А.Е.Авдюшина, В.А. Симахин
Курганский государственный университет, г. Курган

УДК 519.233.22

ИССЛЕДОВАНИЕ АДАПТИВНЫХ ОЦЕНОК ПАРАМЕТРА СДВИГА

Аннотация. Вводятся новые адаптивные (робастные непараметрические) оценки параметра сдвига для семипараметрических супермоделей Тьюки. Методом статистических испытаний проведён анализ потенциальной эффективности адаптивных оценок. Исследования показали высокую потенциальную эффективность адаптивных оценок на симметричных и ассиметричных супермоделях Тьюки.

Ключевые слова. Робастный, непараметрический, адаптивные оценки сдвига, супермодели Тьюки.

Развитие научных исследований характеризуется всё более возрастающей ролью принятия решений в условиях статистической неопределенности. На первый план выходит проблема построения интеллектуальных автоматизированных систем обработки данных в условиях полупараметрических и непараметрических математических моделей. Именно этим обусловлено активное развитие математических методов робастной и непараметрической статистик [1]- [3] и их применений [4], [5]. Классический подход к построению робастных процедур [1]- [2] на основе построения минимаксных решений на полупараметрических классах распределений (супермоделях) - это подход «пессимиста». Решения, полученные таким путем, обладают высокой робастностью на супермодели, но могут иметь катастрофически низкую эффективность для отдельных ситуаций из супермодели. Современный подход заключается в построении адаптивных процедур принятия решений [3]. В данной работе рассматриваются и исследуются новые адаптивные робастные оценки на полупараметрических классах распределений Тьюки. Проводится исследование методом статистических испытаний на примере параметра сдвига для различных моделей выбросов.

Пусть имеется выборка н.о.р. $\vec{X}_N = (x_1,...,x_N)$ с функцией распределения (ф.р.) $F(\vec{t},\theta) \in \mathrm{P}_\theta$ - супермодель Тьюки.

$$\mathrm{P}_\theta = \{F(x,\theta) = (1-\varepsilon)G(x,\theta) + \varepsilon \cdot H(x)\}, \qquad (1)$$

где $G(\vec{t},\theta) \in \mathrm{P}_{0\theta}$ - априорная модель ф.р., $H(x)$ - распределение выбросов, $\varepsilon \geq 0$ доля выбросов. Требуется по выборке $\vec{X}_N$ оценить неизвестный параметр θ.

Обозначим через $f(x), g(x), h(x)$ плотности распределений F, G, H, соответственно и $F_N(x)$ эмпирическую ф.р. Супермодель Тьюки используется в качестве модели реальных распределений $F(\vec{t},\theta)$, которые можно считать приближенно совпадающими с априорным распределением $G(\vec{t},\theta)$.

Для модели (1) оценка максимального правдоподобия (о.м.п.) θ_N параметра θ определяется из эмпирического уравнения

$$\int \varphi(x,\theta_N)dF_N(x) = 0, \qquad (2)$$

где оценочная функция $\varphi(x,\theta)$ представима в виде

$$\varphi(x,\theta) = \frac{\partial}{\partial\theta}\ln g(x,\theta) \cdot \frac{g(x,\theta)}{f(x,\theta)} = U(x,\theta) \cdot W(x,\theta), \qquad (3)$$

$U(x,\theta)$ - функция вклада о.м.п. для $G(\vec{t},\theta)$, $W(x,\theta)$ - весовая функция

$$W(x,\theta) = \frac{(1-\varepsilon)g(x,\theta)}{f(x,\theta)} = \left[1 + \frac{\varepsilon \cdot h(x)}{(1-\varepsilon)\cdot g(x,\theta)}\right]^{-1}, \qquad (4)$$

$$\int \varphi(x,\theta)dG(x) = 0, \qquad (5)$$

условие несмещенности оценки.

Исследование выражения (3) позволяет сделать вывод, что в рассматриваемом случае эффективная оценка представляет взвешенную о.м.п. априорного распределения $G(\vec{t},\theta)$ с весами $W(x,\theta)$ (3), которые и определяют расхождение между $G(\vec{t},\theta)$ и $\{H(x),\varepsilon\}$. Если вид $G(\vec{t},\theta)$ и $\{H(x),\varepsilon\}$ известны, то из соотношений (2) - (4) можно синтезировать эффективную взвешенную о.м.п. и свести решение к параметрической задаче. Оценка θ_N параметра θ из (2) является М-оценкой и можно показать [1] - [3], что $\sqrt{N}[\theta_N - \theta]$ имеет асимптотически нормальное распределение.

Так как вид $G(\vec{t},\theta)$ известен, то в (3) можно определить функцию вклада $U(x,\theta)=\frac{\partial}{\partial\theta}\ln g(x,\theta)$, но остаются неизвестными веса (4).

