ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ДЛЯ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ 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 11.11.2015, после переработки 15.11.2015.

В работе получены асимптотические разложения для функции распределения выборочной медианы в случае, когда размер выборки есть случайная величина. Приведена общая теорема, описывающая асимптотические разложения в этом случае. Рассмотрены случаи, когда предельными распределениями выборочной медианы являются распределения Стьюдента и Лапласа.

Ключевые слова: выборочная медиана, выборка случайного объема, асимптотическое разложение, распределение Стьюдента, распределение Коши, распределение Лапласа.

Вестник ТвГУ. Серия: Прикладная математика. 2015. № 4. С. 39–54.

1. Введение

В данной работе получены асимптотические разложения (а.р.) для функций распределения (ф.р.) выборочной медианы, построенной по выборкам случайного объема. Рассмотрены случаи распределения Лапласа, Стьюдента и Коши. Эти результаты продолжают исследования, начатые в работах [1, 5, 7, 9–14] и в дальнейшем будут использованы для нахождения асимптотического дефекта (см. [2, 5]) выборочной медианы, построенной по выборкам случайного объема.

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

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 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ – вариационный ряд, построенный

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

по исходным наблюдениям $X_1, X_2, \dots X_n$ и через M_n – выборочную медину (см., например, [3], [4], [6], [7]), то есть оценку вида

$$M_n = \begin{cases} X_{(m+1)}, & n = 2m + 1, \\ \dfrac{X_{(m)} + X_{(m+1)}}{2}, & n = 2m. \end{cases} \tag{1.1}$$

Асимптотические свойства первого порядка выборочной медианы M_n хорошо известны (см., например, книгу [4], Теорема 5.3.2, стр. 313 или книгу [8], стр. 81)

$$\sqrt{n}\,\big(M_n - \theta\big) \Longrightarrow \mathcal{N}\Big(0,\ \frac{1}{4p^2(0)}\Big), \tag{1.2}$$

$$\mathsf{E}_\theta\big(M_n - \theta\big)^2 = \frac{1}{4np^2(0)} + o\big(n^{-1}\big),\ \ F(0) = 1/2,\ \ p(0) > 0,\quad n \to \infty. \tag{1.3}$$

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

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

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

$$p(x) = \frac{1}{\pi(1 + x^2)},\quad x \in \mathbb{R} \tag{1.4}$$

и распределения Лапласа

$$p(x) = \frac{1}{2}\,e^{-|x|},\quad x \in \mathbb{R}. \tag{1.5}$$

В случае распредления 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.

Теорема 1.1.([6])

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

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

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

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

$$\mathsf{E}_\theta\Big(M_n \ - \ \theta\Big)^2 \ = \ \frac{1}{8p_0^2 k} \ - \ \frac{p_1}{8p_0^4\sqrt{\pi}k^{3/2}} \ -$$

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

Следствие 1.1.

1. *Для распределения Лапласа (1.5) справедливы асимптотические разложения*

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

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

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

2. *Для распределения Коши (1.4) справедливы асимптотические разложения*

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

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

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

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

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

С учетом этих формул основные утверждения Теоремы 1.1 и ее Следствия можно переписать в виде

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

$$+ \ \phi(x) \ \frac{p_1 x|x|\sqrt{2}}{4p_0\sqrt{n}} \ + \ \phi(x) \ \frac{x}{4n}\Big(3 \ + \ x^2 \ + \ \frac{x^2 p_2}{6p_0^3} \ - \ \frac{x^4 p_1^2}{8p_0^4}\Big) \ + \ O(n^{-3/2}), \qquad (1.6)$$

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

$$+\ \frac{1}{4p_0^2n^2}\left(\frac{1\ -\ (-1)^n}{2}\ -\ 3\ -\ \frac{p_2}{4p_0^3}\ +\ \frac{15p_1^2}{16p_0^4}\right)\ +\ O(n^{-5/2}). \tag{1.7}$$

