УДК 519

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ФУНКЦИИ РИСКА И ДЕФЕКТЫ ОЦЕНОК, 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 14.04.2019, после переработки 20.05.2019.

В работе рассматривается случай, когда число наблюдений случайно. Это приводит к возникновению распределений с тяжелыми хвостами и к изменению эффективности обычно используемых статистических процедур. В работе проведено асимптотическое сравнение статистических оценок, основанных на выборках случайного и неслучайного объемов. Для этого используются асимптотические разложения для функции риска и понятие асимптотический дефект (введенное в работе [1]), который имеет смысл добавочного числа наблюдений, необходимого данной оценке для достижения того же качества, что и оптимальной оценке. Получены также асимптотические разложения для функций риска оценок, основанных на выборках случайного объема.

Ключевые слова: статистическая оценка, асимптотический дефект, выборка случайного объема, функция риска, асимптотическое разложение.

Вестник ТвГУ. Серия: Прикладная математика. 2019. № 2. С. 5–25.
https://doi.org/10.26456/vtpmk530

Введение

В классических задачах математической статистики объем выборки или количество наблюдений, доступное 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

[1]Работа выполнена при поддержке РФФИ, проект №18-07-00252.

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, при определении наработки на отказ) разных партий приборов, 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 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 работе развивается подход, предложенный в работах [5, 6]. Напомним кратко постановку задачи и основные обозначения. Рассмотрим сначала задачу статистического оценивания известной параметрической функции $g(\theta)$, зависящей от неизвестного параметра θ и обозначим через $m(n)$ необходимое число наблюдений, которое требуется оценке $\delta_{m(n)}(X_1, \ldots, X_{m(n)})$ для достижения такого же качества (например, среднеквадратичного отклонения или дисперсии), что и оценке $\delta_n^*(X_1, \ldots, X_n)$, основанной на n наблюдениях $X_1, \ldots, X_n$. Мы рассматриваем асимптотический подход, означающий, что $n \to \infty$. Под асимптотической относительной эффективностью (АОЭ) оценки $\delta_n(X_1, \ldots, X_n)$ по отношению к оценке $\delta_n^*(X_1, \ldots, X_n)$ понимается предел (в случае его существования и независимости от последовательности $m(n)$) вида (см., например, [3] стр. 305)

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

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 оценке $\delta_n(X_1, \ldots, X_n)$. 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 [1]). Они назвали разность $m(n) - n$ дефектом (deficiency) конкурирующей оценки δ_n относительно оценки δ_n^* и предложили обозначение

$$d_n = m(n) - n. \tag{1.1}$$

Если предел $\lim\limits_{n\to\infty} d_n$ существует, то он называется *асимптотическим дефектом* оценки δ_n относительно оценки δ_n^* и обозначается символом d. Часто d называют просто дефектом δ_n относительно δ_n^*. Заметим, что если АОЭ $e \neq 1$, то $d = \infty$ и этот случай малоинтересен. В работе [1] также было отмечено, что существуют статистические задачи, в которых типичным образом возникает случай $e = 1$ (см., например, книгу [4]), то есть в этом случае понятие АОЭ не дает ответа на вопрос какая оценка лучше и понятие дефекта проясняет эту ситуацию, поскольку в этом случае асимптотический дефект может, в принципе, быть любым.

Обозначим функции риска оценок δ_n и δ_n^*, соответственно, через $R_n(\theta)$ и $R_n^*(\theta)$, где θ — неизвестный параметр (произвольной природы), тогда по определению величины $d_n(\theta) \equiv d_n = m(n) - n$, для каждого n должно выполняться равенство

$$R_n^*(\theta) = R_{m(n)}(\theta). \tag{1.2}$$

При решении уравнения (1.2) целочисленную величину $m(n)$ следует рассматривать как переменную, принимающую произвольные действительные значения. Для этого можно определить функцию риска $R_{m(n)}(\theta)$ для нецелых значений $m(n)$ по формуле

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

(см. работу [1]).

Типичным образом функции риска $R_n^*(\theta)$ и $R_n(\theta)$ неизвестны точно и используются их аппроксимации. Предположим, что для функций риска $R_n^*(\theta)$ и $R_n(\theta)$ справедливы асимптотические разложения вида

$$R_n^* = \frac{a(\theta)}{n^r} + \frac{b(\theta)}{n^{r+s}} + o\big(n^{-r-s}\big), \tag{1.3}$$

$$R_n = \frac{a(\theta)}{n^r} + \frac{c(\theta)}{n^{r+s}} + o\big(n^{-r-s}\big), \tag{1.4}$$

где $a(\theta)$, $b(\theta)$ и $c(\theta)$ – некоторые постоянные, не зависящие от n, а $r > 0$, $s > 0$ – некоторые константы, определяющие порядок убывания по n этих функций риска. Первый член в этих асимптотических разложениях одинаков, и это отражает тот факт, что АОЭ этих оценок равна единице. Из соотношений (1.1)–(1.4) легко получить, что (см. работу [1] или книгу [3], стр. 310)

$$d_n(\theta) \equiv \frac{c(\theta) - b(\theta)}{r\,a(\theta)}\, n^{1-s} + o\big(n^{1-s}\big). \tag{1.5}$$

Таким образом асимптотический дефект имеет вид

$$d(\theta) \equiv d = \begin{cases} \pm\infty, & 0 < s < 1, \\ \dfrac{c(\theta) - b(\theta)}{r\,a(\theta)}, & s = 1, \\ 0, & s > 1. \end{cases} \tag{1.6}$$

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

Совершенно аналогично определяется дефект и в общем случае асимптотического сравнения двух статистических процедур, соответственно с мерами качества π_n и π_n^*. В том случае необходимое число наблюдений k_n для первой процедуры определяется из равенства (считая k_n непрерывной переменной)

$$\pi_{k_n} = \pi_n^*,$$

а предел вида (в случае его существования)

$$d = \lim_{n\to\infty} (k_n - n)$$

называется асимптотическим дефектом первой процедуры относительно второй. Если для π_n^* и π_n выполняются формулы типа (1.3) и (1.4), то для асимптотического дефекта d справедливы соотношения типа (1.5) и (1.6).

В настоящей работе приведены примеры вычисления асимптотического дефекта в задачах статистического оценивания в случае, когда оценки (статистики) основаны на выборках случайного объема. Используя асимптотические разложения для функций риска и понятие асимптотического дефекта, проведено асимптотическое сравнение качества оценок, основанных на выборках случайного и неслучайного объемов. Приведены конкретные примеры распределения случайного индекса.

1. Асимптотическое поведение функции риска и обратных моментов случайного индекса

Рассмотрим случайные величины (с.в.) $N_1, N_2, \ldots$ и $X_1, X_2, \ldots$, заданные на одном и томже вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$. При этом случайные величины $X_1, X_2, \ldots X_n$ имеют смысл статистических наблюдений, а случайная величина N_n трактуется как случайный объем выборки, зависящий от натурального параметра $n \in \mathbb{N}$. Всюду ниже будет предполагаться, что

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

то есть в среднем объем случайной выборки равен размеру выборки неслучайного размера. При нахождении дефектов статистических процедур, основанных на выборках случайного объема $N_{m(n)}$, и соответствующей процедуры, основанной на выборке неслучайного объема n, мы фактически сравниваем средний объем случайной выборки $m(n)$ и n с помощью величины $d_n = m(n) - n$ и ее предела.

