

Лисицин, Д. В. Использование принципа максимальной энтропии для конструирования робастных оценок при байесовском точечном засорении. Часть II / Д. В. Лисицин, К. В. Гаврилов // Прикладная математика и вопросы управления. – 2024. – № 2. – С. 18–33. DOI 10.15593/2499-9873/2024.2.02

Библиографическое описание согласно ГОСТ Р 7.0.100–2018

Лисицин, Д. В. Использование принципа максимальной энтропии для конструирования робастных оценок при байесовском точечном засорении. Часть II / Д. В. Лисицин, К. В. Гаврилов. – Текст : непосредственный. – DOI 10.15593/2499-9873/2024.2.02 // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2024. – № 2. – С. 18–33.



**пермский
политех** ПРИКЛАДНАЯ МАТЕМАТИКА
И ВОПРОСЫ УПРАВЛЕНИЯ
№ 2, 2024

<https://ered.pstu.ru/index.php/amcs>



Научная статья

DOI: 10.15593/2499-9873/2024.2.02

УДК 519.233.22



Использование принципа максимальной энтропии для конструирования робастных оценок при байесовском точечном засорении. Часть II

Д.В. Лисицин, К.В. Гаврилов

Новосибирский государственный технический университет, Новосибирск, Российская Федерация

О СТАТЬЕ

Получена: 27 декабря 2023

Одобрена: 19 января 2024

Принята к публикации:

25 января 2024

Финансирование

Исследование не имело спонсорской поддержки.

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Вклад авторов

равноценен.

Ключевые слова:

оценивание параметров, робастность, функция влияния, сниженные оценки, принцип максимальной энтропии, энтропия Шеннона, дивергенция Кульбака – Лейблера, энтропия Реньи, дивергенция Реньи, энтропия Цаллиса, дивергенция Цаллиса, альфа-дивергенция, энтропия Берга, расстояние Хеллингера, косинусно-степенное распределение.

АННОТАЦИЯ

Рассматривается развитие теории робастного оценивания параметров статистических моделей с привлечением аппарата теории информации. Анализируется подход А.М. Шурыгина, основанный на модели серии выборок со случайным точечным засорением (модели байесовского точечного засорения). В первой части нашей работы описан непараметрический способ выбора распределения засоряющей точки – посредством максимизации энтропии Шеннона или перекрестной энтропии в окрестности модельного распределения, ограниченной величиной дивергенции Кульбака – Лейблера. Такой способ нахождения плотности распределения засоряющей точки позволяет рассматривать получаемые оценки как робастные, причем обладающие свойством оптимальности. Полученные оценки мы называем обобщенными радикальными, поскольку их частным случаем являются радикальные оценки А.М. Шурыгина.

Во второй части работы получено другое оптимальное решение на основе формализма А. Реньи (или эквивалентного с точки зрения нашей задачи формализма К. Цаллиса), дающее новое семейство оценок, частными случаями которого также являются некоторые известные оценки. Для выбора одной оценки из семейства, определяемого разными ограничениями на дивергенцию, предложен оптимизационный подход. Основные теоретические результаты, полученные в работе, иллюстрируются на примере оценивания параметра сдвига косинусного распределения.

© Лисицин Даниил Валерьевич – доктор технических наук, профессор, профессор кафедры теоретической и прикладной информатики, e-mail: lisitsin@ami.nstu.ru, ORCID: 0009-0007-0723-6285.

Гаврилов Константин Викторович – кандидат технических наук, доцент кафедры автоматки, e-mail: k.gavrilov@corp.nstu.ru, ORCID: 0009-0007-0723-6285.



Эта статья доступна в соответствии с условиями лицензии Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

Perm Polytech Style: Lisitsin D.V., Gavrilov K.V. The use of maximum entropy principle to construct robust estimators under point Bayesian contamination. Part II. *Applied Mathematics and Control Sciences*. 2024, no. 2, pp. 18–33. DOI: 10.15593/2499-9873/2024.2.02

MDPI and ACS Style: Lisitsin, D.V.; Gavrilov, K.V. The use of maximum entropy principle to construct robust estimators under point Bayesian contamination. Part II. *Appl. Math. Control Sci.* **2024**, *2*, 18–33. <https://doi.org/10.15593/2499-9873/2024.2.02>

Chicago/Turabian Style: Lisitsin, Daniil V., and Konstantin V. Gavrilov. 2024. “The use of maximum entropy principle to construct robust estimators under point Bayesian contamination. Part II”. *Appl. Math. Control Sci.* no. 2: 18–33. <https://doi.org/10.15593/2499-9873/2024.2.02>



APPLIED MATHEMATICS
AND CONTROL SCIENCES

№ 2, 2024

<https://ered.pstu.ru/index.php/amcs>



Article

DOI: 10.15593/2499-9873/2024.2.02

UDC 519.233.22



The use of maximum entropy principle to construct robust estimators under point Bayesian contamination. Part II

D.V. Lisitsin, K.V. Gavrilov

Novosibirsk State Technical University, Novosibirsk, Russian Federation

ARTICLE INFO

Received: 27 December 2023

Approved: 19 January 2024

Accepted for publication:

25 January 2024

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

equivalent.

Keywords:

parameter estimation, robustness, influence function, redescending estimators, principle of maximal entropy, Shannon entropy, Kulback – Leibler divergence, Rényi entropy, Rényi divergence, Tsallis entropy, Tsallis divergence, alpha-divergence, Burg entropy, Hellinger distance, power-of-cosine distribution.

ABSTRACT

The work is devoted to the theory of robust parameter estimation of statistical models using the apparatus of information theory. The approach of A.M. Shurygin based on the model of a series of samples with random point contamination (point Bayesian contamination model) is considered. The first part of our work describes a non-parametric method of selecting the contamination point distribution – by maximizing Shannon entropy or cross entropy in the neighborhood of the model distribution limited by the value of Kulbak – Leibler divergence. This way of finding the distribution density of the contamination point allows us to consider the resulting estimators as robust, and, moreover, having the optimality property. We call the obtained estimators generalized radical, since their special case is the radical estimators of A.M. Shurygin.

In the second part of the work, another optimal solution is obtained on the basis of the formalism of A. Rényi (or the formalism of C. Tsallis equivalent in terms of our problem) that gives a new family of estimators, the special cases of which are also some well-known estimators. To select one estimate from a family defined by different divergence constraints, an optimization approach is proposed. The main theoretical results obtained in the paper are illustrated by the example of location estimating for the cosine distribution.

© Daniil V. Lisitsin – Doctor of Technical Sciences, Professor, Professor of the department of Theoretical and Applied Computer Science, e-mail: lisitsin@ami.nstu.ru, ORCID: 0009-0007-0723-6285.

Konstantin V. Gavrilov – CSc of Technical Sciences, Associate Professor of the department of Automatics, e-mail: k.gavrilov@corp.nstu.ru, ORCID: 0009-0007-0723-6285.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

Введение

В работе развивается подход А.М. Шурыгина к устойчивому оцениванию параметров статистических моделей, основанный на модели байесовского точечного засорения (БТЗ) – модели серии выборок со случайным точечным засорением и некоторыми дополнительными асимптотическими предположениями [1]. Оптимальная оценка в данном подходе для фиксированного распределения засоряющей точки определяется минимизацией асимптотического квадратичного отклонения оценки.

