

УДК 519.24

В.П. Шуленин

**СВОЙСТВА АДАПТИВНЫХ ОЦЕНОК ХОДЖЕСА – ЛЕМАНА
В АСИМПТОТИКЕ И ПРИ КОНЕЧНЫХ ОБЪЕМАХ ВЫБОРКИ**

Рассматриваются адаптивные оценки модифицированных вариантов оценки Ходжеса – Лемана. Для построения адаптивных оценок параметра сдвига симметричных распределений используются выборочные оценки функционалов, описывающих степень «тяжести хвостов» распределений. Исследуются свойства оценок в асимптотике и при конечных объемах выборки в рамках различных супермоделей, описывающих отклонения от гауссовской модели в сторону «утяжеления хвостов» распределений.

Ключевые слова: робастные оценки, функция влияния, адаптивные оценки, метод статистических испытаний.

Оценка Ходжеса – Лемана (HL -оценка), предложенная в [1], среди большого числа оценок параметра положения случайной величины, принадлежит к группе лидеров по многим характеристикам. Для нормальной модели её абсолютная эффективность $AЭ(\Phi, HL) = 0,955$, то есть она проигрывает оптимальному выборочному среднему $\bar{X}$ менее 5 % в эффективности. Оценка Ходжеса – Лемана является B -робастной, её функция влияния ограничена и, следовательно, она «защищена» от наличия выбросов в выборке, её чувствительность к грубым ошибкам конечна и равна $\gamma^*(\Phi, HL) = \sqrt{\pi} \approx 1,77$, её предел устойчивости достаточно высок и равен $\varepsilon^*(HL) = 0,29$ (см., например [3]). В сравнении с этими характеристиками HL -оценки, выборочное среднее $\bar{X}$, являясь оптимальной оценкой параметра сдвига θ нормального распределения, имеет абсолютную эффективность $AЭ(\Phi, \bar{X}) = 1$, однако она теряет свойства оптимальности даже при небольших отклонениях от нормального распределения, её функция влияния неограниченна и $\gamma^*(\Phi, \bar{X}) = \infty$, $\varepsilon^*(\bar{X}) = 0$. Традиционно используемая альтернативная оценка параметра сдвига θ в виде выборочной медианы $\bar{X}_{1/2}$ является B -робастной, имеет максимально возможный предел устойчивости $\varepsilon^*(\bar{X}_{1/2}) = 1/2$, однако её абсолютная эффективность при нормальном распределении очень низкая и равна $AЭ(\Phi, \bar{X}_{1/2}) = 0,637$. Изучение характеристик HL -оценок в различных супермоделях, описывающих отклонения от нормального распределения в сторону «утяжеления хвостов» распределений показало, что свойства HL -оценок могут быть существенно улучшены путем различных модификаций этих оценок (см. [5, 9]), что и послужило основой для построения адаптивных оценок модифицированных вариантов оценки Ходжеса – Лемана.

В работе [9] показано, что модифицированные оценки Ходжеса – Лемана входят в класс обобщенных L -оценок, асимптотические свойства которых описаны в [10]. В данной работе рассматриваются адаптивные оценки для обобщенных ва-

риантов оценок Ходжеса – Лемана. Изучаются свойства оценок в асимптотике и методом статистических испытаний при конечных объемах выборки.

1. Модифицированные варианты оценок Ходжеса – Лемана

Пусть $X_1, \dots, X_n$ – последовательность н.о.р. случайных величин с функцией распределения $F(x - \theta)$. Предполагаем, что ф.р. F непрерывна, симметрична, то есть $F \in \mathfrak{S}_S = \{F : F(x) = 1 - F(-x), \forall x \in R^1\}$ и имеет плотность $f(x)$, $x \in R^1$. Оценка Ходжеса – Лемана параметра сдвига θ определяется в виде медианы средних Уолша $(X_i + X_j)/2$, $1 \leq i < j \leq n$, общее число которых равно $n(n+1)/2$, и записывается в виде

$$HL = med\{(X_i + X_j)/2, 1 \leq i < j \leq n\}, \quad (1)$$

где символ «*med*» обозначает выборочную медиану. Свойства HL -оценки хорошо изучены как в асимптотике, так и при конечных объемах выборки, и подробно описаны в литературе (см., например, [1 – 3]). Для описания модифицированных оценок Ходжеса – Лемана обозначим через $X_{(1)}, \dots, X_{(n)}$ упорядоченную статистику выборки $X_1, \dots, X_n$, и пусть $X_{([\alpha n]+1)}, \dots, X_{(n-[\beta n])}$ обозначает $\alpha\beta$ -урезанную выборку, α и β заданные пропорции урезания выборки, причем $0 \leq \alpha, \beta \leq 1/2$. Обозначим через $C_{\alpha\beta}$ множество m -наборов индексов $(i_1, \dots, i_m)$, удовлетворяющих условию $\{[\alpha n] + 1 \leq i_1 < \dots < i_m \leq n - [\beta n]\}$, то есть множество $C_{\alpha\beta}$ определяется в виде

$$C_{\alpha\beta} = \{(i_1, \dots, i_m) : [\alpha n] + 1 \leq i_1 < \dots < i_m \leq n - [\beta n]\}. \quad (2)$$

Модифицированные оценки Ходжеса – Лемана параметра θ , основанные на $\alpha\beta$ -урезанной выборке, были предложены в [5] и названы обобщенными оценками Ходжеса – Лемана. Эти оценки записываются в виде

$$HL_{\alpha\beta}(m) = med\{(X_{(i_1)} + \dots + X_{(i_m)})/m : (i_1, \dots, i_m) \in C_{\alpha\beta}\}. \quad (3)$$

В классе этих обобщенных оценок выделим два случая. Пусть обе пропорции урезания выборки равны нулю, то есть $\alpha = \beta = 0$. В этом случае обобщенные оценки Ходжеса – Лемана записываются в виде

$$HL(m) = med\{(X_{i_1} + \dots + X_{i_m})/m, 1 \leq i_1 < \dots < i_m \leq n\}. \quad (4)$$

Далее, пусть $m = 2$ и $\alpha = \beta$. В этом случае α -урезанный вариант оценки Ходжеса – Лемана, предложенный в [4], записывается в виде

$$HL_{\alpha} = med\{(X_{(i)} + X_{(j)})/2, k+1 \leq i \leq j \leq n-k\}, \quad k = [\alpha n], \quad 0 \leq \alpha < 1/2. \quad (5)$$

Асимптотические свойства $HL_{\alpha\beta}(m)$ -оценок вида (3) описаны в работе [9]. Некоторые из этих результатов для оценок (4) и (5) понадобятся ниже.

