ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

О ПОВЕДЕНИИ АСИМПТОТИЧЕСКОГО ДЕФЕКТА КВАНТИЛЕЙ РАСПРЕДЕЛЕНИЙ СТАТИСТИК, ОСНОВАННЫХ НА ВЫБОРКАХ СЛУЧАЙНОГО ОБЪЕМА[1]

Бенинг В.Е.
МГУ им. М.В. Ломоносова, г. Москва
Институт проблем информатики Российской академии наук, г. Москва

Поступила в редакцию 28.06.2018, после переработки 09.08.2018.

В работе доказана общая теорема, позволяющая получать асимптотический дефект и 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: асимптотический дефект, квантиль, ненормированная статистика, функция распределения, выборка случайного объема, асимптотическое разложение, трехточечное симметричное распределение.

Вестник ТвГУ. Серия: Прикладная математика. 2018. № 3. С. 42–57.
https://doi.org/10.26456/vtpmk509

Введение

В классических задачах математической статистики объем выборки или количество наблюдений, доступной 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 (например, месяцев или лет) происходит разное число страховых событий (страховых выплат 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,

[1]Работа выполнена при финансовой поддержке РФФИ (проект № 18-07-00252).

при определении наработки на отказ) разных партий приборов (изделий), число 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 исследуется асимптотическое поведение дефектов (то есть добавочного числа наблюдений, см. также работу [11]) квантилей статистик, основанных на выборках случайного объема. Получено также 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 и исследуется поведение асимптотического дефекта.

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.

1. Асимптотический дефект и его свойства

Рассмотрим две статистические процедуры Π_n^* и Π_n с мерами качества β_n^* и β_n соответственно. Здесь n – число наблюдений $X_1, \ldots, X_n$, на которых основаны эти процедуры. При этом предполагается, что статистическая процедура Π_n^* является в некотором смысле «оптимальной», а процедура Π_n – конкурирующей. Например, в задаче статистического оценивания в качестве меры качества обычно выступает среднеквадратичное отклонение оценки от оцениваемой функции, а в задаче проверки статистических гипотез в качестве меры качества критериев рассматривают их мощность.

Обозначим через $m(n)$ необходимое число наблюдений, которое требуется процедуре $\Pi_{m(n)}$, основанной на наблюдениях $X_1, \ldots, X_{m(n)}$, для достижения такого же качества, что и «лучшей» процедуре Π_n^*, основанной на n наблюдениях $X_1, \ldots, X_n$. Ниже рассматривается асимптотический подход, означающий, что $n \to \infty$. Под асимптотической относительной эффективностью (АОЭ) процедуры Π_n по отношению к процедуре Π_n^* понимается предел (в случае его существования и независимости от последовательности $m(n)$) вида (см., например, [9] стр. 305)

$$e \equiv \lim_{n\to\infty} \frac{n}{m(n)}.$$

Например, предположим, что $e = 1/3$, тогда при больших значениях числа наблюдений n величина $m(n)$ приближенно равна $3n$, поэтому процедуре Π_n для достижения такого же качества, что и процедуре Π_n^*, требует примерно в три раза больше наблюдений.

Вместо отношения необходимого числа наблюдений, естественно, можно было бы рассматривать разность вида $m(n) - n$, которая тоже имеет наглядный смысл необходимого дополнительного числа наблюдений, требующихся процедуре Π_n. Однако, исторически сложилось так, что многие авторы сначала исследовали асимптотические свойства отношения $n/m(n)$ (возможно, в силу относительной простоты его поведения).

Впервые общее асимптотическое исследование поведения разности $m(n) - n$ было предпринято в 1970 году Ходжесом и Леманом (см. работу [7]). Они назвали разность $m(n) - n$ дефектом (deficiency) конкурирующей процедуры Π_n относительно процедуры Π_n^* и предложили обозначение

$$d_n = m(n) - n. \tag{2.1}$$

Если предел $\lim_{n\to\infty} d_n$ существует, то он называется *асимптотическим дефектом* процедуры Π_n относительно процедуры Π_n^* и обозначается символом d. Часто d называют просто дефектом Π_n относительно Π_n^*. Заметим, что если АОЭ $e \neq 1$, то $d = \infty$ и этот случай малоинтересен. В работе [7] также было отмечено, что существуют статистические задачи, в которых типичным образом возникает случай $e = 1$ (см., например, книгу [10]), то есть в этом случае понятие АОЭ не дает ответа на вопрос какая процедура лучше и понятие дефекта проясняет эту ситуацию, поскольку в этом случае асимптотический дефект может, в принципе, быть любым.

