ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

О ДЕФЕКТЕ ВЫБОРОЧНОЙ МЕДИАНЫ В СЛУЧАЕ ВЫБОРОК СЛУЧАЙНОГО ОБЪЕМА[1]

Бенинг В.Е.[*,**], **Савушкин В.А.**[***]
[*]МГУ имени М.В. Ломоносова, г. Москва
[**]Институт проблем информатики РАН, г. Москва
[***]Международный университет природы, общества и человека «Дубна», г. Дубна

Поступила в редакцию 28.04.2016, после переработки 15.05.2016.

В работе доказаны теоремы, позволяющие находить асимптотический дефект выборочной медианы, основанной на выборках случайного объема. Это делает возможным сравнивать в терминах добавочного числа наблюдений качество выборочной медианы, основанной на выборках случайного и неслучайного объемов. Рассмотрены случаи биномиального и распределения Пуассона.

Ключевые слова: выборочная медиана, асимптотический дефект, выборка случайного объема, биномиальное и распределение Пуассона.

Вестник ТвГУ. Серия: Прикладная математика. 2016. № 2. С. 5–30.

1. Введение

Рассмотрим задачу статистического оценивания известной параметрической функции $g(\theta)$, зависящей от неизвестного параметра θ и обозначим через $m(n)$ необходимое число наблюдений, которое требуется оценке $\delta_{m(n)}(X_1, \dots, X_{m(n)})$ для достижения такого же качества (например, среднеквадратичного отклонения или дисперсии), что и «лучшей» оценке $\delta_n^*(X_1, \dots, X_n)$, основанной на n наблюдениях $X_1, \dots, X_n$. Мы рассматриваем асимптотический подход, означающий, что $n \to \infty$. Под асимптотической относительной эффективностью (АОЭ) оценки $\delta_n(X_1, \dots, X_n)$ по отношению к оценке $\delta_n^*(X_1, \dots, X_n)$ понимается предел (в случае его существования и независимости от последовательности $m(n)$) вида (см., например, [3], стр. 305)

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

Предположим, что $e = 1/3$. Тогда при больших значениях числа наблюдений n величина $m(n)$ приближенно равна $3n$, поэтому оценка $\delta_n(X_1, \dots, X_n)$ для достижения такого же качества, что и оценка $\delta_n^*(X_1, \dots, X_n)$, требует примерно в три раза больше наблюдений.

Вместо отношения необходимого числа наблюдений, естественно, можно было бы рассматривать разность вида $m(n) - n$, которая тоже имеет наглядный смысл необходимого дополнительного числа наблюдений, требующихся 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 $\lim_{n\to\infty} d_n$ 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.

Предположим, например, что $d = 5$. Тогда для больших значений n величина $m(n)$ равна приближенно $n+5$. Чтобы получить ту же величину критерия качества оценке δ_n требуется примерно на пять наблюдений больше, чем оценке δ_n^*.

Таким образом, дефект оценки δ_n относительно оценки δ_n^* показывает, сколько добавочных наблюдений требуется, если мы настаиваем на использовании оценки δ_n вместо оценки δ_n^*, и поэтому создает естественный базис для их асимптотического сравнения в случае $e = 1$. Исследование асимптотического поведения дефекта d_n технически более сложно, чем нахождение предела e. Часто оно требует построения асимптотических разложений (а.р.) для соответствующих функций, характеризующих качество оценок (см., например, книгу [4]).

Обозначим функции риска оценок δ_n и δ_n^*, соответственно, через

$$R_n(\theta) = \mathsf{E}_\theta \left(\delta_n(X_1, \ldots, X_n) - g(\theta)\right)^2, \quad R_n^*(\theta) = \mathsf{E}_\theta \left(\delta_n^*(X_1, \ldots, X_n) - g(\theta)\right)^2,$$

где $g(\theta)$ – оцениваемая функция, а 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

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

Асимптотический дефект обладает следующим очевидным свойством транзитивности: если дана третья оценка $\bar{\delta}_n$, имеющая функцию риска $\bar{R}_n(\theta)$, которая имеет а.р. типа (1.4), тогда дефект d оценки $\bar{\delta}_n$ относительно оценки δ_n^* удовлетворяет равенству

$$d \;=\; d_1 \;+\; d_2,$$

где d_1 – дефект оценки $\bar{\delta}_n$ относительно оценки δ_n и d_2 – дефект оценки δ_n относительно оценки δ_n^*.

Случай, когда выполняется равенство $s \;=\; 1$, представляется наиболее интересным, поскольку тогда асимптотический дефект конечен. Ходжес и Леман в работе [1] привели ряд простых примеров, показывающих естественность возникновения этого случая в математической статистике (см., также книгу [4]).

Совершенно аналогично определяется дефект и в общем случае асимптотического сравнения двух статистических процедур, соответственно с мерами качества π_n и π_n^*. При этом необходимое число наблюдений $m(n)$ для первой процедуры относительно «оптимальной» второй процедуры определяется из равенства (считая $m(n)$ непрерывной переменной)

$$\pi_{m(n)} \;=\; \pi_n^*,$$

а предел вида (в случае его существования)

$$d \;=\; \lim_{n\to\infty}\big(m(n)\;-\;n\big)$$

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 данной работе получены формулы для асимптотического дефекта выборочной медианы, построенной по выборкам случайного объема. Рассмотрены случаи распределения Лапласа, Стьюдента и Коши. Эти результаты продолжают исследования, начатые в работах [1, 5, 7, 9–15].

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

2. Асимптотические свойства выборочной медианы

Пусть $X_1, X_2, \dots X_n$ – независимые одинаково распределенные наблюдения с функцией распределения $F(x-\theta)$ и плотностью $p(x-\theta)$, где θ – неизвестный параметр сдвига, подлежащий оцениванию на основе выборки $X_1, X_2, \dots X_n$. Обозначим через $X_{(1)} \leqslant X_{(2)} \leqslant \dots \leqslant X_{(n)}$ – вариационный ряд, построенный по исходным наблюдениям $X_1, X_2, \dots X_n$ и через M_n – выборочную медину (см., например, [2], [3], [6], [7], [8]), то есть оценку вида

$$M_n = \begin{cases} X_{(m+1)}, & n = 2m+1, \\ \dfrac{X_{(m)} + X_{(m+1)}}{2}, & n = 2m. \end{cases} \tag{2.1}$$

Асимптотические свойства первого порядка выборочной медианы M_n хорошо известны (см., например, книгу [2], Теорема 5.3.2, стр. 313 или книгу [8], стр. 81)

$$\sqrt{n}\,\big(M_n - \theta\big) \Longrightarrow \mathcal{N}\Big(0,\ \frac{1}{4p^2(0)}\Big), \tag{2.2}$$

$$\mathsf{E}_\theta\big(M_n - \theta\big)^2 = \frac{1}{4np^2(0)} + o\big(n^{-1}\big),\ \ F(0) = 1/2,\ \ p(0) > 0,\quad n \to \infty. \tag{2.3}$$

Асимптотические свойства выборочной медианы второго порядка изучались в работе [6]. Сформулируем основные результаты этой работы. С этой целью приведем условия регулярности на плотность $p(x)$ из работы [6]:

A1. Плотность $p(x)$ симметрична относительно нуля, то есть $p(-x) = p(x)$, $x \in \mathbb{R}$ и $p(0) > 0$.