1.1. Для описания свойств $HL(m)$ -оценок вида (4) обозначим через $F^{(m)*}$ m -кратную свертку ф.р. F с собой, то есть $F^{(m)*} = F * F * \dots * F$ – (m раз), причем $F^{(1)*}(x) = F(x)$, $F^{(2)*}(x) = F * F = \int F(x-y)dF(y)$ и $F^{(m)*} = F^{(m-1)*} * F$. Плот-

ность ф.р. $F^{(m)*}(x)$ обозначим через $f^{(m)*}(x)$, то есть $f^{(m)*}$ является m -кратной сверткой плотности f для ф.р. F . Отметим, что с учетом введенных обозначений, функционал $\tilde{T}(F) = \theta$ для $F \in \mathfrak{F}_S$, соответствующий оценке (4), задается неявно выражением вида

$$\int F^{(m-1)*}(m\tilde{T}(F) - x)dF(x) = 1/2 \quad \text{или} \quad F^{(m)*}(m\tilde{T}(F)) = 1/2. \quad (6)$$

В работе [9] показано, что для $F \in \mathfrak{F}_S$ и $\int f^{(m-1)*}(x)dF(x) > 0$ $HL(m)$ -оценки асимптотически нормальны, то есть случайные величины $\sqrt{n}\{(HL(m) - \tilde{T}(F))/\sigma(F, HL(m))\}$ имеют асимптотически стандартное нормальное распределение, где асимптотическая дисперсия $\sqrt{n}HL(m)$ -оценки вычисляется по формуле

$$\begin{aligned} \sigma^2(F, HL(m)) &= \int IF^2(x; F, HL(m))dF(x) = \\ &= \frac{\int \left(\int_{-\infty}^x f^{(m-1)*}(y)dy - (1/2) \right)^2 dF(x)}{\left(\int_{-\infty}^{\infty} f^{(m-1)*}(x)dF(x) \right)^2}, \quad F \in \mathfrak{F}_S. \end{aligned} \quad (7)$$

Функция влияния $HL(m)$ -оценок определяется выражением

$$IF(x; F, HL(m)) = \frac{F^{(m-1)*}(x) - (1/2)}{\int_{-\infty}^{+\infty} f^{(m-1)*}(x)dF(x)}, \quad x \in R^1. \quad (8)$$

Для нормального распределения, то есть при $F = \Phi$, функция влияния $HL(m)$ -оценок записывается в виде

$$\begin{aligned} IF(x; \Phi, HL(m)) &= \frac{\Phi^{(m-1)*}(x) - (1/2)}{\int \phi^{(m-1)*}(x)d\Phi(x)} = \sqrt{m2\pi} \left[\Phi\left(\frac{x}{\sqrt{m-1}}\right) - \frac{1}{2} \right] = \\ &= \sqrt{m\pi/2} \tilde{\Phi}(x/\sqrt{2(m-1)}), \end{aligned}$$

где $\tilde{\Phi}(x) = 2\Phi(\sqrt{2}x) - 1$. Асимптотическая дисперсия $\sqrt{n}HL(m)$ -оценки вычисляется по формуле

$$\begin{aligned} \sigma^2(\Phi, HL(m)) &= \int IF^2(x; \Phi, HL(m))d\Phi(x) = m\sqrt{\pi/2} \int_0^{\infty} \tilde{\Phi}^2(x/\sqrt{2(m-1)})e^{-x^2/2}dx = \\ &= m \arctg(1/\sqrt{m^2-1}). \end{aligned}$$

Чувствительность $HL(m)$ -оценки к грубым ошибкам равна $\gamma^*(\Phi, HL(m)) = \sqrt{m\pi/2}$, и чувствительность $HL(m)$ -оценки к локальным изменениям наблюдений вычисляется по формуле $\lambda^*(\Phi, HL(m)) = \sqrt{m/(m-1)}$. Численные значения характеристик $HL(m)$ -оценок при $F = \Phi$ приведены в табл. 1.

Таблица 1

Числовые характеристики робастности $HL(m)$ -оценок при $F = \Phi$

Характеристики	m						
	1	2	3	4	5	10	$m \approx n \rightarrow \infty$
$\gamma^*(\Phi, HL(m))$	1,253	1,772	2,171	2,507	2,802	3,693	∞
$\lambda^*(\Phi, HL(m))$	∞	1,414	1,255	1,155	1,118	1,054	1,000
$\sigma^2(\Phi, HL(m))$	1,571	1,047	1,019	1,011	1,007	1,002	1,000

Отметим, что при гауссовском распределении чувствительность $HL(m)$ -оценок к грубым ошибкам возрастает с увеличением m и достигает бесконечного значения (это является следствием неограниченной функции влияния при $m \approx n \rightarrow \infty$). Далее, чувствительность $HL(m)$ -оценок к локальным изменениям наблюдений и их асимптотическая дисперсия убывают до своих предельных значений, соответственно равных единице. Отметим также, что глобальная характеристика робастности оценок, называемая «точкой срыва» или пределом устойчивости, для $HL(m)$ -оценок вычисляется по формуле $\varepsilon^*(HL(m)) = 1 - 2^{-1/m}$. В частности, при $m = 2$ получаем известный результат для точки срыва оценки Ходжеса – Лемана в виде $\varepsilon^*(HL) = 1 - 2^{-1/2} = 0,29$ (детали см. в [3, 6]). Отметим, что в классе обобщенных оценок Ходжеса – Лемана существует асимптотически эффективная оценка параметра сдвига θ в одновыборочной задаче для симметричных распределений, то есть, если ф.р. $F \in \mathfrak{S}_S$ и её плотность f является решением дифференциального уравнения

$$d^2 \{-\ln f(x)\} / dx^2 = a f^{(m-1)*}(x), \quad (9)$$

где a – постоянная величина, тогда $\sigma^2(F, HL(m)) = 1 / I(f)$. Здесь $I(f)$ – количество информации Фишера относительно параметра сдвига θ распределения с плотностью f . Убедимся в справедливости данного факта. Пусть

$$\psi(x) = -f'(x) / f(x) = d\{-\ln f(x)\} / dx,$$

тогда уравнение (9) запишется в виде

$$(1/a)\psi'(x) = f^{(m-1)*}(x).$$

Далее, пусть $a = 2 \int_{-\infty}^0 \psi'(x) dx < \infty$, тогда

$$\int_{-\infty}^{\infty} f^{(m-1)*}(x) dF(x) = (1/a) \int_{-\infty}^{\infty} \psi'(x) dF(x) = (1/a) I(f),$$

а внутренний интеграл в числителе (7) равен

$$\begin{aligned} \int_{-\infty}^x f^{(m-1)*}(y) dy &= (1/a) \left\{ \int_{-\infty}^0 \psi'(y) dy + \int_0^x d\psi(y) \right\} = (1/2) + (1/a) [\psi(x) - \psi(0)] = \\ &= (1/2) + (1/a) \psi(x). \end{aligned}$$

Учитывая, что $\int_{-\infty}^{\infty} \psi^2(x) dF(x) = I(f)$, окончательно из (7) получаем

$$\begin{aligned}\sigma^2(F, HL(m)) &= \frac{\int_{-\infty}^{\infty} \left(\int_{-\infty}^x f^{(m-1)*}(y) dy - (1/2) \right)^2 dF(x)}{\left(\int_{-\infty}^{\infty} f^{(m-1)*}(x) dF(x) \right)^2} = \\ &= \frac{(1/a)^2 \int_{-\infty}^{\infty} \psi^2(x) dF(x)}{(1/a)^2 I^2(f)} = \frac{1}{I(f)}.\end{aligned}\quad (10)$$

Отметим, что решением приведенного уравнения $d^2\{-\ln f(x)\}/dx^2 = a f^{(m-1)*}(x)$ при $m = 2$ является логистическая плотность $f(x) = \exp(-x)/\{1 + \exp(-x)\}^2$. Для этого распределения оценка Ходжеса – Лемана вида (1) является асимптотически эффективной оценкой параметра сдвига θ . Для больших объемов выборки n и при $m \approx n$ $HL(m)$ -оценки совпадают с выборочным средним $\bar{X}$, которое является эффективной оценкой параметра сдвига θ при нормальном распределении.