Для распределения Лапласа (1.5) справедливы равенства

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

$$-\ \phi(x)\ \frac{x|x|}{2\sqrt{n}}\ +\ \phi(x)\ \frac{x}{24n}\Big(18\ +\ 10x^2\ -\ 3x^4\Big)\ +\ O(n^{-3/2}), \tag{1.8}$$

$$\mathsf{E}_\theta\Big(M_n\ -\ \theta\Big)^2\ =\ \frac{1}{n}\ +\ \frac{2\sqrt{2}}{\sqrt{\pi}n^{3/2}}\ +\ \frac{1}{2n^2}\Big((1\ -\ (-1)^n)\ -\ \frac{1}{2}\Big)\ +\ O(n^{-5/2}), \tag{1.9}$$

а для распределения Коши (1.4)

$$\mathsf{P}_\theta\Big(\frac{2\sqrt{2k}}{\pi}(M_n-\theta)\ <\ x\Big)\ =\ \Phi(x)+\phi(x)\ \frac{x}{12n}\Big(9+x^2(3-\pi^3)\Big)+O(n^{-3/2}), \tag{1.10}$$

$$\mathsf{E}_\theta\Big(M_n\ -\ \theta\Big)^2\ =\ \frac{\pi^2}{4n}\ +\ \frac{\pi^2\big((1\ -\ (-1)^n)\ +\ \pi^2\ -\ 6\big)}{8n^2}\ +\ O(n^{-5/2}). \tag{1.11}$$

2. Асимптотические разложения для функции распределения выборочной медианы, основанной на выборках случайного объема

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

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

то есть среднее значение случайного объема выборки равно n.

Предположим, что при каждом $n\geq 1$ с.в. N_n принимают только натуральные значения (то есть $N_n\in\ \mathbb{N}$) и независимы от последовательности с.в. $X_1,\ X_2,\ \dots$. Всюду далее считаем с.в. $X_1,\ X_2,\ \dots$ независимыми и одинаково распределенными.

Обозначим через $T_n=T_n(X_1,\dots,X_n)$ некоторую статистику, то есть действительную измеримую функцию от наблюдений $X_1,\dots,X_n$. Для каждого $n\geq 1$ определим с.в. T_{N_n}, полагая

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

Можно сказать, что T_{N_n} – это статистика, построенная на основе статистики T_n по выборке случайного объема N_n.

Сформулируем условие, определяющее а.р. для ф.р. статистики T_n при неслучайном объеме выборки.

Условие 2.1. *Существуют константы $l \in \mathbb{N}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\alpha > l/2$, $\gamma \geq 0$, $C_1 > 0$, дифференцируемая ф.р. $F(x)$ и дифференцируемые ограниченные функции $f_j(x), j = 1, ..., l$ такие, что*

$$\sup_x \Big| \mathsf{P}\big(\sigma n^\gamma (T_n - \mu) < x\big) - F(x) - \sum_{j=1}^{l} n^{-j/2} f_j(x) \Big| \leq \frac{C_1}{n^\alpha}, \quad n \in \mathbb{N}.$$

Следующее условие определяет а.р. ф.р. нормированного случайного индекса N_n.

Условие 2.2. *Существуют константы $m \in \mathbb{N}$, $\beta > m/2$, $C_2 > 0$, функция $0 < g(n) \uparrow \infty$, $n \to \infty$, ф.р. $H(x)$, $H(0+) = 0$ и функции ограниченной вариации $h_i(x)$, $i = 1, ..., m$ такие, что*

$$\sup_{x \geq 0} \Big| \mathsf{P}\Big(\frac{N_n}{g(n)} < x\Big) - H(x) - \sum_{i=1}^{m} n^{-i/2} h_i(x) \Big| \leq \frac{C_2}{n^\beta}, \quad n \in \mathbb{N}.$$

В работе [9] доказано следующее утверждение

Теорема 2.1. *Пусть статистика $T_n = T_n(X_1, ..., X_n)$ удовлетворяет условию* 2.1, *а случайный объем выборки N_n – условию* 2.2. *Тогда существует константа $C_3 > 0$ такая, что справедливо неравенство*

$$\sup_x \big| \mathsf{P}\big(\sigma\, g^\gamma(n)(T_{N_n} - \mu) < x\big) - G_n(x)\big| \leq C_1\, \mathsf{E}\, N_n^{-\alpha} + \frac{C_3 + C_2 D_n}{n^\beta},$$

где

$$D_n = \sup_x \int\limits_{1/g(n)}^{\infty} \Big| \frac{\partial}{\partial\, y} \Big(F(xy^\gamma) + \sum_{j=1}^{l} (yg(n))^{-j/2} f_j(xy^\gamma)\Big) \Big| \, dy$$

