УДК 519.2

ОБ АППРОКСИМАЦИИ РАСПРЕДЕЛЕНИЙ СТАТИСТИК, ОСНОВАННЫХ НА ВЫБОРКАХ СЛУЧАЙНОГО ОБЪЁМА

Бенинг В.Е.[*], **Савушкин В.А.**[**]
[*] МГУ им. М.В. Ломоносова, г. Москва
[**] Международный университет Природы, Общества и Человека «Дубна», г. Дубна

Поступила в редакцию 20.01.2014, после переработки 25.01.2014.

В работе доказана общая теорема, позволяющая получать асимптотические разложения для функций распределения произвольно нормированных статистик, основанных на выборках случайного объема, из асимптотических разложений для функций распределения соответственно нормированного случайного объема выборки и асимптотических разложений для функций распределения нормированных статистик, основанных на выборках неслучайного объема.

Ключевые слова: нормированная статистика, функция распределения, выборка случайного объема, асимптотическое разложение, смесь вероятностных законов, распределение Стьюдента, распределение Лапласа.

Вестник ТвГУ. Серия: Прикладная математика. 2014. №1. С. 91–111.

1. Введение

Хорошо известно, что в классических задачах математической статистики количество наблюдений или объем выборки, 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 (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 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 [2], 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, 2, 6–8, 10, 12, 14]). Такого рода суммы находят широкое применение в страховании, экономике, биологии и т.п. (см., [2, 4, 8, 14]). 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 [24], 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.

Более того, в математической статистике и ее приложениях часто встречаются статистики, которые не являются суммами наблюдений. Примерами являются ранговые статистики, U-статистики, линейные комбинации порядковых статистик (L-статистики) и т.п.

В данной работе получены асимптотические разложения (а.р.) для функций распределения (ф.р.) статистик, построенных по выборкам случайного объема. Эти а.р. непосредственно зависят от а.р. ф.р. 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. 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 приведен эвристический вывод основного результата, в разделах 3, 4 и 5 содержатся, соответственно, строгая формулировка основной теоремы, ее доказательство и примеры.

2. Ненормированные статистики

Рассмотрим случайные величины (с.в.) $N_1, N_2, \ldots$ и $X_1, X_2, \ldots$, заданные на одном и томже вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$. В статистике с.в. $X_1, X_2, \ldots X_n$ 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

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

то

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

то есть среднее значение случайного объема выборки равно n. Предположим, что при каждом $n \geq 1$ с.в. N_n принимают только натуральные значения, то есть $N_n \in \mathbb{N}$ и независят от последовательности с.в. $X_1, X_2, \ldots$. Всюду далее считаем с.в. $X_1, X_2, \ldots$ независимыми одинаково распределенными и имеющими ф.р. $F(x)$. Обозначим через $T_n = T_n(X_1, \ldots, X_n)$ некоторую статистику, то есть действительную измеримую функцию от наблюдений $X_1, \ldots, X_n$. Рассмотрим случай большого числа наблюдений, то есть пусть $n \to \infty$. Предположим, что функция распределения ненормированной статистики T_n слабо сходится к некоторой ф.р. $G(x)$, то есть в каждой точке нерерывности ф.р. $G(x)$ имеет место сходимость

$$\mathsf{P}\big(T_n \ < \ x\big) \ \longrightarrow \ G(x), \quad n \to \infty. \tag{2.1}$$

Пусть случайный объем выборки N_n стремится к бесконечности по вероятности при $n \ \to \ \infty$, то есть для любого числа $M \ > \ 0$

$$\mathsf{P}\big(N_n \ > \ M\big) \ \longrightarrow \ 1, \quad n \to \infty. \tag{2.2}$$

Рассмотрим предельное поведение ф.р. статистики, основанной на выборке случайного объема, то есть статистики

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

Как показывает приведенная ниже лемма, никакого нового предельного закона при условиях (2.1) и (2.2) не возникает.

Лемма 2.1. *Пусть выполнены условия (2.1) и (2.2), тогда в каждой точке непрерывности ф.р. $G(x)$ справедливо соотношение*

$$\mathsf{P}\big(T_{N_n} \ < \ x\big) \ \longrightarrow \ G(x), \quad n \to \infty.$$

Доказательство. Пусть точка x является точкой непрерывности ф.р. $G(x)$, тогда по формуле полной вероятности для любого натурального числа $M \in \mathbb{N}$ имеем

$$\Big|\mathsf{P}\big(T_{N_n} \ < \ x\big) \ - \ G(x)\Big| \ =$$

$$= \ \Big|\sum_{k=1}^{\infty} \ \mathsf{P}\big(T_k \ < \ x\big) \ \mathsf{P}\big(N_n \ = \ k\big) \ - \ G(x)\Big| \ \le$$

$$\le \sum_{k=1}^{\infty} \Big|\mathsf{P}\big(T_k \ < \ x\big) \ - \ G(x)\Big| \ \mathsf{P}\big(N_n \ = \ k\big) \ \le$$

$$\le \ \sum_{k=1}^{M} \ \mathsf{P}\big(N_n \ = \ k\big) \ + \ \sum_{k=M+1}^{\infty} \ \Big|\mathsf{P}\big(T_k \ < \ x\big) \ - \ G(x)\Big| \ \mathsf{P}\big(N_n \ = \ k\big) \ =$$

$$= \ \mathsf{P}\big(N_n \ \le \ M\big) \ + \ \sum_{k=M+1}^{\infty} \ \Big|\mathsf{P}\big(T_k \ < \ x\big) \ - \ G(x)\Big| \ \mathsf{P}\big(N_n \ = \ k\big) \ =$$

$$= \ 1 \ - \ \mathsf{P}(N_n \ > \ M) \ + \ \sum_{k=M+1}^{\infty} \ \Big|\mathsf{P}\big(T_k \ < \ x\big) \ - \ G(x)\Big| \ \mathsf{P}\big(N_n \ = \ k\big). \qquad (2.3)$$

Из Условия (2.1) следует, что для любого $\varepsilon > 0$ существует натуральное число M_ε, такое, что при всех натуральных $k \ > \ M_\varepsilon$ справедливо неравенство

$$\Big|\mathsf{P}\big(T_k \ < \ x\big) \ - \ G(x)\Big| \ < \ \varepsilon.$$

Полагая в неравенстве (2.3) $M \ = \ M_\varepsilon$, получаем оценку

$$\Big|\mathsf{P}\big(T_{N_n} \ < \ x\big) \ - \ G(x)\Big| \ \le \big(1 \ - \ \mathsf{P}\big(N_n \ > \ M_\varepsilon\big)\big) + \varepsilon.$$

Первое слагаемое в правой части этого неравенство может быть сделано сколь угодо малым в силу условия (2.2), поэтому в силу произвольности ε лемма доказана. □

Предположим теперь, что для функции распределения ненормированной статистики T_n справедливо асимптотическое разложение, описываемое следующим условием.

