

УДК 519.23: 528.854

С. К. АБРАМОВ

*Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина*

## УСТОЙЧИВОЕ ОЦЕНИВАНИЕ ПАРАМЕТРА СДВИГА РАСПРЕДЕЛЕНИЙ В ЗАДАЧАХ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

*Рассмотрена задача определения параметра сдвига закона распределения выборок данных, характерных для задачи автоматической обработки изображений. Показано, что закон распределения таких выборок имеет ярко выраженный несимметричный характер, обусловленный присутствием тяжелого одностороннего «хвоста». Применение известных робастных оценок параметра сдвига в таких условиях не обеспечивает требуемой точности. В связи с этим предложено несколько новых подходов, основанных на использовании мириадной и квантильных оценок. Показано, что предложенные решения обеспечивают высокую точность и робастность. Даны рекомендации по их применению на практике.*

**Ключевые слова:** устойчивое оценивание, параметр сдвига, автоматическая обработка изображений.

### Введение

Современные системы аэрокосмического дистанционного зондирования (ДЗ) и обзора окружающего пространства позволяют решать важные практические задачи: контроль состояния сельскохозяйственных культур, прогнозирование урожайности, проведение природоохранного мониторинга, анализ последствий чрезвычайных ситуаций и пр. [1].

Особенностью данных, предоставляемых современными системами ДЗ, является их многоканальность (одновременная регистрация сигнала в нескольких спектральных диапазонах электромагнитного излучения), высокое пространственное разрешение и большая площадь зоны обзора. В совокупности это приводит к тому, что объем, занимаемый одним снимком, может составлять от сотен мегабайт до единиц гигабайт, что существенно затрудняет передачу таких снимков по каналам связи, их обработку и хранение [1].

Еще одной особенностью таких данных является присутствие в них помех. И если уровень шумов аппаратуры благодаря усовершенствованию технологии удалось существенно снизить, то помехи, обусловленные самим принципом формирования информационного сигнала, все равно имеют место и могут быть достаточно интенсивными. Игнорирование этого факта может привести к существенному снижению эффективности извлечения полезной информации из полученных данных [1]. Поэтому целесообразным является проведение предварительной обработки с целью определения характеристик помех и их последующего устранения [2], а также некоторых характеристик самих изображений. При этом такую предобработку в силу большого объема

данных необходимо проводить автоматически [1].

В задачах предварительного анализа характеристик, искажающих изображения помех и их фильтрации, довольно часто приходится иметь дело со статистическим анализом распределений выборок данных [1]. Наибольший интерес, как правило, представляет определение параметра сдвига или связанной с ним моды распределения. При этом сами распределения выборок часто характеризуются несимметричностью и наличием длинного «хвоста», что приводит к невозможности применения оценки среднего для определения параметра сдвига в силу ее существенной смещенности. В результате решение задачи возможно только с применением устойчивых (робастных) оценок параметра сдвига [3].

Задача определения параметра сдвига часто возникает и в других приложениях обработки изображений. Например, это имеет место при фильтрации в скользящем окне [4], определении статистических характеристик информационной составляющей изображений (контрастности, наличия фона и его характеристик и пр.) [1].

*Целью данной работы* является анализ применимости существующих устойчивых оценок параметра сдвига и разработка новых эффективных оценок применительно к распределениям выборок данных, характерных для задач автоматической обработки изображений, формируемых системами ДЗ аэрокосмического базирования и комплексами контроля окружающего пространства, и других приложений.

### Постановка задачи

Для определения эффективности различных оценок параметра сдвига проводился статистиче-

ский эксперимент с использованием численного моделирования. Для описания распределения выборки данных, размер которой составлял  $N$  элементов, использовалась известная модель загрязненного (составного) распределения [3]

$$\rho(x) = p \cdot \rho_{\text{норм}}(x - m, \sigma_m^2) + q \cdot \rho_{\text{аном}}(x), \quad (1)$$

где  $p$  – вероятность появления «нормальных» элементов выборки, т.е. значений, подчиняющихся основному закону распределения  $\rho_{\text{норм}}$ ;

$q = 1 - p$  – вероятность появления «аномальных» элементов выборки, т.е. значений, подчиняющихся загрязняющему закону распределения  $\rho_{\text{аном}}$ ;

$m$ ,  $\sigma_m^2$  – соответственно параметры сдвига и масштаба (дисперсия) основного распределения.

Полагалось, что плотность распределения вероятности (ПРВ)  $\rho_{\text{норм}}$  является унимодальной и симметричной относительно моды, которая совпадает с  $m$ , а ПРВ загрязняющего распределения –  $\rho_{\text{аном}}$ , наоборот, несимметричной, что делало несимметричной всю функцию  $\rho(x)$ , поскольку такой случай, как указано выше, является типичным для задач автоматической обработки изображений.

Параметры модели (1) выбирались на основе анализа статистических характеристик реальных выборок, характерных для рассматриваемых задач [1, 5]. В качестве основного распределения  $\rho_{\text{норм}}$  использовалось Гауссово. Для задач фильтрации это соответствует наиболее распространенному случаю присутствия на изображении аддитивного Гауссова шума, а для задач оценивания характеристик помех – проведению оценивания в блоках большого размера (тогда закон распределения оценок дисперсии Хи-квадрат можно аппроксимировать Гауссовым).