и а.р. $G_n(x)$ определено по формуле

$$G_n(x) = \int\limits_{1/g(n)}^{\infty} F(x\, y^\gamma)\, dH(y) + \sum_{i=1}^{m} n^{-i/2} \int\limits_{1/g(n)}^{\infty} F(xy^\gamma)\, dh_i(y) +$$

$$+ \sum_{j=1}^{l} g^{-j/2}(n) \int\limits_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma)\, dH(y) +$$

$$+ \sum_{j=1}^{l} \sum_{i=1}^{m} n^{-i/2} g^{-j/2}(n) \int\limits_{1/g(n)}^{\infty} y^{-j/2} f_j(xy^\gamma)\, dh_i(y).$$

Учитывая теорему 1.1, нетрудно видеть, что выборочная медиана M_n удовлетворяет условию 2.1 с

$$\gamma = \frac{1}{2}, \quad \alpha = \frac{3}{2}, \quad l = 2, \quad \mu = \theta, \quad \sigma = 2p_0\sqrt{2}, \quad g(n) = \sqrt{k}, \tag{2.1}$$

$$F(x) = \Phi(x), \quad f_1(x) = \phi(x)\, \frac{p_1 x|x|\sqrt{2}}{4p_0},$$

$$f_2(x) \;=\; \phi(x)\,\frac{x}{4}\Big(3 \;+\; x^2 \;+\; \frac{x^2p_2}{6p_0^3} \;-\; \frac{x^4p_1^2}{8p_0^4}\Big). \tag{2.2}$$

Аналогично доказательству леммы 5.1 [9] можно доказать, что существует константа $D \;>\; 0$ такая, что

$$D_n \leq\; D, \quad n \;\in\; \mathbb{N}. \tag{2.3}$$

В работе [10] получена подобная теорема, в которой отсутствует нормировка статистики, и в которой используется следующее уловие регулярности.

Условие 2.3. *Существуют константы $l \in \mathbb{N}$, $\alpha \;>\; l/2$, $C_1 \;>\; 0$, дифференцируемая ф.р. $G(x)$ и дифференцируемые ограниченные функции $g_i(x), i = 1,\ldots,l$ такие, что*

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

Теорема 2.2. *Пусть выполнены условия 2.2 и 2.3. Тогда справедливо неравенство*

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

где $G_n(x)$ есть

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

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

3. Распределение Стьюдента

В работе [11] показано, что если случайный объем выборки N_n имеет отрицательно биномиальное распределение с параметрами $p \;=\; 1/n$ и $r \;>\; 0$, то есть

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

(при $r \;=\; 1$ имеем геометрическое распределение), то для асимптотически нормальной статистики T_n справедливо предельное соотношение ([11], следствие 2.1, стр. 426)

$$\mathsf{P}(\sigma\sqrt{n}\;(T_{N_n} \;-\; \mu) \;<\; x) \;\longrightarrow\; G_{2r}(x\sqrt{r}), \quad n \;\to\; \infty, \tag{3.1}$$

где $G_f(x)$ – функция распределения Стьюдента с параметром $f \;=\; 2r$, соответствующая плотности вида

$$p_f(x) \;=\; \frac{\Gamma(f+1/2)}{\sqrt{\pi f}\;\Gamma(f/2)}\Big(1 \;+\; \frac{x^2}{f}\Big)^{-(\gamma+1)/2}, \quad x \;\in\; \mathbb{R},$$

где $\Gamma(\cdot)$ – эйлерова гамма-функция, а $f > 0$ – параметр формы (если параметр f натурален, то он называется числом степеней свободы). В нашей ситуации он может быть произвольно мал, то есть мы имеем типичное распределение с тяжелыми хвостами. Если $f = 2$, то есть $r = 1$, то ф.р. $G_2(x)$ выражается в явном виде

$$G_2(x) = \frac{1}{2}\Big(1 + \frac{x}{\sqrt{2+x^2}}\Big), \quad x \in \mathbb{R}.$$

При $r = 1/2$ имеем распределение Коши.