Условие 1. *Существуют константы $l \in \mathbb{N}$, $\alpha \ > \ l/2$, $C_1 \ > \ 0$, дифференцируемая ф.р. $G(x)$ и дифференцируемые ограниченные функции $g_i(x), i \ = \ 1, \ldots, l$ такие, что*

$$\sup_x \Big|\mathsf{P}\big(T_n \ < \ x\big) \ - \ G(x) \ - \sum_{i=1}^{l} n^{-i/2} \ g_i(x)\Big| \ \le \ \frac{C_1}{n^\alpha}, \quad n \in \mathbb{N}.$$

Пусть также для функции распределения нормированного случайного индекса N_n справедливо асимптотическое разложение, которое задается следующим условием.

Условие 2. *Существуют константы* $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, *функции* $0 < v(n) \uparrow \infty$, $n \to \infty$, $u(n) \in \mathbb{R}$, *функция распределения* $H(x), H(0+) = 0$ *и функции ограниченной вариации* $h_j(x)$, $j = 1, \ldots, m$ *такие, что*

$$\sup_{x \geqslant 0} \left| \mathsf{P}\Big(\frac{N_n}{v(n)} - u(n) < x \Big) - H(x) - \sum_{j=1}^{m} n^{-j/2}\, h_j(x) \right| \leq \frac{C_2}{n^{\beta}}, \quad n \in \mathbb{N}.$$

Докажем теорему, уточняющую Лемму 2.1, с использованием асимптотического разложения для функции распределения ненормированной статистики, построенной по выборке случайного объема.

Теорема 2.1. *Пусть выполнены Условия 1 и 2. Тогда справедливо неравенство*

$$\sup_x \left| \mathsf{P}\big(T_{N_n} < x\big) - G_n(x) \right| \leq C_1 \,\mathsf{E}\, N_n^{-\alpha} + 2\, \frac{C_2}{n^{\beta}} \sup_x \sum_{i=1}^{l} |g_i(x)|,$$

где

$$G_n(x) = G(x) +$$

$$+ \sum_{i=1}^{l} \big(v(n)\big)^{-i/2} g_i(x) \int\limits_{1/v(n)}^{\infty} y^{-i/2}\, d\Big(H(y - u(n)) + \sum_{j=1}^{m} n^{-j/2}\, h_j(y - u(n)) \Big) =$$

$$= G(x) + \sum_{i=1}^{l} g_i(x) \int\limits_{1}^{\infty} z^{-i/2}\, d\Big(H(z/v(n) - u(n)) + \sum_{j=1}^{m} n^{-j/2}\, h_j(z/v(n) - u(n)) \Big).$$

Доказательство. Оценим рассматриваемую разность сверху следующим образом

$$\sup_x \left| \mathsf{P}\big(T_{N_n} < x\big) - G_n(x) \right| \leq I_{1n} + I_{2n}, \tag{2.4}$$

где

$$I_{1n} \equiv \sup_x \left| \mathsf{P}\big(T_{N_n} < x\big) - \mathsf{E}\,\Big(G(x) + \sum_{i=1}^{l} N_n^{-i/2}\, g_i(x) \Big) \right|,$$

$$I_{2n} \equiv \sup_x \left| \mathsf{E}\Big(G(x) + \sum_{i=1}^{l} N_n^{-i/2}\, g_i(x) \Big) - G_n(x) \right|.$$

Оценим выражение I_{1n} с помощью Условия 1 и формулы полной вероятности. Имеем

$$I_{1n} = \sup_x \left| \sum_{k=1}^{\infty} \mathsf{P}\big(N_n = k\big) \Big(\mathsf{P}\big(T_k < x\big) - G(x) - \sum_{i=1}^{l} k^{-i/2}\, g_i(x) \Big) \right| \leq$$

$$\leq \sum_{k=1}^{\infty} \mathsf{P}\big(N_n = k\big) \sup_x \left| \mathsf{P}\big(T_k < x\big) - G(x) - \sum_{i=1}^{l} k^{-i/2}\, g_i(x) \right| \leq$$

$$\leq \ C_1 \ \sum_{k=1}^{\infty} \ \frac{1}{k^{\alpha}} \ \mathsf{P}\big(N_n \ = \ k\big) \ = \ C_1 \ \mathsf{E} \ N_n^{-\alpha}. \tag{2.5}$$

Для оценки величины I_{2n} используем Условие 2 и формулу интегрирования по частям. Имеем

$$I_{2n} \ = \ \sup_x \Big|\mathsf{E}\Big(G(x) \ + \ \sum_{i=1}^{l} \ N_n^{-i/2} \ g_i(x)\Big) \ - \ G(x) \ -$$

$$- \ \sum_{i=1}^{l} \ \big(v(n)\big)^{-i/2} \ g_i(x) \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} d\Big(H(y \ - \ u(n)) \ + \ \sum_{j=1}^{m} \ n^{-j/2} h_j(y \ - \ u(n))\Big)\Big| \ =$$

$$= \sup_x \Big|\sum_{i=1}^{l} \mathsf{E} \ N_n^{-i/2} g_i(x) \ -$$

$$- \ \sum_{i=1}^{l} \ \big(v(n)\big)^{-i/2} \ g_i(x) \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} \ d \ \Big(H(y \ - \ u(n)) \ + \ \sum_{j=1}^{m} \ n^{-j/2} \ h_j(y \ - \ u(n))\Big)\Big| \ =$$

$$= \ \sup_x \Big|\sum_{i=1}^{l} \ g_i(x) \ \int\limits_{1}^{\infty} \ y^{-i/2} \ d \ \mathsf{P}\big(N_n \ < \ y\big) \ -$$

$$- \ \sum_{i=1}^{l} \ g_i(x) \ \big(v(n)\big)^{-i/2} \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} \ d \ \Big(H(y \ - \ u(n)) \ - \ \sum_{j=1}^{m} \ n^{-j/2} \ h_j(y \ - \ u(n))\Big)\Big|.$$

После замены переменной в первом интеграле, получим

$$I_{2n} \ = \ \sup_x \Big|\sum_{i=1}^{l} \ g_i(x) \ \big(v(n)\big)^{-i/2} \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} \ d \ \Big(\mathsf{P}\Big(\frac{N_n}{v(n)} \ < y \ \Big) \ - \ H(y \ - \ u(n)) \ -$$

$$- \ \sum_{j=1}^{m} \ n^{-j/2} \ h_j(y \ - \ u(n))\Big)\Big|.$$

Используя формулу интегрирования по частям (см., например, [5], теорема 2.6.11, стр. 222 или [15], теорема 18.4, стр. 236), ограниченность функций $g_i(x), i = 1, ..., l$ и Условие 2, получим, что справедливы неравенства

$$I_{2n} \ \leqslant \ \frac{C_2}{n^{\beta}} \ \sup_x \ \sum_{i=1}^{l} \ \big|g_i(x)\big| \ +$$

$$+ \ \sup_x \Big|\sum_{i=1}^{l} \ g_i(x)(v(n))^{-i/2} \ \int\limits_{1/v(n)}^{\infty} \ \Big(\mathsf{P}\Big(\frac{N_n}{v(n)} \ - \ u(n) \ < \ y \ - \ u(n)\Big) \ -$$

$$- \ H(y \ - \ u(n)) \ - \ \sum_{j=1}^{m} n^{-j/2} h_j(y \ - \ u(n))\Big) \ d \ y^{-i/2}\Big)\Big| \leqslant$$

$$\leqslant\ 2\,\frac{C_2}{n^{\beta}}\ \sup_x\ \sum_{i=1}^{l}\big|g_i(x)\big|. \tag{2.6}$$

Теперь утверждение теоремы следует из неравенств (2.4), (2.5) и (2.6). Теорема доказана. □

3. Примеры

Приведем два примера применения Теоремы 2.1 для статистик, построенных по конкретным выборкам случайного объема. Мы рассмотрим асимптотические разложения для ф.р. выборочного среднего, построенного по выборкам случайного объема. Аналогичные результаты могут быть получены для статистик, допускающих асимптотические разложения типа Эджворта для ф.р. при неслучайном объеме выборки. Например, используя результаты работ [16–21], можно получить, соответственно, асимптотические разложения для ф.р. R-статистик, L-статистик и U-статистик.

