КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ
COMPUTER SIMULATION HISTORY

УДК 519.21
DOI 10.52575/2687-0932-2022-49-3-546-557

Оценивание числа слагаемых суммы независимых случайных величин при моделировании гауссовских случайных величин

[1] Ганичева А.В., [2] Ганичев А.В.
[1] Тверская государственная сельскохозяйственная академия,
Россия, 170904, Тверь, ул. Маршала Василевского, 7
[2] Тверской государственный технический университет,
Россия, 170026, Тверь, наб. Аф. Никитина, 22
E-mail: TGAN55@yandex.ru, alexej.ganichev@yandex.ru

Аннотация. Одной из важнейших проблем теории вероятностей является оценка числа слагаемых центральной предельной теоремы, необходимых, чтобы сумма имела нормальный закон распределения вероятностей. В статье данная проблема решена для любых, заранее неизвестных законов распределения независимых слагаемых. Использован аппарат характеристических функций, представленных комплексным рядом Маклорена. Получена оценка погрешности такого представления. Выведено аналитическое выражение для плотности распределения средней выборочной наблюдений. Разработан алгоритм определения необходимого числа наблюдений в зависимости от точности оценки. Для пояснения работы алгоритма рассмотрен пример практической реализации разработанного метода. Результаты моделирования оценки необходимого числа слагаемых сведены в таблицу и представлены на графике.

Ключевые слова: сумма случайных величин, средняя выборочная, число слагаемых, характеристическая функция, ряд Маклорена, точность, погрешность, интеграл

Для цитирования: Ганичева А.В., Ганичев А.В. 2022. Оценивание числа слагаемых суммы независимых случайных величин при моделировании гауссовских случайных величин. Экономика. Информатика, 49(3): 546–557. DOI 10.52575/2687-0932-2022-49-3-546-557

Estimating the Number of Terms in the Sum of Independent Random Variables when Modeling Gaussian Random Variables

[1] Ganicheva A.V., [2] Ganichev A.V.
[1] Tver State Agricultural Academy, 7 Marshal Vasilevsky St, Tver, 170904, Russia
[2] Tver State Technical University, 22 AF. Nikitin Qy, Tver, 170026, Russia
E-mail: TGAN55@yandex.ru, alexej.ganichev@yandex.ru

Abstract. One of the most important problems in probability theory and statistics is the estimation of the number of terms of the central limit theorem necessary for the sum to have a normal probability distribution law. This problem becomes especially relevant for any previously unknown distribution laws of random variables. To solve the problem, the apparatus of characteristic functions is used. The characteristic function is represented by a complex Maclaurin series. The representation of the residual term of the series in the form of Cauchy is used. An estimate of the error of such a representation is obtained. An analytical expression is derived for the distribution density of the average sample of observations. An algorithm has been developed to determine the required number of observations depending on the accuracy of the assessment. To explain the operation of the algorithm, an example of the practical implementation of the developed method is shown. The obtained simulation results for estimating the required number of terms are summarized in a table. For clarity, the results are presented in the graph. The statements proved in the work can be used in multivariate data analysis, systems for diagnostics, monitoring, and statistical control of manufactured products.

Keywords: sum of random variables, average sample, number of terms, characteristic function, Maclaurin series, accuracy, error, integral

For citation: Ganicheva A.V., Ganichev A.V. 2022. Estimating the Number of Terms in the Sum of Independent Random Variables when Modeling Gaussian Random Variables. Economics. Information technologies, 49(3): 546–557 (in Russian). DOI 10.52575/2687-0932-2022-49-3-546-557

Введение

Центральная предельная теорема (ЦПТ) занимает особое место в теории вероятностей. Она лежит в основе решения множества прикладных задач в статистике [Chatterjee, Diaconis, 2017], науке об обработке данных [Draper, Guo, 2021] и составляет краеугольный камень современной статистики [Kwak, Kim, 2017.]. История ЦПТ обстоятельно изложена в книге [Fischer, 2011]. Показана связь истории ЦПТ с развитием теории вероятностей от ее классической до современной формы. В работе [Arras et al., 2020] отмечается многоаспектность данного направления научных исследований, например: исследование скорости сходимости в многомерной ЦПТ [Волгин, 2017], обобщение на случаи симметрических функций от случайных [Гринь, 2021] и гиперслучайных величин [Gorban, 2017], последовательностей случайных величин [Formanov et al., 2021; Senatov, 2007.], марковских цепей [Garet, 2021]. Статья [Roos, 2022] посвящена разработке нового метода характеристических функций, применяемого в ЦПТ.

Задача оценки числа слагаемых, при котором сумма независимых случайных величин имеет нормальный закон распределения вероятностей, решается в работе [Пименов, Тинаев, 2017] путем моделирования. Приведен пример, когда при трех слагаемых график плотности вероятности приближается к нормальному закону, а при более 30 слагаемых он близок к теоретическому нормальному. В статьях [Ганичева, 2020; 2022] получена теоретическая оценка числа слагаемых ЦПТ, необходимых, чтобы сумма имела нормальный закон распределения вероятностей при любых законах распределения слагаемых. В первом случае, когда слагаемые имеют одинаковые математические ожидания и одинаковые дисперсии, во втором случае – разные дисперсии.

Целью статьи является оценка числа n независимых слагаемых ЦПТ для получения нормального закона распределения вероятностей суммы и средней выборочной с заданной точностью ε при любых законах распределения слагаемых, а также получение алгоритмического описания зависимости n от ε.

Объекты и методы исследования

1. Характеристическая функция средней выборочной

Известно, что при сложении n независимых слагаемых случайных величин, имеющих одинаковые математические ожидания m_j и дисперсии σ_j^2, причем $m_{01} \le m_j \le m_{02}$, $\sigma_{01} \le \sigma_j \le \sigma_{02}$, сумма будет иметь нормальный закон распределения при достаточно большом n.