В книге [12] (формула (6.112), стр. 233) приведена следующая оценка скорости сходимости:

$$\sup_{x\geq 0}\Big|\mathsf{P}\Big(\frac{N_n}{\mathsf{E}\,N_n} < x\Big) - H_r(x)\Big| \leq \begin{cases}\dfrac{C_r}{n}, & r \geq 1,\\ \dfrac{C_r}{n^r}, & r \in (0,1),\end{cases} \qquad C_r > 0, \quad n \in \mathbb{N}, \tag{3.2}$$

где $H_r(x)$ – функция гамма-распределения с параметром $r > 0$:

$$H_r(x) = \frac{r^r}{\Gamma(r)} \int\limits_0^x e^{-ry}\, y^{r-1}\, dy, \qquad x \geq 0. \tag{3.3}$$

При этом

$$\mathsf{E}\, N_n = r(n-1) + 1. \tag{3.4}$$

Таким образом, из соотношений (3.2) – (3.4) следует, что случайный индекс N_n удовлетворяет условию 2.2 с

$$g(n) = r(n - 1) + 1, \quad H(x) = H_r(x), \quad m = 1, \tag{3.5}$$

$$h_1(x) \equiv 0, \quad C_2 = C_r > 0, \tag{3.6}$$

$$\beta = \begin{cases}1, & r \geq 1,\\ r, & r \in (0,1).\end{cases}$$

Далее, используя равенство

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

нетрудно получить, что

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

Для случая $r = 1$, имеем

$$\mathsf{E}N_n^{-1} = \frac{1}{n - 1}\log n, \quad n > 1. \tag{3.8}$$

Таким образом, учитывая теорему 2.1, имеем

$$\int\limits_{(r(n-1)+1)^{-1}}^{\infty} \Phi(x\sqrt{y})dH_r(y) = \int\limits_{0}^{\infty} \Phi(x\sqrt{y})dH_r(y) + \mathcal{O}\Big(\frac{1}{n}\Big) = G_{2r}(x) + \mathcal{O}\Big(\frac{1}{n}\Big), \quad (3.9)$$

$$x|x| \int\limits_{(r(n-1)+1)^{-1}}^{\infty} \varphi(x\sqrt{y})\sqrt{y}\; dH_r(y) \;=\; x|x| \int\limits_{0}^{\infty} \varphi(x\sqrt{y})\;\sqrt{y}\; dH_r(y) \;+\; \mathcal{O}\Big(\frac{1}{n}\Big) \;\equiv$$

$$\equiv \frac{x\;|x|\;r^r\;\Gamma(r+1/2)}{\sqrt{2\pi}\;(r+x^2/2)^{r+1/2}\;\Gamma(r)} \;+\; \mathcal{O}\Big(\frac{1}{n}\Big), \quad (3.10)$$

получаем следующее утверждение.

Теорема 3.1. *Пусть выполнены условия A1 и A2 и при некотором $r > 0$ случайная величина N_n имеет распределение вида*

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

Тогда при $r > 1/2$ для ф.р. нормированной выборочной медианы при $n \to \infty$ справедливо а.р. вида

$$\sup_x \Big|\mathsf{P}_\theta\Big(2p_0\sqrt{2m}\;(M_{N_n} - \theta) < x\Big) - G_{2r}(x) -$$

$$- \frac{p_1\;\Gamma(r+1/2)\;x\;|x|\;r^r}{2p_0\;(r+x^2/2)^{r+1/2}\;\Gamma(r)\;\sqrt{2\pi}\;\sqrt{r(n-1)+1}}\Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2, \end{cases}$$

где функция $G_{2r}(x)$ определена в соотношении (3.9) *и* $m = [(r(n-1)+1)/2]$.

Следствие 3.1.

1. *В случае распределения Лапласа (1.3) для ф.р. выборочной медианы M_n справедливо асимптотическое разложение*

$$\sup_x \Big|\mathsf{P}_\theta\Big(\sqrt{2m}\;(M_{N_n} - \theta) < x\Big) - G_{2r}(x) -$$

$$- G_{2r}(x) - \frac{\Gamma(r+1/2)\;x\;|x|\;r^r}{2\Gamma(r)\;\sqrt{\pi}\;(r+x^2/2)^{r+1/2}\;\sqrt{r(n-1)+1}}\Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2, \end{cases}$$

где функция $G_{2r}(x)$ определена в соотношении (3.9).

