

DOI: 10.18721/JPM.10309

УДК 519.246.2

ВЫСОКОЭФФЕКТИВНЫЕ РОБАСТНЫЕ M -ОЦЕНКИ ПАРАМЕТРА МАСШТАБА НА БАЗЕ Q -ОЦЕНКИ

П.О. Смирнов, И.С. Широков, Г.Л. Шевляков

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Широко используемая высокоэффективная робастная Q -оценка параметра масштаба, предложенная в работе Руссива и Крукса (1993), аппроксимирована с помощью «быстрых» хьюберовских M -оценок. Показано, что предложенные нами M -оценки являются высокоэффективными и робастными на произвольном распределении, благодаря правильному выбору параметров аппроксимации. Особое внимание уделено случаям гауссовского распределения и распределения Коши.

Ключевые слова: M -оценка; Q -оценка; робастность; параметр масштаба; гауссовское распределение; распределение Коши

Ссылка при цитировании: Смирнов П.О., Широков И.С., Шевляков Г.Л. Высокоэффективные робастные M -оценки параметра масштаба на базе Q -оценки // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2017. Т. 10. № 3. С. 95–99. DOI: 10.18721/JPM.10309

HIGHLY-EFFICIENT AND ROBUST M -ESTIMATES OF THE SCALE PARAMETER ON THE Q -ESTIMATE BASIS

P.O. Smirnov, I.S. Shirokov, G.L. Shevlyakov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

The highly efficient and robust Q -estimate of the scale parameter proposed by Rousseeuw and Croux (1993) and commonly employed has been approximated using computationally fast Huber M -estimates. The suggested M -estimates were shown to be robust and highly efficient for an arbitrary underlying data distribution due to right choosing the approximation parameters. The following indicators of the efficiency and robustness of M -estimates of scale were computed: their asymptotic variances, influence functions and breakdown points. A special attention was given to the particular cases of the Gaussian and Cauchy distributions. It is noteworthy that for the Cauchy distribution, the suggested robust estimate of scale coincides with the maximal likelihood estimate. Finally, the computation time of these highly-efficient and robust estimates of scale is 3–4 times less than for the corresponding Q -estimates.

Key words: M -estimate; Q -estimate, robustness; scale parameter; Gaussian distribution; Cauchy distribution

Citation: P.O. Smirnov, I.S. Shirokov, G.L. Shevlyakov, Highly-efficient and robust M -estimates of the scale parameter on the Q -estimate basis, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 10 (3) (2017) 95–99. DOI: 10.18721/JPM.10309

Введение

Проблема оценивания параметра масштаба является одной из наиболее важных в статистическом анализе [1, 2, 6, 8, 10]. В настоящее время наилучшей считается ро-

бастная высокоэффективная Q_n -оценка параметра масштаба (n – размер выборки) [7]. Эта оценка определяется как первый квартиль расстояния между наблюдениями:

$$Q_n = c\{ |x_i - x_j| \}_{(k)},$$

где c – константа, обеспечивающая состоятельность оценки; $k = C_h^2$ ($h = [n/2] + 1$).

Q_n -оценка является робастной, с максимально возможной пороговой точкой $\varepsilon^* = 0,5$ и высокой для нормального распределения асимптотической эффективностью (82 %). К ее недостаткам следует отнести высокую асимптотическую сложность алгоритма вычисления, требующую $O(n \ln(n))$ времени и столько же памяти.

Напротив, для робастных M -оценок параметра масштаба характерна низкая вычислительная сложность. Кроме того, имеются возможности для повышения их эффективности.

Исходя из этого, целями данной работы являлись построение робастной высокоэффективной аппроксимации Q_n -оценки с низкой вычислительной сложностью и адаптация полученной аппроксимации к различным типам распределений.

Постановка задачи

Рассмотрим класс M -оценок $\hat{S}$ параметра масштаба, заданных решением уравнения [5]:

$$\sum \chi(x_i / \hat{S}) = 0, \quad (1)$$

где $\chi(x)$ – оценочная функция, которая обычно является четной и неубывающей для $x > 0$.

Аппроксимация Q_n -оценки

Общий случай. Важным инструментом статистического анализа оценок служит функция влияния $IF(x; S, F)$, которая определяет меру изменения значения функционала оценки $S = S(F)$ на распределении F от «засорения» в точке x [4]. При известной функции влияния асимптотическая дисперсия оценки $\hat{S}$ имеет вид

$$AV(\hat{S}, F) = \int IF(x; S, F)^2 dF(x).$$

Класс M -оценок параметра масштаба (1) имеет важное свойство: функция влияния $IF(x; S, F)$ с точностью до коэффициента равна оценочной функции $\chi(x)$:

$$IF(x; S, F) \propto \chi(x).$$

Таким образом, можно построить M -оценку с любой допустимой функцией

влияния и, соответственно, эффективно-стью.

