

Об асимптотическом приближении статистики, построенной по выборке объема отрицательно биномиального распределения

Р.В. Пирогов

Аннотация—В работе доказана теорема о предельном распределении отрицательно биномиальных случайных сумм. Продемонстрированы результаты моделирования отрицательно биномиальных случайных сумм с помощью языка программирования R.

Как известно, в классических задачах математической статистики объем выборки является известным параметром. В таких задачах обычно статистика случайных сумм сходится к распределению Гаусса. Дополнительно можно утверждать, что математические ожидания независимых случайных слагаемых не влияют на структурные особенности самого предельного распределения, не считая математического ожидания предельного распределения. В случае случайных сумм, где объем выборки также является статистикой, предельное распределение случайных сумм со слагаемыми с нулевым математическим ожиданием от предельного распределения случайных сумм со слагаемыми с ненулевым математическим ожиданием может отличаться. Сформулированная теорема показывает, что в случае, если накопление слагаемых в случайных суммах имеет характер отрицательно биномиального распределения, то структура предельного распределения зависит от симметричности случайных слагаемых.

Для оценки скорости сходимости доказана Лемма, что отрицательно биномиальные случайные суммы являются смешанными пуассоновскими случайными суммами со смешивающим гамма распределением. Поэтому точность асимптотических моделей отрицательно биномиальных случайных сумм можно оценивать по оценкам скорости сходимости смешанных пуассоновских сумм.

Ключевые слова—Смешанная пуассоновская случайная сумма; отрицательно биномиальная случайная сумма; геометрическая случайная сумма; функция распределения; выборка случайного объема; распределение Лапласа; гамма распределение; экспоненциальное распределение.

I. ВВЕДЕНИЕ

В классических задачах математической статистики объем выборки считается известным параметром. На практике же объем выборки чаще всего также неизвестен и является случайной величиной. Обычно задачи с подобной особенностью могут предоставить исследователю временной интервал, в течение которого с определенной вероятностью происходят случайные события, то есть статистические данные накапливаются в течение фиксированного промежутка времени. Примеры выборов со случайным количеством исходов проявляются в страховании, когда страховые случаи случайным образом накапливаются за фиксированный страховой период, и их число варьируется от периода к периоду. Данная особенность может быть справедлива и к другим областям, таким как медицина, где число пациентов варьируется от

года к году, технике, где фиксируется количество отказов, и прочим сферам, где вместо детерминированного объема дан промежуток времени, в течение которого фиксируется наступления случайных событий. Таким образом, число наблюдений перестает быть параметром и становится наблюдением, то есть статистикой. В силу указанных обстоятельств вполне естественным становится изучение асимптотического поведения статистик достаточно общего вида, основанных на выборках случайного объема.

Примером использования количества наблюдений в качестве статистики является работа Б.В. Гнеденко [1], в которой рассматриваются асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема и показано, что при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут быть радикально разными. К примеру, если объем выборки является геометрически распределенной случайной величиной, то вместо ожидаемого, в соответствии с классической теорией нормального закона, в качестве асимптотического распределения выборочной медианы возникает распределение Стьюдента с двумя степенями свободы, хвосты которого столь тяжелы, что у него отсутствуют моменты больше второго порядка. «Тяжесть» хвостов асимптотических распределений играет ключевую роль в задачах проверки гипотез и расчета квантилей.

В данной работе получены предельные распределения для сумм наблюдений со случайным объемом отрицательно биномиального распределения. Далее будем называть такие суммы отрицательно биномиальными случайными суммами. Предельный закон для отрицательно биномиальных случайных сумм отличается от классических предельных законов. Предельное распределение зависит от математического ожидания случайных слагаемых и имеет тяжелые хвосты.

II. ПРЕДЕЛЬНОЕ РАСПРЕДЕЛЕНИЕ ДЛЯ ОТРИЦАТЕЛЬНО БИНОМИАЛЬНЫХ СЛУЧАЙНЫХ СУММ

Для начала дадим определение случайным суммам и напомним некоторые их свойства.