2. Для распределения Коши (1.4) справедливо а.р. вида

$$\sup_x \Big| \mathsf{P}_\theta\Big(\frac{2\sqrt{2m}}{\pi}\ (M_{N_n} - \theta) < x \Big) - G_{2r}(x) \Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2. \end{cases}$$

Применим теперь теорему 2.2. С этой целью определим нормированную выборочную медиану по формуле

$$\bar{M}_n = \begin{cases} \sqrt{n-1}\ X_{(m+1)}, & n = 2m + 1, \\ \sqrt{n}\ \dfrac{X_{(m)} + X_{(m+1)}}{2}, & n = 2m. \end{cases} \tag{3.11}$$

Теперь учитывая формулу

$$\int\limits_{(r(n-1)+1)^{-1}}^{\infty} \frac{\sqrt{2}\ p_1\ x\ |x|}{4\ p_0 \sqrt{n}}\ \frac{\varphi(x)}{\sqrt{(r(n-1)+1)}}\ \frac{1}{\sqrt{y}}\ dH_r(y) \equiv$$

$$\equiv \frac{\varphi(x)\ p_1\ x\ |x|\ \sqrt{2\ r}}{4p_0 \sqrt{rn(n-1)+n}} \frac{\Gamma(r-1/2)}{\Gamma(r)} + \mathcal{O}\Big(\frac{1}{n}\Big), \tag{3.12}$$

и, применяя теорему 2.2, получаем следующее утверждение.

Теорема 3.2. *Пусть выполнены условия A1 и A2 и при некотором $r > 0$ случайная величина N_n имеет распределение вида*

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

Тогда при $r > \frac{1}{2}$ для ф.р. нормированной выборочной медианы при $n \to \infty$ справедливо а.р. вида

$$\sup_x \Big| \mathsf{P}_\theta\Big(2p_0\ \Big(M_{N_n} - \sqrt{2m}\theta\Big) < x\Big) - \Phi(x) - \frac{\varphi(x)\ p_1\ x\ |x|\ \sqrt{2\ r}}{4p_0\sqrt{rn(n-1)+n}} \frac{\Gamma(r-1/2)}{\Gamma(r)} + \mathcal{O}\Big(\frac{1}{n}\Big) \Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2. \end{cases}$$

Следствие 3.2.

1. *В случае распределения Лапласа (1.5) для ф.р. нормированной выборочной медианы $\bar{M}_n$ справедливо асимптотическое разложение*

$$\sup_x \Big| \mathsf{P}_\theta\Big(\Big(\bar{M}_{N_n} - \sqrt{2m}\theta\Big) < x\Big) - \Phi(x) - \frac{\varphi(x)\ x\ |x|\ \sqrt{r}}{2\ \sqrt{rn(n-1)+n}} \frac{\Gamma(r-1/2)}{\Gamma(r)} \Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2. \end{cases}$$

2. *Для распределения Коши (1.4) справедливо а.р. вида*

$$\sup_x \Big| \mathsf{P}_\theta\Big(\frac{2}{\pi}\ \Big(\bar{M}_{N_n} - \sqrt{2m}\theta\Big) < x\Big) - \Phi(x) \Big| =$$

$$= \begin{cases} O\Big(\dfrac{\log n}{n}\Big), & r = 1, \\ O\Big(\dfrac{1}{n}\Big), & r > 1/2. \end{cases}$$

4. Распределение Лапласа

Рассмотрим распределение Лапласа с ф.р. $\Lambda_\theta(x)$ и плотностью

$$\lambda_\theta(x) = \frac{1}{\theta\sqrt{2}} \exp\Big\{-\frac{\sqrt{2}|x|}{\theta}\Big\}, \quad \theta > 0, \quad x \in \mathbb{R}.$$

В работе [13] была построена последовательность с.в. $N_n(s)$, зависящая от параметра $s \in \mathbb{N}$ следующего вида. Пусть $Y_1,\ Y_2, \ldots$ – независимые одинаково распределенные с.в., имеющие непрерывную ф.р. Определим с.в.