Для выбора распределения засоряющей точки в первой части статьи предложено использовать непараметрический способ – максимизацию энтропии Шеннона или перекрестной энтропии в окрестности модельного распределения, ограниченной величиной дивергенции Кульбака – Лейблера. Это позволяет рассматривать получаемые оценки как робастные, причем обладающие свойством оптимальности. Полученные оценки мы называем обобщенными радикальными (ООР¹), поскольку их частным случаем являются радикальные оценки А.М. Шурыгина. К задаче максимизации перекрестной энтропии сводится задача максимизации математического ожидания функции потерь оценок максимального правдоподобия (ОМП) в модели БТЗ. По этой причине ООР могут интерпретироваться как защищенные от намеренного искажения оценок максимального правдоподобия.

Во второй части работы получено другое оптимальное решение на основе формализма А. Реньи (или эквивалентного с точки зрения нашей задачи формализма К. Цаллиса), дающее новое семейство оценок, частными случаями которого также являются некоторые известные оценки. Эти результаты описаны в п. 1. Каждое из полученных в работе решений представляет собой семейство оценок, задаваемых разными ограничениями на дивергенцию. В п. 2 предложено использовать оптимизационный подход для обоснованного выбора одной оценки из семейства. Основные теоретические результаты, полученные в работе, иллюстрируются на примере оценивания параметра сдвига косинусного распределения: в п. 3 строятся и исследуются ООР, в п. 4 – новые оценки из п. 1.

В ссылках на формулы из первой части статьи перед номером формулы будет использоваться префикс «I.».

1. Использование формализмов Реньи и Цаллиса

Функционалы энтропии Шеннона (I.14) и дивергенции Кульбака – Лейблера (I.15) обобщаются в рамках формализма А. Реньи [2; 3] соответственно как

$$H_q(s) = \frac{1}{1-q} \ln \int_x s^q dx, \quad q > 0, \quad q \neq 1; \quad (1)$$

$$D_q(s, f) = \frac{1}{q-1} \ln \int_x s^q f^{1-q} dx, \quad q > 0, \quad q \neq 1. \quad (2)$$

Кроме того, $H_1(s) \stackrel{df}{=} \lim_{q \rightarrow 1} H_q(s) = H(s)$ и $D_1(s, f) \stackrel{df}{=} \lim_{q \rightarrow 1} D_q(s, f) = D(s, f)$.

Аналогично функционалы (I.14) и (I.15) обобщаются в рамках формализма К. Цаллиса [4; 5] соответственно как

$$\tilde{H}_q(s) = \frac{1}{q-1} \left(1 - \int_x s^q dx \right), \quad q > 0, \quad q \neq 1; \quad (3)$$

¹ ООР – оценка обобщенная радикальная.

$$\tilde{D}_q(s, f) = \frac{1}{q-1} \left(\int_x s^q f^{1-q} dx - 1 \right), \quad q > 0, \quad q \neq 1. \quad (4)$$

Величины $\tilde{H}_1(s)$ и $\tilde{D}_1(s, f)$ определяются аналогично $H_1(s)$ и $D_1(s, f)$.

Нетрудно видеть, что энтропии Реньи (1) и Цаллиса (3) порождаются соответствующими дивергенциями (2) и (4) согласно (I.13).

Задачу оптимизации энтропии Шеннона (I.14) при ограничении (I.16) на дивергенцию Кульбака – Лейблера, рассмотренную в п. 2 первой части статьи, можно считать частным случаем задач оптимизации функционалов (1) и (3) с ограничениями на величины (2) и (4) соответственно при $q = 1$. Решим теперь эти две задачи при $q \neq 1$. И первым делом приведем результат об эквивалентности их формулировок.

Лемма. Задача $H_q(s) \rightarrow \max_s$ при ограничении $D_q(s, f) \leq \Delta$ эквивалентна задаче $\tilde{H}_q(s) \rightarrow \max_s$ при ограничении $\tilde{D}_q(s, f) \leq \tilde{\Delta}$, где $\tilde{\Delta} = Y_q(\Delta)$, $Y_q(x) = [e^{(q-1)x} - 1]/(q-1)$.

Доказательство. Функция $Y_q(x)$ при $q \neq 1$ определена на всей числовой оси и непрерывно возрастает, причем $Y_q(0) = 0$. Следовательно, ограничения $D_q(s, f) \leq \Delta$ и $Y_q(D_q(s, f)) \leq Y_q(\Delta)$ эквивалентны. Эквивалентность ограничений в утверждении леммы имеет место, поскольку $\tilde{D}_q(s, f) = Y_q(D_q(s, f))$ и $\tilde{\Delta} = Y_q(\Delta)$.

Покажем теперь, что результат максимизации функционалов $\tilde{H}_q(s)$ и $H_q(s)$ совпадает. Действительно, в силу соотношения $\tilde{H}_q(s) = Y_{2-q}(H_q(s))$ их значения связаны непрерывно возрастающей функциональной зависимостью. Лемма доказана.

Поскольку функционал (3) является вогнутым, а функционал (4) – выпуклым, соответствующая задача оптимизации будет выпуклой, что удобно для ее решения, в отличие от задачи с функционалами (1), (2). Функционалы (3) и (4) также имеют более удобную аналитическую форму. Таким образом, будем максимизировать функцию Лагранжа

$$\tilde{H}_q(s) + a[\tilde{\Delta} - \tilde{D}_q(s, f)] + b \left(\int_x s dx - 1 \right),$$

где $a \geq 0$, b – множители Лагранжа. Лагранжиан данной задачи имеет вид

$$L = -s^q (1 + f^{1-q} a)/(q-1) + bs.$$

Запишем условие стационарности функции Лагранжа:

$$\partial L / \partial s = -s^{q-1} (1 + f^{1-q} a) q / (q-1) + b = 0,$$

откуда после преобразований получаем

$$s = \tilde{k} (1 + f^{1-q} a)^{1/(1-q)} = k (1/l^{1-q} + f^{1-q} \omega)^{1/(1-q)}, \quad (5)$$

где $k = l\tilde{k} = l [b(q-1)/q]^{1/(q-1)}$ – нормировочная константа (предполагается, что плотность (5) существует, следовательно, $k > 0$), l – положительный параметр, некоторый смысл которого будет выяснен ниже, $\omega = al^{q-1}$ – параметр формы плотности (5). В качестве «значения по умолчанию» удобно (см. п. 4) взять l равным параметру масштаба f .

Подставив (5) в (I.6), находим оптимальные оценочные функции:

$$\psi(x, \theta) = \frac{c(\theta, q, \omega, l) f(x, \theta)}{[1/l^{1-q} + f^{1-q}(x, \theta)\omega]^{1/(1-q)}} \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) + \beta(\theta, q, \omega, l) \right], \quad (6)$$

где функции c и β имеют тот же смысл, что и в (I.6), $q \neq 1$. Нормировочная константа k не фигурирует в (6), поскольку поглощается множителем c . Параметр ω определяет компромисс между оценкой максимальной устойчивости (ОМУ) [1] и ОМП [6]: значение $\omega = 0$ соответствует ОМУ, значение $\omega = \infty$ – ОМП.

