

ОЦЕНКА ОТКЛОНЕНИЯ ОТ НОРМАЛЬНОГО ЗАКОНА В ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЕ

Бессарабская Ирина Эдуардовна

кандидат технических наук, доцент, МИРЭА —

Российский Технологический университет

irina.bessarabskaya@gmail.com

ESTIMATION OF THE DEVIATION FROM THE NORMAL LAW IN THE CENTRAL LIMIT THEOREM

I. Bessarabskaya

Summary. This work investigates the rate of convergence in the Central Limit Theorem for sums of independent and identically distributed random variables. Numerical experiments are conducted for the uniform, exponential, and beta distributions to evaluate the Kolmogorov distance between the distribution of the standardized sum and the standard normal law for various numbers of summands. A logarithmic regression approach is used to estimate the convergence order and an empirical analogue of the Berry–Esseen constant. The obtained estimates are consistent with the leading terms of the Edgeworth expansion and demonstrate the influence of the asymmetry of the underlying distribution on the rate of normal approximation. The proposed approach provides quantitative characteristics of the asymptotic behavior of distributional convergence and can be applied to assess the quality of normal approximation for a wide class of distributions.

Keywords: central limit theorem, edgeworth expansion, berry–esseen inequality, continuous distributions, rate of convergence.

Аннотация. В работе исследуется скорость сходимости суммы независимых одинаково распределённых случайных величин к нормальному распределению в рамках Центральной предельной теоремы. Для равномерного, показательного и бета-распределений выполнено численное моделирование оценки Колмогоровского расстояния между распределением стандартизированной суммы и нормальным законом при различных значениях числа слагаемых. Использован анализ логарифмических зависимостей, позволяющий оценить порядок сходимости и аналог константы Берри–Эссеена. Полученные оценки согласуются с ведущими членами разложения Эджворта и демонстрируют влияние асимметрии исходного распределения на скорость приближения к нормальной аппроксимации. Предложенный подход позволяет получать количественные характеристики асимптотического поведения распределений сумм и может быть применён для анализа качества нормальной аппроксимации в широком классе распределений.

Ключевые слова: центральная предельная теорема, разложение Эджворта, неравенство Берри–Эссеена, непрерывные распределения, скорость сходимости.

Введение

Центральная предельная теорема (ЦПТ) является одним из ключевых результатов теории вероятностей, обеспечивая фундаментальную основу для нормальной аппроксимации сумм независимых одинаково распределённых случайных величин. Однако ЦПТ не предоставляет оценок скорости сходимости, тогда как для практических задач статистического моделирования такая информация является весьма важной.

Существуют два основных теоретических инструмента, позволяющих анализировать величину отклонения стандартизированной суммы случайных величин от нормального закона. Первым является неравенство Берри–Эссеена [1,2], которое даёт верхнюю оценку для колмогоровского расстояния между распределением суммы и нормальным распределением. Эта оценка выражается через третий абсолютный момент исходного распределения и константу Берри–Эссеена, точное значение которой до сих пор неизвестно. Современные теоретические оценки [3] дают диапазон возможных значений константы Берри–Эссеена: $0.409 \leq C \leq 0.4748$.

Вторым инструментом является разложение Эджворта [4], уточняющее структуру отклонений стандартизированной суммы случайных величин от нормального закона и связывающее скорость сходимости с коэффициентами асимметрии и эксцесса распределения.

Численные исследования скорости сходимости в ЦПТ позволяют получить оценки аналога константы, присутствующей в неравенстве Берри–Эссеена, и оценить скорость сходимости для конкретных распределений. Однако такое моделирование проводилось ограниченно. Для симметричных распределений, например равномерного или Бернулли при $p = 0.5$ известны отдельные вычислительные результаты. В работах [5,6] исследования равномерного распределения ограничиваются визуализацией сходимости и не содержат систематического регрессионного анализа зависимости колмогоровского расстояния от числа слагаемых n . Для несимметричных распределений подобных данных тоже немного. Например, в работе [7] представлены оценки для распределения Бернулли, включая случаи сильной асимметрии. В статье [8] исследованы выборочные средние для гамма-распределения. В целом, до настоящего

времени численный анализ скорости сходимости в ЦПТ остается ограниченным.

Цель настоящей работы — численная оценка скорости сходимости стандартизованных сумм для ряда непрерывных распределений: равномерного, показательного и бета-распределений с различной степенью асимметрии. Рассматриваются зависимости колмогоровского расстояния от числа суммируемых случайных величин, а также вычисляются численные аналоги константы Берри–Эссеена.

Для анализа используется метод логарифмических зависимостей, позволяющий исследовать асимптотический характер сходимости и получить количественные оценки порядка скорости сходимости и соответствующих констант в моделируемых зависимостях. Такой подход позволяет выделить поведение, соответствующее первому члену разложения Эджворта.

Полученные результаты расширяют существующие численные данные для классических непрерывных распределений и конкретизируют применение разложения Эджворта в случаях асимметричных распределений. Показано, что метод логарифмической регрессии $\ln(D_n) - \ln(n)$ является эффективным инструментом для оценки порядка сходимости и соответствующих констант в ЦПТ.

Особое внимание уделено показательному распределению, для которого ранее приводились численные оценки константы Берри–Эссеена, полученные как максимум констант для различных n . Такие оценки, как правило, завышены из-за существенного вклада в результат неасимптотических членов, шума и случайных выбросов. В настоящей работе используется логарифмическая регрессия, позволяющая получать значения, отражающие асимптотический порядок сходимости; для показательного распределения такие оценки оказываются существенно меньшими и согласуются с ведущим членом разложения Эджворта.