Пусть $X_1, X_2, \ldots$ – независимые одинаково распределенные случайные величины с $\mathsf{E}X_1 = \mu$, $0 < \mathsf{D}X_1 = \sigma^2$, $\mathsf{E}|X_1|^{3+2\delta} < \infty$, $\delta \in (0,\frac{1}{2})$ и $\mathsf{E}\big(X_1-\mu\big)^3 = \mu_3$. Для натурального n обозначим

$$T_n\ =\ \frac{X_1+\ldots\ +X_n\ -\ n\mu}{\sigma\sqrt{n}}. \tag{3.1}$$

Предположим также, что случайная величина X_1 удовлетворяет условию Крамéра (C)

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

тогда с учетом теоремы 6.3.2 из книги [22] (стр. 207), получаем, что

$$\sup_x\Big|\mathsf{P}\big(T_n\ <\ x\big)\ -\ \Phi(x)\ -\ \frac{\mu_3}{6\sqrt{n}\sigma^3}\,(1-x^2)\ \varphi(x)\Big|\ \leq$$

$$\leq\ \frac{C_1}{n^{1/2+\delta}},\ \ C_1\ >\ 0,\ \ \delta\in(0,\tfrac{1}{2}),\ \ n\ \in\ \mathbb{N}. \tag{3.2}$$

Таким образом, статистика (3.1) удовлетворяет Условию 1 Теоремы 2.1 с

$$\alpha\ =\ \frac{1}{2}\ +\ \delta,\quad l\ =\ 1,\quad G(x)\ =\ \Phi(x),\quad g_1(x)\ =\ \frac{\mu_3}{6\sigma^3}\,(1\ -\ x^2)\varphi(x). \tag{3.3}$$

При этом не трудно видеть, что

$$\sup_x\ |g_1(x)|\ <\ \infty.$$

3.1 Отрицательно биномиальный индекс

Пусть случайный объем выборки N_n (случайный индекс) имеет отрицательно биномиальное распределение с параметрами $p = 1/n$ и $r > 0$, то есть

$$\mathsf{P}\big(N_n\ =\ k\big)\ =\ \frac{(k+r-2)\cdots r}{(k-1)!}\ \frac{1}{n^r}\ \Big(1\ -\ \frac{1}{n}\Big)^{k-1},\quad k\ \in\ \mathbb{N}.$$

При $r = 1$ имеем геометрическое распределение. В книге [23] (формула (6.112), стр. 233) приведена следующая оценка скорости сходимости для нормированного случайного индекса

$$\sup_{x\geqslant 0}\left|\mathsf{P}\Big(\frac{N_n}{\mathsf{E}\,N_n} < x\Big) - H_r(x)\right| \leq \begin{cases}\dfrac{C_r}{n}, & r \geqslant 1,\\ \dfrac{C_r}{n^r}, & r \in (0,1),\end{cases} \tag{3.4}$$

где $C_r > 0, \;\; n \in \mathbb{N}$ и

$$H_r(x) = \frac{r^r}{\Gamma(r)}\int_0^x e^{-ry}y^{r-1}\,dy, \;\; x \geqslant 0 \tag{3.5}$$

есть функция гамма-распределения с параметром $r > 0$. При этом

$$\mathsf{E}N_n = r(n-1)+1. \tag{3.6}$$

Таким образом, из соотношений (3.4)–(3.6) следует, что случайный индекс N_n удовлетворяет Условию 2 с

$$v(n) = r(n-1)+1, \;\; H(x) = H_r(x), \;\; m = 1,$$

$$h_1(x) \equiv 0, \;\; C_2 = C_r > 0, \; u(n) = 0, \;\; n \in \mathbb{N},$$

$$\beta = \begin{cases}1, & r \geqslant 1,\\ r, & r \in (1/2,1).\end{cases}$$

Далее, используя равенство

$$\big(1+x\big)^{\gamma} = \sum_{k=0}^{\infty}\frac{\gamma(\gamma-1)\cdots(\gamma-k+1)}{k!}\,x^k, \quad |x| < 1, \;\; \gamma \in \mathbb{R},$$

нетрудно получить, что

$$\mathsf{E}N_n^{-1} = \frac{1}{(n-1)\,(1-r)}\Big(\frac{1}{n^{r-1}}-1\Big) = O(n^{-r}), \;\; r > 0, \; r\neq 1, \;\; n \in \mathbb{N}. \tag{3.7}$$

Для случая $r = 1$, имеем

$$\mathsf{E}N_n^{-1} = \frac{1}{n-1}\,\log n, \quad n > 1.$$

Теперь, используя неравенство Гельдера, получим

$$\mathsf{E}\,N_n^{-\alpha} \leqslant \big(\mathsf{E}\,N_n^{-1}\big)^{\alpha}, \;\; \alpha \leqslant 1,$$

$$\mathsf{E}\,N_n^{-\alpha} = O(n^{-r(1/2+\delta)}), \quad r > 0, \;\; r\neq 1,$$

$$\mathsf{E}\,N_n^{-\alpha} = O\Big(\Big(\frac{\log n}{n}\Big)^{1/2+\delta}\Big), \;\; \alpha = \frac{1}{2}+\delta, \; r = 1, \;\; n \in \mathbb{N}. \tag{3.8}$$

Таким образом, учитывая Теорему 2.1, формулы (3.2), (3.3), а также соотношения (3.4)–(3.8) и равенство

$$\int\limits_{(r(n-1)+1)^{-1}}^{\infty} \sqrt{y}\, dH_r(y) \;=\; \int\limits_0^{\infty} \sqrt{y}\, dH_r(y) \;+\; O\big(1/n\big) =$$

$$= \int\limits_0^{\infty} \sqrt{y}\, \frac{r^r}{\Gamma(r)}\, e^{-ry}\, y^{r-1}\, dy \;+\; O\big(1/n\big) \;=$$

$$= \frac{r^r}{\Gamma(r)} \int\limits_0^{\infty} e^{-y}\, \frac{y^{r-1/2}}{r^{r+1/2}}\, dy \;+\; O\big(1/n\big) \;=\; \frac{\Gamma(r+1/2)}{\Gamma(r)\,\sqrt{r}} \;+\; O\big(1/n\big),$$

