ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

ОБ АСИМПТОТИЧЕСКОМ ПОВЕДЕНИИ КВАНТИЛЕЙ РАСПРЕДЕЛЕНИЙ СТАТИСТИК, ОСНОВАННЫХ НА ВЫБОРКАХ СЛУЧАЙНОГО ОБЪЕМА[1]

Бенинг В.Е.
МГУ имени М.В. Ломоносова, г. Москва
Институт проблем информатики РАН, г. Москва

Поступила в редакцию 23.06.2017, после переработки 25.07.2017.

В работе доказана общая теорема, позволяющая получать асимптотические разложения для квантилей функций распределения ненормированных статистик, основанных на выборках случайного объема, из асимптотических разложений для обратных моментов нормированного случайного объема выборки и асимптотических разложений для функций распределения статистик, основанных на выборках неслучайного объема.

Ключевые слова: квантиль, ненормированная статистика, функция распределения, выборка случайного объема, асимптотическое разложение, трехточечное симметричное распределение.

Вестник ТвГУ. Серия: Прикладная математика. 2017. № 3. С. 5–12.
https://doi.org/10.26456/vtpmk175

1. Введение

В классических задачах математической статистики объем выборки или количество наблюдений, доступных исследователю, традиционно считается детерминированным и в асимптотических постановках играет роль (как правило, неограниченно возрастающего) *известного* параметра. Однако на практике часто возникают ситуации, когда размер выборки не является заранее определенным и может рассматриваться как случайный. Подобного рода ситуации, как правило, связаны с тем, что статистические данные накапливаются в течение фиксированного «времени». Это имеет место, в частности, в страховании, когда в течение разных отчетных периодов одинаковой длины (скажем, месяцев или лет) происходит разное число страховых событий (страховых выплат и/или заключений страховых контрактов), в медицине, когда число пациентов с тем или иным заболеванием варьируется от года к году, в технике, когда при испытании на надежность (скажем, при определении наработки на отказ) разных партий приборов (изделий), число

[1]Работа выполнена при финансовой поддержке РФФИ (проект № 15-07-02652).

отказавших приборов в разных партиях будет разным и заранее неопределенным. В таких ситуациях число наблюдений, которые будут доступны исследователю, и заранее не известное, разумно считать случайной величиной. Другими словами, в таких ситуациях объем выборки является не (известным) параметром, а сам становится *наблюдением*, то есть статистикой. В силу указанных обстоятельств вполне естественным становится изучение асимптотического поведения распределений статистик достаточно общего вида, основанных на выборках случайного объема.

На естественность такого подхода, в частности, обратил внимание Б.В. Гнеденко в работе [1], в которой рассматривались асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема, и было продемонстрировано, что при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут радикально измениться. 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 порядков, больших второго. «Тяжесть» хвостов асимптотических распределений же имеет критически важное значение, в частности, в задачах проверки гипотез.

Простейшей статистикой является сумма наблюдений. Для выборок случайного объема число слагаемых в таких суммах само становится случайным. Асимптотическим свойствам распределений сумм случайного числа случайных величин посвящено много работ (см., например, [1–3]). Такого рода суммы находят широкое применение в страховании, экономике, биологии и т.п. (см. [4–5]). В классической статистике суммирование наблюдений как правило возникает при определении выборочных средних. При статистическом анализе, основанном на моделях, в которых объем выборки считается неслучайным, асимптотическое поведение статистик типа сумм и статистик типа средних арифметических одинаково – эти статистики после нормировки, обязательной для получения нетривиальных предельных распределений, становятся неразличимыми. Однако, как уже говорилось, в реальной практике очень часто объем выборки сам является статистикой, и, как недавно показано, например, в работе [5], асимптотическое поведение статистик типа сумм и статистик типа средних арифметических при их неслучайной нормировке оказывается различным. Заметим, что, конечно же, формально допустима и случайная нормировка, но для построения разумных асимптотических аппроксимаций для распределений статистик (а именно это и является целью асимптотической статистики), она неприменима. Именно использованием неслучайной нормировки и объясняется возникновение не «чистого» нормального закона, а (разных!) смешанных нормальных предельных распределений у статистик типа сумм и типа средних арифметических. При этом различие этих предельных законов может дать дополнительную информацию о структуре исходных данных.