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, ... Всюду далее предполагается, что случайные величины $X_1, X_2, \ldots$ независимы, одинаково распределены и имеют распределение, зависящее от неизвестного параметра $\theta \in \Theta$, при этом в принципе множество Θ может иметь произвольную природу.

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

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

Далее, под статистикой в этом разделе будем понимать оценку действительной известной функции $g(\theta)$, зависящей от неизвестного параметра $\theta \in \Theta$, и будем обозначать ее символами типа $\delta_n(X_1, \ldots, X_n)$. Всюду далее будем рассматривать функции риска вида

$$\begin{aligned} R_n(\theta) \;&=\; \mathsf{E}_\theta \left|\delta_{N_n}(X_1, \ldots, X_{N_n}) \;-\; g(\theta)\right|^r, \\ R_n^*(\theta) \;&=\; \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) \;-\; g(\theta)\right|^r, \qquad r \;>\; 0, \end{aligned} \tag{2.2}$$

то есть это риски оценок, основанных на выборках случайного и неслучайного объемов.

Сформулируем условия на асимптотическое поведение функций распределения нормированной исходной оценки и случайного индекса.

Условие 1. *Для некоторых* $\alpha \;>\; 0,\; l \;>\; 0,\; \delta_1 \;>\; 0,\; C_1 \;>\; 0$ *и* $k \;\in\; \{0, 1, \ldots\}$ *справедливо неравенство*

$$\Big|\mathsf{P}\big(n^\alpha(\delta_n \;-\; g(\theta)) \;<\; x\big) \;-\; G(x) \;-\; \sum_{i=1}^{k} \frac{g_i(x)}{n^{i/2}}\Big| \;\leq\; \frac{C_1}{n^{k/2+\delta_1}(1+|x|)^l},$$

где $G(x)$ *и* $g_i(x)$ $i = 1, \ldots, k$ *– некоторые измеримые функции такие, что*

$$\int_0^\infty x^{r-1}\left|1 \;-\; G(x) \;+\; G(-x)\right| dx \;<\; \infty,\; \int_{-\infty}^\infty x^{r-1}\left|g_i(x)\right| dx \;<\; \infty,\;\; i \;=\; 1, \ldots, k.$$

Условие 2. *Для некоторых* $\beta > 0$, $\delta_2 > 0$, $C_2 > 0$ *и* $m \in \{0, 1, \ldots\}$ *справедливо неравенство*

$$\Big|\mathsf{P}\big(N_n/n^{\beta} < x\big) - H(x) - \sum_{j=1}^{m} \frac{h_j(x)}{n^{j/2}}\Big| \le \frac{C_2}{n^{m/2+\delta_2}}\, h_{m+1}(x),\ h_{m+1}(x) \ge 0,\ x \in \mathbb{R},$$

где $H(x)$ *и* $h_j(x)$ $j = 1, \ldots, m+1$ *– некоторые измеримые ограниченные функции.*

Лемма 2.1. *Пусть выполнено Условие 1 и* $l > r$, *тогда для функции риска* $R_n^*(\theta)$ *справедливо асимптотическое разложение вида*

$$R_n^*(\theta) = \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) - g(\theta)\right|^r =$$

$$= rn^{-r\alpha} \int_0^{\infty} x^{r-1} \big(1 - G(x) + G(-x)\big)\, dx +$$

$$+ rn^{-r\alpha} \sum_{i=1}^{k} n^{-i/2} \int_0^{\infty} x^{r-1} \big(g_i(-x) - g_i(x)\big)\, dx + \rho^*_{nk},$$

где

$$\left|\rho^*_{nk}\right| \le 2r\, C_1\, n^{-\alpha r - k/2 - \delta_1} \int_0^{\infty} \frac{x^{r-1}}{(1 + x)^l}\, dx.$$

Доказательство. Интегрируя по частям, имеем

$$R_n^*(\theta) = n^{-r\alpha}\, \mathsf{E}_\theta \left|n^{\alpha}(\delta_n(X_1, \ldots, X_n) - g(\theta))\right|^r =$$

$$= rn^{-r\alpha} \int_0^{\infty} x^{r-1} \Big(\mathsf{P}_\theta\big(n^{\alpha}(\delta_n(X_1, \ldots, X_n) - g(\theta)) \ge x\big) +$$

$$+ \mathsf{P}_\theta\big(n^{\alpha}(\delta_n(X_1, \ldots, X_n) - g(\theta)) < -x\big)\Big)\, dx,$$

отсюда, используя Условие 1, получаем утверждение Леммы. Лемма доказана. □

Приведем пример использования Леммы 2.1.

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

$$\mathsf{E}_\theta\, X_1 = \theta \in \mathbb{R},\ \mathsf{D}_\theta\, X_1 = \sigma^2,\ \mathsf{E}_\theta\, |X_1|^{k+2+\delta_1} < \infty,\ k \ge 1,\ k \in \mathbb{N},\ \delta_1 > 0. \quad (2.3)$$

Для каждого n пусть

$$\delta_n = \frac{1}{n}\,(X_1 + \ldots + X_n). \quad (2.4)$$

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

$$\limsup_{|t| \to \infty} |\mathsf{E}_\theta\ \exp\{it\ X_1\}| < 1. \quad (2.5)$$

При выполнении условий (2.3) и (2.5) из Теоремы 6.3.2 книги [7] следует выполнение неравенства для $n \in \mathbb{N}$ и $C_{k,\theta} > 0$

$$\left| \mathsf{P}_\theta\big(\sqrt{n}\sigma^{-1} \, (\delta_n - \theta) < x\big) - \Phi(x) - \sum_{i=1}^{k} n^{-i/2} \, Q_i(x) \right| \le \frac{C_{k,\theta}}{n^{k/2+\delta_1}(1 + |x|)^{k+2}}, \tag{2.6}$$

где функции $Q_1(x), \ldots, Q_k(x)$ определены в книге [7], например,

$$Q_1(x) = -(x^2 - 1) \, \phi(x) \, \frac{\mathsf{E}\, Y_1^3}{6},$$

$$Q_2(x) = -(x^3 - 3x) \, \phi(x) \, \frac{\mathsf{E}\, Y_1^4 - 3}{24} -$$

$$- (x^5 - 10x^3 + 15x) \, \phi(x) \, \frac{(\mathsf{E}\, Y_1^3)^2}{72}, \quad Y_1 = \frac{X_1 - \theta}{\sigma}. \tag{2.7}$$

Таким образом, Условие 1 выполнено с

$$\alpha = \frac{1}{2}, \quad g(\theta) = \theta, \quad G(x) = \Phi(x),$$

$$g_i(x) = Q_i(x), \quad i = 1, \ldots, k, \quad l = k + 2, \quad C_k = C_{k,\theta}.$$

Следствие 2.1. *Пусть выполнены условия (2.3) и (2.5), тогда для любого $0 < r < k+2$ для функции риска $R_n^*(\theta)$ оценки (2.4) справедливо асимптотическое разложение вида*

$$R_n^*(\theta) = \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) - \theta\right|^r =$$

$$= \sigma^r n^{-r/2} \left(\mathsf{E}\, |\eta|^r - 2r \sum_{i=1}^{[k/2]} n^{-i} \int_0^\infty x^{r-1} \, Q_{2i}(x) \, dx\right) + \rho_{nk}^*,$$

