ДОКЛАДЫ РОССИЙСКОЙ АКАДЕМИИ НАУК. МАТЕМАТИКА, ИНФОРМАТИКА, ПРОЦЕССЫ УПРАВЛЕНИЯ, 2022, том 503, с. 45–47

МАТЕМАТИКА

УДК 519.212.2

О ПОВЕДЕНИИ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ ВБЛИЗИ МЕДИАНЫ

© 2022 г. Н. А. Волков[1], Д. И. Дмитриев[2], М. Е. Жуковский[1,*]

Представлено академиком РАН В.В. Козловым 27.12.2021 г.
Поступило 06.01.2022 г.
После доработки 06.01.2022 г.
Принято к публикации 10.02.2022 г.

В работе исследуется поведение функции распределения биномиальной случайной величины с параметрами n и $b/(n+c)$ в точке $b-1$ при натуральных $b \le n$ и $c \in [0, 1]$. Полученные результаты имеют непосредственное следствие в широко известной задаче о малых отклонениях сумм независимых случайных величин от их математического ожидания. Кроме того, мы ответили на вопрос о монотонности функции Рамануджана для биномиального распределения, который сформулировали в своей работе Джогдео и Самуэльс в 1968 г.

Ключевые слова: биномиальное распределение, медиана, функция Рамануджана, малые отклонения сумм независимых случайных величин

DOI: 10.31857/S2686954322020199

1. ПУАССОНОВСКОЕ РАСПРЕДЕЛЕНИЕ И ЗАДАЧА РАМАНУДЖАНА

Пусть $b \in \mathbb{N}$ – натуральное число. Разложим e^b в ряд Тэйлора: $e^b = \sum_{j=0}^{\infty} \frac{b^j}{j!}$. Зададимся вопросом: каково наименьшее натуральное μ такое, что $\sum_{j=0}^{\mu} \frac{b^j}{j!} \ge \frac{1}{2} e^b$? Ответ на этот вопрос хорошо известен.

Пусть ξ – неотрицательная целочисленная случайная величина. Медианой ξ называется наименьшее целое неотрицательное число $\mu := \mu(\xi)$ такое, что $P(\xi \le m) \ge \frac{1}{2}$. Для пуассоновской случайной величины η_b с натуральным параметром b известно [1], что $\mu(\eta_b) = b$, что и является ответом на поставленный выше вопрос. Но насколько близка вероятность $\mathrm{P}(\eta_b \le b)$ к $\frac{1}{2}$? Рамануджан выдвинул гипотезу [2], что

$$y_b := \frac{\frac{1}{2} - \mathrm{P}(\eta_b < b)}{\mathrm{P}(\eta_b = b)} \in \left(\frac{1}{3}, \frac{1}{2}\right) \text{ и убывает.}$$

Эта гипотеза была доказана независимо Сего в [3] и Ватсоном в [4]. С тех пор поведение функции y_b было хорошо изучено. В 1913 г. в своем письме Харди Рамануджан выдвинул еще одну гипотезу:

$$y_b = \frac{1}{3} + \frac{4}{135(b + \alpha_b)},$$

где $8/45 \ge \alpha_b \ge 2/21$. Эта гипотеза была подтверждена лишь в 1995 г. Флажолетом и соавт. [5]. В 2003 г. Алм [6] доказал, что α_b убывает, а в 2004 г. Алзер [7] усилил гипотезу Рамануджана:

$$y_b = \frac{1}{3} + \frac{4}{135b} - \frac{8}{2835(b^2 + \beta_b)},$$

где

$$-\frac{1}{3} < \beta_b \le -1 + \frac{4}{\sqrt{21(368 - 135e)}},$$

причем указанные границы являются точными.

2. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И ЗАДАЧА САМУЭЛЬСА

Пусть $\xi_{b,n}$ – биномиальная случайная величина с параметрами n и b/n, где $b \le n$ – натуральные числа. Хорошо известно, что $\xi_{b,n}$ сходится по рас-

[1] *Московский физико-технический институт (национальный исследовательский университет), Долгопрудный, Московская обл., Россия*
[2] *ETH Zürich, ETH AI Center, Zurich, Switzerland*
**E-mail: zhukmax@gmail.com*

пределению к случайной величине η_b при $n \to \infty$. Поэтому естественно ожидать, что описанные в предыдущем разделе свойства пуассоновского распределения справедливы и для биномиального распределения при достаточно больших n. Но можно ли ответить на аналогичные вопросы при всех n?

