УДК 519.21

О численных исследованиях точности аппроксимаций в центральной предельной теореме

В. В. Сенатов*

* *Кафедра теории вероятностей,*
Московский государственный университет им. М.В.Ломоносова,
Ленинские Горы 1, Москва, Россия, 119992

Аннотация. Рассматриваются вопросы, связанные с точностью аппроксимации в ЦПТ. Приводятся численные и графические иллюстрации.

Ключевые слова: центральная предельная теорема, точность аппроксимации, асимптотические разложения.

Цель данной работы — изучение аппроксимаций, которые строятся с помощью асимптотических разложений в центральной предельной теореме (ЦПТ). Для них будут указаны явные оценки точности, которые они гарантируют, и мы сравним их с реальной точностью для тех исключительных случаев, когда многократные свёртки $F^{*n}(x)$ функции распределения $F(x)$ исходных случайных величин вычисляются в явном виде. Интерес к таким аппроксимациям связан с тем, что в ЦПТ аппроксимация "чистым" нормальным законом оказывается слишком грубой.

Пусть $X_1, X_2, \ldots$ — независимые одинаково распределённые случайные величины с нулевым средним, единичной дисперсией и функцией распределения $F(x)$. Обозначим $F_n(x) = F^{*n}(\sqrt{n}x)$ функцию распределения нормированной суммы

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

и $G(x)$ — функцию распределения нормального закона с плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Мы рассмотрим локальную форму ЦПТ для плотностей гладких распределений. Здесь мы называем гладкими функции распределения $F(x)$, для которых при некотором $\nu > 0$

$$\int_{-\infty}^{+\infty} |f(t)|^\nu \, dt < \infty, \tag{1}$$

E-mail: v.senatov@yandex.ru.

где $f(t)$ — характеристическая функция для $F(x)$. Выполнение этого условия гарантирует существование непрерывных и ограниченных плотностей $p_n(x) = F_n'(x)$ при $n \geq \nu$ и справедливость локальной формы ЦПТ: $p_n(x) \to \varphi(x)$, $n \to \infty$, $-\infty < x < +\infty$. Нам понадобятся моменты α_k и абсолютные моменты β_k функции распределения $F(x)$. Мы будем использовать многочлены Чебышева – Эрмита

$$H_k(x) = (-1)^k \frac{\varphi^{(k)}}{\varphi(x)}$$

и моменты Чебышева – Эрмита

$$\theta_k = \int_{-\infty}^{+\infty} H_k(x)\, dF(x), \quad k = 1, 2, \ldots,$$

функции распределения $F(x)$. Отметим, что $H_4(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, $H_6(x) = x^6 - 15x^4 + 45x^2 - 15$, $H_8(x) = x^8 - 28x^6 + 210x^4 - 420x^2 - 105$ и для распределений с нулевым средним и единичной дисперсией $\theta_3 = \alpha_3$, $\theta_4 = \alpha_4 - 3$.

Теорема 1. *Если для функции распределения $F(x)$ выполнено условие* (1) *и её пятый момент конечен, то*

$$\begin{gathered}\left| p_n(x) - \left(1 + \frac{\theta_3}{3!\sqrt{n}} H_3(x)\right)\varphi(x)\right| \leq \\ \leq \left|\frac{\theta_4}{4!n} H_4(x)\varphi(x) + \frac{1}{2}\frac{n-1}{n}\left(\frac{\theta_3}{3!\sqrt{n}}\right)^2 H_6(x)\varphi(x)\right| + O\left(\frac{1}{n^{3/2}}\right),\end{gathered} \tag{2}$$

где

$$\begin{aligned}&O\left(\frac{1}{n^{3/2}}\right) \leq \\ &\leq \frac{\|\theta_5\|}{5!n^{3/2}} B_{5,n} + \frac{1}{2}\left(\frac{|\theta_3|}{3!\sqrt{n}}\frac{|\theta_4|}{4!n} + \frac{|\theta_3|}{3!\sqrt{n}}\frac{\|\theta_4\|}{4!n}\right) B_{7,n} + \frac{1}{6}\left|\frac{\theta_3}{3!\sqrt{n}}\right|^3 B_{9,n} + \\ &+ \left(\frac{|\theta_4|}{4!2} + \frac{1}{48}\right)\frac{B_{6,n}}{n^2} + \frac{1}{2}\frac{|\theta_4|}{4!n}\frac{\|\theta_4\|}{4!n} B_{8,n} + \frac{1}{6}\left(\frac{\theta_3}{3!\sqrt{n}}\right)^2 \frac{\|\theta_4\|}{4!n} B_{10,n} + \\ &+ \frac{\alpha^{n-\nu}(T)}{\pi}\int_T^{+\infty} |f(t)|^\nu\, dt + \frac{1}{\pi T} e^{-T^2 n/2}.\end{aligned}$$