Параметры  $m$  и  $\sigma_m^2$  в зависимости от решаемой задачи могут меняться в широких пределах. Для задачи фильтрации величина  $m$  ограничена только динамическим диапазоном представления изображений (обычно от 0 до 255), а величина  $\sigma_m^2$  – диапазоном реальных дисперсий помех (обычно от 5 до 400), а сами параметры основного распределения являются независимыми. Для задачи оценивания дисперсии, наоборот, между этими параметрами есть взаимосвязь  $\sigma_m^2 \sim m^2$ , а параметр  $m$  соответствует величине оцениваемой дисперсии (т.е. лежит в диапазоне от 5 до 400).

В качестве  $\rho_{\text{аном}}$  использовалось равномерное распределение в диапазоне  $[0, M]$ , где для обеспечения несимметричности  $M \gg m$ . Нулевая нижняя граница обусловлена положительной определенно-

стью элементов выборок для обеих задач, как фильтрации, так и оценивания характеристик помех.

Параметры  $p$  и  $q$  характеризуют доли элементов выборок, соответствующих основному и загрязняющему распределениям. На практике величина параметра  $p$  определяется процентным содержанием однородных (не содержащих границ, объектов и текстуры) участков изображения и может меняться в широких пределах: от 0,9 (практически однородное изображение) до 0,1 (почти полностью текстурное).

Для проведения всестороннего анализа параметры модели (1) задавались в соответствии с семью вариантами, описание которых дано в табл. 1. Для получения статистически достоверных результатов оценивание проводилось по 10 000 реализаций, размер выборки составлял  $N = 1000$ .

Таблица 1

Наборы параметров моделирования

Параметры	$p$	$m$	$\sigma_m^2$	$M/m$
Вариант 1	0,9	100	800	100
Вариант 2	0,8	300	7200	100
Вариант 3	0,65	100	800	100
Вариант 4	0,5	200	3200	100
Вариант 5	0,15	300	7200	100
Вариант 6	0,1	400	12800	100
Вариант 7	0,65	400	12800	10

В качестве численных критериев эффективности применялись: смещенность  $\Delta = \langle \hat{m} \rangle - m$ , дисперсия  $\theta^2 = \langle (\hat{m} - \langle \hat{m} \rangle)^2 \rangle$ , приведенная  $\varepsilon = \Delta^2 + \theta^2$  и относительная  $\varepsilon\% = (\sqrt{\varepsilon} / m) \cdot 100\%$  приведенная ошибка оценивания параметра сдвига  $m$ .

Основными требованиями, предъявляемыми к оценке параметра сдвига, являются: несмещенность ( $\Delta \rightarrow 0$ ) и минимальная дисперсия оценивания ( $\theta^2 \rightarrow 0$ ), что в совокупности должно обеспечивать и минимальную приведенную ошибку ( $\varepsilon \rightarrow 0$ ). С практической точки зрения (в частности, для задач оценивания характеристик помех), приемлемой можно считать точность  $\varepsilon\% \leq 15...20\%$  [6].

### Рассматриваемые оценки параметра сдвига

Вначале рассмотрим две устойчивые оценки параметра сдвига, позволяющие получать приемлемые результаты для рассматриваемой модели (1): квантильную и мириадную.

Квантильная оценка  $\hat{Q}_n$  принадлежит к классу устойчивых  $L$ -оценок [3] с единственным ненулевым коэффициентом. Фактически, эта оценка явля-

Таблица 2

Численные критерии точности оценок параметра сдвига для различных анализируемых вариантов

Вариант	Квантильная оценка ( $\hat{Q}_{30\%}$ )				Мириадная оценка (при оптимальном $k/\sigma_{m\text{opt}}$ )				Квантильно-мириадная оценка ( $k=0,3 \hat{Q}_{30\%}$ )			
	$\Delta_Q$	$\theta_Q^2$	$\varepsilon_Q$	$\varepsilon_Q^{\%}$	$\Delta_M$	$\theta_M^2$	$\varepsilon_M$	$\varepsilon_M^{\%}$	$\Delta_{QM}$	$\theta_{QM}^2$	$\varepsilon_{QM}$	$\varepsilon_{QM}^{\%}$
1	-12,34	1,56	153,73	12,40	0,05	8,78	8,78	2,96	0,18	1,20	153,73	12,40
2	-27,70	16,91	784,35	9,34	0,65	129,09	129,51	3,79	0,65	12,08	784,35	9,34
3	-3,08	2,60	12,10	3,48	0,13	12,62	12,64	3,55	0,57	1,74	12,10	3,48
4	13,18	17,85	191,47	6,92	0,63	77,63	78,03	4,42	2,14	8,36	191,47	6,92
5	5021	$3 \times 10^5$	$3 \times 10^7$	1682	6,88	465,91	513,21	7,55	$1 \times 10^4$	$4 \times 10^6$	$3 \times 10^7$	1682
6	8534	$4 \times 10^5$	$7 \times 10^7$	2139	10539	$7 \times 10^7$	$18 \times 10^7$	3426	$2 \times 10^4$	$3 \times 10^6$	$7 \times 10^7$	2139
7	-25,54	39,46	691,70	6,58	4,32	212,49	231,11	3,80	10,69	31,51	691,70	6,58

ется аналогом медианы с возможностью выбора произвольного номера квантиля, отличного от  $p_{\%}=50\%$ , что позволяет распространить свойство медианы определять моду симметричных распределений и на случай несимметричных ПРВ. Для обеспечения устойчивости квантильной оценки к присутствию тяжелого правостороннего «хвоста» (до  $p \approx 0,6$ ) и не слишком заниженных значений в случае больших  $p$ ,  $p_{\%}$  был установлен равным 30%.