Методы анализа и данные

Математическая модель

Пусть X_1, X_2, \dots, X_n — независимые и одинаково распределённые случайные величины с математическим ожиданием μ , дисперсией $\sigma^2 > 0$ и конечным третьим центральным моментом. Определим сумму этих случайных величин и её стандартизованную форму:

$$S_n = X_1 + \dots + X_n, Z_n = (S_n - n\mu)/(\sigma\sqrt{n}). \quad 1)$$

Центральная предельная теорема утверждает, что Z_n сходится по распределению к стандартному нормальному закону $N(0,1)$, однако не даёт информации о скорости

этой сходимости. Эту информацию дают неравенство Берри–Эссеена и разложение Эджворта.

Неравенство Берри–Эссеена имеет вид:

$$\sup_x |F_n(x) - \Phi(x)| \leq (C\rho_{abs})/\sqrt{n}, \quad 2)$$

где $F_n(x)$ — функция распределения вероятностей стандартизованной суммы n независимых случайных величин, $\Phi(x)$ — стандартная нормальная функция распределения вероятностей, $\rho_{abs} = M(|X - M(X)|^3) / \sigma^3$, а C — константа Берри–Эссеена. Для константы C неизвестно точное значение. Обозначим:

$$D_n = \sup_x |F_{Z_n}(x) - \Phi(x)|, \quad 3)$$

где $F_{Z_n}(x)$ — функция распределения вероятностей стандартизованной суммы случайных величин Z_n . D_n — оценка величины расхождения двух распределений по Колмогорову. Тогда можно записать: $D_n \leq (C\rho_{abs}) / \sqrt{n}$. Таким образом, неравенство Берри–Эссеена устанавливает верхнюю границу отклонения распределения стандартизованной суммы независимых случайных величин от стандартного нормального закона и не зависит от того, какое распределение имеют суммируемые случайные величины.

Ф.И. Эджвортом [3] было введено разложение, названное его именем, которое описывает не только предельную форму распределения суммы случайных величин, но и структуру отклонений распределения такой суммы от нормального.

Пусть μ_3 и μ_4 — третий и четвертый центральные моменты соответственно исходного распределения. Обозначим $\gamma_1 = \frac{\mu_3}{\sigma^3}$ — коэффициент асимметрии, $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$ — коэффициент эксцесса.

Согласно разложению Эджворта плотность распределения вероятностей стандартизованной суммы случайных величин, $f_{Z_n}(x)$ может быть представлена в виде ряда по степеням $1/\sqrt{n}$:

$$f_{Z_n}(x) = \varphi(x) \left(1 + \frac{P_1(x)}{\sqrt{n}} + \frac{P_2(x)}{n} \right) + o\left(\frac{1}{n}\right). \quad 4)$$

Здесь $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in (-\infty; +\infty)$ — плотность стандартного нормального распределения, $P_1(x)$, $P_2(x)$ — некоторые многочлены.

Первый многочлен, $P_1(x)$, выражается через коэффициент асимметрии γ_1 :

$$P_1(x) = (\gamma_1/6)(x^3 - 3x). \quad 5)$$

Второй многочлен $P_2(x)$ зависит от коэффициентов асимметрии и эксцесса:

$$P_2(x) = (\gamma_2/24)(x^4 - 6x^2 + 3) + (\gamma_1^2/72)(x^6 - 15x^4 + 45x^2 - 15). \quad 6)$$

Функция распределения вероятностей стандартизованной суммы случайных величин Z_n выражается аналогично:

$$F_{Z_n}(x) = \Phi(x) - \varphi(x) \frac{\gamma_1(x^2 - 1)}{6\sqrt{n}} - \varphi(x) \left(\frac{\gamma_2(x^3 - 3x)}{24n} + \frac{\gamma_1^2(x^5 - 10x^3 + 15x)}{72n} \right) + o\left(\frac{1}{n}\right) \quad 7)$$

Если $\gamma_1 \neq 0$ (распределение несимметричное), то главный член поправки имеет порядок $1/\sqrt{n}$, и скорость сходимости распределения к нормальному также порядка $O(1/\sqrt{n})$. Если же $\gamma_1 = 0$, но $\gamma_2 \neq 0$, то следующий член разложения становится ведущим, сходимость ускоряется и становится порядка $O(1/n)$. Этим объясняется быстрая сходимость для симметричных распределений. Для симметричных распределений $D_n \sim b/n$, где b — некоторая константа, так же, как и C , не зависящая от n .

Таким образом, разложение Эджворта связывает асимптотическую скорость сходимости с асимметрией и коэффициентом эксцесса исходного распределения, а неравенство Берри-Эссеена даёт верхнюю не асимптотическую оценку этой скорости.

Оценим с помощью разложения Эджворта скорость сходимости суммы показательного распределённых случайных величин к нормальному. Запишем разложение (7) в виде разности $F_{Z_n}(x)$ и $\Phi(x)$, оставив в разложении член порядка $O(1/\sqrt{n})$:

$$F_{Z_n}(x) - \Phi(x) = \varphi(x) \frac{\gamma_1(x^2 - 1)}{6\sqrt{n}} + O\left(\frac{1}{n}\right). \quad 8)$$

$\max_x |F_{Z_n}(x) - \Phi(x)|$ достигается при $x=0$ так как $\varphi(x)|x^2 - 1| \leq \varphi(0)$. Отсюда получаем, что $\max_x |F_{Z_n}(x) - \Phi(x)| \leq \frac{\gamma_1}{6\sqrt{2\pi n}} + O\left(\frac{1}{n}\right) \approx \frac{0.0665\gamma_1}{\sqrt{n}} + O\left(\frac{1}{n}\right)$.

Для показательного распределения коэффициент асимметрии $\gamma_1 = 2$. Подставляя величину $\gamma_1 = 2$ в полученное выше неравенство, получаем:

$$\max_x |F_{Z_n}(x) - \Phi(x)| \leq \frac{0.133}{\sqrt{n}}. \quad 9)$$

Заметим, что данная оценка получена на основе первого члена разложения Эджворта и, таким образом, является асимптотической оценкой, справедливой при больших n . Разложение Эджворта не является строгим неасимптотическим неравенством. Поэтому данная оценка используется в статье как теоретический ориентир для сопоставления с результатами численного моделирования. Вывод строгого неасимптотического неравенства выходит за рамки настоящей работы.

