

Лисицин, Д. В. Использование принципа максимальной энтропии для конструирования робастных оценок при байесовском точечном засорении. Часть I / Д.В. Лисицин, К.В. Гаврилов // Прикладная математика и вопросы управления. – 2024. – № 1. – С. 55–72. – DOI 10.15593/2499-9873/2024.1.04

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 I / Д.В. Лисицин, К.В. Гаврилов. – Текст : непосредственный // Прикладная математика и вопросы управления / Applied Mathematics and Control Sciences. – 2024. – № 1. – С. 55–72. – DOI 10.15593/2499-9873/2024.1.04



ПРИКЛАДНАЯ МАТЕМАТИКА
И ВОПРОСЫ УПРАВЛЕНИЯ
№ 1, 2024

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



Научная статья

DOI: 10.15593/2499-9873/2024.1.04

УДК 519.233.22



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

Д.В. Лисицин, К.В. Гаврилов

Новосибирский государственный технический университет, Новосибирск, Российская Федерация

О СТАТЬЕ

Получена: 03 октября 2023

Одобрена: 18 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

Развивается теория робастного оценивания параметров статистических моделей с привлечением аппарата теории информации. Рассмотрен подход А.М. Шурыгина, основанный на модели серии выборок со случайным точечным засорением (модели байесовского точечного засорения). Из предложенных А.М. Шурыгиным оценок, пожалуй, наиболее интересными свойствами обладают стойкие оценки. Хотя данный подход можно связать с подходом Ф. Хампеля к робастному оцениванию, необходимость при нахождении стойких оценок постулировать параметрический вид распределения засоряющей точки не позволяет считать это робастной процедурой.

В первой части нашей работы предложен непараметрический способ выбора указанного распределения – посредством максимизации энтропии Шеннона 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. Обобщенные радикальные оценки широко известны в зарубежных публикациях как оценки минимума логарифмической дивергенции степени плотности (гамма-дивергенции), при этом вопрос их оптимальности там не исследуется. К обобщенным радикальным относятся некоторые популярные оценки параметра сдвига: оценки Мешалкина (Уэлша), Эндрюса, Смита, Бернулли, бивес-оценка Тьюки, оценка Хьюбера типа урезанного среднего, обобщенные оценки Шарбонье.

Также в первой части работы предложено использовать функционал перекрестной энтропии. Перекрестная энтропия, применяемая в качестве оптимизируемого функционала вместо энтропии Шеннона, позволяет получить семейство оценок с наиболее широким диапазоном значений параметра, задающего это семейство. К задаче максимизации перекрестной энтропии сводится задача максимизации математического ожидания функции потерь оценок максимального правдоподобия в модели байесовского точечного засорения. По этой причине обобщенные радикальные оценки 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. Для выбора одной оценки из семейства, определяемого разными ограничениями на дивергенцию, предложен оптимизационный подход, аналогичный таковому, приводящему к стойким оценкам, но, в отличие от последнего, остающийся непараметрическим.

Основные теоретические результаты, полученные в работе, иллюстрируются во второй ее части на примере оценивания параметра сдвига косинусного распределения.

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

Гаврилов Константин Викторович (Новосибирск, Россия) – кандидат технических наук, доцент кафедры автоматизации, e-mail: k.gavrilov@corp.nstu.ru, ORCID: 0009-0001-9252-9942.



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 I. *Applied Mathematics and Control Sciences*. 2024, no. 1, pp. 55–72. DOI: 10.15593/2499-9873/2024.1.04

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 I. *Appl. Math. Control Sci.* **2024**, *1*, 55–72. <https://doi.org/10.15593/2499-9873/2024.1.04>

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 I”. *Appl. Math. Control Sci.* no. 1: 55–72. <https://doi.org/10.15593/2499-9873/2024.1.04>



APPLIED MATHEMATICS
AND CONTROL SCIENCES

№ 1, 2024

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



Article

DOI: 10.15593/2499-9873/2024.1.04

UDC 519.233.22



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

D.V. Lisitsin, K.V. Gavrilov

Novosibirsk State Technical University, Novosibirsk, Russian Federation

ARTICLE INFO

Received: 03 October 2023

Approved: 18 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, Tsallis divergence, Tsallis entropy, Tsallis divergence, alpha-divergence, Burg entropy, Hellinger distance, power-of-cosine distribution.

ABSTRACT

The paper develops 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. Perhaps, stoyki (steady) estimators have the most interesting properties among the estimators proposed by A. M. Shurygin. Although this approach can be associated with F. Hampel's approach to robust estimation, the need to postulate a parametric form of the contamination point distribution when finding stoyki estimators does not allow this to be considered a robust procedure.

In the first part of our work, a non-parametric method of selecting said distribution is proposed – by maximizing Shannon 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. Generalized radical estimators are widely known in foreign publications as estimators of the minimum logarithmic density power divergence (gamma-divergence), while the question of their optimality is not investigated there. Generalized radical estimators include some popular estimators of the location: the estimators of Meshalkin (Welsh), Andrews, Smith, Bernoulli, Tukey's biweight estimator, Huber-type skipped mean and the generalized Charbonnier estimators.

Also in the first part of the work it is proposed to use the cross entropy functional. Cross entropy, used as an optimized functional instead of Shannon entropy, allows us to get a family of estimators with the widest range of values of the parameter specifying this family. Since the problem of maximizing the expectation of the loss function of the maximum likelihood estimators in the point Bayesian contamination model is reduced to the problem of maximizing the cross entropy, the generalized radical estimators can be interpreted as being defended against malicious modifying of the maximum likelihood estimates.

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 that is similar to the approach leading to stoyki estimators, but, unlike the latter, remaining non-parametric.

The main theoretical results obtained in the paper are illustrated in its second part 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: 0009-0001-9252-9942.



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

Введение

Одним из направлений развития статистической теории являются методы оценивания параметров моделей, направленные на обеспечение устойчивости оценок к отклонению распределения наблюдений от постулируемого (модельного) [1–4]. Важность таких оценок обусловлена тем, что на практике модель неизбежно расходится с реальной ситуацией по причине наличия некачественных наблюдений или неточности самой модели, а многие классические методы оценивания, например, метод максимального правдоподобия [1] являются неустойчивыми.

В данной работе развивается подход А.М. Шурыгина, связанный с моделью серии выборок со случайным точечным засорением [2], где каждая выборка из серии имеет распределение в виде смеси модельного распределения, параметры которого оцениваются, и распределения, сосредоточенного в одной фиксированной точке. Случайная засоряющая точка в серии выборок имеет некоторое распределение, которое может постулироваться или определяться в результате решения той или иной оптимизационной задачи. Описанные построения будем называть моделью *байесовского точечного засорения* (БТЗ) [2, 5]. Оптимальная оценка в данном подходе для фиксированного распределения засоряющей точки определяется минимизацией асимптотического квадратичного отклонения оценки при некоторых дополнительных асимптотических предположениях. Более подробно данный подход описан в п. 1.