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

Хорошо известно, что так определенные с.в. имеют распределение вида

$$\mathsf{P}(N(s) \geq k) = \frac{s}{s+k-1}, \quad k \geq 1 \tag{4.1}$$

(см., например, [16] и [17]). Пусть теперь $N^{(1)}(s), N^{(2)}(s), \ldots$ – независимые одинаково распределенные с.в., имеющие распределение (4.1). Определим с.в.

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

тогда, как показано в работе [13],

$$\lim_{n\to\infty} \mathsf{P}\Big(\frac{N_n(s)}{n} < x\Big) = e^{-s/x}, \quad x > 0, \tag{4.2}$$

и для асимптотически нормальной статистики T_n справедливо соотношение

$$\mathsf{P}\Big(\sigma\sqrt{n}(T_{N_n(s)} - \mu) < x\Big) \longrightarrow \Lambda_{1/s}(x), \quad n \to \infty, \ \ x \in \mathbb{R},$$

где $\Lambda_{1/s}(x)$ – функция распределения Лапласа с параметром $\theta = 1/s$.

В работе [14] была получена следующая оценка скорости сходимости в соотношении (4.2):

$$\sup_{x\geq 0}\left|\,\mathsf{P}\Big(\frac{N_n(s)}{n} < x\Big) - e^{-s/x}\right| \leq \frac{C_s}{n}, \quad C_s > 0, \quad n \in \mathbb{N}. \tag{4.3}$$

Таким образом, из соотношения (4.3) следует, что случайный индекс $N_n(s)$ удовлетворяет условию 2.2 с

$$g(n) = n, \quad H(x) = e^{-s/x}, \quad m = 1, \tag{4.4}$$

$$h_1(x) \equiv 0, \quad C_2 = C_s > 0, \quad \beta = 1. \tag{4.5}$$

Рассмотрим более подробно величину $\mathsf{E}N_n^{-1}(s)$. Из определения с.в. $N_n(s)$ и равенства (4.1) имеем

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

поэтому

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

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

Для вычисления последнего интеграла используем формулу (см. [15], формула 856.12, стр. 184)

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

Получим

$$\mathsf{E}\,N_n^{-1}(s) \leq sn\frac{\Gamma(n-1)\Gamma(2)}{s^2\Gamma(n+1)} = \frac{1}{s(n-1)} = \mathcal{O}(n^{-1}).$$

Таким образом, учитывая теорему 1.1 и формулы

$$\int\limits_{n^{-1}}^{\infty}\Phi(x\sqrt{y})\,de^{-s/y} = \int\limits_{0}^{\infty}\Phi(x\sqrt{y})\,de^{-s/y} + \mathcal{O}\Big(\frac{1}{n}\Big) = \Lambda_{1/s}(x) + \mathcal{O}\Big(\frac{1}{n}\Big), \tag{4.5}$$

$$x|x|\int\limits_{n^{-1}}^{\infty}\varphi(x\sqrt{y})\,\sqrt{y}\,de^{-s/y} \;=\; x|x|\int\limits_{0}^{\infty}\varphi(x\sqrt{y})\sqrt{y}\,de^{-s/y} \;+\; \mathcal{O}\Big(\frac{1}{n}\Big) \;\equiv$$

$$\equiv\; l_s(x) \;+\; \mathcal{O}\Big(\frac{1}{n}\Big), \tag{4.6}$$

непосредственно получаем следующую теорему.

Теорема 4.1. *Пусть выполнены условия A1 и A2 и предположим, что при некотором $s \in \mathbb{N}$ с.в. $N_n(s)$ имеет распределение вида*

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

Тогда для ф.р. нормированной статистики медианы $M_{N_n(s)}$ справедливо а.р. вида

$$\sup_x\Big|\mathsf{P}_\theta\big(2p_0\sqrt{2k}\;(M_{N_n(s)}\;-\;\theta)\;<\;x\big)\;-\;\Lambda_{1/s}(x)\;-\;\frac{p_1 l_s(x)}{2p_0\sqrt{n}}\Big| \;=\; \mathcal{O}\Big(\frac{1}{n}\Big), \quad n\;\to\;\infty,$$

где функции $\Lambda_{1/s}(x)$ и $l_s(x)$ определены соответственно в соотношениях (4.5) *и* (4.6).

Следствие 4.1.

