ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2016 Управление, вычислительная техника и информатика № 2 (35)

УДК 519.24
DOI: 10.17223/19988605/35/7

В.П. Шуленин

АСИМТОТИЧЕСКИЕ СВОЙСТВА РОБАСТНЫХ ОЦЕНОК МАСШТАБНОГО ПАРАМЕТРА

Изучаются свойства робастных оценок масштабного параметра. Показано, что оценка медианы абсолютных разностей имеет асимтотически нормальное распределение, является B-робастной и имеет ограниченную функцию влияния. Приводятся результаты сравнения оценок масштабного параметра в рамках гауссовской модели с засорением.
Ключевые слова: масштабный параметр; робастные оценки; функция влияния.

Масштабный параметр используется в качестве меры, характеризующей степень разброса случайной величины, и определяется в виде функционала от функции распределения наблюдений. Общие требования, предъявляемые к таким функционалам, сформулированы в работах [1, 2]. Традиционно используемые на практике оценки масштабного параметра, такие как выборочная оценка $\hat{S}_1(0)$ стандартного отклонения $S_1(F)$ и оценка $\hat{S}_2(0)$ среднего абсолютных отклонений $S_2(F)$, имеют неограниченные функции влияния и очень чувствительны к наличию выбросов в выборке. Урезанные варианты этих оценок $\hat{S}_1(\alpha)$ и $\hat{S}_2(\alpha)$, $0 \le \alpha < 1/2$, которые вычисляются не по исходной выборке $X_1,...,X_n$, а на основе упорядоченной статистики $X_{(1)},...,X_{(n)}$, из которой предварительно удаляются $[\alpha n]$ наименьших и наибольших порядковых статистик, имеют ограниченные функции влияния, и их характеристики существенно зависят от параметра α, что на практике приводит к дополнительным усилиям по выбору этого параметра, например с помощью адаптивного подхода, и это усложняет оценку. Выборочные оценки $\hat{S}_3(\alpha) = [X_{(n-[\alpha n])} - X_{([\alpha n])}]/2$, $0 < \alpha < 1/2$, интер-α-квантильных размахов имеют ограниченные функции влияния, но их асимптотические эффективности по отношению к оценке стандартного отклонения при нормальном распределении Φ очень низкие. Например, для оценки интерквартильного размаха $\hat{S}_3(0{,}25)$ асимптотическая относительная эффективность равна $\text{АОЭ}_\Phi(\hat{S}_3(0{,}25):\hat{S}_1(0)) = 0{,}37$. В теории робастного оценивания параметров (см., например, [3–6]) при построении оценок масштабного параметра обычно используют медиану абсолютных отклонений от медианы, выборочная оценка которой записывается в виде $\hat{S}_3^* = \text{med}\{|X_i - \text{med}(X)|,\ 1 \le i \le n\}$. Эта оценка имеет ограниченную функцию влияния, для неё «точка срыва» (breakdown point) равна максимально возможному значению, равному $1/2$, но при этом её эффективность невелика, и при нормальном распределении Φ также имеем $\text{АОЭ}_\Phi(\hat{S}_3^* : \hat{S}_1(0)) = 0{,}37$. Таким образом, мы приходим к необходимости построения новых оценок масштабного параметра, и с ограниченными функциями влияния, и с высокой эффективностью при нормальном распределении. К числу таких оценок относится рассматриваемая в работе медиана абсолютных разностей $\hat{S}_4 = \text{med}\{|X_i - X_j|,\ 1 \le i < j \le n\}$, которая входит в семейство обобщённых L-оценок [7]. В данной работе доказана асимптотическая нормальность этой оценки, отмечено, что она имеет ограниченную функцию влияния и высокую эффективность при нормальном распределении. Приведены результаты сравнения оценок в рамках гауссовской модели с масштабным засорением.

1. Обсуждение общего подхода к построению оценок масштабного параметра

Введем необходимые понятия и обозначения. Пусть X – изучаемая случайная величина с функцией распределения $F(x)$, $x \in R^1$, которая абсолютно непрерывна, имеет плотность $f(x)$, $x \in R^1$, и симметрична относительно точки θ_x, т.е. $F \in \Im_{S|\theta}$, где $\Im_{S|\theta} = \{F : F(x) = 1 - F(2\theta_x - x), \forall x \in R^1\}$. Масштабный параметр функции распределения F используется в качестве меры, характеризующей степень разброса случайной величины (с.в.) X с функцией распределения (ф.р.) F. Рассмотрим такие меры, которые могут быть представлены в виде функционала $S(F)$, $F \in \Im$, заданного на множестве допустимых распределений $\Im$ в условиях эксперимента, связанного с изучением с.в. X по статистическим данным $X_1,...,X_n$, полученным в серии n независимых и повторных наблюдений над с.в. X. Выборочная оценка $\hat{S}(X_1,...,X_n)$ функционала $S(F)$, $F \in \Im$, построенная методом подстановки, записывается в виде $\hat{S}(X_1,...,X_n) = S(F_n)$, где $F_n(x)$ – эмпирическая функция распределения, построенная по выборке $X_1,...,X_n$. Общие требования, которым должен удовлетворять функционал $S(F)$, описывающий разброс случайной величины X, сформулированы Бикелем и Леманом [1, 2]. Для формулировки этих требований напомним определения.