Предположим, например, что $d = 7$. Тогда для больших значений n величина $m(n)$ равна приближенно $n + 7$. Чтобы получить ту же величину критерия качества процедуре Π_n требуется примерно на семь наблюдений больше, чем процедуре Π_n^*.

Таким образом дефект процедуры Π_n относительно процедуры Π_n^* показывает сколько добавочных наблюдений примерно требуется, если мы настаиваем на использовании процедуры Π_n вместо процедуры Π_n^*, и поэтому создает естественный базис для их асимптотического сравнения в случае $e = 1$. Исследование асимптотического поведения дефекта d_n технически более сложно, чем нахождение предела e. Часто оно требует построения асимптотических разложений (а.р.) для соответствующих функций, характеризующих качество оценок (см., например, книги [8], [9] и [10]).

Напомним, что статистические процедуры Π_n^* и Π_n имеют меры качества β_n^* и β_n соответственно, тогда по определению величины $d_n = m(n) - n$, для каждого n должно выполняться равенство

$$\beta_n^* = \beta_{m(n)}. \tag{2.2}$$

При решении уравнения (2.2) целочисленную величину $m(n)$ следует рассматривать как переменную, принимающую произвольные действительные значения. Для этого можно определить функцию $\beta_{m(n)}$ для нецелых значений $m(n)$ по формуле

$$\beta_{m(n)} = \big(1 - m(n) + [m(n)]\big)\,\beta_{[m(n)]} + \big(m(n) - [m(n)]\big)\,\beta_{[m(n)]+1}$$

(см. работу [7]).

Типичным образом функции β_n^* и β_n неизвестны точно и используются их аппроксимации. Предположим, что справедливы асимптотические разложения вида

$$\beta_n^* = \frac{a}{n^r} + \frac{b}{n^{r+s}} + o\big(n^{-r-s}\big), \tag{2.3}$$

$$\beta_n = \frac{a}{n^r} + \frac{c}{n^{r+s}} + o\big(n^{-r-s}\big), \tag{2.4}$$

где $a,\ b$ и c – некоторые постоянные не зависящие от n, а $r > 0,\ s > 0$ – некоторые константы, определяющие порядок убывания по n этих критериев качества. Первый член в этих асимптотических разложениях одинаков и это отражает тот факт, что АОЭ этих процедур равна единице. Из соотношений (2.1) – (2.4) легко получить, что (см. работу [7] или книгу [9], стр. 310)

$$d_n = \frac{c - b}{r\ a}\, n^{1-s} + o\big(n^{1-s}\big). \tag{2.5}$$

Таким образом асимптотический дефект имеет вид

$$d = \begin{cases} \pm\infty, & 0 < s < 1, \\ \dfrac{c - b}{r\ a}, & s = 1, \\ 0, & s > 1. \end{cases} \tag{2.6}$$

Асимптотический дефект обладает следующим очевидным свойством транзитивности: если дана третья статистическая процедура $\bar{\Pi}_n$, имеющая меру качества $\bar{\beta}_n$, которая имеет а.р. типа (2.4), тогда дефект d процедуры $\bar{\Pi}_n$ относительно процедуры Π_n^* удовлетворяет равенству

$$d = d_1 + d_2,$$

где d_1 – дефект процедуры $\bar{\Pi}_n$ относительно процедуры Π_n и d_2 – дефект процедуры Π_n относительно процедуры Π_n^*.

Случай, когда выполняется равенство $s = 1$, представляется наиболее интересным, поскольку в этом случае асимптотический дефект конечен. Ходжес и Леман в работе [7] привели ряд простых примеров, показывающих естественность возникновения этого случая в математической статистике (см., также книгу [10]).

2. Асимптотический дефект и квантили

Рассмотрим некоторую статистику $T_n = T_n(X_1 \ldots, X_n)$, основанную на n наблюдениях $X_1, \ldots, X_n$. Назовем асимптотической α – квантилью ($\alpha \in (0,1)$) статистики T_n величину $c_\alpha^*(n)$, удовлетворяющую асимптотическому равенству

$$\mathsf{P}\big(\sqrt{n}\, T_n \geq c_\alpha^*(n)\big) = \alpha + o(n^{-1}), \quad n \to \infty. \tag{3.1}$$

Если интерпретировать статистику $\sqrt{n}\, T_n$ как страховые требования, предъявляемые 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 3.1. *Пусть для функции распределения статистики $\sqrt{n}\, T_n$ равномерно по $x \in \mathbb{R}$ справедливо а.р. вида*