Определение 1. Пусть N, X_1, X_2, \dots – независимые случайные величины на некотором вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, N принимает целые неотрицательные значения, что формально означает $P(N \in \mathbb{N}_0) = 1$. Тогда случайная величина

$$S_N(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega), \omega \in \Omega \quad (1)$$

где, по определению будем полагать $\sum_{i=1}^0(\cdot) := 0$, называется случайной суммой.

Свойства случайных сумм:

- 1) Математическое ожидание и дисперсия случайной суммы

$$\begin{aligned} E[S_N] &= E[N] E[X], \\ D[S_N] &= E[N] D[X] + D[N](E[X])^2, \end{aligned} \quad (2)$$

если соответствующие моменты случайных величин N и X конечны.

- 2) Характеристическая функция случайной суммы

$$f_{S_N}(t) := E[e^{itS_N}] = \psi(f(z)), t \in \mathbb{R}, \quad (3)$$

где $\psi(\cdot)$ – производящая функция случайной величины N , $f(\cdot)$ – характеристическая функция случайной величины X .

Предельное распределение отрицательно биномиальных случайных сумм с нулевым математическим ожиданием у слагаемых отличается от предельного распределения отрицательно биномиальных случайных сумм с ненулевым математическим ожиданием у слагаемых. Сформулируем следующую теорему и докажем её методом характеристических функций.

Теорема 1. Пусть X_1, X_2, \dots – независимые одинаково распределенные случайные величины с $E[X^2] = 1$, $N_{r,p} \sim NB(r, p)$ и $N_{r,p}, X_1, X_2, \dots$ независимы на некотором вероятностном пространстве $(\Omega, \mathfrak{A}, P)$.

Если $E[X_1] = 0$, то

$$S_1 = \sqrt{p} \sum_{i=1}^{N_{r,p}} X_i \Rightarrow \xi_1 + \dots + \xi_r, p \rightarrow 0, \quad (4)$$

где $\xi_{i=1}^r$ – независимые одинаково распределенные случайные величины, имеющие распределение Лапласа.

Если $E[X_1] \neq 0$, то

$$S_2 = \frac{p}{E[X_1]} \sum_{i=1}^{N_{r,p}} X_i \Rightarrow \Gamma(r, 1) \stackrel{d}{=} \xi_1 + \dots + \xi_r, p \rightarrow 0, \quad (5)$$

где $\xi_{i=1}^r$ – независимые одинаково распределенные случайные величины, имеющие экспоненциальное распределение $Exp(1)$.

Доказательство. Если $E[X_1] = 0$. Используя 3 получим характеристическую функцию суммы S_1 :

$$f_{S_1}(t) = \psi_{N_{r,p}}(f_{X_1}(t\sqrt{p})), \quad (6)$$

где $\psi_{N_{r,p}}(\cdot)$ – производящая функция случайной величины $N_{r,p}$, $f_{X_1}(\cdot)$ – характеристическая функция случайной величины X_1 . Подставим производящую функцию отрицательного биномиального распределения $\psi_{N_{r,p}}(z) = \left(\frac{p}{1-(1-p)z}\right)^r$ в 6

$$\begin{aligned} f_{S_1}(t) &= \left(\frac{p}{1-(1-p)f_{X_1}(t\sqrt{p})}\right)^r = \\ &= \left(\frac{p}{1-(1-p)E[e^{itX_1\sqrt{p}}]}\right)^r \end{aligned} \quad (7)$$

Воспользуемся разложением Тейлора для асимптотического разложения характеристической функции случайной величины X_1 и получим

$$\begin{aligned} f_{S_1}(t) &= \left(\frac{p}{1-(1-p)E[e^{itX_1\sqrt{p}}]}\right)^r = \\ &= \left(\frac{p}{1-(1-p)\left(1+it\sqrt{p}E[X] - \frac{t^2pE[X^2]}{2} + o(t^3p^{3/2})\right)}\right)^r = \\ &= \left(\frac{p}{1-(1-p)\left(1 - \frac{t^2p}{2} + o(t^3p^{3/2})\right)}\right)^r = \\ &= \left(\frac{1}{1+(1-p)\frac{t^2}{2} - o(t^3p^{1/2})}\right)^r p \xrightarrow{p \rightarrow 0} 0 \\ &= \left(\frac{2}{2+t^2}\right)^r = (f_{\xi_1}(t))^r \end{aligned}$$

где $f_{\xi_1}(t)$ – характеристическая функция распределения Лапласа. Первая часть теоремы доказана.