где η — стандартная нормальная с.в. с параметрами $(0,1)$,

$$\left|\rho_{nk}^*\right| \le 2\sigma^r \, r \, C_{k,\theta} \, n^{-r/2-k/2-\delta_1} \int_0^\infty \frac{x^{r-1}}{(1 + x)^{k+2}} \, dx.$$

В частности, при $k = 2$ справедливо равенство

$$R_n^*(\theta) = \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) - \theta\right|^r =$$

$$= \sigma^r n^{-r/2} \left(\mathsf{E}\, |\eta|^r - 2r \, n^{-1} \int_0^\infty x^{r-1} \, Q_2(x) \, dx\right) + \rho_{n2}^*.$$

Лемма 2.2. *Пусть выполнено Условие 2, тогда для для любого $p > 0$ справедливо асимптотическое разложение вида*

$$\mathsf{E}\, N_n^{-p} = p \int_1^\infty \frac{H(xn^{-\beta})}{x^{p+1}} \, dx + p \sum_{j=1}^{m} n^{-j/2} \int_1^\infty \frac{h_j(xn^{-\beta})}{x^{p+1}} \, dx + \rho_{nm},$$

где

$$\left|\rho_{nm}\right| \le p \, C_2 \, n^{-m/2-\delta_2} \int_1^\infty \frac{h_{m+1}(xn^{-\beta})}{x^{p+1}} \, dx.$$

Доказательство. Интегрируя по частям, имеем

$$\mathsf{E}\, N_n^{-p} \;=\; p \int_1^{\infty} \frac{1}{x^{p+1}}\, \mathsf{P}\big(N_n \;<\; x\big)\, dx \;=\; p \int_1^{\infty} \frac{1}{x^{p+1}}\, \mathsf{P}\big(n^{-\beta} N_n \;<\; x n^{-\beta}\big)\, dx,$$

теперь с учетом Условия 2 отсюда следует утверждение Леммы. Лемма доказана. □

Применим Лемму 2.2 к конкретным случайным индексам N_n.

В работе [9] предложена модель, в которой в качестве предельных законов распределений статистик, основанных на выборках случайного объема, возникает распределение Стьюдента с произвольным числом степеней свободы. При этом возникал случайный индекс $N_n(s),\ s \;>\; 1$, имеющий отрицательно биномиальное распределение с вероятностью успеха $1/n$ вида

$$\mathsf{P}\;\big(N_n(s) \;=\; j\big) \;=\; \frac{\Gamma(j+s-1)}{(j-1)!\Gamma(s)} \left(\frac{1}{n}\right)^{j} \left(1 \;-\; \frac{1}{n}\right)^{j-1}, \quad j \;\in\; \mathbb{N}, \quad s \;>\; 0, \tag{2.8}$$

где $\Gamma(\cdot)$ — гамма–функция. В работе [11] (Теорема 1 и Лемма 1) была получена следующая оценка скорости сходимости для ф.р. нормированного случайного индекса $N_n(s)$: для всех $x \;>\; 0$ и $n \;\in\; \mathbb{N}$ существует константа $C(s) \;>\; 0$ такая, что справедливо неравенство

$$\sup_{x>0} \left| \mathsf{P}\Big(\frac{N_n(s)}{s(n-1)+1} \;<\; x\Big) \;-\; H_{n,2}(x) \right| \;\le\; C(s)\; n^{-\min\{s,2\}},$$

$$\mathsf{E}\; N_n^{-3/2}(s) \;=\; \begin{cases} C_s\; n^{-s}, & 1 \;<\; s \;<\; 3/2, \\ C_s\; \log n\; n^{-3/2}, & s \;=\; 3/2, \\ C_s\; n^{-3/2}, & s \;>\; 3/2, \end{cases} \tag{2.9}$$

где

$$H_{n,2}(x) \;=\; G_s(x) \;+\; \frac{g_s(x)\big((x-1)(2-s) \;+ 2L(x(s(n-1)+1))\big)}{2sn}\; \mathbb{I}_{[2,\infty)}(n),$$

$$L(y) \;=\; \frac{1}{2} \;-\; (y \;-\; [y]),$$

$$G_s(x) \;=\; \int_{-\infty}^{x} g_s(t)\; dt, \quad g_s(x) \;=\; \frac{s^s}{\Gamma(s)}\; x^{s-1}\; e^{-sx}\; \mathbb{I}_{(0,\infty)}(x), \quad x \;\in\; \mathbb{R}.$$

Таким образом, Условие 2 выполнено при

$$\beta \;=\; 1,\;\; m \;=\; 2,\;\; C_2 \;=\; C(s),\; H(x) \;=\; G_s(xn/s(n{-}1){+}1),\;\; h_1(x) \;\equiv\; 0, \;\; h_3(x) \;\equiv\; 1,$$

$$h_2(x) \;=\; n\; \big(H_{n,2}(xn/s(n-1)+1) \;-\; G_s(xn/s(n-1)+1)\big).$$

Следствие 2.2.1. *Пусть* $0 \;<\; p \;\le\; s \;-\; 1,\; s \;>\; 1$, *тогда для случайного индекса вида (2.8) справедливы асимптотические разложения*

$$\sup_{x>0} \left| \mathsf{P}\Big(\frac{N_n(s)}{n} \;<\; x\Big) \;-\; H_{n,2}\Big(\frac{xn}{s(n-1)+1}\Big) \right| \;\le\; C(s)\; n^{-\min\{s,2\}},$$

$$\mathsf{E}\ N_n^{-p}(s) \ = \ \frac{s^p\ \Gamma(s-p)}{(s(n-1)+1)^p\ \Gamma(s)} \ +$$
$$+\ \frac{p}{2s\ n^{p+1}\Gamma(s)}\Big((2\ -\ s)\big(\Gamma(s-p)\ -\ s\Gamma(s-p-1)\big)\ +\ \frac{2\Gamma(s)}{s^p}\ \int_0^\infty\ \frac{g_s(u)L(nus)}{u^{p+1}}\ du\Big)\ +$$
$$+o(n^{-\min\{p+1,2\}}).$$

Доказательство Следствия следует из равенств

$$\mathsf{E}\ N_n^{-p}(s)\ =$$
$$=p\ \int_1^\infty\ \frac{G_s\big(x/(s(n-1)+1)\big)}{x^{p+1}}\ dx\ +\ p\ n^{-1}\ \int_1^\infty\ \frac{h_2(xn^{-1})}{x^{p+1}}\ dx\ +\ O(n^{-\min\{s,2\}})\ =$$
$$=\ \frac{p}{(s(n-1)+1)^p}\ \int_{\frac{1}{s(n-1)+1}}^\infty\ \frac{G_s(u)}{u^{p+1}}\ du\ +\ \frac{p}{n^{p+1}}\ \int_{\frac{1}{n}}^\infty\ \frac{h_2(u)}{u^{p+1}}\ du\ +\ O(n^{-\min\{s,2\}}),$$
$$\frac{p}{(s(n-1)+1)^p}\ \int_{\frac{1}{s(n-1)+1}}^\infty\ \frac{G_s(u)}{u^{p+1}}\ du\ =\ \frac{s^p\ \Gamma(s-p)}{(s(n-1)+1)^p\ \Gamma(s)}\ +\ O(n^{-s}),$$
$$p\ n^{-1}\ \int_1^\infty\ \frac{h_2(xn^{-1})}{x^{p+1}}\ dx\ =$$
$$=\ \frac{p}{2s\ n^{p+1}\Gamma(s)}\Big((2\ -\ s)\big(\Gamma(s-p)\ -\ s\Gamma(s-p-1)\big)\ +\ \frac{2\Gamma(s)}{s^p}\ \int_0^\infty\ \frac{g_s(u)L(nus)}{u^{p+1}}\ du\Big)\ +$$
$$+o(n^{-p-1}).$$