Оценки с оценочными функциями вида (6) будем называть *компромиссными степенными оценками* (ОКС²).

Условие дополняющей нежесткости для множителя Лагранжа a , по аналогии с задачей, рассмотренной в п. 2 первой части статьи, позволяет установить связь параметров a и ω с размером окрестности Δ , который определяется через значение функционала (4) для наихудшей плотности (5):

$$\tilde{D}_q(s, f) = \frac{1}{1-q} \left(1 - \int_X \tilde{k}^q (1 + f^{1-q} a)^{q/(1-q)} f^{1-q} dx \right) = \frac{1}{1-q} \left(1 - \int_X k^q (1/l^{1-q} + f^{1-q} \omega)^{q/(1-q)} f^{1-q} dx \right).$$

Как и в случае ООР, вместо параметра Δ удобно задавать параметр ω .

Семейство (6) по построению определено для $q > 0$. Между тем данное семейство оценочных функций может быть доопределено для $q = 0$ как предел функции (6) при $q \rightarrow 0$. В этом случае решение (6) совпадает с выражением (I.4) для условно оптимального семейства, где $\lambda = a = \omega l^{1-q}$. Для обоснования (I.4) согласно [1; 7] не используется модель БТЗ, тогда как в нашем подходе невозможность нормирования s при неограниченном X делает обоснование оптимальности данной оценочной функции до некоторой степени формальным.

Кроме того, семейство оценочных функций (6) может быть формально продолжено для значений $q < 0$ и для значений ω , таких что

$$-\min_{x \in X} [f(x, \theta) l]^{q-1} < \omega < 0. \quad (7)$$

Для некоторых моделей оценочные функции из семейства (6) могут быть корректно доопределены не только значениями ω , удовлетворяющими (7), но и его значениями, совпадающими с нижней гранью неравенства (7). В частности, для косинусно-степенного распределения (I.12) при $\kappa = 3$ и $q \leq -1$ имеем нижнюю грань в (7), равную -1 ; оценочные функции (6) при $\omega = -1$ являются ограниченными, хотя и разрывными.

Что касается продолжения семейства (6) в область $q \leq 0$, то эти решения тоже могут быть получены в результате решения оптимизационной задачи, но с применением обобщенных на $q \in R$ функционалов энтропии и дивергенции.

Рассмотрим функционал *альфа-дивергенции* (непрерывной версии степенной дивергенции) [5], обобщающий дивергенцию Цаллиса (4),

$$\tilde{D}_q(s, f) = \frac{1}{q(q-1)} \left(\int_X s^q f^{1-q} dx - 1 \right), \quad q \neq 0, \quad q \neq 1. \quad (8)$$

² ОКС – оценка компромиссная степенная.

При $q = 2$ функционал (8) с точностью до несущественного множителя соответствует дивергенции $\chi^2(s, f)$, известной как расстояние хи-квадрат Пирсона, при $q = -1$ – дивергенции $\chi^2(f, s)$ (расстояние хи-квадрат Неймана), при $q = 1/2$ – квадрату расстояния Хеллингера [5]. Аналогично обобщаются дивергенция и энтропия Реньи и энтропия Цаллиса [5], причем энтропии могут быть получены как порождаемые соответствующими дивергенциями.

В частности, для *альфа-энтропии* в соответствии с (I.13) на основе выражения (8) возможно следующее представление:

$$\check{H}_q(s) = \varphi + \frac{1}{q(q-1)} \left(1 - \int_X s^q dx \right), \quad q \neq 0, \quad q \neq 1. \quad (9)$$

Чтобы перейти к пределу при $q \rightarrow 1$, зафиксируем здесь значение $\varphi = 0$, тогда в пределе получим энтропию Шеннона (I.14). Таким образом, при $q = 1$ функционалы (8) и (9) доопределяются соответствующими предельными функционалами подобно (1)–(4). Для перехода к пределу при $q \rightarrow 0$ возьмем

$$\varphi = \frac{1}{q(q-1)} \left(\int_X dx - 1 \right),$$

однако обратим внимание, что такое значение φ существует только для ограниченного X . Таким образом, при $q = 0$ функционал (8) непрерывно доопределяется величиной $D(f, s)$, а функционал (9) – *энтропией Берга* $\int_X \ln s \, dx$, порождаемой указанной предельной дивергенцией.

Повторяя рассуждения, приведенные в этом параграфе, для функционалов (8) и (9), можно показать, что формальное решение задачи $\check{H}_q(s) \rightarrow \max_s$ при ограничении $\check{D}_q(s, f) \leq \check{\Delta}$, где $\check{\Delta}$ – некоторое неотрицательное число, с точностью до выражения для константы k совпадает с (5), а соответствующая данной s оценочная функция – с решением (6). И это решение справедливо для всех $q \in R$. Найденная стационарная точка функции Лагранжа – точка максимума, поскольку задача оптимизации выпуклая.

Рассмотрим поведение оценочной функции (6) при значениях x , для которых $f(x, \theta) \rightarrow 0$. Для $q < 1$ имеем

$$\begin{aligned} \psi &= \frac{c(\dot{f} + \beta f)}{(1/l^{1-q} + f^{1-q}\omega)^{1/(1-q)}} \sim \frac{c(\dot{f} + \beta f)}{1/l} = lc \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) f = \\ &= \psi_{\text{ОМУ}} + lc(\beta - \beta_{\text{ОМУ}})f, \end{aligned}$$

где $\psi_{\text{ОМУ}}$ и $\beta_{\text{ОМУ}}$ – соответственно оценочная функция и функция β для ОМУ, величины c и l – те же, что в (6). В этом случае, если $\psi_{\text{ОМУ}}$ удовлетворяет условию (I.8), то функция (6) тоже будет ему удовлетворять. Для $q > 1$ и $\omega > 0$ имеем

$$\begin{aligned} \psi &= \frac{c(\dot{f} + \beta f)}{(1/l^{1-q} + f^{1-q}\omega)^{1/(1-q)}} \sim \frac{c(\dot{f} + \beta f)}{f\omega^{1/(1-q)}} = c\omega^{1/(q-1)} \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) = \\ &= \psi_{\text{ОМП}} + c\beta\omega^{1/(q-1)}. \end{aligned}$$

В этом случае функция (6) будет удовлетворять условию (I.8), если ему удовлетворяет $\psi_{\text{ОМП}}$ и $\beta = 0$.

Отметим, что для многих широко используемых на практике распределений ОМУ оказываются удовлетворяющими условию (I.8), а ОМП – нет. По этой причине ОКС будут представлять интерес главным образом при $q < 1$.

Рассмотрим поведение плотности (5) и оценочной функции (6) при $q \rightarrow \pm\infty$ и фиксированном $\omega > 0$. При $q \rightarrow -\infty$ имеем

$$s = k \begin{cases} 1/l, & f \leq 1/l \\ f, & f \geq 1/l \end{cases}$$

$$\psi = c \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) \begin{cases} f l, & f \leq 1/l \\ 1, & f \geq 1/l \end{cases}.$$

При $q \rightarrow \infty$ имеем

$$s = k \begin{cases} f, & f \leq 1/l \\ 1/l, & f \geq 1/l \end{cases}$$

$$\psi = c \left(\frac{\partial}{\partial \theta} \ln f + \beta \right) \begin{cases} 1, & f \leq 1/l \\ f l, & f \geq 1/l \end{cases}.$$

В случае, когда $\psi_{\text{ОМУ}}$ обладает свойством (I.8), им будет обладать и функция (6) при $q \rightarrow -\infty$. При $q \rightarrow \infty$ функция (6) будет обладать свойством (I.8), если этим свойством обладает $\psi_{\text{ОМП}}$ и $\beta = 0$.