Если $E[X_1] \neq 0$. Используя 3 получим характеристическую функцию суммы S_2 и произведем те же операции, что и в доказательстве первой части теоремы

$$\begin{aligned} f_{S_2}(t) &= \psi_{N_{r,p}}\left(f_{X_1}\left(t\frac{p}{E[X_1]}\right)\right) = \\ &= \left(\frac{p}{1-(1-p)f_{X_1}\left(t\frac{p}{E[X_1]}\right)}\right)^r = \\ &= \left(\frac{p}{1-(1-p)E[e^{itX_1\frac{p}{E[X_1]}}]}\right)^r = \\ &= \left(\frac{p}{1-(1-p)\left(1+itE[X\frac{p}{E[X_1]}] - o(t^2p^2)\right)}\right)^r = \\ &= \left(\frac{p}{1-(1-p)(1+itp - o(t^2p^2))}\right)^r = \\ &= \left(\frac{p}{1-1-itp+p+itp^2+o(t^2p^2)}\right)^r = \\ &= \left(\frac{1}{1-it+itp - o(t^3p)}\right)^r p \xrightarrow{p \rightarrow 0} 0 \\ &= \left(\frac{1}{1-it}\right)^r = (f_{\xi_1}(t))^r \end{aligned}$$

где $f_{\xi_1}(t)$ – характеристическая функция экспоненциального распределения. Вторая часть теоремы доказана. \square

Для моделирования отрицательно биномиальных случайных сумм, в качестве распределения объема возьмем геометрическое распределение, то есть $N \sim Geom(p) \stackrel{d}{=} NB(1, p)$ с $p = \frac{1}{n}$, где $n = 10^6$ и назовем такую сумму геометрической случайной суммой. Смоделируем оба случая из теоремы 1:

1) В качестве слагаемых с нулевым математическим ожиданием возьмем $\xi - 2$, где $\xi \sim Bi(4, 0.5)$. Частотная гистограмма (рисунок 1) и нормальный qq-график (рисунок 2) указывают на симметричное распределение с тяжелыми хвостами.

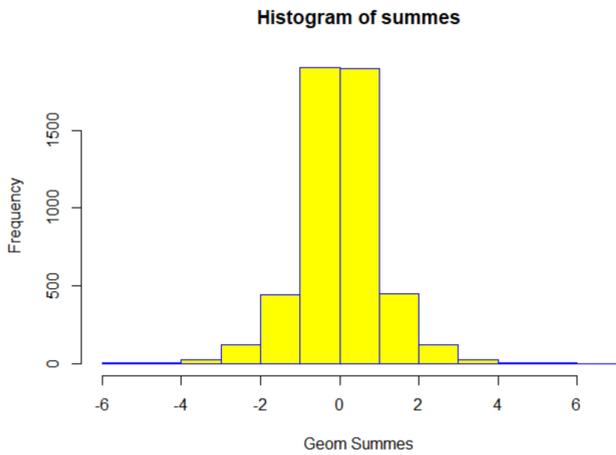


Рис. 1. Частотная диаграмма выборки геометрических случайных сумм с нулевым математическим ожиданием у слагаемых.