A2. Плотность $p(x)$ имеет три непрерывные ограниченные производные в некоторой окрестности нуля вида $(0,\ \delta)$, $\delta > 0$.

A3. Существуют постоянные $C > 0$ и $\alpha > 0$ такие, что функция распределения $F(x)$ удовлетворяет неравенству

$$1 - F(x) \leqslant C\,x^{-\alpha},\ \ x > 0.$$

Заметим, что эти условия регулярности выполняются, например, для распределения Коши

$$p(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R} \tag{2.4}$$

и распределения Лапласа

$$p(x) = \frac{1}{2}\, e^{-|x|}, \quad x \in \mathbb{R}. \tag{2.5}$$

В случае распределения Лапласа выборочная медиана M_n совпадает с оценкой максимального правдоподобия параметра θ (см., например, [6]).

Далее используем следующие обозначения

$$k = \big[n/2\big], \;\; p_0 = p(0) > 0, \; p_1 = p'(0+), \; p_2 = p''(0+),$$

где $[\cdot]$ – целая часть числа.

В следующей Теореме получены а.р. для ф.р. и среднеквадратичного отклонения выборочной медианы (см. формулы (2.6) – (2.11)).

Теорема 2.1 ([6]).

1. Пусть плотность $p(x)$ удовлетворяет условиям регулярности A1 и A2, тогда равномерно по $x \in \mathbb{R}$ справедливо равенство

$$\mathsf{P}_\theta\big(2p_0\sqrt{2k}(M_n - \theta) < x\big) = \Phi(x) +$$

$$+ \varphi(x)\, \frac{p_1 x|x|\sqrt{2}}{8p_0\sqrt{k}} + \varphi(x)\, \frac{x}{8k}\Big(3 + x^2 + \frac{x^2 p_2}{6p_0^3} - \frac{x^4 p_1^2}{8p_0^4}\Big) + O(n^{-3/2}).$$

2. Если выполнены условия регулярности A1 – A3, то для среднеквадратичного отклонения выборочной медианы M_n справедливо асимптотическое разложение

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{8p_0^2 k} - \frac{p_1}{8p_0^4\sqrt{\pi}k^{3/2}} - \frac{1}{16p_0^2k^2}\Big(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\Big) + O(n^{-5/2}).$$

Следствие 2.1.

1. Для распределения Лапласа (2.5) справедливы асимптотические разложения

$$\mathsf{P}_\theta\big(\sqrt{2k}(M_n - \theta) < x\big) = \Phi(x) -$$

$$- \varphi(x)\, \frac{x|x|\sqrt{2}}{4\sqrt{k}} + \varphi(x)\, \frac{x}{48k}\Big(18 + 10x^2 - 3x^4\Big) + O(n^{-3/2}),$$

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{2k} + \frac{1}{\sqrt{\pi}k^{3/2}} - \frac{1}{16k^2} + O(n^{-5/2}).$$

2. Для распределения Коши (2.4) справедливы асимптотические разложения

$$\mathsf{P}_\theta\Big(\frac{2\sqrt{2k}}{\pi}(M_n - \theta) < x\Big) = \Phi(x) + \varphi(x)\, \frac{x}{24k}\Big(9 + x^2(3 - \pi^3)\Big) + O(n^{-3/2}),$$

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{\pi^2}{8k} + \frac{\pi^2(\pi^2 - 6)}{32k^2} + O(n^{-5/2}).$$

Нетрудно видеть, что если $k = [n/2]$, то справедливы следующие асимптотические формулы

$$\frac{1}{\sqrt{k}} = \frac{\sqrt{2}}{\sqrt{n}} + O(n^{-3/2}), \quad \frac{1}{k} = \frac{2}{\sqrt{n}} + \frac{1 - (-1)^n}{n^2} + O(n^{-3}),$$

$$\frac{1}{k^{3/2}} = \frac{2^{3/2}}{n^{3/2}} + O(n^{-5/2}), \quad \frac{1}{k^2} = \frac{4}{n^2} + O(n^{-3}).$$

С учетом этих формул основные утверждения Теоремы 2.1 и ее Следствия можно переписать в виде

$$\mathsf{P}_\theta\big(2p_0\sqrt{2k}(M_n - \theta) < x\big) = \Phi(x) +$$

$$+ \varphi(x)\,\frac{p_1 x|x|\sqrt{2}}{4p_0\sqrt{n}} + \varphi(x)\,\frac{x}{4n}\Big(3 + x^2 + \frac{x^2 p_2}{6p_0^3} - \frac{x^4 p_1^2}{8p_0^4}\Big) + O(n^{-3/2}), \tag{2.6}$$

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} +$$

$$+ \frac{1}{4p_0^2 n^2}\,\Big(\frac{1 - (-1)^n}{2} - 3 - \frac{p_2}{4p_0^3} + \frac{15p_1^2}{16p_0^4}\Big) + O(n^{-5/2}). \tag{2.7}$$

Для распределения Лапласа (2.5) справедливы равенства

$$\mathsf{P}_\theta\big(\sqrt{2k}(M_n - \theta) < x\big) = \Phi(x) -$$

$$- \varphi(x)\,\frac{x|x|}{2\sqrt{n}} + \varphi(x)\,\frac{x}{24n}\Big(18 + 10x^2 - 3x^4\Big) + O(n^{-3/2}), \tag{2.8}$$

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} + \frac{1}{2n^2}\Big((1 - (-1)^n) - \frac{1}{2}\Big) + O(n^{-5/2}), \tag{2.9}$$

а для распределения Коши (2.4)

$$\mathsf{P}_\theta\Big(\frac{2\sqrt{2k}}{\pi}(M_n - \theta) < x\Big) = \Phi(x) + \varphi(x)\,\frac{x}{12n}\Big(9 + x^2(3 - \pi^3)\Big) + O(n^{-3/2}), \tag{2.10}$$

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2\big((1 - (-1)^n) + \pi^2 - 6\big)}{8n^2} + O(n^{-5/2}). \tag{2.11}$$

Из этих формул (см. (2.7), (2.9) и (2.11)) непосредственно получается следующее утверждение.