Нахождение оптимальной оценки сближает рассматриваемый подход с классическими подходами П. Хьюбера [3] и Ф. Хампеля [4] к робастному оцениванию. Последние дают оптимальные оценки в непараметрической окрестности модельного распределения, при этом часто окрестность формируется путем засорения модельного распределения. Отметим, что в подходе на основе модели БТЗ окрестность задается несколько более сложным образом. Классические решения теории робастности часто являются недостаточно устойчивыми (например, при оценивании параметра сдвига симметричного распределения в условиях его несимметричного засорения). Стойкие оценки, получаемые в рамках подхода на основе модели БТЗ, при подходящем распределении засоряющей точки могут обладать большей устойчивостью. Однако задавать такое распределение приходится в параметрическом виде, что не позволяет рассматривать данную схему как робастную в классическом смысле.

Большой практический интерес представляют оценки, которые соответствуют выбору плотности засоряющей точки, пропорциональной некоторой степени (как правило, не превышающей единицы) модельной плотности [6–8]. Мы будем называть такие оценки *обобщенными радикальными* или, сокращенно, ООР¹, поскольку их частным случаем являются радикальные оценки [2]. С точки зрения теории Шурыгина, но вне контекста модели БТЗ, обсуждаемые оценки или их частные случаи изучались, например, в работах [9–11]. В зарубежных публикациях данные оценки обычно называют оценками минимума логарифмической дивергенции степени плотности или гамма-дивергенции, т.е. они рассматриваются как частный случай оценок метода минимума расстояния, при этом вопрос их оптимальности не исследуется (см., например, [12–16]). Частными случаями ООР оказываются известные робастные оценки параметра сдвига, введенные их авторами в большей или меньшей степени эвристически, например, оценки Мешалкина (в зарубежной литературе известные как оценки Уэлша), Эндрюса, Смита, Бернулли, бивес-оценка Тьюки, оценка Хьюбера типа урезанного среднего, обобщенные оценки Шарбонье [2; 8; 17]. Так, в [4] высказано предположение, что бивес-оценка Тьюки и оценка Эндрюса не имеют каких-либо свойств оптимально-

¹ ООР – оценка обобщенная радикальная.

сти. В связи со сказанным обоснование ООР как *робастных* оценок, обладающих определенными *свойствами оптимальности*, представляется нам актуальным.

Для такого обоснования в рамках подхода на основе модели БТЗ необходим способ непараметрического задания плотности засоряющей точки указанного вида. Оказывается, возможный способ состоит в использовании широко известного *принципа максимальной энтропии* [18]: нужная плотность является решением задачи максимизации *энтропии Шеннона* [19, 20] в окрестности модельного распределения, которая определяется ограничением на *дивергенцию Кульбака – Лейблера* [1] между распределением засоряющей точки и модельным распределением [21]. Изложению этих результатов посвящен п. 2.

Описанный подход является довольно общим: вместо энтропии Шеннона и дивергенции Кульбака – Лейблера в оптимизационной задаче можно использовать иные виды энтропий и дивергенций. Здесь и далее имеются в виду непрерывные версии этих функционалов. В наиболее общем случае можно говорить о выборе плотности, оптимальной по какому-либо целевому критерию (характеризующему, например, разброс случайной величины или неточность модели) на некотором непараметрически заданном множестве допустимых плотностей. В результате будут получаться другие плотности засоряющей точки и, как следствие, – новые семейства оценок. В данной работе предлагается ограничить произвол в выборе функционалов: в рамках одной вариационной задачи будут использоваться функционалы энтропии и дивергенции, которые определенным образом связаны между собой. Так, во второй части статьи (п. 1) используются энтропия Реньи в паре с дивергенцией Реньи [22] или эквивалентные им с точки зрения оптимизационной задачи энтропия и дивергенция Цаллиса [23]. Получаемое семейство оценок охватывает многие полезные на практике устойчивые решения, включая ООР и условно оптимальные оценки Шурыгина, которые описываются ниже в п. 1 [2; 24] (см. также [15] как пример получения этих оценок вне теории Шурыгина). Условно оптимальные оценки являются оптимальным решением в рамках другого – локально устойчивого – подхода Шурыгина. Его отличие от робастного подхода Хампеля в значительной мере состоит в использовании другой нормы функции влияния: вместо L_∞ используется L_2 .

Рассматриваемые нами оценки являются семействами с одним или несколькими параметрами. Такое положение является обычным в теории робастности, однако интерес может представлять и обоснованный выбор какого-то одного члена семейства. Способ такого выбора, аналогичный способу получения стойких оценок Шурыгина, но, в отличие от последнего, остающийся непараметрическим, описан во второй части статьи (п. 2).

Основные теоретические результаты, полученные в работе, проиллюстрированы на примере оценивания параметра сдвига *косинусного распределения* [7; 8] и приведены во второй части статьи (п. 3 и 4).

1. Элементы теории устойчивого оценивания

Пусть x_1, \dots, x_m – наблюдения непрерывной случайной величины ξ , распределенной с плотностью $f(x, \theta)$, где $x \in X \subseteq R$ и параметр $\theta \in \Theta \subseteq R$. Будем предполагать, что функция $f(x, \theta)$ непрерывна на $X \times \Theta$. Если X имеет конечную длину, такую модель будем называть *финитной*. M -оценка неизвестного параметра θ может определяться [1; 3] как решение оптимизационной задачи

$$\hat{\theta} = \arg \min_t \sum_{i=1}^m \rho(x_i, t),$$

где $\rho(x, \theta) : X \times \Theta \rightarrow R$ – непрерывная, дифференцируемая почти всюду функция, удовлетворяющая условию асимптотической несмещенности [1] $\theta = \arg \min_t \mathbf{E} \rho(\xi, t)$, \mathbf{E} – оператор математического ожидания. Условие равенства нулю производной по θ оптимизируемой функции служит альтернативной (хотя и не эквивалентной [1]) формулировкой задачи M -оценивания:

$$\sum_{i=1}^m \psi(x_i, \hat{\theta}) = 0.$$

Данное уравнение называется *оценочным*, а функция

$$\psi(x, \theta) = c(\theta) \dot{\rho}(x, \theta) \quad (1)$$

называется *оценочной функцией* для параметра θ , $c(\theta)$ – произвольная непрерывная функция, не равная нулю для всех $\theta \in \Theta$. Множитель $c(\theta)$ в оценочной функции (очевидно, его введение не влияет на решение оценочного уравнения) задает семейство *эквивалентных оценочных функций*. Здесь и далее точкой сверху обозначено дифференцирование по оцениваемому параметру.