Мириадная оценка является устойчивой оценкой максимального правдоподобия (М-оценкой [3]) для класса  $\alpha$ -стабильных распределений [7]. Ее основной особенностью является наличие настроечного параметра  $k$ , при определенном выборе которого мириадная оценка соответствует моде распределения выборки данных. Для устранения масштабной неинвариантности мириадной оценки [7] исследование ее свойств проводилось в зависимости от  $k/\sigma_m$ .

Численные критерии точности квантильной (с индексом Q) и мириадной (с индексом M) оценок для различных анализируемых вариантов приведены в табл. 2. Данные по мириадной оценке приведены только для оптимальных (обеспечивающих наименьшие величины приведенной ошибки) значений настроечного параметра  $k/\sigma_{m\text{opt}}$ .

Как видно из табл. 2, квантильная оценка с фиксированным  $p_{\%}=30\%$  практически для всех ситуаций дает существенно смещенные оценки. Смещенность  $\Delta_Q$  может быть как отрицательной (при  $p < 0,6$ ), так и положительной (при  $p \geq 0,6$ ). Особенно большие смещенности наблюдаются при крайне малых значениях  $p$ , т.е. в ситуациях почти полного отсутствия «нормальных» элементов выборки.

Мириадная оценка с оптимальным параметром  $k/\sigma_{m\text{opt}}$  оказывается более точной. Она обеспечивает малые смещенности  $\Delta_M$  практически для всех рассматриваемых вариантов. Исключение составляет вариант 6, для которого мириадная оценка оказы-

вается неспособной обеспечить приемлемые результаты из-за недостаточной робастности к столь тяжелому «хвосту» распределения ( $q=0,9$ ). По дисперсии оценивания  $\theta_M^2$  и, соответственно, приведенной ошибке  $\varepsilon_M$  мириадная оценка при  $k/\sigma_{m\text{opt}}$  обеспечивает существенно лучшие показатели по сравнению с квантильной при фиксированном  $p_{\%}=30\%$ .

Основной проблемой применения мириадной оценки при известном  $k/\sigma_{m\text{opt}}$  является необходимость определения параметра  $\sigma_m$ , используемого при расчете  $k_{\text{opt}}=\sigma_m \cdot k/\sigma_{m\text{opt}}$ . Сложность определения масштаба распределения  $\sigma_m$  связана с отсутствием достаточно устойчивых оценок для рассматриваемых несимметричных распределений с тяжелым односторонним «хвостом». Для решения этой проблемы воспользуемся упомянутой выше зависимостью между параметрами  $\sigma_m^2 \sim m^2$  и определим  $k_{\text{opt}}$  пропорционально предварительной квантильной оценке  $k=0,3\hat{Q}_{30\%}$ . Такая оценка получила название «квантильно-мириадной» и ее критерии точности приведены в правой части табл. 2.

Для этой оценки приемлемые результаты наблюдаются для пяти из семи рассматриваемых вариантов. Для вариантов 5 и 6 устойчивости квантильно-мириадной оценки оказывается недостаточно, что для варианта 5 связано с существенной смещенностью предварительной оценки  $\hat{Q}_{30\%}$ , а для варианта 6 – со смещенностью обеих оценок.

Отсюда вытекают следующие предварительные выводы. Во-первых, использование квантильной оценки с фиксированным  $p_{\%}$  не позволяет обеспечить приемлемую точность оценивания параметра сдвига в широком диапазоне варьирования параметров модели (1). Во-вторых, характеристики комбинированной квантильно-мириадной оценки можно

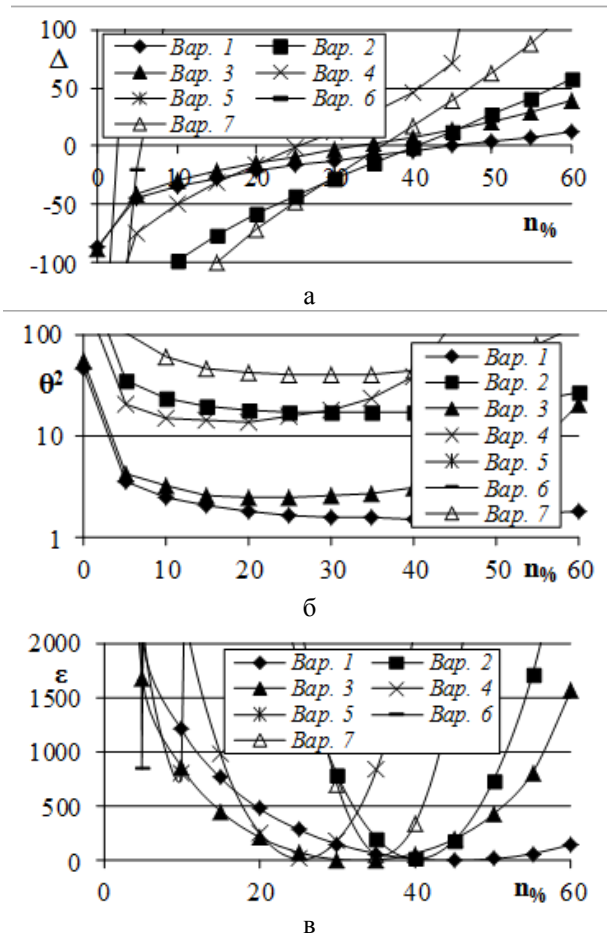


Рис. 1. Характеристики точности квантильной оценки в зависимости от номера квантиля

считать приемлемыми только при  $p \geq 0,15$ , что существенно сужает область ее использования.