В данной работе получены асимптотические разложения (а.р.) для квантилей функций распределений (ф.р.) статистик, построенных по выборкам случайного объема. Эти а.р. непосредственно зависят от а.р. обратных моментов случайного объема выборки и а.р. ф.р. статистики, основанной на неслучайной выборке. Подобного рода утверждения принято называть теоремами переноса. Таким образом, в данной работе доказаны теоремы переноса для а.р. квантилей функций распределения статистик, построенных по выборкам случайного объема.

В работе приняты следующие обозначения: $\mathbb{R}$ – множество вещественных чисел, $\mathbb{N}$ – множество натуральных чисел, $\Phi(x)$, $\varphi(x)$ – соответственно ф.р. и плотность стандартного нормального закона.

В разделе 2 приведена основная теорема, касающаяся а.р. квантилей, в разделе 3 рассмотрен пример.

2. Асимптотические разложения для функций распределения ненормированных статистик

В этом разделе будет построено асимптотическое разложение для функций распределения ненормированных статистик, основанных на выборках случайного объема из соответствующего а.р. для случая неслучайного размера выборки. Такого рода преобразования а.р. назовем теоремами переноса для а.р.

Рассмотрим случайные величины (с.в.) $N_1, N_2, ...$ и $X_1, X_2, ...$, заданные на одном и том же вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$. В статистике с.в. $X_1, X_2, ...X_n$ имеют смысл наблюдений, n – неслучайный объем выборки, а с.в. N_n – случайный объем выборки, зависящий от натурального параметра $n \in \mathbb{N}$. Например, если с.в. N_n имеет геометрическое распределение вида

$$\mathsf{P}(N_n = k) = \frac{1}{n}\Big(1 - \frac{1}{n}\Big)^{k-1}, \ k \in \mathbb{N},$$

то

$$\mathsf{E}N_n = n, \tag{2.1}$$

то есть среднее значение случайного объема выборки равно n.

Предположим, что для каждого $n \geq 1$ с.в. N_n принимает только натуральные значения (то есть, $N_n \in \mathbb{N}$) и не зависит от последовательности с.в. $X_1, X_2, \ldots$. Всюду далее предполагается, что случайные величины $X_1, X_2, \ldots$ независимы одинаково распределены.

Для каждого $n \geq 1$ обозначим через $T_n = T_n(X_1, \ldots, X_n)$ некоторую статистику, то есть действительную измеримую функцию, зависящую от наблюдений $X_1, \ldots, X_n$. Для каждого $n \geq 1$ определим статистику T_{N_n}, зависящую от выборки случайного объема как

$$T_{N_n}(\omega) \equiv T_{N_n(\omega)}(X_1(\omega), \ldots, X_{N_n(\omega)}(\omega)), \ \omega \in \Omega.$$

Следующее условие описывает а.р. для ф.р. статитсики T_n в случае неслучайного объема выборки.

Условие А. *Существуют числа* $k \in \mathbb{N}, k \geq 2$, $\alpha_{in} \in \mathbb{R}, i = 1, \ldots, k, \beta_n > 0$, $C_k > 0$, *дифференцируемая ф.р.* $F(x)$ *и измеримые функции* $f_j(x), j = 1, ..., k$ *такие, что*

$$\beta_n \to 0, \ \max_{1 \leq i \leq k} |\alpha_{in}| \to 0, \ n \to \infty,$$

$$\sup_x \Big| \mathsf{P}\big(T_n < x\big) - F(x) - \sum_{i=1}^{k} \alpha_{in} f_i(x) \Big| \leq C_k \beta_n, \ n \in \mathbb{N}.$$

Лемма 2.1. *Пусть статистика $T_n = T_n(X_1, ..., X_n)$ удовлетворяет условию А. Тогда*

$$\sup_x \left| \mathsf{P}\big(T_{N_n} < x\big) - F(x) - \sum_{i=1}^{k} \mathsf{E}\alpha_{iN_n} f_i(x) \right| \le C_k \mathsf{E}\beta_{N_n}.$$

Доказательство непосредственно следует из формулы полной вероятности.