Оценочная функция также должна удовлетворять условию асимптотической несмещенности, которое принимает вид [1; 25]

$$\mathbf{E} \psi(\xi, \theta) = \int_X \psi(x, \theta) f(x, \theta) dx = 0. \quad (2)$$

Дифференцируя (2) по θ и допуская возможность изменения порядка дифференцирования и интегрирования, можно записать следующие равенства [3; 25]:

$$N(\theta) \stackrel{df}{=} - \lim_{t \rightarrow \theta} \frac{\partial}{\partial t} \mathbf{E} \psi(\xi, t) = - \mathbf{E} \dot{\psi}(\xi, \theta) = \int_X \psi(x, \theta) \dot{f}(x, \theta) dx. \quad (3)$$

В рассматриваемой теории можно предложить различные формулировки условий регулярности (см., например, [1; 3]), однако их детальное обсуждение выходит за рамки статьи. Далее будет приведен набор условий, который выглядит наиболее естественно с точки зрения используемых в статье построений. Приведенные ниже условия должны выполняться локально – в некоторой окрестности истинного значения параметра θ . Итак, потребуем, чтобы:

- 1) выполнялось условие асимптотической несмещенности (2);
- 2) были справедливы равенства (3), функция $N(\theta)$ была непрерывной и не равной нулю [6];
- 3) выполнялось условие $\left| \lim_{t \rightarrow \theta} \frac{\partial^2}{\partial t^2} \mathbf{E} \psi(\xi, t) \right| < \infty$;
- 4) выполнялось условие $\mathbf{E} \psi^2(\xi, \theta) < \infty$.

При выписанных условиях M -оценка $\hat{\theta}$ является \sqrt{m} -состоятельной и асимптотически нормальной [26]. Ее нормированная асимптотическая дисперсия [1] определяется выражением

$$V(\psi, f) = \frac{1}{N^2(\theta)} \int_X \psi^2(x, \theta) f(x, \theta) dx.$$

В теории робастного оценивания конструируются оценки, имеющие высокое качество не только при постулируемом распределении (плотности f), но и при отклонении от него. Одним из наиболее широко используемых инструментов для анализа качества оценки в

теории робастности является *функция влияния* Хампеля [4]. Она отражает воздействие на оценку бесконечно малого точечного засорения модельного распределения. Оценки с ограниченной функцией влияния называют *B-робастными*.

Функция влияния *M-оценок* задается как $IF(x; \psi, f) = \psi(x, \theta) / N(\theta)$, откуда следует $V(\psi, f) = \mathbf{E} IF^2(\xi; \psi, f)$.

Перейдем к рассмотрению подходов Шурыгина, и первым рассмотрим локально устойчивый подход, основанный на показателе *неустойчивости* [2; 5; 9]

$$W(\psi, f) = \frac{1}{N^2(\theta)} \int_X \psi^2(x, \theta) dx = \int_X IF^2(x; \psi, f) dx.$$

Показатель *W* представляет собой квадрат L_2 -нормы функции влияния (подобно тому, как в [4] одним из критериев оптимальности является L_∞ -норма функции влияния). Оценку и соответствующую оценочную функцию будем называть устойчивыми по Шурыгину, если $W(\psi, f) < \infty$.

Эффективностью и *устойчивостью* оценки называются соответственно характеристики $eff \psi = V_{\text{ОМП}} / V(\psi, f)$ и $stb \psi = W_{\text{ОМУ}} / W(\psi, f)$, где $V_{\text{ОМП}}$ – асимптотическая дисперсия оценки максимального правдоподобия (ОМП), минимизирующей $V(\psi, f)$, $W_{\text{ОМУ}}$ – неустойчивость оценки максимальной устойчивости (ОМУ) [2], минимизирующей $W(\psi, f)$.

В рамках данного подхода построено семейство *условно оптимальных* оценочных функций [2, 5, 24, 27]

$$\psi(x, \theta, \lambda) = c(\theta, \lambda) \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) + \beta(\theta, \lambda) \right] \frac{f(x, \theta)}{1 + \lambda f(x, \theta)}, \quad \lambda \geq 0, \quad (4)$$

где функция $c(\theta, \lambda)$ имеет тот же смысл, что и в (1), функция $\beta(\theta, \lambda)$ определяется из условия (2). Значение параметра $\lambda = 0$ в (4) соответствует ОМУ, значение $\lambda = \infty$ – ОМП. Промежуточные значения параметра λ позволяют получить различные условно оптимальные оценки, имеющие минимальную неустойчивость при ограничении сверху на величину асимптотической дисперсии либо минимальную асимптотическую дисперсию при ограничении сверху на величину неустойчивости [9; 27].

Множество значений параметра λ в выражении (4) можно дополнить отрицательными значениями, такими, что $-1 / \max_{x \in X} f(x, \theta) < \lambda < 0$. Соответствующее семейство оценочных функций является решением задачи минимизации неустойчивости *W* при ограничении равенстве $V = V_0$ на асимптотическую дисперсию оценки [24; 27], где $V_{\text{ОМП}} \leq V_0 < V_{\text{max}}$, V_{max} – предел (конечный или бесконечный) асимптотической дисперсии оценки при $\lambda \rightarrow -1 / \max_{x \in X} f(x, \theta)$.

Рассмотрим подход Шурыгина, основанный на модели БТЗ [2; 5]. Пусть плотность распределения случайной величины имеет вид

$$(1 - \alpha) f(x, \theta) + \alpha \delta(x - v),$$

где α – доля аномальных наблюдений, $0 \leq \alpha < 1$; δ – функция Дирака; v – засоряющая точка такая, что в пределах одной выборки $v = \text{const}$, а в серии выборок представляет собой случайную величину, распределенную на X с плотностью $s = s(x, \theta)$. В дополнение к условиям,

описанным в [2], будем предполагать, что функция плотности $s(x, \theta)$ непрерывна на $X \times \Theta$. Пусть α – бесконечно малая величина с порядком малости, меньшим $1/2$, т.е. $\alpha = \gamma m^{-\zeta/2}$, где $\gamma > 0$ и $0 < \zeta < 1$. Тогда нормированное асимптотическое квадратичное отклонение оценки $\gamma^{-2} \lim_{m \rightarrow \infty} m^\zeta E(\hat{\theta} - \theta)^2$ не зависит от параметров γ и ζ и определяется выражениями [2]

$$U(\psi, s) = \frac{1}{N^2(\theta)} \int_X \psi^2(x, \theta) s(x, \theta) dx = \int_X \text{IF}^2(x; \psi, f) s(x, \theta) dx. \quad (5)$$

В [2] показано, что функционал (5) достигает минимума на оценочной функции

$$\psi_s(x, \theta) = c(\theta) \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) + \beta(\theta) \right] \frac{f(x, \theta)}{s(x, \theta)}, \quad (6)$$

где функция $c = c(\theta)$ имеет тот же смысл, что и в (1), функция $\beta = \beta(\theta)$ определяется из условия (2).

Функционал (5) можно интерпретировать как квадрат нормы функции влияния в весовом L_2 -пространстве с весом s , т.е. постулировать вне модели БТЗ. В этом случае весовая функция s не интерпретируется как плотность, но должна быть неотрицательной.

Заметим также, что в выражение (6) допустимо подставлять функции s , не удовлетворяющие условию нормировки плотности из-за того, что нормировочная константа плотности поглощается сомножителем $c(\theta)$ (справедливость такого подхода для неинтегрируемых функций s будет показана ниже). Нетрудно видеть, что во всех этих случаях решение (6) по-прежнему доставляет минимум функционалу (5). Но если функция s не является интегрируемой, то модель БТЗ не может быть применена для обоснования показателя (5), а он не будет иметь смысла асимптотического квадратичного отклонения оценки (хотя сохраняет смысл квадрата весовой L_2 -нормы функции влияния).