Здесь и далее $T > 0$ — число, выбор которого находится в нашем распоряжении,

$$B_{k,n} = \frac{1}{2\pi}\int_{-T\sqrt{n}}^{T\sqrt{n}} |t|^k \mu^{n-1}\left(\frac{t}{\sqrt{n}}\right) dt,$$

$\mu(t) = \max\{|f(t)|, e^{-t^2/2}\}$ и существует $T > 0$ такое, что

$$B_{k,n} \to B_k = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |x|^k \varphi(x)\, dx \quad n \to \infty,$$

$\alpha(T) = \sup\{|f(t)| : t \geq T\} < 1$, $\|\theta_4\| = \alpha_4 + 3$, $\|\theta_5\| = \beta_5 + 10|\alpha_3|$.

Величина $\frac{\theta_3}{3!\sqrt{n}} H_3(x)\varphi(x)$ является первым членом асимптотического разложения разности $p_n(x) - \varphi(x)$, она эквивалентна этой разности при $n \to \infty$ для всех x, где $H_3(x)$ не обращается в нуль. Функция, стоящая под знаком модуля в правой части (2) является вторым членом разложения Грама – Шарлье разности $p_n(x) - \varphi(x)$, она эквивалентна функции

$$\left| p_n(x) - \left(1 + \frac{\theta_3}{3!\sqrt{n}} H_3(x)\right) \varphi(x) \right|$$

при $n \to \infty$ для всех x, где она не обращается в нуль. Нас будет интересовать вопрос о том, при каких значениях n левая часть (2) не превосходит 10^{-3} и связанные с этим величины.

Рассмотрим функцию распределения $F(x)$ с плотностью $p(x) = e^{-(x+1)}$, $x \geq -1$, и $p(x) = 0$, $x < -1$. Хорошо известно, что многократные свёртки плотности $q(x) = e^{-x}$, $x \geq 0$, и $q(x) = 0$, $x < 0$, вычисляются в явном виде, для натуральных n они суть

$$q^{*n}(x) = \frac{x^{n-1}}{(n-1)!} e^{-x}, \quad x \geq 0; \quad q^{*n}(x) = 0, \quad x < 0,$$

это — плотности распределений Эрланга. Отсюда легко следует, что

$$p_n(x) = \sqrt{n} \frac{n^n}{n! e^n} \left(1 + \frac{x}{\sqrt{n}}\right)^{n-1} e^{-\sqrt{n} x},\ x \geq -\sqrt{n}; \quad p_n(x) = 0,\ x < -\sqrt{n}.$$

Вычисления по этой формуле не представляют труда, более того, с помощью пакета программ, который доступен на сайте ru.numberempire.com, в котором есть графопостроитель, легко нарисовать график функции $p_n(x)$, равно как и графики других функций, участвующих в формуле (2). При этом следует учесть, что для указанной функции $F(x)$ величины $\alpha_3 = 2$, $\alpha_4 = 9$, поэтому $\theta_3/3! = 1/3$, $\theta_4/4! = 1/4$. При использовании упомянутого графопостроителя следует учесть, что он не понимает, что такое $n!$, поэтому величину $\sqrt{n}\frac{n^n}{n! e^n}$ нужно вычислять отдельно, например, при $n = 100$ её значение приближённо равно 0.39861.