В работе [10] рассмотрен случай, когда в качестве предельных законов распределений статистик, основанных на выборках случайного размера, возникает распределение Лапласа. Там же описан случайный индекс N_{ns}, $s\ >\ 0$, имеющий распределение вида

$$\mathsf{P}\ \big(N_{ns}\ =\ j\big)\ =\ \Big(\frac{j}{s+j}\Big)^n\ -\ \Big(\frac{j-1}{s+j-1}\Big)^n,\ \ j\ \in\ \mathbb{N},\ \ s\ >\ 0. \tag{2.10}$$

В работе [11] (Теорема 4 и Лемма 3) доказана следующая оценка скорости сходимости для ф.р. нормированного случайного индекса N_{ns}: для всех $x\ >\ 0$ и $n\ \in\ \mathbb{N}$ существует константа $C_s\ >\ 0$ такая, что при $s\ \geq\ s_0\ >\ 0$, s_0 произвольно мало, справедливы следующие соотношения

$$\sup_{x>0}\Big|\mathsf{P}\Big(\frac{N_{ns}}{n}\ <\ x\Big)\ -\ \bar{H}_{n,2}(x)\Big|\ \leq\ C_s\ n^{-2}, \tag{2.11}$$

где

$$\bar{H}_{n,2}(x)\ =\ \exp\{-s/x\}\Big(1\ +\ \frac{s\big(s-1\ +\ 2L(xn)\big)}{2x^2n}\Big),$$
$$L(y)\ =\ \frac{1}{2}\ -\ (y\ -\ [y]),$$
$$\mathsf{E}\ N_{ns}^{-3/2}\ =\ O(n^{-3/2}). \tag{2.12}$$

Таким образом, Условие 2 выполнено при

$$\beta = 1,\ \ m = 2,\ \ C_2 = C_s,\ H(x) = \exp\{-s/x\},\ \ h_1(x) \equiv 0,\quad h_3(x) \equiv 1,$$

$$h_2(x) = \exp\{-s/x\}\ \frac{s\big(s-1 + 2L(xn)\big)}{2x^2}.$$

Следствие 2.2.2. *Пусть* $0 < p \le 1$, *тогда для случайного индекса вида (2.10) справедливо асимптотическое разложение*

$$\mathsf{E}\ N_n^{-1}(s) = \frac{1}{sn} + o(n^{-1}),$$

$$\mathsf{E}\ N_n^{-p}(s) = \frac{p\ \Gamma(p)}{s^p\ n^p} +$$

$$+\frac{p}{2s^{p+1}\ n^{p+1}}\Big((s-1)\,\Gamma(p+2) + 2\int_0^\infty e^{-t}\,t^{p+1}\,L(sn/t)\,dt\Big) + o(n^{-p-1}),\ 0<p<1.$$

Доказательство Следствия следует из равенств

$$p\int_1^\infty \frac{e^{-sn/x}}{x^{p+1}}\,dx = \frac{p\ \Gamma(p)}{s^p\ n^p} + o(n^{-p-1}),$$

$$pn\int_1^\infty \frac{e^{-sn/x}s(s-1+2L(x))}{2x^{p+3}}\,dx =$$

$$= \frac{p}{2s^{p+1}\ n^{p+1}}\Big((s-1)\,\Gamma(p+2) + 2\int_0^\infty e^{-t}\,t^{p+1}\,L(sn/t)\,dt\Big) + o(n^{-p-1}).$$

Из этих двух Лемм можно получить асимптотическое разложение для функции риска оценок, основанных на выборках случайного объема.

Теорема 2.3. *Пусть выполнены Условия 1 и Условие 2, тогда справедливо асимптотическое разложение для функции риска оценки, основанной на выборке случайного объема:*

$$R_n(\theta) = \mathsf{E}_\theta\ \big|\delta_{N_n}(X_1,\dots,X_{N_n}) - g(\theta)\big|^r =$$

$$= r\ \mathsf{E}\ N_n^{-r\alpha}\int_0^\infty x^{r-1}\,\big(1 - G(x) + G(-x)\big)\,dx +$$

$$+ r\sum_{i=1}^k \mathsf{E}\ N_n^{-r\alpha-i/2}\int_0^\infty x^{r-1}\,\big(g_i(-x) - g_i(x)\big)\,dx + \bar\rho_{nk} =$$

$$= r^2\alpha\,\Big(\int_1^\infty \frac{H(xn^{-\beta})}{x^{r\alpha+1}}\,dx + \sum_{j=1}^m n^{-j/2}\int_1^\infty \frac{h_j(xn^{-\beta})}{x^{r\alpha+1}}\,dx + \rho_{nm}\Big)\times$$

$$\times\int_0^\infty x^{r-1}\,\big(1 - G(x) + G(-x)\big)\,dx +$$

$$+ r^2\alpha\sum_{i=1}^k \Big(\Big(\int_1^\infty \frac{H(xn^{-\beta})}{x^{r\alpha+i/2}}\,dx + \sum_{j=1}^m n^{-j/2}\int_1^\infty \frac{h_j(xn^{-\beta})}{x^{r\alpha+i/2}}\,dx + \rho_{nm}\Big)\times$$

$$\times \int_0^\infty x^{r-1} \left(g_i(-x) - g_i(x)\right) dx\Big) + \bar\rho_{nk},$$

где

$$\left|\bar\rho_{nk}\right| \le 2r\, C_1\, \mathsf{E}\, N_n^{-\alpha r - k/2 - \delta_1} \int_0^\infty \frac{x^{r-1}}{(1 + x)^l}\, dx.$$

Доказательство. По формуле полной вероятности имеем

$$R_n(\theta) = \mathsf{E}_\theta \left|\delta_{N_n}(X_1, \ldots, X_{N_n}) - g(\theta)\right|^r =$$

$$= \sum_{s=1}^{\infty} \mathsf{E}_\theta \left|\delta_s(X_1, \ldots, X_s) - g(\theta)\right|^r \mathsf{P}\Big(N_n = s\Big).$$

Применяя здесь для каждого k Лемму1, получаем, что

$$R_n(\theta) = r \sum_{s=1}^{\infty} \Big(s^{-r\alpha} \int_0^\infty x^{r-1} \left(1 - G(x) + G(-x)\right) dx +$$

$$+ s^{-r\alpha} \sum_{i=1}^{k} s^{-i/2} \int_0^\infty x^{r-1} \left(g_i(-x) - g_i(x)\right) dx\, \mathsf{P}\big(N_n = s\big)\Big) + \bar\rho_{nk},$$

где

$$\left|\bar\rho_{nk}\right| \le 2r\, C_1\, \mathsf{E}\, N_n^{-\alpha r - k/2 - \delta_1} \int_0^\infty \frac{x^{r-1}}{(1 + x)^l}\, dx. \qquad (2.13)$$