Для использования оценочной функции (6) необходимо выбрать плотность $s(x, \theta)$, но на практике вряд ли можно это сделать объективно. Возможное решение в этой ситуации – в соответствии с принципами теории робастности произвести поиск наилучшей оценки в наихудших условиях [3], например, использовать минимаксную формулировку

$$\Psi_{\text{med}} = \arg \min_{\psi} \max_{s \in S} U(\psi, s), \quad (7)$$

где S – множество плотностей, ограниченное лишь условиями регулярности. Нижний индекс «med» в обозначении минимаксной оценочной функции обусловлен тем, что она является *медианной* [2; 28], т.е. имеет вид

$$\Psi_{\text{med}}(x, \theta) = c(\theta) \operatorname{sgn} \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) + \beta(\theta) \right].$$

Медианная оценка, уже известная в рамках классических подходов к робастному оцениванию [3; 4], является B -робастной, однако часто она недостаточно устойчива, например, при оценивании параметра сдвига симметричного распределения в условиях его несимметричного засорения [2].

Таким образом, свойства B -робастности оценки не всегда достаточно для ее устойчивости на практике. Большой устойчивостью обладают оценки со свойством

$$\text{IF}(x, \theta) \approx 0 \text{ при } f(x, \theta) \approx 0,$$

обобщающим свойство $\psi(x, \theta) \rightarrow 0$ при $|x| \rightarrow \infty$, которому удовлетворяют *сниженные оценки* параметра сдвига для $X = R$ [4]. Данное свойство выглядит естественно, например, в рамках борьбы с выбросами [13]: признаком выброса может служить то, что соответствующее значение модельной плотности близко к нулю; и если функция влияния тоже близка к нулю, то влияние выброса на оценку нивелируется. Поскольку мы работаем с непрерывными распределениями, данное условие для рассмотренных в статье построений будем использовать в усиленном виде:

$$IF(x, \theta) \rightarrow 0 \text{ при } f(x, \theta) \rightarrow 0. \quad (8)$$

Другой путь построения оптимальных оценок в модели БТЗ – решение максиминной задачи [2, 28]

$$s_*(x, \theta) = \arg \max_{s \in S} \min_{\psi} U(\psi, s) = \arg \max_{s \in S} U(\psi_s, s), \quad (9)$$

откуда находим $\psi_* = \psi_{s_*}$, подставляя (9) в (6). Здесь плотность (9) – это наихудшая плотность распределения засоряющей точки [2] на некотором множестве плотностей S . Если S представляет собой параметрическое семейство плотностей, включающее модельную плотность f , тогда решение ψ_* соответствует *стойким оценкам* [2; 5]. Стойкие оценки часто обладают свойством (8), важным для обеспечения устойчивости. Однако параметрический способ задания множества S не отвечает методологии теории робастности, где модель засорения определяется непараметрически.

Как правило, свойством (8) обладают также ООР, получаемые при выборе плотности засоряющей точки, пропорциональной степени модельной плотности:

$$s_\nu(x, \theta) = k f^{1-\nu}(x, \theta), \quad (10)$$

где $k = k(\theta, \nu)$ – нормировочная константа, $\nu \geq 0$ [6, 7]; они обобщают радикальные оценки Шурыгина с плотностью $s_{1/2}(x, \theta) = k \sqrt{f(x, \theta)}$ [2; 5].

Формально подставив плотность вида (10) в выражение (6), получаем семейство оценочных функций

$$\psi(x, \theta, \nu) = c(\theta, \nu) \left[\frac{\partial}{\partial \theta} \ln f(x, \theta) + \beta(\theta, \nu) \right] f^\nu(x, \theta). \quad (11)$$

В этом выражении нормировочная константа k поглощается функцией c . В частности, при $\nu = 0$ имеем ОМП – наименее устойчивую в семействе оценку, при $\nu = 1$ имеем ОМУ – наиболее устойчивую оценку в рамках локально устойчивого подхода Шурыгина.

Обозначим через ν_0 верхнюю грань множества значений ν , для которых существует плотность (10). Здесь возможны следующие варианты.

1. Неограниченный (с одной или обеих сторон) интервал X . В этом случае $\nu_0 \leq 1$. Если при $|x| \rightarrow \infty$ имеет место $\ln f(x, \theta) \sim -\tau \ln |x|$, где $\tau > 1$, то $\nu_0 = 1 - 1/\tau$. Так, для распределения Коши $\tau = 2$, $\nu_0 = 1/2$. Если в левой и правой окрестностях бесконечности плотность $f(x, \theta)$ стремится к нулю с различной скоростью, следует взять наименьшую из двух τ .

2. Ограниченный интервал X . В этом случае $\nu_0 > 1$. Если при этом $\inf_{x \in X} f(x, \theta) > 0$, то $\nu_0 = \infty$. Например, рассмотрим *косинусно-степенное распределение* с плотностью [29]

$$f(x, \theta, l, \kappa) = \frac{\sqrt{\pi} \Gamma((\kappa+1)/2)}{2l \Gamma(\kappa/2)} \cos^{\kappa-1} \frac{\pi(x-\theta)}{2l}, \quad |x-\theta| < l, \quad \kappa > 0, \quad (12)$$

где $\Gamma(\cdot)$ – гамма-функция. Для $\kappa > 1$ имеем $\nu_0 = \kappa/(\kappa-1)$, для $0 < \kappa \leq 1$ имеем $\nu_0 = \infty$.

При $\nu \geq \nu_0$, поскольку выражения (5) и (6) это допускают, вместо плотности (10) можно использовать ненормированную функцию $\tilde{s}_\nu(x, \theta) = f^{1-\nu}(x, \theta)$. Соответствующие решения также могут рассматриваться как формальное продолжение семейства (11) в область $\nu \geq \nu_0$. Семейство (11) допускает распространение и на отрицательные значения параметра ν , однако в рамках теории робастности изучение оценок, менее устойчивых, чем ОМП, как правило, малоинтересно.

Интерпретировать описанный выше подход можно следующим образом. Строго говоря, модель БТЗ неприменима, если плотность распределения засоряющей точки не существует. Однако понятие модели БТЗ можно обобщить и на такие распределения, рассматривая неинтегрируемую функцию \tilde{s} как предел некоторой последовательности интегрируемых весовых функций \tilde{s}_n , где $n = 1, 2, \dots$. Действительно, рассмотрим оценочную функцию (6). Далее для краткости часто будем опускать аргументы функций. Так, оценочная функция (6) может быть записана в виде $\psi = c(\dot{f} + \beta f)/s$. Пусть для определенности величина c здесь выбрана таким образом, что оценочная функция совпадает с функцией влияния (это не уменьшает общности изложения, так как все эквивалентные оценочные функции отличаются от соответствующей функции влияния только ненулевым множителем c). Тогда для элементов последовательности \tilde{s}_n имеем соответственно следующие БТЗ-оптимальные решения:

$$\psi_n = \frac{\frac{\dot{f} + \beta_n f}{\tilde{s}_n} \int_X \tilde{s}_n dx}{\int_X \frac{\dot{f} + \beta_n f}{\tilde{s}_n} \dot{f} dx \int_X \tilde{s}_n dx} = \frac{\dot{f} + \beta_n f}{\tilde{s}_n \int_X \frac{\dot{f} + \beta_n f}{\tilde{s}_n} \dot{f} dx},$$

где β_n являются решениями уравнения (2) относительно β для оценочных функций ψ_n . Поскольку в этом выражении нормировочная константа плотности $1/\int_X \tilde{s}_n dx$ сокращается, предельный переход может быть выполнен непосредственной подстановкой предельных функций:

$$\psi = \lim_{n \rightarrow \infty} \psi_n = (\dot{f} + \beta f) / \left(\tilde{s} \int_X \frac{\dot{f} + \beta f}{\tilde{s}} \dot{f} dx \right).$$

При этом предполагаем, что предельная функция $\beta = \lim_{n \rightarrow \infty} \beta_n$ существует и удовлетворяет равенству (2) для $\psi = c(\dot{f} + \beta f)/\tilde{s}$ (выбор функции c не влияет на β). Таким образом, в пределе получается выражение для функции влияния, соответствующей неинтегрируемой функции \tilde{s} .