В данной работе рассмотрим вопрос о сумме и средней выборочной таких случайных величин. Докажем, что средняя выборочная будет иметь нормальное распределение при $n \ge n_0$, где n_0 определяется на основе разработанного алгоритма.

Найдем зависимость числа слагаемых n от точности ε. И, как следствие, будет получен нормальный закон распределения суммы этих величин с указанием соответствующего числа слагаемых для заданной точности.

Оценим n_0 на основе характеристических функций. Сначала рассмотрим непрерывные случайные величины.

Обозначим через $q_{x_j}\left(\frac{1}{n_0}t\right)=\int_{-\infty}^{\infty} e^{\frac{itx_j}{n_0}}\cdot f\left(x_j\right)dx_j$ случайной величины $\frac{X_j}{n_0}\left(j=\overline{1,n_0}\right)$. Так как случайные величины X_j независимы, то характеристическая функция случайной величины $\bar{x}$ будет:

$$q_{\bar{x}}\left(t\right)=\prod_{j=1}^{n_0} q_{x_j}\left(\frac{1}{n_0}t\right)=\prod_{j=1}^{n_0}\int_{-\infty}^{\infty} e^{\frac{itx_j}{n_0}}\cdot f\left(x_j\right)dx_j. \tag{1}$$

Представим функцию $q_{x_j}\left(\frac{1}{n_0}t\right)$ рядом Маклорена:

$$q_{x_j}\left(0\right)=1;\ \ q'_{x_j}\left(0\right)=\frac{i}{n_0}\int_{-\infty}^{\infty} x_j f\left(x_j\right)dx_j=\frac{i}{n_0}m_j.$$

Найдем вторую производную $q''_{x_j}\left(0\right)=\frac{i^2}{n_0^2}\int_{-\infty}^{\infty} x_j^2 f\left(x\right)dx=-\frac{1}{n_0^2}\sigma_j^2$ и остаточную сумму ряда $S_{ост}\left(t\right)=q_{x_j}\left(t\right)-1+\frac{\sigma_j^2}{2n_0^2}\cdot t^2-\frac{i}{n_0}m_j t.$ Так как $q_{x_j}\left(0\right)=1$, то $S_{ост}\left(0\right)=0$.

Следовательно, $q_{x_j}\left(t\right)=1-\frac{\sigma_j^2}{2n_0^2}\cdot t^2+\frac{i}{n_0}m_j t$. Тогда

$$q_{\bar{x}}\left(t\right)=\prod_{j=1}^{n_0}\left(1+\left(\frac{i}{n_0}m_j t-\frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)\right). \tag{2}$$

Прологарифмируем выражение (2):

$$\ln q_{\bar{x}}\left(t\right)=\sum_{j=1}^{n_0}\ln\left(1+\left(\frac{i}{n_0}m_j t-\frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)\right). \tag{3}$$

Ряд Маклорена от выражения (3) будет:

$$\ln q_{\bar{x}}\left(t\right)=\sum_{j=1}^{n_0}\sum_{l=0}^{\infty}\frac{(-1)^l}{l+1}\left(\frac{i}{n_0}m_j t-\frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)^{l+1}. \tag{4}$$

Пусть $z_j=\frac{i}{n_0}m_j t-\frac{\sigma_j^2}{2n_0^2}\cdot t^2$. Обозначим: $a_j=\frac{\sigma_j^2}{2n_0^2}\cdot t^2$, $b_j=\frac{m_j\cdot t}{n_0}$. Тогда $z_j=-a_j+ib_j$.

При этом

$$\left|z_j\right|=\left|-\frac{\sigma_j^2}{2n_0^2}\cdot t^2+\frac{i}{n_0}m_j t\right|<1,\ \ j=\overline{1,n_0}\text{, т. е.}$$

$$\frac{\sigma_j^4}{4n_0^4}\cdot t^4+\frac{{m_j}^2}{{n_0}^2}t^2<1,\ \ j=\overline{1,n_0}. \tag{5}$$

Отсюда $n_0^4-{m_j}^2t^2{n_0}^2-\frac{1}{4}\sigma_j^4t^4>0,\ \ j=\overline{1,n_0}$. Следовательно,

$$n_0 > \frac{t}{\sqrt{2}} \cdot \sqrt{{m_j}^2 + \sqrt{{m_j}^4 + \sigma_j^4}} \quad (j = \overline{1, n_0}). \tag{6}$$

Считаем, что $-3\sigma_{02} \le t \le 3\sigma_{02}$.

Тогда
$$n_0 > n_1 = 2{,}12\sigma_{02} \cdot \sqrt{{m_{02}}^2 + \sqrt{{m_{02}}^4 + \sigma_{02}^4}}. \tag{7}$$

Пусть $t^2\sigma_{02}^2 < 2n_0^2$, тогда

$$n_0 > n_2 = 2{,}12\sigma_{02}^2. \tag{8}$$

Из (7) и (8) имеем
$$n_0 \ge n_3 = \max\{n_2, n_1\}.$$

Поскольку ряд (4) равен сумме n_0 рядов, то погрешность суммарного ряда (4) равна сумме погрешностей слагаемых. При этом ряд будет комплексным. Модуль суммарной погрешности

$$|R_l| \le \sum_{j=1}^{n_0} \left(\frac{|z_j - c_j|}{r_j} \right)^l \cdot \frac{r_j \cdot M(r_j)}{r_j - |z_j - c_j|},$$

где $|z_j - c_j| = r_j$ – окружность K_j радиуса r_j с центром в точке c_j, $M(r_j)$ – верхняя граница $|f(z_j)|$ на K_j.