Для выявления путей устранения указанных недостатков квантильной оценки проведем более детальный анализ ее свойств. Для этого представим ее смещенность, дисперсию и приведенную ошибку как функции от номера квантиля  $n\%$  (рис. 1).

Как видно из рис. 1, а, для всех рассматриваемых вариантов смещенность меняет свой знак с отрицательного (при малых  $n\%$ ) на положительный (при больших  $n\%$ ) и при этом существует такое  $n\% = n\%_{\Delta \min}$ , отличное для различных рассматриваемых вариантов, для которого  $\Delta_Q \approx 0$ . Номер квантиля  $n\%_{\Delta \min}$  существенно зависит от  $p$  и  $n\%_{\Delta \min} \approx 100\% \cdot p/2$ . Кроме того, кривая дисперсии оценивания для каждого рассматриваемого варианта (см. рис. 1, б) также имеет минимум в своей, отличной от других, точке  $n\% = n\%_{\theta^2 \min}$ , определяемой параметрами модели (1). Следует отметить, что  $n\%_{\Delta \min} \approx n\%_{\theta^2 \min}$ , т.е. квантильная оценка, математическое ожидание которой совпадает с  $m$ , одно-

временно обладает и наименьшей дисперсией.

В результате кривые приведенной ошибки (см. рис. 1, в) имеют глобальные минимумы в точках  $n\%_{opt} \approx n\%_{\Delta \min} \approx n\%_{\theta^2 \min}$ . При этом значения  $n\%_{opt}$  для разных вариантов могут существенно отличаться от  $n\% = 30\%$ , используемого в рассмотренной ранее оценке  $\hat{Q}_{30\%}$ .

Основной проблемой, затрудняющей использование выявленных закономерностей на практике, является существенная зависимость  $n\%_{opt}$  от параметров распределения выборки, которые неизвестны априорно, а их оценивание является весьма сложной задачей. В связи с этим воспользуемся другим подходом на основе бутстреппирования [8], основной идеей которого является формирование набора псевдовыборок данных из исходной выборки с сохранением ее закона распределения.

Тогда, применяя к сформированному набору псевдовыборок квантильные оценки с различными  $n\%$ , можно найти среди них такую с  $n\%_{\theta^2 \min}$ , которая обеспечит наименьшую дисперсию оценивания, а, следовательно, и наименьшие смещенность и приведенную ошибку.

Для получения предлагаемой бутстреп оценки из исходной выборки размером 1000 элементов формировалось 1000 псевдовыборок, каждая из которых состояла из 100 элементов. Такой выбор был сделан в соответствии с рекомендациями, приведенными в основополагающей работе [8]. В качестве оценок использовалось 100 квантилей с номерами от 1% до 100% с шагом в 1%.

Проанализируем статистические характеристики предлагаемой бутстреп оценки в зависимости от изменения параметра  $p$  модели (1) при фиксированных  $m=10$ ;  $\sigma_m^2=8$  и  $M/m=100$ . Поскольку именно от параметра  $p$  в наибольшей степени зависит номер наиболее точной квантильной оценки. Помимо использованных ранее численных критериев  $\Delta_B$ ,  $\theta_B^2$ ,  $\epsilon_B$  и  $\epsilon_B^{\%}$  (индекс B соответствует бутстреп оценке), также определялись среднее  $\bar{n}\% = \langle \hat{n}\% \rangle$  и дисперсия  $\theta_{n\%}^2 = \langle (\hat{n}\% - \langle \hat{n}\% \rangle)^2 \rangle$  номера квантиля  $\hat{n}\%$  оценки, выбираемой в качестве итоговой. Для сравнения приведем также результаты для мириадной оценки, как наиболее точной (при оптимальном  $k/\sigma_m$ ) из рассмотренных выше. Результаты моделирования представлены в табл. 3.

Анализ этих данных показывает, что абсолютные значения смещенности бутстреп оценки увеличиваются с уменьшением  $p$ , а ее знак всегда остается

Таблица 3

Численные критерии точности оценок для различных значений параметра  $p$  модели (1)

$p$	Бутстреп оценка						Мириадная оценка				
	$\bar{n}_{\%}$	$\theta_{n\%}^2$	$\Delta_B$	$\theta_B^2$	$\varepsilon_B$	$\varepsilon_B^{\%}$	$k/\sigma_{m\text{opt}}$	$\Delta_M$	$\theta_M^2$	$\varepsilon_M$	$\varepsilon_M^{\%}$
0,75	26,94	75,51	-1,04	0,91	1,99	14,12	1,50	0,048	0,012	0,014	1,19
0,65	21,69	52,13	-1,25	0,90	2,45	15,64	1,25	0,063	0,016	0,020	1,42
0,50	15,39	27,82	-1,46	0,87	3,00	17,32	0,90	0,085	0,023	0,030	1,73
0,25	6,82	5,51	-1,75	0,84	3,91	19,77	0,47	0,213	0,064	0,109	3,30
0,15	3,67	1,53	-2,02	0,74	4,82	21,96	0,27	0,322	0,143	0,247	4,97
0,10	1,90	0,73	-2,61	1,08	7,89	28,09	0,06	269,63	47232	119932	3463

ся отрицательным. Это означает, что даже при очень малых  $p$ , когда «хвост» распределения выборки оказывается крайне тяжелым, это не приводит к «утягиванию» оценки в его сторону, как это происходит с большинством других оценок, в частности, мириадной, для которой смещенность всегда остается положительной. Приведенная и относительная приведенная ошибки бутстреп оценки при  $p \geq 0,15$  оказываются существенно большими, чем для мириадной оценки с оптимальным  $k/\sigma_{m\text{opt}}$ . Однако при  $p \leq 0,1$  мириадная оценка теряет работоспособность, в то время как точность бутстреп оценки остается вполне приемлемой.