получаем следующее утверждение

Теорема 3.1 *Пусть статистика* T_n *имеет вид* (3.1), *причем* $X_1, X_2, \ldots$ *– независимые одинаково распределенные случайные величины с* $\mathsf{E}X_1 = \mu$, $0 < \mathsf{D}X_1 = \sigma^2$, $\mathsf{E}|X_1|^{3+2\delta} < \infty$, $\delta \in (0, \frac{1}{2})$ *и* $\mathsf{E}\big(X_1 - \mu\big)^3 = \mu_3$. *Предположим также, что случайная величина* X_1 *удовлетворяет условию Крамéра* (C)

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

Пусть при некотором $r > 0$ *случайная величина* N_n *имеет отрицательно биномиальное распределение вида*

$$\mathsf{P}\big(N_n \;=\; k\big) \;=\; \frac{(k+r-2)\cdots r}{(k-1)!}\;\frac{1}{n^r}\;\Big(1 \;-\; \frac{1}{n}\Big)^{k-1}, \quad k \;\in\; \mathbb{N}.$$

Тогда при $r > 1/(1+2\delta)$ *для функции рапределения статистики* T_{N_n} *при* $n \to \infty$ *справедливо асимптотическое разложение вида*

$$\sup_x \Big|\mathsf{P}\Big(T_{N_n} \;<\; x\Big) \;-\; \Phi(x) \;-\; \frac{\mu_3\;\Gamma(r+1/2)}{6\;\sigma^3\;\Gamma(r)\sqrt{r^2(n-1)+r}}\;(1 \;-\; x^2)\;\varphi(x)\Big| \;=$$

$$= \begin{cases} O\Big(\Big(\dfrac{\log n}{n}\Big)^{1/2+\delta}\Big), & r \;=\; 1, \\ O\Big(\dfrac{1}{n^{\min(1, r(1/2+\delta))}}\Big), & r \;>\; 1, \\ O\Big(\dfrac{1}{n^{r(1/2+\delta)}}\Big), & \dfrac{1}{1+2\delta} \;<\; r \;<\; 1. \end{cases}$$

3.2 Дискретное распределение Парето

В работе [9] была построена последовательность с.в. $N_n(s)$, зависящая от натурального параметра $s \in \mathbb{N}$ следующего вида. Пусть $Y_1, Y_2, \ldots$ – независимые одинаково распределенные случайные величины, имеющие непрерывную функцию распределения. Определим случайные величины

$$N(s) \;=\; \min\,\{i \;\geq\; 1 \;:\; \max_{1\leq j\leq s}\; Y_j \;<\; \max_{s+1\leq k\leq s+i}\; Y_k\}.$$

Хорошо известно, что так определенные случайные величины имеют дискретное распределение Парето вида

$$\mathsf{P}\big(N(s) \ \geq \ k\big) \ = \ \frac{s}{s+k-1}, \quad k \geq \ 1 \tag{3.9}$$

(см., например, [26] и [27]). Пусть теперь $N^{(1)}(s), N^{(2)}(s), ...$ – независимые одинаково распределенные случайные величины, имеющие распределение (3.9). Определим случайные величины

$$N_n(s) \ = \ \max_{1\leq \ j\leq \ n} \ N^{(j)}(s).$$

Тогда, как показано в работе [9],

$$\lim_{n\to\infty} \ \mathsf{P}\Big(\frac{N_n(s)}{n} \ < \ x\Big) \ = \ e^{-s/x}, \quad x \ > \ 0. \tag{3.10}$$

В работе [11] была получена следующая оценка скорости сходимости в соотношении (3.10)

$$\sup_{x\geqslant 0}\Big|\mathsf{P}\Big(\frac{N_n(s)}{n} \ < \ x\Big) \ - \ e^{-s/x}\Big| \ \leq \ \frac{C_s}{n}, \quad C_s \ > \ 0, \ \ n \ \in \ \mathbb{N}. \tag{3.11}$$

Таким образом, из соотношения (3.11) следует, что случайный индекс $N_n(s)$ удовлетворяет Условию 2 Теоремы 2.1 с

$$v(n) \ = \ n, \ \ H(x) \ = \ e^{-s/x}, \ \ m \ = \ 1, \ \ \ h_1(x) \ \equiv \ 0, \ \ C_2 \ = \ C_s \ > \ 0,$$

$$u(n) \ = \ 0, \ n \ \in \ \mathbb{N}, \ \ \beta \ = \ 1. \tag{3.12}$$

Рассмотрим более подробно величину $\mathsf{E}N_n^{-1}(s)$. Из определения случайной величины $N_n(s)$ и равенства (3.9) имеем

$$\mathsf{P}\big(N_n(s) \ = \ k\big) \ = \ \Big(\frac{k}{s+k}\Big)^n \ - \ \Big(\frac{k-1}{s+k-1}\Big)^n \ = \ sn \int\limits_{k-1}^{k} \frac{x^{n-1}}{(s \ + \ x)^{n+1}} \ d\,x,$$

поэтому

$$\mathsf{E}\ N_n^{-1}(s) \ = \ \sum_{k=1}^{\infty} \frac{1}{k}\,\mathsf{P}\big(N_n(s) \ = \ k\big) \ = \ sn\sum_{k=1}^{\infty}\frac{1}{k}\int\limits_{k-1}^{k} \frac{x^{n-1}}{(s \ + \ x)^{n+1}} \ d\,x \ \leq$$

$$\leq \ sn\sum_{k=1}^{\infty}\int\limits_{k-1}^{k} \frac{x^{n-2}}{(s \ + \ x)^{n+1}} \ d\,x \ = \ sn\int\limits_{0}^{\infty} \frac{x^{n-2}}{(s \ + \ x)^{n+1}} \ d\,x.$$

Для вычисления последнего интеграла используем формулу (см. [13] формула 856.12, стр. 184)

$$\int\limits_{0}^{\infty} \frac{x^{s-1}}{(a \ + \ bx)^{s+n}} \ d\,x \ = \ \frac{\Gamma(s)\ \Gamma(n)}{a^n\ b^s\ \Gamma(s+n)}, \quad a, \ b, \ s, \ n \ > \ 0.$$

Получим

$$\mathsf{E}\, N_n^{-1}(s) \;\le\; sn\,\frac{\Gamma(n-1)\,\Gamma(2)}{s^2\,\Gamma(n+1)} \;=\; \frac{1}{s(n\;-\;1)} \;=\; O(n^{-1}). \tag{3.13}$$

Теперь, используя неравенство Гельдера, получим

$$\mathsf{E}\, N_n^{-\alpha} \;\leqslant\; \big(\mathsf{E}\, N_n^{-1}\big)^{\alpha}, \quad \alpha \;\leqslant\; 1,$$

$$\mathsf{E}\, N_n^{-\alpha} \;=\; O(n^{-(1/2+\delta)}), \quad \alpha \;=\; 1/2 \;+\; \delta. \tag{3.14}$$

Таким образом, учитывая Теорему 2.1, формулы (3.11)–(3.14) и равенство

$$\int\limits_{n^{-1}}^{\infty} \sqrt{y}\; d\, e^{-s/y} \;=\; \sqrt{s}\int\limits_{0}^{\infty} y^{-1/2}\, e^{-y}\; d\,y \;+\; O\Big(\frac{1}{n}\Big) \;=$$

$$=\; \sqrt{s}\;\Gamma(1/2) \;+\; O\Big(\frac{1}{n}\Big) \;=\; \sqrt{s\;\pi} \;+\; O\Big(\frac{1}{n}\Big),$$

непосредственно получаем следующую теорему.