Для рассматриваемого случая $c_j = 0$, $f(z_j) = \ln(1 + z_j)$, $r_j = 1$, т. е.

$$|R_l| \le \left| \sum_{j=1}^{n_0} |z_j|^l \cdot \frac{M(1)}{1 - |z_j|} \right|. \tag{10}$$

Оценим $M(1)$.

Имеем: $f(z_j) = f(-a_j + ib_j) = \ln(1 + (-a_j + ib_j))$. При этом $f(z_j)$ можно представить в виде: $f(z_j) = x_j + iy_j$. Отсюда получаем:

$$\begin{cases} e^{x_j} \cdot \cos y_j = 1 - a_j, \\ e^{x_j} \cdot \sin y_j = b_j. \end{cases} \tag{11}$$

Следовательно, $tg\, y_j = \frac{b_j}{1 - a_j}$, $y_j = arctg \frac{b_j}{1 - a_j}$, т. е. $\frac{\sin y_j}{\cos y_j} = \frac{\sqrt{1 - \cos^2 y_j}}{\cos y_j} = \frac{b_j}{1 - a_j}$.

Отсюда находим: $\cos y_j = \frac{1 - a_j}{\sqrt{(1 - a_j)^2 + b_j^2}}$. Здесь учтено, что $a_j = \frac{\sigma_j^2}{2n_0^2} \cdot t^2 < 1$ при $n_0 \ge 2{,}12\sigma_{02}^2$ (условие (8)). Тогда с учетом первого уравнения из (11) получаем:

$$x_j = \frac{1}{2} \ln((1 - a_j)^2 + b_j^2).$$

Таким образом,

$$f(z_j) = \ln(1 + (-a_j + ib_j)) = \frac{1}{2} \ln(1 - a_j)^2 + b_j^2) + i \cdot arctg \frac{b_j}{1 - a_j}. \tag{12}$$

Фиксируем n_0 и подставим в (12) вместо a_j и b_j их выражения через n_0, m_j, σ_j^2 и t. Получим:

$$f(z_j) = \ln\left(1 - \frac{\sigma_j^2}{2n_0^2}\cdot t^2 + \frac{i}{n_0} m_j t\right) = \frac{1}{2}\ln\left(\left(1 - \frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)^2 + \frac{m_j^2 t^2}{n_0^2}\right) + i\cdot arctg\,\frac{m_j t}{n_0\left(1 - \frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)}.$$

При фиксированных n_0, m_j, σ_j^2 полученное выражение является функцией *t*. Находим:

$$\left|f(z_j)\right| = \sqrt{\frac{1}{4}\ln^2\left(\left(1 - \frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)^2 + \frac{m_j{}^2 t^2}{n_0^2}\right) + arctg^2\,\frac{m_j t}{n_0\left(1 - \frac{\sigma_j^2}{2n_0^2}\cdot t^2\right)}}. \quad (13)$$

Оценим $\left|f(z_j)\right|$. Заметим, что $\ln(1+x) \approx x$ с точностью $\frac{x^2}{2(1-\xi x)}$, где $0 < \xi < 1$. Здесь использовано представление остаточного члена в форме Коши:

$$r(x) = \frac{x^{k+1}}{k+1}\cdot\frac{(1-\xi)^k}{(1-\xi x)^{k+1}}.$$

При этом $\left(\frac{1-\xi}{1-\xi x}\right) < 1$, т. к. согласно (5), $x \le 1$ при $n_0 \ge n_1$, и тогда $\xi x < \xi$. Следовательно,

$$\ln((1-a_j)^2 + b_j^2) = \ln(1 + (a_j^2 + b_j^2 - 2a_j)) \approx a_j^2 + b_j^2 - 2a_j \quad (14)$$

с точностью $(a_j^2 + b_j^2 - 2a_j)^2 \cdot \frac{n_0}{2} < \varepsilon$ т.к. $\frac{1}{1-\xi x} < n_0$ при $n_0 > 2$ и $\xi x < 1$. Отсюда

$$\left|a_j^2 + b_j^2 - 2a_j\right| < \frac{\sqrt{2\varepsilon}}{\sqrt{n_0}}, \quad (15)$$

т. е. $-\frac{\sqrt{2\varepsilon}}{\sqrt{n_0}} < \frac{\sigma_j^4}{4n_0^4}t^4 - \left(\frac{\sigma_j^2}{n_0^2} - \frac{m_j{}^2}{n_0^2}\right)t^2 < \frac{\sqrt{2\varepsilon}}{\sqrt{n_0}}$,

$$\begin{cases} \sqrt{2\varepsilon}\cdot n_0^{3,5} + (\sigma_j^2 - m_j^2)\cdot n_0^2 t^2 - \frac{\sigma_j^4}{4}t^4 \ge 0, \\ \sqrt{2\varepsilon}\cdot n_0^{3,5} - (\sigma_j^2 - m_j^2)\cdot n_0^2 t^2 + \frac{\sigma_j^4}{4}t^4 \ge 0. \end{cases} \quad (16)$$

Система верна для любого $n_0 > 2$ и любого $j = \overline{1, n_0}$.

Получим оценку для n_0, при которой аргумент арктангенса будет не больше $\sqrt{\varepsilon}$. Имеем

$$x = 3\sigma_{02}m_{02} \Big/ n_0\left(1 - \frac{9\sigma_{02}^4}{2n_0^2}\right) < \sqrt{\varepsilon}, \quad (17)$$

здесь взято $3\sigma_{02}$ (максимальное значение *t*). Получим

$3\sigma_{02}m_{02} < \sqrt{\varepsilon}n_0\cdot\left(1 - \frac{9\sigma_{02}^4}{2n_0^2}\right)$, т.к. $1 - \frac{9\sigma_{02}^4}{2n_0^2} > 0$ при

$$n_0 > n_2 = 2{,}12\sigma_{02}^2. \quad (18)$$

Отсюда находим

$$n_0 > n_4 = \frac{3m_{02}\sigma_{02} + 3\sigma_{02}\sqrt{m_{02}^2 + 2\varepsilon\sigma_{02}^2}}{2\sqrt{\varepsilon}}. \quad (19)$$

Следовательно,

$$arctg\, x \le arctg\sqrt{\varepsilon} \approx \sqrt{\varepsilon} \tag{20}$$

с погрешностью $x^3/3 = \sqrt{\varepsilon}^3/3$ (ряд Лейбница) при $n_0 > n_4$.