Лемма 2.1. *Пусть выполнены условия Теоремы 2.1, тогда*

1. Если $n = 2l,\ l \in \mathbb{N}$, то

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2 n^2}\,\Big(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\Big) + O(n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} - \frac{1}{4n^2} + O(n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2\big(\pi^2 - 6\big)}{8n^2} + O(n^{-5/2}).$$

2. Если $n = 2l + 1$, $l \in \mathbb{N}$, *то*

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2 n^2}\Big(2 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\Big) + O(n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} + \frac{1}{4n^2} + O(n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_n - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2\big(\pi^2 - 4\big)}{8n^2} + O(n^{-5/2}).$$

3. Оценки, основанные на выборках случайного объема

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

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

то

$$\mathsf{E}\, N_n = n, \tag{3.1}$$

то есть среднее значение случайного объема выборки равно n. При нахождении дефектов статистических процедур, основанных на выборках случайного объема $N_{m(n)}$, для которых выполнено соотношение (3.1), 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. $X_1, X_2, \ldots$. 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 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, указанной выше, будем понимать оценку действительной известной функции $g(\theta)$, зависящей от неизвестного параметра $\theta \in \Theta$ и будем обозначать ее символами типа $\delta_n(X_1, \ldots, X_n)$. В работе [5] доказана следующая теорема.

Теорема 3.1.

1. Если $\delta_n = \delta_n(X_1, \ldots, X_n)$ любая несмещенная оценка функции $g(\theta)$, то есть справедливо тождество

$$\mathsf{E}_\theta\, \delta_n \equiv g(\theta), \quad \theta \in \Theta$$

и $\delta_{N_n} \equiv \delta_{N_n}(X_1, \ldots, X_{N_n})$ – оценка, построенная по выборке случайного объема, то она также является несмещенной оценкой функции $g(\theta)$, то есть

$$\mathsf{E}_\theta\, \delta_{N_n} \equiv g(\theta), \quad \theta \in \Theta.$$

2. Предположим, что существуют числа $a(\theta)$, $b(\theta)$ и $C(\theta) > 0$, $\alpha > 0$, $r > 0$, $s > 0$ такие, что для функции риска оценки $\delta_n(X_1, \ldots, X_n)$ справедливо соотношение

$$\left| R_n^*(\theta) - \frac{a(\theta)}{n^r} - \frac{b(\theta)}{n^{r+s}} \right| \leqslant \frac{C(\theta)}{n^{r+s+\alpha}},$$

где

$$R_n^*(\theta) = \mathsf{E}_\theta \big(\delta_n(X_1, \ldots, X_n) - g(\theta)\big)^2,$$

тогда для функции риска оценки, построенной по выборке случайного объема, выполнено неравенство

$$\left| R_n(\theta) - a(\theta)\, \mathsf{E}\, N_n^{-r} - b(\theta)\, \mathsf{E}\, N_n^{-r-s} \right| \leqslant C(\theta)\, \mathsf{E}\, N_n^{-r-s-\alpha},$$

где

$$R_n(\theta) = \mathsf{E}_\theta \big(\delta_{N_n}(X_1, \ldots, X_{N_n}) - g(\theta)\big)^2.$$

Следствие 3.1. *Пусть существуют числа $a(\theta)$, $b(\theta)$ and $r > 0$, $s > 0$ такие, что*

$$R_n^*(\theta) = \frac{a(\theta)}{n^r} + \frac{b(\theta)}{n^{r+s}},$$

тогда

$$R_n(\theta) = a(\theta)\, \mathsf{E}\, N_n^{-r} + b(\theta)\, \mathsf{E}\, N_n^{-r-s}.$$

Теорема 3.2 ([9]). *Пусть существуют числа $a(\theta)$, $b(\theta)$, $\gamma > 0$ и k_1, k_2 такие, что справедливы соотношения*

$$R_n^*(\theta) = \mathsf{E}_\theta\big(\delta_n(X_1,\ldots,X_n) - g(\theta)\big)^2 =$$

$$= \frac{a(\theta)}{n} + \frac{b(\theta)}{n^2} + O\big(n^{-2-\gamma}\big)$$

и

$$\mathsf{E}\ N_n^{-1} = \frac{1}{n} + \frac{k_1}{n^2} + o\big(n^{-2}\big),$$

$$\mathsf{E}\ N_n^{-2} = \frac{k_2}{n^2} + o\big(n^{-2}\big),$$

$$\mathsf{E}\ N_n^{-2-\gamma} = o\big(n^{-2}\big),$$

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

$$d(\theta) = \frac{k_1\ a(\theta) + b(\theta)\ (k_2 - 1)}{a(\theta)}.$$

Теорема 3.3 ([9]). *Пусть существуют числа $a(\theta)$, $b(\theta)$, $\gamma > 0$ такие, что для функции риска оценки $\delta_n(X_1,\ldots,X_n)$ справедливо соотношение*

$$R_n^*(\theta) = \mathsf{E}_\theta\big(\delta_n(X_1,\ldots,X_n) - g(\theta)\big)^2 =$$

$$= \frac{a(\theta)}{n} + \frac{b(\theta)}{n^2} + O\big(n^{-2-\gamma}\big).$$

Пусть случайные величины N_{ni}, $i = 1,2$ принимают натуральные значения и не зависят от наблюдений X_1, $X_2,\ldots$ Предположим, что для некоторых чисел k_{1i}, k_{2i}, $i = 1,2$ справедливы равенства

$$\mathsf{E}\ N_{ni}^{-1} = \frac{1}{n} + \frac{k_{1i}}{n^2} + o\big(n^{-2}\big),$$

$$\mathsf{E}\ N_{ni}^{-2} = \frac{k_{2i}}{n^2} + o\big(n^{-2}\big),$$

$$\mathsf{E}\ N_{ni}^{-2-\gamma} = o\big(n^{-2}\big), \quad i = 1,2.$$

Тогда для асимптотического дефекта оценки $\delta_n^{(2)} \equiv \delta_{N_{n2}}(X_1,\ldots,X_{N_{n2}})$ относительно оценки $\delta_n^{(1)} \equiv \delta_{N_{n1}}(X_1,\ldots,X_{N_{n1}})$ справедливо равенство

$$d_{21}(\theta) = \frac{a(\theta)(k_{12} - k_{11}) + b(\theta)\ (k_{22} - k_{21})}{a(\theta)}.$$

Асимптотические разложения для функции распределения выборочной медианы M_n в случае выборок случайного объема были получены в работе [19].