Известно, что функция влияния Q_n -оценки задается следующим выражением [7]:

$$IF(x; Q, F) = c[(1/4) - F(x + c^{-1}) + F(x - c^{-1})] [\int f(y + c^{-1}) f(y) dy]^{-1}. \quad (2)$$

Поскольку оценочная функция χ в выражении (1) определяется с точностью до произвольного множителя, нормирующий интеграл в выражении (2) можно опустить. Тогда Q_n -оценке соответствует M -оценка, порожденная оценочной функцией

$$\chi_Q(x) = \frac{c}{4} - c \cdot (F(x + c^{-1}) - F(x - c^{-1})), \quad (3)$$

и, следовательно, функция влияния $IF(x; \chi_Q, F)$ совпадает с $IF(x; Q, F)$, обеспечивая совпадение и производных от нее характеристик.

Преобразуем выражение (3): произведем замену $\alpha = c^{-1}$, при этом не будем фиксировать α , рассматривая эту величину как параметр настройки оценки; далее разложим функцию распределения F в ряд Тейлора, оставив только первые три члена:

$$F(x \pm \alpha) = F(x) \pm \alpha f(x) + \frac{1}{2} \alpha^2 f'(x) \pm \frac{1}{6} \alpha^3 f''(x) + o(\alpha^3). \quad (4)$$

Объединяя выражения (3) и (4), получим следующий результат.

Определение. Пусть плотность вероятности $f(x)$ – аналитическая на $\mathbf{R}$ функция. Однопараметрическое семейство M -оценок с оценочными функциями вида

$$\chi_\alpha(x) = c_\alpha - 2f(x) - \frac{1}{3} \alpha^2 f''(x) \quad (5)$$

назовем семейством MQ_n -оценок.

Константа c_α в выражении (5) обеспечивает состоятельность MQ_n -оценок.

Распределение Гаусса. Рассмотрим M -оценку в случае гауссовской плотности распределения:

$$f(x) = \varphi(x) = (2\pi)^{-1/2} \exp(-x^2 / 2).$$

Тогда

$$\varphi''(x) = (x^2 - 1)\varphi(x),$$

и оценочная функция принимает следую-

щий вид:

$$\chi_{\alpha}(x) = c_{\alpha} - \frac{1}{3}(6 + \alpha^2(x^2 - 1))\varphi(x),$$

$$c_{\alpha} = \frac{12 - \alpha^2}{12\sqrt{\pi}}. \quad (6)$$

В важном частном случае при $\alpha = 0$ выражение (6) принимает вид:

$$\chi_0(x) = \frac{1}{\sqrt{\pi}} - 2\varphi(x). \quad (7)$$

Этот результат напоминает обобщенную оценку Уэлша [3], задаваемую выражением

$$\chi(x) = \sqrt{\frac{d}{d+2}} - \exp\left(-\frac{x^2}{d}\right), \quad d > 0.$$

При $d = 2$ эта оценка совпадает с M -оценкой параметра масштаба, определенной выражением (7). Максимально возможное значение асимптотической дисперсии для оценки, задаваемой выражением (7), равно 95,9 %.

Отсюда мы имеем следующий результат.

Теорема 1. MQ_n -оценки на гауссовском распределении при $\alpha \in [0; \sqrt{2}]$ являются B -робастными с ограниченной функцией влияния, задаваемой выражением

$$IF(x; MQ, \Phi) = \frac{2(12 - \alpha^2) - 8\sqrt{\pi}(6 + \alpha^2(x^2 - 1))\varphi(x)}{3(4 - \alpha^2)}.$$

Доказательство теоремы приведено в работе [9]. Оно основано на интегрировании оценочной функции рассматриваемой MQ_n -оценки.

Асимптотическая эффективность подобных MQ_n -оценок с оценочной функцией вида (7) составляет 81 %, что лишь на 1 % меньше асимптотической эффективности Q_n -оценок на гауссовском распределении. При этом достигается существенный

выигрыш в скорости вычислений [9].

Распределение Коши. Теперь рассмотрим MQ_n -оценки для распределения Коши, имеющего «тяжелые хвосты», с оценочной функцией вида (5). В этом случае плотность распределения принимает вид

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Чтобы получить небольшую сложность алгоритма, рассмотрим MQ_n -оценку в случае $\alpha = 0$. Остальные значения параметра нас не интересуют, так как дают худшие результаты.

Теорема 2. MQ_n -оценка на распределении Коши, определенная выражением (5), является оценкой максимального правдоподобия для данного распределения и имеет оценочную функцию вида

$$\chi_0(x) = \frac{1}{\pi} \cdot \frac{x^2 - 1}{x^2 + 1}.$$

Доказательство теоремы также приведено в работе [9].