Для указанной функции распределения легко вычислить все величины из (2). Нетрудно проверить, что $|f(t)| = 1/\sqrt{1+t^2}$, поэтому можно взять $\nu = 2$ легко понять, что $\mu(T) = 1/\sqrt{1+T^2}$ для любого $T > 0$,

$$B_{k,n} = \frac{1}{2\pi}\int_{-T\sqrt{n}}^{T\sqrt{n}} \frac{|t|^k}{(1+t^2/n)^{(n-1)/2}}\,dt,$$

максимальное значение $k = 10$ и для $n > 12$ можно переходить к пределу по $T \to \infty$, при этом $B_{k,n} \to B_k$, $n \to \infty$. Последние интегралы можно выразить через бета-функции. Числа B_k при нечётных $k = 2l + 1$ суть $2^l l!/\pi$, при чётных k они суть $(k-1)!!/\sqrt{2\pi}$. Учитывая сказанное, и заменяя $B_{k,n}$ на B_k, нетрудно проверить, что в (2) при больших n

$$O\left(\frac{1}{n^{3/2}}\right) \le \frac{4.03}{n^{3/2}} + \frac{7}{n^2},$$

при этом мы учли, что $\beta_5 < 44.5$.

Для нашей функции распределения $F(x)$ левая часть (2) не превосходит 10^{-3} при $n \ge 114$, для этого n максимум первого слагаемого в правой части (2) равен 0.00088 и величина $O(1/n^{3/2})$ меньше 0.0039. Оценку последней величины можно существенно улучшить, привлекая третий член асимптотического разложения разности $p_n(x) - \varphi(x)$ (см. например, [1]). Отметим, что для нашего распределения при $n = 114$ величина $\max_x |p_n(x) - \varphi(x)|$ равна 0.018, она становится меньше 10^{-3} лишь при $n > 34000$.

Оценка величины $O(1/n^{3/2})$ из правой части (2) является довольно громоздкой. Для левой части (2) справедлива более простая оценка, справедливая, к тому же, для функций распределения $F(x)$ с конечным четвёртым моментом. Эта оценка имеет вид

$$\begin{aligned}&\left|p_n(x) - \left(1 + \frac{\theta_3}{3!\sqrt{n}}H_3(x)\right)\varphi(x)\right| \le \\ &\le \frac{\|\theta_4\|}{4!n}B_{4,n} + \frac{1}{2}\frac{n-1}{n}\left(\frac{\theta_3}{3!\sqrt{n}}\right)^2 B_{6,n} + O\left(\frac{1}{n^{3/2}}\right),\end{aligned} \tag{3}$$

где

$$O\left(\frac{1}{n^{3/2}}\right) \le \frac{\|\theta_5^{(3)}\|}{5!n^{3/2}} + \frac{1}{2}\frac{|\theta_3|}{3!\sqrt{n}}\frac{\|\theta_4\|}{4!n} + \frac{\alpha^{n-\nu}(T)}{\pi}\int_T^{+\infty}|f(t)|^\nu\,dt + \frac{e^{-T^2 n/2}}{\pi T},$$

$\|\theta_5^{(3)}\| = 10|\alpha_3|$.

Сумма двух первых слагаемых из правой чисти (3) оценивает первое слагаемое из правой части (2), однако, эта сумма почти в десять раз больше первого слагаемого из правой части (2). То есть, в оценке (2) мы, усилив (по сравнению с оценкой (3)) условие на моменты $F(x)$, улучшили величину в правой части, убывающую при росте n

как $1/n$ и сделали её эквивалентной оцениваемой величине. При этом ухудшилась оценка величины, убывающей при росте n как $1/n^{3/2}$. Отметим, что в отмеченном улучшении решающую роль сыграло то, что функции под знаком модуля в (2) таковы, что в точках локальных экстремумов одной из них другая имеет противоположные знаки и при сложении они компенсируют друг друга.

Заканчивая обсуждение теоремы 1, отметим, что распределения нормированных сумм независимых случайных величин с указанной функцией распределения очень плохо приближаются «чистым» нормальным законом и это связано с тем, что указанное распределение очень асимметрично.

Рассмотрим теперь случай симметричных распределений.

Теорема 2. *Если для симметричной функции распределения* $F(x)$ *выполнено условие* (1) *и её восьмой момент конечен, то*

$$\begin{aligned} &\left| p_n(x) - \left(1 + \frac{\theta_4}{4!n} H_4(x)\right)\varphi(x) \right| \le \\ &\le \left| \frac{\theta_6}{6!n^2} H_6(x)\varphi(x) + \frac{1}{2}\left(\frac{\theta_4}{4!n}\right)^2 H_8(x)\varphi(x) \right| + O\left(\frac{1}{n^3}\right). \end{aligned} \tag{4}$$