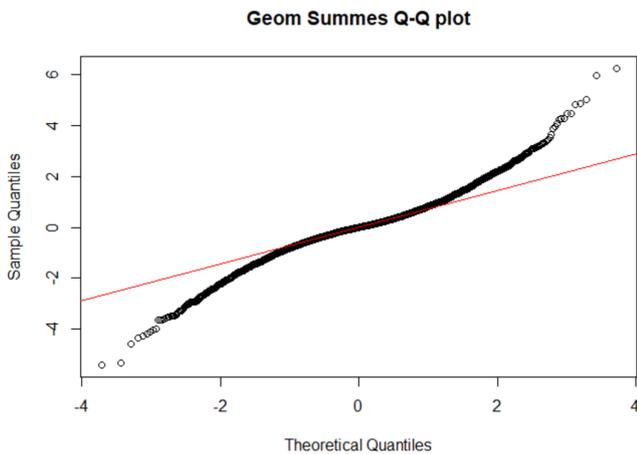


Рис. 2. Нормальный qq-график выборки геометрических случайных сумм с нулевым математическим ожиданием у слагаемых.

Вероятностная гистограмма выборки геометрических случайных сумм с нулевым математическим ожиданием у слагаемых графически аппроксимируется распределением Лапласа (рисунок 3). На диаграммах разброса (рисунок 4) у выборки пик выражен сильнее, чем у распределения Лапласа. По теореме 1 пик будет сглаживаться при $p \rightarrow 0$. Дополнительно можно сделать вывод, что в районе высоких квантилей аппроксимация по распределению Лапласа даст оценки сверху, что с практической точки зрения важнее.

2) В качестве слагаемых с ненулевым математическим ожиданием возьмем $\xi - 1$, где $\xi \sim Bi(4, 0.5)$. Частотная гистограмма (рисунок 5) и нормальный qq-график (рисунок 6) показывают ассиметрию и тяжесть единственного хвоста. Вероятностная гистограмма выборки геометрических случайных сумм с ненулевым математическим

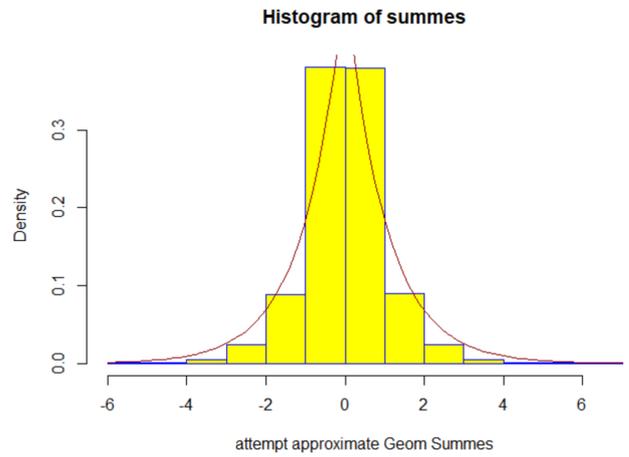


Рис. 3. Вероятностная гистограмма выборки геометрических случайных сумм с нулевым математическим ожиданием у слагаемых, с плотностью распределения Лапласа.

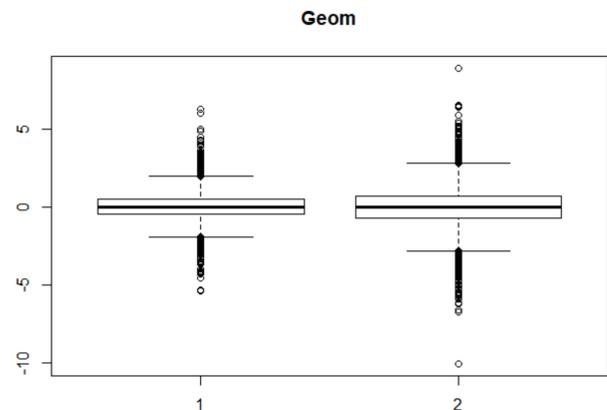


Рис. 4. Диаграммы разброса выборки геометрических случайных сумм с нулевым математическим ожиданием у слагаемых и выборки распределения Лапласа.