1.2. Асимптотические свойства HL_α -оценок вида (5) описаны в [6]. Приведем необходимые результаты. Отметим, что в отличие от HL -оценки вида (1), которая вычисляется по исходной выборке $X_1, \dots, X_n$ путем определения медианы средних значений Уолша $(X_i + X_j)/2$, $1 \leq i \leq j \leq n$, α – урезанная оценка Ходжеса – Лемана (HL_α -оценка вида (5)) вычисляется на основе упорядоченной статистики $X_{(1)}, \dots, X_{(n)}$, из которой предварительно удалены $k = [\alpha n]$ наименьших и наибольших порядковых статистик. Функционал $T_\alpha(F)$, выборочной оценкой которого является оценка вида (5), задается неявно выражением

$$\int_{\alpha}^{1-\alpha} F[2T_\alpha(F) - F^{-1}(t)] dt = (1 - 2\alpha)/2, \quad 0 \leq \alpha < 1/2. \quad (11)$$

В работе [3] показано, что для $F \in \mathfrak{S}_S$ и $\int_{\alpha}^{1-\alpha} f(F^{-1}(t)) dt > 0$ HL_α -оценки асимптотически нормальны, то есть случайные величины $\sqrt{n}\{(HL_\alpha - T_\alpha(F))/\sigma(F, HL_\alpha)\}$ имеют асимптотически стандартное нормальное распределение, где функционал $T_\alpha(F) = \theta$ определен в (11) и асимптотическая дисперсия $\sqrt{n}HL_\alpha$ -оценки вычисляется по формуле

$$\sigma_F^2(HL_\alpha) = \frac{(1 + 4\alpha)(1 - 2\alpha)^2}{12 \left(\int_{\alpha}^{1-\alpha} f(F^{-1}(t)) dt \right)^2}. \quad (12)$$

Функция влияния HL_α -оценки записывается в виде

$$IF(x; F, HL_\alpha) = \frac{1}{2 \int_{\alpha}^{1-\alpha} f(F^{-1}(t)) dt} \begin{cases} (1 - 2\alpha) \operatorname{sign}(x), & |x| > F^{-1}(1 - \alpha) \\ (2F(x) - 1), & |x| \leq F^{-1}(1 - \alpha) \end{cases}, \quad x \in R^1. \quad (13)$$

Ниже (см. табл. 4 и 6) приводятся вычисления асимптотических дисперсий модифицированной оценки Ходжеса – Лемана (5) по формуле (12) для супермоделей, описывающих различные отклонения от нормального распределения.

2. Описание супермоделей

Изучение эффективности многих статистических процедур при изменении распределения вероятности наблюдений в некотором заданном классе (в рамках заданной супермодели) показывает (см., например, [3]), что эффективность часто зависит монотонно от некоторых общих свойств распределений. В частности, к таким общим свойствам относится «затянутость хвостов» распределений или «тяжесть хвостов» распределений. В литературе (см., например, [3]) описаны различные подходы для упорядочивания распределений в заданном классе по степени тяжести хвостов. Следуя работе [13], используем следующее определение. Пусть $F, G \in \mathfrak{Z}_{S|0}$. Говорят, что хвосты ф.р. F легче хвостов ф.р. G (или G имеет хвосты тяжелее, чем F , и это записывают в виде $F <_S G$), если функция $G^{-1}(F(x))$ выпуклая для $x \geq \theta$, где θ – точка симметрии для F и G . Отметим, что если $F <_S G$, то также говорят, что распределения F и G являются S -упорядоченными.

2.1. Рассмотрим супермодель $\mathfrak{Z}_S^*$ в виде конечного семейства заданных симметричных распределений, то есть

$$\mathfrak{Z}_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}\}, \quad (14)$$

где $F_{(1)} = \Phi$ – стандартное нормальное распределение, информация Фишера $I(f_{(1)}) = 1$; $F_{(2)}$ – логистическое, $I(f_{(2)}) = 1/3$; $F_{(3)}$ – Лапласа, $I(f_{(3)}) = 1$; $F_{(4)}$ – Коши, $I(f_{(2)}) = 1/2$. Можно убедиться (см. [12]), что данная супермодель $\mathfrak{Z}_S^*$ содержит S -упорядоченные распределения, причем выполняется выражение

$$F_{(1)} <_S F_{(2)} <_S F_{(3)} <_S F_{(4)}. \quad (15)$$

Следуя работе [14], рассмотрим меру «тяжести хвостов» распределения $F(x)$, $x \in R^1$, в виде функционала $Q_F(v, \mu)$, определяемого выражением

$$Q_F(v, \mu) = \frac{(1/v) \left\{ \int_{1-v}^1 F^{-1}(t) dt - \int_0^v F^{-1}(t) dt \right\}}{(1/\mu) \left\{ \int_{1-\mu}^1 F^{-1}(t) dt - \int_0^\mu F^{-1}(t) dt \right\}}, \quad 0 < v < \mu \leq 0,5. \quad (16)$$

Отметим, что для класса симметричных распределений $\mathfrak{Z}_S$ функционал $Q_F(v, \mu)$ записывается в виде

$$Q_F(v, \mu) = \left\{ (1/v) \int_0^v F^{-1}(t) dt \right\} / \left\{ (1/\mu) \int_0^\mu F^{-1}(t) dt \right\}, \quad 0 < v < \mu \leq 0,5, \quad F \in \mathfrak{Z}_S. \quad (17)$$

Например, для нормального распределения функционал $Q_F(v, \mu)$ равен $Q_F(v, \mu) = \mu \phi[\Phi^{-1}(v)] / v \phi[\Phi^{-1}(\mu)]$. Численные значения функционала $Q_F(v, \mu)$ для значений $v = 0,2$, $\mu = 0,5$ и S -упорядоченных распределений из супермодели $\mathfrak{S}_S^*$ приведены в табл. 2.

Таблица 2

Значения функционала $Q_F(0,2;0,5)$ для $F \in \mathfrak{S}_S^*$

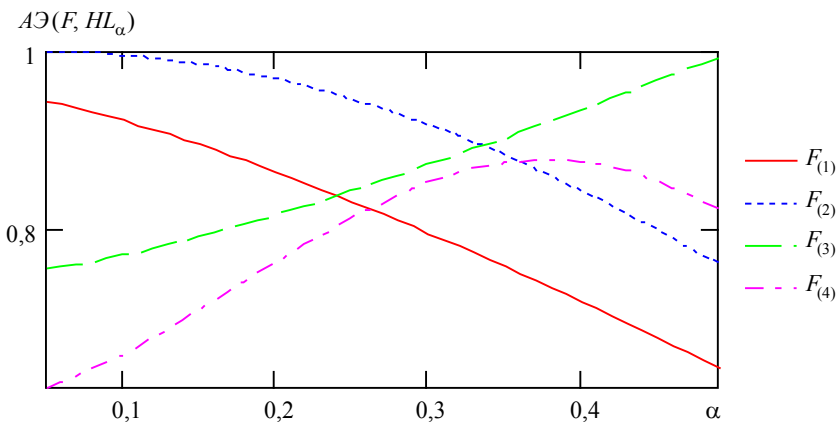
Ф.р.	$F_{(1)}$	$F_{(2)}$	$F_{(3)}$	$F_{(4)}$
$Q_F(0,2;0,5)$	1,755	1,805	1,916	2,500

Таким образом, в рамках супермодели $\mathfrak{S}_S^*$ выполняется выражение

$$F <_S G \Rightarrow Q_F(v, \mu) < Q_G(v, \mu).$$

Асимптотические дисперсии $\sqrt{n}HL_\alpha$ -оценок, вычисленные для распределений из супермодели $\mathfrak{S}_S^*$ по формуле (12), приведены в табл. 6.

