АДАПТИВНЫЕ MD-ОЦЕНКИ КРАМЕРА – МИЗЕСА

В. П. Шуленин

Томский государственный университет
Томск, Россия
E-mail: shulenin-vp@rambler.ru

Оценки параметров, построенные методом минимума расстояний, в литературе кратко называют *MD*-оценками. Метод минимума расстояний был предложен Вольфовитцем [1] в 1957 году. Обширная библиография работ по *MD*-оценкам составлена и опубликована Парром [2]. В данной работе рассматриваются эффективные *MD*-оценки параметра сдвига, основанные на использовании взвешенного расстояния Крамера – Мизеса и изучаются их свойства робастности в различных супермоделях, описывающих отклонения от гауссовской модели. Предлагается адаптивный вариант оценки Крамера – Мизеса, для которой весовая функция выбирается с использованием выборочной оценки функционала, характеризующего степень затянутости «хвостов» распределений.

Ключевые слова: эффективные, робастные, адаптивные оценки параметров, метод минимума расстояний.

Пусть статистическая модель эксперимента, связанного с изучением случайной величины X задана в параметрической форме, то есть в виде $(\mathrm{X}, \Im_\theta)$, где $\mathrm{X} = \{\vec{x}\}$ обозначает выборочное пространство, элементами которого является наблюденные реализации $\vec{x} = (x_1,..., x_n)$ случайного вектора $\vec{X} = (X_1,..., X_n)$, а $\Im_\theta = \{F : F(x,\theta), \theta \in \Theta\}$ – заданное параметрическое множество допустимых распределений вероятностей в условиях данного эксперимента, X_1, …, X_n – последовательность н.о.р. случайных величин с функцией распределения $F(x,\theta)$ и с плотностью $f(x,\theta)$, $x \in R^1$, $\theta \in \Theta$. Функциональный вид распределения задан с точностью до *неизвестного параметра* θ (скалярного, либо векторного), который принадлежит заданному параметрическому множеству $\theta \in \Theta$. Требуется построить по выборке X_1, …, X_n, порождённой распределением $F(x,\theta)$, оценку неизвестного параметра $\theta \in \Theta$.

Метод минимума расстояний состоит в том, что на множестве непрерывных функций распределений $\Im$, для пары распределений F, $G \in \Im$, задается метрика (или расстояние) $\rho(F,G)$. Оценка параметра θ, полученная методом выбранного расстояния $\rho(F,G)$, определяется из условия минимума этого расстояния между эмпирической функцией распределения $F_n(x)\, F_n(x)$, построенной по выборке X_1, …, X_n, и функцией распределения $F_\theta(x) = F_X(x,\theta)$, принятой в модели $(\mathrm{X}, \Im_\theta)$. *MD*-оценка параметра θ определяется для выбранного расстояния $\rho(F,G)$ выражением вида $\hat{\theta} = \arg\min_\theta \{\rho(F_n, F_\theta)\}$. При построении *MD*-оценок могут быть использованы различные расстояния (см. [5]). Заметим, что метод максимального правдоподобия основан на использовании расстояния вида $\rho(F_n, F_\theta) = -\int \ln f(x,\theta) dF_n(x)$.

В данной работе мы рассмотрим *MD*-оценки, которые основаны на использовании взвешенного расстояния Крамера – Мизеса, определяемого при $G = F_n$, в виде

$$\rho_W(F_n, F_\theta) = \int [F_n(x) - F_\theta(x)]^2 W_\theta(x, F_\theta)\, dF_\theta(x), \qquad (1)$$

где $W_\theta = W(x, F_\theta)$ – заданная весовая функция, которая в общем случае может зависеть от ф.р. F_θ (или от плотности f_θ). Предполагая, что $\rho_W(F_n, F_\theta)$ дифференцируемая по параметру θ функция, обозначим её производную через $\tilde{\lambda}_{F_n}(\theta) = \partial\rho_W(F_n, F_\theta)/\partial\theta$. С учетом этого обозначения, *MD*-оценка θ_n параметра θ, основанная на использовании взвешенного расстояния Крамера – Мизеса вида (1), является решением уравнения $\tilde{\lambda}_{F_n}(\theta) = 0$, где

$$\tilde{\lambda}_{F_n}(\theta) = -2\int [F_n(x) - F_\theta(x)] \frac{\partial F_\theta(x)}{\partial\theta} W_\theta(x) dF_\theta(x) + \int [F_n(x) - F_\theta(x)]^2 \frac{\partial}{\partial\theta}[W_\theta f_\theta(x)] dx. \qquad (2)$$