4. Биномиальный случай

В этом разделе мы применим результаты предыдущих разделов к вычислению дефекта выборочной медианы, основанной на выборках, объем которых имеет биномиальное распределение.

Лемма 4.1. *Пусть неотрицательная целочисленная случайная величина N имеет вид $N = B + 1$, где случайная величина B имеет биномиальное распределение вида*

$$\mathsf{P}\Big(B = k\Big) = \binom{n}{k}\, p^k\, q^{n-k}, \quad k = 0, \ldots, n, \quad p \in (0,1), \quad q = 1 - p.$$

Тогда

$$\mathsf{E}\, N^{-1} = \frac{1 - q^{n+1}}{p(n+1)},$$

$$\mathsf{E}\, N^{-3/2} = \frac{1}{(np)^{3/2}} + O(n^{-5/2}),$$

$$\mathsf{E}\, N^{-2} = \frac{1}{p^2 n^2} + O(n^{-3}),$$

$$\mathsf{E}\, N^{-5/2} = O(n^{-3}),$$

$$\mathsf{E}\, N^{-3} = \frac{q^n}{n+1} \sum_{k=0}^{n} \left(\frac{p}{q}\right)^k \frac{1}{(k+1)^2}.$$

Доказательство. Первое и третье утверждения Леммы доказаны в работе [9] (см. Лемму 3.2 и Следствие 3.1). Для доказательства пятого утверждения применим Лемму 3.1 [9], согласно которой имеем

$$\mathsf{E}\, N^{-3} = \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} \frac{ds_2}{s_2} \int_0^{s_2} \frac{\Psi_N(s_3)}{s_3}\, ds_3,$$

где производящая функция случайной величины N имеет вид

$$\Psi_N(s) = \mathsf{E}\, s^N = s\,(ps + q)^n.$$

Таким образом, имеем

$$\mathsf{E}\, N^{-3} = \frac{1}{p(n+1)} \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} \frac{\big((ps_2 + q)^{n+1} - q^{n+1}\big)}{s_2}\, ds_2 =$$

$$= \frac{1}{n+1} \sum_{k=0}^{n} q^{n-k} \int_0^1 \frac{ds_1}{s_1} \int_0^{s_1} (ps_2)^k\, ds_2 =$$

$$= \frac{q^n}{n+1} \sum_{k=0}^{n} \left(\frac{p}{q}\right)^k \frac{1}{(k+1)^2}.$$

Далее, используя неравенство Иенсена, имеем

$$\mathsf{E}\ N^{-3/2} \ \geqslant \ \frac{1}{(\mathsf{E}\ N)^{3/2}} \ = \ \frac{1}{(np)^{3/2}},$$

с другой стороны неравенство Гельдера дает оценку сверху вида

$$\mathsf{E}\ N^{-3/2} \ \leqslant \ \Big(\mathsf{E}\ N^{-2}\Big)^{3/4} \ = \ \Big(\frac{1}{n^2p^2} \ + \ O(n^{-3})\Big)^{3/4}.$$

Аналогично с учетом неравенства Гельдера имеем

$$\mathsf{E}\ N^{-5/2} \ \leqslant \ \Big(\mathsf{E}\ N^{-3}\Big)^{5/6} \ = \ O(n^{-3}).$$

□

Из этой Леммы непосредственно получаем следующее утверждение.

Лемма 4.2. *Пусть случайная величина B_n имеет биномиальное распределение с параметрами $m(n-1)$, $n \geqslant 2$ и $p = 1/m$, где $m \geqslant 2$ – фиксированное натуральное число, тогда для случайной величины*

$$N_n \ = \ B_n \ + \ 1$$

справедливы равенства

$$\mathsf{E}\ N_n \ = \ n, \quad \mathsf{E}\ N_n^{-1} \ = \ \frac{1}{n} \ + \ \frac{m \ - \ 1}{mn^2} \ + \ O(n^{-3}),$$

$$\mathsf{E}\ N_n^{-3/2} \ = \ \frac{1}{n^{3/2}} \ + O(n^{-5/2}), \quad \mathsf{E}\ N_n^{-2} \ = \ \frac{1}{n^2} \ + O(n^{-3}),$$

$$\mathsf{E}\ N_n^{-5/2} \ = \ O(n^{-3}), \quad \mathsf{E}\ N_n^{-3} \ = \ O\Big(\frac{(1 \ - \ 1/m)^n}{n+1}\Big), \quad n \ \to \ \infty.$$

Из первой части Леммы 2.1 и Леммы 4.2 непосредственно получаем следующую

Теорема 4.1. *Пусть случайная величина B_n имеет биномиальное распределение с параметрами $m(n-2)$, $n \geqslant 3$ и $p = 1/2m$, где $m \geqslant 2$ – фиксированное натуральное число и*

$$N_n = 2(B_n + 1).$$

Предположим также, что выполнены Условия $A1 - A3$ и $n = 2l$, $l \in \mathbb{N}$, тогда

$$\mathsf{E}\ N_n \ = \ n, \quad \mathsf{E}\ N_n^{-1} \ = \ \frac{1}{n} \ + \ \frac{2m \ - \ 1}{mn^2} \ + \ O(n^{-3}),$$

$$\mathsf{E}\ N_n^{-3/2} \ = \ \frac{1}{n^{3/2}} \ + O(n^{-5/2}), \quad \mathsf{E}\ N_n^{-2} \ = \ \frac{1}{n^2} \ + O(n^{-3}),$$

$$\mathsf{E}\ N_n^{-5/2} \ = \ O(n^{-3}), \quad \mathsf{E}\ N_n^{-3} \ = \ O(n^{-3}),$$

$$\mathsf{E}_\theta\Big(M_{N_n} \ - \ \theta\Big)^2 \ =$$

$$= \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2n^2}\Big(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4} - \frac{2m-1}{m}\Big) + O(n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} - \frac{1}{4n^2}\Big(1 - \frac{4(2m-1)}{m}\Big) + O(n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2}{8n^2}\Big(\pi^2 - 6 + \frac{2(2m-1)}{m}\Big) + O(n^{-5/2}), \qquad n \to \infty.$$

Теорема 4.2. *Пусть выполнены условия Теоремы 4.1 и $n = 2l$, $l \in \mathbb{N}$, тогда если $p_1 \neq 0$, то дефект выборочной медианы M_{N_n} относительно M_n имеет вид*

$$d_n = o(\sqrt{n}), \quad n \to \infty,$$

если $p_1 = 0$, то

$$d = \lim_{n\to\infty} d_n = \frac{2m - 1}{m}.$$

В частности, для распределения Лапласа (2.5) справедливо соотношение

$$d_n = o(\sqrt{n}),$$

а для распределения Коши (2.4)

$$d = \frac{2m - 1}{m}.$$

Аналогично предыдущему можно получить следующие два утверждения.