Абсолютные эффективности $\sqrt{n}HL_\alpha$ -оценок, вычисленные для распределений супермодели $\mathfrak{S}_S^*$ по формуле $AЭ(F, HL_\alpha) = \{I(f) \cdot \sigma^2(F, HL_\alpha)\}^{-1}$, приведены на рис. 1.

Рис. 1. Зависимости $AЭ(F, HL_\alpha)$ от параметра α для $F \in \mathfrak{S}_S^*$

Из рисунка наглядно видно, что качество HL_α -оценок существенно зависит от параметра α . Для распределений с «тяжелыми хвостами» параметр α следует выбирать близким к 0,5. Например, для распределения Коши при $\alpha = 0,375$ достигается максимальная абсолютная эффективность. Для нормального распределения следует выбрать параметр $\alpha = 0$. Эти качественные соображения будут использованы ниже при адаптивном выборе параметра α , основываясь на выборочной оценке функционала $Q_F(v, \mu)$, построенной по исходной выборке $X_1, \dots, X_n$.

2.2. Рассмотрим супермодель в виде семейства распределений Стюдента $\mathfrak{F}_r \in \mathfrak{S}_S$, для которого ф.р. $F_r(x)$ имеет плотность распределения $f_r(x)$ с r степенями свободы и записывается в виде

$$f_r(x) = A(r)(1 + (x^2/r))^{-(r+1)/2}, \quad x \in R^1, \quad A(r) = \Gamma((r+1)/2) / \sqrt{r\pi} \Gamma(r/2).$$

Можно убедиться, что функционал $Q_F(v; \mu)$ для распределений Стюдента, то есть для $F(x) = F_r(x)$, вычисляется по формуле

$$Q_{F_r}(v, \mu) = \frac{\mu}{v} \cdot \frac{f_{r-2}[\sqrt{(r-2)/r} \cdot F_r^{-1}(v)]}{f_{r-2}[\sqrt{(r-2)/r} \cdot F_r^{-1}(\mu)]}.$$

В частности, для $v = 0,2$ и $\mu = 0,5$ с учетом того, что $F_r^{-1}(1/2) = 0$ и $f_{r-2}(0) = A(r-2)$, получаем

$$\begin{aligned} Q_{F_r}(0, 2; 0, 5) &= 5 f_{r-2}[\sqrt{(r-2)/r} F_r^{-1}(0, 2)] / 2 A(r-2) = \\ &= 2,5 \cdot (1 + [F_r^{-1}(0, 2)]^2 / r)^{-(r-1)/2}. \end{aligned}$$

Численные значения функционала $Q_F(0, 2; 0, 5)$ приведены в табл. 3.

Таблица 3

Значения функционала $Q_F(0, 2; 0, 5)$ для $F \in \mathfrak{F}_r$

r	1	2	3	4	5	7	9	25	$r \rightarrow \infty$
$Q_F(0, 2; 0, 5)$	2,50	2,00	1,90	1,85	1,83	1,81	1,79	1,77	1,75

Итак, для семейства распределений Стюдента функционал $Q_F(v; \mu)$ монотонно зависит от числа степеней свободы r , то есть выполняется неравенство $Q(F_{r_1}; v, \mu) < Q(F_{r_2}; v, \mu)$, для $r_2 < r_1$.

В работе [3] показано, что для $F_r(x) \in \mathfrak{F}_r$ асимптотическая дисперсия $\sqrt{n}HL_\alpha$ -оценок для $F_r(x) \in \mathfrak{F}_r$ вычисляется по формуле

$$\sigma^2(F_r, HL_\alpha) = \frac{r\pi}{12} \left[\frac{\Gamma(r+1)}{\Gamma(r+(1/2))} \right]^2 \left[\frac{\Gamma(r/2)}{\Gamma((r+1)/2)} \right]^4 \frac{(1+4\alpha)(1-2\alpha)^2}{\{2F_{2r+1}[\sqrt{(2r+1)/r} F_r^{-1}(1-\alpha)] - 1\}^2}. \quad (18)$$

В частности, из формулы (18) при $\alpha = 0$ получаем формулу для вычисления асимптотической дисперсии $\sqrt{n}HL$ -оценки в виде

$$\sigma^2(F_r, HL) = \frac{r\pi}{12} \cdot \left[\frac{\Gamma(r+1)}{\Gamma(r+(1/2))} \right]^2 \left[\frac{\Gamma(r/2)}{\Gamma((r+1)/2)} \right]^4. \quad (19)$$

Далее, из (18) при $\alpha \rightarrow 1/2$ следует формула для вычисления асимптотической дисперсии выборочной медианы $\bar{X}_{1/2}$. Для $\sqrt{n}\bar{X}_{1/2}$ -оценки асимптотическая дисперсия для $F_r(x) \in \mathfrak{F}_r$ вычисляется по формуле

$$\sigma^2(F_r, \bar{X}_{1/2}) = \frac{\pi r}{4} \cdot \left[\frac{\Gamma(r/2)}{\Gamma((r+1)/2)} \right]^2. \quad (20)$$

Отметим, что для выборочного среднего $\bar{X}$ асимптотическая дисперсия $\sqrt{n}\bar{X}$ -оценки для $F_r(x) \in \mathfrak{F}_r$ вычисляется по формуле

$$\sigma_{F_r}^2(\bar{X}) = A(r) \int_{-\infty}^{\infty} x^2 (1 + x^2/r)^{\frac{r+1}{2}} dx = r/(r-2), \quad r > 2. \quad (21)$$

Численные значения асимптотических значений дисперсий оценок, рассчитанные по приведенным формулам (19) – (21) для $F_r(x) \in \mathfrak{F}_r$ при различных степенях свободы r , приведены в табл. 4.

Таблица 4

Асимптотическая дисперсия и абсолютная эффективность оценок для $F_r(x) \in \mathfrak{F}_r$

Ст. св. r $I(f)$	$r=1$, $I(f_1)=0,50$	$r=2$ $I(f_2)=0,60$	$r=3$ $I(f_3)=0,67$	$r=5$ $I(f_5)=0,75$
$\sigma^2(F_r, HL)$	3,29 (0,61)	1,92 (0,87)	1,58 (0,95)	1,34 (0,99)
$\sigma^2(F_r, \bar{X}_{1/2})$	2,47 (0,81)	2,00 (0,81)	1,85 (0,80)	1,73 (0,77)
$\sigma^2(F_r, \bar{X})$	∞ (0,00)	∞ (0,00)	3,00 (0,50)	1,67 (0,80)
Ст. св. r $I(f)$	$r=7$ $I(f_7)=0,80$	$r=9$ $I(f_9)=0,83$	$r \rightarrow \infty$, $I(f_\infty)=1,00$	
$\sigma^2(F_r, HL)$	1,25 (1,00)	1,20 (1,00)	1,047 (0,96)	
$\sigma^2(F_r, \bar{X}_{1/2})$	1,67 (0,75)	1,66 (0,72)	1,571 (0,64)	
$\sigma^2(F_r, \bar{X})$	1,40 (0,89)	1,28 (0,94)	1,000 (1,00)	

В этой таблице в скобках приведена абсолютная эффективность оценок $\hat{\theta}$, вычисляемая по формуле

$$A\mathfrak{D}(F_r, \hat{\theta}) = \{I(f_r) \cdot \sigma_{F_r}^2(\hat{\theta})\}^{-1},$$

где $I(f_r)$ – информация Фишера. Из данных таблицы (4) следует, что оценка Ходжеса – Лемана обеспечивает высокую абсолютную эффективность в рамках всего семейства распределений Стьюдента для числа степеней свободы $r \geq 3$.