В данной работе рассмотрим *MD*-оценки параметра сдвига θ в одновыборочном варианте, то есть в этом случае $F_\theta(x) = F(x-\theta)$. Введем опорное семейство распределений в виде $\Im_0 = \{F : F_\theta(x) = F_0(x-\theta),\ \theta \in R^1\}$, где F_0 – заданное опорное распределение с плотностью f_0. Перепишем (1) в виде

$$\rho_{F_n,F_0}(\theta, W) = \int [F_n(x) - F_0(x-\theta)]^2 W(x-\theta) dx. \qquad (3)$$

Отметим, что выбор весовой функции *W* в виде плотности опорного распределения, то есть в виде $W(x) = f_0(x)$, соответствует расстоянию Крамера – Мизеса, выбор весовой функции $W(x) = f_0(x)/F_0(x)(1 - F_0(x))$ соответствует расстоянию Андерсона – Дарлинга (см., например [2 – 4]). Предполагая, что $\rho_{F_n,F_0}(\theta, W)$ из (3) дифференцируемая по параметру θ функция, обозначим её производную через $\lambda_{F_n}(\theta) = \partial\rho_{F_n,F_0}(\theta, W)/\partial\theta$. Уравнение $\lambda_{F_n}(\theta) = 0$ для нахождения *MD*-оценки записывается в виде

$$\frac{2}{n}\sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F_0(X_{(i)} - \theta)\right] W(X_{(i)} - \theta) = 0, \qquad (4)$$

где $X_{(1)},..., X_{(n)}$ – упорядоченная статистика исходной выборки $X_1,..., X_n$.

Асимптотические свойства *MD*-оценок изучались различными авторами (см., например [3, 4, 7]). В данной работе обсуждаются асимптотические свойства *MD*-оценок θ_n параметра сдвига θ, которые при заданной опорной ф.р. F_0, принятой в качестве исходной модели, и заданной весовой функции *W* являются решением уравнения (4). При этом различаются следующие два варианта оценивания параметра θ.

Вариант 1. Функция распределения *F* наблюдений $X_1,..., X_n$ *известна* и она совпадает с опорной функцией распределения F_0, то есть $F = F_0$ (или $F \in \Im_0$).

Вариант 2. Функция распределения *F* наблюдений $X_1,..., X_n$ *неизвестна* и она не обязательно совпадает с опорной функцией распределения F_0, то есть $F \neq F_0$ (или $F \notin \Im_0$).

Отметим, что *MD*-оценки θ_n параметра сдвига θ, которые являются решением уравнения (4), могут быть записаны в виде функционала от эмпирической функции распределения, то есть в виде $\theta_n = \theta(F_n)$, где функционал $\theta(F)$ определяется выражением

$\min\limits_{\theta} \rho_{F,F_0}(\theta, W) = \rho_{F,F_0}(\theta(F), W)$ или, с принятым обозначением $T(F)$ для функционала, этот функционал $T(F) = \theta(F)$ задается неявно с помощью выражения вида

$$2\int [F(x+T(F)) - F_0(x)] f_0(x) W(x) dx - \int [F(x+T(F)) - F_0(x)]^2 W'(x) dx = 0 . \quad (5)$$

Для изучения асимптотических свойств *MD*-оценок $\theta_n = \theta(F_n)$ параметра сдвига θ воспользуемся подходом Мизеса (см. [10, 14]) и рассмотрим разложение вида

$$\theta(F_n) = \theta(F) + V_{1n} + R_{1n} , \quad (6)$$

где $R_{1n} = \theta(F_n) - \theta(F) - V_{1n}$ – остаточный член разложения (6) и V_{1n} – аппроксимационная статистика, которая определяется через дифференциал Гато первого порядка $d_1T(F; G-F)$ функционала $T(F)$, заданного выражением (5), в виде

$$V_{1n} = d_1T(F; F_n - F) = n^{-1} \sum IF(X_i; F, F_0, W) ,$$

здесь $IF(u; F, F_0, W)$, $0 \le u < \infty$, – функция влияния Хампеля для *MD*-оценки $\theta_n = \theta(F_n)$ параметра сдвига θ, которая при заданной опорной ф.р. F_0 и заданной весовой функции W является решением уравнения (4).