Теорема 3.2 *Пусть статистика T_n имеет вид* (3.1), *причем $X_1, X_2, \ldots$ – независимые одинаково распределенные случайные величины с $\mathsf{E}X_1 \;=\; \mu$, $0 \;<\; \mathsf{D}X_1 \;=\; \sigma^2$, $\mathsf{E}|X_1|^{3+2\delta} \;<\; \infty$, $\delta \;\in\; (0, \frac{1}{2})$ и $\mathsf{E}\big(X_1 \;-\; \mu\big)^3 \;=\; \mu_3$. Предположим также, что случайная величина X_1 удовлетворяет условию Крамéра* (C)

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

Пусть также при некотором $s \in \mathbb{N}$ случайная величина $N_n(s)$ имеет дискретное распределение Парето вида

$$\mathsf{P}\big(N_n(s) \;=\; k\big) \;=\; \Big(\frac{k}{s+k}\Big)^n \;-\; \Big(\frac{k-1}{s+k-1}\Big)^n, \quad k \;\in\; \mathbb{N}.$$

Тогда для функции распределения статистики $T_{N_n(s)}$ справедливо асимптотическое разложение вида

$$\sup_x \Big|\mathsf{P}\big(T_{N_n(s)} \;<\; x\big) \;-\; \Phi(x) \;-\; \frac{\mu_3\,\sqrt{s\;\pi}}{6\sigma^3\,\sqrt{n}}\;(1\;-\;x^2)\varphi(x)\Big| \;=\; O\Big(\frac{1}{n^{1/2+\delta}}\Big), \quad n \;\to\; \infty.$$

4. Нормированные статистики

Предположим теперь, что функция распределения нормированной статистики T_n стремится к некоторой функции распределения $G(x)$, то есть существуют постоянные $\sigma \;>\; 0$, $\mu \;\in\; \mathbb{R}$, $\gamma \;>\; 0$ такие, что в каждой точке непрерывности функции распределения $G(x)$ справедливо соотношение

$$\mathsf{P}\big(\sigma n^{\gamma}\;(T_n \;-\; \mu) \;<\; x\big) \;\longrightarrow\; G(x), \quad n \to \infty. \tag{4.1}$$

Здесь мы рассматриваем степенные нормирующие множители типа n^γ. Из доказанной выше Леммы 2.1 следует, что при условии (2.2) в каждой точке непрерывности функции распределения $G(x)$ справедлива сходимость

$$\mathsf{P}\big(\sigma {N_n}^\gamma\ (T_{N_n}\ -\ \mu)\ <\ x\big)\ \longrightarrow\ G(x),\quad n\to\infty. \tag{4.2}$$

В соотношении (4.2) мы имеем случайный нормирующий множитель ${N_n}^\gamma$, что неудобно на практике, поскольку соотношение (4.1) означает, что при больших n

$$\sigma n^\gamma\ (T_n\ -\ \mu)\ \approx\ \xi$$

или асимптотически статистика T_n ведет себя примерно как

$$T_n\ \approx\ \frac{1}{\sigma n^\gamma}\,\xi\ +\ \mu,$$

где случайная величина ξ имеет функцию распределения $G(x)$. В тоже время соотношение (4.2) означает лишь, что при больших n

$$\sigma N_n^\gamma\ (T_{N_n}\ -\ \mu)\ \approx\ \xi$$

и аппроксимация самой статистики T_{N_n} затруднена. Это делает естественной постановку, в которой при условии (4.1) ищутся предельные законы функции распределения случайной величины $\sigma n^\gamma\ (T_{N_n}\ -\ \mu)$, то есть нормированной статистики T_{N_n}, но неслучайным нормирующим множителем n^γ. Эвристически ясно, что если нормированный индекс $\frac{N_n}{n}$ имеет слабый предел, то в силу тождества

$$\sigma n^\gamma\ (T_{N_n}\ -\ \mu)\ =\ \sigma N_n^\gamma\ (T_{N_n}\ -\ \mu)\,\frac{n}{N_n}$$

и доказанной Леммы 2.1 случайная величина $\sigma N_n^\gamma\ (T_{N_n}\ -\ \mu)$ также имеет слабый предел, уже отличный от функции распределения $G(x)$, в виде масштабной смеси предельных законов. Точнее, если предположить, что

$$\mathsf{P}\Big(\frac{N_n}{n}\ <\ x\Big)\ \longrightarrow\ H(x),\quad n\to\infty, \tag{4.3}$$

то при выполнении условия (4.1) справедливы приближенные равенства, описывающие предельный закон этой случайной величины

$$\mathsf{P}\big(\sigma n^\gamma\ (T_{N_n}\ -\ \mu)\ <\ x\big)\ =\ \sum_{k=1}^{\infty}\ \mathsf{P}\big(\sigma n^\gamma\ (T_k\ -\ \mu)\ <\ x\big)\ \mathsf{P}\big(N_n\ =\ k\big)\ =$$

$$=\ \sum_{k=1}^{\infty}\ \mathsf{P}\big(\sigma k^\gamma\ (T_k\ -\ \mu)\ <\ \big(k/n\big)^\gamma\ x\big)\ \mathsf{P}\big(N_n\ =\ k\big)\ \approx$$

$$\approx\ \sum_{k=1}^{\infty}\ G\big(\big(k/n\big)^\gamma\ x\big)\ \mathsf{P}\big(N_n\ =\ k\big)\ =\ \mathsf{E}\,G\big(\big(N_n/n\big)^\gamma\ x\big)\ =$$

$$=\ \int_0^\infty\ G\big(xy^\gamma\big)\ d\mathsf{P}\big(N_n/n\ <\ y\big)\ \approx$$

$$\approx \int_0^\infty G(xy^\gamma)\, dH(y). \tag{4.4}$$

Более того, если предположить, что существуют функции (см. Условие 2) $v(n) > 0$, $u(n) \in \mathbb{R}$, $n \in \mathbb{N}$ такие, что

$$\mathsf{P}\Big(\frac{N_n}{v(n)} - u(n) < x\Big) \longrightarrow H(x), \quad n \to \infty,$$

и функция $w(n) > 0$, $n \in \mathbb{N}$ такая, что при некоторых числах $\sigma > 0$, $\mu \in \mathbb{R}$ справедливо соотношение

$$\mathsf{P}\Big(\sigma\, w(n)(T_{N_n} - \mu) < x\Big) \longrightarrow G(x), \quad n \to \infty,$$

то справедлива аппроксимация

$$\mathsf{P}\big(\sigma w(n)\,(T_{N_n} - \mu) < x\big) = \sum_{k=1}^\infty \mathsf{P}\big(\sigma w(n)\,(T_k - \mu) < x\big)\, \mathsf{P}\big(N_n = k\big) =$$

$$= \sum_{k=1}^\infty \mathsf{P}\big(\sigma w(k)\,(T_k - \mu) < \big(w(k)/w(n)\big)\, x\big)\, \mathsf{P}\big(N_n = k\big) \approx$$

$$\approx \sum_{k=1}^\infty G\big(\big(w(k)/w(n)\big)\, x\big)\, \mathsf{P}\big(N_n = k\big) = \mathsf{E}\, G\big(\big(w(N_n)/w(n)\big)\, x\big) =$$

$$= \int_0^\infty G\Big(x\, \frac{w(yv(n))}{w(n)}\Big)\, d\mathsf{P}\big(N_n/v(n) < y\big) \approx$$

$$\approx \int_0^\infty G\Big(x\, \frac{w(yv(n))}{w(n)}\Big)\, dH\big(y - u(n)\big).$$

Теперь сформулируем и докажем теорему об асимптотических разложениях, непосредственно обобщающую Теорему 2.1 на случай произвольно нормированных статистик, основанных на выборках случайного объема и обосновывающую эти аппроксимации. Сформулируем сначала непосредсвенное обобщение Условия 1.