Асимптотическая относительная эффективность HL_α -оценки относительно HL -оценки Ходжеса – Лемана, вычисляемая по формуле

$$AO\mathfrak{D}_F(HL_\alpha : HL) = \sigma^2(F, HL) / \sigma^2(F, HL_\alpha), \quad \text{для } F_r(x) \in \mathfrak{F}_r,$$

записывается в виде

$$AO\mathfrak{D}_{F_r}(HL_\alpha : HL) = \frac{\{2F_{2r+1}[\sqrt{(2r+1)/r} F_r^{-1}(1-\alpha)] - 1\}^2}{(1+4\alpha)(1-2\alpha)^2}, \quad (22)$$

где $F_r(x)$ – табулированная функция распределения Стьюдента с r степенями свободы и F_r^{-1} также табулированная квантильная функция. Из формулы (22) получаем при $r \rightarrow \infty$ асимптотическую относительную эффективность $AO\mathfrak{D}_F(HL_\alpha : HL)$ для нормального распределения, то есть для $F = \Phi$ имеем

$$AO\mathfrak{D}_\Phi(HL_\alpha : HL) = \frac{\{2\Phi[\sqrt{2}\Phi^{-1}(1-\alpha)] - 1\}^2}{(1+4\alpha)(1-2\alpha)^2}. \quad (23)$$

Из формулы (22) получаем при $r = 1$ асимптотическую относительную эффективность $AO\mathfrak{E}_F(HL_\alpha : HL)$ для распределения Коши, то есть для $F = F_1$ имеем

$$AO\mathfrak{E}_{F_1}(HL_\alpha : HL) = \frac{\{2F_3[\sqrt{3} F_1^{-1}(1-\alpha)] - 1\}^2}{(1+4\alpha)(1-2\alpha)^2} = \frac{\{1-2\alpha + \pi^{-1} \sin(\pi(1-2\alpha))\}^2}{(1+4\alpha)(1-2\alpha)^2}. \quad (24)$$

Например, при $\alpha = 0,4$ из (23) и (24) получаем $AO\mathfrak{E}_\Phi(HL_{0,4} : HL) = 0,75$ и $AO\mathfrak{E}_{F_1}(HL_{0,4} : HL) = 1,44$.

2.3. Рассмотрим теперь гауссовскую модель с масштабным засорением, которая обычно используется при изучении влияния выбросов в выборке на оценку и определяется в виде

$$\mathfrak{F}_{\varepsilon,\tau}(\Phi) = \{F : F_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)\}, \quad 0 \leq \varepsilon \leq 1, \quad \tau \geq 1. \quad (25)$$

Можно убедиться (см. [3]), что функционал $Q_F(v; \mu)$ для распределений $F \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$ вычисляется по формуле

$$Q_{F_{\varepsilon,\tau}}(v; \mu) = \frac{\mu}{v} \cdot \frac{(1-\varepsilon)\phi[F_{\varepsilon,\tau}^{-1}(v)] + \varepsilon\tau\phi[F_{\varepsilon,\tau}^{-1}(v)/\tau]}{(1-\varepsilon)\phi[F_{\varepsilon,\tau}^{-1}(\mu)] + \varepsilon\tau\phi[F_{\varepsilon,\tau}^{-1}(\mu)/\tau]}, \quad (26)$$

где ϕ – стандартная нормальная плотность. В частности, для значений $v = 0,2$ и $\mu = 0,5$ функционал $Q_F(0,2;0,5)$ для $F \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$ вычисляется по формуле

$$Q_{F_{\varepsilon,\tau}}(0,2;0,5) = \frac{5\sqrt{2\pi}\{(1-\varepsilon)\phi(x^*) + \varepsilon\tau\phi(x^*/\tau)\}}{2(1-\varepsilon + \varepsilon\tau)}, \quad (27)$$

где $x^* = F_{\varepsilon,\tau}^{-1}(0,2)$ – квантиль уровня $v = 0,2$ функции распределения $F_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)$. Отметим, что для нормального распределения из формул (26) и (27) при $\varepsilon = 0$ получаем

$$Q_\Phi(v, \mu) = \mu\phi[\Phi^{-1}(v)]/v\phi[\Phi^{-1}(\mu)],$$

$$Q_\Phi(0,2;0,5) = 2,5 \cdot \sqrt{2\pi} \phi[\Phi^{-1}(0,2)] \approx 1,75.$$

Численные значения функционала $Q_F(0,2;0,5)$ для $F \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$ приведены в табл. 5.

Таблица 5

Значения функционала $Q_F(0,2;0,5)$ для $F \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$

τ	ε								
	0,001	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35
$\tau = 3$	1,756	1,764	1,797	1,830	1,865	1,876	1,892	1,903	1,911
$\tau = 5$	1,757	1,777	1,850	1,916	1,962	1,996	2,020	2,036	2,046
$\tau = 7$	1,758	1,790	1,898	1,985	2,042	2,080	2,107	2,124	2,134
$\tau = 10$	1,760	1,809	1,959	2,064	2,126	2,166	2,192	2,208	2,218

В работе [3] показано, что асимптотическая дисперсия $\sqrt{n}HL_\alpha$ -оценки для $F(x) \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$ вычисляется по формуле

$$\sigma^2(F_{\varepsilon,\tau}, HL_\alpha) = \frac{\pi}{3} \cdot \frac{(1+4\alpha)(1-2\alpha)^2}{[(1-\varepsilon)^2 A + (2\sqrt{2}B\varepsilon(1-\varepsilon)/\sqrt{1+\tau^2}) + (\varepsilon^2/\tau)C]^2}, \quad (28)$$

где $A = 1 - 2\Phi(x^* \sqrt{2})$, $B = 1 - 2\Phi(x^* \sqrt{1+\tau^2}/\tau)$, $C = 1 - 2\Phi(x^* \sqrt{2}/\tau)$, $x^* = F_{\varepsilon,\tau}^{-1}(\alpha)$ – квантиль уровня α для ф.р. $F_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)$. Численные расчеты по формуле (28) приведены в табл. 8. Зависимости абсолютной эффективности $AЭ(F_{\varepsilon,\tau}, HL_\alpha) = \{\sigma^2(F_{\varepsilon,\tau}, HL_\alpha)I(f_{\varepsilon,\tau})\}^{-1}$ от параметра α при различных значениях параметра ε и фиксированном значении $\tau = 3$ приведены на рис. 2.

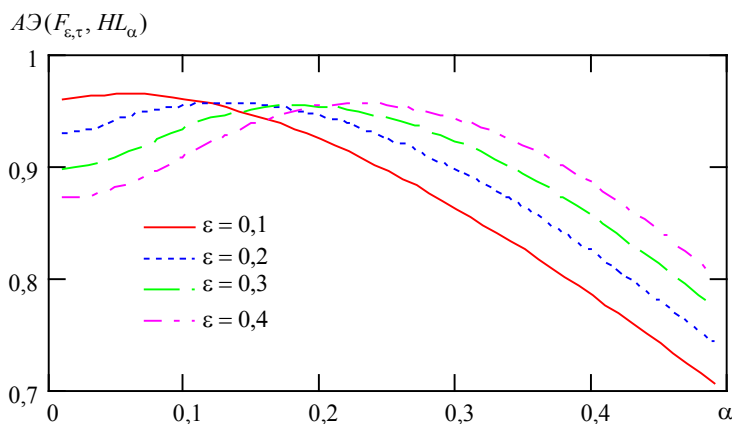


Рис. 2. Зависимости $AЭ(F_{\varepsilon,\tau}, HL_\alpha)$ от параметра α при различных ε и $\tau = 3$

Асимптотическая относительная эффективность оценки HL_α -оценки относительно HL -оценки Ходжеса – Лемана, вычисляемая по формуле $AOЭ_F(HL_\alpha : HL) = \sigma^2(F, HL) / \sigma^2(F, HL_\alpha)$, для $F(x) \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$, приведена на рис. 3.