Последнее равенство можно записать в виде

$$R_n(\theta) = r\, \mathsf{E}\, N_n^{-r\alpha} \int_0^\infty x^{r-1} \left(1 - G(x) + G(-x)\right) dx +$$

$$+ r \sum_{i=1}^{k} \mathsf{E}\, N_n^{-r\alpha - i/2} \int_0^\infty x^{r-1} \left(g_i(-x) - g_i(x)\right) dx + \bar\rho_{nk}.$$

Теперь к каждому математическому ожиданию в этой формуле применим Лемму 2.

$$R_n(\theta) = r^2\alpha \Big(\int_1^\infty \frac{H(xn^{-\beta})}{x^{r\alpha+1}}\, dx + \sum_{j=1}^{m} n^{-j/2} \int_1^\infty \frac{h_j(xn^{-\beta})}{x^{r\alpha+1}}\, dx + \rho_{nm}\Big) \times$$

$$\times \int_0^\infty x^{r-1} \left(1 - G(x) + G(-x)\right) dx +$$

$$+ r^2\alpha \sum_{i=1}^{k} \Big(\Big(\int_1^\infty \frac{H(xn^{-\beta})}{x^{r\alpha+i/2}}\, dx + \sum_{j=1}^{m} n^{-j/2} \int_1^\infty \frac{h_j(xn^{-\beta})}{x^{r\alpha+i/2}}\, dx + \rho_{nm}\Big) \times$$

$$\times \int_0^\infty x^{r-1} \left(g_i(-x) - g_i(x)\right) dx\Big) + \bar\rho_{nk}.$$

Из этой формулы и неравенства (2.13) следует утверждение Теоремы. Теорема доказана. □

Заметим, что, используя доказательство этой Теоремы, можно получить асимптотическое разложение для функции риска без использования Условия 2.

Замечание 2.3. Пусть выполнено Условие 1 и $0 < r < l$, тогда справедливо асимптотическое разложение для функции риска оценки, основанной на выборке случайного объема:

$$R_n(\theta) \;=\; \mathsf{E}_\theta \left|\delta_{N_n}(X_1,\ldots,X_{N_n}) \;-\; g(\theta)\right|^r \;=$$

$$=\; r\,\mathsf{E}\,N_n^{-r\alpha} \int_0^\infty x^{r-1}\,\big(1 \;-\; G(x) \;+\; G(-x)\big)\,dx \;+$$

$$+\, r\sum_{i=1}^k \mathsf{E}\,N_n^{-r\alpha-i/2} \int_0^\infty x^{r-1}\,\big(g_i(-x) \;-\; g_i(x)\big)\,dx \;+\; O\big(\mathsf{E}\,N_n^{-r\alpha-k-\delta_1}\big).$$

Следствие 2.3.1. *Пусть выполнены условия Следствия 2.1 при $k = 2$ и Следствия 2.2.1, тогда при $s > 2$ и $0 < r < 2\min\{1-\delta_1, s-2-\delta_1\}$ для случайного индекса вида (2.8) справедливо асимптотическое разложение*

$$R_n(\theta) \;=\; \mathsf{E}_\theta \left|\delta_{N_n(s)}(X_1,\ldots,X_{N_n(s)}) \;-\; \theta\right|^r \;=$$

$$=\; \sigma^r\,\mathsf{E}\,N_n^{-r/2}(s)\,\mathsf{E}\,|\eta|^r \;-\; 2r\,\sigma^r\,\mathsf{E}\,N_n^{-r/2-1}(s) \int_0^\infty x^{r-1}\,Q_2(x)\,dx \;+\; o\big(n^{-r/2-1}\big),$$

где

$$\mathsf{E}\,N_n^{-r/2}(s) \;=\; \frac{s^{r/2}\,\Gamma(s-r/2)}{(s(n-1)+1)^{r/2}\,\Gamma(s)} \;+\; \frac{r/2}{2s\;n^{r/2+1}\Gamma(s)} \;\times$$

$$\times\, \Big((2-s)\big(\Gamma(s{-}r/2) - s\Gamma(s{-}r/2{-}1)\big) + \frac{2\Gamma(s)}{s^{r/2}} \int_0^\infty \frac{g_s(u)L(nus)}{u^{r/2+1}}\,du\Big) + o\big(n^{-r/2-1}\big),$$

$$\mathsf{E}\,N_n^{-r/2-1}(s) \;=\; \frac{\Gamma(s-r/2-1)}{n^{r/2+1}\Gamma(s)} \;+\; o\big(n^{-r/2-1}\big).$$

Следствие 2.3.2. *Пусть выполнены условия Следствия 2.1 при $k = 2$ и Следствия 2.2.2, тогда при $0 < r < 2$, $s \ge s_0 > 0$ для случайного индекса вида (2.10) справедливо асимптотическое разложение*

$$R_n(\theta) \;=\; \mathsf{E}_\theta \left|\delta_{N_n}(X_1,\ldots,X_{N_n}) \;-\; \theta\right|^r \;=$$

$$=\; \sigma^r\,\mathsf{E}\,N_n^{-r/2}\,\mathsf{E}\,|\eta|^r \;-\; 2r\,\sigma^r\,\mathsf{E}\,N_n^{-r/2-1} \int_0^\infty x^{r-1}\,Q_2(x)\,dx \;+\; o\big(n^{-r/2-1}\big),$$

где

$$\mathsf{E}\,N_n^{-r/2}(s) \;=\; \frac{r\,\Gamma(r/2)}{2s^{r/2}\;n^{r/2}} \;+$$

$$+\; \frac{r}{4s^{r/2+1}\;n^{r/2+1}}\Big((s\;-\;1)\,\Gamma(r/2{+}2) \;+\; 2\int_0^\infty e^{-t}\,t^{r/2+1}\,L(sn/t)\,dt\Big) + o(n^{-r/2-1}),$$

$$\mathsf{E}\,N_n^{-r/2-1}(s) \;=\; \frac{(r+2)\Gamma(r/2+1)}{2s^{r/2+1}n^{r/2+1}} \;+\; o\big(n^{-r/2-1}\big).$$

2. Асимптотические дефекты оценок, основанных на выборках случайного объема

В этом разделе иллюстрируется применение Замечания 3.2. Напомним, что рассматривается оценивание действительной известной функции $g(\theta)$, зависящей от неизвестного параметра $\theta \in \Theta$, с помощью оценок $\delta_n(X_1, \ldots, X_n)$, основанной на выборке неслучайного объема $X_1, \ldots, X_n$ и оценки $\delta_{N_n}(X_1, \ldots, X_{N_n})$, основанных на выборке случайного объема $X_1, \ldots, X_{N_n}$. Качество оценок сравнивается с помощью функций риска вида

$$R_n(\theta) = \mathsf{E}_\theta \left|\delta_{N_n}(X_1, \ldots, X_{N_n}) - g(\theta)\right|^r,$$

$$R_n^*(\theta) = \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) - g(\theta)\right|^r, \quad r > 0, \tag{3.1}$$

то есть это риски оценок, основанных на выборках случайного и неслучайного объемов. Непосредственно из Замечания 2.3 следует