Определение 1. О случайных величинах X_1 и X_2 с функциями распределений F_1 и F_2 говорят, что с.в. X_2 стохастически больше, чем с.в. X_1 (при этом используют обозначение в виде $F_1 <_{St} F_2$), если выполняется неравенство $P(X_1 > x) \le P(X_2 > x)$, $\forall x \in R^1$. Отметим, что $F_1 <_{St} F_2 \Rightarrow F_1(x) \ge F_2(x)$, $\forall x \in R^1$ и $F_1^{-1}(t) \le F_2^{-1}(t)$, $0 \le t \le 1$.

Определение 2. Говорят, что выборочная оценка $\hat{S}(X_1,...,X_n) = S(F_n)$ функционала $S(F)$, $F \in \Im$, является эквивариантной относительно линейных преобразований наблюдений $X_1,...,X_n$, если выполняется равенство $\hat{S}(aX_1 + b,...,aX_n + b) = |a|\hat{S}(X_1,...,X_n)$.

Определение 3. Разброс с.в. X относительно θ_x (масштабный параметр с.в. X) определяют в терминах «расстояния» X от θ_x, т.е. с помощью величины $|X - \theta_x|$, при этом говорят, что с.в. X_1 имеет больший разброс относительно θ_{x_1}, чем с.в. X относительно θ_x, если с.в. $|X_1 - \theta_{x_1}|$ стохастически больше с.в. $|X - \theta_x|$.

Определение 4. Согласно условиям Бикеля и Лемана [1, 2] функционал $S(F)$, $F \in \Im$, определяет меру разброса, или масштабный параметр ф.р. F, если его выборочная оценка $\hat{S}(X_1,...,X_n) = S(F_n)$ является эквивариантной относительно линейных преобразований наблюдений $X_1,...,X_n$ и он удовлетворяет условию монотонности относительно стохастического возрастания распределений, т.е. выполняется выражение $S(F_1) \le S(F_2)$ для $F_1 <_{St} F_2$, где F_1 и F_2 – функции распределения вероятностей случайных величин $|X_1 - \theta_{x_1}|$ и $|X_2 - \theta_{x_2}|$.

Замечание 1. Отметим, что выбор конкретного функционала для описания масштабного параметра может быть продиктован различными требованиями. Так, в работе [2] кроме условия эквивариантности оценки относительно линейных преобразований накладывается требование непрерывности функционала относительно метрики, порождающей слабую сходимость. Выполнение этого требования приводит к оценкам функционалов, удовлетворяющих условиям качественной робастности [3, 4]. Окончательный выбор может осуществляться путем сравнения точностей, с которыми каждый функционал может быть оценен по наблюдениям в рамках заданной супермодели. Множество различных функционалов, характеризующих масштабный параметр, условно можно разделить на следующие группы.

К первой группе относятся функционалы, построенные с помощью отклонений каждого члена генеральной совокупности от некоторого «центрального» (типичного значения) с.в. X с ф.р. F. Обычно в качестве такого значения используется параметр положения, определённый функционалом $T(F)$, либо в виде среднего $T_1(F) = \int x dF(x) = M(X)$, либо в виде медианы $T_2(F) = F^{-1}(1/2) = MED(X)$. Обозначим ф.р. $|X - T(F)|$ через F_1 и ф.р. $|X_1 - X_2|$, где X_1 и X_2 – независимые с ф.р. F, через F_2. Многих представителей первой группы можно описать с помощью функционалов вида

$$\left\{ \int_0^1 [F_i^{-1}(t)]^\gamma \, dV(t) \right\}^{1/\gamma}, \quad i = 1, 2, \tag{1}$$

где $V(t)$ – некоторая функция распределения на $[0, 1]$ и $\gamma > 0$. Например, если в (1) положить $i = 1$, в качестве параметра положения выбрать среднее значение $T_1(F)$ и положить $V(t) = t$, $0 \le t \le 1$, то при $\gamma = 1$ получим среднее абсолютных отклонений $S_2(F)$, выборочная оценка которого запишется в виде $\hat{S}_2(0) = n^{-1}\Sigma | X_i - \overline{X} |$. При $\gamma = 2$ будем иметь стандартное отклонение. Если же положить $V(t) = t/(1-\alpha)$, $0 \le t \le 1-\alpha$, $0 \le \alpha < 1/2$, то получим α-урезанные варианты указанных мер масштабного параметра. Другая часть этой группы определяется функционалом $F_1^{-1}(1/2)$. Например, при использовании в качестве параметра положения $T_2(F) = F^{-1}(1/2) = MED(X)$ получаем широко используемую в теории робастного оценивания медиану абсолютных отклонений от медианы $MED | X - MED(X) |$ [6–8].

Ко второй группе относятся функционалы, построенные с помощью отклонений между всеми членами генеральной совокупности. Некоторые представители этой группы также выражаются с помощью функционалов вида (1). Например, при $i = 2$, $\gamma = 1$, и $V(t) = t$, $0 \le t \le 1$, из (1) получим среднюю разность Джини, выборочная оценка которой записывается в виде $\hat{\Delta}_0 = [n(n-1)]^{-1}\Sigma | X_i - X_j |$. При $\gamma = 2$ получаем стандартное отклонение, умноженное на $\sqrt{2}$. К этой же группе относится и медиана абсолютных разностей $MED | X_i - X_j |$, $1 \le i, j \le n$, определяемая с помощью функционала $S_4(F) = F_2^{-1}(1/2)$ [8, 10].