Единственный параметр l полученных предельных оценочных функций может выбираться, например, исходя из вероятности, соответствующей значениям плотности, бóльшим величины $1/l$, иначе говоря, значениям случайной величины, лежащим в области около моды (или мод).

Предельные оценочные функции могут быть интуитивно привлекательными из-за простоты их структур и очевидности геометрического смысла параметра l , особенно в случае $\beta = 0$. Однако стыковка в граничных точках, вообще говоря, является негладкой, что может быть нежелательно. С данной точки зрения на оценочную функцию ОКС при большом конечном значении $|q|$ можно смотреть как на гладкую, когда это возможно, аппроксимацию предельной оценочной функции. В последнем случае, исходя из поведения функции (5) при увеличении $|q|$, наиболее предпочтительным выбором для параметра ω представляется значение 1.

Когда для всех значений x справедливо $f \leq 1/l$, оценочная функция совпадает с $\psi_{\text{ОМУ}}$ для случая $q \rightarrow -\infty$ и с $\psi_{\text{ОМП}}$ для случая $q \rightarrow \infty$. Наоборот, когда для всех значений x справедливо $f \geq 1/l$, оценочная функция совпадает с $\psi_{\text{ОМП}}$ для случая $q \rightarrow -\infty$ и с $\psi_{\text{ОМУ}}$ – для случая $q \rightarrow \infty$.

Поскольку в области с бóльшими значениями плотности, например, в районе моды, естественнее иметь оценочную функцию, близкую к $\psi_{\text{ОМП}}$, а в области с малыми значениями плотности, особенно при $f \rightarrow 0$, – близкую к $\psi_{\text{ОМУ}}$, случай $q \rightarrow -\infty$ представляется более предпочтительным. Таким образом, для получения робастных оценок на практике чаще будут представлять интерес малые значения параметра q .

Заметим также, что такая структура оценочной функции схожа со структурой оценочной функции дисперсионно-минимаксной оценки в теории Хьюбера [8].

Для решения рассмотренных вариационных задач необходимо, чтобы функция s (5) являлась плотностью, т.е. удовлетворяла условию нормировки. Однако при $q < 1$ и неограниченном интервале X это не имеет места. В таком случае следует применить обобщенную модель БТЗ: рассмотреть последовательность усеченных интервалов $X_n \rightarrow X$ (где n – натуральные числа) подобно тому, как описано в п. 2. первой части статьи. Тогда формально найденное решение (6) в тех случаях, когда плотность (5) не существует, может интерпретироваться как предел решений, соответствующих усеченным интервалам.

2. Максимальные робастные оценки

Рассмотренные ранее ООР и ОКС являются семействами, имеющими от одного до трех параметров. Такое положение является обычным в теории робастных оценок [9], однако также представляет интерес аргументированный выбор какого-то одного члена семейства [1]. И такой выбор в соответствии с идеями теории робастности целесообразно сделать, исходя из наихудших условий. Подобная логика используется при построении стойких оценок. Применим ее и в нашем случае, используя (I.6). Для ООР решим задачу

$$v_* = \arg \max_v \min_{\psi} U(\psi, s_v) = \arg \max_v U(\psi_v, s_v),$$

где ψ_v определена в (I.11), функционал U – в (I.5). Для ОКС решим задачу

$$(q_*, \omega_*) = \arg \max_{q, \omega} \min_{\psi} U(\psi, s_{q, \omega}) = \arg \max_{q, \omega} U(\psi_{q, \omega}, s_{q, \omega}),$$

где $s_{q, \omega}$ – плотность (5), $\psi_{q, \omega}$ определена в (6); или задачу

$$\omega_* = \arg \max_{\omega} \min_{\psi} U(\psi, s_{q, \omega}) = \arg \max_{\omega} U(\psi_{q, \omega}, s_{q, \omega}),$$

если параметр q фиксирован. Оценки с таким образом выбранными параметрами будем называть *максимальными*, так как s_{v_*} и s_{q_*, ω_*} (или s_{q, ω_*} , если q фиксирован) – частные случаи плотности (I.9).

Поскольку в выражении (5) параметр l не обязательно является фиксированным, для него тоже можно выбирать наихудшее значение. Например, мы могли бы оптимизировать параметры q и l , зафиксировав значение ω равным единице.

Важно подчеркнуть, что описанный способ формирования плотности s , в отличие от случая стойких оценок, остается непараметрическим.

В [1] приведены примеры стойких оценок для параметров ряда одномерных распределений: параметра масштаба экспоненциального распределения, параметра сдвига распределения Лапласа, параметров нормального распределения. Можно заметить, что все эти оценки оказываются также максимальными ООР. Причины этого следующие. Во-первых, параметрическое семейство плотностей S во всех примерах совпадает с семейством модельных распределений, имеющих единственной оптимизируемой переменной параметр масштаба. Во-вторых, для названных распределений возведение плотности в некоторую положительную степень (с последующей нормировкой) равносильно лишь изменению параметра масштаба. И хотя множество допустимых функций s для максимальных ООР мо-

жет быть шире S за счет использования обобщенной модели БГЗ, оптимальные решения в обоих случаях оказываются совпадающими.

Примеры стойких оценок, отличных от максиминных ООР, можно найти в [10]. Один из этих примеров примечателен тем, что стойкая оценка оказывается устойчивой к менее широкому множеству искажений, чем ООР [11] (в том числе максиминная ООР, близкая к радикальной оценке), и, более того, может быть менее устойчивой, чем ОМП.

3. Пример: обобщенные радикальные оценки

Рассмотрим задачу оценивания параметра сдвига для косинусного распределения – частного случая (I.12) при $k = 3$ – с плотностью [12; 13]

$$f(x, \theta) = \frac{1}{l} \cos^2 \frac{\pi(x - \theta)}{2l} = \frac{1}{2l} \left[1 + \cos \frac{\pi(x - \theta)}{l} \right], \quad |x - \theta| \leq l, \quad (10)$$

где θ – параметр сдвига, l – параметр масштаба.