Следует отметить, что оптимальное значение  $k/\sigma_{m\text{opt}}$  для мириадной оценки существенно зависит от  $p$ , что демонстрирует рис. 2, а. Что касается бутстреп оценки, то для нее также четко видна тенденция изменения значений  $\bar{n}_{\%}$  и их дисперсия  $\theta_{n\%}^2$  от величины параметра  $p$  (см. рис. 2, а).

Другими словами, номер квантиля итоговой оценки, выбранной в результате бутстреппирования, может служить индикатором веса «хвоста» распределения (характеризуемого параметром  $p$ ). А с уче-

том четкой взаимосвязи параметров  $k/\sigma_{m\text{opt}}$  и  $\Delta_B$  от номера индекса  $\bar{n}_{\%}$  (представленной в виде графиков на рис. 2, б) ее можно достаточно хорошо аппроксимировать линейными функциями. Тогда полученный бутстреппированием номер квантиля  $\bar{n}_{\%}$  может быть использован для определения оценки параметра  $p$  и, следовательно, зависящих от него смещенности  $\Delta_B$  и оптимального  $k/\sigma_{m\text{opt}}$ .

Однако приведенные на рис. 2, а и б графики отражают зависимость  $\Delta_B$  и  $k/\sigma_{m\text{opt}}$  лишь от одного параметра  $p$  модели (1). Проведем также исследование влияния параметров  $m$  и  $\sigma_m^2$ , зафиксировав остальные параметры:  $p=0,75$ ;  $M/m=100$ . Результаты моделирования приведены в табл. 4.

Как видно из табл. 4, смещенность бутстреп оценки существенно зависит от истинного значения оцениваемого параметра  $m$ , а также практически линейно зависит от среднего значения бутстреп оценки  $\langle \hat{m}_B \rangle$ , что может быть использовано для ее коррекции.

С применением метода наименьших квадратов были получены следующие аппроксимирующие выражения. Для определения скорректированной (с учетом смещенности) бутстреп оценки –  $\hat{m}_B^{\text{corr}} = 1,115\hat{m}_B - 0,035\hat{n}_{\%} + 0,931$ , для определения оптимального параметра мириадной оценки –  $(k/\sigma_m)_{q\text{-opt}} = 0,053\hat{n}_{\%} - 0,030\sqrt{N} + 1,042$  (справедливо для  $p \geq 0,15$ ). Для мириадной оценки также учтена зависимость от размера выборки  $N$ , полученная в результате дополнительных исследований.

Было исследовано влияние на характеристики оценок и параметра  $M/m$ . Выяснилось, что на бутстреп оценку он влияет слабо, а на мириадную – довольно существенно. Однако учесть это влияние на практике не представляется возможным ввиду отсутствия адекватных оценок параметра  $M/m$ .

С учетом указанных выше свойств бутстреп и мириадной оценок предлагается следующая комби-

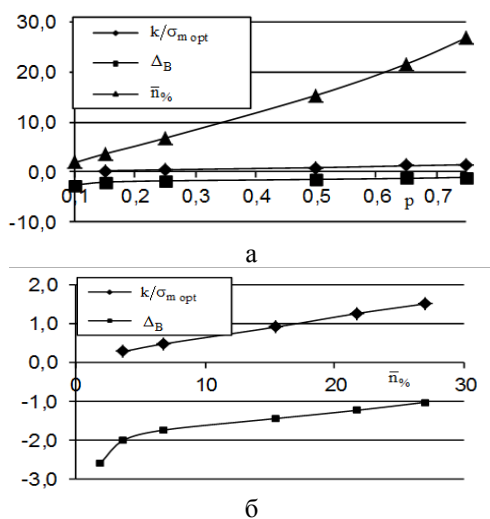


Рис. 2. Зависимости  $k/\sigma_{m\text{opt}}$ ,  $\bar{n}_{\%}$  и  $\Delta_B$  от параметра  $p$  (а), и их взаимная зависимость (б)

Таблица 4

Численные критерии точности оценок для различных значений параметров  $m$  и  $\sigma_m^2$  модели (1)

m	$\sigma_m^2$	Бутстреп оценка					Мириадная оценка			
		$\bar{n}_\%$	$\theta_{n_\%}^2$	$\Delta_B$	$\theta_B^2$	$\varepsilon_B$	$k/\sigma_{m\text{opt}}$	$\Delta_M$	$\theta_M^2$	$\varepsilon_M$
10	8	26,94	75,51	-1,04	0,91	1,99	1,50	0,048	0,012	0,014
100	800	27,02	74,77	-10,29	90,37	196,3	1,31	0,400	1,322	1,482
200	3200	27,09	74,49	-20,50	361,3	781,5	1,35	0,939	4,930	5,812
300	7200	26,89	75,85	-31,42	822,8	1810	1,15	1,530	12,13	14,47
400	12800	27,05	75,62	-41,16	1461	3155	1,37	1,812	20,56	23,84

нированная бутстреп-мириадная оценка, определяемая в несколько этапов.