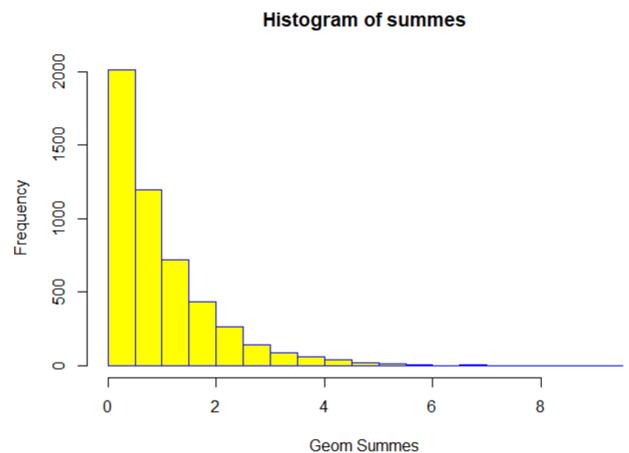


Рис. 5. Частотная диаграмма выборки геометрических случайных сумм с ненулевым математическим ожиданием у слагаемых.

ским ожиданием у слагаемых графически аппроксимируется плотностью от распределения $Exp(1)$ (рисунок 7). Диаграммы разброса (рисунок 8) показывают, что распределения имеют одинаковую структуру.

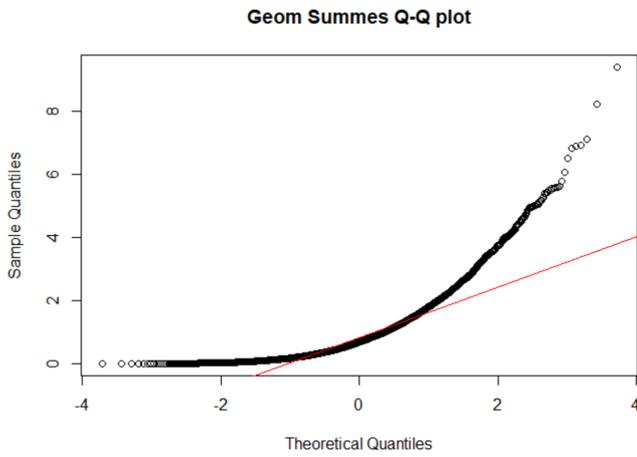


Рис. 6. Нормальный qq-график выборки геометрических случайных сумм с ненулевым математическим ожиданием у слагаемых.

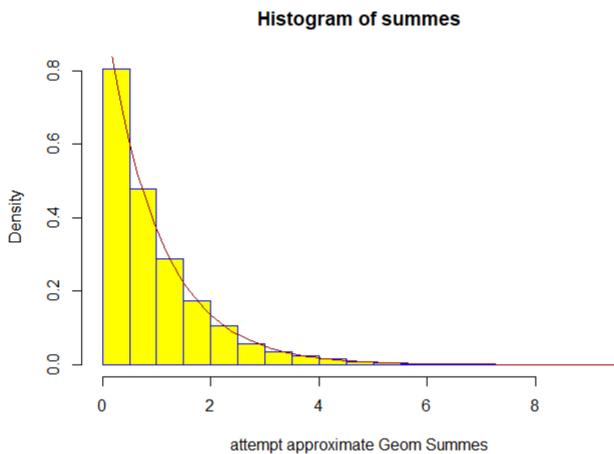


Рис. 7. Вероятностная гистограмма выборки геометрических случайных сумм, с ненулевым математическим ожиданием у слагаемых, с плотностью от распределения $Exp(1)$.

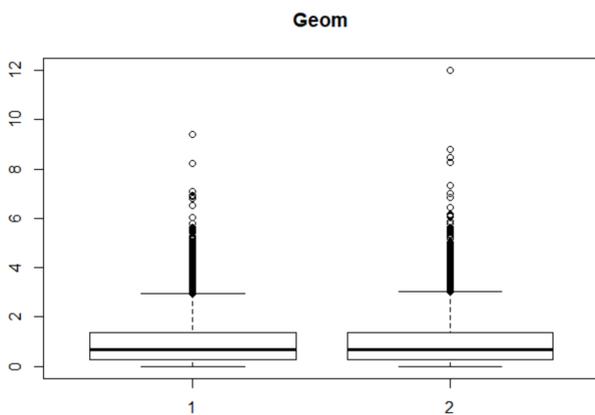


Рис. 8. Диаграммы разброса выборки геометрических случайных сумм с ненулевым математическим ожиданием у слагаемых и выборки распределения $Exp(1)$.

