УДК 519.214.4, 519.671, 519.651.1

О вычислении абсолютной константы в неравенстве Берри – Ессеена для двухточечных распределений

А. Я. Золотухин*, **С. В. Нагаев**†, **В. И. Чеботарев**‡

* *Кафедра вычислительной механики и математики, Тульский государственный университет, пр. Ленина, 92а, Тула, Россия, 300012*

† *Институт математики им. С. Л. Соболева, пр. Коптюга, 4, Новосибирск, Россия, 630090*

‡ *Вычислительный центр ДВО РАН, ул. Ким Ю Чена, 65, Хабаровск, Россия, 680000*

Аннотация. В работе показано, что абсолютная константа в неравенстве Берри – Эссеена для бернуллиевских случайных величин, обозначим ее C_{02}, строго меньше, чем постоянная Эссеена C_E, если $n \leq 500000$, где n – число слагаемых. В силу аналитического результата Нагаева и Чеботарева 2011 г. это означает, что $C_{02} < 0.4099539$ (при всех $n \geq 1$).

Ключевые слова: оптимальное значение абсолютной константы в неравенстве Берри – Эссеена, биномиальное распределение, численные методы.

1. Введение

Рассмотрим неравенство Берри – Эссеена в случае независимых одинаково распределенных слагаемых. Абсолютную постоянную в этом неравенстве обозначим C_0. В 1956 г. К.-Г. Эссееном [1] доказал, что $C_0 \geq C_E \equiv \frac{3+\sqrt{10}}{6\sqrt{2\pi}} = 0.40973218\ldots$ Верхние оценки, найденные к тому времени, и этот результат послужили аргументом для предположения $C_0 = C_E$, которое высказал В.М. Золотарев в 1966 г. [2].

С тех пор был получен ряд верхних оценок C_0, исторический обзор которых можно найти, например, в [3] и [4]. Наилучшая верхняя оценка, известная к настоящему времени, принадлежит Шевцовой: $C_0 \leq 0.469$ [4]. Заметим, что в получении верхних оценок, начиная с оценок в [2, 5], существенную роль играло использование вычислительной техники.

Тот факт, что нижняя оценка Эссеена достигается на некотором специальном двухточечном распределении, послужил толчком к рассмотрению задачи о константе C_0 в частном случае двухточечных распределений. В этом случае мы вместо C_0 будем писать C_{02}. В 2007 г. в работе [6] опубликовано доказательство неравенства $C_{02} \leq \frac{1}{\sqrt{2\pi}}$ симметричном случае. В 2016 г. Й. Шульц доказал, что если условие симметричности нарушается, то имеет место равенство $C_{02} = C_E$ [7].

Отметим, что еще в 2011 г. в работе [8, теорема 1.1] была получена новая оценка погрешности нормального приближения для биномиального распределения. Главная часть найденной там мажоранты погрешности содержит в качестве множителя функцию $\mathcal{E}(p) = \frac{2-p}{3\sqrt{2\pi}\,[p^2+(1-p)^2]}$, где p – параметр исходной случайной величины, $0 \le p \le 0.5$. При этом остаточная часть мажоранты положительна, убывает по n и при $n \to \infty$ стремится к нулю быстрее главной части. Так как максимальное значение функции $\mathcal{E}(p)$ равно постоянной C_E, то этот результат из [8] позволяет находить верхнюю границу для C_{02}, сколь угодно близкую к C_E, при условии, что n достаточно велико.

Вместо C_{02} мы будем писать $C_{02}(\underline{N})$, если предполагается, что $n \ge N$, и $C_{02}(\overline{N})$, если предполагается, что $n \le N$.

Таким образом, теорема 1.1 из [8] позволяет по данному $\varepsilon > 0$ находить такое N_ε, что для всех $n \ge N_\varepsilon$ выполняется неравенство $C_{02} \le C_E + \varepsilon$.

В [8], в частности, было доказано, что если $\varepsilon = 0.4215 - C_E = 0.0117678\ldots$, то $N_\varepsilon = 200$, т. е. $C_{02}(\underline{200}) \le 0.4215$. В дополнение к этой теоретической оценке были проведены вычисления (см. [9]), которые показали, что $C_{02}(\overline{200}) < C_E$. Следовательно, при всех $n \ge 1$ верна оценка $C_{02} \le 0.4215$. Заметим, что вычисления проводились на обычном персональном компьютере.

Однако при достаточно малом $\varepsilon > 0$ номер N_ε может быть настолько большим, что проверка неравенства $C_{02}(\overline{N_\varepsilon}) < C_E$ потребует более мощной вычислительной техники.