1. Получение предварительной бутстреп оценки  $\hat{m}_B$  с одновременным определением  $\hat{n}_\%$ .

2. Проведение коррекции бутстреп оценки согласно  $\hat{m}_B^{\text{corr}} = 1,115\hat{m}_B - 0,035\hat{n}_\% + 0,931$ .

3. Если  $\hat{n}_\% < 8$  использовать  $\hat{m}_B^{\text{corr}}$  в качестве итоговой оценки.

4. В противном случае использовать в качестве итоговой мириадную оценку, полученную при  $k = (k/\sigma_m)_{q\text{-opt}} \cdot \hat{\sigma}_m$ , где  $(k/\sigma_m)_{q\text{-opt}} = 0,053\hat{n}_\% - 0,030\sqrt{N} + 1,042$ , а  $\hat{\sigma}_m = \hat{m}_B^{\text{corr}} \sqrt{2/N_{\text{бл}}}$ ,  $N_{\text{бл}}$  – размер блоков, в которых производятся оценивания дисперсии помех.

Результаты моделирования для скорректированной бутстреп (индекс Вс) и комбинированной бутстреп-мириадной (индекс ВМ) оценок для разных вариантов приведены в табл. 5.

Как видно из табл. 5, скорректированная бутстреп оценка имеет некоторую смещенность, которая, тем не менее, существенно меньше смещенности первичной бутстреп оценки (см. табл. 4). Эта смещенность сильно зависит от истинного значения оцениваемого параметра  $m$ . Величина  $|\Delta/m|$  обычно меньше 0,05 и только для наиболее сложных вариантов 5 и 6 может достигать 0,11...0,17. Дисперсия скорректированной бутстреп оценки оказывается существенно выше, чем у мириадной с оптимальным  $k/\sigma_{m\text{opt}}$  (см. табл. 2). Основной причиной является сам принцип бутстреппирования, когда оценивание проводится по псевдовыборкам существен-

но меньшего размера (100 элементов), чем исходная выборка (1000 элементов). Несмотря на это, результирующая точность, характеризуемая  $\varepsilon_{Вс}^\%$ , оказывается удовлетворительной для практики [6].

В то же время, предложенная комбинированная бутстреп-мириадная оценка обеспечивает существенно лучшую точность (меньшие  $|\Delta_M|$ ,  $\theta_M^2$  и  $\varepsilon_M$ ,  $\varepsilon_M^\%$ ) по сравнению со скорректированной бутстреп оценкой (за исключением варианта 6, где мириадная оценка не применима и итоговой выступает скорректированная бутстреп оценка).

Таким образом, среди рассмотренных выше оценок параметра сдвига комбинированная бутстреп-мириадная оценка оказывается наиболее точной. Однако это справедливо только для ситуации, когда величина настроечного параметра  $k$  мириадной оценки определена правильно. В описанной выше процедуре параметр  $k$  определяется на основе оцененного заранее квази-оптимального значения  $(k/\sigma_m)_{q\text{-opt}}$  и оценки параметра  $\hat{\sigma}_m$ , основанной

на пропорциональности параметров  $m$  и  $\sigma_m^2$ . Однако такая пропорциональность имеет место только для выборок, получаемых в задачах оценки дисперсии помех [5]. Для других задач эти параметры оказываются независимыми и для определения параметра  $k$  необходимо проводить оценивания непосредственно параметра  $\sigma_m$ , что, учитывая несимметричность и тяжесть хвоста рассматриваемых распределений, является весьма сложной задачей.

В связи с этим применение бутстреп-мириадной оценки оказывается несколько ограни-

Таблица 5

Численные критерии точности оценок для различных вариантов параметров модели (1)

Вариант	Скорректированная бутстреп оценка				Бутстреп-мириадная оценка			
	$\Delta_{Вс}$	$\theta_{Вс}^2$	$\varepsilon_{Вс}$	$\varepsilon_{Вс}^\%$	$\Delta_{ВМ}$	$\theta_{ВМ}^2$	$\varepsilon_{ВМ}$	$\varepsilon_{ВМ}^\%$
1	4,32	119,8	138,5	11,77	0,30	1,02	1,11	1,05
2	3,84	1016	1030	10,70	1,36	11,53	13,38	1,22
3	-2,38	102,5	108,1	10,40	0,69	1,82	2,30	1,52
4	-9,24	441,0	526,3	11,47	1,95	11,91	15,73	1,98
5	-31,79	802,8	1813	14,19	9,34	177,9	265,2	5,43
6	-68,70	2021	6742	20,53	-68,70	2021	6742	20,53
7	1,43	1581	1583	9,95	16,8	74,57	358,4	3,75

ченным и необходимо искать альтернативные пути определения параметра сдвига. Возьмем за основу квантильное оценивание, как потенциально наиболее перспективное решение.

