_{k_1,k_2,...,k_l}^{m} \left|\theta_{3+k_1}\theta_{3+k_2}...\theta_{3+k_l}\right| \left\|\theta_{m+3-l-k_1-k_2-...-k_l}^{(m+2-l-k_1-k_2-...-k_l)}\right\|.$$

The notation $\sum\limits_{k_1,k_2,...,k_l}^{m}$ implies summation over all sets $k_1, k_2, ..., k_l$ of non-negative numbers, such that $0 \leqslant k_1 + k_2 + ... + k_l \leqslant m - l - 1$.

We will also use the following quantities of the moment type

$$L_l(u) = \frac{1}{2\pi} \int\limits_{|t|\geqslant u} |t|^l e^{-\frac{t^2}{2}}\, dt\,, \qquad B_{l,n-k} = \frac{1}{2\pi} \int\limits_{-T\sqrt{n}}^{+T\sqrt{n}} |t|^l \mu^{n-k}\left(\frac{t}{\sqrt{n}}\right) dt\,,$$

where the function $\mu(t) = \max\left\{|f(t)|, e^{-t^2/2}\right\}$ was introduced by V.Yu. Korolev.

Below, we formulate our main theorem that provides a CLT-asymptotic expansion with an improved estimation of the remainder under fairly general conditions.

Theorem 1. *Let identically distributed independent random variables $X_1, X_2, ...$ with zero mean and unit variance each follow the same distribution P. Suppose that P has a finite absolute moment of order $m+2$ and $\int_{-\infty}^{\infty} |f(t)|^{\nu} dt < \infty$. Where $f(t)$ is the characteristic function of P. Then,* for any $n \geqslant \max(\nu, m+1)$ and for all $x \in \mathbb{R}$

$$p_n(x) = \varphi(x) \sum_{l=0}^{m-1} \frac{C_n^l}{\left(\sqrt{n}\right)^{3l}} \sum_{k_1,k_2,...,k_l}^{m} \frac{\theta_{3+k_1}\theta_{3+k_2}...\theta_{3+k_l}}{\left(\sqrt{n}\right)^{k_1+k_2+...+k_l}} H_{3l+k_1+...+k_l}(x) + R_{n,m}(x)\,,$$

where

$$|R_{n,m}(x)| \leqslant \frac{1}{\left(\sqrt{n}\right)^{m+2}} \sum_{l=0}^{m-2} \frac{C_n^{l+1}}{n^l} \left(B_{m+2+2l,n-1} S_{l,2}(\Theta) + B_{m+3+2l,n-1} S_{l,3}(\Theta) \frac{1}{\sqrt{n}} \right)$$

$$+\Lambda_n(T) + \bar{\Lambda}_n(T)\,.$$

and two last terms in the remainder's estimate are

$$\Lambda_n(T) = \frac{\sqrt{n}}{\pi} \alpha^{n-\nu}(T) \int\limits_{T}^{+\infty} |f(t)|^{\nu} dt\,,$$

$$\bar{\Lambda}_n(T) = \sum_{l=0}^{m-1} C_n^l \sum_{k_1,k_2,...,k_l}^{m} \frac{\left|\theta_{3+k_1}\theta_{3+k_2}...\theta_{3+k_l}\right|}{\left(\sqrt{n}\right)^{3l+k_1+k_2+...+k_l}} L_{3l+k_1+k_2+...+k_l}\left(T\sqrt{n}\right),$$

which decay exponentially fast.

3. Conclusions

We obtain new explicit estimates for accuracy of approximation in the CLT-expansions. Our approach can be used for further refinement of allied asymptotic expansions.

Author Contributions: Conceptualization, V.S. and S.R.; methodology, V.S. and S.R.; software, V.S. and S.R.; validation, V.S. and S.R.; formal analysis, V.S. and S.R.; investigation, V.S. and S.R.; resources, V.S. and S.R.; data curation, V.S. and S.R.; writing—original draft preparation, V.S. and S.R.; writing—review and editing, V.S. and S.R.; visualization, V.S. and S.R.; supervision, V.S. and S.R.; project administration, V.S. and S.R.; funding acquisition, V.S. and S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CLT	Central Limit Theorem
i.i.d.	independent and identically distributed

References

1. Senatov, V.V. *The Central Limit Theorem: Approximation Accuracy and Asymptotic Expansions*; Librokom: Moscow, Russia, 2009. (In Russian)
2. Shevtsova, I.G. Optimizaciya struktury momentnyh ocenok tochnosti normal'noj approksimacii dlya raspredelenij summ nezavisimyh sluchajnyh velichin [Optimization of the Structure of Moment Estimates for the Accuracy of Normal Approximation of the Distribution of Sums of Independent Random Variables]. Ph.D. Dissertation, Moscow State Univiversity, Moscow, Russia, 2013; 354p. (In Russian)
3. Sobolev, V.N. On the estimation of the accuracy of asymptotic expansions in the central limit theorem. *Surv. Appl. Ind. Math.* **2011**, *18*, 45–51. (In Russian)
4. Prohorov, Y.V. Predel'nye teoremy dlya summ nezavisimyh sluchajnyh velichin [Limit Theorems for Sums of Independent Random Variables]. Ph.D. Thesis, Moscow State Univiversity, Moscow, Russia, 1952. (In Russian)
5. Cramér, H. On an asymptotic expansion occurring in the theory of probability. *J. Lond. Math. Soc.* **1927**, *2*, 262–265. [CrossRef]
6. Esseen, C.-G. Fourier analysis of distribution functions: A mathematical study of the Laplace–Gaussian law. *Acta Math.* **1945**, *106*, 1–125. [CrossRef]
7. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
8. Edgeworth, F.Y. The law of error. *Camb. Phil. Soc. Proc.* **1905**, *20*, 36–141. [CrossRef]
9. Petrov, V.V. On some polynomials encountered in probability. *Vestnik Leningrad Univ. Math.* **1962**, *19*, 150–153.
10. Sobolev, V.N.; Kondratenko, A.E. On Senatov Moments in Asymptotic Expansions in the Central Limit Theorem. *Theory Probab. Appl.* **2022**, *67*, 154–157. [CrossRef]
11. Senatov, V.V. On Asymptotic Expansions in the Central Limit Theorem with Explicit Estimates of Remainder Terms. *Theory Probab. Appl.* **2006**, *51*, 810–816. [CrossRef]
12. Senatov, V.V.; Sobolev, V.N. New forms of asymptotic expansions in the central limit theorem. *Theory Probab. Appl.* **2013**, *57*, 82–96. [CrossRef]
13. Sobolev, V.N. On asymptotic expansions in CLT. *Vestn. TVGU. Seriya Prikl. Mat. [Herald Tver State Univ. Ser. Appl. Math.]* **2010**, *18*, 35–48. (In Russian)