Методика численного моделирования

Целью описываемого далее численного эксперимента являлась оценка скорости сходимости к нормальному закону сумм независимых одинаково распределённых случайных величин с различной степенью асимметрии исходного распределения, а именно: равномерного, показательного и бета-распределений, а также оценка численного аналога константы Берри-Эссеена. Под аналогом константы Берри-Эссеена понимается коэффициент в зависимостях вида $D_n \sim C_{\text{abs}}/n^\alpha$.

Для каждого рассматриваемого распределения моделировалась серия значений Z_n стандартизованной суммы n независимых одинаково распределённых случайных величин. Далее количество моделируемых значений суммарной случайной величины обозначается через M . Для каждого n вычислялось значение D_n . Для всех n и M эксперимент повторялся с различными начальными условиями генератора случайных чисел для контроля устойчивости результатов.

Следует учесть, что определение $D_n = \sup_x |F_{Z_n}(x) - \Phi(x)|$ справедливо для теоретического значения D_n . Однако, в численном эксперименте оценивается приближение этого значения, с помощью эмпирической функции распределения, построенной по M смоделированным значениям суммарной случайной величины. Таким образом, смоделированная величина D_n является случайной величиной, ошибка которой имеет порядок $\frac{1}{\sqrt{M}}$. Следовательно, масштаб статистического шума можно оценить как $O(1/\sqrt{M})$.

Для суммарных распределений строились следующие графики.

1. Гистограммы смоделированной плотности распределения стандартизованной суммы Z_n , приведенная на одном графике с кривой, соответствующей плотности стандартного нормального распределения.
2. Q-Q-график. На этих графиках сравнивались квантили эмпирического распределения сумм S_n

с квантилями стандартного нормального распределения.

По оси X отложены теоретические квантили, по оси Y — эмпирические квантили. Чем ближе к прямой $y = x$ расположены точки Q–Q-графика, тем ближе смоделированное распределение к нормальному. Единицы измерения на Q–Q-графиках — стандартные отклонения σ .

- Графики зависимости $\ln(D_n)$ от $\ln(n)$ (ln–ln-графика). Поскольку в работе проверяется зависимость вида $D_n \sim \tilde{C}n^{-\alpha}$, где $\alpha = 1$ или $\alpha = \frac{1}{2}$ (в зависимости от симметричности или несимметричности изначального распределения), то предпочтительнее рассматривать эту зависимость в логарифмической форме, в которой получаем линейную зависимость:

$$\ln D_n \approx \ln \tilde{C} - \alpha \ln n. \quad (10)$$

Здесь α — наклон прямой, а $\ln \tilde{C}$ — свободный член, откуда можно получить оценку \tilde{C} — аналога константы Берри–Эссеена для каждого конкретного распределения. На логарифмическом графике моделируемая зависимость при численных расчетах должна приближаться к прямой, параметры которой можно оценить методом наименьших квадратов, построив по полученным точкам линейную регрессию. Регрессия на ln–ln-графиках позволяет определить показатель степени и константу в зависимости D_n от n или \sqrt{n} и оценить константу в этой зависимости.

Для показательного распределения приведённая выше теоретическая оценка скорости сходимости использовалась далее в статье как эталонная скорость сходимости, позволяющая сопоставить результат моделирования с ожидаемым главным членом Эджворта и оценить корректность вычисленных углов наклона на ln–ln-графиках.

В первую серию моделей вошли распределения с нулевой асимметрией: равномерное на отрезке $[-0,5; 0,5]$ и симметричное бета-распределение $\beta(2,2)$. Для этих распределений рассматривались значения n до 32, и большое количество смоделированных значений случайных величин $M=5 \cdot 10^6$ соответственно. Выбор такого M обусловлен тем, что для быстро сходящихся распределений, которыми являются симметричные распределения, статистический шум (порядка $1/\sqrt{M}$) вносит заметный вклад в D_n и коэффициенты регрессии уже при $n > 30$.

Для каждого распределения были построены следующие графики.

- Q–Q-графики, единица измерения — стандартная ошибка отклонения.
- Графики зависимости $\ln(D_n)$ от $\ln(n)$ (ln–ln-графика), безразмерные единицы.

Во второй серии экспериментов рассматривались показательное распределение, и несимметричные бета-распределения (α, β) с $\alpha \neq \beta$. Для этих распределений использовались существенно большие значения n (до $n = 500$) и $M = 1 \cdot 10^6$. Такое сочетание обеспечивает баланс между численной точностью и диапазоном исследуемых n .

Для каждой фиксированной степени асимметрии были построены следующие графики.

- Гистограммы смоделированной плотности распределения стандартизированной суммы Z_n/n случайных величин, безразмерные единицы.
- Q–Q-графики, единица измерения — стандартные отклонения σ .
- Графики зависимости $\ln(D_n)$ от $\ln(n)$ (ln–ln-графика), безразмерные координаты.

Численные расчёты выполнялись на языке Python (версия 3.11) с использованием библиотек NumPy и SciPy. Для генерации независимых реализаций случайных величин применялся генератор псевдослучайных чисел библиотеки NumPy с фиксированными начальными значениями (seed) для каждого набора параметров, что обеспечивает воспроизводимость результатов моделирования. Полный исходный код программы, реализующей описанную методику, опубликован в открытом доступе [9].

Результаты

Результаты моделирования для равномерного распределения и симметричного бета-распределения $\beta(2,2)$ показали быструю сходимость сумм к нормальному распределению. На рис. 1 показаны Q–Q-графики, определённые при $n=14$, для равномерного распределения (слева) и для бета распределения $\beta(2,2)$ (справа). Видно, что соответствующие значения эмпирических и теоретических квантилей располагаются вдоль прямой линии, что говорит о близости суммы к нормальному закону.