III. Оценка точности асимптотических моделей отрицательно биномиальных случайных сумм

Отрицательно биномиальные случайные суммы относятся к смешанным пуассоновским случайным суммам.

Смешанные пуассоновские случайные суммы охватывают целый класс случайных сумм, где закон накопления случайных слагаемых может описываться различными распределениями.

Определение 2. Пусть Λ – почти наверное положительная случайная величина с функцией распределения G . Будем говорить, что целочисленная неотрицательная случайная величина N имеет смешанное распределение Пуассона со структурным (смешивающим) распределением Λ (или G) и писать $N \sim MP(\Lambda)$, если

$$P(N = k) = \frac{1}{k!} \int_0^\infty \lambda^k e^{-\lambda} dG(\lambda), \quad k = 0, 1, 2, \dots \quad (8)$$

Если ввести случайную величину $N_\lambda \sim Pois(\lambda)$, независимую от Λ при каждом $\lambda > 0$, то можно записать $N \stackrel{d}{=} N_\Lambda$

Определение 3. Если X_1, X_2, \dots одинаково распределены, $N \sim MP(\Lambda)$ и N, X_1, X_2, \dots независимы на некотором вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, то случайная величина

$$S(\omega) := X_1(\omega) + \dots + X_{N(\omega)}(\omega), \quad \omega \in \Omega, \quad (9)$$

называется смешанной пуассоновской случайной суммой, а её распределение – обобщенным смешанным пуассоновским (compound mixed Poisson). Примем, что если $N = 0$, то $S = 0$.

Пусть теперь $\Lambda = \Lambda_t \stackrel{п.н.}{>} 0$ – случайная величина, распределение которой зависит от параметра $t > 0$, и случайная величина $N(t) \sim MP(\Lambda_t)$ такова, что $N(t), X_1, X_2, \dots$ независимы при каждом $t > 0$. Тогда такую случайную сумму будем обозначать следующим образом

$$S(t) := X_1 + \dots + X_{N(t)}. \quad (10)$$

Для оценки смешанных пуассоновских случайных сумм существуют аналоги неравенства Берри-Эссеена [2]. К примеру для оценки скорости сходимости смешанных пуассоновских случайных сумм с нулевым математическим ожиданием у слагаемых есть следующие две теоремы, взятые из [3]:

Теорема 2. Пусть X_1, X_2, \dots – одинаково распределенные случайные величины. Предположим, что $E[X] = 0$, $E[X^2] = 1$ и $\Lambda_t \xrightarrow{P} \infty$ при $t \rightarrow \infty$. Тогда для положительной неограниченно возрастающей функции $d(t)$ имеет место слабая сходимость

$$\frac{S(t)}{\sqrt{d(t)}} \Rightarrow Y$$

к некоторой случайной величине Y , тогда и только тогда, когда найдется такая случайная величина Λ , что при той же функции $d(t)$

$$\frac{\Lambda_t}{d(t)} \Rightarrow \Lambda \text{ и } Y \stackrel{d}{=} Z\sqrt{\Lambda}, \quad (11)$$

где $Z \sim N(0, 1)$.

Теорема 3. Пусть $E[X] = 0$, $E[X^2] = 1$, $E[|X|^{2+\delta}] < \infty$ для некоторого $\delta \in (0, 1]$, случайные величины $Z \sim$

$N(0, 1)$, $\Lambda > 0$, Λ_t – независимы при каждом $t > 0$, и $d(t)$ – положительная функция, определенная для всех $t > 0$. Положим

$$\Delta_t := \rho \left(\frac{S(t)}{\sqrt{d(t)}}, Z\sqrt{\Lambda} \right),$$

$$\delta_t := \rho \left(Z\sqrt{\frac{\Lambda_t}{d(t)}}, Z\sqrt{\Lambda} \right), \quad t > 0.$$

В частности для одинаково распределенных слагаемых

$$\delta_t \leq \frac{1}{2}\rho \left(\frac{\Lambda_t}{d(t)}, \Lambda \right), \quad t > 0.$$

Тогда для всех $t > 0$

$$\Delta_t \leq M(\delta) E[|X|^{2+\delta}] E[\Lambda_t^{-\frac{\delta}{2}}] + \delta_t, \quad (12)$$

где $M(\delta)$ – константа, которая минимизируется по δ .