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

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

$$c_\alpha^*(n) = c_\alpha - \frac{g_1(c_\alpha)}{\sqrt{n}\, G^{(1)}(c_\alpha)} -$$

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

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

Следствие 3.1. *Пусть $\delta_n \to 0$, $n \to \infty$, тогда в условиях Леммы 3.1 равномерно по $x \in \mathbb{R}$ справедливо равенство*

$$\mathsf{P}\big(\sqrt{n}\, T_n < x + \delta_n\big) =$$

$$= \mathsf{P}\big(\sqrt{n}\, T_n < x\big) + \delta_n\, G^{(1)}(x) + \frac{\delta_n^2}{2}\, G^{(2)}(x) + \frac{\delta_n}{\sqrt{n}}\, g_1^{(1)}(x) + o\Big(\max\big(\delta_n^2,\, \frac{\delta_n}{\sqrt{n}},\, n^{-1}\big)\Big).$$

Рассмотрим теперь другую статистику $S_n = S_n(X_1 \ldots, X_n)$, основанную на n наблюдениях $X_1, \ldots, X_n$, имеющую α – квантиль $c_\alpha(n)$, удовлетворяющую асимптотическому равенству

$$\mathsf{P}\big(\sqrt{n}\, S_n \geq c_\alpha(n)\big) = \alpha + o(n^{-1}), \quad n \to \infty. \tag{3.2}$$

Предположим, что а.р. для ф.р. статистики $\sqrt{n}\, S_n$ имеет вид

$$\mathsf{P}\big(\sqrt{n}\, S_n < x\big) = G(x) + \frac{1}{\sqrt{n}}\, g_1(x) + \frac{1}{n}\, \bar{g}_2(x) + o(n^{-1}), \tag{3.3}$$

где $G(x)$, $g_1(x)$, $\bar{g}_2(x)$ – достаточно гладкие функции, то есть это а.р. отличается от а.р. для ф.р. статистики $\sqrt{n}\ T_n$ только членом порядка $1/n$ (см. Лемму 3.1). Определим последовательность натуральных чисел $\{m(n) = n + d + o(1),\ d \in \mathbb{R},\ \ n = 1, 2, \ldots\}$ равенством (d – асимптотический дефект)

$$\mathsf{P}\big(\sqrt{n}\, S_{m(n)} \geq c_\alpha^*(m(n))\big) = \alpha + o(n^{-1}), \quad n \to \infty, \tag{3.4}$$

то есть это необходимое число наблюдений $m(n)$ для того, чтобы статистика $\sqrt{n}\, S_{m(n)}$ превзошла α – квантиль $c_\alpha^*(m(n))$ статистики $\sqrt{m(n)}\, T_{m(n)}$.

Теорема 3.1. *Пусть выполнены условия Леммы 3.1 и условие (3.3). Тогда асимптотический дефект d имеет вид*

$$d = \frac{2\big(g_2(c_\alpha) - \bar{g}_2(c_\alpha)\big)}{G^{(1)}(c_\alpha)\ c_\alpha} + o(1).$$

Доказательство. Из Леммы 3.1 и условия (3.3) непосредственно следует, что

$$c_\alpha(n) = c_\alpha - \frac{g_1(c_\alpha)}{\sqrt{n}\, G^{(1)}(c_\alpha)} -$$

$$- \frac{1}{n}\Big(\frac{G^{(2)}(c_\alpha) g_1^2(c_\alpha)}{2(G^{(1)}(c_\alpha))^3} + \frac{G^{(1)}(c_\alpha)\bar{g}_2(c_\alpha) - g_1(c_\alpha) g_1^{(1)}(c_\alpha)}{(G^{(1)}(c_\alpha))^2}\Big) + o(n^{-1}) \tag{3.5}$$

и поэтому

$$\delta_n \equiv \sqrt{\frac{m(n)}{n}}\, c_\alpha^*(m(n)) - c_\alpha(m(n)) = \frac{d}{2n}\, c_\alpha - \frac{1}{n}\, \frac{\big(g_2(c_\alpha) - \bar{g}_2(c_\alpha)\big)}{G^{(1)}(c_\alpha)} + o(n^{-1}). \tag{3.6}$$