Теорема 4.3. *Пусть случайная величина B_n имеет биномиальное распределение с параметрами $m(n-1)$, $n \geqslant 2$ и $p = 1/2m$, где $m \geqslant 2$ – фиксированное натуральное число и*

$$N_n = 2B_n + 1.$$

Предположим также, что выполнены Условия A1 − A3 и $n = 2l + 1$, $l \in \mathbb{N}$, тогда

$$\mathsf{E}\,N_n = n, \quad \mathsf{E}\,N_n^{-1} = \frac{1}{n} + \frac{m - 1}{mn^2} + O(n^{-3}),$$

$$\mathsf{E}\,N_n^{-3/2} = \frac{1}{n^{3/2}} + O(n^{-5/2}), \quad \mathsf{E}\,N_n^{-2} = \frac{1}{n^2} + O(n^{-3}),$$

$$\mathsf{E}\,N_n^{-5/2} = O(n^{-3}), \quad \mathsf{E}\,N_n^{-3} = O(n^{-3}),$$

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2n^2}\Big(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4} - \frac{m-1}{m}\Big) + O(n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} - \frac{1}{4n^2}\Big(1 - \frac{4(m-1)}{m}\Big) + O(n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2}{8n^2}\Big(\pi^2 - 6 + \frac{2(m-1)}{m}\Big) + O(n^{-5/2}), \qquad n \to \infty.$$

Теорема 4.4. *Пусть выполнены условия Теоремы 4.3 и* $n = 2l + 1$, $l \in \mathbb{N}$, *тогда если* $p_1 \neq 0$, *то дефект выборочной медианы* M_{N_n} *относительно* M_n *имеет вид*

$$d_n = o(\sqrt{n}), \quad n \to \infty,$$

если $p_1 = 0$, *то*

$$d = \lim_{n\to\infty} d_n = \frac{m - 1}{m}.$$

В частности для распределения Лапласа (2.5) справедливо соотношение

$$d_n = o(\sqrt{n}),$$

а для распределения Коши (2.4)

$$d = \frac{m - 1}{m}.$$

5. Трехточечное симметричное распределение

В этом разделе будет рассмотрен случай, когда случайный индекс N_n (случайный объем выборки) имеет симметричное распределение вида

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

где последовательность натуральных чисел $h_n < n$ удовлетворяет условию

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

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

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

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

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

$$\mathsf{E}\ N_n^{-5/2} \ = \ \frac{1}{n^{5/2}} \ + \ O\Big(\frac{1}{n^{5/2}}\ \Big(\frac{h_n}{n}\Big)^2\Big),$$

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

Доказательство. Доказательство Леммы следует из равенств

$$\mathsf{E}\ N_n^{-1} \ = \ \frac{3n^2 \ - \ h_n^2}{3n(n^2 \ - \ h_n^2)} \ =$$

$$= \ \frac{1}{n}\ \Big(1 \ - \ \frac{h_n^2}{3n}\Big)\ \Big(1 \ + \ \frac{h_n^2}{n^2} \ + \ O\Big(\frac{h_n^4}{n^4}\Big)\Big) \ =$$

$$= \ \frac{1}{n} \ + \frac{2}{3n}\ \Big(\frac{h_n}{n}\Big)^2 \ + \ O\Big(\frac{1}{n}\ \Big(\frac{h_n}{n}\Big)^4\Big),$$

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

$$= \ \frac{1}{n^{3/2}} \ + \ O\Big(\frac{1}{n^{3/2}}\ \Big(\frac{h_n}{n}\Big)^2\Big),$$

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

$$= \ \frac{1}{n^2} \ + \ O\Big(\frac{1}{n^2}\ \Big(\frac{h_n}{n}\Big)^2\Big).$$

Аналогично доказываются оставшиеся соотношения. Лемма доказана. □

Предположим теперь, что последовательность натуральных чисел $h_n \ < \ n$ удовлетворяет следующему условию

$$\frac{h_n}{n} \ = \ \alpha_0 \ + \ \alpha_1\ n^{-\delta} \ + \ o(n^{-\delta}), \quad n \ \to \ \infty, \tag{5.3}$$

где $\alpha_0 \ \in \ (0,1)$, $\delta \ > \ 0$ и α_1 – некоторые числа. Тогда аналогично предыдущему утверждению, получаем следующую Лемму.

Лемма 5.2. *Пусть случайная величина N_n имеет распределение (5.1) и выполнено условие (5.3), тогда для любого числа $p \ > \ 0$ справедливы равенства*

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

$$\mathsf{E}\ N_n^{-p} \ =$$

$$= \ \frac{1}{3n^p}\ \left(1 \ + \ \frac{(1+\alpha_0)^p \ + \ (1-\alpha_0)^p}{(1-\alpha_0^2)^p}\right) +$$

$$+ \ \frac{p\alpha_1\ \big((1+\alpha_0)^{p+1} \ - \ (1-\alpha_0)^{p+1}\big)}{3n^{p+\delta}(1-\alpha_0^2)^{p+1}} \ + \ o(n^{-p-\delta}).$$

Доказательство. Доказательство Леммы следует из равенств

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

$$f(x) \equiv \frac{1}{(1 \pm x)^p} =$$

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

причем последняя формула применяется с

$$x = \frac{h_n}{n} = \alpha_0 + \alpha_1\, n^{-\delta} + o(n^{-\delta}).$$

Лемма доказана. □

Из Теоремы 3.1, Леммы 2.1 и Лемм 5.1, 5.2 непосредственно получаем следующие утверждения.