На рис. 2 представлены ln–ln-графики для обоих симметричных распределений. Регрессия, построенная по ln–ln-графику для равномерного распределения имеет вид: $\ln(D_n) = -3.481(\pm 0,09) - 0.994(\pm 0,035) \cdot \ln(n)$. Угол наклона $\alpha = -0,994(\pm 0,035)$. Для бета-распределения $\beta(2,2)$ уравнение регрессии: $\ln(D_n) = -3,883(\pm 0,087) - 0,959(\pm 0,034) \cdot \ln(n)$. Угол наклона $\alpha = -0,959(\pm 0,034)$. В обоих случаях расчёты показывают, что скорость сходимости

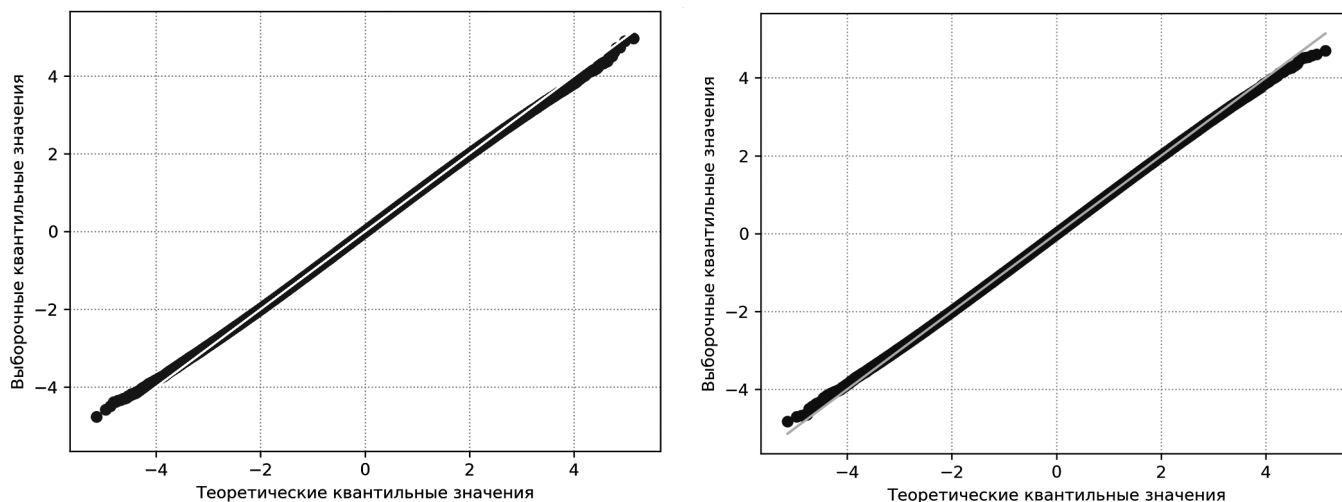


Рис. 1. Q–Q-графики для симметричных распределений: равномерного (слева) и $\beta(2, 2)$ (справа)

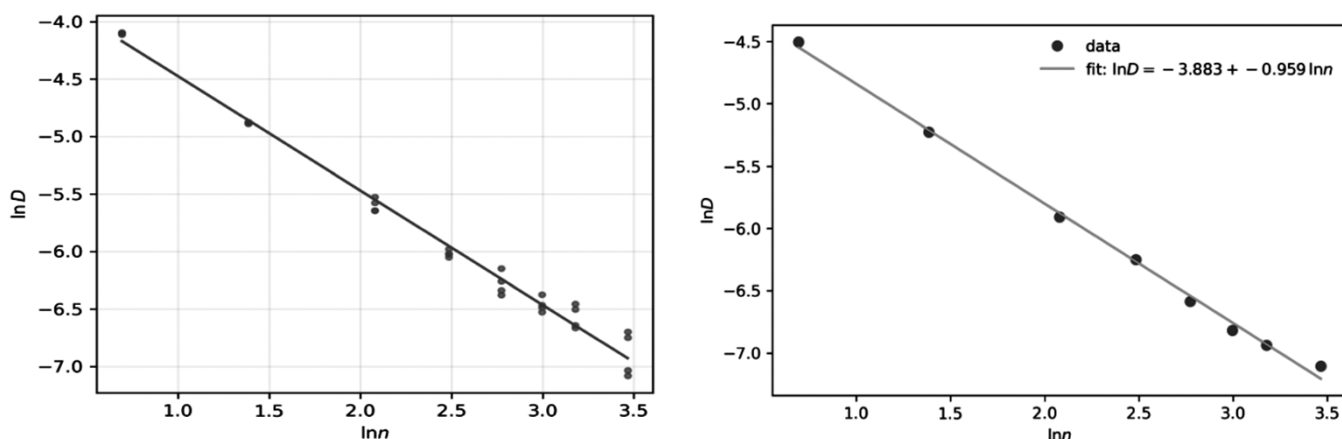


Рис. 2. ln–ln-графики для симметричных распределений: равномерного и $\beta(2, 2)$

димости симметричных распределений значительно выше, чем $O(1/\sqrt{n})$ и имеет порядок $\approx O\left(\frac{1}{n}\right)$.

Чтобы проиллюстрировать рост влияния шума Монте-Карло на результаты эксперимента при увеличении n , на левом графике для каждого значения n показаны 4 значения $\ln(D_n)$, полученные при моделировании с 4 разными seeds (начальное значение генератора случайных чисел). Видно, что при небольших n точки практически сливаются, но при росте n растет разброс экспериментальных точек. Это объясняется тем, что для симметричных распределений с ростом n , D_n быстро стремится к нулю и добавка шума становится существенной.

Аналогичные расчеты были проведены для несимметричных распределений. Показательное распределение рассматривалось с параметром 1. На рис. 3 слева направо показаны гистограмма распределения плотности вероятности суммы случайных величин, Q–Q-график и ln–ln-график. Расчеты проведены при $n=100$.