Далее, с учетом определений $m(n)$ (см. (3.4)) и δ_n, имеем

$$\alpha + o(n^{-1}) = \mathsf{P}\big(\sqrt{n}\, S_{m(n)} \ge c^*_\alpha(m(n))\big) =$$

$$= \mathsf{P}\big(\sqrt{m(n)}\, S_{m(n)} \ge c_\alpha(m(n)) + \delta_n\big). \tag{3.7}$$

Применяя к правой части равенства (3.7) Следствие 3.1, получим

$$\alpha + o(n^{-1}) = \mathsf{P}\big(\sqrt{m(n)}\, S_{m(n)} \ge c_\alpha(m(n))\big) - \delta_n\, G^{(1)}(c_\alpha) + o(n^{-1}).$$

Теперь из (3.2) и (3.6) следует, что

$$d = \frac{2\big(g_2(c_\alpha) - \bar{g}_2(c_\alpha)\big)}{G^{(1)}(c_\alpha)\; c_\alpha} + o(1).$$

Теорема доказана. □

Приведем пример применения Теоремы 3.1.

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

$$\mathsf{E}\, X_1 = 0, \;\; \mathsf{E}\, X_1^2 = 1, \; \mathsf{E}\, |X_1|^{k+\delta} < \infty, \;\; k \ge 3, \; k \in \mathbb{N}, \; \delta > 0. \tag{3.8}$$

Для каждого n пусть

$$T_n = \frac{1}{n}\,(X_1 + \ldots + X_n). \tag{3.9}$$

Предположим, что с.в. X_1 удовлетворяет условию Крамера (C)

$$\limsup_{|t|\to\infty} |\mathsf{E}\, \exp\{itX_1\}| < 1. \tag{3.10}$$

При выполнении условий (3.8) и (3.10) из Теоремы 6.3.2 книги [6] следует справедливость неравенства

$$\sup_x \Big|\mathsf{P}\big(\sqrt{n}\, T_n < x\big) - \Phi(x) - \sum_{i=1}^{k-2} n^{-i/2}\, Q_i(x)\,\Big| \le \frac{C_{k,\delta}}{n^{(k-2+\delta)/2}},\; C_{k,\delta} > 0, \;\; n \in \mathbb{N}, \tag{3.11}$$

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

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

$$Q_2(x) = -(x^3 - 3x)\,\phi(x)\,\frac{\mathsf{E}\, X_1^4 - 3}{24} - (x^5 - 10x^3 + 15x)\,\phi(x)\,\frac{(\mathsf{E}\, X_1^3)^2}{72}. \tag{3.12}$$

Учитывая неравенство (3.11) и Лемму 3.1, получаем следующее утверждение.

Лемма 3.2. *Пусть выполнены условия (3.8) – (3.10) с $k = 3$, тогда для асимптотической α – квантили $c^*_\alpha(n)$ статистики T_n (см. (3.9)) справедливо а.р.*

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

$$+ \frac{1}{12n} \left(\frac{\mathsf{E}^2 \, X_1^3}{3} \, (5u_\alpha \, - \, 2u_\alpha^3) \, + \, \frac{\mathsf{E} \, X_1^4 \, - \, 3}{2} \, (u_\alpha^3 \, - \, 3u_\alpha) \right) \, + \, o(n^{-1}),$$

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

Пусть теперь $Y_1, Y_2, ...$ независимые одинаково распределенные с.в. такие, что

$$\mathsf{E} \, Y_1 \, = \, 0, \;\; \mathsf{E} \, Y_1^2 \, = \, 1, \; \mathsf{E} \, |Y_1|^{4+\delta} \, < \, \infty, \;\; \delta \, > \, 0. \tag{3.13}$$

Для каждого n пусть

$$S_n \, = \, \frac{1}{n} \, (Y_1 \, + \, \ldots \, + \, Y_n). \tag{3.14}$$

Предположим, что

$$\mathsf{E} \, Y_1^3 \, = \, \mathsf{E} \, X_1^3, \tag{3.15}$$

(например, условие (3.15) выполнено, если распределения с.в. X_1 и Y_1 симметричны) и с.в. Y_1 удовлетворяет условию Крамера (C)

$$\limsup_{|t| \to \infty} \, |\mathsf{E} \, \exp\{itY_1\}| \, < \, 1. \tag{3.16}$$

Теперь из Теоремы 3.1 непосредственно вытекает следующее утверждение.