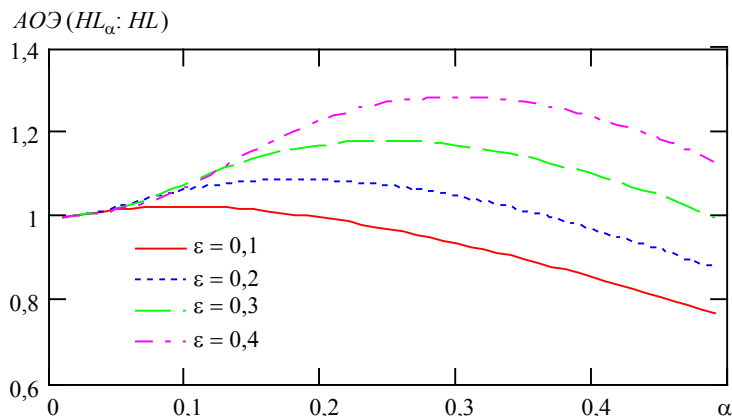


Рис. 3. Зависимости $AOЭ_F(HL_\alpha : HL)$, $F(x) \in \mathfrak{F}_{\varepsilon,\tau}(\Phi)$, от параметра α при разных ε и $\tau = 5$

Из рис. 3 следует, что относительные достоинства HL_α -оценок возрастают при «утяжелении хвостов» распределений, то есть при увеличении ε -пропорции засорения грубыми ошибками исходной выборки. Этот факт является проявлением более общего свойства R_α -оценок, рассмотренных в [3].

3. Адаптивные HL_α -оценки

Изучение асимптотических свойств HL_α -оценок показало, что качество этих оценок существенно зависит от выбора пропорции урезания исходной выборки, которая характеризуется параметром α , $0 \leq \alpha < 1/2$. Например, если мы заинтересованы в уменьшении асимптотической дисперсии, то для распределений «близких по затянутости хвостов» к нормальному величину α следует выбирать близкой к нулю, для распределений с «тяжелыми хвостами» (например, Лапласа, Коши) параметр α следует выбирать близким к $1/2$. Таким образом, выбор параметра α можно связать с поведением функционала $Q_F(v; \mu)$ вида (16), который характеризует степень затянутости «хвостов» распределений при их изменении в заданной супермодели. Однако на практике функция распределения F наблюдений $X_1, \dots, X_n$ обычно неизвестна, поэтому естественно использовать вместо функционала $Q_F(v; \mu)$ его выборочную оценку, построенную по исходной выборке $X_1, \dots, X_n$. Выборочная оценка $Q(F_n)$ функционала $Q_F(v; \mu)$ вида (16), построенная по выборке $X_1, \dots, X_n$ методом подстановки, записывается в виде

$$Q(F_n; v, \mu) = \frac{m}{k} \left(\sum_{i=n-k+1}^n X_{(i)} - \sum_{i=1}^k X_{(i)} \right) / \left(\sum_{i=n-m+1}^n X_{(i)} - \sum_{i=1}^m X_{(i)} \right),$$

$$k = [vn], \quad m = [\mu n], \quad (29)$$

где $X_{(1)}, \dots, X_{(n)}$ порядковые статистики выборки $X_1, \dots, X_n$. Отметим, что $Q(F_n) \xrightarrow{p} Q(F; v, \mu)$ при $n \rightarrow \infty$. Кроме того, результаты моделирования (см. [3]) показывают, что уже при объемах выборки $n \geq 20$ статистика $Q(F_n)$ приемлема для определения типов распределений, различающихся степенью «затянутости хвостов» распределений. Для построения адаптивной оценки Ходжеса – Лемана ($HL_{\hat{\alpha}}$ -оценки), определим параметр $\hat{\alpha}(X_1, \dots, X_n)$ в виде

$$\hat{\alpha}(X_1, \dots, X_n) = \begin{cases} \alpha_1, & Q(F_n) \leq Q_1, \\ \frac{\alpha_2 - \alpha_1}{Q_2 - Q_1} \{Q(F_n) - Q_1\} + \alpha_1, & Q_1 < Q(F_n) < Q_2, \\ \alpha_2, & Q(F_n) \geq Q_2, \end{cases} \quad (30)$$

где параметры α_1 , α_2 , Q_1 и Q_2 задаются в соответствии с рассматриваемым типом супермодели и выборочная оценка $Q(F_n)$ функционала $Q_F(v; \mu)$ определена в (29). Следуя [14], везде ниже полагаем $v = 0,2$ и $\mu = 0,5$.

Пример 1. Рассмотрим супермодель $\mathfrak{S}_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}\}$, определенную в (14). Примем следующие значения параметров: $\alpha_1 = 0$, $\alpha_2 = 0,5$, $Q_1 = 1,75$,

$Q_2 = 2,50$. В табл. 6 приведены асимптотические дисперсии (строки $n = \infty$) для HL_α -оценок, вычисленные по формуле (12), и выборочные дисперсии (строки $n = 20$) для HL_α -оценок, полученные методом статистических испытаний при числе испытаний $N = 10\,000$ и $n = 20$.

Таблица 6

Асимптотические и выборочные дисперсии HL_α -оценок для $\mathfrak{Z}_S^* = \{F_1, F_2, F_3, F_4\}$

Ф	n	HL	$HL_{0,05}$	$HL_{0,10}$	$HL_{0,20}$	$HL_{0,30}$	$HL_{0,40}$	$HL_{0,50}$	$HL_{\hat{\alpha}}$
F_1 – Гаусс	$n = 20$	1,04	1,07	1,09	1,15	1,25	1,39	1,43	1,04
	$n = \infty$	1,047	1,060	1,085	1,156	1,256	1,390	1,571	1,047
	$20/\infty$	0,99	1,01	1,00	0,99	1,00	1,00	0,91	0,99
F_2 – логист.	$n = 20$	3,06	3,09	3,11	3,24	3,50	3,87	3,96	3,14
	$n = \infty$	3,000	3,002	3,016	3,099	3,273	3,561	4,000	3,000
	$20/\infty$	1,02	1,03	1,03	1,05	1,07	1,09	0,99	1,05
F_3 – Лаплас	$n = 20$	1,40	1,37	1,33	1,26	1,23	1,23	1,25	1,39
	$n = \infty$	1,333	1,322	1,296	1,224	1,146	1,070	1,000	1,306
	$20/\infty$	1,05	1,04	1,03	1,03	1,07	1,15	1,25	1,06
F_4 – Коши	$n = 20$	3,93	3,66	3,36	2,81	2,47	2,43	2,45	2,66
	$n = \infty$	3,290	3,208	3,025	2,616	2,345	2,283	2,467	2,467
	$20/\infty$	1,19	1,14	1,11	1,07	1,05	1,06	0,99	1,08