Таким образом, $\left|f(z_j)\right| \le \sqrt{\frac{1}{4}\left(\frac{\sqrt{2\varepsilon}}{\sqrt{n_0}}\right)^2 + \varepsilon} = \sqrt{\varepsilon}\sqrt{\frac{1}{2n_0}+1} \le \sqrt{\varepsilon}\sqrt{\varepsilon+1}$ при $n_0 \ge n_6 = \max\{n_3, n_4, n_5\}$,

где

$$n_5 = \frac{1}{2\varepsilon}. \tag{21}$$

Получим оценку для $|R_l|$ на основе (14), (15), (20):

$$|R_l| \le \sum_{j=1}^{n_0} |z_j|^l \cdot \sqrt{\varepsilon}\sqrt{\varepsilon+1} \Big/ \left(1-|z_j|\right).$$

Отсюда

$$|R_l| \le n_0 \cdot \left(\frac{81\sigma_{02}^4}{4n_0^4} + \frac{9m_{02}^2 \cdot \sigma_{02}^2}{n_0^2}\right)^{l/2} \cdot \sqrt{\varepsilon}\sqrt{\varepsilon+1} \Bigg/ \left(1 - \sqrt{\frac{81\sigma_{02}^4}{4n_0^4} + \frac{9m_{02}^2 \cdot \sigma_{02}^2}{n_0^2}}\right). \tag{22}$$

Пусть в (22) при l=2 правая часть не превосходит ε, т. е. $|R_2| < \varepsilon$. Оценим соответствующее значение n_0. Из (22) имеем:

$$\left(\frac{81\sigma_{02}^8}{4n_0^3} + \frac{9m_{02}^2\sigma_{02}^2}{n_0}\right) \cdot \sqrt{\varepsilon+1} + \frac{\varepsilon^{0,5}}{\sqrt{n_0}}\sqrt{\frac{81\sigma_{02}^8}{4n_0^3} + \frac{9m_{02}^2\sigma_{02}^2}{n_0}} - \varepsilon^{0,5} \le 0.$$

Для краткости обозначим выражение в скобках первого слагаемого через V, т. е. $\sqrt{\varepsilon+1} \cdot V + \frac{\varepsilon^{0,5}}{\sqrt{n_0}}\sqrt{V} - \varepsilon^{0,5} \le 0$. Находим $\sqrt{V} \le \frac{-\varepsilon^{0,5}/\sqrt{n_0} + \sqrt{\varepsilon/n_0 + 4\sqrt{\varepsilon+1}\varepsilon^{0,5}}}{2\sqrt{\varepsilon+1}}$. Усилим последнее неравенство: $\sqrt{(\varepsilon+1)}\sqrt{V} \le -\frac{\varepsilon^{0,5}}{2\sqrt{n_0}} + \varepsilon^{0,5}$ и

$$(\varepsilon+1) \cdot \left(\frac{81\sigma_{02}^8}{2n_0^3} + \frac{9m_{02}^2\sigma_{02}^2}{n_0}\right) \le \left(1 - \frac{1}{2\sqrt{n_0}}\right)^2 \cdot \varepsilon^{0,25}.$$

Положим $d = \left(1 - \frac{1}{2\sqrt{n_0}}\right)^2 \cdot \varepsilon^{0,25}$. Тогда последнее неравенство преобразуется к виду:

$$2dn_0^3 - 18m_{02}^2\sigma_{02}^2 n_0^2(\varepsilon+1) - 81\sigma_{02}^8(\varepsilon+1) \ge 0. \tag{23}$$

Решая данное неравенство, находим сначала решение уравнения – некоторое n_7, тогда $n_0 \ge n_8 = \max\{n_6, n_7\}$. Таким образом, при $n_0 \ge n_8$ погрешность R_2 по абсолютной величине не превосходит ε, т. е. $|R_2| = \varepsilon_1 \le \varepsilon$. Тогда при $n_0 \ge n_8$ $\ln q_{\bar{x}}(t) = -\sum_{j=1}^{n_0}\left(-\frac{i}{n_0}m_j t + \frac{\sigma_j^2}{2n_0^2} \cdot t^2\right)$ с погрешностью R_2, т. е.

$$q_{\bar{x}}(t) = \exp\left(-\sum_{j=1}^{n_0}\left(-\frac{i}{n_0}m_j t + \frac{\sigma_j^2}{2n_0^2} \cdot t^2\right) + R_2\right). \tag{24}$$

В то же время в общем случае R_2 является комплексной величиной, зависящей от γ и $R_2 = \varepsilon_1 \cdot (\cos\gamma + i\sin\gamma)$, где γ – аргумент R_2, отсюда $|R_2 - i\varepsilon_1 \sin\gamma| = \varepsilon_1 \le \varepsilon$. Тогда и при максимальном значении $\varepsilon = 0{,}9$ можно считать, что $R_2 = i\varepsilon_1 \sin\gamma$ и $e^{R_2} = e^{\varepsilon_1 \cdot \cos\gamma} \cdot e^{i\varepsilon_1 \sin\gamma} = e^{i\varepsilon_1 \sin\gamma}$.