К третьей группе относятся функционалы, построенные с помощью расстояний между точками, в которых ф.р. F имеет характерные особенности. К таким точкам могут относиться, например, квантили заданных уровней. Некоторых представителей этой группы можно описать с помощью функционалов вида

$$\left\{ \int_0^1 | F^{-1}(t) - F^{-1}(1-t) |^{\gamma} \, dK(t) \right\}^{1/\gamma}, \tag{2}$$

где $K(t)$ – некоторая функция распределения на $[0, 1]$ и $\gamma > 0$. В частности, к этой группе относятся интер-α-квантильные размахи, определяемые в виде $[F^{-1}(1-\alpha) - F^{-1}(\alpha)]/2$, $0 < \alpha < 1/2$. Отметим, что при $\alpha = 0{,}25$ получаем интерквартильный размах, который для симметричных распределений совпадает с медианой абсолютных отклонений от медианы, т.е. определяется функционалом $F_1^{-1}(1/2)$ при использовании в качестве параметра положения функционала $T_2(F) = F^{-1}(1/2) = MED(X)$. Подводя итог, отметим, что один из общих подходов построения различных (мер) функционалов от распределений, описывающих масштабный параметр, сводится, по существу, к следующему. Над исходной случайной величиной X осуществляется некоторое преобразование вида $| X - T_1(F) |^{\gamma}$, $| X - T_2(F) |^{\gamma}$, $| X_1 - X_2 |^{\gamma}$, $\gamma > 0$, и т.п. Затем к преобразованным случайным величинам применяется либо «операция усреднения», либо «операция вычисления медианы», либо «операция вычисления оценки Ходжеса–Лемана» и т.п. Другими словами, функционал, описывающий масштабный параметр, определяется с помощью функционала, характеризующего параметр положения для преобразованных случайных величин. Например, «медианная операция», примененная к $| X - T_2(F) |^{\gamma}$ при $\gamma = 1$, приводит к медиане абсолютных отклонений от медианы; «медианная операция», примененная к $| X_1 - X_2 |^{\gamma}$ при $\gamma = 1$, приводит к медиане абсолютных разностей; «операция усреднения» в этом случае приводит к средней разности Джини; «операция вычисления оценки Ходжеса–Лемана» приводит к еще не изученным оценкам масштабного параметра, например, такого вида:

$$\mathrm{med}\{ [| X_i - \mathrm{med}(X) | + | X_j - \mathrm{med}(X) |]/2,\ 1 \le i, j \le n \},\ \mathrm{med}\{ [| X_i - X_j | + | X_k - X_l |]/2,\ 1 \le i, j, k, l \le n \}.$$

При этом также возможно использование обобщенных оценок Ходжеса–Лемана и их урезанных вариантов. Применение этой схемы открывает большие возможности при построении новых мер масштабного параметра, при этом могут использоваться обширные результаты, полученные при изучении оценок параметра положения, включая их общие классы *M*-, *L*- и *R*-оценок, а также обобщённые *L*-оценки [7, 11–15].

Замечание 2. Отметим, что для выбранного функционала $S(F)$, $F \in \Im$, описывающего масштабный параметр с.в. X с ф.р. F, его выборочная оценка строится методом подстановки и записывается в виде $\hat{S}(X_1,...,X_n) = S(F_n)$, где $F_n(x)$ – эмпирическая функция распределения, построенная по выборке $X_1,...,X_n$. Асимптотическая нормальность таких оценок масштабного параметра изучается методом Мизеса с использованием теоремы Слуцкого и центральной предельной теоремы [6, 15] на основе разложения

$$S(F_n) = S(F) + \frac{1}{n}\sum_{i=1}^{n} IF(X_i; F, S) + o_p(n^{-1/2}), \tag{3}$$

где $IF(x;F,S)$ – функция влияния Хампеля [5, 9] оценки $\hat{S} = S(F_n)$ функционала $S(F)$, $F \in \Im$, которая определяется в виде

$$IF(x;F,S) = \lim_{\lambda \to 0+} \frac{S[(1-\lambda)F + \lambda\Delta_x] - S(F)}{\lambda}, \; 0 \le \lambda \le 1, \; x \in R^1, \tag{4}$$

для тех $x \in R^1$, при которых предел существует. Здесь Δ_x обозначает вырожденную функцию распределения в точке $x \in R^1$. Асимптотическая дисперсия $\sqrt{n}\hat{S}$-оценки обозначается через $\sigma^2(F,\hat{S})$ и вычисляется по формуле

$$\sigma^2(F,\hat{S}) = \int_{-\infty}^{\infty} IF^2(x;F,S)dF(x). \tag{5}$$

Для сравнения различных оценок масштабного параметра при заданной ф.р. F будем использовать понятие асимптотической относительной эффективности, определенное через обратное отношение стандартизованных асимптотических дисперсий. Асимптотическую эффективность оценки $\hat{S}_1$ относительно $\hat{S}_2$ при заданной ф.р. F обозначим через $\text{АОЭ}_F(\hat{S}_1 : \hat{S}_2)$ и, следуя работам [1, 2], определим в виде

$$\text{АОЭ}_F(\hat{S}_1 : \hat{S}_2) = \frac{\tilde{\sigma}^2(F,\hat{S}_2)}{\tilde{\sigma}^2(F,\hat{S}_1)}, \tag{6}$$