Лемма 3.3. *Пусть выполнены условия Леммы 3.2 и (3.13) – (3.16). Тогда асимптотический дефект α – квантили статистики S_n относительно α – квантили статистики T_n имеет вид*

$$d \, = \, \frac{\left(\mathsf{E} \, X_1^4 \, - \, \mathsf{E} \, Y_1^4\right) \, \left(3 \, - u_\alpha^2\right)}{12} \, + \, o(1).$$

3. Выборки случайного объема

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

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

то

$$\mathsf{E} \, N_n \, = \, n, \tag{4.1}$$

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 и одинаково распределены.

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

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

Следующее условие описывает а.р. для ф.р. статистики H_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. $G(x)$ и измеримые функции $g_j(x), j = 1, ..., k$ такие, что*

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

$$\sup_x \Big| \mathsf{P}\big(H_n \;<\; x\big) \;-\; G(x) - \sum_{i=1}^{k} \alpha_{in}\; g_i(x) \;\Big| \;\leq\; C_k\; \beta_n \quad n \;\in\; \mathbb{N}.$$

Лемма 4.1. *Пусть статистика $H_n = H_n(X_1, ..., X_n)$ удовлетворяет Условию А. Тогда*

$$\sup_x \big|\; \mathsf{P}\big(H_{N_n} \;<\; x\big) \;-\; G(x) \;-\; \sum_{i=1}^{k} \mathsf{E}\; \alpha_{iN_n}\; g_i(x)\big| \;\leq\; C_k\; \mathsf{E}\; \beta_{N_n}.$$

Доказательство непосредственно следует из формулы полной вероятности.

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

$$\mathsf{E}\; X_1 \;=\; 0, \;\; \mathsf{E}\; X_1^2 \;=\; 1, \; \mathsf{E}\; |X_1|^{k+\delta} \;<\; \infty, \;\; k \;\geq\; 3, \; k \;\in\; \mathbb{N}, \; \delta \;>\; 0. \tag{4.2}$$

Для каждого n пусть

$$H_n \;=\; \frac{1}{\sqrt{n}}\; (X_1 \;+\; ... \;+\; X_n). \tag{4.3}$$

Предположим, что с.в. X_1 удовлетворяет условию Крамера (C)

$$\limsup_{|t| \to \infty}\; |\mathsf{E}\; \exp\{itX_1\}| \;<\; 1. \tag{4.4}$$

При выполнении условий (4.2) и (4.4) из Теоремы 6.3.2 книги [6] следует выполнение неравенства

$$\sup_x \Big| \mathsf{P}\big(H_n \;<\; x\big) \;-\; \Phi(x) - \sum_{i=1}^{k-2} n^{-i/2}\; Q_i(x) \;\Big| \;\leq\; \frac{C_{k,\delta}}{n^{(k-2+\delta)/2}}, \; C_{k,\delta} \;>\; 0, \quad n \;\in\; \mathbb{N}, \tag{4.5}$$

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

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

$$Q_2(x) \;=\; -(x^3 \;-\; 3x)\,\phi(x)\,\frac{\mathsf{E}\,X_1^4 \;-\; 3}{24} \;-\; (x^5 \;-\; 10x^3 \;+\; 15x)\,\phi(x)\,\frac{(\mathsf{E}\,X_1^3)^2}{72}. \quad (4.6)$$

Учитывая неравенство (4.5) и Лемму 4.1, получаем следующее утверждение.

Лемма 4.2. *Пусть выполнены условия (4.2) – (4.4), тогда*

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

Из соотношения (4.5) и Леммы 4.2 непосредственно вытекает следующее утверждение.

Лемма 4.3. *Пусть для* $k = 4,\ \delta > 0$ *выполнены условия (4.2) – (4.4) и*

$$\mathsf{E}\,N_n \;=\; n, \quad \mathsf{E}\,N_n^{-1/2} \;=\; \frac{1}{\sqrt{n}} \;+\; \frac{a}{n} \;+\; o(n^{-1}), \quad a \;\in\; \mathbb{R},$$

$$\mathsf{E}\,N_n^{-1} \;=\; \frac{b}{n} \;+\; o(n^{-1}), \quad \mathsf{E}\,N_n^{-(2+\delta)/2} \;=\; o(n^{-1}), \quad b \;\in\; \mathbb{R},$$

тогда

$$\sup_x \Big|\,\mathsf{P}\big(H_n \;<\; x\big) \;-\; \Phi(x) \;-\; \frac{Q_1(x)}{\sqrt{n}} \;-\; \frac{Q_2(x)}{n}\Big| \;=\; o(n^{-1})$$