Для получения последовательности функций \tilde{s}_n бывает удобно рассматривать последовательность усеченных интервалов X_n , когда функции \tilde{s}_n совпадают с \tilde{s} в центральной части интервала X и равны 0 в окрестности его границ. В отличие от описанного выше

подхода, здесь каждой \tilde{z}_n ставится в соответствие модельная плотность f_n , полученная аналогично в результате усечения плотности f . Поскольку последовательность f_n имеет пределом f , введение этой последовательности не влияет на предельную оценочную функцию. Однако справедливость равенств $\psi = \lim_{n \rightarrow \infty} \psi_n$ и $\beta = \lim_{n \rightarrow \infty} \beta_n$ не очевидна и требует постулирования. Кроме того, для получения БТЗ-оптимальной ψ_n требуются те же условия регулярности, что и для исходной модели.

Расширение модели БТЗ на неинтегрируемые функции \tilde{z} , играющие здесь роль несобственных плотностей распределения засоряющей точки (см., например, [30]), будем называть *обобщенной моделью БТЗ*.

Применение описанной теории к финитным моделям, в том числе полученным выше в результате усечения исходного множества X , имеет свою специфику. Появление наблюдений за пределами интервала финитности X возможно только с нулевой вероятностью (в незасоренной модели), и можно выделить два подхода к их обработке: традиционный, в котором такие наблюдения запрещаются, и обобщенный, при котором они допускаются, а оценочная функция за пределами интервала финитности полагается равной нулю [8]. Обобщенный подход имеет приблизительно тот же прикладной смысл, что и условие (8): маловероятные наблюдения должны игнорироваться.

С целью обеспечения регулярности финитной модели в данной работе подразумевается применение обобщенного подхода к оцениванию параметров. Это означает, что в модели БТЗ плотности f и s определяются на $x \in R$. Плотность f на $x \in R \setminus X$ доопределяется нулем, поэтому из (6) следует БТЗ-оптимальность оценочной функции в обобщенном подходе при условии, что функция s на $x \in R \setminus X$ доопределяется произвольным образом, но не равна нулю. При этом происходит нарушение нормировки s , что, как уже было сказано, не влияет на справедливость (6). Таким образом, плотность s представляет интерес только при $x \in X$, поэтому нормировать ее (если это возможно) и приводить выражения для нее будем только на этом интервале.

2. Оптимальность обобщенных радикальных оценок

Для обоснования ООР как робастных оценок в рамках подхода Шурыгина, связанного с байесовским точечным засорением, покажем, что плотность (10) может быть получена как результат непараметрической оптимизации. Такой способ нахождения оптимальной плотности, использующий аппарат теории информации, описан в [21] и основан на принципе максимальной энтропии [18].

Определение. Будем говорить, что энтропия $H(s)$ порождается дивергенцией $D(s, f)$, если выполнено формальное равенство [20]

$$H(s) = \varphi - D(s, 1) \quad (13)$$

для некоторой $\varphi = \text{const}$.

Эта формула отражает свойство энтропии, согласно которому ее безусловный максимум достигается на равномерном распределении вероятностей. Условие (13) обеспечивает единообразие подходов к характеристике неопределенности распределения и непохожести между распределениями в рамках одной оптимизационной задачи.

Итак, плотность (10) является решением задачи максимизации *энтропии Шеннона* (дифференциальной энтропии) [19; 20]

$$H(s) = - \int_x s \ln s \, dx, \quad (14)$$

порождаемой дивергенцией Кульбака – Лейблера [1]

$$D(s, f) = \int_x s \ln \frac{s}{f} \, dx, \quad (15)$$

при ограничении

$$D(s, f) \leq \Delta, \quad (16)$$

где Δ – неотрицательная величина. Действительно, лагранжиан данной задачи имеет вид

$$L = -(a+1)s \ln s + a s \ln f + bs,$$

где $a \geq 0$, b – множители Лагранжа, и условие стационарности –

$$\partial L / \partial s = -(a+1)(\ln s + 1) + a \ln f + b = 0.$$

Отсюда после преобразований находим $s = k f^{1-v}$, где $k = e^{vb-1}$ и $v = 1/(a+1)$, причем $0 \leq v \leq 1$. Однако в случае $v_0 \leq 1$ функционалы (14), (15) и модель БТЗ при $v_0 \leq v \leq 1$ не имеют смысла, а формально найденное решение задачи с такими значениями параметра v (им соответствует $k = 0$) нуждается в дополнительном объяснении. Это объяснение строится на основе обобщенной модели БТЗ и будет рассмотрено в этом параграфе ниже для более общего случая $v \geq 0$.

Одним крайним членом получаемого семейства является ОМП при $\Delta = 0$, другим – для случая $v_0 > 1$ – ОМУ, которая соответствует безусловному максимуму (14). Последний достигается при любых $\Delta \geq D(s_1, f)$, где s_1 – плотность равномерного на X распределения. В случае $v_0 \leq 1$ при $\Delta \rightarrow \infty$ соответствующим решением (11) в пределе будет $\psi(x, \theta, v_0)$. Таким образом, рассмотренная оптимизационная задача позволяет получать решения из семейства (11) в диапазоне $0 \leq v \leq \tilde{v}$ (с возможностью расширения до $0 \leq v \leq 1$ за счет обобщенной БТЗ), где

$$\tilde{v} = \min \{v_0; 1\}.$$

С этой точки зрения случай $s = s_1$ (или $\tilde{s} = \text{const}$), названный нами максимально неопределенным байесовским точечным засорением [7], соответствует наиболее слабым предположениям в рамках рассматриваемого подхода – отсутствию ограничения (16) – и одновременно наихудшей с точки зрения критерия (14) ситуации. В свою очередь, интерпретация неустойчивости W как показателя (5) при максимально неопределенном байесовском точечном засорении позволяет представить локально устойчивый подход Шурыгина частным случаем его же подхода на основе модели БТЗ при наиболее слабых предположениях (с той лишь разницей, что в случае $v_0 \leq 1$ используется обобщенная модель БТЗ, как описано в п. 1).