Условие 3. *Существуют константы $l \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > l/2$, $C_3 > 0$, функция $w(n) > 0$, $n \in \mathbb{N}$, дифференцируемая функция распределения $G(x)$ и дифференцируемые ограниченные функции $g_i(x)$, $i = 1, \ldots, l$ такие, что*

$$\sup_x \Big|\mathsf{P}\big(\sigma\, w(n)(T_n - \mu) < x\big) - G(x) - \sum_{i=1}^{l} n^{-i/2} g_i(x)\Big| \leq \frac{C_3}{n^\alpha}, \quad n \in \mathbb{N}.$$

Теорема 4.1 *Пусть выполнены Условия 2 и 3. Тогда справедливо неравенство*

$$\sup_x \big|\mathsf{P}\big(\sigma\, w(n)(T_{N_n} - \mu) < x\big) - G_n(x)\big| \leq$$

$$\leq C_3\, \mathsf{E}\, N_n^{-\alpha} + \frac{C_2(1 + M_n)}{n^\beta}\, \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big),$$

где

$$M_n \ = \ \sup_x \int\limits_{1/v(n)}^{\infty} \ \Big|\frac{\partial}{\partial y}\Big(G\Big(x\frac{w(yv(n))}{w(n)}\Big) \ + \ \sum_{i=1}^{l} \ (yv(n))^{-i/2} \ g_i\Big(x\frac{w(yv(n))}{w(n)}\Big)\Big)\Big| \ dy$$

и асимптотическое разложение $G_n(x)$ *имеет следующий вид*

$$G_n(x) \ = \ \int\limits_{1/v(n)}^{\infty} \ G\Big(x\frac{w(yv(n))}{w(n)}\Big) \ dH(y \ - \ u(n)) \ +$$

$$+\sum_{j=1}^{m} \ n^{-j/2} \ \int\limits_{1/v(n)}^{\infty} \ G\Big(x\frac{w(yv(n))}{w(n)}\Big) \ dh_j(y \ - \ u(n)) \ +$$

$$+\sum_{i=1}^{l} v^{-i/2}(n) \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} g_i\Big(x\frac{w(yv(n))}{w(n)}\Big) \ dH(y \ - \ u(n)) \ +$$

$$+ \ \sum_{i=1}^{l} \ \sum_{j=1}^{m} \ n^{-j/2} v^{-i/2}(n) \ \int\limits_{1/v(n)}^{\infty} \ y^{-i/2} g_i\Big(x\frac{w(yv(n))}{w(n)}\Big) \ dh_j(y \ - \ u(n)).$$

Доказательство. Сначала эвристически выведем формулу для асимптотического разложения $G_n(x)$. Далее этот вывод будет использоваться в формальном доказательстве По формуле полной вероятности имеем

$$\mathsf{P}\big(\sigma w(n)(T_{N_n} \ - \ \mu) \ < \ x\big) \ = \ \mathsf{P}\Big(\sigma w(N_n)(T_{N_n} \ - \ \mu) \ < \ \frac{w(N_n)}{u(n)} \ x\Big) \ =$$

$$= \ \mathsf{E}\,\mathsf{P}\Big(w(N_n)(T_{N_n} \ - \ \mu) \ < \ \frac{w(N_n)}{w(n)} \ x\Big| \ N_n\Big) \ =$$

$$= \ \sum_{k=1}^{\infty} \ \mathsf{P}\Big(\sigma w(k)(T_k \ - \ \mu) \ < \frac{w(k)}{w(n)} \ x\Big) \ \mathsf{P}\big(N_n \ = \ k\big). \tag{4.5}$$

Используя Условие 3, вероятность под знаком ряда в формуле (4.5) аппроксимируем следующим образом

$$\mathsf{P}\Big(\sigma w(n)(T_{N_n} \ - \ \mu) \ < \ x\Big) \ \approx$$

$$\approx \ \sum_{k=1}^{\infty} \ \Big(G\Big(x \ \frac{w(k)}{w(n)}\Big) \ + \ \sum_{i=1}^{l} \ k^{-i/2} g_i\Big(x \ \frac{w(k)}{w(n)} \ \Big)\Big) \ \mathsf{P}\big(N_n \ = \ k\big) \ =$$

$$= \ \mathsf{E} \ \Big(G\Big(x \ \frac{w(N_n)}{w(n)} \ \Big) \ + \ \sum_{i=1}^{l} \ N_n^{-i/2} \ g_i\Big(x \ \frac{w(N_n)}{w(n)}\Big)\Big) \ =$$

$$= \int\limits_{1/v(n)}^{\infty} \Big(G\Big(x\ \frac{w(yv(n))}{w(n)}\Big) + \sum_{i=1}^{l} \big(yv(n)\big)^{-i/2}\ g_i\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\Big)\ d\,\mathsf{P}\Big(\frac{N_n}{v(n)} < y\Big). \tag{4.6}$$

Теперь, аппроксимируя вероятность под знаком последнего интеграла с помощью Условия 2, получим формулу для $G_n(x)$

$$\mathsf{P}\big(\sigma w(n)(T_{N_n} - \mu) < x\big) \approx G_n(x) =$$

$$= \int\limits_{1/v(n)}^{\infty} \Big(G\Big(x\ \frac{w(yv(n))}{w(n)}\Big) +$$

$$+ \sum_{i=1}^{l} \big(yv(n)\big)^{-i/2} g_i\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\Big)\ d\ \Big(H(y - u(n)) + \sum_{j=1}^{m} n^{-j/2}\ h_j(y - u(n))\Big) =$$

$$= \int\limits_{1/v(n)}^{\infty} G\Big(x\frac{w(yv(n))}{w(n)}\Big)\ dH(y - u(n)) +$$

$$+ \sum_{j=1}^{m} n^{-j/2} \int\limits_{1/v(n)}^{\infty} G\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\ dh_j(y - u(n)) +$$

$$+ \sum_{i=1}^{l} v^{-i/2}(n) \int\limits_{1/v(n)}^{\infty} y^{-i/2} g_i\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\ dH(y - u(n)) +$$

$$+ \sum_{i=1}^{l} \sum_{j=1}^{m} n^{-j/2} v^{-i/2}(n) \int\limits_{1/v(n)}^{\infty} y^{-i/2} g_i\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\ dh_i(y - u(n)). \tag{4.7}$$