и

$$\sup_x \Big|\,\mathsf{P}\big(H_{N_n} \;<\; x\big) \;-\; \Phi(x) \;-\; \frac{Q_1(x)}{\sqrt{n}} \;-\; \frac{bQ_2(x) \;+\; aQ_1(x)}{n}\Big| \;=\; o(n^{-1}).$$

Ниже эти результаты будут применены для аппроксимации α – квантилей и нахождения асимптотического дефекта. Назовем асимптотической α – квантилью ($\alpha \;\in\; (0,1)$) статистики H_n величину $h^*_\alpha(n)$, удовлетворяющую асимптотическому равенству

$$\mathsf{P}\big(H_n \;\ge\; h^*_\alpha(n)\big) \;=\; \alpha \;+\; o(n^{-1}), \quad n \;\to\; \infty, \qquad (4.7)$$

соответственно α – квантилью статистики H_{N_n} назовем величину $h_\alpha(n)$ такую, что

$$\mathsf{P}\big(H_{N_n} \;\ge\; h_\alpha(n)\big) \;=\; \alpha \;+\; o(n^{-1}), \quad n \;\to\; \infty. \qquad (4.8)$$

Из соотношений (4.5), (4.6) и Леммы 3.1 непосредственно следуют а.р. для асимптотических α – квантилей.

Лемма 4.4. *Пусть для* $k = 4,\ \delta > 0$ *выполнены условия (4.2) – (4.4) и условия Леммы 4.3, тогда для асимптотических* α *– квантилей* $h^*_\alpha(n)$ *и* $h_\alpha(n)$ *справедливы а.р.*

$$h^*_\alpha(n) \;=\; u_\alpha \;+\; \frac{\mathsf{E}\,X_1^3}{6\sqrt{n}}\,(u_\alpha^2 \;-\; 1) \;+$$

$$+\;\frac{1}{12n}\left(\frac{\mathsf{E}^2\,X_1^3}{3}\,(5u_\alpha \;-\; 2u_\alpha^3) \;+\; \frac{\mathsf{E}\,X_1^4 \;-\; 3}{2}\,(u_\alpha^3 \;-\; 3u_\alpha)\right) \;+\; o(n^{-1}),$$

$$h_\alpha(n) \;=\; u_\alpha \;+\; \frac{\mathsf{E}\,X_1^3}{6\sqrt{n}}\,(u_\alpha^2 \;-\; 1) \;+$$

$$+\frac{1}{12n}\left(\frac{\mathsf{E}^2\ X_1^3}{3}\,(5u_\alpha-2u_\alpha^3)+\frac{b(\mathsf{E}\ X_1^4\ -\ 3)}{2}\,(u_\alpha^3-3u_\alpha)+2a\,\mathsf{E}\,X_1^3\,(u_\alpha^2-1)\right)+o(n^{-1}),$$

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

Определим теперь последовательность натуральных чисел $\{m(n)\ =\ n\ +\ d^*\ +\ o(1),\ \ d^*\ \in\mathbb{R},\ \ \ n\ =\ 1,2,\ldots\}$ равенством (d^* – асимптотический дефект)

$$\mathsf{P}\Big(H_{N_{m(n)}}\ \geq\ \sqrt{m(n)/n}\ h_\alpha^*(m(n))\Big)\ =\ \alpha\ +\ o(n^{-1}),\quad n\ \to\ \infty,\tag{4.9}$$

то есть это необходимое число $m(n)$ для того, чтобы статистика $H_{N_{m(n)}}$ превзошла нормированную α – квантиль $h_\alpha^*(m(n))$ статистики $H_{m(n)}$.

Аналогично Теореме 3.1 можно доказать следующее утверждение.

Теорема 4.5. *Пусть выполнены следующие условия*

$$\mathsf{E}\ N_n\ =\ n,\quad \mathsf{E}\ N_n^{-1/2}\ =\ \frac{1}{\sqrt{n}}\ +\ \frac{a}{n}\ +\ o(n^{-1}),\quad a\ \in\ \mathbb{R},$$

$$\mathsf{E}\ N_n^{-1}\ =\ \frac{b}{n}\ +\ o(n^{-1}),\quad \mathsf{E}\ N_n^{-(2+\delta)/2}\ =\ o(n^{-1}),\quad b\ \in\ \mathbb{R}$$

и

$$\sup_x\Big|\ \mathsf{P}\big(H_n\ <\ x\big)\ -\ G(x)\ -\ \frac{g_1(x)}{\sqrt{n}}\ -\ \frac{g_2(x)}{n}\Big|\ \leq\ \frac{C}{n^{(2+\delta)/2}},\quad \delta\ >\ 0,$$

тогда асимптотический дефект d^* (см. (4.9)) α – квантилей статистик H_{N_n} и H_n имеет вид

$$d^*\ =\ \frac{2\big(g_2(c_\alpha)\ (1\ -\ b)-\ a\ g_1(c_\alpha)\big)}{G^{(1)}(c_\alpha)\ c_\alpha}\ +\ o(1),$$

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

Из этой Теоремы непосредственно получается следующее утверждение.