Пусть $\varepsilon = 0.4099539 - C_E = 0.0002217\ldots$. Вследствие [8, теорема 1.1] имеем $N_\varepsilon = 500000$ и $C_{02}(\underline{N_\varepsilon}) \le 0.4099539$. Чтобы доказать справедливость оценки $C_{02} \le 0.4099539$ при всех $n \ge 1$, остается проверить это неравенство для $1 \le n \le 500000$. На самом деле, как показали вычисления и дополнительный анализ, справедливо более сильное неравенство: $C_{02}(\overline{500000}) < C_E$. Вычисления проводились на суперкомпьютере Blue Gene/P ВЦ ВМК МГУ им. М.В. Ломоносова.

2. Кратко о вычислениях и аргументации

Пусть X, X_1, $X_2, \ldots, X_n$ последовательность независимых случайных величин с одним и тем же распределением: $\mathbf{P}(X=1)=p$, $\mathbf{P}(X=0)=q=1-p$. Далее мы будем использовать следующие обозначения:

$$F_{n,p}(x)=\mathbf{P}\Big(\sum_{i=1}^{n} X_i < x\Big),\ \ \Phi(x)=\int\limits_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt,\ \ G_{n,p}(x)=\Phi\Big(\frac{x-np}{\sqrt{npq}}\Big),$$

$$\Delta_n(p)=\sup_{x\in\mathbb{R}}\big|F_{n,p}(x)-G_{n,p}(x)\big|,\ \ \rho(p)=\frac{p^2+q^2}{\sqrt{pq}},\ \ K_n(p)=\frac{\Delta_n(p)\sqrt{n}}{\rho(p)}.$$

Величина $\sup\limits_{n\geq 1}\sup\limits_{p\in(0,0.5]} K_n(p)$ суть абсолютная постоянная в неравенстве Берри – Эссеена, если его рассматривать не для всех распределений с конечным абсолютным третьим моментом, а только для класса двухточечных распределений. В настоящей работе решается задача вычисления $K(n) = \sup_{p\in(0,0.5]} K_n(p)$ для всех таких n, что $1 \leq n \leq 5\cdot 10^5$.

Заметим, что при фиксированных n и p супремум $\sup\limits_{x\in\mathbb{R}} |F_{n,p}(x) - G_{n,p}(x)|$ достигается в некоторой точке разрыва функции $F_{n,p}(x)$. Таким образом, можно искать сначала

$$\Delta_n(p) = \max_{0\leq i\leq n} \left\{ |F_{n,p}(i) - G_{n,p}(i)|, |F_{n,p}(i+0) - G_{n,p}(i)| \right\}, \qquad (1)$$

где i – целые, а затем $K(n)$.

Отметим, что вместо отрезка $[0, 0.5]$ (для параметра p) мы можем иметь дело с более узким отрезком, причем отделенным от нуля, а именно, с отрезком $I = [0.1689, 0.5]$. Основанием для этого служит следующее утверждение.

Лемма 1. *Если $0 < p \leq 0.1689$, то $K_n(p) < 0.4096$ при любом $n \geq 1$.*

Лемма 1 выводится с помощью модифицированного неравенства Берри – Эссеена с числовыми константами, доказанного в [10].

Справедливо следующее утверждение.

Теорема 1. *Пусть S – равномерная сетка на I с шагом $h = 10^{-12}$. Если $p \in I$, а p' – узел сетки S, ближайший к p, то при всех $1 \leq n \leq$ 500000*

$$|K_n(p) - K_n(p')| \leq 2\cdot 10^{-8}.$$

Доказательство теоремы 1 весьма громоздко и здесь не приводится из-за ограничений на объем статьи.

Результат вычислений. *При всех $1 \leq n \leq 500000$*

$$\max_{p_j\in S} K_n(p_j) < 0.40973214. \qquad (2)$$

Алгоритм счета представляет собой тройной цикл: цикл по параметру i (см. (1)) вложен в цикл по параметру p, который в свою очередь вложен в цикл по параметру n.

С ростом n время вычислений быстро росло. К примеру, при $2000 \leq n \leq 2100$ вычисления заняли более 3 часов на компьютере с процессором Core2Due E6400. Для $2101 \leq n \leq 500000$ вычисления проводились на суперкомпьютере Blue Gene/P ВЦ ВМК МГУ им. М.В. Ломоносова.

Из [11, Следствие 7] вытекает, что в цикле по i нужно перебирать не все значения i от 0 до n, а лишь те, которые удовлетворяют неравенству $np - (\nu+1)\sqrt{npq} \leq i \leq np + \nu\sqrt{npq}$, где $\nu = \sqrt{3+\sqrt{6}}$ при условии, что $n > 200$. Это привело к значительному сокращению времени вычислений. Например, для промежутка $10000 \leq n \leq 11024$ машинное время составило около 3 минут (без учёта ожидания очереди). Заметим, что вычисления на суперкомпьютере для $490000 \leq n \leq 500000$ продолжались 7 часов. Программа написана на языке программирования C+MPI и зарегистрирована [12].