Поскольку работа [21] не связана с робастным оцениванием параметров, дадим интерпретацию рассматриваемого решения с точки зрения нашей задачи. В теории робастности решения часто формулируются как минимаксные, подобно (7): находится наилучшая оценка при наихудшем засорении [3]. Отличие нашего подхода от классического состоит в том, что мы используем два критерия: по одному критерию находится наихудшее засоре-

ние, а по другому – наилучшая оценка. Отказываясь от минимаксной задачи (7), мы теряем гарантированность результата, зато приобретаем возможность получать оценки, удовлетворяющие условию (8).

С точки зрения теории принятия решений в условиях неопределенности подход на основе двух критериев соответствует игре с противоположными интересами [31], где исследователь «играет» с природой. Действительно, природа о задаче оценивания модели ничего не знает, поэтому в общем случае не имеет со статистиком противоположные интересы. Но разумно предположить, что она стремится к увеличению неопределенности наблюдаемой системы, выражаемой в нашем случае энтропией Шеннона. Иначе говоря, стратегия природы – максимизация энтропии, стратегия статистика – получение оптимальной оценки при данной стратегии природы. Применительно к модели БТЗ стратегия природы направлена на серию выборов, т.е. определяет плотность s .

Получение решения в виде семейства, разные члены которого полезны при разной степени искажения модельного распределения, также является характерным для теории робастности. Поскольку решением в случае $s = f$ является ОМП, которую целесообразно принять в качестве крайнего (наименее устойчивого) члена семейства, степень искажения определяется степенью непохожести плотности s на модельное распределение, а сама степень непохожести формализуется через ограничение на дивергенцию. Подобно тому, как это делается в подходе Хьюбера, ограничение (16) формирует окрестность модельного распределения размера Δ , в которой ведется поиск наихудшего распределения.

На практике указанное ограничение может формироваться на основе априорной информации или, если ее нет, на основе формальных процедур, например, с использованием максиминного подхода, как в стойких оценках (см. п. 2 второй части статьи); возможен и субъективный выбор нескольких значений с целью исследования изменения оценок при изменении ограничения (размера окрестности).

Вместо размера окрестности Δ на практике задают параметр ν оценочной функции, что проще. Чтобы найти размер окрестности Δ при известном значении ν , подставим (10) в (15). Получаем

$$D(s_\nu, f) = \ln k - k \nu \int_x f^{1-\nu} \ln f \, dx = \frac{\ln k + \nu H(s_\nu)}{1-\nu}, \quad 0 \leq \nu < \tilde{\nu}. \quad (17)$$

Если $\nu_0 > 1$, формулы (17) позволяют находить $D(s_\nu, f)$ и для $1 \leq \nu < \nu_0$, однако результат в этом случае не может интерпретироваться как размер окрестности Δ . Кроме того, $D(s_1, f)$ определяет лишь нижнюю границу размера окрестности.

Заметим, что даже при наличии устойчивости по Шурыгину функция влияния оптимальной оценки может оказаться неограниченной (см. п. 3 второй части статьи), что будет означать отсутствие робастности с точки зрения подхода Хампеля [4] (отсутствие качественной робастности оценки). С практической точки зрения такие оценки тоже нельзя считать устойчивыми. Таким образом, робастными мы будем считать лишь те БТЗ-оптимальные оценки, которые удовлетворяют условию B -робастности.

Для обоснования семейства (11) при всех $\nu \geq 0$ в соответствии с [8, 32], необходимо максимизировать вместо энтропии Шеннона (14) *перекрестную энтропию* (неточность Керриджа) [13; 33]

$$H(s, f) = - \int_x s \ln f \, dx. \quad (18)$$

Заметим, что в [34] такая задача сформулирована в рамках принципа максимума функции неточности. В интерпретации новой задачи с точки зрения игры с противоположными интересами исследователь «играет», скорее, уже не с безразличной природой, а с целенаправленно действующим противником (см., например, [35, 36]). Ограничение на дивергенцию тогда можно понимать как один из инструментов маскировки последнего.

Легко увидеть, что к задаче максимизации по s перекрестной энтропии сводится задача максимизации математического ожидания функции потерь ОМП $\rho(x, \theta) = -\ln f(x, \theta)$ в модели БТЗ при фиксированной плотности f и истинном значении параметра θ (параметр α ненастраиваемый).

Заметим в связи с этим, что максимизация критерия на основе функции потерь часто используется во вредоносном машинном обучении (*adversarial machine learning*) с целью формирования наихудшего искажения («отравления») обучающих данных с последующим получением оценок, защищенных от этого искажения [36]. С рассматриваемой точки зрения задача максимизации перекрестной энтропии соответствует наличию у противника предположения об использовании исследователем ОМП, в то время как исследователь выбирает оценку, защищенную от искажающего воздействия противника. Цель противника тогда состоит в компрометации истинного значения параметра как согласованного с данными. Таким образом, ООР могут быть охарактеризованы как защищенные от намеренного искажения ОМП.

Отметим также, что функционал (14) может рассматриваться как предельный случай (18) [37]:

$$\min_f H(s, f) = H(s).$$

Использование энтропии тогда можно интерпретировать как формализацию незнания противником модельной плотности.

Рассмотрение задачи максимизации функционала (18) при ограничении (16) может натолкнуть на следующую трудность. Решение (10) формально может быть записано для всех $\nu \geq 0$, однако при $\nu \geq \nu_0$ никакие значения множителей Лагранжа не удовлетворяют условию нормировки плотностей s_ν , поскольку соответствующие функции \tilde{s}_ν неинтегрируемые. По этой же причине функционалы (15) и (18) не имеют смысла, а модель БТЗ неприменима.

Для преодоления указанной трудности вместо интервала X рассмотрим последовательность усеченных (с одной или двух сторон по необходимости) интервалов X_n (где n – натуральные числа), таких что $\lim_{n \rightarrow \infty} X_n = X$ (под сходимостью последовательности интервалов понимается сходимость последовательностей соответствующих границ) и для каждого X_n существуют плотности вида (10) $s_{\nu,n} = k_n f_n^{1-\nu}$. Плотности f_n получены в результате соответствующего усечения плотности f и могут рассматриваться как ее аппроксимации.

Решая задачу оптимизации на интервале X_n аналогично [21] и подставляя найденную оптимальную плотность $s_{\nu,n}$ в (6), получаем семейство оценочных функций вида (11) для модельной плотности f_n и $\nu \geq 0$. Таким образом, семейство (11) в случае, когда плотность s_ν не существует, т.е. $\nu \geq \nu_0$, может интерпретироваться как предел аппроксимирующих семейств, построенных для интервалов X_n при $n \rightarrow \infty$. Как и в п. 1, будем предполагать, что изложенный здесь способ построения обобщенной модели БТЗ возможен.

Заключение

В первой части работы предложен непараметрический способ выбора плотности засоряющей точки для модели БТЗ, что позволяет рассматривать решения, получаемые в рамках соответствующего подхода А.М. Шурыгина, как решения теории робастности.