Теорема 5.1. *Пусть случайная величина N_n имеет распределение (5.1) и*

1. n и h_n – четные числа и выполнено условие (5.2), тогда

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2n^2}\left(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4} - \frac{2h_n^2}{3n}\right) +$$

$$+ O\Big(\max\Big\{\frac{1}{n}\Big(\frac{h_n}{n}\Big)^4, \frac{1}{n^{3/2}}\Big(\frac{h_n}{n}\Big)^2\Big\}\Big).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} - \frac{1}{4n^2}\Big(1 - \frac{8h_n^2}{3n}\Big) +$$

$$+ O\Big(\max\Big\{\frac{1}{n}\Big(\frac{h_n}{n}\Big)^4, \frac{1}{n^{3/2}}\Big(\frac{h_n}{n}\Big)^2\Big\}\Big).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2}{8n^2}\Big(\pi^2 - 6 + \frac{4h_n^2}{3n}\Big) +$$

$$+ O\Big(\max\Big\{\frac{1}{n}\Big(\frac{h_n}{n}\Big)^4, \frac{1}{n^{3/2}}\Big(\frac{h_n}{n}\Big)^2\Big\}\Big), \qquad n \to \infty.$$

2. n – нечетное число, а h_n – четное, тогда

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{4p_0^2 n} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}n^{3/2}} - \frac{1}{4p_0^2n^2}\left(2 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4} - \frac{2h_n^2}{3n}\right) +$$

$$+ \, O\Big(\max\Big\{\frac{1}{n}\,\Big(\frac{h_n}{n}\Big)^4,\ \frac{1}{n^{3/2}}\,\Big(\frac{h_n}{n}\Big)^2\Big\}\Big).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{1}{n} + \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}} + \frac{1}{4n^2}\Big(1 + \frac{8h_n^2}{3n}\Big) +$$

$$+ \, O\Big(\max\Big\{\frac{1}{n}\,\Big(\frac{h_n}{n}\Big)^4,\ \frac{1}{n^{3/2}}\,\Big(\frac{h_n}{n}\Big)^2\Big\}\Big).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4n} + \frac{\pi^2}{8n^2}\Big(\pi^2 - 4 + \frac{4h_n^2}{3n}\Big) +$$

$$+ \, O\Big(\max\Big\{\frac{1}{n}\,\Big(\frac{h_n}{n}\Big)^4,\ \frac{1}{n^{3/2}}\,\Big(\frac{h_n}{n}\Big)^2\Big\}\Big), \qquad n \to \infty.$$

Теорема 5.2. *Пусть случайная величина N_n имеет распределение (5.1), n, h_n – четные числа и выполнено условие (5.3), тогда*

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{4p_0^2}\,\mathsf{E}\,N_n^{-1} - \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}}\,\mathsf{E}\,N_n^{-3/2} - \frac{1}{4p_0^2}\,\mathsf{E}\,N_n^{-2}\,\Big(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\Big) + O(\mathsf{E}\,N_n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \mathsf{E}\,N_n^{-1} + \frac{2\sqrt{2}}{\sqrt{\pi}}\,\mathsf{E}\,N_n^{-3/2} - \frac{1}{4}\,\mathsf{E}\,N_n^{-2} + O(\mathsf{E}\,N_n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4}\,\mathsf{E}\,N_n^{-1} + \frac{\pi^2(\pi^2 - 6)}{8}\,\mathsf{E}\,N_n^{-2} + O(\mathsf{E}\,N_n^{-5/2}).$$

2. Если n нечетное число, а h_n – четное, то

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{1}{4p_0^2}\,\mathsf{E}\,N_n^{-1} -$$

$$- \frac{p_1\sqrt{2}}{4p_0^4\sqrt{\pi}}\,\mathsf{E}\,N_n^{-3/2} - \frac{1}{4p_0^2}\,\mathsf{E}\,N_n^{-2}\,\Big(2 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\Big) + O(\mathsf{E}\,N_n^{-5/2}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \mathsf{E}\,N_n^{-1} + \frac{2\sqrt{2}}{\sqrt{\pi}}\,\mathsf{E}\,N_n^{-3/2} + \frac{1}{4}\,\mathsf{E}\,N_n^{-2} + O(\mathsf{E}\,N_n^{-5/2}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 = \frac{\pi^2}{4}\,\mathsf{E}\,N_n^{-1} + \frac{\pi^2(\pi^2 - 4)}{8}\,\mathsf{E}\,N_n^{-2} + O(\mathsf{E}\,N_n^{-5/2}).$$

Подставляя в формулы из этой теоремы асимптотические разложения для обратных моментов из Леммы 5.2, непосредственно получаем следующее утверждение.

Следствие 5.1. *Пусть выполнены условия Теоремы 5.2, тогда*

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{12p_0^2\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{4\alpha_1\,\alpha_0}{12p_0^2\, n^{1+\delta}(1-\alpha_0^2)^2} -$$

$$- \frac{p_1\sqrt{2}}{12p_0^4\sqrt{\pi}n^{3/2}}\left(1 + \frac{(1+\alpha_0)^{3/2} + (1-\alpha_0)^{3/2}}{(1-\alpha_0^2)^{3/2}}\right) -$$

$$- \frac{p_1\sqrt{2}\alpha_1\,\left((1+\alpha_0)^{5/2} - (1-\alpha_0)^{5/2}\right)}{8p_0^4\sqrt{\pi}n^{3/2+\delta}(1-\alpha_0^2)^{5/2}} -$$

$$- \frac{1}{12p_0^2n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right)\left(3 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\right) + o(n^{-1-\delta}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{3\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{4\alpha_1\,\alpha_0}{3\, n^{1+\delta}(1-\alpha_0^2)^2} +$$

$$+ \frac{2\sqrt{2}}{3\sqrt{\pi}n^{3/2}}\left(1 + \frac{(1+\alpha_0)^{3/2} + (1-\alpha_0)^{3/2}}{(1-\alpha_0^2)^{3/2}}\right) -$$

$$- \frac{\sqrt{2}\alpha_1\,\left((1+\alpha_0)^{5/2} - (1-\alpha_0)^{5/2}\right)}{\sqrt{\pi}n^{3/2+\delta}(1-\alpha_0^2)^{5/2}} -$$

$$- \frac{1}{12n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right) + o(n^{-1-\delta}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{\pi^2}{12\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{\pi^2\alpha_1\,\alpha_0}{3\, n^{1+\delta}(1-\alpha_0^2)^2} +$$

$$+ \frac{\pi^2(\pi^2 - 6)}{24n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right) + o(n^{-1-\delta}).$$

2. Если n нечетное число, а h_n – четное, то

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{12p_0^2\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{4\alpha_1\, \alpha_0}{12p_0^2\, n^{1+\delta}(1-\alpha_0^2)^2} -$$

$$- \frac{p_1\sqrt{2}}{12p_0^4\sqrt{\pi}n^{3/2}}\left(1 + \frac{(1+\alpha_0)^{3/2} + (1-\alpha_0)^{3/2}}{(1-\alpha_0^2)^{3/2}}\right) -$$

$$- \frac{p_1\sqrt{2}\alpha_1\,\big((1+\alpha_0)^{5/2} - (1-\alpha_0)^{5/2}\big)}{8p_0^4\sqrt{\pi}n^{3/2+\delta}(1-\alpha_0^2)^{5/2}} -$$

$$- \frac{1}{12p_0^2n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right)\left(2 + \frac{p_2}{4p_0^3} - \frac{15p_1^2}{16p_0^4}\right) + o(n^{-1-\delta}).$$