Пусть $X = \varepsilon_1 \cdot \sin\gamma$, тогда X – случайная величина и $q_{1x}(t) = M\left[e^{itX}\right]$ – характеристическая функция случайной величины X. Отсюда получаем:

$$q_{1x}(t) \approx 1 + itM[X] \text{ с точностью } \frac{\xi^2}{2} M\left[X^2\right],\ 0 < \xi < x.$$

Считаем, что угол γ распределен равномерно в промежутке $[-\pi/2, \pi/2]$.

Для примера рассмотрим этот случай, тогда $M[X] = \frac{\varepsilon_1}{\pi} \int\limits_{-\pi/2}^{\pi/2} \sin\gamma d\gamma = 0;$ $M[X^2] = D_x = \frac{\varepsilon_1^2}{\pi} \int\limits_{-\pi/2}^{\pi/2} \sin^2 \gamma d\gamma = 0.$ Таким образом, погрешность $\frac{\xi^2}{2} M[X^2]$ равна нулю.

Следовательно, $q_{1x}(t) = 1,\ R_2 = 0$ и $e^{R_2} = 1$. Рассмотрен случай, когда γ имеет равномерное распределение, но такой же результат будет и при других распределениях $f_1(\gamma)$, т. к.

$$M[X] = \int\limits_{-\pi/2}^{\pi/2} \sin\gamma \cdot f_1(\gamma) d\gamma = \begin{vmatrix} u = \sin\gamma,\ du = \cos\gamma d\gamma, \\ v = \int f_1(\gamma) d\gamma \end{vmatrix} =$$

$$= \sin\gamma \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cdot \int\limits_{-\pi/2}^{\pi/2} f_1(\gamma) d\gamma - \int\limits_{-\pi/2}^{\pi/2} \cos\gamma d\gamma \cdot \int\limits_{-\pi/2}^{\pi/2} f_1(\gamma) d\gamma = 0.$$

Аналогично можно показать, что $M[X^2] = 0$. Следовательно, в (24) $R_2 = 0$.

2. Плотность распределения средней выборочной. Оценка числа слагаемых

Подвергнем $q_{\bar{x}}(t)$ обратному преобразованию Фурье, получим:

$$f(\bar{x}) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-it\bar{x}} \cdot q_{\bar{x}}(t) dt. \tag{25}$$

На основе (24) имеем:

$$f(\bar{x}) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-it\bar{x}} \cdot e^{-\sum\limits_{j=1}^{n_0} \left(-\frac{i}{n_0} m_j t + \frac{\sigma_j^2}{2n_0^2} \cdot t^2 \right)} dt. \tag{26}$$

Используем известную формулу, сводящуюся к интегралу Эйлера – Пуассона: $\int\limits_{-\infty}^{\infty} e^{-Ax \pm 2Bx - C} dx = \sqrt{\frac{\pi}{A}} \cdot e^{-\frac{AC - B^2}{A}}$. Тогда

$$f(\bar{x}) = \frac{1}{\sqrt{2\pi} \sqrt{\sum\limits_{j=1}^{n_0} \frac{\sigma_j^2}{2n_0^2}}} e^{-\sum\limits_{j=1}^{n_0} \left(\bar{x} - \frac{m_j}{n_0} \right)^2 \Big/ \sum\limits_{j=1}^{n_0} \frac{\sigma_j^2}{n_0^2}}. \tag{27}$$

Заметим, что при вычислении $f(\bar{x})$ используется несобственный интеграл при $t \in (-\infty, \infty)$. Однако $t \in [-3\sigma_{02}, 3\sigma_{02}]$. Покажем, что с любой заданной точностью ε при соответствующем n_0 интеграл (26) в пределах от $-\infty$ до ∞ будет отличаться на ε от соответствующего интеграла с границами от $-3\sigma_{02}$ до $3\sigma_{02}$.

Из (26) имеем: $f(\bar{x}) = \dfrac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-t^2\sum\limits_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}+it\sum\limits_{j=1}^{n_0}\left(\frac{m_j}{n_0}-\bar{x}\right)}dt$.

Преобразуем подынтегральное выражение. Выделим в показателе полный квадрат:

$$-\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}\left(t^2-2i\frac{t}{2}\sum_{j=1}^{n_0}\left(\frac{m_j}{n_0}-\bar{x}\right)\Big/\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}-\frac{1}{4}\sum_{j=1}^{n_0}\left(\frac{m_j}{n_0}-\bar{x}\right)^2\Big/\sum_{j=1}^{n_0}\left(\frac{\sigma_j^2}{2n_0^2}\right)^2+\right.$$

$$\left.+\frac{1}{4}\sum_{j=1}^{n_0}\left(\frac{m_j}{n_0}-\bar{x}\right)^2\Big/\sum_{j=1}^{n_0}\left(\frac{\sigma_j^2}{2n_0^2}\right)^2\right)=-\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}\left(t-\frac{i}{2}\left(\frac{m_j}{n_0}-\bar{x}\right)\Big/\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}\right)^2-\frac{1}{4}\left(\sum_{j=1}^{n_0}\frac{m_j}{n_0}-\bar{x}\right)^2\Big/\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}.$$

Обозначим: $a=\sum\limits_{j=1}^{n_0}\dfrac{\sigma_j^2}{2n_0^2}$, $b=\sum\limits_{j=1}^{n_0}\dfrac{m_j}{n_0}-\bar{x}$. Отсюда получаем:

$$f(\bar{x})=\frac{1}{2\pi}e^{-\frac{1}{4a}b^2}\int\limits_{-\infty}^{\infty}e^{-a\left(t-i\frac{b}{a}\right)^2}dt. \tag{28}$$

Покажем, что с точностью $\varepsilon/2$ интеграл (28) при соответствующем n_0 в пределах от $3\sigma_{02}$ до ∞ будет сколь угодно мало отличаться от 0 (на величину $\varepsilon/2$).