Так как $|\mu(\xi_{b,n}) - b| \le \ln 2$ [8], медиана биномиальной случайной величины равна $\mu(\xi_{b,n}) = b$. В 1968 г. Джогдео и Самуэльс [9] рассмотрели величину, аналогичную той, которую ввел Рамануджан для пуассоновского распределения:

$$z_{b,n} := \frac{1/2 - P(\xi_{b,n} < b)}{P(\xi_{b,n} = b)}.$$

Теорема 1 (K. Jogdeo, S.M. Samuels, 1968 [9]). *$z_{b,n}$ убывает при $n \ge 2b$, $z_{b,n} \to y_n$ при $n \to \infty$. Более того, для всех $n > 2b$ справедливо $\frac{1}{3} < z_{b,n} < \frac{1}{2}$; при $b < n < 2b$ справедливо $\frac{1}{2} < z_{b,n} < \frac{2}{3}$; $z_{b,\,2b} = \frac{1}{2} = z_{b,b}$.*

Они также заметили, что $z_{b+1,n} < z_{b,n}$ для всех достаточно больших n, но не смогли уточнить этот результат.

Рассмотрим теперь биномиальную случайную величину $\xi_{b,n,c}$ с параметрами n и $\frac{b}{n+c}$, где $b < n$ – натуральные числа и $c \in [0, 1]$. Обозначим $p_{b,n,c} := P(\xi_{b,n,c} < b)$. Исследование монотонности $p_{b,n,c}$ по b мотивировано широко известной задачей о неравенстве малых отклонений, поставленной Самуэльсом [10], которая может быть сформулирована следующим образом: найти минимум $P(\xi_1 + \xi_2 + \ldots + \xi_n < n + c)$ по всем множествам независимых неотрицательных случайных величин $\{\xi_1, \ldots, \xi_n\}$ с одинаковым математическим ожиданием, равным 1. Эта задача до сих пор не решена. Тем не менее известно, что оптимальными случайными величинами являются величины, принимающие два значения с вероятностью 1 (как говорится, с двумя атомами). Если мы далее ограничимся одинаково распределенными случайными величинами с двумя атомами, то сведем исходную задачу к анализу монотонности $p_{b,n,c}$ по b.

Из упомянутого результата Сего и Ватсона следует, что $P(\eta_b < b)$ возрастает. Так как $\lim\limits_{n\to\infty} p_{b,n,c} = P(\eta_b < b)$, то при достаточно больших n справедливо $p_{b+1,n,c} > p_{b,n,c}$. С другой стороны, например, при $n = b + 1$ и $c = 0$ выполнено $0 = p_{b+1,n,c} < p_{b,n,c}$. Таким образом, с ростом n монотонность $p_{b,n,c}$ (как функции от b) меняется.

3. НОВЫЕ РЕЗУЛЬТАТЫ

Нам удалось решить задачу, поставленную Джогдео и Самуэльсом о монотонности $z_{b,n}$ по b.

Теорема 2. *Пусть $\varepsilon > 0$. Существует такое n_0, что при всех $n \ge n_0$ справедливо следующее.*

1. *Если $n - (1+\varepsilon)\sqrt{\frac{77}{360}n} > b > (1+\varepsilon)\sqrt{\frac{77}{360}n}$, то $z_{b+1,n} > z_{b,n}$;*

2. *Если либо $b > n - (1-\varepsilon)\sqrt{\frac{77}{360}n}$, либо $b < (1-\varepsilon)\sqrt{\frac{77}{360}n}$, то $z_{b+1,n} < z_{b,n}$.*

Кроме того, мы исследовали функцию $p_{b,n,c}$ на монотонность по b.

Теорема 3. *Справедливы следующие утверждения*:

1. *Если $n \ge 3b + 2$, то $p_{b+1,n,0} > p_{b,n,0}$. Если $n \le 3b + 1$,* то $p_{b+1,n,0} < p_{b,n,0}$;

2. *При всех $1 \le b < n$ справедливо $p_{b+1,n,1} > p_{b,n,1}$;*

3. *Если $n \ge 3b + 2$ и $c \in (0,1)$, то $p_{b+1,n,c} > p_{b,n,c}$.*

К сожалению, исчерпывающий результат удалось получить лишь для $c = 0$ и $c = 1$. Тем не менее мы нашли асимптотическое значение порога, после которого меняется монотонность.

Теорема 4. *Для всех положительных δ, ε, достаточно больших n и целых $b \in (\varepsilon n, n)$ справедливо следующее*:

1. *Если $b < \frac{n(1-\delta)}{3(1-c)}$, то $p_{b+1,n} > p_{b,n}$;*

2. *Если $b > \frac{n(1+\delta)}{3(1-c)}$, то $p_{b+1,n} < p_{b,n}$.*

Заметим, что из теоремы 3 вытекает следствие в задаче о малых отклонениях при $c = 1$. Пусть $\alpha \in (0, 1)$, $\beta > 1$. Пусть b – целое число такое, что $b < \frac{n+1-n\alpha}{\beta-\alpha} \le b + 1$. Тогда $P(\xi_1 + \ldots + \xi_n < n + 1) \ge p_{b,n,1}$, где $\xi_1, \ldots, \xi_n$ – независимые одинаково распределенные случайные величины со средним 1, и равенство выполняется тогда и только тогда, когда $\alpha = 0$ и $\frac{n+1}{\beta} = b + 1$. Аналогичные утверждения можно было бы сформулировать для любого $c \in (0, 1)$, если бы удалось асимптотический результат в теореме 4 сделать точным.