Для распределения Лапласа (2.5) справедливо соотношение

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{1}{3\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{4\alpha_1\, \alpha_0}{3\, n^{1+\delta}(1-\alpha_0^2)^2} +$$

$$+ \frac{2\sqrt{2}}{3\sqrt{\pi}n^{3/2}}\left(1 + \frac{(1+\alpha_0)^{3/2} + (1-\alpha_0)^{3/2}}{(1-\alpha_0^2)^{3/2}}\right) -$$

$$- \frac{\sqrt{2}\alpha_1\,\big((1+\alpha_0)^{5/2} - (1-\alpha_0)^{5/2}\big)}{\sqrt{\pi}n^{3/2+\delta}(1-\alpha_0^2)^{5/2}} +$$

$$+ \frac{1}{12n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right) + o(n^{-1-\delta}).$$

В случае распределения Коши (2.4)

$$\mathsf{E}_\theta\Big(M_{N_n} - \theta\Big)^2 =$$

$$= \frac{\pi^2}{12\, n}\left(1 + \frac{2}{1-\alpha_0^2}\right) + \frac{\pi^2\alpha_1\, \alpha_0}{3\, n^{1+\delta}(1-\alpha_0^2)^2} +$$

$$+ \frac{\pi^2(\pi^2 - 4)}{24n^2}\left(1 + \frac{2 + 2\alpha_0^2}{(1-\alpha_0^2)^2}\right) + o(n^{-1-\delta}).$$

Теорема 5.3. *Пусть выполнены условия Теоремы 5.1, тогда если $p_1 \neq 0$, то дефект выборочной медианы M_{N_n} относительно M_n имеет вид*

$$d_n = o(\sqrt{n}), \quad n \to \infty,$$

если $p_1 \; = \; 0$, *и*

$$\frac{h_n^2}{n} \; \to \; h, \quad n \; \to \; \infty,$$

то

$$d \; = \; \lim_{n \to \infty} \; d_n \; = \; \frac{2}{3} \; h.$$

В частности для распределения Лапласа (2.5) справедливо соотношение

$$d_n \; = \; o(\sqrt{n}),$$

а для распределения Коши (2.4)

$$d \; = \; \frac{2}{3} \; h.$$

Замечание 5.1. *Заметим, что в условии Теоремы 5.3*

$$\frac{h_n^2}{n} \; \to \; h, \quad n \; \to \; \infty,$$

h может быть сколь угодно большой величиной (для этого достаточно рассмотреть случай, когда h_n – последовательность натуральных чисел порядка $\sqrt{h\ n}$), поэтому указанный в Теореме 5.3 асимптотический дефект d в принципе может быть произвольно большим.

Напомним теперь понятие предельной риск-эффективности (ПРЭ) последовательности оценок (см., например, [3] стр. 308). Пусть $\delta_n \; \equiv \; \delta_n(X_1, \ldots, X_n)$, $\delta_n^* \; \equiv \; \delta_n^*(X_1, \ldots, X_n)$ – две последовательности оценок параметрической функции $g(\theta)$ соответственно с функциями риска

$$R(\theta) \; = \; \mathsf{E}_\theta \; (\delta_n \; - \; g(\theta))^2, \quad R^*(\theta) \; = \; \mathsf{E}_\theta \; (\delta_n^* \; - \; g(\theta))^2.$$

Под предельной риск-эффективностью 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 $\bar{m}(n)$) вида (см., например, [3] стр. 308)

$$\bar{e} \; \equiv \; \lim_{n \to \infty} \; \frac{n}{\bar{m}(n)},$$

где последовательность натуральных чисел $\bar{m}(n)$ определяется из условия

$$\lim_{n \to \infty} \; n^r \; R_n^*(\theta) \; = \; \lim_{n \to \infty} \; n^r \; R_{\bar{m}(n)}(\theta) \; < \; \infty,$$

при некотором $r > 0$. Таким образом, если ПРЭ равна, например, $1/3$, то оценке $\delta_n(X_1, \ldots, X_n)$ для достижения того же качества, что и оценке $\delta_n^*(X_1, \ldots, X_n)$ требуется примерно втрое больше ($\bar{m}(n) \; \approx \; 3n$) наблюдений.

В книге [3] (Теорема 2.3, стр. 309) доказано следующее утверждение: если при некотором $r > 0$

$$\lim_{n \to \infty} \; n^r \; R_n^*(\theta) \; = \; \tau^* \; > \; 0, \quad \lim_{n \to \infty} \; n^r \; R_n(\theta) \; = \; \tau \; > \; 0,$$

то ПРЭ оценки δ_n относительно оценки δ_n^* равна

$$\bar{e} = \left(\frac{\tau^*}{\tau}\right)^{1/r}.$$

Применяя это утверждение с

$$r = 1, \ \ \tau = \frac{1}{12\, p_0^2}\left(1 + \frac{2}{1-\alpha_0^2}\right), \ \ \tau^* = \frac{1}{4\, p_0^2}$$

и Следствие 5.1 к выборочным медианам M_{N_n}, M_n, непосредственно получаем следующий результат.

Теорема 5.4. *Пусть выполнены условия Теоремы 5.2, тогда ПРЭ выборочной медианы M_{N_n} относительно M_n имеет вид*

$$\bar{e} = \frac{3 - 3\alpha_0^2}{3 - \alpha_0^2} < 1.$$

6. Случай распределения Пуассона

В этом разделе будет найден дефект выборочной медианы M_n, основанной на выборках, объем которых имеет распределение Пуассона. Всюду далее будем предполагать, что

$$N_n = \Pi_n + 1,$$

где случайная величина Π_n имеет распределение Пуассона вида

$$\mathsf{P}\left(\Pi_n = k\right) = \frac{e^{-\lambda_n}\ \lambda_n^k}{k!}, \quad k = 0,\ 1,\ 2,\ \ldots, \quad \lambda_n = n-1, \quad n \geq 2,$$

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

В случае, если исходные наблюдения имеют распределение Коши, то, используя результаты Леммы 2.1, нетрудно видеть, что

$$k_1 = 1, \ \ k_2 = 1, \ \ a(\theta) = \pi^2/4.$$

Кроме того, $b(\theta) = \pi^2(\pi^2-6)/8$ для четных n, и $b(\theta) = \pi^2(\pi^2-4)/8$ для нечетных. Теперь, применяя Теорему 3.2, непосредственно получаем, что

$$d(\theta) = 1. \tag{6.1}$$

В общем случае имеем

$$r = 1,\ s = \frac{1}{2},\ \alpha = \frac{1}{2}.$$

В работе [9] получены асимптотические разложения следующих обратных моментов

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

Аналогично предыдущему отсюда следует, что

$$d_n(\theta) \ = \ \frac{\sqrt{2}p_1}{\sqrt{\pi}p_0^2}\ n^{1/2} \ + \ O(\sqrt{n}). \tag{6.2}$$

Заключение

В работе рассмотрен случай статистического оценивания неизвестного параметра сдвига, основанный на выборочной медиане, построенной по выборкам случайного объема. С помощью понятия дефекта проведено асимптотическое сравнение качества такой оценки. Получены явные формулы для асимптотического дефекта, имеющего смысл необходимого добавочного числа наблюдений. Подробно рассмотрен случай, когда случайный объем выборки имеет биномиальное, пуассоновское и трехточечное симметричное распределения.