Положим $u=\left(t-i\dfrac{b}{a}\right)\sqrt{a}$, тогда $du=\sqrt{a}dt$. Пусть $b_1=\left(3\sigma_{02}-i\dfrac{b}{a}\right)\sqrt{a}$. Имеем: $\int\limits_{3\sigma_{02}}^{\infty}e^{-a\left(t-\frac{i}{2}\frac{b}{a}\right)^2}dt=\dfrac{1}{\sqrt{a}}\int\limits_{b_1}^{\infty}e^{-u^2}du$. Для вычисления последнего интеграла возведем его в квадрат, затем перейдем к двойному интегралу в полярных координатах:

$$\left(\int\limits_{b_1}^{\infty}e^{-u^2}du\right)^2=\int\limits_{b_1}^{\infty}e^{-u^2}du\cdot\int\limits_{b_1}^{\infty}e^{-v^2}dv=\int\limits_{0}^{\pi}d\varphi\cdot\int\limits_{|b_1|}^{\infty}r\cdot e^{-r^2}dr,$$

где $r^2=u^2+v^2$,

т. е. $\dfrac{1}{\sqrt{a}}\int\limits_{b_1}^{\infty}e^{-u^2}du=\sqrt{\dfrac{\pi}{2a}}\cdot e^{-\frac{1}{2}|b_1|^2}$. Потребуем, чтобы этот интеграл не превосходил $\dfrac{\varepsilon}{2}$ и оценим соответствующее n_0.

Отсюда
$$-\frac{1}{2}|b_1|^2\le\ln\left(\frac{\varepsilon}{2}\Big/\sqrt{\frac{\pi}{2a}}\right), \tag{29}$$

или

$$9\sigma_{02}^2\cdot\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}+\left(\sum_{j=1}^{n_0}\frac{m_j}{n_0}-\bar{x}\right)^2\Big/\sum_{j=1}^{n_0}\frac{\sigma_j^2}{2n_0^2}\ge-2\ln\left(\frac{\varepsilon}{2\sqrt{\pi}}\sqrt{\sum_{j=1}^{n_0}\frac{\sigma_j^2}{n_0^2}}\right).$$

Усилим последнее неравенство:

$$4,5\sigma_{02}^2\cdot\sigma_{01}^2+2n_0^2\left(m_{01}^2-2m_{02}\bar{x}+\bar{x}^2\right)\Big/\sigma_{02}^2+2n_0\ln\frac{\varepsilon\cdot\sigma_{02}}{2\sqrt{\pi}\sqrt{n_0}}\ge 0. \tag{30}$$

С учетом масштабирования (полагаем $\sigma_{02}=1$), сдвига по горизонтальной оси (полагаем $m_{02}=0$) последнее неравенство можно усилить следующим образом:

$$2n_0\cdot m_{01}^2\ge\ln n_0-2\cdot\ln\varepsilon+2\cdot\ln 3,544. \tag{31}$$

Так, при минимальном значении $\varepsilon=0,002$ неравенство (31) будет иметь вид;

$$2n_0\cdot m_{01}^2\ge\ln n_0-2\cdot\ln 0,00056,$$

т. е.

$$2n_0 \cdot m_{01}^2 \ge \ln n_0 + 14{,}9751 .$$

Из (31) для данных m_{01} и ε находим соответствующее значение n_9 – решение (31), рассматриваемого как равенство. Тогда при $n_0 \ge n_9$ интеграл (28) в границах от $3\sigma_{02}$ до ∞ будет сколь угодно мало отличаться от 0 (на $\frac{\pi}{2}$). Аналогичное утверждение будет верно для этого интеграла и в границах от $-\infty$ до $-3\sigma_{02}$. Следовательно, при $n_0 \ge n_9$ интеграл (26) будет равен соответствующему интегралу с границами от $-3\sigma_{02}$ до $3\sigma_{02}$ с точностью $\varepsilon > 0$.

Итак, доказано следующее утверждение.

Теорема. Пусть имеется *n* независимых случайных величин, математические ожидания m_j $(j = \overline{1,n})$ которых попадают в интервал (m_{01}, m_{02}), средние квадратические отклонения попадают в интервал $(\sigma_{01}, \sigma_{02})$. Тогда для любого $\varepsilon > 0$ найдется $n_0 \ge \max\{n_8, n_9\}$, определяемое с учетом (7), (8), (19), (21), (23), (30), (31) и такое, что при $n \ge n_0$ случайная величина $\overline{x}$ будет иметь нормальное распределение с математическим ожиданием $\sum_{j=1}^{n} \frac{m_j}{n}$ и средним квадратическим отклонением $\sqrt{\sum_{j=1}^{n} \frac{\sigma_j^2}{2n^2}}$ с погрешностью ε.

Следствие 1. В рамках доказанной теоремы следует, что сумма *n* случайных независимых величин будет иметь нормальный закон распределения, т. к. $Y = n\overline{x}$ имеет нормальное распределение при $n \ge n_0$.

Следствие 2. В случае *n* независимых дискретных случайных величин для каждой из них с любой заданной степенью точности строится эконометрическая модель, при этом число значений каждой случайной величины должно быть не менее 7. Затем применяется изложенный в подразделе 1 метод для непрерывных случайных величин.

Результаты и их обсуждение

Рассмотрим конкретный пример. Пусть $\sigma_{01} = 0{,}8$, $\sigma_{02} = 1$, $\varepsilon = 0{,}1$, $m_{01} = -0{,}8$, $m_{02} = 0$, $m = 6$.

Алгоритм построения $f(\overline{x})$ заключается в последовательности шагов (справа указан номер формулы).

1. $n_1 = 2{,}12\sigma_{02}\sqrt{m_{02}^2 + \sqrt{m_{02}^4 + \sigma_{02}^4}}$. (7)

Для рассматриваемого примера имеем: $n_1 = 2{,}12$.