Регрессия на ln–ln-графике для показательного распределения имеет вид: $\ln(D_n) = -2,026(\pm 0,012) - 0,496(\pm 0,003) \cdot \ln(n)$. Угол наклона $\alpha = -0,496(\pm 0,003)$, что практически полностью совпадает с теоретической скоростью сходимости порядка $O(n^{-0,5})$.

Чтобы сравнить величину численной константы в зависимости $D_n \sim \tilde{C}n^{-\alpha}$ с теоретической оценкой, константой Берри–Эссена, найдем величину $\tilde{C} = e^{-2,022} \approx 0,132$. Затем полученную величину \tilde{C} следует разделить на величину ρ_{abs} , которая для показательного распределения равна: $\rho_{abs} \approx 2,414$. Получаем, что для показательного распределения оценка аналога константы Берри–Эссена равна $\tilde{C} / \rho_{abs} \approx 0,0547$.

Для несимметричного бета-распределения $\beta(2,5)$ коэффициент асимметрии $\gamma_1 \approx 0,596$. На рис. 4 слева направо показаны гистограмма распределения плотности вероятности суммы, Q–Q-график и ln–ln-график. Гистограмма и Q–Q-график приведены для $n=100$. Гистограмма хорошо соответствует плотности нормального распределения, так же, как и Q–Q-график. Регрессия

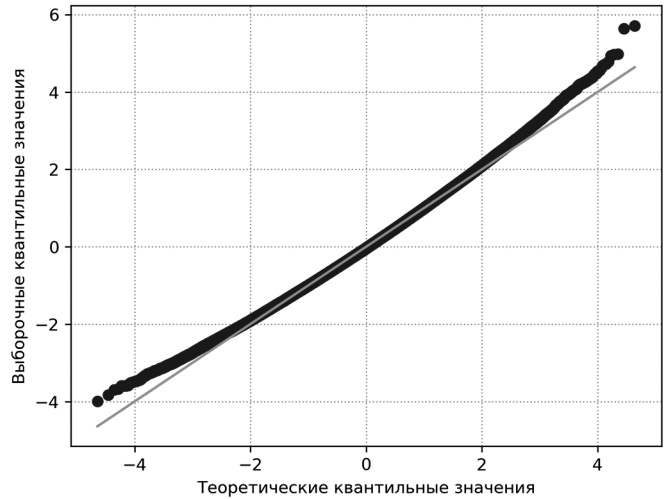
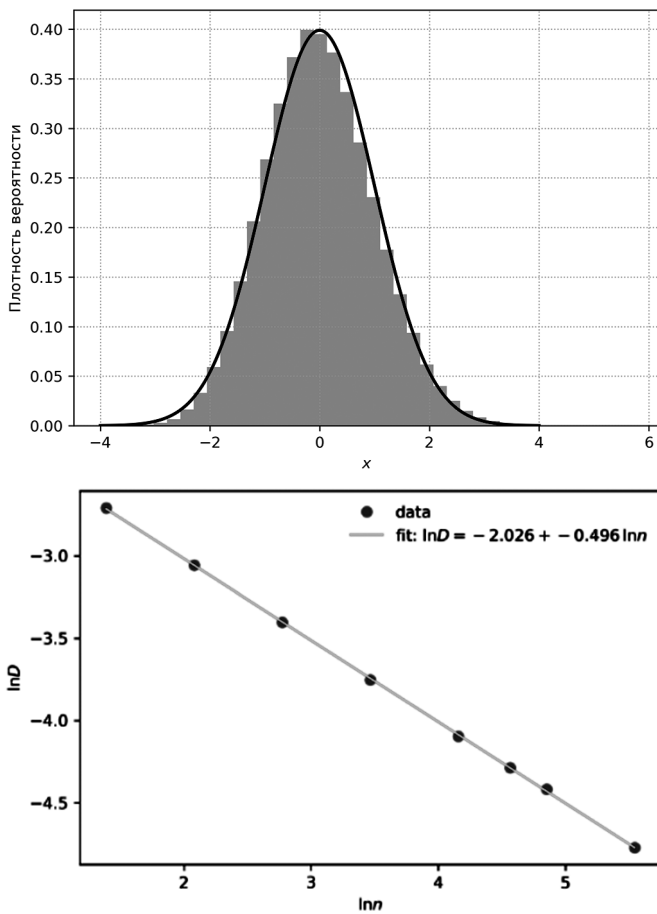


Рис. 3. Результаты для показательного распределения (n = 100): гистограмма плотности вероятности суммы, Q-Q-график и ln-ln-график

на ln-ln-графике для этого распределения имеет вид: $\ln(D_n) = -3,267(\pm 0,123) - 0,473(\pm 0,028) \cdot \ln(n)$. Параметр $\alpha = -0,473(\pm 0,028)$ близок к теоретической скорости сходимости порядка $O(n^{-0,5})$.

Величина численной константы \tilde{C} в зависимости $D_n \sim \tilde{C}n^{-\alpha}$ равна $\tilde{C} = e^{-3,263} \approx 0,0381$. Коэффициент ρ_{abs} для бета-распределения с параметрами (2,5) равен: $\rho_{abs} \approx 1,55$. Отсюда оценка аналога константы Берри-Эссеена $\tilde{C} / \rho_{abs} \approx 0,0246$. Эта величина меньше аналогичной для показательного распределения, что объясняется меньшим коэффициентом асимметрии.

Последний расчет был выполнен для бета-распределения с параметрами (0,5,7). Для этого варианта бета-распределения коэффициент асимметрии $\gamma_1 \approx 2,13$. На рис. 5 слева направо расположены гистограмма распределения плотности вероятности суммы (n=128), Q-Q-график (n=128) и ln-ln-график. Гистограмма в целом хорошо соответствует плотности нормального распределения, однако, заметна небольшая перекошенность, вызванная асимметрией начального распределения.

На Q-Q-графике в центральной части, по-прежнему, наблюдается хорошее совпадение эмпирических и теоретических квантилей,

Отклонение эмпирических квантилей от теоретических существенно лишь на хвостах распределения, вне интервала $\pm 3\sigma$ от математического ожидания.