Теорема 3.1. *Пусть выполнены условия Замечания 2.3, $k = 2$ и существуют числа k_{10}, k_{11}, k_{12} k_{21}, k_{22}, k_{31} такие, что справедливы соотношения*

$$\mathsf{E}\, N_n^{-r\alpha} = \sum_{i=0}^{2} \frac{k_{1i}}{n^{r\alpha+i/2}} + o\big(n^{-r\alpha-1}\big),$$

$$\mathsf{E}\, N_n^{-r\alpha-1/2} = \sum_{i=1}^{2} \frac{k_{2i}}{n^{r\alpha+i/2}} + o\big(n^{-r\alpha-1}\big),$$

$$\mathsf{E}\, N_n^{-r\alpha-1} = \frac{k_{31}}{n^{r\alpha+1}} + o\big(n^{-r\alpha-1}\big),$$

тогда

$$R_n^*(\theta) = \mathsf{E}_\theta \left|\delta_n(X_1, \ldots, X_n) - g(\theta)\right|^r =$$

$$= rn^{-r\alpha} \int_0^\infty x^{r-1} \big(1 - G(x) + G(-x)\big)\, dx +$$

$$+ rn^{-r\alpha} \sum_{i=1}^{2} n^{-i/2} \int_0^\infty x^{r-1} \big(g_i(-x) - g_i(x)\big)\, dx + o\big(n^{-r\alpha-1}\big),$$

$$R_n(\theta) = \mathsf{E}_\theta \left|\delta_{N_n}(X_1, \ldots, X_{N_n}) - g(\theta)\right|^r =$$

$$= \sum_{i=0}^{2} \frac{u_i}{n^{r\alpha+i/2}} + o\big(n^{-r\alpha-1}\big),$$

где

$$u_0 = r\, k_{10} \int_0^\infty x^{r-1} \big(1 - G(x) + G(-x)\big)\, dx,$$

$$u_1 = r\, k_{11} \int_0^\infty x^{r-1} \big(1 - G(x) + G(-x)\big)\, dx +$$

$$+ r\, k_{21} \int_0^\infty x^{r-1} \big(g_1(-x) - g_1(x)\big)\, dx,$$

$$u_2 = r\,k_{12}\int_0^\infty x^{r-1}\bigl(1 - G(x) + G(-x)\bigr)\,dx +$$

$$+\,r\,k_{22}\int_0^\infty x^{r-1}\bigl(g_1(-x) - g_1(x)\bigr)\,dx + r\,k_{31}\int_0^\infty x^{r-1}\bigl(g_2(-x) - g_2(x)\bigr)\,dx.$$

Применим эту Теорему для нахождения асимптотического дефекта оценки $\delta_{N_n}(X_1,\ldots,X_{N_n})$ относительно оценки $\delta_n(X_1,\ldots,X_n)$ при условии, что

$$\mathsf{E}N_n = n, \tag{3.2}$$

то есть средний объем выборки случайного объема равен n. При нахождении дефекта оценки, основанной на выборке случайного объема $N_{m(n)}$, для которого согласно соотношению (3.2)

$$\mathsf{E}N_{m(n)} = m(n)$$

и соответствующей оценки, основанной на выборке неслучайного объема n, фактически сравнивается средний объем случайной выборки $m(n)$ и n с помощью величины $d_n = m(n) - n$ и ее предела.

Из Теоремы 3.1 и формулы (1.6) непосредственно следует

Теорема 3.2. *Пусть выполнены условия Теоремы 3.1, $g_1(-x) \equiv g_1(x)$, $u_1 = 0$ и $k_{10} = 1$, тогда для асимптотического дефекта оценки $\delta_{N_n}(X_1,\ldots,X_{N_n})$ относительно оценки $\delta_n(X_1,\ldots,X_n)$ справедливо равенство*

$$d = \frac{k_{12}}{r\alpha} + \frac{(k_{31} - 1)\int_0^\infty x^{r-1}\bigl(g_2(-x) - g_2(x)\bigr)\,dx}{r\alpha\int_0^\infty x^{r-1}\bigl(1 - G(x) - G(-x)\bigr)\,dx}.$$

Следствие 3.2. *Пусть выполнены условия Следствия 2.1 при $k = 2$ и Теоремы 3.2, тогда для асимптотического дефекта оценки (2.4) справедливо равенство*

$$d = \frac{2k_{12}}{r} + \frac{4(k_{31} - 1)\int_0^\infty x^{r-1}Q_2(x)\,dx}{\mathsf{E}\,|\eta|^r}.$$

Приведем конкретное применение этой Теоремы и ее Следствия.

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

$$N_n :\ \begin{matrix} n - h_n, & n, & n + h_n \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3}, \end{matrix} \tag{3.3}$$

где последовательность натуральных чисел $h_n < n$ удовлетворяет условию

$$\frac{h_n}{n} = \alpha_0 + \frac{\alpha_1}{\sqrt{n}} + \frac{\alpha_2}{n} + o\bigl(n^{-1}\bigr), \tag{3.4}$$

где $\alpha_0 \in [0,1)$, α_1, α_2 — некоторые числа.

Лемма 3.3. *Пусть случайная величина N_n имеет распределение (3.3) и выполнено условие (3.4), тогда для любого $p > 0$ справедливы равенства*

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

$$\mathsf{E}\; N_n^{-p} \;=\; \frac{1}{3n^p}\,\Big(1 \;+\; \frac{(1+\alpha_0)^p + (1-\alpha_0)^p}{(1-\alpha_0^2)^p} \;+$$
$$+\; \frac{p\alpha_1\;\big((1+\alpha_0)^{p+1} - (1-\alpha_0)^{p+1}\big)}{\sqrt{n}\;(1-\alpha_0^2)^{p+1}} \;+$$
$$+\; \frac{p(p+1)\alpha_1^2\;\big((1-\alpha_0)^{p+2} - (1+\alpha_0)^{p+2}\big)}{2\;n\;(1-\alpha_0^2)^{p+2}} \;+$$
$$+\; \frac{p\alpha_2\;\big((1+\alpha_0)^{p+1} - (1-\alpha_0)^{p+1}\big)}{n\;(1-\alpha_0^2)^{p+1}}\Big) \;+\; o\big(n^{-p-1}\big).$$

Доказательство. Доказательство Леммы непосредственно следует из равенств

$$\mathsf{E}\; N_n^{-p} \;=\; \frac{1}{3n^p}\,\Big(1 \;+\; \frac{1}{\big(1 \;+\; h_n/n\big)^p} \;+\; \frac{1}{\big(1 \;-\; h_n/n\big)^p}\Big),$$

и

$$\frac{1}{(1\;\pm\;x)^p} \;=\; \frac{1}{(1\;\pm\;\alpha_0)^p} \;\mp\; \frac{p\;(x\;-\;\alpha_0)}{(1\;\pm\;\alpha_0)^{p+1}} \;\pm\; \frac{p(p+1)\;(x\;-\;\alpha_0)^2}{2(1\;\pm\;\alpha_0)^{p+2}} \;+\; O\big((x\;-\;\alpha_0)^3\big).$$

Лемма доказана. □

Из симметрии распределения (3.3) случайного индекса N_n следует

Теорема 3.4. *Пусть выполнены условия Следствия 3.2, случайный индекс N_n имеет распределение (3.3) и выполнено условие (3.4) с $\alpha_0 = \alpha_1 = 0$, тогда асимптотический дефект оценки (2.4) равен нулю, то есть $d = 0$.*