Отметим, что общие условия регулярности, накладывающие ограничения на поведение хвостов ф.р. F и весовой функции W, при которых выполняется выражение $\sqrt{n}R_{1n} \to^p 0$, $n \to \infty$, и при которых *MD*-оценки состоятельны и асимптотически нормальны, приводятся в [3, 4]. Отметим также, что рассматриваемые здесь *MD*-оценки входят в семейство MD_α-оценок, которые вычисляются по «α-урезанной выборке» и асимптотические свойства которых обсуждаются в [7].

Для формулировки дальнейших результатов обозначим через $\Im_S$ семейство абсолютно непрерывных симметричных распределений. Выделим класс W_S весовых функций, для которых предполагаем, что они четные, дифференцируемые, то есть $W(-x) = W(x)$, и $\int \{F(x)(1-F(x))\}^p W(x+c) dx < \infty$, $p > 0$, $c \in (-\infty, +\infty)$.

Теорема. Пусть $(F, F_0) \in \Im_S$ и $W \in W_S$. Тогда при выполнении неравенств

$$0 < \sigma^2(F; F_0, W) = \int IF^2(x; F, F_0, W) dF(x) < \infty$$

выполняется асимптотическое выражение вида

$$L\{\sqrt{n}[\theta(F_n) - \theta(F)] / \sigma(F; F_0, W)\} = N(0, 1), \quad n \to \infty ,$$

где асимптотическая дисперсия *MD*-оценки с опорной ф.р. F_0 и весовой функцией W при распределении F наблюдений $X_1, ..., X_n$ равна $D(F; F_0, W) = \sigma^2(F; F_0, W)/n$ и функция влияния Хампеля $IF(u; F, F_0, W) = -IF(-u; F, F_0, W)$ для *MD*-оценки вычисляется по формулам

$$IF(u; F, F_0, W) = A_{F,F_0}(u; W) / B_{F,F_0}(W), \quad 0 \le u < \infty , \quad (7)$$

$$A_{F,F_0}(u; W) = \int_0^u W(x) dF(x) - W(u)[F(u) - F_0(u)] , \quad (8)$$

$$B_{F,F_0}(W) = \int_{-\infty}^{\infty} f_0(x) W(x) dF(x) - \int_{-\infty}^{\infty} [F(x) - F_0(x)] W'(x) dF(x) . \quad (9)$$

Доказательство этой теоремы может быть найдено в [3, 4, 10].

Отметим, что для первого варианта оценивания параметра θ, когда $F \in \Im_0$, функция влияния $IF(u;F,W)$, $0 \le u < \infty$, определяется выражением

$$IF(u;F,W) = \frac{\int_{-\infty}^{+\infty}\{F(x) - I[u \le x]\}W(x)dF(x)}{\int_{-\infty}^{+\infty} f^2(x)W(x)dx} = \frac{\int_0^u W(x)dF(x)}{\int_{-\infty}^{+\infty} f(x)W(x)dF(x)},\ 0 \le u < \infty. \quad (10)$$

и асимптотическая дисперсия $\sqrt{n}\,MD$-оценки вычисляется по формуле

$$\sigma^2(F,W) = \frac{\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\{F(y) - I[u \le y]\}W(y)dF(y)\right)^2 dF(u)}{\left(\int_{-\infty}^{+\infty} f(x)W(x)dF(x)\right)^2} = \frac{\int_{-\infty}^{+\infty}\left(\int_0^x W(y)dF(y)\right)^2 dF(x)}{\left(\int_{-\infty}^{+\infty} f^2(x)W(x)dx\right)^2}. \quad (11)$$

Отметим также, что для первого варианта оценивания параметра θ, когда функция распределения F наблюдений $X_1, \ldots, X_n$ известна и совпадает с опорной симметричной функцией распределения F_0, в классе *MD*-оценок существует эффективная оценка параметра θ, асимптотическая дисперсия которой равна обратной величине информации Фишера $I(f)$. Эта эффективная *MD*-оценка определяется весовой функцией $W^*(x)$ вида

$$W^*(x) = a\ \frac{d^2\{-\ln f_0(x)\}}{dx^2} \cdot \frac{1}{f_0(x)}. \quad (12)$$