где $\tilde{\sigma}^2(F,\hat{S})$ – стандартизованная дисперсия $\sqrt{n}\hat{S}$-оценки, равная отношению асимптотической дисперсии к квадрату функционала, т.е.

$$\tilde{\sigma}^2(F,\hat{S}) = \sigma^2(F,\hat{S}) / S^2(F). \tag{7}$$

2. Асимптотическая нормальность выборочной медианы абсолютных разностей

Пусть $X_1,...,X_n$ – независимые одинаково распределённые случайные величины, порождённые ф.р. $F(x)$ с плотностью $f(x)$, $\forall x \in R^1$.

Обозначим функцию распределения с. в. $Y = |X_1 - X_2|$ через $H_F(y)$ и определим её в виде

$$H_F(y) = \int\int I[|x_1 - x_2| \le y]dF(x_1)dF(x_2) = \int [F(x+y) - F(x-y)]\, dF(x), \; y \in R^1.$$

Плотность $h_F(y)$ ф.р. $H_F(y)$ равна

$$h_F(y) = \int [f(x+y) + f(x-y)]\, dF(x).$$

Эмпирическая функция распределения $H_n(y)$, значений $|X_i - X_j|$, $1 \le i < j \le n$, записывается в виде

$$H_n(y) = \frac{1}{n(n-1)} \sum_{i<j} I[|X_i - X_j| \le y].$$

Определим функционал $S_4(H_F)$ в виде

$$S_4(H_F) = H_F^{-1}(1/2), \tag{8}$$

где H_F^{-1} квантильная функция для ф.р. $H_F(y)$. Выборочная оценка $\hat{S}_4 = S_4(H_n)$ функционала $S_4(H_F)$, называемая выборочной медианой абсолютных разностей, входит в класс обобщенных L-оценок [7] и определяется в виде

$$S_4(H_n) = \mathrm{med}\{\, |X_i - X_j|,\ 1 \le i < j \le n\,\} = \begin{cases} W_{(r+1)}\,,\ N = 2r+1, \\ \{W_{(r)} + W_{(r+1)}\}/2\ ,\ N = 2r, \end{cases} \tag{9}$$

где $W_{(1)},..., W_{(N)}$, $N = n(n-1)$, – упорядоченные значения «абсолютных разностей Джини» $|X_i - X_j|$, $1 \le i < j \le n$, число которых равно $N = n(n-1)$. Отметим, что при выполнении неравенства $h(H_F^{-1}(1/2)) > 0$ оценка $\hat{S}_4 = S_4(H_n)$ является асимптотически нормальной (см. [10], а также теорему (8.4.26) в [15]). Отметим также, что при изучении свойств робастности оценок важную роль играет функция влияния Хампеля, определяемая выражением (4). Приведем эту функцию для асимптотически эквивалентной оценки $\tilde{T}(F_n) = \mathrm{med}\{|X_i - X_j|,\ 1 \le i, j \le n\}$. Эта оценка была предложена в [10]. Она является выборочной оценкой эквивалентного функционала $\widetilde{T}(F)$, который в данном случае может быть определен через исходную ф.р. F неявно с помощью выражения

$$\int\!\!\int I[|x_1 - x_2| \le \widetilde{T}(F)]dF(x_1)dF(x_2) = \frac{1}{2}, \text{ или } \int [F(x + \widetilde{T}(F)) - F(x - \widetilde{T}(F)) - F(x)]dF(x) = 0. \tag{10}$$

Можно убедиться [14, 15], что дифференциал Гато первого порядка функционала $\widetilde{T}(F)$, заданного выражением (10), вычисляется по формуле

$$d_1\widetilde{T}(F; G - F) = \frac{(1/2) - \int [F(x + \widetilde{T}) - F(x - \widetilde{T})]\, dG(x)}{\int [f(x + \widetilde{T}) + f(x - \widetilde{T})]\, dF(x)}. \tag{11}$$

Отсюда следует, что функция влияния оценки $\widetilde{T}(F_n)$ функционала $\widetilde{T}(F)$, заданного выражением (10), определяется в виде

$$IF(x; F, \widetilde{T}) = d_1\widetilde{T}(F; \Delta_x - F) = \frac{1 + 2F(x - \widetilde{T}) - 2F(x + \widetilde{T})}{2\int [f(x + \widetilde{T}) + f(x - \widetilde{T})]\, dF(x)},\ x \in R^1. \tag{12}$$