На рисунке 1 показано поведение весов $W(x,\theta)$ в разных ситуациях в случае нормальных распределений $G(\vec{t},\theta)$, $H(x)$ для оценки параметра сдвига.

Из Рис.1 (а) наглядно видно, что только в случае удаленных выбросов для получения о.м.п. необходима интуитивно понятная операция усечения функции вклада $U(x,\theta)$. Эвристические алгоритмы, основанные на операции усечения давно и широко используется для построения адаптивных робастных оценок [3]. Все усложняется, когда выбросы находятся внутри области определения $G(\vec{t},\theta)$(Рис.1(б, в)). В данном случае эвристически понятная операция усечения не приводит к эффективным оценкам.

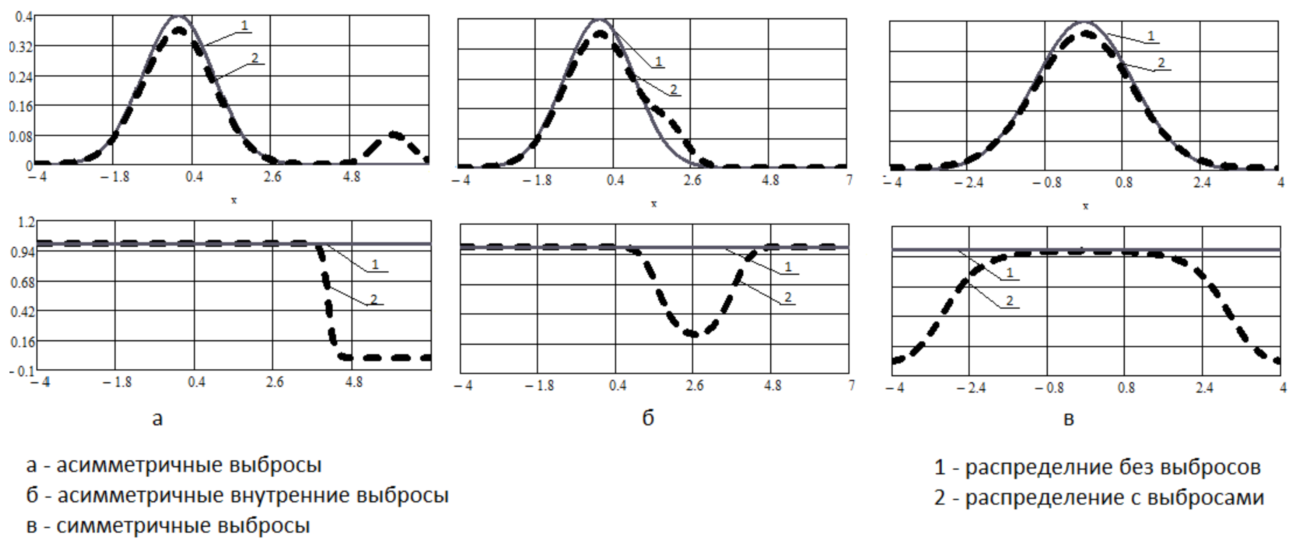

Рис.1

На основе выборки $\vec{X}_N=(x_1,\ldots,x_N)$ построим непараметрическую оценку плотности Розенблатта-Парзена

$$f_N(x,h_N)=\frac{1}{N\sqrt{2\pi}\cdot h_N}\sum_{i=1}^{N}\exp\left(-\frac{1}{2}\left(\frac{x-x_i}{h_N}\right)^2\right) \tag{6}$$

Для получения непараметрической оценки $W_N(x,\theta)$ весовой функции $W(x,\theta)$ подставим в выражение (4) оценку (6)

$$W_N(x,\theta)=\frac{(1-\varepsilon)g(x,\theta)}{f_N(x,\theta)} \tag{7}$$

Можно показать [3], что в результате получим состоятельную непараметрическую оценку $W_N(x,\theta)$ весовой функции $W(x,\theta)$. Адаптивная оценка максимального правдоподобия (а.о.) θ_N параметра θ определяется из эмпирического уравнения

$$\int \varphi_N(x,\theta_N)dF_N(x)=0\,, \tag{8}$$

где оценочная функция $\varphi_N(x,\theta)$ представима в виде

$$\varphi_N(x,\theta)=\frac{\partial}{\partial\theta}\ln g(x,\theta)\cdot\frac{g(x,\theta)}{f_N(x,\theta)}=U(x,\theta)\cdot W_N(x,\theta)\,, \tag{9}$$

Методом статистического моделирования исследовались свойства адаптивной оценки (8), (9) при следующих условиях:

Модель Тьюки – суперпозиция нормальных распределений с $\theta=0,\varepsilon=0{,}1$

$$f(x,\theta)=0{,}9\cdot\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]+0{,}1\cdot\frac{1}{\sqrt{2\pi\cdot\lambda}}\exp\left[-\frac{(x-\mu)^2}{2\cdot\lambda}\right] \tag{10}$$