С константами в неравенстве Берри-Эссеена можно ознакомиться в [5].

С доказательством теоремы 3 можно ознакомиться в [10].

Получается для того, чтобы оценить отрицательно биномиальные случайные суммы, требуется найти смешивающее распределение для объема. Сформулируем и докажем следующую Лемму.

Лемма 1. *Отрицательно биномиальная случайная сумма с $N \sim NB \left(r, \frac{1}{\theta+1} \right)$, $\forall \theta, r \in \mathbb{N}$ – смешанная пуассоновская случайная сумма со смешивающим гамма распределением $\Gamma(r, \theta)$ с плотностью $f(x) = \frac{e^{-\frac{x}{\theta}}}{\theta^r \Gamma(r)} \mathbb{1}(x \geq 0)$, где $\Gamma(r)$ – гамма-функция Эйлера.*

Доказательство. Докажем Лемму подставив $\Gamma(r, \theta)$ в 8.

$$\begin{aligned} P(N = k) &= \frac{1}{k!} \int_0^\infty \lambda^k e^{-\lambda} \lambda^{r-1} \frac{e^{-\frac{\lambda}{\theta}}}{\theta^r \Gamma(r)} = \\ &= \frac{1}{\theta^r k! \Gamma(r)} \int_0^\infty \lambda^{k+r-1} e^{-\lambda(1+\frac{1}{\theta})} d\lambda = \\ &= \frac{\Gamma(r+k)}{k! \Gamma(r)} \frac{1}{\theta^r (1+\frac{1}{\theta})^{r+k}} = \\ &= C_{r+k-1}^k \left(\frac{1}{\theta+1} \right)^r \left(\frac{\theta}{\theta+1} \right)^k = NB \left(r, \frac{1}{\theta+1} \right) \end{aligned}$$

□

IV. ЗАКЛЮЧЕНИЕ

В работе была сформулирована и доказана теорема о предельном распределении отрицательно биномиальных случайных сумм. Была проанализирована выборка отрицательно биномиальных случайных сумм. Результаты анализа не противоречат сформулированной теореме.

Была сформулирована и доказана Лемма о принадлежности отрицательно биномиальных случайных сумм смешанным пуассоновским случайным суммам со смешивающим гамма распределением. Был описан способ оценки точности асимптотических моделей отрицательно биномиальных случайных сумм.

СПИСОК ЛИТЕРАТУРЫ

- [1] Гнеденко Б.В. *Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений* 1989: Труды Тбилисского Государственного Университета.
- [2] Berry A. C. *The accuracy of the Gaussian approximation to the sum of independent variates* // Trans. Amer. Math. Soc. – 1941. – Vol. 49. – P. 122-136.
- [3] Шевцова И.Г. *Оценки точности асимптотических вероятностных моделей* 2018.
- [4] Шевцова И.Г. *Об абсолютных константах в неравенстве Берри-Эссеена и его структурных и неравномерных уточнениях* // Информ. и её примен. – 2013. – Т. 7, № 1. – С. 124-125.
- [5] Шевцова И.Г. *Об абсолютных константах в неравенствах типа Берри-Эссеена* // ДАН. – 2014. – Т. 456, № 6. – С. 650-654.
- [6] Grandell J. *Mixed Poisson processes*. – London : Chapman and Hall, 1997.
- [7] Б. В. Гнеденко, Гуссейн Фахим, *Об одной теореме переноса*, Докл. АН СССР, 1969, том 187, номер 1, С. 15-17.
- [8] Bening V. E., Korolev V. Y. *Generalized Poisson Models and their Applications in Insurance and Finance*. – Utrecht, The Netherlands : VSP, 2002.
- [9] Korolev V. Y. *A general theorem on the limit behavior of superpositions of independent random processes with applications to Cox processes*. *Journal of Mathematical Sciences* // J. Math. Sci. – 1996. – Vol. 81, no. 5. – P. 2951-2956.
- [10] Гавриличенко С.В., Королев В.Ю. *Оценки скорости сходимости смешанных пуассоновских случайных сумм* // Системы и средства информатики. Специальный выпуск. – 2006. – С. 248-257.
- [11] Петров В.В. *Суммы независимых случайных величин*. // М.: Наука, 1972.
- [12] <https://www.rdocumentation.org/RDocumentation> 2018.