Рассмотрим теперь наряду с оценкой (2.4) оценку вида

$$\delta_{n,\gamma} \;=\; \frac{1}{n+\gamma}\;(X_1 \;+\; \ldots \;+\; X_n) \;=\; \frac{n}{n+\gamma}\;\delta_n \quad \gamma \;\in\; \mathbb{R}. \tag{3.5}$$

Заметим, что все последующие результаты непосредственно переносятся на оценки более общего вида

$$\delta_{n,\gamma_1,\gamma_2} \;=\; \frac{1}{n+\gamma_1}\;\sum_{j=1}^{n+\gamma_2} X_j, \quad \gamma_1,\gamma_2 \;\in\; \mathbb{R},$$

которые имеют более наглядный вид в смысле добавочного числа наблюдений. Тогда с учетом равенства

$$\mathsf{P}\big(\sigma^{-1}\sqrt{n}\;(\delta_{n,\gamma}\;-\;\theta)\;<\;x\big) \;=\; \mathsf{P}\big(\sigma^{-1}\sqrt{n}\;(\delta_n\;-\;\theta)\;<\;(n+\gamma)x/n \;+\; \gamma\theta/\sigma\sqrt{n}\big)$$

при выполнении условий (2.3), (2.5) при $k = 2$ справедливо соотношение (см. (2.7))

$$\mathsf{P}\big(\sigma^{-1}\sqrt{n}\;(\delta_{n,\gamma}\;-\;\theta)\;<\;x\big) \;=\; \Phi(y) \;+\; \frac{1}{\sqrt{n}}\;Q_1(y) \;+\; \frac{1}{n}\;Q_2(y) \;+\; o(n^{-1}),$$

где

$$y \;=\; \frac{x(n+\gamma)}{n} \;+\; \frac{\gamma\theta}{\sigma\sqrt{n}}.$$

Поэтому

$$\mathsf{P}\big(\sigma^{-1}\sqrt{n}\ (\delta_n(\gamma)\ -\ \theta)\ <\ x\big)\ =$$

$$=\ \Phi(x)\ +\ \frac{1}{\sqrt{n}}\ \Big(Q_1(x)\ +\ \gamma\theta\sigma^{-1}\ \phi(x)\Big)\ +$$

$$+\ \frac{1}{n}\ \Big(Q_2(x)\ +\ \gamma x\ \phi(x)\ +\ \gamma\theta\sigma^{-1}\ Q_1'(x)\ +\ 2^{-1}\gamma^2\theta^2\sigma^{-2}\ \phi'(x)\Big)\ +\ o(n^{-1}). \quad (3.6)$$

Непосредственно из Следствия 2.1 вытекает

Теорема 3.5. *Пусть выполнены условия (2.3) и (2.5) при $k = 2$, тогда для любого $0 < r < 4$ для функций риска $R_n^*(\theta)$ и $R_{n,\gamma}^*(\theta)$ оценок (2.4) и (3.5) справедливы асимптотические разложения вида*

$$R_n^*(\theta)\ =\ \mathsf{E}_\theta\ \big|\delta_n(X_1,\ldots,X_n)\ -\ \theta\big|^r\ =$$

$$=\ \sigma^r n^{-r/2}\ \Big(\mathsf{E}\ |\eta|^r\ -\ 2r\ n^{-1}\ \int_0^\infty\ x^{r-1}\ Q_2(x)\ dx\Big)\ +\ o(n^{-r/2-1}),$$

$$R_{n,\gamma}^*(\theta)\ =\ \mathsf{E}_\theta\ \big|\delta_{n,\gamma}(X_1,\ldots,X_n)\ -\ \theta\big|^r\ =\ \sigma^r n^{-r/2}\ \Big(\mathsf{E}\ |\eta|^r\ -$$

$$-\ 2r\ n^{-1}\ \int_0^\infty\ x^{r-1}\ \Big(Q_2(x)\ +\ \gamma x\ \phi(x)\ +\ \gamma\theta\sigma^{-1}\ Q_1'(x)\ +\ 2^{-1}\gamma^2\theta^2\sigma^{-2}\ \phi'(x)\Big)\ dx\Big)\ +$$

$$+o(n^{-r/2-1}).$$

Асимптотический дефект оценки $\delta_{n,\gamma}(X_1,\ldots,X_n)$ относительно оценки $\delta_n(X_1,\ldots,X_n)$ равен

$$d\ =\ -\frac{4\gamma\ \int_0^\infty\ x^{r-1}\ \Big(x\ \phi(x)\ +\ \theta\sigma^{-1}\ Q_1'(x)\ +\ 2^{-1}\gamma\theta^2\sigma^{-2}\ \phi'(x)\Big)\ dx}{\sigma^r\ \mathsf{E}\ |\eta|^r}.$$

В частности, если

$$\mathsf{E}\ (X_1\ -\ \theta)^3\ =\ 0,$$

то

$$d\ =\ \frac{\gamma\ (\gamma\theta^2\ -\ 2\sigma^2)}{\sigma^{r+2}}.$$

Рассмотрим теперь выборки со случайным числом наблюдений N_n, которое имеет распределение (3.3).

Теорема 3.6. *Пусть выполнены условия (2.3) и (2.5) с $k = 2$ и случайный индекс имеет распределение (3.3), (3.4), тогда для любого $0\ <\ r\ <\ 4$ для функций риска $R_n(\theta)$ и $R_{n,\gamma}(\theta)$ оценок (2.4) и (3.5), построеным по выборкам случайного объема N_n, справедливы асимптотические разложения вида*

$$R_n(\theta)\ =\ \mathsf{E}_\theta\ \big|\delta_{N_n}(X_1,\ldots,X_{N_n})\ -\ \theta\big|^r\ =$$

$$=\ \frac{h_1}{n^{r/2}}\ +\ \frac{h_2}{n^{r/2+1/2}}\ +\ \frac{h_3}{n^{r/2+1}}\ +\ o(n^{-r/2-1}),$$

где

$$h_1 \;=\; \frac{1}{3}\;\sigma^r\;\mathsf{E}\;|\eta|^r\;\Big(1\;+\;\frac{(1+\alpha_0)^{r/2}+(1-\alpha_0)^{r/2}}{(1-\alpha_0^2)^{r/2}}\Big),$$

$$h_2 \;=\; \frac{1}{6}\;\sigma^r\;\mathsf{E}\;|\eta|^r\;\frac{r\alpha_1\;\big((1+\alpha_0)^{r/2+1}-(1-\alpha_0)^{r/2+1}\big)}{(1-\alpha_0^2)^{r/2+1}},$$

$$h_3 \;=\; \frac{1}{6}\;\sigma^r\;\mathsf{E}\;|\eta|^r\;\Big(\frac{r(r/2+1)\alpha_1^2\;\big((1-\alpha_0)^{r/2+2}-(1+\alpha_0)^{r/2+2}\big)}{2(1-\alpha_0^2)^{r/2+2}}\;+$$

$$+\;\frac{r\alpha_2\;\big((1+\alpha_0)^{r/2+1}-(1-\alpha_0)^{r/2+1}\big)}{(1-\alpha_0^2)^{r/2+1}}\Big)\;-$$

$$-\;\frac{2r\sigma^r}{3}\;\int_0^\infty\;x^{r-1}\;Q_2(x)\;dx\;\Big(1\;+\;\frac{(1+\alpha_0)^{r/2+1}+(1-\alpha_0)^{r/2+1}}{(1-\alpha_0^2)^{r/2+1}}\Big),$$

$$R_{n,\gamma}(\theta)\;=\;\mathsf{E}_\theta\;\big|\delta_{N_n,\gamma}(X_1,\ldots,X_{N_n})\;-\;\theta\big|^r\;=$$

$$=\;\frac{h_1}{n^{r/2}}\;+\;\frac{h_2}{n^{r/2+1/2}}\;+\;\frac{h_{3,\gamma}}{n^{r/2+1}}\;+\;o(n^{-r/2-1}),$$