Пусть $X_1, X_2, ...$ независимые одинаково распределенные с.в. такие, что

$$\mathsf{E}X_1 = 0,\ \mathsf{E}X_1^2 = 1,\ \mathsf{E}|X_1|^{k+\delta} < \infty,\ k \ge 3,\ k \in \mathbb{N},\ \delta > 0. \tag{2.2}$$

Для каждого n пусть

$$T_n = \frac{1}{\sqrt{n}}(X_1 + ... + X_n). \tag{2.3}$$

Предположим, что с.в. X_1 удовлетворяет условию Крамера

$$\limsup_{|t|\to\infty} |\mathsf{E}\exp\{itX_1\}| < 1. \tag{2.4}$$

При выполнении условий (2.2) и (2.4) из Теоремы 6.3.2 [6] следует выполнение неравенства

$$\sup_x \left| \mathsf{P}\big(T_n < x\big) - \Phi(x) - \sum_{i=1}^{k-2} n^{-i/2} Q_i(x) \right| \le \frac{C_{k,\delta}}{n^{(k-2+\delta)/2}},\ C_{k,\delta} > 0,\ n \in \mathbb{N}, \tag{2.5}$$

где функции $Q_1(x), \ldots, Q_{k-2}(x)$ определены в [6], например,

$$Q_1(x) = -(x^2 - 1)\phi(x)\frac{\mathsf{E}X_1^3}{6},$$

$$Q_2(x) = -(x^3 - 3x)\phi(x)\frac{\mathsf{E}X_1^3}{24} - (x^5 - 10x^3 + 15x)\phi(x)\frac{(\mathsf{E}X_1^3)^2}{72}. \tag{2.6}$$

Учитывая неравенство (2.5) и Лемму 2.1, получаем следующее утверждение.

Лемма 2.2. *Пусть выполнены условия (2.2)–(2.4), тогда*

$$\sup_x \left| \mathsf{P}\big(T_{N_n} < x\big) - \Phi(x) - \sum_{i=1}^{k} \mathsf{E}N_n^{-i/2} Q_i(x) \right| \le C_{k,\delta} \mathsf{E}N_n^{-(k-2+\delta)/2}.$$

Ниже эти результаты будут применены для аппроксимации α-квантилей. Назовем асимптотической α-квантилью ($\alpha \in (0,1)$) статистики S_n величину $c^*_\alpha(n)$, удовлетворяющую асимптотическому равенству

$$\mathsf{P}\big(S_n \ge c^*_\alpha(n)\big) = \alpha + o(n^{-1}),\ n \to \infty. \tag{2.7}$$

Если интерпретировать статистику S_n как страховые требования страховой компании, то асимптотическая α-квантиль $c^*_\alpha(n)$ может рассматриваться как резервный капитал страховой компании, который ей с большой вероятностью $1-\alpha$ желательно не превышать.

Применяя формулу Тейлора, несложно получить следующий результат.

Лемма 2.3. *Пусть для функции распределения статистики S_n равномерно по $x \in \mathbb{R}$ справедливо а.р. вида*

$$\mathsf{P}\big(S_n < x\big) = G(x) + \frac{1}{\sqrt{n}} g_1(x) + \frac{1}{n} g_2(x) + o(n^{-1}),$$

где $G(x), g_1(x), g_2(x)$ – достаточно гладкие функции, тогда для асимптотической α-квантили $c^*_\alpha(n)$ справедливо а.р.

$$c^*_n(\alpha) = c_\alpha - \frac{g_1(c_\alpha)}{\sqrt{n} G'(c_\alpha)} -$$

$$-\frac{1}{n}\Big(\frac{G''(c_\alpha) g_1^2(c_\alpha)}{2(G'(c_\alpha))^3} + \frac{G'(c_\alpha) g_2(c_\alpha) - g_1(c_\alpha) g_1'(c_\alpha)}{(G'(c_\alpha))^2}\Big) + o(n^{-1}),$$

где c_α удовлетворяет уравнению $G(c_\alpha) = 1 - \alpha$.

Из соотношений (2.5), (2.6) и Леммы 2.3 непосредственно следует а.р. для асимптотической α-квантили.

Лемма 2.4. *Пусть для $k = 4, \delta > 0$ выполнены условия (2.2)–(2.4), тогда для асимптотической α-квантили $c^*_\alpha(n)$ справедливо а.р.*

$$c^*_n(\alpha) = u_\alpha + \frac{\mathsf{E} X_1^3}{6\sqrt{n}}(u_\alpha^2 - 1) +$$

$$+\frac{1}{12n}\Big(\frac{\mathsf{E}^2 X_1^3}{3}(5u_\alpha - 2u_\alpha^3) + \frac{\mathsf{E} X_1^3 - 3}{2}(u_\alpha^3 - 3u_\alpha)\Big) + o(n^{-1}),$$

где u_α удовлетворяет уравнению $\Phi(u_\alpha) = 1 - \alpha$.