1. Модель1 – без выбросов (**БВ**) - $\theta=0,\varepsilon=0$;
2. Модель2 – симметричные выбросы (**СВ**) - $\theta=0,\varepsilon=0{,}1,\mu=0,\lambda=3$;
3. Модель3 - ассимметричные выбросы (**АВ1**) - $\theta=0,\varepsilon=0{,}1,\mu=6,\lambda=1$;
4. Модель3 - ассимметричные выбросы (**АВ2**) - $\theta=0,\varepsilon=0{,}1,\mu=2,\lambda=1$;

Для всех моделей находились доверительные интервалы (ДИ), средние квадратические ошибки СКО($\theta_{МП}$) эффективных оценок максимального правдоподобия, средние квадратические ошибки СКО($\theta_{АО}$) адаптивных оценок (8), (9) для N=100, M=500 – число итераций моделирования.

Вычислялась потенциальная эффективность $E^*=\frac{СКО_{МП}}{СКО_{АО}}$ для разных ситуаций и доверительные интервалы ДИ для адаптивной оценки.

Результаты моделирования приводятся в таблице 1

Таблица 1

Потенциальная эффективность E^* адаптивной оценки.

	БВ	СВ	АС1	АС2
ДИ	[-0.002; 0.016]	[-0.023; 0.015]	[0.154; 0.171]	[0.135; 0.152]
E^*	1	0.91	0.83	0.86

На рис. 2 приведены графики весовых функций потенциальной (4) и оценочной (7).

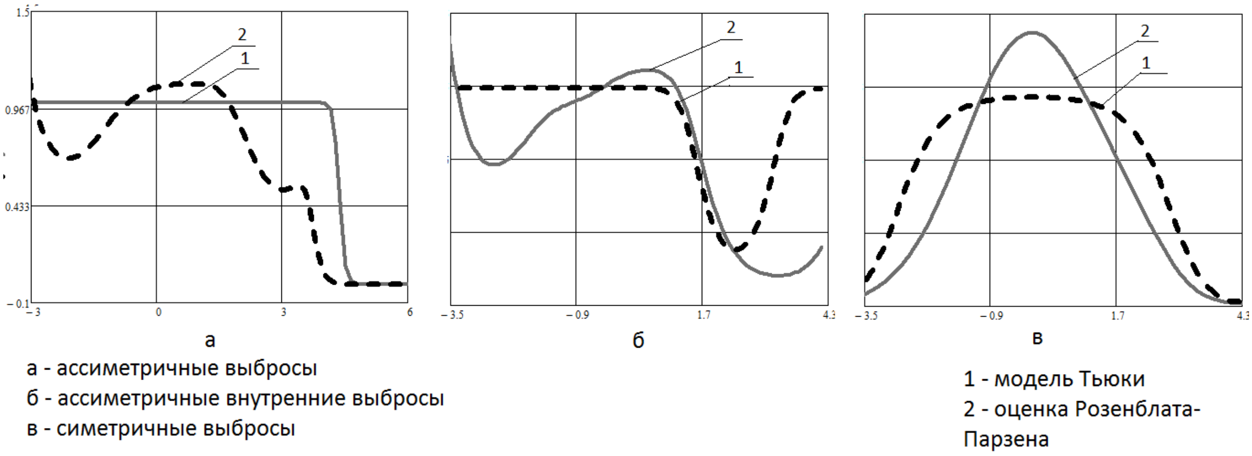

Рис.2

Оценка плотности Розенблатта-Парзена и моделирование производились в системе **R.** Известно [3], что эффективность оценки плотности Розенблатта-Парзена (6) сильно зависит от выбора параметра размытости h_N, поэтому при выборе h_N используют различные критерии оптимальности. Как следует из анализа рис.2 выбор h_N в системе **R** производится не наилучшим образом.

Выводы.

Результаты моделирования показывают, что исследуемые новые адаптивные оценки имеют высокую потенциальную эффективность для нормальной модели Тьюки. При оптимальном выборе h_N потенциальная эффективность адаптивных оценок может быть существенно повышена.

СПИСОК ЛИТЕРАТУРЫ

1. Хьюбер П. Робастность в статистике. – М.: Мир, 1984, 303 с.
2. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. – М.: Мир, 1989, 512 с.
3. Симахин В. А. Робастные непараметрические оценки. – Germany: LAMBERT Academic Publishing, 2011. – 292 с.
4. Сызранцев В. Н., Невелев Я. П., Голофаст С. Л. Расчет прочностной надежности изделий на основе методов непараметрической статистики. – Новосибирск: Наука, 2008, 216 с.
5. Симахин В.А., Черепанов О.С., Шаманаева Л.Г. Пространственно-временная динамика скорости ветра по результатам мини-содарных измерений // Известия вузов. Физика. 2015. Т. 58. № 12. С. 176–181.