Рассмотрено решение, в котором оптимальная плотность определяется в результате максимизации функционала энтропии при ограничении на величину дивергенции Кульбака – Лейблера между искомой и модельной плотностями. Характерно, что данные функционалы энтропии и дивергенции связаны между собой. В результате получается известное семейство ООР при условии $0 \leq v \leq 1$, где v – параметр, задающий семейство. Использование в качестве оптимизируемого функционала вместо энтропии перекрестной энтропии позволяет вывести семейство ООР для всех $v \geq 0$. Таким образом, получено обоснование семейства ООР как робастных оценок, обладающих свойством оптимальности.

Список литературы

1. Боровков, А.А. Математическая статистика / А.А. Боровков. – СПб.: Издательство «Лань», 2021. – 704 с.
2. Шурыгин, А.М. Математические методы прогнозирования: учебное пособие для вузов / А.М. Шурыгин. – М.: Горячая линия – Телеком, 2009. – 180 с.
3. Хьюбер, П. Робастность в статистике / П. Хьюбер. – М.: Мир, 1984. – 303 с.
4. Робастность в статистике: подход на основе функций влияния / Ф. Хампель, Э. Ронchetti, П. Рауссеу, В. Штаэль. – М.: Мир, 1989. – 512 с.
5. Shurygin, A.M. New approach to optimization of stable estimation / A.M. Shurygin // Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach. – 1994. – Vol. 3. Dordrecht: Kluwer. – P. 315–340. DOI: 10.1007/978-94-011-0854-6_15
6. Лисицин, Д.В. Об устойчивом оценивании параметров модели при асимметричном засорении данных / Д.В. Лисицин, К.В. Гаврилов // Научный вестник НГТУ. – 2008. – № 1(32). – С. 33–40.
7. Лисицин, Д.В. Устойчивое оценивание параметров модели при асимметричном засорении данных / Д.В. Лисицин, К.В. Гаврилов // Известия МАН ВШ. – 2006. – № 1(35). – С. 60–73.
8. 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
9. Shevlyakov, G. Redescending M-estimators / G. Shevlyakov, S. Morgenthaler, A. Shurygin // J. Statist. Plann. Inference. – 2008. – Vol. 138, no. 10. – P. 2906–2917. DOI: 10.1016/J.JSPI.2007.11.008
10. Asymptotically stable detection of a weak signal / G. Shevlyakov, V. Shin, S. Lee, K. Kim // Int. J. Adapt. Control Signal Process. – 2014. – Vol. 28. – P. 848–858. DOI: 10.1002/acs.2405
11. Симахин, В.А. Адаптивные оценки параметра сдвига / В.А. Симахин, О.С. Черепанов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2013. – № 1(22). – С. 131–137.
12. Windham, M.P. Robustifying model fitting / M.P. Windham // J. R. Statist. Soc. B. – 1995. – Vol. 57, no. 3. – P. 599–609. DOI: 10.1111/J.2517-6161.1995.TB02050.X

13. Eguchi, S. Entropy and divergence associated with power function and the statistical application / S. Eguchi, S. Kato // *Entropy*. – 2010. – Vol. 12. – P. 262–274. DOI: 10.3390/e12020262
14. A comparison of related density-based minimum divergence estimators / M.C. Jones, N.L. Hjort, I.R. Harris, A. Basu // *Biometrika*. – 2001. – Vol. 88, no. 3. – P. 865–873. DOI: 10.1093/biomet/88.3.865
15. Fujisawa, H. Normalized estimating equation for robust parameter estimation / H. Fujisawa // *Electronic Journal of Statistics*. – 2013. – Vol. 7. – P. 1587–1606. DOI: 10.1214/13-EJS817
16. Eguchi, S. Minimum divergence methods in statistical machine learning: from an information geometric viewpoint / S. Eguchi, O. Komori. – Springer: Japan KK. – 2020. – 221 p. DOI: 10.1007/978-4-431-56922-0
17. Lisitsin, D.V. Minimum gamma-divergence estimation for non-homogeneous data with application to ordered probit model / D.V. Lisitsin, A.G. Usol'tsev // *Applied Methods of Statistical Analysis. Statistical Computation and Simulation: Proc. of the Intern. Workshop*. – Novosibirsk: NSTU, 2019. – P. 227–234.
18. Джейнс, Э.Т. О логическом обосновании методов максимальной энтропии / Э.Т. Джейнс // *ТИИЭР*. – 1982. – Т. 70, № 9. – С. 33–51.
19. Michalowicz, J.V. Handbook of differential entropy / J.V. Michalowicz, J.M. Nichols, F. Bucholtz. – Boca Raton: CRC Press, 2014. – 226 p. DOI: 10.1201/b15991
20. 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
21. Farhadi, A. Robust coding for a class of sources: applications in control and reliable communication over limited capacity channels / A. Farhadi, C.D. Charalambous // *Systems & Control Letters*. – 2008. – Vol. 57. – P. 1005–1012. DOI: 10.1016/j.sysconle.2008.06.006
22. Principe, J.C. Information theoretic learning: Renyi's entropy and kernel perspectives / J.C. Principe. – Springer: Information Science and Statistics. – 2010. – 515 p. DOI: 10.1007/978-1-4419-1570-2
23. 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
24. 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
25. Смоляк, С.А. Устойчивые методы оценивания: статистическая обработка неоднородных совокупностей / С.А. Смоляк, Б.П. Титаренко. – М.: Статистика, 1980. – 208 с.
26. Van der Vaart, A.W. Asymptotic statistics / A.W. Van der Vaart. – Cambridge: Cambridge University Press, 1998. – 443 p. DOI: 10.1017/CBO9780511802256
27. Лисицин, Д.В. Устойчивое оценивание параметров модели по многомерным неоднородным неполным данным / Д.В. Лисицин // *Научный вестник НГТУ*. – 2013. – № 1(50). – С. 17–30.
28. Лисицин, Д.В. Максиминая задача оценивания параметров в условиях байесовского точечного засорения / Д.В. Лисицин, К.В. Гаврилов // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. – 2023. – № 1(62). – С. 56–64. DOI: 10.17223/19988605/62/6
29. 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.

30. Зельнер А. Байесовские методы в эконометрии / А. Зельнер. – М.: Статистика, 1980. – 439 с.
31. Гермейер, Ю.Б. Игры с противоположными интересами / Ю.Б. Гермейер. – М.: Наука, 1976. – 328 с.
32. Glasserman, P. Robust risk measurement and model risk / P. Glasserman, X. Xu // *Quantitative Finance*. – 2014. – Vol. 14. – P. 29–58. DOI: 10.2139/ssrn.2167765
33. Kerridge, D.F. Inaccuracy and inference / D.F. Kerridge // *Journal of Royal Statistical Society, Ser. B*. – 1961. – Vol. 23. – P. 184–194. DOI: 10.1111/j.2517-6161.1961.tb00404.x
34. Трухаев, Р.И. Модели принятия решений в условиях неопределенности / Р.И. Трухаев. – М.: Наука, 1981. – 258 с.
35. Wagner, D. Resilient aggregation in sensor networks / D. Wagner // *Proceedings of the 2nd ACM Workshop on Security of Ad hoc and Sensor Networks (SASN'04)*. – 2004. – P. 78–87. DOI: 10.1145/1029102.1029116
36. Biggio, B. Wild patterns: ten years after the rise of adversarial machine learning / B. Biggio, F. Roli // *Pattern Recognition*. – 2018. – Vol. 84. – P. 317–331. DOI: 10.1016/j.patcog.2018.07.023
37. Walley, P. Statistical reasoning with imprecise probabilities / P. Walley. – New York: Chapman and Hall, 1991. – 706 p.