Обобщенной радикальной оценке соответствует плотность

$$s_\nu(x, \theta) = k f^{1-\nu}(x, \theta) = k \left[\frac{1}{l} \cos^2 \frac{\pi(x - \theta)}{2l} \right]^{1-\nu}, \quad (11)$$

где $\nu < \nu_0 = 3/2$ и

$$k = \frac{\sqrt{\pi}}{2} \frac{\Gamma(2 - \nu)}{\Gamma(3/2 - \nu)} l^{-\nu}.$$

Энтропия Шеннона плотности (10) определяется выражением

$$H(s) = (1 - \nu)[\ln l + \Psi(2 - \nu) - \Psi(3/2 - \nu)] - \ln k,$$

где $\Psi(\cdot)$ – дигамма-функция. В частности, при $\nu = 0$ получаем

$$H(f) = \ln(4l) - 1.$$

Подставляя найденную $H(s)$ в (I.17), получаем выражение

$$\begin{aligned} D(s, f) &= \ln k + \nu[\ln l + \Psi(2 - \nu) - \Psi(3/2 - \nu)] = \\ &= \ln \frac{\sqrt{\pi}}{2} + \ln \Gamma(2 - \nu) - \ln \Gamma(3/2 - \nu) + \nu[\Psi(2 - \nu) - \Psi(3/2 - \nu)], \end{aligned}$$

которое в случае $0 \leq \nu < 1$ определяет размер окрестности Δ , соответствующий заданному параметру ν , а при $\nu = 1$ – ее минимальный размер. Функционал перекрестной энтропии для плотности (11) имеет вид

$$H(s, f) = \ln l + \Psi(2 - \nu) - \Psi(3/2 - \nu).$$

На рис. 1 приведены графики величин $H(s)$, $H(s, f)$ и $D(s, f)$ в зависимости от ν при $l = 1$. Максимум энтропии Шеннона $\max_s H(s) = \ln(2l)$ достигается при равномерном распределении, которому соответствует $\nu = 1$.

Оценочная функция семейства обобщенных радикальных оценок, в данном случае совпадающая за счет выбора $c(\theta)$ с функцией влияния, задается формулой

$$\Psi_{\text{OOP}}(x, \theta, \nu) = \frac{l}{\sqrt{\pi}} \frac{\Gamma(\nu+2)}{\Gamma(\nu+1/2)} \sin \frac{\pi(x-\theta)}{2l} \cos^{2\nu-1} \frac{\pi(x-\theta)}{2l}, \quad |x-\theta| < l. \quad (12)$$

Условия регулярности из п. 1 первой части статьи для семейства (12) выполняются при $\nu > 0$. Для ООР при $\nu = 0$ (аналог ОМП в обобщенном подходе) характеристики условимся понимать в смысле пределов соответствующих выражений при $\nu \rightarrow 0$. В частности, $\lim_{\nu \rightarrow 0} V(\Psi_{\text{OOP}}, f) = l^2/\pi^2$, что соответствует нижней границе в неравенстве Рао – Крамера [6; 14]. Отношение данной величины к величине V некоторой оценки в дальнейшем и будем использовать для вычисления эффективности последней.

Отметим также, что ОМУ ($\nu = 1$) связана с известной оценкой Эндрюса [9; 13]. Графики оценочных функций (12) для $\nu = 0$ (аналог ОМП), $\nu = 1/2$ (радикальная оценка), $\nu = 1$ (ОМУ) и $\nu = 2$ при $l = 1$ показаны на рис. 2.

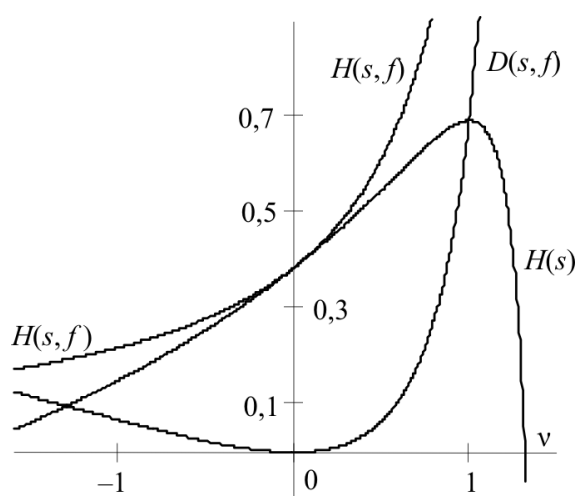


Рис. 1. Дифференциальная, перекрестная энтропии и расстояние Кульбака – Лейблера при $l = 1$

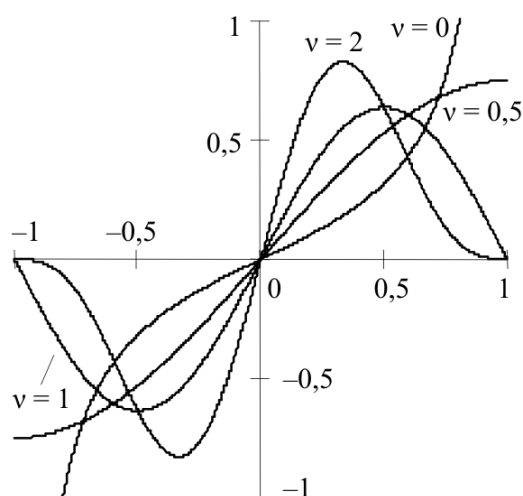


Рис. 2. Некоторые функции влияния ООР при $\theta = 0, l = 1$

Оценки, соответствующие (12), имеют следующие эффективность eff и устойчивость stb (см. п. 1 первой части статьи):

$$\text{eff } \Psi = \frac{4^\nu (2\nu+1)}{\pi(\nu+1)} \frac{\Gamma^3(\nu+1/2)}{\Gamma(\nu+2)\Gamma(2\nu+1/2)}, \quad \nu \geq 0;$$

$$\text{stb } \Psi = \frac{4^{\nu+1}}{\pi(\nu+1)} \frac{\Gamma^3(\nu+1/2)}{\Gamma(\nu+2)\Gamma(2\nu-1/2)} = \left(4 - \frac{6}{2\nu+1}\right) \text{eff } \Psi, \quad \nu > \frac{1}{4}.$$

Графики зависимостей эффективности и устойчивости от параметра ν представлены на рис. 3.

При $\nu \leq 1/4$ оценки являются неустойчивыми по Шурыгину ($\text{stb } \Psi = 0$). Кроме того, при $\nu < 1/2$ функция влияния не ограничена, что означает отсутствие у оценки качественной робастности. Значению $\nu = 1/2$ соответствует окрестность модельной плотности с довольно малым значением $\Delta \approx 0,065288$ (примерно на порядок меньшим, чем для ОМУ). Значениям $\nu < 1/2$ соответствуют еще меньшие окрестности. Это отражает общую закономерность снижения устойчивости и даже потери качественной робастности оценки при сокращении рассматриваемой окрестности модельного распределения.

Сконструируем максиминную ООР параметра θ . Для этого оценочную функцию (12) и плотность засоряющей точки (11) подставим в (I.5). Получаем

$$U(v) = \frac{v(1-v)(1+v)}{\pi(1-2v)} l^2 \operatorname{ctg}(\pi v). \quad (13)$$

При $v = 0$, $v = 1/2$ и $v = 1$ неопределенные значения функции $U(v)$ раскрываются соответственно как l^2/π^2 , $3l^2/16$ и $2l^2/\pi^2$. График функции (13) при $l = 1$ приведен на рис. 4. Ее максимум достигается в точке $v_* \approx 0,833603$. Соответствующая максиминная ООР имеет характеристики $\operatorname{eff} \psi_* \approx 56,07\%$, $\operatorname{stb} \psi_* \approx 98,14\%$. Как видно по значению устойчивости, найденная оценка оказалась весьма близкой к ОМУ, показатели которой составляют 50 и 100% соответственно. Максиминная ООР является также стойкой оценкой при плотности (11) с единственным оптимизируемым параметром v .