Список литературы

[1] Hodges J.L., Lehmann E.L. Deficiency // 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 объема // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 5–14.

[6] Бурнашев М.В. Асимптотические разложения для медианной оценки параметра // Теория вероятностей и ее применения. 1996. Т. 41, № 4. С. 738–753.

[7] Гнеденко Б.В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Труды Тбилисского Математического Института. 1989. Т. 92. С. 147–150.

[8] Lehmann E.L. Elements of Large – Sample Theory. Springer, 1999. 631 p.

[9] Бенинг В.Е. Об асимптотическом поведении обратных моментов некоторых случайных величин // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 2. С. 47–65.

[10] 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. 75–91.

[11] 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.

[12] Бенинг В.Е., Королев В.Ю. Об использовании распределения Стьдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения. 2008. Т. 2, № 2. С. 19–34.

[13] 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.

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

[15] 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.

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

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

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

[19] Бенинг В.Е., Савушкин В.А. Асимптотические разложения для функции распределения выборочной медианы в случае выборок случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 4. С. 39–54.

Библиографическая ссылка

Бенинг В.Е., Савушкин В.А. О дефекте выборочной медианы в случае выборок случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2016. № 2. С. 5–30.

Сведения об авторах

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

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

2. **Савушкин Владислав Андреевич**
ассистент кафедры прикладной математики и информатики факультета естественных и инженерных наук государственного университета Природы, Общества и Человека «Дубна».

Россия, 141982, Московская область, г. Дубна, улица Университетская, д. 19, Университет Дубна. E-mail: savushkinva@mail.ru.

ON THE DEFICIENCY OF SAMPLE MEDIANE BASED ON THE SAMPLE WITH RANDOM SIZE

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

Savushkin Vladislav Andreevich
Assistant at Applied Mathematics and Computer Science department,
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 28.04.2016, revised 15.05.2016.

In the paper general theorems concerning the asymptotic deficiencies of sample mediane based on the sample of random size are proved.

Keywords: sample mediane, asymptotic deficiency, sample with random size, binomial and Poisson distributions.

Bibliographic citation

Bening V.E., Savushkin V.A. On the deficiency of sample mediane based on the sample with random size. *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2016, no. 2, pp. 5–30. (in Russian)

References

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

[2] Kramer G. *Matematicheskie Metody Statistiki* [Mathematical Methods of Statistics]. Mir Publ., Moscow, 1976. 648 p. (in Russian)

[3] Leman E. *Teoriya Tochechnogo Ocenivaniya* [Theory of Point Estimation]. Nauka, FizMatLit Publ., Moscow, 1991. 444 p. (in Russian)

[4] Bening V.E. *Asymptotic theory of testing statistical hypotheses: efficient statistics, optimality, power loss, and deficiency.* VSP, Utrecht, 2000. 277 p.

[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, no. 1, pp. 5–14. (in Russian)

[6] Burnashev M.V. The asymptotic expansions for median estimate parameter. *Teoriya Veroyatnostei i ee Primeneniya* [Probability Theory and its Applications], 1996, vol. 41(4), pp. 738–753. (in Russian)

[7] Gnedenko B.V. An estimate of the distribution of the unknown parameters with a random number of independent observations. *Trudy Tbilisskogo Matematicheskogo Instituta* [Proceedings of the Tbilisi Institute of Mathematics], 1989, vol. 92, pp. 147–150. (in Russian)

[8] Lehmann E.L. *Elements of Large – Sample Theory.* Springer, 1999. 631 p.

[9] Bening V.E. Asymptotic behavior of inverse moments of some discrete random variables. *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2015, no. 2, pp. 47–65. (in Russian)

[10] Bening V.E., Galieva N.K., Korolev V.Yu. Asymptotic expansions for the distribution of the statistics functions, constructed from samples with random sizes. *Informatika i ee Primeneniya* [Informatics and its Applications], 2013, vol. 7(2), pp. 75–91. (in Russian)

[11] 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)

[12] Bening V.E., Korolev V.Yu. On the use of Student distribution in the problems of probability theory and mathematical statistics. *Teoriya Veroyatnostei i ee Primeneniya* [Probability Theory and its Applications], 2008, vol. 2(2), pp. 19–34. (in Russian)

[13] Bening V.E., Korolev V.Yu., Sokolov I.A., Shorgin S.Ya. *Randomizirovannye Modeli i Metody Teorii Nadezhnosti Informatsionnykh i Tekhnicheskikh Sistem* [Randomized Models and Methods of the Theory of Reliability of Information and Technical Systems]. TORUS PRESS Publ., Moscow, 2007. (in Russian)

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

[15] Lyamin O.O. On the rate of convergence of the distributions of some statistics to the Laplace and Student distributions. *Vestnik Moskovskogo Universiteta. Seriya 15: Vychislitel'naya Matematika i Kibernetika* [Bulletin of Moscow University. Series 15: Computational Mathematics and Cybernetics], 2011, no. 1, pp. 39–47. (in Russian)

[16] Dwait G.B. *Tablitsy Integralov i Drugie Matematicheskie Formuly* [Tables of Integrals and Other Mathematical Formulas]. Nauka Publ., Moscow, 1977. (in Russian)

[17] Wilks S.S. Recurrence of extreme observations. *Journal of American Mathematical Society*, 1959, vol. 1(1), pp. 106–112.

[18] Nevzorov V.B. *Rekordy. Matematicheskaya Teoriya* [Records. Mathematical Theory]. Fazis Publ., Moscow, 2000. (in Russian)

[19] Bening V.E., Savushkin V.A. Asymptotic expansions for the distribution function of sample median based on the sample with random size. *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2015, no. 4, pp. 39–54. (in Russian)