References

1. Borovkov A. A. *Matematicheskaya statistika [Mathematical statistics]*. Saint Petersburg, Izdatel'stvo «Lan'», 2021, 704 p.
2. Shurygin A. M. *Matematicheskie metody prognozirovaniia [Mathematical methods of prediction]*. Moscow, Goriachaia Liniia – Telekom, 2009, 180 p.
3. Huber P. *Robastnost' v statistike [Robust statistics]*. Moscow, Mir, 1984, 303 p.
4. 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.
5. Shurygin A. M. New approach to optimization of stable estimation // *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, 1994, vol. 3, pp. 315-340. DOI: 10.1007/978-94-011-0854-6_15
6. Lisitsin D. V., Gavrilov K. V. Ob ustoychivom otsenivanii parametrov modeli pri asimmetrichnom zasoreнии dannykh [On stable estimation of models parameters in presence of asymmetric data contamination] // *Science Bulletin of the NSTU*, 2008, vol. 32, no. 1, pp. 33-40.
7. Lisitsin D. V., Gavrilov K. V. Ustoychivoe otsenivanie parametrov modeli pri asimmetrichnom zasoreнии dannykh [Stable parameter estimation of model at the asymmetric contamination of the data] // *Proceedings of IHEAS*, 2006, vol. 35, no. 1, pp. 60-73.
8. Lisitsin D. V., Gavrilov K. V. Otsenivanie parametrov raspredeleniia ogranichennoi sluchainoi velichiny, robastnoe 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
9. Shevlyakov G., Morgenthaler S., Shurygin A. Redescending M-estimators // *J. Statist. Plann. Inference*, 2008, vol. 138, no. 10, pp. 2906-2917. DOI: 10.1016/J.JSPI.2007.11.008
10. Shevlyakov G., Shin V., Lee S., Kim K. Asymptotically stable detection of a weak signal // *Int. J. Adapt. Control Signal Process*, 2014, vol. 28, pp. 848-858. DOI: 10.1002/acs.2405
11. Simakhin V. A., Cherepanov O. S. Adaptivnye otsenki parametra sdviga [Adaptive estimates of location parameter] // *Tomsk State University Journal of Control and Computer Science*, 2013, vol. 22, no. 1, pp. 131-137.

12. Windham M. P. Robustifying model fitting // *J. R. Statist. Soc. B*, 1995, vol. 57, no. 3, pp. 599-609. DOI: 10.1111/J.2517-6161.1995.TB02050.X
13. Eguchi S., Kato S. Entropy and divergence associated with power function and the statistical application // *Entropy*, 2010, vol. 12, pp. 262-274. DOI: 10.3390/e12020262
14. Jones M. C., Hjort N. L., Harris I. R., Basu A. A comparison of related density-based minimum divergence estimators // *Biometrika*, 2001, vol. 88, no. 3, pp. 865-873. DOI: 10.1093/biomet/88.3.865
15. Fujisawa H. Normalized estimating equation for robust parameter estimation // *Electronic Journal of Statistics*, 2013, vol. 7, pp. 1587-1606. DOI: 10.1214/13-EJS817
16. Eguchi S., Komori O. Minimum divergence methods in statistical machine learning: from an information geometric viewpoint. Springer, Japan KK, 2020, 221 p. DOI: 10.1007/978-4-431-56922-0
17. Lisitsin D. V., Usol'tsev A. G. Minimum gamma-divergence estimation for non-homogeneous data with application to ordered probit model // *Proc. of the Intern. Workshop on AMSA'2019, Novosibirsk, NSTU, 2019*, pp. 227-234.
18. Jaynes E. T. O logicheskoy obosnovanii metodov maksimal'noi entropii [On the rationale of maximum-entropy methods] // *Proceedings of the Institute of Electrical and Radio Electronics Engineers*, 1982, vol. 70, no. 9, pp. 33-51.
19. Michalowicz J. V., Nichols J. M., Bucholtz F. Handbook of differential entropy. Boca Raton, CRC Press, 2014, 226 p. DOI: 10.1201/b15991
20. 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
21. Farhadi A., Charalambous C. D. Robust coding for a class of sources: applications in control and reliable communication over limited capacity channels // *Systems & Control Letters*, 2008, vol. 57, pp. 1005-1012. DOI: 10.1016/j.sysconle.2008.06.006
22. 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
23. Tsallis C. Introduction to nonextensive statistical mechanics: approaching a complex world. New York, Springer, 2009, 382 p. DOI: 10.1007/978-0-387-85359-8
24. 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
25. Smolyak S. A., Titarenko B. P. Ustoichivye metody otsenivaniia: statisticheskaya obrabotka neodnorodnykh sovokupnostei [Stable estimation methods: inhomogeneous-set processing]. Moscow, Statistika, 1980, 208 p.
26. Van der Vaart A. W. Asymptotic statistics. Cambridge, Cambridge University Press, 1998, 443 p.
27. 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.
28. Lisitsin D. V., Gavrilov K. V. Maksiminnaia zadacha otsenivaniia parametrov v usloviakh baiesovskogo tochechnogo zasoreniia [Maximin problem of parameter estimation in conditions of point Bayesian contamination] // *Tomsk State University Journal of Control and Computer Science*, 2023, vol. 62, no. 1, pp. 56-64. DOI: 10.17223/19988605/62/6

29. 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.
30. Zellner A. Baiesovskie metody v ekonometrii [Bayesian analysis in econometrics]. Moscow, Statistika, 1980, 439 p.
31. Germeier Yu. B. Iгры s neprotivopolozhnymi interesami [Non-antagonistic games]. Moscow, Nauka, 1976, 328 p.
32. Glasserman P., Xu X. Robust risk measurement and model risk // *Quantitative Finance*, 2014, vol. 14, pp. 29-58. DOI: 10.2139/ssrn.2167765
33. Kerridge D. F. Inaccuracy and inference // *Journal of Royal Statistical Society, Ser. B*, 1961, vol. 23, pp. 184-194. DOI: 10.1111/j.2517-6161.1961.tb00404.x
34. Trukhaev R. I. Modeli priniatiia reshenii v usloviakh neopredelennosti [Models of decision-making in conditions of uncertainty]. Moscow, Nauka, 1981, 258 p.
35. Wagner D. Resilient aggregation in sensor networks // *Proceedings of the 2nd ACM Workshop on Security of Ad hoc and Sensor Networks (SASN'04)*, 2004, pp. 78-87. DOI: 10.1145/1029102.1029116
36. Biggio B., Roli F. Wild patterns: ten years after the rise of adversarial machine learning // *Pattern Recognition*, 2018, vol. 84, pp. 317-331. DOI: 10.1016/j.patcog.2018.07.023
37. Walley P. Statistical reasoning with imprecise probabilities. New York, Chapman and Hall, 1991, 706 p.