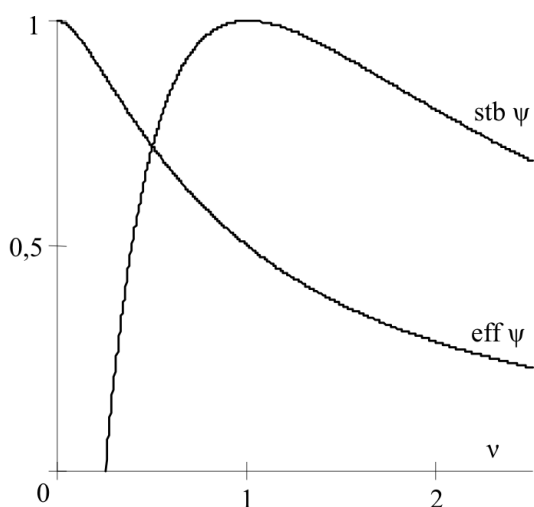


Рис. 3. Эффективность и устойчивость ООР отклонение ООР при $l = 1$

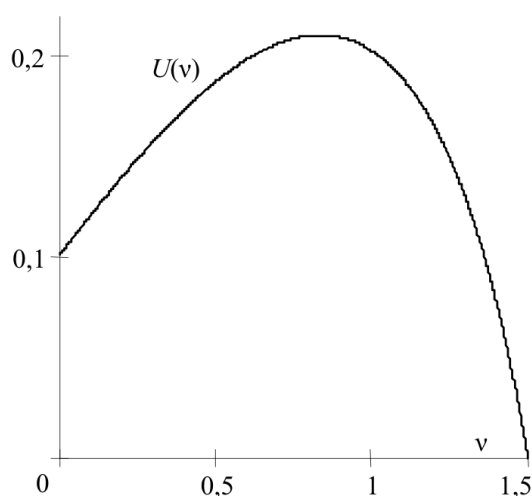


Рис. 4. Асимптотическое квадратичное отклонение ООР при $l = 1$

4. Пример: компромиссные степенные оценки

Рассмотрим ОКС с параметром $q = 1/2$ для распределения с плотностью (10). Подставим выражение (10) и значение $q = 1/2$ в (6), где $\omega = a/\sqrt{l}$, а в качестве параметра l плотности (5) используем l из (10). В результате получаем следующее семейство эквивалентных оценочных функций:

$$\psi(x, \theta, \omega) = l c(\omega) \sin \frac{\pi(x-\theta)}{l} \cos \frac{\pi(x-\theta)}{2l} \left[1 + \omega \cos \frac{\pi(x-\theta)}{2l} \right]^{-2}, \quad (14)$$

где $\omega > -1$ согласно условию (7). Функции влияния соответствует выбор

$$c(\omega) = \frac{\omega^4}{\pi \omega^2 + 12\omega - 6\pi + (12 - 8\omega^2)g(\omega)},$$

$$\text{где } g(\omega) = \begin{cases} \operatorname{arch} \omega/\sqrt{\omega^2 - 1} & \text{при } \omega > 1, \\ 1 & \text{при } \omega = 1, \\ \operatorname{arccos} \omega/\sqrt{1 - \omega^2} & \text{при } -1 < \omega < 1. \end{cases}$$

Значение $q = 1/2$ определяет середину промежутка $[0; 1]$, и оценочная функция (14) лежит в этом смысле посередине между оценочными функциями ООР и условно оптимальных оценок. Кроме того, она примечательна тем, что окрестность плотности f определяется ограничением на расстояние Хеллингера.

Оценочной функции (14) соответствует плотность распределения засоряющей точки

$$s(x, \theta, \omega) = \frac{1}{(2 + \omega^2 + 8\omega/\pi)l} \left[1 + \omega \cos \frac{\pi(x - \theta)}{2l} \right]^2. \quad (15)$$

Параметр l здесь имеет смысл масштаба плотности. Подставив (14) и (15) в (1.5), находим асимптотическое квадратичное отклонение оценки $U(\omega)$; график этой функции при $l = 1$ изображен на рис. 5 (аналитическая формула не приводится ввиду ее громоздкости). Максимум функции $U(\omega)$ достигается при $\omega_* \approx 0,224726$. Соответствующая этому значению максиминная ОКС имеет характеристики $\text{eff } \psi_* \approx 53,91\%$ и $\text{stb } \psi_* \approx 99,37\%$. Она оказалась весьма близкой к ОМУ (отличие в соответствующих характеристиках составляет 3,91 и $-0,63\%$) – чуть более эффективная и чуть менее устойчивая. Почти столь же похожей максиминная ОКС оказалась на максиминную ООР (отличие составляет $-2,16$ и $1,23\%$ соответственно) – чуть более устойчивая и чуть менее эффективная.

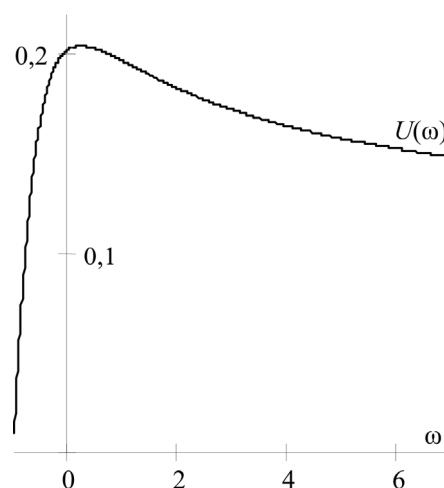


Рис. 5. Асимптотическое квадратичное отклонение ОКС при $l = 1, q = 1/2$

Графики функций влияния (14) для $\omega = -0,75$, $\omega = \omega_*$ и $\omega = 2$ при $l = 1$ показаны на рис. 6. На рис. 7 изображены графики характеристик эффективности и устойчивости оценочных функций (14) в зависимости от параметра ω .