Лемма 4.6. *Пусть выполнены условия Леммы 3.4. Тогда асимптотический дефект d^* имеет вид (см. (3.12))*

$$d^*\ =\ \frac{2\big((1\ -\ b)\ Q_2(u_\alpha)\ -\ a\ Q_1(u_\alpha)\big)}{\phi(u_\alpha)\ u_\alpha}\ +\ o(1).$$

Если

$$\mathsf{E}\ X_1^3\ =\ 0,$$

то

$$d^*\ =\ \frac{(1\ -\ b)\ (3\ -\ u_\alpha^2)\ (\mathsf{E}\ X_1^4\ -\ 3)}{12}\ +\ o(1).$$

4. Случай трехточечного симметричного распределения

Применим Лемму 4.2 и Теорему 4.5 для получения асимптотического дефекта и а.р. асимптотических α – квантилей в случае, если объем выборки N_n имеет трехточечное симметричное распределение.

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

$$N_n \ : \ \begin{matrix} n - h_n, & n, & n + h_n \\ \frac{1}{3} & \frac{1}{3}, & \frac{1}{3}, \end{matrix} \tag{5.1}$$

где последовательность натуральных чисел $h_n < n$ удовлетворяет условию

$$\lim_{n\to\infty} \frac{h_n}{n} = 0. \tag{5.2}$$

Лемма 5.1. *Пусть случайная величина N_n имеет распределение (5.1) и выполнено условие (5.2), тогда справедливы равенства*

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

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. □

Применяя Лемму 4.1, непосредственно получаем следующее утверждение.

Лемма 5.2. *Пусть выполнены условия (4.2) – (4.4) с $k = 4$ и $0 < \delta \le 1$ и условия (5.1) и (5.2). Тогда*

$$\sup_x \Big| \mathsf{P}\big(H_{N_n} < x\big) - \Phi(x) - \frac{1}{\sqrt{n}}\Big(1 - \frac{h_n^2}{4n^2}\Big) Q_1(x) -$$

$$- \frac{1}{n}\Big(1 + \frac{2h_n^2}{3n^2}\Big) Q_2(x)\Big| = O\Big(\frac{1}{n^{(2+\delta)/2}} \Big(\frac{h_n}{n}\Big)^{(4+2\delta)/3}\Big).$$

Следствие 5.2. *Пусть выполнены условия Леммы 5.2 и* $h_n = n^{3/4}$. *Тогда равномерно по* $x \in \mathbb{R}$

$$\mathsf{P}\big(H_{N_n} < x\big) = \Phi(x) + \frac{1}{\sqrt{n}}\, Q_1(x) + \frac{1}{n}\Big(Q_2(x) - \frac{1}{4} Q_1(x)\Big) + o(n^{-1}).$$

Теперь из этих Лемм непосредственно вытекает следующая Теорема.

Теорема 5.3. *Пусть выполнены условия Следствия 5.2, тогда для асимптотических* α – *квантилей* $h^*_\alpha(n)$ *и* $h_\alpha(n)$ *статистик* H_n *и* H_{N_n} *справедливы а.р.*

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

$$+ \frac{1}{12n}\Big(\frac{\mathsf{E}^2\, X_1^3}{3}\, (5u_\alpha - 2u_\alpha^3) + \frac{\mathsf{E}\, X_1^4 - 3}{2}\, (u_\alpha^3 - 3u_\alpha)\Big) + o(n^{-1}),$$

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

$$+ \frac{1}{12n}\Big(\frac{\mathsf{E}^2\, X_1^3}{3}\, (5u_\alpha - 2u_\alpha^3) + \frac{\mathsf{E}\, X_1^4 - 3}{2}\, (u_\alpha^3 - 3u_\alpha) - \frac{1}{2}\,\mathsf{E}\, X_1^3\, (u_\alpha^2 - 1)\Big) + o(n^{-1}),$$