2. $n_2 = 2{,}12\sigma_{02}$, (8)

т. е. $n_2 = 2{,}12$ – для рассматриваемого примера.

3. $n_3 = \max\{n_1, n_2\}$, т. е. $n_3 = 2{,}12$.

4. $n_4 = 2{,}12\sigma_{02}^2$, $n_4 = 2{,}12$. (19)

5. $n_5 = \dfrac{1}{2\varepsilon}$, $n_5 = 5$ (21)

6. $n_6 = \max\{n_3, n_4, n_5\}$, $n_6 = 5$.

7. n_7 находится из неравенства (23):

$$2dn_0^3 - 18m_{02}^2\sigma_{02}^2 n_0^2(\varepsilon + 1) - 81\sigma_{02}^8(\varepsilon + 1) \ge 0,$$

где $d=\left(1-1/2\sqrt{n_0}\right)^2\cdot\varepsilon^{0,25}$. Для рассматриваемого примера $n_7=5$

8. $n_8=\max\{n_6,n_7\}=\max\{5,5\}=5.$

9. n_9 определяется из неравенства (31):

$$2n_0\cdot m_{01}^2\geq\ln n_0-2\cdot\ln\varepsilon+2\cdot\ln 3{,}544,\ n_9=5.$$

10. $n_0=\max\{n_8,n_9\}=\max\{5,5\},\ n_0=5$.

11. Построим таблицу (см. таблицу) и график зависимости n_0 от ε (рисунок).

Точность ε и количество шагов n_0
Accuracy ε and number of steps n_0

ε	0,002	0,004	0,005	0,01	0,02	0,05	0,07	0,08	0,1	0,5
n_0	250	125	100	50	25	10	7	6	5	5

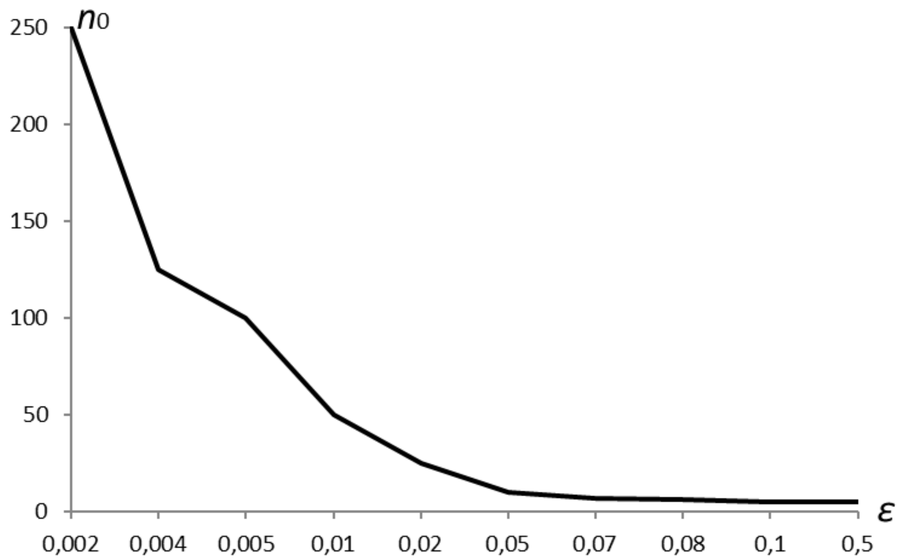

График зависимости n_0 от ε
Graph of dependence n_0 on ε

Изображенная на рисунке зависимость аналогична зависимостям, полученным в статьях [Ганичева, 2020; Ганичева, 2022] для нормального закона распределения вероятностей. Таким образом, настоящее исследование является обобщением результатов, полученных в [Ганичева, 2020; Ганичева, 2022] при любых законах распределения слагаемых ЦПТ.

Заключение

Основные результаты работы:

1) получено аналитическое выражение для плотности распределения средней выборочной любых законов распределения;

2) разработан алгоритм определения числа слагаемых ЦПТ в зависимости от точности оценки;

3) приведена практическая реализация разработанного метода;

4) результаты моделирования проиллюстрированы графиком.

Список литературы

Волгин А.В. 2017. Улучшение оценки скорости сходимости в многомерной центральной предельной теореме для сумм локально зависимых случайных векторов. Прикладная дискретная математика, 36: 13–24.

Ганичева А.В. 2022. Оценка числа слагаемых нормальной аппроксимации сумм независимых случайных величин. Вестник Бурятского государственного университета. Математика, информатика, 1: 26–34. DOI: 10.18101/2304-5728-2022-1-26-34.

Ганичева А.В. 2020. Оценка числа слагаемых центральной предельной теоремы. Прикладная математика и вопросы управления, 4: 7–19. DOI: 10.15593/2499-9873/2020.4.01.

Гринь А.Г. 2021. О центральной предельной теореме с нелинейной масштабной нормировкой. Математические структуры и моделирование, 4 (60): 9–16. DOI: 10.24147/2222-8772.2021.4.9-16.

Пименов С.Ю., Тинаев В.В. 2017. Применение центральной предельной теоремы для компьютерного моделирования случайных сигналов. Наука и образование: новое время, 2 (19): 227–231.

Arras B., Breton J.-C., Aurelia Deshayes A., Durieu O., Lachièze-Rey R. 2020. Some recent advances for limit theorems. ESAIM Proceedings and Surveys, 68: 73–96. DOI:10.1051/proc/202068005.

Chatterjee S., Diaconis P. 2017. A central limit theorem for a new statistic on permutations. Indian J. Pure Appl. Math., 48(4): 561–573. DOI: 10.1007/s13226-017-0246-3.