Из теоремы 1, леммы 1 и (2) следует, что при всех $1 \leq n \leq 500000$ и $p \in (0, 0.5]$ справедлива оценка $K_n(p) < 0.40973216$. Нетрудно обосновать, что это неравенство верно и при $p \in (0.5, 1)$.

3. Заключение

Установлено, что для всех $n \leq 500000$ справедлива оценка $C_{02} < C_E$. В силу аналитического результата Нагаева и Чеботарева 2011 г. [8] тогда справедлива оценка $C_{02} < 0.4099539$ (при всех $n \geq 1$). Заметим, что $0 < 0.4099539 - C_E < 0.0002218$.

Благодарности

Мы благодарим В.Ю. Королева, зав. кафедрой математической статистики ВМК МГУ, И.Г. Шевцову, доцента той же кафедры, за содействие в получении возможности использовать суперкомпьютер, Гуляева А.В., зам. декана факультета ВМК МГУ, и Коробкова С.В., главного администратора ЦОД, за предоставленную возможность проводить вычисления без очереди.

Литература

1. *Esseen C.-G.* A moment inequality with an application to the central limit theorem // Scand. Aktuarietidskr. J. — 1956. — Vol. 3–4. — P. 1–170.
2. *Золотарев В.М.* Абсолютная оценка остаточного члена в центральной предельной теореме // Теория вероятн. и ее примен. — 1966. — Т. 11, вып. 1. — С. 108–119.
3. *Королев В. Ю., Шевцова И. М.*, О верхней оценке абсолютной постоянной в неравенстве Берри–Эссеена // Теория вероятн. и ее примен. — 2009. — Т. 53, вып. 4. — С. 671–695.
4. *Шевцова И. М.*, Оптимизация структуры моментных оценок точности нормальной аппроксимации для распределений сумм независимых случайных величин. – Дисс. на соискание уч. степ. д. физ.-мат. наук. М.: МГУ, 2013.
5. *Zolotarev V. M.*, A sharpening of the inequality of Berry – Esseen // Z. Wahrscheinlichkeitstheor. verw. Geb. — 1967. — Vol. 8. — P. 332–342.

6. *Hipp C., Mattner L.* On the normal approximation to symmetric binomial distribution // Теория вероятн. и ее примен. — 2007. — Т. 52, вып. 3. — С. 610–617.
7. *Schulz J.* The optimal Berry – Esseen constant in the binomial case. – Dissertation. Universität Trier, 2016.
8. *Нагаев С. В., Чеботарев В. И.* Об оценке близости биномиального распределения к нормальному // Теория вероятн. и ее примен. — 2011. —Т. 56, вып. 2. — С. 248–278.
9. *Нагаев С. В., Михайлов К. В., Чеботарев В. И.* Об оценке близости биномиального распределения к нормальному для ограниченного числа наблюдений. — Препринт 2010/160. Хабаровск: Вычислительный центр ДВО РАН, 2010.
10. *Королев В. Ю., Шевцова И. М.*, Уточнение неравенства Берри – Эссеена с приложениями к пуассоновским и смешанным пуассоновским суммам // Обозрение прикладной и промышленной математики. — 2010. — Т. 17, вып. 1. — С. 25–56.
11. *S. V. Nagaev, V. I. Chebotarev, A. Ya. Zolotukhin* On a non-uniform bound of the remainder term in Central Limit Theorem for Bernoulli random variables // Journal of Mathematical Sciences. — 2016. — Vol. 214, no. 1. — P. 83–100.
12. *Золотухин А. Я.* Программа вычисления оценки главного члена в центральной предельной теореме распределения Бернулли для ограниченного числа наблюдений. – Свидет. о госуд. рег. программы для ЭВМ № 2015617151 от 01.06.2015.

UDC 519.214.4, 519.671, 519.651.1

On calculating the absolute constant in the Berry – Esseen inequality for two-point distributions

A. Ya. Zolotukhin*, **S. V. Nagaev**†, **V. I. Chebotarev**‡

* *Department of Computational Mechanics and Mathematics,*
Tula State University,
Lenin avenue, 92a, Tula, 300012, Russia
† *Sobolev Institute of mathematics,*
pr. Koptyuga, 4, Novosibirsk, 630090, Russia
‡ *Computing Center FEB RAS,*
Kim Yu Chen st., 65, Khabarovsk, 680000, Russia

It is shown in the paper that the absolute constant C_{02} in the Berry – Esseen inequality for Bernoulli random variables is less than the Esseen constant C_E, if $n \leq 500000$, where n is the number of summands. Then the bound $C_{02} < 0.4099539$ (for all $n \geq 1$) follows from the analitic result obtained by Nagaev and Chebotarev in 2011.

Keywords: optimal value of the absolute constant in the Berry – Esseen inequality, binomial distribution, computational methods.