Изучение эффективности и свойств робастности *MD*-оценок (см. работы [6, 8–12]) при изменении распределения вероятности наблюдений в некотором заданном классе (в рамках заданной супермодели) показывает, 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 [10, 11, 13]) описаны различные подходы для упорядочивания распределений в заданном классе по степени тяжести хвостов. Следуя работе [13], рассмотрим количественную меру «тяжести хвостов» распределения $F(x)$, $x \in R^1$ в виде функционала $Q(F;\nu,\mu)$

$$Q(F;\nu,\mu) = \frac{(1/\nu)\{\int_{1-\nu}^{1} F^{-1}(t)dt - \int_0^{\nu} F^{-1}(t)dt\}}{(1/\mu)\{\int_{1-\mu}^{1} F^{-1}(t)dt - \int_0^{\mu} F^{-1}(t)dt\}},\ \ 0 < \nu < \mu \le 0{,}5.$$

Значения этого функционала для различных супермоделей вычислены в [10]. Изучение асимптотических свойств *MD*-оценок показало, что качество этих оценок существенно зависит от выбора весовой функции *W*. Выбор этой весовой функции можно связать с поведением функционала $Q(F;\nu,\mu)$, который характеризует степень «тяжести хвостов» распределений при их изменении в заданной супермодели. Однако на практике функция распределения F наблюдений $X_1,\ldots,X_n$ обычно неизвестна, поэтому естественно использовать вместо функционала $Q(F;\nu,\mu)$ его выборочную оценку, построенную по исходной выборке $X_1,\ldots,X_n$. Выборочная оценка $Q(F_n)$ функционала $Q(F;\nu,\mu)$, построенная по выборке $X_1,\ldots,X_n$ методом подстановки, записывается в виде

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

где $0 < \nu < \mu \le 0{,}5$ и $X_{(1)},..., X_{(n)}$ – порядковые статистики выборки $X_1,..., X_n$. Следуя работе [13], везде ниже полагаем $\nu = 0{,}2$ и $\mu = 0{,}5$. Отметим, что $Q(F_n) \xrightarrow{p} Q(F;\nu,\mu)$ при $n \to \infty$. Кроме того, результаты моделирования (см., например [10]) показали, что уже при объемах выборки $n \ge 20$ статистика $Q(F_n)$ может быть использована для определения типов распределений, различающихся их степенью «тяжести хвостов». Итак, сопоставляя поведение абсолютных эффективностей оценок в зависимости от выбора весовой функции *W*, и поведение функционала $Q_F(\nu;\mu)$ в рамках заданной супермодели, можно предложить процедуру выбора *W* на основе информации, содержащейся в исходной выборке $X_1,..., X_n$, точнее на основе выборочной оценки $Q(F_n)$ функционала $Q(F;\nu,\mu)$, построенной по исходной выборке $X_1,..., X_n$ из распределения *F*.

В качестве примера рассмотрим супермодель с засорением

$$\Im_{\varepsilon,\tau}(\Phi) = \{F : F(x) = \Phi_{\varepsilon,\tau}(x)\},$$

где $\Phi_{\varepsilon,\tau}(x) = (1-\varepsilon)\Phi(x) + \varepsilon\Phi(x/\tau)$, здесь $\Phi(x)$ – стандартная нормальная функция распределения и ϕ обозначает её плотность. Предполагаем, что пропорция засорения ε может изменяться и удовлетворяет неравенствам $0 \le \varepsilon \le 0{,}3$, а масштабный параметр τ известен и $\tau = 3$. В рамках данной супермодели свойства *MD*-оценок существенно зависят от выбора весовой функции *W*. В связи с этим, адаптивный выбор весовой функции *W* позволяет обеспечивать требуемые качества *MD*-оценки для заданной супермодели. При построении *MD*-оценки в рамках супермодели $\Im_{\varepsilon,\tau}(\Phi)$, которая описывает отклонения от гауссовской модели наблюдений, выберем опорное распределение $F_0 = \Phi$, и определим адаптивную весовую функцию в виде

$$\hat{W}(x;\ X_1,...,X_n) = \begin{cases} 1/\phi(x), & 1{,}71 < Q(F_n) \le 1{,}76; \\ 1, & 1{,}76 < Q(F_n) \le 1{,}86; \\ \phi(x), & 1{,}86 < Q(F_n) \le 1{,}91. \end{cases}$$