Регрессия на ln-ln-графике для этого распределения имеет вид:

$\ln(D_n) = -1,91(\pm 0,045) - 0,506(\pm 0,01) \cdot \ln(n)$. Параметр $\alpha = -0,506(\pm 0,01)$, то есть, скорость сходимости по-прежнему близка к теоретической порядка $O(n^{-0,5})$.

Величина численной константы \tilde{C} в зависимости $D_n \sim \tilde{C}n^{-\alpha}$ равна $\tilde{C} = e^{-1,91} \approx 0,148$. Коэффициент ρ_{abs} для бета-распределения с параметрами (0,5,7) равен: $\rho_{abs} \approx 2,446$. Отсюда оценка аналога константы Берри-Эссеена $\tilde{C} / \rho_{abs} \approx 0,061$.

Обсуждение и выводы

Проведённые исследования позволили получить новые численные оценки скорости сходимости законов распределения сумм ряда независимых одинаково распределённых случайных величин с различной степенью асимметрии к нормальному закону. Полученные резуль-

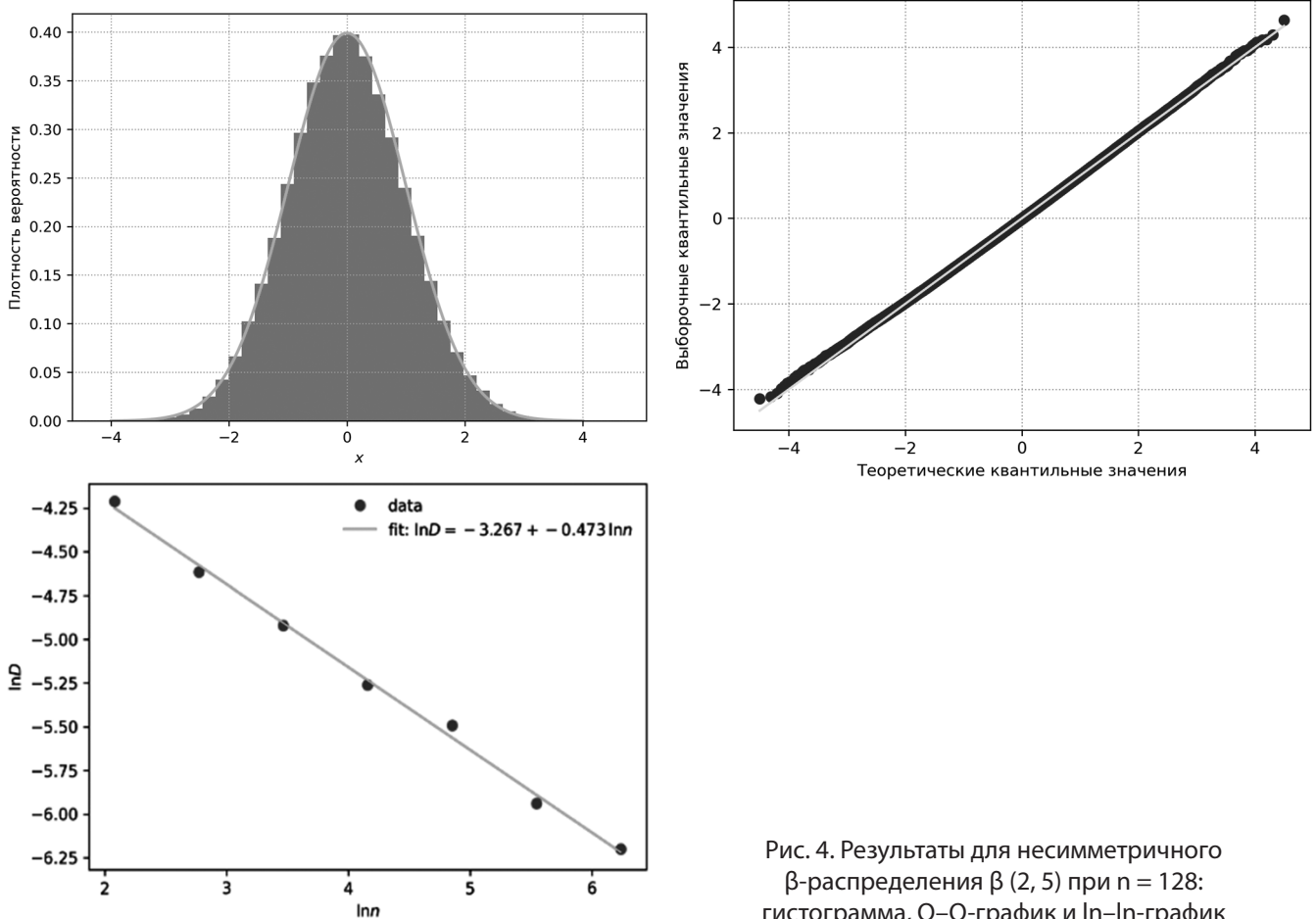


Рис. 4. Результаты для несимметричного β -распределения $\beta(2, 5)$ при $n = 128$: гистограмма, Q-Q-график и \ln - \ln -график

таты согласуются с теоретическими оценками первых членов разложения Эджворта, подробно изложенными в работах [11,12], и подтверждают влияние асимметрии распределения на скорость сходимости [10].

Результаты моделирования для равномерного распределения и симметричного бета-распределение $\beta(2,2)$ показали, что скорость сходимости сумм к нормальному распределению $D_n \approx O(1/n)$. Именно такой характер сходимости для распределений с нулевой асимметрией следует из разложения Эджворта.

Для показательного распределения визуальный анализ гистограмм и Q-Q-графиков показывает хорошее согласование с нормальным распределением. Отклонение эмпирических квантилей от теоретических существенно лишь для хвостовых областей, вне интервала $\pm 3\sigma$ от математического ожидания, который соответствует области почти всех возможных значений суммарной случайной величины.