1. *Для распределения Лапласа (1.5) справедливо а. р. вида*

$$\sup_x\Big|\mathsf{P}_\theta\big(\sqrt{2k}\;(M_{N_n(s)}\;-\;\theta)\;<\;x\big)\;-\;\Lambda_{1/s}(x)\;-\;\frac{l_s(x)}{\sqrt{2}\sqrt{n}}\Big| \;=\; \mathcal{O}\Big(\frac{1}{n}\Big), \quad n\;\to\;\infty,$$

   где функции $\Lambda_{1/s}(x)$ и $l_s(x)$ определены соответственно в соотношениях (4.5) *и* (4.6).

2. *В случае распределения Коши (1.4) справедливо равенство*

$$\sup_x\Big|\mathsf{P}_\theta\Big(\frac{2\sqrt{2k}}{\pi}\;(M_{N_n(s)}\;-\;\theta)\;<\;x\Big)\;-\;\Lambda_{1/s}(x)\Big| \;=\; \mathcal{O}\Big(\frac{1}{n}\Big), \quad n\;\to\;\infty.$$

Для нормированной выборочной медианы $\bar{M}_n$ (см. (3.11)) с использованием формулы

$$\int\limits_{n^{-1}}^{\infty}\frac{x\;|x|}{\sqrt{y}}\;de^{-s/y}\;\equiv l_s(x), \tag{4.7}$$

получаем следующую теорему.

Теорема 4.2. *Пусть выполнены условия A1 и A2 и предположим, что при некотором $s \in \mathbb{N}$ с.в. $N_n(s)$ имеет распределение вида*

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

Тогда для ф.р. нормированной выборочной медианы $\bar{M}_{N_n(s)}$ справедливо а.р. вида

$$\sup_x\Big|\mathsf{P}_\theta\big(2p_0\;(\bar{M}_{N_n(s)}\;-\;\sqrt{2m}\theta)\;<\;x\big)\;-\;\Phi(x)\;-\;\frac{\sqrt{2}p_1 l_s(x)}{4p_0\sqrt{n}}\Big| \;=\; \mathcal{O}\Big(\frac{1}{n}\Big), \quad n\;\to\;\infty,$$

где функция $l_s(x)$ определена в соотношении (4.7).

Следствие 4.2.

1. *Для распределения Лапласа (1.3) справедливо а. р. вида*

$$\sup_x \Big|\mathsf{P}_\theta\big((\bar{M}_{N_n(s)} - \sqrt{2m}\theta) < x\big) - \Phi(x) - \frac{l_s(x)}{\sqrt{2}\sqrt{n}}\Big| = \mathcal{O}\Big(\frac{1}{n}\Big), \quad n \to \infty.$$

2. *В случае распределения Коши (1.4) справедливо равенство*

$$\sup_x \Big|\mathsf{P}_\theta\Big(\frac{2}{\pi}\,(\bar{M}_{N_n(s)} - \sqrt{2m}\theta) < x\Big) - \Phi(x)\Big| = \mathcal{O}\Big(\frac{1}{n}\Big), \quad n \to \infty.$$

Заключение

В работе рассмотрена выборочная медиана, построенная по выборкам случайного объема. Получены асимптотические разложения для функции распределения выборочной медианы в этом случае. Подробно рассмотрены случаи, когда предельными распределениями для функции распределения выборочной медианы выступают распределения Стьюдента (с произвольно малым числом степеней свободы) и Лапласа. Рассмотрен случай нормированной выборочной медианы. Приведены явные выражения для асимптотческих разложений в случаях, когда исходные наблюдения имеют «тяжелохвостые» распределения Коши и Лапласа.

Список литературы

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

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

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

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

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

[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., 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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. 39а–47.

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

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

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

Библиографическая ссылка

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.

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 Андреевича**
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.

ASYMPTOTIC EXPANSIONS FOR THE DISTRIBUTION FUNCTION OF SAMPLE MEDIAN BASED ON THE SAMPLE WITH RANDOM SIZE

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

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

Received 11.11.2015, revised 15.11.2015.

In the paper asymptotic expansions for the distribution function of sample median based on the sample with random size are considered. General theorem concerning the asymptotic expansion in this case is proved. Examples concerning Cauchy, Student and Laplace distributions are presented.

Keywords: sample median, sample with random size, asymptotic expansion, Student distribution, Cauchy distribution, Laplace distribution.

Bibliographic citation

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)