При таком выборе весовой функции $\hat{W}(x;\ X_1,...,X_n)$, абсолютная эффективность $АЭ(F,W) = \{\sigma^2(F,F_0,W)I(f)\}^{-1}$ адаптивной *MD*-оценки не опускается ниже уровня 0,95 при изменении пропорции засорения ε, $0 \le \varepsilon \le 0{,}3$, то есть в рамках заданной супермодели для абсолютной эффективности выполняются неравенства $0{,}95 \le АЭ(\Phi_{\varepsilon,\tau},\hat{W}) \le 1$ при $\tau = 3$, $0 \le \varepsilon \le 0{,}3$, $n \ge 40$. Если же не использовать адаптивную процедуру, и выбрать оптимальную весовую функцию в виде (12) для гауссовской модели, то есть в виде $W^*(x) = 1/\phi(x)$, то абсолютная эффективность *MD*-оценки в рамках супермодели $\Im_{\varepsilon,\tau}(\Phi)$ может опуститься до уровня 0,55. Если же выбрать весовую функцию Андерсона – Дарлинга в виде $\tilde{W}(x,\phi) = \phi(x)/\Phi(x)(1-\Phi(x))$, то абсолютная эффективность *MD*-оценки с такой весовой функцией в рамках данной супермодели с засорением может опуститься до уровня 0,47.

ЛИТЕРАТУРА

1. *Wolfowitz, J.* The minimum distance method / J. Wolfowitz // Ann. Math. Statist. 1957. V. 28. P. 75–88.

2. *Parr, W. C.* Minimum distance estimation: a bibliography / W. C. Parr / Comm. Statist. 1981. A10. P. 1205–1224.

3. *Boos, D. D.* Minimum distance estimators for location and goodness of fit / D. D. Boos // J. Amer. Statist. Assoc. 1981. V. 76. № 375. P. 663–670.

4. *Parr, W. C.* On minimum weighted Cramer-von Mists statistical estimation / W. C. Parr, De Wet // Comm. Statist. 1981. A10(12). P. 1149–1166.

5. *Parr, W. S.* Minimum distance and robust estimation / W. S. Parr, W. R. Schucany // J. Amer. Statist. Assoc. 1980. V. 75, № 371. P. 616–624.

6. *Shulenin, V. P.* Connections of MD-estimates with classes of robust estimates of location parameter / V. P. Shulenin, F. P. Tarasenko // 12th Prague Conf. on Inform. Theory. August 29 – September 2, 1994. P. 220–223.

7. *Шуленин, В. П.* Асимптотические свойства и робастность MD-оценок // В. П. Шуленин //. Теория вероятностей и её применение. 1992. Т. 37. В. 4. С. 816–818.

8. *Шуленин, В. П.* Границы эффективности оценок, построенных методом минимума расстояния Крамера-Мизеса / В. П. Шуленин // Известия вузов Физика. 1995. № 9. С.84–89.

9. *Шуленин, В. П.* Эффективные и робастные MD-оценки Крамера–Мизеса / В. П. Шуленин // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 3(12). С. 107–121.

10. *Шуленин, В. П.* Математическая статистика. Ч. 3. Робастная статистика: учебник / В. П. Шуленин. Томск : Изд-во НТЛ, 2012. 520 с.

11. *Шуленин, В. П.* Введение в робастную статистику / В. П. Шуленин. Томск : Изд-во Томского ун-та, 1993. 227 с.

12. *Shulenin, V. P.* On estimation of parameters by the minimum distance method / V. P. Shulenin // Reliability: Theory and Applications. Gnedenko-Forum, San Diego, 2013. V. 8. № 2. P. 24–38.

13. *Hogg, R. V.* Adaptive robust procedures: A partial review and some suggestions for future applications and theory / R. V. Hogg // J. Amer. Statist. Assoc. 1974. V. 69. P. 909–923.

14. *Serfling, R. J.* Approximation Theorems of Mathematical Statistics / R. J. Serfling. N.-Y. : Wiley, 1980. 371 p.