Теперь, используя формулы (4.5)–(4.7), получаем оценку

$$\sup_x \big|\mathsf{P}\big(\sigma w(n)(T_{N_n} - \mu) < x\big) - G_n(x)\big| \le I_{1n} + I_{2n}, \tag{4.8}$$

где

$$I_{1n} = \sum_{k=1}^{\infty} \sup_z \Big|\mathsf{P}\big(\sigma w(k)(T_k - \mu) < z\big) - G(z) - \sum_{i=1}^{l} k^{-i/2} g_i(z)\Big|\ \mathsf{P}\big(N_n = k\big), \tag{4.9}$$

$$I_{2n} = \sup_x \Big| \int\limits_{1/v(n)}^{\infty} \Big(G\Big(x\ \frac{w(yv(n))}{w(n)}\Big) + \sum_{i=1}^{l} \big(yv(n)\big)^{-i/2} g_i\Big(x\ \frac{w(yv(n))}{w(n)}\Big)\Big) \times$$

$$\times\ d\ \Big(\mathsf{P}\Big(\frac{N_n}{v(n)} < y\Big) - H(y - u(n)) - \sum_{j=1}^{m} n^{-j/2}\ h_j(y - u(n))\Big)\Big|. \tag{4.10}$$

Ряд в определении I_{1n} (см. (4.9)) оценим с помощью Условия 3. Имеем

$$I_{1n} \le C_3 \sum_{k=1}^{\infty} \frac{1}{k^\alpha} \mathsf{P}\big(N_n = k\big) = C_3 \,\mathsf{E}\, N_n^{-\alpha}. \tag{4.11}$$

Для оценки величины I_{2n} используем равенство (4.10), Условие 2, формулу интегрирования по частям (см., например, [5], теорема 2.6.11, стр. 222 или [15], теорема 18.4, стр. 236) и ограниченность функций $g_i(z)$, $i = 1, \ldots, l$. Получим, что справедливы неравенства

$$I_{2n} \leqslant \frac{C_2}{n^\beta} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big) +$$

$$+ \sup_x \Big| \int\limits_{1/v(n)}^{\infty} \Big(\mathsf{P}\Big(\frac{N_n}{v(n)} < y\Big) - H(y - u(n)) - \sum_{j=1}^{m} n^{-j/2} h_j(y - u(n))\Big) \times$$

$$\times\, d\,\Big(G\Big(x\,\frac{w(yv(n))}{w(n)}\Big) + \sum_{i=1}^{l} \big(yv(n)\big)^{-i/2} g_i\Big(x\,\frac{w(yv(n))}{w(n)}\Big)\Big)\Big| \leqslant$$

$$\leqslant \frac{C_2}{n^\beta} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big) +$$

$$+ \sup_x \int\limits_{1/v(n)}^{\infty} \Big|\mathsf{P}\Big(\frac{N_n}{v(n)} < y\Big) - H(y - u(n)) - \sum_{j=1}^{m} n^{-j/2} h_j(y - u(n))\Big| \times$$

$$\times \Big|\frac{\partial}{\partial\, y} \Big(G\Big(x\,\frac{w(yv(n))}{w(n)}\Big) + \sum_{i=1}^{l} (yv(n))^{-i/2} g_i\Big(x\,\frac{w(yv(n))}{w(n)}\Big)\Big)\Big|\, dy \leqslant$$

$$\leqslant \frac{C_2(1 + M_n)}{n^\beta} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big). \tag{4.12}$$

Теперь утверждение теоремы следует из неравенств (4.6)–(4.12). Теорема доказана. □

Следствие 4.1. *Если моменты* $\mathsf{E}\ (N_n/v(n))^{-\alpha}$ *равномерно по* n *ограничены, то есть*

$$\mathsf{E}\left(\frac{N_n}{v(n)}\right)^{-\alpha} \le C_4, \quad C_4 > 0,\ n \in \mathbb{N},$$

то правая часть неравенства в формулировке Теоремы 4.1 приобретает вид

$$\frac{C_3\ C_4}{v^\alpha(n)} + \frac{C_2(1 + M_n)}{n^\beta} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big).$$

Следствие 4.2. *В силу неравенства Гельдера при* $0 < \alpha \leq 1$ *справедлива оценка*

$$\mathsf{E}\, N_n^{-\alpha} \leq \left(\mathsf{E}\, N_n^{-1}\right)^{\alpha},$$

которая может быть полезной при практическом применении Теоремы 4.1. В этом случае правая часть неравенства из формулировки этой теоремы может быть записана в виде

$$C_3 \left(\mathsf{E}\, N_n^{-1}\right)^{\alpha} + \frac{C_2(1 + M_n)}{n^{\beta}} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big).$$

Для вычисления $\mathsf{E}\, N_n^{-1}$ можно использовать следующую формулу (см., например, [25], стр. 93, задача 40 (б)). Если неотрицательная целочисленная с.в. N имеет производящую функцию

$$\Psi(s) = \mathsf{E}\, s^N, \qquad |s| \leqslant 1,$$

то из теоремы Фубини непосредственно следует, что

$$\mathsf{E}\, N^{-1} = \int\limits_0^1 \frac{\Psi(s)}{s}\, ds. \tag{4.13}$$

Используя это соотношение, оценку из формулировки Теоремы 4.1 можно представить в виде

$$C_3 \Bigg(\int\limits_0^1 \frac{\Psi_n(s)}{s}\, ds\Bigg)^{\alpha} + \frac{C_2(1 + M_n)}{n^{\beta}} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big), \tag{4.14}$$

где $\Psi_n(s)$ – производящая функция с.в. N_n.

Приведем пример использования формулы (4.13). Пусть с.в. N_n имеет геометрическое распределение

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

В этом случае производящая функция $\Psi_n(s)$ имеет вид

$$\Psi_n(s) = \mathsf{E}\, s^{N_n} = \frac{s}{n}\Big[1 - s\Big(1 - \frac{1}{n}\Big)\Big]^{-1} = \frac{s}{n - s(n-1)}, \quad |s| \leqslant 1,$$

поэтому

$$\mathsf{E}\, N_n^{-1} = \int\limits_0^1 \frac{\Psi_n(s)}{s}\, ds = \frac{1}{n - 1} \log n, \quad n > 1. \tag{4.15}$$

С учетом формулы (4.15) оценка (4.14) принимает вид

$$C_3 \frac{\log^{\alpha} n}{(n - 1)^{\alpha}} + \frac{C_2(1 + M_n)}{n^{\beta}} \Big(1 + \sup_x \sum_{i=1}^{l} |g_i(x)|\Big), \quad n > 1.$$

Замечание 4.1. Заметим, что из Условия 2 в частности вытекает, что если $u(n) = 0$, $n \in \mathbb{N}$, то с.в. $N_n/v(n)$ слабо сходится к с.в. V, имеющей ф.р. $H(x)$. Из определения слабой сходимости с функцией $x^{-\alpha}$, $x \geqslant 1$ в случае, если $N_n \geqslant v(n)$, $n \in \mathbb{N}$, следует, что

$$\mathsf{E}\left(\frac{N_n}{v(n)}\right)^{-\alpha} \longrightarrow \mathsf{E}\,\frac{1}{V^{\alpha}}, \quad n \to \infty,$$

то есть моменты $\mathsf{E}\,(N_n/v(n))^{-\alpha}$ равномерно ограничены по n и справедливо утверждение из Следствия 4.1.