Пирогов Роман Владимирович, магистрант кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М.В.Ломоносова.

Россия, 119992, г. Москва, ГСП - 1, Воробьевы горы, МГУ им. М.В. Ломоносова.

E - mail: pirogov1007@gmail.com.

Asymptotic approximation of statistics based on the sample of negative binomial distribution

Roman Pirogov

Abstract—The paper proves the theorem on the limit distribution of negative binomial random sums. The results of modeling negative binomial random sums using the R programming language are demonstrated.

As is well-known, in classical problems of mathematical statistics sample volume is a known parameter. In these problems, the statistics of random sums usually converges to the distribution of Gauss. Additionally, it can be stated that mathematical expectations of independent random terms do not affect the structural features of the limit distribution itself, except mathematical expectations of the limit distribution. In case of random sums, where the sample volume is also a statistic, the limit distribution of random sums with the sums of terms with zero mathematical expectation from the limit distribution of random sums with the sums of terms with non-zero mathematical expectation may be different. The formulated theorem shows that if the accumulation of terms in random sums has the character of a negative binomial distribution, the structure of the limit distribution depends on the symmetry of the random terms.

To estimate the convergence rate, Lemma proved that negative binomial random sums are mixed Poisson random sums with a mixing gamma distribution. Therefore, accuracy of asymptotic models of negative binomial random sums can be estimated by estimating the convergence rate of mixed Poisson sums.

Keywords—Mixed poisson random sum; negative binomial random sum; geometric random sum; distribution function; random volume sampling; Laplace distribution; gamma distribution; exponential distribution.

References

- [1] Gnedenko B.V. *Estimation of unknown distribution parameters at a random number of independent observations* 1989: Works of Tbilisi State University.
- [2] Berry A. C. *The accuracy of the Gaussian approximation to the sum of independent variates* // Trans. Amer. Math. Soc. – 1941. – Vol. 49. – P. 122-136.
- [3] Shevtsova I.G. *Estimates of the accuracy of asymmetric probabilistic models* 2018.
- [4] Shevtsova I.G. *Absolute constants in the Berry-Esseen inequality and its structural and non-uniform clarifications* – 2013. – vol. 7, no. 1. – p. 124-125.
- [5] Shevtsova I.G. *Absolute constants in inequalities like Berry-Esseen* – 2014. – vol. 456, no. 6. – p. 650-654.
- [6] Grandell J. *Mixed Poisson processes*. – London : Chapman and Hall, 1997.
- [7] Gnedenko B.V., Gussein Fakhim, *One transfer theorem* 1969, vol. 187, no. 1, p. 15-17.
- [8] Bening V. E., Korolev V. Y. *Generalized Poisson Models and their Applications in Insurance and Finance*. – Utrecht, The Netherlands : VSP, 2002.
- [9] Korolev V. Y. *A general theorem on the limit behavior of superpositions of independent random processes with applications to Cox processes*. *Journal of Mathematical Sciences* // J. Math. Sci. – 1996. – Vol. 81, no. 5. – P. 2951-2956.
- [10] Gavrilichenko S.V., Korolev V.U. *Estimates of the convergence rate of mixed Poisson random sums* // Systems and tools of informatics. Special edition. – 2006. – p. 248-257.
- [11] Petrov V.V. *Sums of independent random variables*. – 1972.
- [12] <https://www.rdocumentation.org/RDocumentation> 2018.