Предположим, что из имеющегося набора элементов выборки  $x_1, 1=1, \dots, N$  была сформирована выборка, элементы которой отсортированы по возрастанию  $X^{(t)}, t=1, \dots, N$ , где  $X^{(t)}$  обозначает  $t$ -ю порядковую статистику ( $t$ -й порядковый элемент в отсортированной выборке). Такую упорядоченную выборку можно использовать для получения аппроксимации  $\hat{P}(x)$  закона распределения

$$P(x) = \int_{-\infty}^x \rho(y) dy, \text{ положив начальное значение}$$

$\hat{P}(x) = 0$  при  $x = -\infty$  и добавляя шаг  $1/N$  по оси амплитуд для каждого  $x = X^{(t)}, t=1, \dots, N$ .

Таким образом, функция  $\hat{P}(x)$  будет расти наиболее быстро там, где порядковые статистики расположены наиболее плотно. По аналогии, используя упорядоченную выборку  $X^{(t)}, t=1, \dots, N$ , можно получить аппроксимацию функции плотности распределения вероятности  $\hat{\rho}(x)$ . Для этого разобьем интервал  $x \in [X^{(1)}; X^{(N_b)}]$  на подынтервалы фиксированной ширины  $\Delta x$ . Тогда  $\hat{\rho}(x) = (O(x) - O(x - \Delta x)) / \Delta x N$ , где  $O(x)$  – число порядковых статистик меньших  $x$ .

Моде распределения будет соответствовать такой подынтервал, для которого  $\hat{\rho}(x)$  будет наибольшим или, другими словами, где  $\Delta x N / (O(x) - O(x - \Delta x))$  будет наименьшим. Для случая фиксированных  $\Delta x$  и  $N$  это будет эквивалентно тому, что в подынтервале окажется наибольшее число порядковых статистик. Следовательно, можно ожидать, что расстояние между соседними порядковыми статистиками  $X^{(t+s)} - X^{(t)}$  (где  $s$  некоторое целое) в пределах такого интервала будет наименьшим. Тогда найдя такое  $t_{\text{est}} \in [1, N-s]$ , для которого  $X^{(t+s)} - X^{(t)}$  будет наименьшим, можно определить окрестность максимума функции  $\rho(x)$ , т.е. моду распределения.

Возникают два практических вопроса. Во-первых, как определить точную оценку положения моды  $\hat{X}$ , зная интервал  $[X^{(t_{\text{est}})}; X^{(t_{\text{est}}+s)}]$ ? Во-вторых, каким должен быть оптимальный (или, по крайней мере, рациональный) выбор  $s$ ?

Наиболее простым решением первого вопроса является выбор в качестве оценки моды

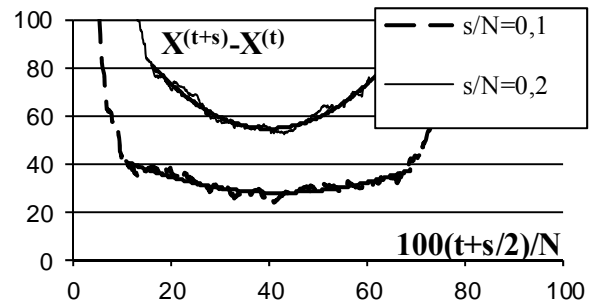


Рис. 3. Функция  $\delta(t+s/2) = X^{(t+s)} - X^{(t)}$  от нормированного аргумента  $100(t+s/2)/N$  при разных  $s/N$  и ее аппроксимация

$\hat{X} = X^{(t_{\text{est}}+s/2)}$  ( $s$  должно быть четным целым). Ответ на второй вопрос требует проведения дополнительных исследований. Вначале проверим работоспособность предложенного метода на выборке с параметрами, соответствующими варианту 2 (см. табл. 1). Рассмотрим функцию  $\delta(t+s/2) = X^{(t+s)} - X^{(t)}$ , представленную на рис. 3, при двух нормированных значениях параметра  $s$ :  $s/N = 0,1$  и  $s/N = 0,2$ . Для удобства восприятия по горизонтальной оси будем также откладывать нормированные значения аргумента  $100(t+s/2)/N$ .

Как видно из рис. 3, для обоих значений  $s/N$  функции  $\delta(t+s/2)$  имеют глобальные минимумы при аргументе  $\sim 40$ , т.е. при  $(t+s/2)/N \approx 0,4 \approx p/2$ . А как следует из рис. 1, в этому квантилю для варианта 2, как раз, и соответствует наименьшая приведенная ошибка  $\varepsilon$ . Таким образом, несмотря на «шумный» характер функции  $\delta(t+s/2)$ , она позволяет достаточно точно определить какой квантиль следует выбрать в качестве итоговой оценки параметра сдвига. Из сравнения графиков на рис. 3 следуют еще два вывода. Во-первых, для  $s/N = 0,2$  соответствующие значения функции  $\delta(t+s/2)$  примерно в два раза больше, чем для  $s/N = 0,1$ . Во-вторых, кривая для  $s/N = 0,2$  менее «шумная», чем для  $s/N = 0,1$ . Это косвенно свидетельствует о том, что точность оценивания зависит от выбора  $s$ .

Вначале проведем исследования предлагаемого метода (получившего название интерквантильного, ИК) (результаты обозначены индексом IQR), полагая оптимальное (обеспечивающее наименьшее значение приведенной ошибки оценивания) значение  $s^{\text{opt}}$  (или нормированное  $s_{\%}^{\text{opt}} = 100 \cdot s^{\text{opt}}/N$ ) известным. Применим методику, использованную для исследования свойств бутстреп оценки. Начнем с анализа влияния параметра  $p$ , зафиксировав остальные

параметры модели (1):  $m=10$ ;  $\sigma_m^2=8$ ;  $M/m=100$ . Результаты моделирования приведены в табл. 6. По аналогии с бутстреп-оценкой для ИК оценки вводятся параметры  $\bar{n}_{IQR\%}$  и  $\theta_{n_{IQR\%}}^2$ , характеризующие среднее и дисперсию номера квантиля, выбираемого в качестве итоговой оценки.