где

$$h_{3,\gamma}=h_3-\frac{2r\sigma^r}{3}\int_0^\infty\;x^{r-1}\;\Big(\gamma x\;\phi(x)\;+\;\gamma\theta\sigma^{-1}\;Q_1'(x)\;+\;2^{-1}\gamma^2\theta^2\sigma^{-2}\;\phi'(x)\Big)\;dx\times$$

$$\times\Big(1\;+\;\frac{(1+\alpha_0)^{r/2+1}+(1-\alpha_0)^{r/2+1}}{(1-\alpha_0^2)^{r/2+1}}\Big),$$

Если $\alpha_1\;=\;0$, *то асимптотический дефект оценки* $\delta_{N_n,\gamma}(X_1,\ldots,X_{N_n})$ *относительно оценки* $\delta_{N_n}(X_1,\ldots,X_{N_n})$ *равен*

$$d\;=\;\frac{2\big(h_{3,\gamma}\;-\;h_3\big)}{r\;h_1}\;=$$

$$=\;-\frac{4\gamma\;\int_0^\infty\;x^{r-1}\;\Big(x\;\phi(x)\;+\;\theta\sigma^{-1}\;Q_1'(x)\;+\;2^{-1}\gamma\theta^2\sigma^{-2}\;\phi'(x)\Big)\;dx}{\sigma^r\;\mathsf{E}\;|\eta|^r}\;\times$$

$$\times\;\Big(1\;+\;\frac{(1+\alpha_0)^{r/2+1}+(1-\alpha_0)^{r/2+1}}{(1-\alpha_0^2)^{r/2+1}}\Big)\;\Big(1\;+\;\frac{(1+\alpha_0)^{r/2}+(1-\alpha_0)^{r/2}}{(1-\alpha_0^2)^{r/2}}\Big)^{-1}.$$

В частности, если

$$\mathsf{E}\;(X_1\;-\;\theta)^3\;=\;0,$$

то

$$d\;=\;\frac{\gamma\;(\gamma\theta^2\;-\;2\sigma^2)}{\sigma^{r+2}}\;\Big(1\;+\;\frac{(1+\alpha_0)^{r/2+1}+(1-\alpha_0)^{r/2+1}}{(1-\alpha_0^2)^{r/2+1}}\Big)\times$$

$$\times\Big(1\;+\;\frac{(1+\alpha_0)^{r/2}+(1-\alpha_0)^{r/2}}{(1-\alpha_0^2)^{r/2}}\Big)^{-1}.$$

Заключение

Таким образом, в работе рассмотрены проблемы точечного оценивания, касающиеся неизвестных параметров распределения. При этом предполагается, что статистики, используемые в качестве статистических оценок, основаны на выборках случайного объема. Получены асимптотические разложения для функций риска таких оценок. Рассмотрены конкретные распределения случайного индекса, возникающие в случаях предельного распределения Стьюдента и Лапласа. С помощью понятия дефекта проведено асимптотическое сравнение качества оценок, основанных на выборках случайного объема. Получены явные формулы для асимптотического дефекта, имеющего смысл необходимого добавочного числа наблюдений. В качестве иллюстрации рассмотрено также трехточечное распределение случайного индекса.

Список литературы

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

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

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

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

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

[6] Бенинг В.Е., Королев В.Ю., Савушкин В.А. О сравнении качества оценок, основанных на выборках случайного объема, с помощью понятия дефект // Статистические методы оценивания и проверки гипотез: межвуз. сб. 2015. С. 26–42.

[7] Петров В.В. Суммы независимых случайных величин. М.: Наука, 1972. 414 с.

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

[9] 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

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

[11] Кристоф Г., Монахов М.М., Ульянов В.В. Разложения Чебышева - Эджворта и Корниша - Фишера второго порядка для распределений статистик, построенных по выборкам случайного размера // Записки научных семинаров Санкт-Петербургского отделения математического института им. В.А. Стеклова РАН. 2017. Т. 466. С. 167–207.

Образец цитирования

Бенинг В.Е. Асимптотические разложения функции риска и дефекты оценок, основанных на выборках случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2019. № 2. С. 5–25. https://doi.org/10.26456/vtpmk530

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 THE ASYMPTOTIC EXPANSIONS FOR THE RISK FUNCTION AND DEFICIENCIES OF SOME STATISTICAL ESTIMATORS BASED ON THE SAMPLES WITH RANDOM SIZES

Bening Vladimir Evgenyevich
Professor in the Department of Mathematical Statistics,
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 14.04.2019, revised 20.05.2019.

Statistical regularities of the information flows in contemporary communication, computational and other information systems are characterized by the presence of the so-called "heavy tails". The random character of the intensity of the flow of informative events results in that the available sample size (traditionally this is the number of observations registered within a certain time interval) is random. The randomness of the sample size crucially changes the asymptotic properties of the statistical procedures (e.g., estimators). The present paper consists of a number of applications of the deficiency concept, i.e., the number of additional observations required by the less effective procedure and thereby provides a basis for deciding whether or not the price is too high. The deficiency was introduced and initiated in its study by Hodges and Lehmann in 1970 [1]. In this paper asymptotic deficiencies of statistical estimators based on the samples with random sizes are considered. Asymptotic expansions for the risk function of some estimators based on the samples with random sizes are presented.

Keywords: statistical estimator, asymptotic deficiency, sample with random size, risk function, asymptotic expansion.

Citation

Bening V.E., "On the asymptotic expansions for the risk function and deficiencies of some statistical estimators based on the samples with random sizes", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2019, № 2, 5–25 (in Russian). https://doi.org/10.26456/vtpmk530

References

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

[2] Cramer H., *Mathematical Method of Statistics*, Princeton University Press, Princeton, NJ, 1946, 656 pp.

[3] Lehmann E.L., Casella G., *Theory of Point Estimation*, 2nd edition, Springer-Verlag, New York; FizMatLit, 1998, 470 pp.

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

[5] 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).

[6] Bening V.E., Korolev V.Yu., Savushkin V.A., “On the comparison of statistical estimators based on samples with random sizes with the help of deficiency concept”, *Statisticheskie Metody Otsenivaniya i Proverki Gipotez: Mezhvuz. Sb. [Interuniversity Transactions on Statistical Method of Estimation and Testing Hypotheses]*, 2015, 26–42 (in Russian).

[7] Petrov V.V., *Sums of Independent Random Variables*, Springer-Verlag, Berlin, 1975, 414 pp.

[8] 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).

[9] 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.

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

[11] Christoph G., Monakhov M.M., Ulyanov V.V., “Second order Chebyshev - Edgeworth and Cornish - Fisher expansions for distributions of statistics constructed from samples with random sizes”, *Zapiski nauchnykh seminarov Sankt-Peterburgskogo otdeleniya matematicheskogo instituta im. V.A. Steklova RAN [Journal of Mathematical Sciences]*, **466** (2017), 167–207 (in Russian).