Регрессионный анализ зависимости D_n от n в логарифмических координатах демонстрирует устойчивый наклон, близкий к $-0,5$, что подтверждает теоретический порядок сходимости $O(n^{-1/2})$, кроме того, величи-

на численной константы в этой зависимости оказалась равной $\tilde{C} \approx 0,132$, что практически совпадает с полученной выше теоретической оценкой, см. (9), из разложения Эджворта. Скорость сходимости сумм показательно распределенных независимых случайных величин значительно превышает общую оценку из неравенства Берри-Эссеена.

Для несимметричных бета-распределений $\beta(2,5)$ и $\beta(0,5,7)$ также получено хорошее соответствие между распределением сумм и нормальной аппроксимацией.

На Q-Q-графиках в центральной части, наблюдается хорошее совпадение эмпирических и теоретических квантилей. Отклонение эмпирических квантилей от теоретических возникает лишь на хвостах распределения и усиливается с ростом асимметрии. Значительное расхождение квантилей наблюдается вне интервала $\pm 3\sigma$ от математического ожидания.

Анализ уравнения регрессии на \ln - \ln -графиках для обоих распределений показал, что скорость сходимости соответствует теоретической скорости порядка $O(n^{-0.5})$. Увеличение асимметрии приводит к росту константы при $n^{-0.5}$, что согласуется с теоретической зави-

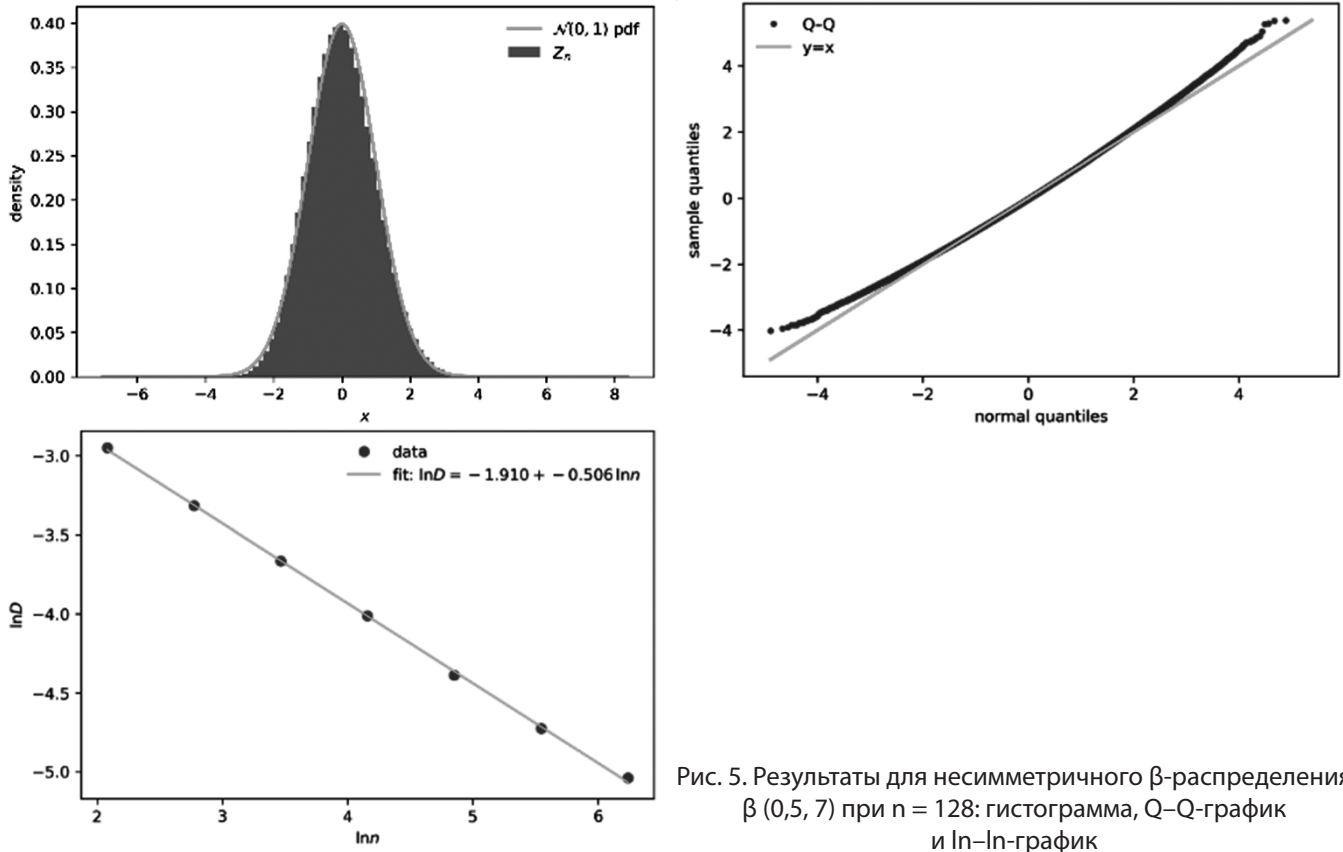


Рис. 5. Результаты для несимметричного β -распределения $\beta(0,5, 7)$ при $n = 128$: гистограмма, Q–Q-график и \ln – \ln -график

симостью отклонения от третьего центрального момента [11,12].

Следует подчеркнуть, что полученные в настоящей работе значения аналога константы Берри–Эссеена (0,024 – 0,061) для конкретных несимметричных непрерывных распределений закономерно оказались существенно ниже универсальной константы 0,4748 [4]. Универсальная константа является верхней границей в неравенстве Берри–Эссеена, то есть оценивает наилучший возможный случай по всем распределениям с конечным третьим моментом. Поэтому индивидуальные коэффициенты, рассчитанные для конкретных распределений, должны быть меньше универсальной верхней оценки.