где u_α *удовлетворяет уравнению* $\Phi(u_\alpha) = 1 - \alpha$. *Соответствующий асимптотический дефект* d^* *равен*

$$d^* = \frac{Q_1(u_\alpha)}{2\phi(u_\alpha)\, u_\alpha} + o(1) = \frac{(1 - u_\alpha^2)\,\mathsf{E}\, X_1^3}{12\, u_\alpha} + o(1).$$

Заключение

Таким образом в работе исследуется асимптотическое поведение квантилей статистик с помощью понятия дефект, 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 и вычисления асимптотического дефекта. 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/10.1137/S0040585X97981159

[4] Bening V.E., Korolev V.Yu. Generalized Poisson Models and Their Applications in Insurance and Finance. Utrecht: VSP, 2002. 434 p.

[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. 414 с.

[7] Hodges J.L., Lehmann E.L. Deficiency // The Annals of Mathematical Statistics. 1970. Vol. 41, № 5. Pp. 783–801.

[8] Крамер Г. Математические методы статистики. М.: Мир, 1976. 648 с.

[9] Леман Э. Теория точечного оценивания. М.: Наука, ФизМатЛит, 1991. 444 с.

[10] Bening V.E. Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency. Utrecht: VSP, 2000. 277 p.

[11] Бенинг В.Е. О дефекте некоторых оценок, основанных на выборках случайного объёма // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 5–14.

Образец цитирования

Бенинг В.Е. О поведении асимптотического дефекта квантилей распределений статистик, основанных на выборках случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2018. № 3. С. 42–57. https://doi.org/10.26456/vtpmk509

Сведения об авторах

1. **Бенинг Владимир Евгеньевич**
профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени 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 DEFICIENCIES OF THE DISTRIBUTIONS OF STATISTICS BASED ON THE SAMPLES WITH RANDOM SIZES

Bening Vladimir Evgenyevich
Professor at Mathematical Statistics department,
Lomonosov Moscow State University
Senior Researcher, Institute of Informatics Problems
of the Russian Academy of Sciences
Russia, 119992, Moscow, GSP-1, Vorobyovi gory, Lomonosov MSU.
E-mail: bening@yandex.ru

Received 28.06.2018, revised 09.08.2018.

In the paper general theorem concerning the asymptotic deficiency and 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: asymptotic deficiency, quantile, sample of random size, asymptotic expansions, three-point symmetric distribution, transfer theorem.

Citation

Bening V.E., “On asymptotic behavior of quantiles deficiencies 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], 2018, no. 3, 42–57. (in Russian) https://doi.org/10.26456/vtpmk509

References

[1] Gnedenko B.V., “On the estimation of unknown distribution parameters for a random number of independent observations”, *Trudy Tbilisskogo Matematicheskogo instituta [Proceedings of the Tbilisi Mathematical Institute]*, **92** (1989), 146–150 (in Russian).

[2] Gnedenko B.V., Fakhim Kh., “On a transfer theorem”, *Doklady Mathematics*, **187** (1969), 15–17 (in Russian).

[3] Bening V.E., Korolev V.Y., “On an application of the Student distribution in the theory of probability and mathematical statistics”, *Theory of Probability and its Applications*, **49**:3 (2005), 377–391, https://doi.org/10.1137/S0040585X97981159.

[4] Bening V.E., Korolev V.Yu., *Generalized Poisson Models and Their Applications in Insurance and Finance*, VSP, Utrecht, 2002, 434 pp.

[5] Bening V.E., Korolev V.Yu., "Some statistical problems related to the Laplace distribution", *Informatics and Applications*, **2**:2 (2008), 19–34 (in Russian).

[6] Petrov V.V., *Summy nezavisimykh sluchajnykh velichin [Sums of independent random variables]*, Nauka Publ., Moscow, 1972 (in Russian), 414 pp.

[7] Hodges J.L., Lehmann E.L., "Deficiency", *The Annals of Mathematical Statistics*, **41**:5 (1970), 783–801.

[8] Kramer G., *Matematicheskie metody statistiki [Mathematical methods of statistics]*, Mir Publ., Moscow, 1976 (in Russian), 648 pp.

[9] Leman E., *Teoriya Tochechnogo Otsenivaniya [Theory of Point Estimation]*, Nauka Publ., Moscow; FizMatLit Publ., 1991 (in Russian), 444 pp.

[10] Bening V.E., *Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency*, VSP, Utrecht, 2000, 277 pp.

[11] Bening V.E., "On deficiencies of some estimators based on samples of random size", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2015, № 1, 5–14 (in Russian).