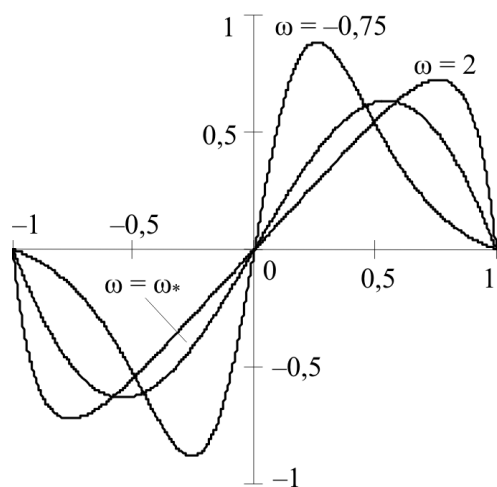


Рис. 6. Некоторые функции влияния ОКС при $\theta = 0, l = 1, q = 1/2$

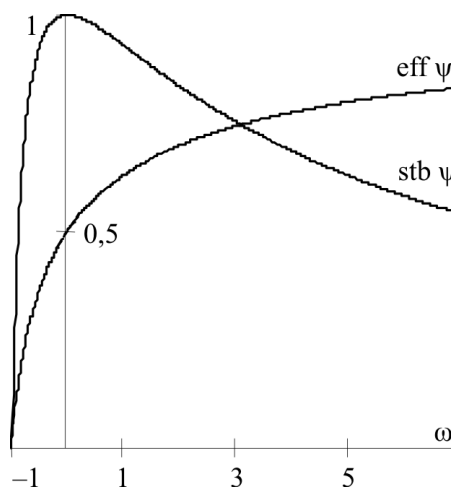


Рис. 7. Эффективность и устойчивость ОКС при $q = 1/2$

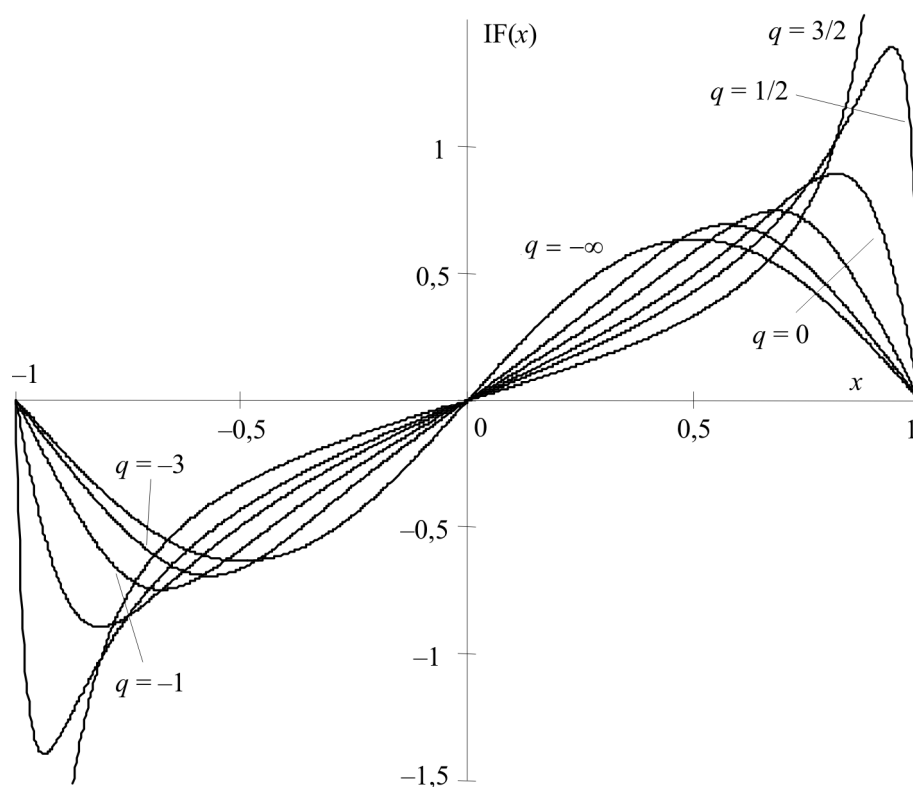


Рис. 8. Функции влияния ОКС для некоторых q при $\theta = 0$, $l = 1$, $\omega = 10$

За счет использования членов семейства оценочных функций (14) с разными значениями q , включая рассмотренный случай $q = 1/2$, можно произвести более тонкую настройку свойств ОКС.

Параметр q оказывает заметное воздействие на степень изменчивости функции влияния, обусловленной вариацией параметра ω : является тенденцией уменьшение изменчивости при увеличении величины $|q - 1|$, а в пределе при $q \rightarrow \pm\infty$ (в соответствии с результатами п. 1) для случая $\omega > 0$ этой изменчивости нет. Можно предположить, что и для многих других распределений параметр q будет обладать таким же свойством. Отметим также, что в пределе оценочная функция не будет гладкой в точке, где $f = 1/l$.

Поскольку для плотности (10) справедливо $f \leq 1/l$ при всех $x \in X$, ОКС параметра θ при фиксированном $\omega > 0$ будет стремиться к робастной ОМУ при $q \rightarrow -\infty$ и неробастной ОМП – при $q \rightarrow \infty$.

На рис. 8 показаны графики функций влияния ОКС в модели (10) при $\theta = 0$, $l = 1$ и $\omega = 10$ для различных значений параметра q . Значение $q = 3/2$, поскольку оно больше 1, определяет неограниченную оценочную функцию (близкую к оценочной функции ОМП), и соответствующая оценка не является робастной. Значению $q = -\infty$ соответствует ОМУ. Видно, как по мере уменьшения q графики функций влияния приближаются к функции влияния ОМУ.

Заключение

Предложенный в работе способ выбора плотности засоряющей точки для модели БТЗ позволяет рассматривать решения, получаемые в рамках соответствующего подхода А.М. Шурыгина, как решения теории робастности. Оптимальная плотность определяется

непараметрически в результате максимизации энтропии при ограничении на величину дивергенции между искомой и модельной плотностями, причем функционалы энтропии и дивергенции являются связанными между собой.

В рамках предложенного метода были использованы энтропия Шеннона в паре с дивергенцией Кульбака – Лейблера, а также формализмы А. Реньи и К. Цаллиса. Это позволило вывести семейство ООР и двухпараметрическое семейство ОКС, за исключением некоторых поддиапазонов параметров. Частным случаем ОКС при $q = 0$ является условно оптимальное семейство оценок, а при $q = 1$ – ООР. Исследованы свойства данных оценок. Так, установлено, что ОКС имеют наибольшую практическую значимость при $q \leq 1$.

Данный метод не позволяет вывести ОКС с параметром ω , удовлетворяющим (7), поскольку этот параметр, как и параметр a , по смыслу является неотрицательным. Для ООР по той же причине не может быть выведена часть семейства с параметром $\nu > 1$. Однако вывести ООР при любых $\nu \geq 0$ удалось благодаря использованию в качестве критерия оптимальности перекрестной энтропии. Данный подход заслуживает более подробного описания, поскольку похожий прием может быть применен для вывода всего семейства ОКС, а также его интерпретации как защищенного от искажения ООР и близкой к ней оценки минимума дивергенции степени плотности (бета-дивергенции) [15], подобно тому, как семейство ООР оказывается защищенным от искажения ОМП. Изложение этих результатов должно стать темой отдельной работы.

Интерес, на наш взгляд, представляют также оценки, защищенные от искажения L_p -оценок [16]. Среди оценок с таким свойством оказываются названные во введении робастные оценки параметра сдвига (кроме оценки Эндрюса), а также оценки Фэйра, Рамсея и несколько менее известных. Этим результатам также планируется посвятить отдельную работу.

Объяснение приведенных в [1] примеров стойких оценок параметров одномерных распределений как максиминных ООР позволяет считать эти примеры решениями теории робастности.

Рассмотренная нами теория касалась только случая одномерной непрерывной случайной величины при одномерном оцениваемом параметре. Однако это сделано лишь с целью упрощения изложения. Результаты работы можно распространить на случай вектора случайных величин (непрерывных, дискретных или разнотипных) при векторном оцениваемом параметре (в рамках подхода к оцениванию параметров из [17]).