Величина $\frac{\theta_4}{4!n} H_4(x)\varphi(x)$ является первым членом асимптотического разложения разности $p_n(x)-\varphi(x)$, она эквивалентна этой разности при $n \to \infty$ для всех x, где $H_4(x)$ не обращается в нуль. Функция, стоящая под знаком модуля в правой части (4) является вторым членом разложения Эджворта – Крамера разности $p_n(x) - \varphi(x)$, она эквивалентна функции

$$\left| p_n(x) - \left(1 + \frac{\theta_4}{4!n} H_4(x)\right)\varphi(x) \right|$$

при $n \to \infty$ для всех x, где она не обращается в нуль. Нас вновь будет интересовать вопрос о том, при каких значениях n левая часть (4) не превосходит 10^{-3} и связанные с этим величины.

Мы не будем выписывать явную оценку величины $O(1/n^3)$, её можно найти в [2]. Мы ограничимся её оценкой для равномерного распределения на $(-\sqrt{3}, \sqrt{3})$. Для этого распределения (при больших n) $O(1/n^3) < 2.94/n^3$. Для этого распределения известны явные формулы для $p_n(x)$, которые можно найти, например, в [2]. Вычисления по этим формулам показывают, что при $n = 5$ максимальное значение левой части (4) есть $0.9 \cdot 10^{-3}$ (при $n = 4$ оно равно $1.56 \cdot 10^{-3}$). При $n = 5$ максимальное значение первого слагаемого в правой части (4) есть $0.683 \cdot 10^{-3}$, оно существенно меньше (в 34 с лишним раза) величины $O(1/n^3)$. Отметим, что максимальное значение первого слагаемого в правой части (4) сравнимо с величиной $O(1/n^3)$ при n порядка 170.

Отметим также, что при $n = 5$ максимальное значение $|p_n(x) - \varphi(x)|$ чуть больше 0.0123 и оно становится меньше 10^{-3} при $n > 60$.

Оценка величины $O(1/n^3)$ из правой части (4) является довольно громоздкой, поэтому мы её и не выписали. Для левой части (4) справедлива более простая оценка, справедливая, к тому же, для функций распределения $F(x)$ с конечным шестым моментом. Эта оценка имеет вид

$$\left|p_n(x) - \left(1 + \frac{\theta_4}{4!n}H_4(x)\right)\varphi(x)\right| \leq \frac{\|\theta_6\|}{6!n^2}B_{6,n} + \frac{1}{2}\frac{|\theta_4|}{4!n}\frac{\|\theta_4\|}{4!n}B_{8,n} + K, \quad (5)$$

где

$$\frac{\|\theta_6\|}{6!} = \frac{\alpha_6}{6!} + \frac{\alpha_4}{4!2} + \frac{1}{12}$$

и K является суммой трёх слагаемых, убывающих при росте n экспоненциально быстро, сейчас они для нас не интересны. Нетрудно понять, что сумма двух первых слагаемых из правой части (5) оценивает первое слагаемое из правой части (4), но эта оценка примерно в 56 раз грубее оцениваемой величины. То есть, в оценке (4) мы, усилив (по сравнению с оценкой (5)) условие на моменты $F(x)$, улучшили величину в правой части, убывающую при росте n как $1/n^2$ и сделали её эквивалентной оцениваемой величине. При этом в оценке (4) появилась величина, убывающая при росте n как $1/n^3$.

Отметим, что и здесь в отмеченном улучшении решающую роль сыграло то, что функции под знаком модуля в (4) таковы, что в точках локальных экстремумов одной из них другая имеет противоположные знаки и при сложении они компенсируют друг друга.

Литература

1. *Senatov V. V.* On quasi-nonuniform estimates for asymptotic expansions in the central limit theorem // J. Math. Sci. — 2016. — Vol. 218, no. 3. — P. 335–353.
2. *Сенатов В. В.* О реальной точности аппроксимаций в центральной предельной теореме. II // Математические труды. — 2016. — Т. 19, вып. 2. — С. 170–199.

UDC 519.21

On numerical studies of the accuracy of approximations in the central limit theorem

V. V. Senatov*

* *Department of Probability Theory,*
Lomonosov Moscow State University,
Leninskie Gory 1, Moscow, 119991, Russia

We consider the questions connected with accuracy of approximations in the CLT. Numerical and graphic illustrations are given.

Keywords: central limit theorem, accuracy of approximation, asymptotic expansions.