Заметим, что для симметричных распределений функционал $\widetilde{T}(F)$ определяется выражением

$$\int F(x + \widetilde{T})dF(x) = 3/4, \tag{13}$$

т.е. соответствует квантилю уровня $(3/4)$ для ф.р. случайной величины $Y = |X_1 - X_2|$. Используя разложение (3), можно убедиться [14, 15], что оценка $\widetilde{T}(F_n)$ асимптотически нормальная, т.е. справедливо выражение

$$L\{\sqrt{n}[\widetilde{T}(F_n) - \widetilde{T}(F)]/\sigma(F, \widetilde{T})\} = N(0, 1) \text{ при } n \to \infty, \tag{14}$$

причем асимптотическая дисперсия $\sqrt{n}\widetilde{T}(F_n)$-оценки вычисляется по формуле

$$\sigma^2(F, \widetilde{T}) = \int_{-\infty}^{\infty} IF^2(x; F, \widetilde{T})dF(x) = \frac{\int [1 + 2F(x - \widetilde{T}) - 2F(x + \widetilde{T})]^2\, dF(x)}{4\left(\int [f(x + \widetilde{T}) + f(x - \widetilde{T})]dF(x)\right)^2}. \tag{15}$$

Функции влияния различных оценок масштабного параметра при нормальном распределении Ф приведены на рис. 1. Отметим, что функция влияния оценки $\hat{S}_4$ медианы абсолютных разностей является *ограниченной*, следовательно, эта оценка является В-робастной [9] и подвержена меньшему влиянию выбросов в выборке $X_1,..., X_n$, по сравнению с традиционно применяемой оценкой $\hat{S}_1(0)$ стандартного отклонения, оценкой $\hat{S}_2(0)$ среднего абсолютных отклонений и оценкой средних разностей Джини Δ_α, $0 \le \alpha < 1/2$ [13].

Пример 1. Приведем результаты сравнения оценки $S_4(F)$ с оценкой $\hat{S}_1(0)$ стандартного отклонения и оценкой $\hat{S}_2(0)$ среднего абсолютных отклонений в рамках супермодели с засорением вида $\Im_{\varepsilon,\tau}(\Phi) = \{F : F(x) = \Phi_{\varepsilon,\tau}(x)\}$, где $\Phi_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)$. Учитывая, что $\Phi_{\varepsilon,\tau}(x) \in \Im_{S|0}$, выражение (13), определяющее функционал $S_4(F)$, перепишем в виде

$$\int \Phi_{\varepsilon,\tau}(x+S_4)d\Phi_{\varepsilon,\tau}(x)-(3/4)=0\,.$$

Формула (15) для асимптотической стандартизованной дисперсии перепишется в виде

$$\tilde{\sigma}^2(F,\tilde{T})=\frac{\pi^2\int[1+2\Phi_{\varepsilon,\tau}(x-S_4)-2\Phi_{\varepsilon,\tau}(x+S_4]^2\,d\Phi_{\varepsilon,\tau}(x)}{S_4^2\cdot[(1-\varepsilon)J_1+\varepsilon(1-\varepsilon)J_2/\tau\,+\,\varepsilon^2J_3/\tau^2\,]^2},$$

где

$$J_1=\sqrt{\pi}\exp\{-S_4^2/4\}\,,\ J_2=[\tau\sqrt{2\pi}/\sqrt{1+\tau^2}\,]\exp\{-(1/2)(S_4/\tau)^2\}\,,\ J_3=\tau\sqrt{\pi}\exp\{-S_4^2(2-\tau^2)/4\tau^2\}\,.$$

Численные расчеты относительных эффективностей $АОЭ_F(\hat{S}_4:\hat{S}_1(0))$ медианы абсолютных разностей $\hat{S}_4$ по отношению к оценке $\hat{S}_1(0)$ стандартного отклонения и относительных эффективностей $АОЭ_F(\hat{S}_4:\hat{S}_2(0))$ для $F\in\Im_{\varepsilon,\tau}(\Phi)$ при различных ε и τ приведены в табл. 1.

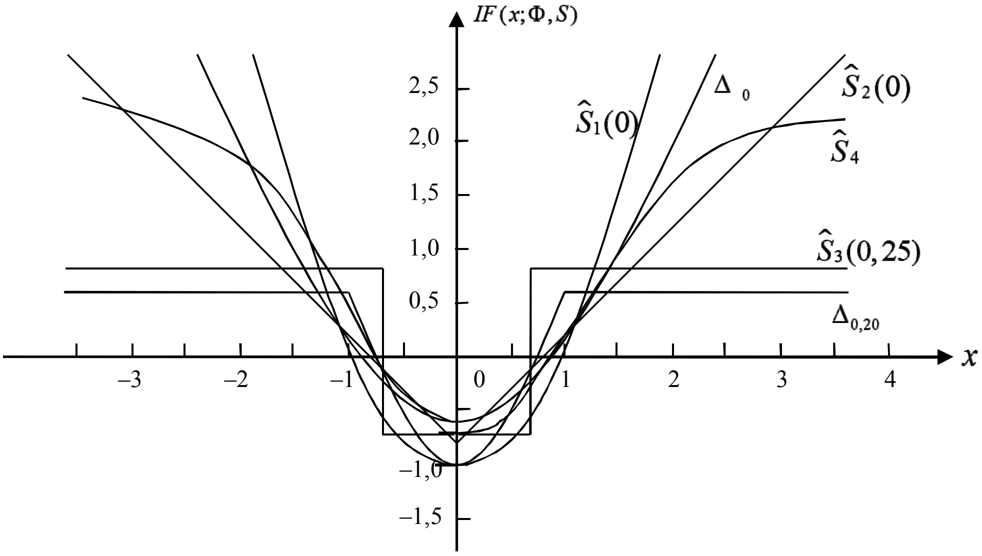

Рис. 1. Функции влияния оценок масштабного параметра

Т а б л и ц а 1

Относительные эффективности $АОЭ_F(\hat{S}_4:\hat{S}_1(0))$ и $АОЭ_F(\hat{S}_4:\hat{S}_2(0))$ для $F\in\Im_{\varepsilon,\tau}(\Phi)$