Draper D., Guo E. 2021. The Practical Scope of the Central Limit Theorem. Other Statistics: 47. DOI: https://doi.org/10.48550/arXiv.2111.12267. Corpus ID: 244527194.

Fischer H. 2011. A History of the Central Limit Theorem From Classical to Modern Probability Theory. Springer Science+Business Media, LLC: 415 p. DOI 10.1007/978-0-387-87857-7.

Formanov S., Khusainova B., Sirozhitdinov A. 2021. On the numerical characteristics in the central limit theorem. AIP Conference Proceedings 2365: 060011. DOI: 10.1063/5.0058101.

Garet O. 2021. A central limit theorem for the number of descents and some urn models. Markov Processes And Related Fields, Polymat Publishing Company, 27 (5): 789–801.

Gorban I.I. 2017. The central limit theorem/ The Statistical Stability Phenomenon: 261–270. DOI:10.1007/978-3-319-43585-5_19.

Kwak S.G., Kim J.H. 2017. Central limit theorem: the cornerstone of modern statistics. Korean journal of anesthesiology. 70(2): 144. DOI:10.4097/kjae.2017.70.2.144.

Roos B. 2022. On the accuracy in a combinatorial central limit theorem: the characteristic function method, 67 (1): 150–175. DOI: 10.4213/tvp5412.

Senatov V.V. 2007. On Asymptotic Expansions in the Central Limit Theorem with Explicit Estimates of Remainder Terms Theory of Probability and Its Applications, 51(4): 729–736. DOI: 10.1137/S0040585X9798275X.

References

Volgin A.V. 2017. Uluchshenie ocenki skorosti shodimosti v mnogomernoj central'noj predel'noj teoreme dlja summ lokal'no zavisimyh sluchajnyh vektorov [Improving the rate of convergence in the multidimensional central limit theorem for sums of locally dependent random vectors]. Prikladnaja diskretnaja matematika [Applied Discrete Mathematics], 36: 13–24.

Ganicheva A.V. 2022. Estimation of the number of terms of the normal approximation of the sums of independent random variables. BSU bulletin. Mathematics, Informatics, 1: 26–34 (in Russian). DOI: 10.18101/2304-5728-2022-1-26-34.

Ganicheva A.V. 2020. Estimation of the number of summands of the central limit theorem. Applied Mathematics and Control Sciences, 4: 7–19 (in Russian). DOI: 10.15593/2499-9873/2020.4.01.

Grin A.G. 2021. On the central limit theorem with nonlinear scale normalization. Mathematical Structures and Modeling, 4 (60): 9–16. DOI: 10.24147/2222-8772.2021.4.9-16.

Pimenov S.Ju., Tinaev V.V. 2017. Primenenie central'noj predel'noj teoremy dlja komp'juternogo modelirovanija sluchajnyh signalov [Application of the central limit theorem for computer simulation of random signals]. Nauka i obrazovanie: novoe vremja [Science and Education: Modern Times], 2(19): 227–231.

Arras B., Breton J.-C., Aurelia Deshayes A., Durieu O., Lachièze-Rey R. 2020. Some recent advances for limit theorems. ESAIM Proceedings and Surveys, 68: 73-96. DOI:10.1051/proc/202068005.

Chatterjee S., Diaconis P. 2017. A central limit theorem for a new statistic on permutations. Indian J. Pure Appl. Math., 48(4): 561–573. DOI: 10.1007/s13226-017-0246-3.

Draper D., Guo E. 2021. The Practical Scope of the Central Limit Theorem. Other Statistics: 47. DOI: https://doi.org/10.48550/arXiv.2111.12267. Corpus ID: 244527194.

Fischer H. 2011. A History of the Central Limit Theorem From Classical to Modern Probability Theory. Springer Science+Business Media, LLC: 415 p. DOI 10.1007/978-0-387-87857-7.

Formanov S., Khusainova B., Sirozhitdinov A. 2021. On the numerical characteristics in the central limit theorem. AIP Conference Proceedings 2365: 060011. DOI: 10.1063/5.0058101.

Garet O. 2021. A central limit theorem for the number of descents and some urn models. Markov Processes And Related Fields, Polymat Publishing Company, 27 (5): 789–801.

Gorban I.I. 2017. The central limit theorem/ The Statistical Stability Phenomenon: 261–270. DOI:10.1007/978-3-319-43585-5_19.

Kwak S.G., Kim J.H. 2017. Central limit theorem: the cornerstone of modern statistics. Korean journal of anesthesiology. 70 (2): 144. DOI:10.4097/kjae.2017.70.2.144.

Roos B. 2022. On the accuracy in a combinatorial central limit theorem: the characteristic function method, 67 (1): 150–175. DOI: 10.4213/tvp5412.

Senatov V.V. 2007. On Asymptotic Expansions in the Central Limit Theorem with Explicit Estimates of Remainder Terms Theory of Probability and Its Applications, 51(4): 729–736. DOI: 10.1137/S0040585X9798275X.

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.
Conflict of interest: no potential conflict of interest related to this article was reported.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Ганичева Антонина Валериановна, кандидат физико-математических наук, доцент, кафедра физико-математических дисциплин и информационных технологий, Тверская государственная сельскохозяйственная академия, Тверь, Россия

ORCID: 0000-0002-0224-8945

Ганичев Алексей Валерианович, доцент, кафедра информатики и прикладной математики, Тверской государственный технический университет, Тверь, Россия

ORCID: 0000-0003-3389-7582

INFORMATION ABOUT THE AUTHORS

Antonina V. Ganicheva, Candidate in Physical and Mathematical Sciences, Associate Professor, Department Physical and Mathematical Disciplines and Information Technology, Tver State Agricultural Academy, Tver, Russian Federation

Aleksey V. Ganichev, Associate Professor, Department of Informatics and Applied Mathematics, Tver State Technical University, Tver, Russian Federation