3. Случай трехточечного симметричного распределения

Применим Лемму 2.2 для получения а.р. асимптотической α-квантили в случае, если объем выборки N_n имеет трехточечное симметричное распределение.

Пусть случайный индекс N_n (случайный объем выборки) имеет симметричное распределение вида

$$N_n : \begin{array}{ccc} n - h_n, & n, & n + h_n \\ \frac{1}{3} & \frac{1}{3}, & \frac{1}{3}, \end{array} \tag{3.1}$$

где последовательность натуральных чисел $h_n < n$ удовлетворяет условию

$$\lim_{n\to\infty} \frac{h_n}{n} = 0. \tag{3.2}$$

Лемма 3.1. *Пусть случайная величина N_n имеет распределение (3.1) и выполнено условие (3.2), тогда справедливы равенства*

$$\mathsf{E} N_n = n,$$

$$\mathsf{E} N_n^{-1/2} = \frac{1}{\sqrt{n}} - \frac{1}{4\sqrt{n}}\Big(\frac{h_n}{n}\Big)^2 + O\Big(\frac{1}{\sqrt{n}}\Big(\frac{h_n}{n}\Big)^3\Big),$$

$$\mathsf{E}N_n^{-1} = \frac{1}{n} + \frac{2}{3n}\Big(\frac{h_n}{n}\Big)^2 + O\Big(\frac{1}{n}\Big(\frac{h_n}{n}\Big)^4\Big),$$
$$\mathsf{E}N_n^{-3/2} = \frac{1}{n^{3/2}} + O\Big(\frac{1}{n^{3/2}}\Big(\frac{h_n}{n}\Big)^2\Big), n \to \infty.$$

Доказательство. Доказательство леммы следует из равенств

$$\mathsf{E}N_n^{-1} = \frac{3n^2 - h_n^2}{3n(n^2 - h_n^2)} =$$
$$= \frac{1}{n}\Big(1 - \frac{h_n^2}{3n}\Big)\Big(1 + \frac{h_n^2}{n^2} + O\Big(\frac{h_n^4}{n^4}\Big)\Big) =$$
$$= \frac{1}{n} + \frac{2}{3n}\Big(\frac{h_n}{n}\Big)^2 + O\Big(\frac{1}{n}\Big(\frac{h_n}{n}\Big)^4\Big),$$
$$\mathsf{E}N_n^{-3/2} = \frac{1}{3n^{3/2}}\Big(\frac{1}{(1 - h_n/n)^{3/2}} + 1 + \frac{1}{(1 + h_n/n)^{3/2}}\Big) =$$
$$= \frac{1}{n^{3/2}} + O\Big(\frac{1}{n^{3/2}}\Big(\frac{h_n}{n}\Big)^2\Big).$$

Аналогично доказываются оставшиеся соотношения. Лемма доказана. □

Теперь из этих лемм непосредственно вытекает следующая теорема.

Теорема 3.1. *Пусть выполнены условия (2.2)–(2.4), (3.1) и с.в. N_n имеет распределение (3.1), тогда для асимптотической α-квантили $c_\alpha(n)$, соответствующей статистике T_{N_n}, справедливо а.р. вида*

$$c_\alpha(n) = c_\alpha^*(n) - \frac{\mathsf{E}X_1^3(u_\alpha^2 - 1)}{24\sqrt{n}}\Big(\frac{h_n}{n}\Big)^2 + o(n^{-1}),\ n \to \infty.$$

Замечание 3.1. *Пусть выполнены условия Теоремы 3.1 и*

$$h_n = \gamma n^\beta + o(n^\beta),\ \gamma \geq 0,\ 0 \leq \beta < 1,$$

тогда

$$n^{5/2-2\beta}\big(c_\alpha^*(n) - c_\alpha(n)\big) \to \frac{\gamma^2}{24}\mathsf{E}X_1^3(u_\alpha^2 - 1),\ n \to \infty.$$

Заключение

В работе получены асимптотические разложения для функций распределения ненормированных статистик, основанных на выборках случайного объема. Эти асимптотические разложения непосредственно зависят от асимптотических разложений обратных моментов случайного индекса и асимптотического разложения функции распределения статистики, основанной на неслучайной выборке. Подобного рода утверждения обычно называются теоремами переноса. Поэтому в работе доказаны теоремы переноса, касающиеся асимптотических разложений. Эти результаты применяются для асимптотического исследования поведения квантилей функций распределения таких статистик. Приведены два примера, иллюстрирующие доказанные теоремы. Первый пример касается сумм независимых случайных величин, а во втором рассматривается трехточечное симметричное распределение. Эти результаты могут применяться в теории риска, например, для изучения асимптотического поведения необходимого резерва страховых компаний.