τ	$АОЭ(\hat{\theta}_1:\hat{\theta}_2)$	$\varepsilon=0,00$	$\varepsilon=0,01$	$\varepsilon=0,05$	$\varepsilon=0,10$	$\varepsilon=0,15$	$\varepsilon=0,20$
$\tau=3$	$\hat{S}_4:\hat{S}_1(0)$	0,864	1,507	2,383	2,211	1,816	1,459
$\tau=3$	$\hat{S}_4:\hat{S}_2(0)$	0,986	1,048	1,172	1,163	1,075	0,967
$\tau=5$	$\hat{S}_4:\hat{S}_1(0)$	0,864	5,471	6,733	4,395	2,751	1,735
	$\hat{S}_4:\hat{S}_2(0)$	0,986	1,335	1,988	1,958	1,616	1,230

Из данных таблицы видно, что оценка $\hat{S}_4$ медианы абсолютных разностей, проигрывая по эффективности оценке $\hat{S}_1(0)$ стандартного отклонения при нормальном распределении $(\varepsilon=0)$ менее 14%, а оценке $\hat{S}_2(0)$ среднего абсолютных отклонений лишь 2%, обладает существенным преимуществом при отклонении от нормального распределения в рамках супермодели $\Im_{\varepsilon,\tau}(\Phi)$. В частности, при изменении пропорции засорения ε в интервале $0<\varepsilon<0,2$ приведенные эффективности больше единицы. Отмеченное преимущество возрастает при «утяжелении хвостов распределений» (при увеличении ε и масштабного параметра τ засорения нормального распределения). Напомним также, что оценка $\hat{S}_4$, в отличие от оценок $\hat{S}_1(0)$ и $\hat{S}_2(0)$, имеет *ограниченную функцию влияния* (см. рис. 1).

Пример 2. Пусть исходная функция распределения F является распределением Лапласа с плотностью $f(x)=(1/2)\exp\{-|x|\}$, $x\in R^1$. В этом случае выражение (13) принимает вид уравнения $\ln(2+S_4)-S_4=0$, решением которого является значение функционала $S_4(F)\approx 1{,}146$. Асимптотическая дисперсия $\sqrt{n}\hat{S}_4$-оценки в данном случае, вычисленная по формуле (15), равна 1,685. Следовательно, стандартизованная дисперсия равна $\tilde{\sigma}^2(F,\hat{S}_4)=1{,}685/(1{,}146)^2=1{,}277$. Для сравнения отметим, что при распределении Лапласа для оценок $\hat{S}_1(0)$ и $\hat{S}_2(0)$ имеем $\tilde{\sigma}^2(F,\hat{S}_1(0))=1{,}250$, $\tilde{\sigma}^2(F,\hat{S}_2(0))=1{,}000$. Следовательно, относительные эффективности при распределении Лапласа равны $\text{АОЭ}_F(\hat{S}_4:\hat{S}_1(0))=0{,}98$, $\text{АОЭ}_F(\hat{S}_4:\hat{S}_2(0))=0{,}78$.

Приведем теперь результаты сравнения оценки $\hat{S}_4$ с оценками $\hat{S}_3(\alpha)$, $0<\alpha<1/2$, семейства интер-α-квантильных размахов. Согласно формуле (15) для распределения Лапласа имеем $\tilde{\sigma}^2(F,\hat{S}_3(\alpha))=(1-2\alpha)/2\alpha(\ln 2\alpha)^2$. Численные значения относительной эффективности $\text{АОЭ}_F(\hat{S}_4:\hat{S}_3(\alpha))$ оценок $\hat{S}_4$ и $\hat{S}_3(\alpha)$, $0<\alpha<1/2$, для распределения Лапласа приведены в табл. 2.

Таблица 2

Относительные эффективности $\text{АОЭ}_F(\hat{S}_4:\hat{S}_3(\alpha))$, *F*-распределение Лапласа

α	$\alpha=0{,}05$	$\alpha=0{,}10$	$\alpha=0{,}15$	$\alpha=0{,}20$	$\alpha=0{,}25$	$\alpha=0{,}30$
$\text{АОЭ}_F(\hat{S}_4:\hat{S}_3(\alpha))$	1,33	1,21	1,26	1,40	1,63	2,00

Из таблицы видно, что оценка $\hat{S}_4$ медианы абсолютных разностей при распределении Лапласа эффективнее любой оценки $\hat{S}_3(\alpha)$, $0<\alpha<1/2$, из семейства интер-α-квантильных размахов. В частности, её эффективность по отношению к оценке $\hat{S}_3^*=\text{med}\{|X_i-\text{med}(X)|,\ 1\le i\le n\ \}$ при распределении Лапласа равна $\text{АОЭ}_F(\hat{S}_4:\hat{S}_3^*)=\text{АОЭ}_F(\hat{S}_4:\hat{S}_3(0{,}25))=1{,}63$.