4. B-ФУНКЦИЯ

Напомним, что $B(x, y) = \int\limits_0^1 t^{x-1}(1-t)^{y-1}$. Доказательства приведенных в предыдущем разделе

утверждений опираются на следующие полученные нами удобные выражения для $p_{b+1,n,c} - p_{b,n,c}$ и $z_{b,n}$.

Утверждение 1. *Для всех натуральных* $b \le n$ *и* $c \in [0, 1]$ *справедливы равенства*

$$p_{b,n,c} = \frac{\int\limits_0^{1-\frac{b}{n+c}} (1-z)^{b-1} z^{n-b} dz}{b\mathrm{B}(b, n-b+1)},$$

$$z_{b,n} = \frac{\frac{1}{2} b \left(\left[\int\limits_{1-b/n}^{1} - \int\limits_0^{1-b/n} \right] (1-z)^{b-1} z^{n-b} dz \right)}{\left(\frac{b}{n}\right)^b \left(1 - \frac{b}{n}\right)^{n-b}}.$$

Анализ этих выражений сводится к исследованию поведения функции $g(z) = (1-z)^{b-1} z^{n-b}$ на $\left[1 - \frac{b+1}{n+c}, 1 - \frac{b}{n+c}\right]$. Ее поведение удается исследовать, используя формулу Тейлора (достаточно разложить до пятого члена) с остаточным членом в форме Лагранжа и следующее удобное представление производных от функции g:

$$\frac{\partial^\ell g}{\partial z^\ell} = (1-z)^{b-1-\ell} z^{n-b-\ell} \times$$
$$\times \sum_{i=0}^{\ell} \binom{\ell}{i} (-1)^{\ell-i} z^{\ell-i} \frac{(n-1-i)!}{(n-1-\ell)!} \frac{(n-b)!}{(n-b-i)!},$$
$$\ell \in \{1, \ldots, \min\{b-1, n-b\}\}.$$

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена за счет гранта Российского научного фонда, проект № 21-71-10092.

СПИСОК ЛИТЕРАТУРЫ

1. *Choi K.P.* On the medians of Gamma distributions and an equation of Ramanujan // Proc. Amer. Math. Soc. 1994. V. 121. № 1. P. 245–251.
2. *Ramanujan S.* Question 294 // Indian Math. Soc. 1911. V. 3. P. 151–152.
3. *Szegö G.* Über einige von S. Ramanujan gestellte Aufgaben // J. Lond. Math. Soc. 1928. V. 3. P. 225–232.
4. *Watson G.N.* Theorems stated by Ramanujan (V): approximations connected with e^x // Proc. Lond. Math. Soc. 1929. V. 29. P. 293–308.
5. *Flajolet P., Grabner P.J., Kirschenhofer P., Prodinger H.* On Ramanujan's Q-function // J. Comput. Appl. Math. 1995. V. 58. P. 103–116.
6. *Alm S.E.* Monotonicity of the difference between median and mean of gamma distributions and of a related Ramanujan sequence // Bernoulli. 2003. V. 9. P. 351–371.
7. *Alzer H.* On Ramanujan's inequalities for exp(k) // J. Lond. Math. Soc. 2004. V. 69. P. 639–656.
8. *Hamza K.* The smallest uniform upper bound on the distance between the mean and the median of the binomial and Poisson distributions // Statist. Probab. Lett. 1995. V. 23. P. 21–25.
9. *Jogdeo K., Samuels S.M.* Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation // Ann. Math. Statist. 1968. V. 39. № 4. P. 1191–1195.
10. *Samuels S.M.* On a Chebyshev-type inequality for sums of independent random variables // The Ann. of Math. Stat. 1966. V. 37. P. 248–259.

BEHAVIOUR OF BINOMIAL DISTRIBUTIONS NEAR ITS MEDIANS

N. A. Volkov[a], D. I. Dmitriev[b], and M. E. Zhukovskii[a]

[a] *Moscow Institute of Physics and Technology (National Reserch University), Dolgoprudnyi, Moscow Region, Russia*

[b] *ETH Zurich and ETH AI Center, Zurich, Switzerland*

Presented by Academician of the RAS V.V. Kozlov

We study a behaviour of the value of cumulative distribution function of a binomial random variable with parameters n and $b/(n+c)$ at point $b-1$ for positive integers $b \le n$ and real $c \in [0, 1]$. Our results can be applied directly to the well known problem about small deviations of sums of independents random variables from their expectations. Moreover, we give an answer to a question about the monotonicity of the Ramanujan function for the binomial distribution posed by Jogdeo and Samuels in 1968.

Keywords: binomial distribution, median, Ramanujan function, small deviations of sums of independent random variables