Полученные по регрессионной модели значения аналога константы Берри–Эссеена для показательного распределения (порядка 0,05) значительно ниже ранее опубликованных аналогичных численных оценок для показательного распределения при умеренных n [13, 14], где соответствующие величины лежали в интервале 0,66 – 0,70. Разница объясняется тем, что в работах [13, 14] численные оценки нормальной аппроксимации для конкретных распределений получались путём прямого вычисления величины $K(n) = D_n \sqrt{n}$ на фиксированной сетке значений n , после чего в качестве оценки «эффективной» константы выбиралось максимальное зна-

чение $K(n)$ на этом диапазоне. Такой подход отражает поведение при умеренных n , но неизбежно завышает оценку коэффициента из-за существенного вклада членов порядка n^{-1} , шумов Монте-Карло и случайных выбросов.

В настоящей работе используется иной метод: коэффициент при $n^{-1/2}$ оценивается по регрессионной зависимости D_n от n в логарифмических координатах. Этот подход задействует всю совокупность значений D_n , сглаживает локальные колебания и позволяет выделить устойчивый наклон, соответствующий теоретической скорости сходимости. Поэтому полученные численные значения аналога константы Берри–Эссеена отражают именно асимптотическое поведение и оказываются существенно меньше оценок, основанных на максимумах $K(n)$ при конечных n .

В настоящей работе представлены детальные численные расчёты аналога константы Берри–Эссеена для нескольких бета-распределений, различающихся степенью асимметрии. Эти результаты расширяют набор распределений, для которых численные характеристики скорости нормальной аппроксимации получены в явном виде.

Рассчитанные значения $(\ln(D_n); \ln(n))$ очень близко сгруппированы относительно прямой регрессии, что

говорит о малой величине статистической ошибки при моделировании случайных величин. Для симметричных распределений отмечено влияние статистического шума, обусловленного очень высокой скоростью сходимости, что отразилось на $\ln\text{-}\ln$ -графиках.

Разработанная методика анализа на основе линейной регрессии в логарифмических координатах показала высокую устойчивость и может быть использована для оценки скорости сходимости и определения констант Берри–Эссеена для широкого класса распределений.

Полученные результаты согласуются с ведущими членами разложения Эджворта и подтверждают корректность оценённого порядка скорости сходимости. Разработка строгих неасимптотических оценок для экс-

поненциального и бета-распределений требует отдельного теоретического анализа на основе методов Эссеена и выходит за рамки настоящей работы. Вместе с тем мы видим, что для распределений с небольшой асимметрией мы можем пользоваться первыми членами разложения Эджворта для оценки скорости сходимости, несмотря на их асимптотический характер.

Полученные результаты могут быть использованы в прикладных задачах статистического моделирования, анализа случайных процессов и оценки точности нормальных аппроксимаций. Исследования есть смысл распространить на уточнение скорости сходимости в ЦПТ для распределений с существенной не симметрией и для оценки возникающих ошибок квантилей, что важно в различных задачах.

ЛИТЕРАТУРА

1. Berry A.C. The accuracy of the Gaussian approximation to the sum of independent variates // *Transactions of the American Mathematical Society*. — 1941. — Vol. 49, No. 1. — P. 122–136. — DOI: 10.2307/1990253.
2. Esseen C.-G. On the Liapunoff limit of error in the theory of probability // *Arkiv för Matematik, Astronomi och Fysik*. — 1942. — Vol. 28A, No. 9. — P. 1–19. — DOI: 10.1007/BF02481954.
3. Шевцова И.Г. Об абсолютных константах в неравенствах типа Берри–Эссеена для одинаково распределённых слагаемых // *Препринт*. — 2011. — arXiv:1111.6554. — URL: <https://arxiv.org/abs/1111.6554>.
4. Edgeworth F.Y. The law of error // *Transactions of the Cambridge Philosophical Society*. — 1905. — Vol. 20. — P. 36–65.
5. Dinov I.D., Christou N., & Sanchez J. (2008). Central Limit Theorem: New SOCR Applet and Demonstration Activity. *Journal of Statistics Education*, 16(2), 1–15.
6. Dambolena I.G. (1984). Teaching the Central Limit Theorem through Computer Simulation. *Mathematics & Computer Education*, 18(2), P.128–132.
7. Золотухин Н., Нагаев С., Чеботарёв А. Константа Берри–Эссеена для распределений Бернулли: численные эксперименты // *Препринт arXiv*. — 2018. — arXiv:1810.09681. — URL: <https://arxiv.org/abs/1810.09681>
8. Chang H.-J., Wu C.-H., Ho J.-F., Chen P.-Y. On sample size in using Central Limit Theorem for Gamma distribution // *Information and Management Sciences*. 2008. Vol. 19, no. 1. P. 153–174.
9. Бессарабская И.Э., Бессарабский А.Ю. Numerical Experiments for CLT. Версия 1: программное обеспечение. — Zenodo, 2025. — DOI 10.5281/zenodo.17783275 — URL: <https://doi.org/10.5281/zenodo.17783275>
10. Hall P. Fast rates of convergence in the central limit theorem // *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*. 1983. Vol. 62, No. 4. P. 491–507. DOI: 10.1007/BF00534201.
11. Bhattacharya R.N., Rao R.R. *Normal Approximation and Asymptotic Expansions*. New York: Wiley, 1976. 300 p.
12. Petrov V.V. *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford: Clarendon Press, 1995. 292 p.
13. Тюрин Ю.Н. Об оценках константы Берри–Эссеена // *Теория вероятностей и её применения*. — 2010. — Т. 55, № 3. — С. 612–628.
14. Королёв В.Ю., Шевцова И.Г. Оценки расстояния между распределением суммы независимых случайных величин и нормальным распределением // *Теория вероятностей и её применения*. — 2010. — Т. 55, № 3. — С. 498–515.

© Бессарабская Ирина Эдуардовна (irina.bessarabskaya@gmail.com)
Журнал «Современная наука: актуальные проблемы теории и практики»