Список литературы

[1] Гнеденко Б.В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Труды Тбилисского Математического института. 1989. Т. 92. С. 146–150.

[2] Гнеденко Б.В., Фахим Х. Об одной теореме переноса // Доклады Академии наук СССР. 1969. Т. 187. С. 15–17.

[3] Бенинг В.Е., Королев В.Ю. Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения. 2004. Т. 49, № 3. С. 417–435. https://doi.org/10.4213/tvp201

[4] Bening V.E., Korolev V.Yu. Generalized Poisson Models and Their Applications in Insurance and Finance. Utrecht: VSP, 2002.

[5] Бенинг В.Е., Королев В.Ю. Некоторые статистические задачи, связанные с распределением Лапласа // Информатика и ее применения. 2008. Т. 2, № 2. С. 19–34.

[6] Петров В.В. Суммы независимых случайных величин. М.: Наука, 1972.

Образец цитирования

Бенинг В.Е. Об асимптотическом поведении квантилей распределений статистик, основанных на выборках случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2017. № 3. С. 5–12. https://doi.org/10.26456/vtpmk175

Сведения об авторах

1. **Бенинг Владимир Евгеньевич**

   профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова, старший научный сотрудник ИПИ РАН.

   Россия, 119992, г. Москва, ГСП-1, Воробьевы горы, МГУ им. М.В. Ломоносова. E-mail: bening@yandex.ru.

ON ASYMPTOTIC BEHAVIOR OF QUANTILES OF THE DISTRIBUTIONS OF STATISTICS BASED ON THE SAMPLES WITH RANDOM SIZES

Bening Vladimir Evgenyevich
Professor at Mathematical Statistics department,
Lomonosov Moscow State University
Russia, 119992, Moscow, GSP-1, Vorobyovi gory, Lomonosov MSU.
E-mail: bening@yandex.ru

Received 23.06.2017, revised 25.07.2017.

In the paper general theorem concerning the asymptotic expansions of quantiles of distribution functions of statistics based on the sample with random size is proved. Two examples (sums of independent random variables and three-point symmetric distribution) are presented.

Keywords: quantile, sample of random size, asymptotic expansions, three-point symmetric distribution, transfer theorem.

Citation

Bening V.E. On asymptotic behavior of quantiles of the distributions of statistics based on the samples with random sizes. *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2017, no. 3, pp. 5–12. (in Russian). https://doi.org/10.26456/vtpmk175

References

[1] Gnedenko B.V. An estimate of the distribution of the unknown parameters with a random number of independent observations. *Trudy Tbilisskogo Matematicheskogo Instituta* [Proceedings of the Tbilisi Institute of Mathematics], 1989, vol. 92, pp. 146–150. (in Russian)

[2] Gnedenko B.V., Fahim H. On one transfer theorem. *Doklady Akademii nauk SSSR* [Reports of the Academy of Sciences of the USSR], 1969, vol. 187, pp. 15–17. (in Russian)

[3] Bening V.E., Korolev V.Yu. On the use of Student distribution in the problems of probability theory and mathematical statistics. *Teoriya Veroyatnostei i ee Primeneniya* [Probability Theory and its Applications], 2004, vol. 49(3), pp. 417–435. (in Russian) https://doi.org/10.4213/ tvp201

[4] Bening V.E., Korolev V.Yu. Generalized Poisson Models and Their Applications in Insurance and Finance. VSP, Utrecht, 2002.

[5] Bening V.E., Korolev V.Yu. Some statistical problems related to the Laplace distribution. *Informatika i ee Primeneniya* [Informatics and its Applications], 2008, vol. 2(2), pp. 19–34. (in Russian)

[6] Petrov V.V. *Sums of Independent Random Variables.* Nauka Publ., Moscow, 1972.