Заключение

В работе обсуждается общий подход к построению робастных оценок масштабного параметра, который основан на использовании робастных оценок параметра положения для предварительно преобразованных наблюдений. Отмечено, что предложенная в [10] выборочная оценка в виде медианы абсолютных разностей, т.е. оценка $\tilde{T}(F_n)=\text{med}\{|X_i-X_j|,\ 1\le i,j\le n\}$, является B-робастной, её функция влияния Хампеля ограничена, она подвержена меньшему влиянию выбросов в выборке, чем традиционно используемые оценки $\hat{S}_1(0)$ стандартного отклонения $S_1(F)$ и оценка $\hat{S}_2(0)$ среднего абсолютных отклонений $S_2(F)$. Для предложенной оценки приведены параметры асимптотически нормального распределения. Проведено сравнение оценок и показано, что предложенная оценка $\tilde{T}(F_n)=\text{med}\{|X_i-X_j|,\ 1\le i,j\le n\}$ имеет высокую эффективность и при нормальном распределении, и в рамках гауссовской модели с масштабным засорением.

ЛИТЕРАТУРА

1. Bickel P.J., Lehmann E.L. Measures of location and scale // Proc. Prague Symp. Asymptotic Statist. 1973. Prague Charles Univ. 1974. V. 1. P. 25–36.

2. Bickel P.J., Lehmann E.L. Descriptive statistics for nonparametric models. III // Dispersion. Ann. Statist. 1976. V. 4, No. 6. P. 1139–1158.
3. Hampel F.R. Contribution to the theory of robust estimation : Ph. D. diss. Berkeley : Univ. California, 1968. 103 p.
4. Hampel F.R. A general qualitative definition of robustness // Ann. Math. Statist. 1971. V. 42. P. 1887–1896.
5. Hampel F.R. The influence curve and its role in robust estimation // J. Amer. Statist. Assoc. 1974. V. 69, No. 346. P. 383–393.
6. Serfling R.J. Approximation Theorems of Mathematical Statistics. N.Y. : Wiley, 1980. 371 p.
7. Serfling R.J. Generalized L-M-R-statistics // Ann. Statist. 1984. V. 12. P. 76–86.
8. Rousseeuw P.J., Croux C. Alternatives to the Median Absolute Deviation // Journal of the American Statistical Association. 1993. V. 88, No. 424. P. 1273–1283.
9. Хампель Ф., Рончетти Э., Рауссей П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М. : Мир, 1989. 512 с.
10. Шуленин В.П. Исследование свойств оценки медианы абсолютных разностей // V Совещание-семинар по непараметрическим и робастным методам статистики в кибернетике. Томск, 1987. Ч. II. С. 460–467.
11. Шуленин В.П. Асимптотические свойства и робастность обобщенных L-оценок // Труды V Международной конференции по теории вероятности и математической статистике. Вильнюс, 1989. Т. 4. С. 377–378.
12. Шуленин В.П. Асимптотические свойства обобщенных L-оценок, вычисляемых по урезанным выборкам // Непараметрические и робастные статистические методы в кибернетике и информатике. Томск : Изд-во Том. ун-та, 1990. С. 564–570.
13. Шуленин В.П. Исследование устойчивости и асимптотических свойств урезанной средней разности Джини // Труды IV Международной конференции по теории вероятности и математической статистике. Вильнюс, 1985. С. 330–332.
14. Шуленин В.П. Введение в робастную статистику. Томск : Изд-во Том. ун-та, 1993. 227 с.
15. Шуленин В.П. Математическая статистика. Ч. 3 : Робастная статистика : учебник. Томск : Изд-во НТЛ, 2012. 520 с.

Шуленин Валерий Петрович, канд. техн. наук, доцент. E-mail: shvp@fpmk.tsu.ru
Томский государственный университет
Поступила в редакцию 15 февраля 2016 г.

Shulenin Valery P. (Tomsk State University, Russian Federation).
Asymptotic properties of robust estimators of scale parameters.
Keywords*:* scale parameter; robust estimator; influence function; asymptotic relative efficiency.

DOI: 10.17223/19988605/35/7

This paper deals with asymptotic robust properties of some estimators of scale parameter by the ε-contamination of the model distributions: $F = \Phi_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)$, ε is a known proportion of contamination $(0 < \varepsilon < 1/2)$, τ is a known scale parameter and Φ is the standard Gaussian distribution function.

Assume that $X_1,...,X_n$ is a random sample with distribution function $F(x)$ and F has a density $f(x)$, $x \in R^1$. Let $T(F)$, $F \in \Im$, is a generic scale functional and $T_n(X_1,...,X_n) = T(F_n)$ is its sample estimator. We consider the functional $T(F)$ defined by $\int [F(x+T(F)) - F(x-T(F)) - F(x)]dF(x) = 0$ and the location invariance and scale equivariance sample estimators of the functional $T(F)$, $F \in \Im$. The sample estimator of this functional $T(F)$ is given by $T_n(X_1,...,X_n) = med\{|X_i - X_j|,\ 1 \le i, j \le n\}$. This estimator is also named as the median of the absolute differences.