Случай, когда $N_n \geqslant v(n)$, возникает, например, если с.в. N_n принимает значения $v(n)$, $2v(n)$, $\ldots$, $kv(n)$ с равными вероятностями $1/k$ при любом фиксированном $k \in \mathbb{N}$. В этом случае с.в. $N_n/v(n)$ вообще не зависит от n и, значит, слабо сходится к с.в. V, которая принимает значения $1, 2, \ldots, k$ с равными вероятностями $1/k$.

Заключение

Таким образом, в работе получены асимптотические разложения для функций распределения статистик, основанных на выборках случайного объема. Эти асимптотические разложения непосредственно зависят от асимптотических разложений функций распределения случайного индекса и статистики, основанной на неслучайной выборке. Подобного рода утверждения обычно называются теоремами переноса. Поэтому в работе доказаны теоремы переноса, касающиеся асимптотических разложений. Приведены два примера, иллюстрирующие доказанные теоремы. Первый пример касается распределения Стьюдента, а второй – распределения Лапласа.

Список литературы

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

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

[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.

[4] Гнеденко Б.В. Курс теории вероятностей. М.: Наука, 1988.

[5] Ширяев А.Н. Вероятность. М.: Наука, 1989.

[6] Круглов В.М., Королев В.Ю. Предельные теоремы для случайных сумм. М.: Издательство Московского университета, 1990.

[7] Gnedenko B.V., Korolev V.Yu. Random summation. Limit theorems and applications. Boca Raton: CRC Press, 1996.

[8] Bening V.E., Korolev V.Yu. Generalized poisson models and their applications in insurance and finance. Utrecht: VSP, 2002.

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

[10] Бенинг В.Е., Королев В.Ю., У Да. Оценки скорости сходимости распределений некоторых статистик к распределению Стьюдента // Вестник Российского университета дружбы народов. Серия: Прикладная математика и информатика. 2004. №1. С. 59–74.

[11] Лямин О.О. О скорости сходимости распределений некоторых статистик к распределению Лапласа и Стьюдента // Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика. 2011. №1. С. 39–47.

[12] Von Ghossy R., Rappl G. Some approximation methods for the distribution of random sums // Insurance: Mathematics and Economics. 1983. Vol. 2. Pp. 251–270.

[13] Двайт Г.Б. Таблицы интегралов и другие математические формулы. М.: Наука, 1977.

[14] Королев В.Ю. Предельные распределения для случайно индексированных последовательностей и их применения: дис. ... докт. физ.-мат. наук. М.: МГУ, 1993.

[15] Billingsley P. Probability and Measure. John Wiley & Sons, 1995.

[16] Bickel P.G. Edgeworth expansions in nonparametric statistics // Annals of Statistics. 1974. Vol. 2. Pp. 1–21.

[17] Albers W. Asymptotic expansions and the deficiency concept in statistics // Mathematical Centre Tracts. 1974. Vol. 58.

[18] Helmers R. Edgeworth expansions for linear combinations of order statistics // Mathematical Centre Tracts. 1984. Vol. 105.

[19] Albers W., Bickel P.G., Van Zwet W.R. Asymptotic expansions for the power of distribution free tests in the one-sample problem // Annals of Statistics. 1976. Vol. 4. Pp. 108–156.

[20] Bickel P.G., Van Zwet W.R. Asymptotic expansions for the power of distribution free tests in the two-sample problem // Annals of Statistics. 1978. Vol. 6. Pp. 947–1004.

[21] Bentkus V., Gotze F., Van Zwet W.R. An Edgeworth expansions for symmetric statistics // Annals of Statistics. 1997. Vol. 25. Pp. 851–896.

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

[23] Бенинг В.Е., Королев В.Ю., Соколов И.А., Шоргин С.Я. Рандомизированные модели и методы теории надежности информационных и технических систем. М.: Торус Пресс, 2007.

[24] Королев В.Ю. О взаимосвязи обобщенного распределения Стьюдента и дисперсионного гамма-распределения при статистическом анализе выборок случайного объема // Доклады РАН. 2012. Т. 445, №6. С. 622–627.

[25] Климов Г.П. Теория вероятностей и математическая статистика. М.: Издательство Московского университета, 1983.

[26] Wilks S.S. Recurrence of extreme observations // Journal of American Mathematical Society. 1959. Vol. 1, №1. Pp. 106–112.

[27] Невзоров В.Б. Рекорды. Математическая теория. М.: Фазис, 2000.

[28] Бенинг В.Е., Галиева Н.К, Королев В.Ю. Асимптотические разложения для функций распределения статистик, построенных по выборкам случайного объема // Информатика и ее применения. 2013. Т. 7, №2. С. 75–83.

Библиографическая ссылка

Бенинг В.Е., Савушкин В.А. Об аппроксимации распределений статистик, основанных на выборках случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2014. №1. С. 91–111.

Сведения об авторах

1. **Бенинг Владимир Евгеньевич**
профессор кафедры математической статистики МГУ им. М.В. Ломоносова.

Россия, 119992, г. Москва, ГСП-1, Воробьевы горы, МГУ, ф-т ВМиК.
E-mail: bening@yandex.ru.

2. **Савушкин Владислав Андреевич**
Международный университет природы, общества и человека «Дубна».

Россия, 141980, Московская область, г.Дубна, ул.Университетская, 19, Международный университет природы, общества и человека «Дубна».
E-mail: savushkinva@mail.ru.

ON THE APPROXIMATION OF THE DISTRIBUTIONS OF STATISTICS BASED ON THE SAMPLES WITH RANDOM SIZE

Bening Vladimir Evgen'evich
Professor of the Department of Mathematical Statistics,
Lomonosov Moscow State University
Russia, 119992, Moscow, GSP-1, Vorob'evy Gory, Lomonosov Moscow State University, faculty of Computational Mathematics and Cybernetics.
E-mail: bening@yandex.ru

Savushkin Vladislav Andreevich
International University of Nature, Society and Man «Dubna»
Russia, 141980, Moscow region, Dubna, 19 Universitetskaya str., International University of Nature, Society and Man «Dubna».
E-mail: savushkinva@mail.ru

Received 20.01.2014, revised 25.01.2014.

In the paper general theorem concerning the asymptotic expansions of the distribution function of the statistics based on the sample of random size is proved.Two examples (Laplace distribution and Student distribution) are presented.

Keywords: sample of random size, asymptotic expansions, mixtures of probability laws, Laplace distribution, Student distribution, transfer theorem.

Bibliographic citation

Bening V.E., Savushkin V.A. On the approximation of the distributions of statistics based on the samples with random size. *Vestnik TvGU. Seriya: Prikladnaya matematika* [Herald of Tver State University. Series: Applied Mathematics], 2014, no. 1, pp. 91–111. (in Russian)