Максимально возможное значение асимптотической эффективности такой оценки параметра масштаба равно 100 % на распределении Коши, но для гауссовского распределения оно не превосходит 50 %.

Заключение

Для широкого класса распределений предложены робастные MQ_n -оценки параметра масштаба, имеющие небольшую сложность алгоритма и высокую скорость вычислений. Предложенные оценки близки по своей эффективности к высокоэффективной робастной Q_n -оценке.

Важные случаи распределений Гаусса и Коши теоретически изучены в асимптотике.

Предложенный нами подход планируется применить на параметрических семействах t - и экспоненциально-степенных распределений.

СПИСОК ЛИТЕРАТУРЫ

1. Balakrishnan N., Cohen A.C. Order statistics & inference: estimation methods. USA: Elsevier, 2014. 377 p.
2. Cramér H. Mathematical methods of statistics (PMS-9). USA: Princeton University Press, 2016. Vol. 9.
3. Genton M.G. Asymptotic variance of M -estimators for dependent Gaussian random variables // Statistics and Probability Letters. 1998. Vol. 38. No. 3. Pp. 255–261.
4. Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. Robust statistics: the approach

based on influence functions. USA, New Jersey: John Wiley, 1986.

5. **Huber P.J.** Robust statistics, USA, New Jersey: John Wiley, 1981.

6. **Launer R.L., Wilkinson G.N.** (eds.). Robustness in statistics. USA: Academic Press, 2014. 296 p.

7. **Rousseeuw P.J., Croux C.** Alternatives to the median absolute deviation // Journal of the American Statistical Association. 1993. Vol. 88. No. 424. Pp. 1273–1283.

8. **Shevlyakov G.L., Oja H.** Robust correlation: Theory and applications, USA, New Jersey: John Wiley & Sons, 2016. Vol. 3.

9. **Smirnov P.O., Shevlyakov G.L.** Fast highly efficient and robust one-step M -estimators of scale based on Q_n // Computational Statistics and Data Analysis. 2014. Vol. 78. Pp. 153–158.

10. **Tarr G.** Quantile based estimation of scale and dependence // Bulletin of the Australian Mathematical Society. 2015. Vol. 92. No. 01. Pp. 173–175.

Статья поступила в редакцию 29.06.16, принята к публикации 10.05.17.

СВЕДЕНИЯ ОБ АВТОРАХ

СМИРНОВ Павел Олегович — кандидат физико-математических наук, доцент кафедры прикладной математики Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Российская Федерация.

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
s.paul@mail.ru

ШИРОКОВ Иван Сергеевич — аспирант кафедры прикладной математики Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Российская Федерация.

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
summertouch@mail.ru

ШЕВЛЯКОВ Георгий Леонидович — доктор физико-математических наук, профессор кафедры прикладной математики Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Российская Федерация.

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
Georgy.Shevlyakov@phmf.spbstu.ru

REFERENCES

[1] **N. Balakrishnan, A.C. Cohen**, Order statistics & inference: estimation methods, Elsevier (2014).

[2] **H. Cramér**, Mathematical methods of statistics (PMS-9), Princeton University Press. (9) (2016).

[3] **M.G. Genton**, Asymptotic variance of M -estimators for dependent Gaussian random variables, Statistics and Probability Letters. 38 (3) (1998) 255–261.

[4] **F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel**, Robust statistics: the approach based on influence functions, John Wiley (1986).

[5] **P.J. Huber**, Robust statistics, USA, New Jersey, John Wiley (1981).

[6] **R.L. Launer, G.N. Wilkinson** (eds), Robustness in statistics, Academic Press (2014).

[7] **P.J. Rousseeuw, C. Croux**, Alternatives to the median absolute deviation, Journal of the American Statistical Association. 88 (424) (1993) 1273–1283.

[8] **G.L. Shevlyakov, H. Oja**, Robust correlation: Theory and applications, John Wiley & Sons, (3) (2016).

[9] **P.O. Smirnov, G.L. Shevlyakov**, Fast highly efficient and robust one-step M -estimators of scale based on Q_n , Computational Statistics and Data Analysis. 78 (2014) 153–158.

[10] **G. Tarr**, Quantile based estimation of scale and dependence, Bulletin of the Australian Mathematical Society. 92 (01) (2015) 173–175.

Received 29.06.2016, accepted 10.05.2017.

THE AUTHORS

SMIRNOV Pavel O.

Peter the Great St. Petersburg Polytechnic University

29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
s.paul@mail.ru



SHIROKOV Ivan S.

Peter the Great St. Petersburg Polytechnic University

29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation

summertouch@mail.ru

SHEVLYAKOV Georgiy L.

Peter the Great St. Petersburg Polytechnic University

29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation

Georgy.Shevlyakov@phmf.spbstu.ru