The purpose of this article is to study asymptotic robust properties T_n – estimators for different models distributions. The formal calculation of the Influence Function $IF(x;F,T)$ is given by

$$IF(x;F,T) = d_1T(F;\Delta_x - F) = \frac{1 + 2F(x-T) - 2F(x+T)}{2\int [f(x+T) + f(x-T)]dF(x)},\ x \in R^1.$$

Note that Influence Function $IF(x;F,T)$ is bounded and looks like as the U-shaped curve. If $\int [f(x+T) + f(x-T)]dF(x) > 0$, then the random variable $\sqrt{n}\{T_n - T(F)\}/\sigma(F,T_n)$ has asymptotically standard normal distribution, where the asymptotic variance of $\sqrt{n}\,T_n$ is given by the following formula:

$$\sigma^2(F,T_n) = \int_{-\infty}^{\infty} IF^2(x;F,T)dF(x) = \frac{\int [1 + 2F(x-T) - 2F(x+T)]^2 dF(x)}{4\left(\int [f(x+T) + f(x-T)]dF(x)\right)^2}.$$

The paper contains numerical comparisons for some estimators of scale parameters by ε – contamination of the model distribution for different values of ε and τ. It is shown that for normal distribution asymptotic relative efficiency T_n – estimator with respect to $\hat{S}_1$ having the classical standard deviation is equal: $ARE_\Phi(T_n : \hat{S}_1) = 0.86$ and $ARE_\Phi(T_n : \hat{S}_2) = 0.98$, where $\hat{S}_2$ has the average absolute deviation.

REFERENCES

1. Bickel, P.J. & Lehmann, E.L. (1974) Measures of location and scale. *Proc. Prague Symp. Asymptotic Statist.* 1. Prague Charles University. pp. 25–36.
2. Bickel, P.J. & Lehmann, E.L. (1976) Descriptive statistics for nonparametric models. III. Dispersion. *Annual Statistics*. 4(6). pp. 1139-1158.
3. Hampel, F.R. (1968) *Contribution to the theory of robust estimation*. Ph. D. Diss. Berkeley, Univ. California.
4. Hampel, F.R. (1971) A general qualitative definition of robustness. *Annual Mathematical Statistics* 42. pp. 1887-1896. DOI: 10.1214/aoms/1177693054
5. Hampel, F.R. (1974) The influence curve and its role in robust estimation. *Journal of American Statistical Association.* 69(346). pp. 383-393. DOI: 10.1080/01621459.1974.10482962
6. Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
7. Serfling, R. J. (1984) Generalized L-M-R-statistics. *Annual Statistics.* 12. pp. 76-86.
8. Rousseeuw, P.J. & Croux, C. (1993) Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*. 88(424). pp. 1273-1283. DOI: 10.1080/01621459.1993.10476408
9. Hampel, F., Ronchetti, E., Rousseeuw, P. & Stahel, W. (1989) *Robastnost' v statistike. Podkhod na osnove funktsiy vliyaniya* [Robust Statistics. The Approach Based on Influence Functions]. Moscow: Mir.
10. Shulenin, V.P. (1987) [Studying the properties of the median absolute differences of assessment]. *V Soveshchanie-seminar po neparametricheskim i robastnym metodam statistiki v kibernetike* [Proc. of the Fifth Conference on nonparametric and robust methods of statistics in cybernetics]. Tomsk. pp. 460-467. (In Russian).
11. Shulenin, V.P. (1989) [The asymptotic properties and robustness of generalized L-assessments]. *Trudy V Mezhdunarodnoy konferentsii po teorii veroyatnosti i matematicheskoy statistike* [Proc. of the Fifth International Conference on the Theory of Probability and Mathematical Statistics]. Vilnius. pp. 377-378. (In Russian).
12. Shulenin, V.P. (1990) Asimptoticheskie svoystva obobshchennykh L-otsenok, vychislyaemykh po urezannym vyborkam [The asymptotic properties of generalized L-estimates calculated by the trimmed samples]. In: Tarasenko, F.P. (ed.) *Neparametricheskie i robastnye statisticheskie metody v kibernetike i informatike* [Nonparametric and robust statistical methods in cybernetics and computer science]. Tomsk: Tomsk State University. pp. 564-570.
13. Shulenin, V.P. (1985) [Investigation of the stability and asymptotic properties of the truncated mean difference Gini]. *Trudy IV Mezhdunarodnoy konferentsii po teorii veroyatnosti i matematicheskoy statistike* [Proc. of the Fourth International Conference on the Theory of Probability and Mathematical Statistics] Vilnius. pp. 330-332. (In Russian).
14. Shulenin, V.P. (1993) *Vvedenie v robastnuyu statistiku* [Introduction into Robust Statistics]. Tomsk: Tomsk State University.
15. Shulenin, V.P. (2012) *Matematicheskaya statistika* [Math statistics]. Part 3. Tomsk: NTL.