В строках $(20/\infty)$ этой таблицы приведены отношения выборочных дисперсий HL_α -оценок при $n = 20$ к асимптотическим дисперсиям. В правом столбце таблицы приведены данные для адаптивной $HL_{\hat{\alpha}}$ -оценки Ходжеса – Лемана, для которой параметр $\hat{\alpha}(X_1, \dots, X_n)$ определен выражением (30) при значении параметров: $\alpha_1 = 0$, $\alpha_2 = 0,5$, $Q_1 = 1,75$, $Q_2 = 2,50$. Из данных таблицы (6) следует, что приведенные выше асимптотические результаты являются вполне приемлемой аппроксимацией дисперсий HL_α -оценок при конечных объемах выборки $n \geq 20$, за исключением некоторых значений при распределениях Лапласа и Коши. Для этих распределений с «тяжелыми хвостами», как показали результаты моделирования, качество асимптотики существенно улучшается при объемах выборки $n \geq 40$. Чтобы проиллюстрировать преимущество предложенных адаптивных $HL_{\hat{\alpha}}$ -оценок, применим критерий сравнения оценок, основанный на понятии дефекта оценки (см. работы [2, 3]). Дефект оценки $\hat{\theta}_i$, $i = 1, \dots, k$, среди сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ параметра θ при заданном распределении F определяется в виде

$$DE(F, \hat{\theta}_i) = 1 - \min\{\sigma^2(F, \hat{\theta}_1), \dots, \sigma^2(F, \hat{\theta}_k)\} / \sigma^2(F, \hat{\theta}_i), \quad i = 1, \dots, k. \quad (31)$$

Отметим, что если среди сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ есть эффективная оценка $\hat{\theta}^*$ параметра θ при заданном распределении F , для которой $\sigma^2(F, \hat{\theta}^*) = 1/I(f)$, тогда $\min\{\sigma^2(F, \hat{\theta}_1), \dots, \sigma^2(F, \hat{\theta}_k)\} = 1/I(f)$ и, следовательно, в

этом случае дефект оценки $\hat{\theta}_i$ равен единица минус ее абсолютная эффективность, то есть

$$DE(F, \hat{\theta}_i) = 1 - A\mathfrak{E}(F, \hat{\theta}_i), \quad i = 1, \dots, k. \quad (32)$$

При изучении свойств робастности сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ параметра сдвига θ в рамках супермодели $\mathfrak{S}$, состоящей из конечного набора симметричных распределений $\mathfrak{S} = \{F_1, \dots, F_r\}$, изучают поведение дефективности оценок на плоскости двух распределений (см. [2]). По оси абсцисс обычно откладывают дефективность для базовой (идеальной модели, обычно гауссовской), а по оси ординат – дефективность для альтернативной модели, входящей в супермодель $\mathfrak{S} = \{F_1, \dots, F_r\}$. При таком наглядном представлении дефективности оценок на плоскости двух распределений предпочтение отдается той оценке, которая окажется ближе к началу координат. Если же мы хотим сделать вывод о предпочтительности оценки среди сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ параметра θ в рамках всей рассматриваемой супермодели $\mathfrak{S} = \{F_1, \dots, F_r\}$, то можно использовать евклидову метрику, которая, с использованием введенных обозначений, запишется в виде

$$d(\hat{\theta}_i; \mathfrak{S}) = \left\{ \sum_{j=1}^r [DE(F_j, \hat{\theta}_i)]^2 \right\}^{1/2}, \quad i = 1, \dots, k. \quad (33)$$

Предпочтение в рамках всей рассматриваемой супермодели $\mathfrak{S} = \{F_1, \dots, F_r\}$ отдается той оценке $\hat{\theta}_i$ среди сравниваемых оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$, для которой вычисленное значение евклидовой метрики $d(\hat{\theta}_i; \mathfrak{S})$ минимальное, то есть

$$d(\hat{\theta}_i; \mathfrak{S}) = \min \{d(\hat{\theta}_1; \mathfrak{S}), \dots, d(\hat{\theta}_k; \mathfrak{S})\}, \quad i = 1, \dots, k. \quad (34)$$

Таблица 7

Значения евклидовой метрики HL_α -оценок для супермодели $\mathfrak{S}_S^* = \{F_{(1)}, F_{(2)}, F_{(3)}, F_{(4)}\}$

$\hat{\theta}$	HL	$HL_{0,05}$	$HL_{0,10}$	$HL_{0,20}$	$HL_{0,30}$	$HL_{0,40}$	$HL_{0,50}$	$HL_{\hat{\alpha}}$
$d(\hat{\theta}, \mathfrak{S}_S^*)$	0,40	0,35	0,29	0,18	0,21	0,33	0,36	0,15

В этой таблице евклидова метрика (33) была вычислена по эмпирическим данным (при $n = 20$) таблицы (6) с использованием формулы (31). Таким образом, в семействе HL_α -оценок для супермодели $\mathfrak{S}_S^*$ предпочтение следует отдать адаптивной $HL_{\hat{\alpha}}$ -оценке, для которой $d(HL_{\hat{\alpha}}, \mathfrak{S}_S^*) = 0,15$.

Пример 2. Рассмотрим супермодель $\mathfrak{S}_{\varepsilon, \tau}(\Phi)$ вида (25). Для определения параметра $\hat{\alpha}(X_1, \dots, X_n)$ по формуле (30) примем следующие значения параметров: $\alpha_1 = 0$, $\alpha_2 = 0,5$, $Q_1 = 1,75$, $Q_2 = 2,00$. В табл. 8 приведены асимптотические дисперсии (строки $n = \infty$) HL_α -оценок, вычисленные по формуле (28), и выборочные дисперсии (строки $n = 20$) HL_α -оценок, полученные методом статистических испытаний при числе испытаний $N = 10000$ и объеме выборки $n = 20$.

Таблица 8

Асимптотические и выборочные дисперсии HL_α -оценок для $F \in \mathfrak{Z}_{\varepsilon, \tau}(\Phi)$ при $\tau = 3$

При	n	HL	$HL_{0,05}$	$HL_{0,10}$	$HL_{0,20}$	$HL_{0,30}$	$HL_{0,40}$	$HL_{0,50}$	$HL_{\hat{\alpha}}$
$\varepsilon = 0,00$	$n = 20$	1,04	1,06	1,10	1,19	1,30	1,42	1,46	1,05
	$n = \infty$	1,047	1,060	1,085	1,156	1,256	1,390	1,571	1,047
	$20/\infty$	0,99	1,00	1,01	1,03	1,04	1,02	0,93	1,00
$\varepsilon = 0,05$	$n = 20$	1,17	1,18	1,20	1,29	1,40	1,55	1,58	1,17
	$n = \infty$	1,171	1,172	1,189	1,252	1,351	1,490	1,681	1,171
	$20/\infty$	1,00	1,01	1,01	1,03	1,04	1,04	0,94	1,00
$\varepsilon = 0,10$	$n = 20$	1,38	1,38	1,40	1,46	1,55	1,67	1,70	1,32
	$n = \infty$	1,311	1,302	1,308	1,360	1,457	1,600	1,803	1,303
	$20/\infty$	1,05	1,06	1,07	1,07	1,06	1,04	0,94	1,01
$\varepsilon = 0,20$	$n = 20$	1,80	1,77	1,75	1,75	1,85	1,94	1,97	1,65
	$n = \infty$	1,651	1,628	1,605	1,622	1,709	1,861	2,091	1,617
	$20/\infty$	1,09	1,09	1,09	1,08	1,08	1,04	0,94	1,02
$\varepsilon = 0,30$	$n = 20$	2,14	2,08	2,00	1,97	2,05	2,21	2,24	1,95
	$n = \infty$	2,090	2,062	2,004	1,966	2,032	2,191	2,454	1,993
	$20/\infty$	1,02	1,01	1,00	1,00	1,01	1,01	0,91	0,98
$\varepsilon = 0,40$	$n = 20$	2,63	2,57	2,49	2,44	2,53	2,79	2,87	2,50
	$n = \infty$	2,655	2,627	2,543	2,425	2,455	2,616	2,921	2,483
	$20/\infty$	0,99	0,98	0,98	1,01	1,03	1,07	0,98	1,01

Для сравнения HL_α -оценок с адаптивной $HL_{\hat{\alpha}}$ -оценкой в рамках супермодели $\mathfrak{Z}_{\varepsilon, \tau}(\Phi)$ воспользуемся критерием, основанным на евклидовой метрике вида (33), которая была вычислена по эмпирическим данным (при $n = 20$) табл. 8 с использованием формулы (31).