Список литературы

1. Шурыгин, А.М. Математические методы прогнозирования: учебное пособие для вузов / А.М. Шурыгин. – М.: Горячая линия – Телеком, 2009. – 180 с.
2. Rényi, A. On measure of entropy and information / A. Rényi // Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. – 1961. – Vol. 1. – P. 547–561.
3. Principe, J.C. Information theoretic learning: Rényi's entropy and kernel perspectives / J.C. Principe. – Springer: Information Science and Statistics. – 2010. – 515 p. DOI: 10.1007/978-1-4419-1570-2
4. Tsallis, C. Introduction to nonextensive statistical mechanics: approaching a complex world / C. Tsallis. – N.Y.: Springer, 2009. – 382 p. DOI: 10.1007/978-0-387-85359-8
5. Cichocki, A. Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarities / A. Cichocki, S.-I. Amari // Entropy. – 2010. – Vol. 12, no. 6. – P. 1532–1568. DOI: 10.3390/e12061532

6. Боровков, А.А. Математическая статистика / А.А. Боровков. – СПб.: Издательство «Лань», 2021. – 704 с.
7. Лисицин, Д.В. О свойствах условно оптимальных оценок / Д.В. Лисицин, К.В. Гаврилов // Научный вестник НГТУ. – 2015. – № 1(58). – С. 76–93. DOI: 10.17212/1814-1196-2015-1-76-93
8. Хьюбер, П. Робастность в статистике / П. Хьюбер. – М.: Мир, 1984. – 303 с.
9. Робастность в статистике: подход на основе функций влияния / Ф. Хампель, Э. Ронchetti, П. Рауссеу, В. Штаэль. – М.: Мир, 1989. – 512 с.
10. Лисицин, Д.В. Исследование стойких оценок параметров распределения минимального значения / Д.В. Лисицин, Д.А. Грюнер // Доклады АН ВШ РФ. – 2010. – № 1(14). – С. 6–17.
11. Лисицин, Д.В. Исследование устойчивых оценок параметров распределения минимальных значений / Д.В. Лисицин, Д.А. Грюнер // Научный вестник НГТУ. – 2010. – № 2(39). – С. 21–30.
12. Cartwright, D.E. The use of directional spectra in studying the output of a wave recorder on a moving ship / D.E. Cartwright // Ocean Wave Spectra: Proc. of Conf. – New Jersey: Prentice Hall, 1963. – P. 203–218.
13. Лисицин, Д.В. Оценивание параметров распределения ограниченной случайной величины, робастное к нарушению границ / Д.В. Лисицин, К.В. Гаврилов // Научный вестник НГТУ. – 2016. – № 2(63). – С. 70–89. DOI: 10.17212/1814-1196-2016-2-70-89
14. Цыпкин, Я.З. Основы информационной теории идентификации / Я.З. Цыпкин. – М.: Наука, 1984. – 320 с.
15. Robust and efficient estimation by minimizing a density power divergence / A. Basu, I.R. Harris, N.L. Hjort, M.C. Jones // Biometrika. – 1998. – Vol. 85. – P. 549–559. DOI:10.1093/BIOMET/85.3.549
16. Демиденко, Е.З. Линейная и нелинейная регрессии / Е.З. Демиденко. – М.: Финансы и статистика, 1981. – 302 с.
17. Лисицин, Д.В. Устойчивое оценивание параметров модели по многомерным неоднородным неполным данным / Д.В. Лисицин // Научный вестник НГТУ. – 2013. – № 1(50). – С. 17–30.

References

1. Shurygin A. M. Matematicheskie metody prognozirovaniia [Mathematical methods of prediction]. Moscow, Goriachaia Liniia – Telekom, 2009, 180 p.
2. Rényi A. On measure of entropy and information. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1961, vol. 1, pp. 547-561.
3. Principe J. C. Information theoretic learning: Renyi's entropy and kernel perspectives. Springer, Information Science and Statistics, 2010, 515 p. DOI: 10.1007/978-1-4419-1570-2
4. Tsallis C. Introduction to nonextensive statistical mechanics: approaching a complex world. New York, Springer, 2009, 382 p. DOI: 10.1007/978-0-387-85359-85.
5. Cichocki A., Amari S.-I. Families of alpha- beta- and gamma-divergences: flexible and robust measures of similarities. Entropy, 2010, vol. 12, no. 6, pp. 1532-1568. DOI: 10.3390/e12061532
6. Borovkov A. A. Matematicheskaia statistika [Mathematical statistics]. Saint Petersburg, Izdatel'stvo «Lan'», 2021, 704 p.

7. Lisitsin D. V., Gavrilov K. V. O svoistvakh uslovno optimal'nykh otsenok [On properties of conditionally optimal estimates]. Science Bulletin of the NSTU, 2015, vol. 58, no. 1, pp. 76-93. DOI: 10.17212/1814-1196-2015-1-76-93
8. Huber P. Robastnost' v statistike [Robust statistics]. Moscow, Mir, 1984, 303 p.
9. Hampel F., Ronchetti E., Rousseeuw P., Stahel W. Robastnost' v statistike: podkhod na osnove funktsii vliianiia [Robust statistics: the approach based on influence functions]. Moscow, Mir, 1989, 512 p.
10. Lisitsin D. V., Gryuner D. A. Issledovanie stoikikh otsenok parametrov raspredeleniia minimal'nogo znacheniiia [The research of sturdy estimators of parameters of the least extreme value distribution]. Proceedings of the RHEAS, 2010, vol. 14, no. 1, pp. 6-17.
11. Lisitsin D. V., Gryuner D. A. Issledovanie ustoichivyykh otsenok parametrov raspredeleniia minimal'nykh znachenii [The research of robust estimators of parameters of the least extreme value distribution]. Science Bulletin of the NSTU, 2010, vol. 39, no. 2, pp. 21-30.
12. Cartwright D. E. The use of directional spectra in studying the output of a wave recorder on a moving ship. Ocean Wave Spectra, Proc. of Conf., New Jersey, Prentice Hall, 1963, pp. 203-218.
13. Lisitsin D. V., Gavrilov K. V. Otsenivanie parametrov raspredeleniia ogranichennoi sluchainoi velichiny, robstnoe k narusheniiu granits [Estimation of distribution parameters of a bounded random variable robust to bound disturbance]. Science Bulletin of the NSTU, 2016, vol. 63, no. 2, pp. 70-89. DOI: 10.17212/1814-1196-2016-2-70-89
14. Tsyppkin Ya. Z. Osnovy informatsionnoi teorii identifikatsii [Basics of information theory of identification]. Moscow, Nauka, 1984, 320 p.
15. Basu A., Harris I. R., Hjort N. L., Jones M. C. Robust and efficient estimation by minimizing a density power divergence. Biometrika, 1998, vol. 85, pp. 549-559. DOI: 10.1093/BIOMET/85.3.549
16. Demidenko E. Z. Lineinaia i nelineinaia regressii [Linear and nonlinear regression]. Moscow, Finansy i statistika, 1981, 302 p.
17. Lisitsin D. V. Ustoichivoe otsenivanie parametrov modeli po mnogomernym neodnorodnym nepolnym dannym [Robust estimation of model parameters in presence of multivariate nonhomogeneous incomplete data]. Science Bulletin of the NSTU, 2013, vol. 50, no. 1, pp. 17-30.