Таблица 9

Значения евклидовой метрики HL_α -оценок для супермодели $\mathfrak{Z}_{\varepsilon, \tau}(\Phi)$

$\hat{\theta}$	HL	$HL_{0,05}$	$HL_{0,10}$	$HL_{0,20}$	$HL_{0,30}$	$HL_{0,40}$	$HL_{0,50}$	$HL_{\hat{\alpha}}$
$d(\hat{\theta}, \mathfrak{Z}_{\varepsilon, \tau}(\Phi))$	0,15	0,12	0,11	0,19	0,32	0,48	0,52	0,03

Таким образом, в семействе HL_α -оценок в рамках супермодели $\mathfrak{Z}_{\varepsilon, \tau}(\Phi)$, согласно критерию (34), предпочтение следует отдать адаптивной $HL_{\hat{\alpha}}$ -оценке, для которой параметр $\hat{\alpha}(X_1, \dots, X_n)$ определяется выражением (30) и $d(HL_{\hat{\alpha}}, \mathfrak{Z}_{\varepsilon, \tau}(\Phi)) = 0,03$. На рис. 4 приведены абсолютные эффективности HL_α -оценок и адаптивной $HL_{\hat{\alpha}}$ -оценки для гауссовской супермодели с масштабным засорением, то есть для $F \in \mathfrak{Z}_{\varepsilon, \tau}(\Phi)$. Из рисунка наглядно видно, что адаптивная $HL_{\hat{\alpha}}$ -оценка обладает преимуществом перед HL_α -оценками с фиксированной пропорцией урезания α исходной выборки $X_1, \dots, X_n$.

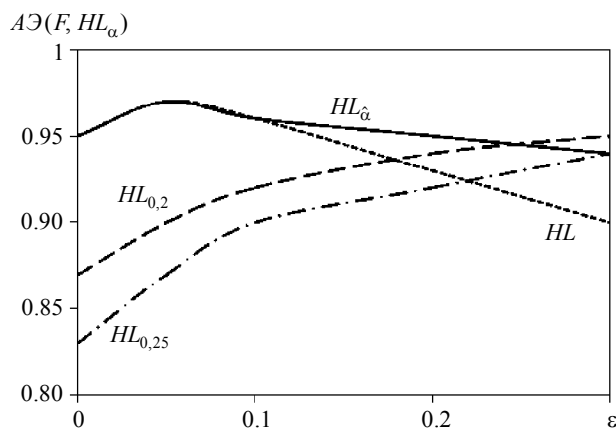


Рис. 4. Абсолютные эффективности HL_{α} -оценок в рамках супермодели $\mathfrak{Z}_{\varepsilon, \tau}(\Phi)$, $\tau = 3$

Заключение

В работе предложены адаптивные оценки α -урезанного варианта (5) оценки Ходжеса – Лемана параметра сдвига симметричных распределений. Описаны асимптотические свойства предложенных оценок. Параметр $\hat{\alpha}(X_1, \dots, X_n)$ адаптивной $HL_{\hat{\alpha}}$ -оценки выбирается на основе информации, содержащейся в исходной выборке $X_1, \dots, X_n$, путем использования выборочной оценки функционала, характеризующего степень «тяжести хвостов» распределений. Приводятся свойства оценок в асимптотике и при конечных объемах выборки в рамках различных супермоделей, описывающих отклонения от нормальной модели в сторону «утяжеления хвостов» распределений. Показано, что приведенные асимптотические результаты являются вполне приемлемой аппроксимацией дисперсий HL_{α} -оценок при конечных объемах выборки $n \geq 20$. Предложен критерий сравнения заданного семейства оценок $\hat{\theta}_1, \dots, \hat{\theta}_k$ параметра сдвига θ в рамках супермодели $\mathfrak{Z}$, состоящей из конечного набора симметричных распределений. Этот критерий записан в виде евклидовой метрики с использованием понятия дефекта оценки при заданном распределении.

ЛИТЕРАТУРА

1. Hodges J.L., Lehmann E.L. Estimation of location based on rank tests // Ann. Math. Statist. 1963. V. 34. P. 598 – 611.
2. Andrews D.F., Bickel P.Z., Hampel F.R., et al. Robust estimation of location: survey and advances. Princeton, N.Y.: Princeton Univ. Press, 1972. 375 p.
3. Шуленин В.П. Введение в робастную статистику. Томск: Изд-во Том. ун-та, 1993. 227 с.
4. Шуленин В.П. Об устойчивости класса оценок Ходжеса – Лемана // Тр. 6-й Всес. конф. по теории кодирования и передачи информ. Москва – Вильнюс, 1978. С. 147 – 151.
5. Shulenin V.P. Asymptotic properties and robustness a generalized Hodges-Lehmann estimate. II-th PRAGUE Conf. On Information Theory. Abstracts. 1990.
6. Shulenin V.P., Deeva T.A. The numerical characteristics of robustness of the class of the Hodges-Lehmann Generalized estimators // Proc. the Third Russian-Korean International Symposium on Science and Technology. KORUS'99. June 22 – 25, 1999 at Novosibirsk State Technical University, Novosibirsk, Russia. V. 2. P. 510 – 513.

7. *Shulenin V.P., Deeva T.A.* Asymptotic efficiency for the generalized Hodges – Lehmann estimator under the normal scale mixture distributions. Computer data analysis and modeling. Minsk, 1998. V. 2. P. 107 – 112.
8. *Shulenin V.P.* Asymptotic properties of the trimmed GL- and U-statistics // 6th Prague Symposium on Asymptotic Statistics. Prague, 1998, August 23 – 28. Prague Stochastics'98. Abstracts. P. 84.
9. *Шуленин В.П.* Асимптотические свойства урезанных GL-и U-статистик // Вестник ТГУ. Приложение. 2004. № 9(II). С. 184 – 190.
10. *Serfling R.J.* Generalized L-, M- and R-statistics // Ann. Statist. 1984. V. 12. P. 76 – 86.
11. *Serfling R.J.* Approximation Theorems of Mathematical Statistics. N.Y.: Wiley, 1980. 371 p.
12. *Хеттсманспергер Т.* Статистические выводы, основанные на рангах. М.: Финансы и статистика, 1987. 334 с.
13. *Van Zwet W.R.* Convex Transformations of Random Variables // Math. Centrum. Amsterdam, 1964.
14. *Hogg R.V.* Adaptive robust procedures: partial review and some suggestions for future applications and theory // J. Amer. Statist. 1974. V. 35. P. 73 – 101.

Шуленин Валерий Петрович

Томский государственный университет

E-mail: shvp@fpmk.tsu.ru

Поступила в редакцию 23 марта 2010 г.