Таблица 6

Численные критерии точности ИК оценки для различных значений параметра  $p$  модели (1)

$p$	$s_{\%}^{opt}$	$\bar{n}_{IQR\%}$	$\theta_{n_{IQR\%}}^2$	$\Delta_{IQR}$	$\theta_{IQR}^2$	$\varepsilon_{IQR}$
0,90	98,0	45,0	0,10	0,013	0,015	0,015
0,75	73,4	37,7	0,25	0,019	0,019	0,019
0,65	63,7	32,8	0,34	0,022	0,022	0,023
0,50	48,2	25,4	0,42	0,023	0,030	0,031
0,25	23,7	13,2	0,44	0,022	0,063	0,063
0,15	13,7	8,3	0,37	0,023	0,114	0,115
0,10	9,1	5,9	0,33	0,023	0,185	0,186

Из сравнения результатов в табл. 6 и 3 видно, что предлагаемая ИК оценка (при оптимальном выборе  $s$ ) существенно превосходит бутстреп и сравнима с мириадной (при оптимальном  $k/\sigma_{m_{opt}}$ ) оценками по всем анализируемым параметрам  $|\Delta|$ ,  $\theta^2$ ,  $\varepsilon$  и  $\varepsilon_{\%}$  для всех значений  $p$ . Интересно отметить, что для ИК характерен более «аккуратный» (по сравнению с бутстреп оценкой) выбор номера квантиля итоговой оценки. Об этом свидетельствует существенно меньшая дисперсия  $\theta_{n_{IQR\%}}^2$  по сравнению с  $\theta_{n_{\%}}^2$ . Кроме того, между средним номером квантиля  $\bar{n}_{IQR\%}$ , параметром  $p$  и оптимальным интерквантильным расстоянием  $s_{\%}^{opt}$  прослеживается четкая взаимосвязь  $\bar{n}_{IQR\%} \approx 100 \cdot p/2 \approx s_{\%}^{opt}/2$ . При этом дополнительный анализ показал, что при изменении других параметров модели (1) ( $m$ ,  $\sigma_m^2$  и  $M/m$ ) эта взаимосвязь сохраняется.

Интересен также тот факт, что среднее значение номера квантиля  $\bar{n}_{IQR\%}$  слабо зависит от величины интерквантильного расстояния, при котором была получена оценка (что можно видеть на рис. 3, где положения минимумов кривых  $\delta(t+s/2)$  при разных  $s/N$  совпадают). Все это в совокупности приводит к идее использования для определения оптимального  $s_{\%}^{opt}$  предварительной интерквантильной оценки  $t_{est}^{init}$  с  $s = [0,1 \cdot N]_2$ , которая обеспе-

чит практически несмещенное оценивание для всего диапазона возможных  $p \geq 0,1$  (что можно считать достаточным для большинства практических ситуаций). Как видно из рис. 4, зависимость  $s_{\%}^{opt}$  от  $\bar{n}_{IQR\%}^{s=0,1N}$  оказывается практически линейной при этом  $s_{\%}^{opt} \approx 2\bar{n}_{IQR\%}^{s=0,1N}$ . Однако на практике среднее значение номера квантиля неизвестно и вместо него приходится использовать оценку  $t_{est}^{init}$ , определяемую с некоторой неточностью. В результате аппроксимированное значение может выходить за допустимый максимум  $s^{opt} \leq N-1$ , кроме того оно может оказываться нечетным. Для устранения этих недостатков будем использовать квазиоптимальное значение  $s^{qopt} = [1,444 \cdot t_{est}^{init} - 0,067 N]_2$ , где  $[\bullet]_2$  обозначает операцию округления до ближайшего четного. Такая аппроксимация является более «осторожной» за счет меньшего угла наклона прямой (см. зависимость  $s_{\%}^{qopt}$  от  $\bar{n}_{IQR\%}^{s=0,1N}$  на рис. 4).

Одной из проблем использования  $t_{est}^{init}$  для расчета квазиоптимального  $s^{qopt}$  является достаточно высокая дисперсия этой оценки, связанная с «шумным» характером функции интерквантильного расстояния  $\delta(t+s/2) = X^{(t+s)} - X^{(t)}$ , характерной для малых  $s$  (см. рис. 3). Для уменьшения флуктуаций функции  $\delta(t+s/2)$  предлагается аппроксимировать ее значения в области минимума регрессионной кривой второго порядка, а затем использовать аналитически найденный минимум  $t_{est}^{flt}$  в качестве скорректированной оценки  $X^{(t_{est}^{flt})}$ . Примеры регрессионных кривых для соответствующих интерквантильных расстояний  $s/N=0,1$  и  $0,2$  приведены жирными линиями на рис. 3.

Для определения границ области аппроксимации было проведено специальное исследование, из которого следует, что наилучшим можно считать выбор  $[0,1(t_{est}^{init} + s/2); 1,9(t_{est}^{init} + s/2)]$ , поскольку он для большинства рассматриваемых вариантов обеспечивает наибольшую точность определения глобального минимума функции интерквантильного расстояния  $\delta(t+s/2)$ . Назовем такую оценку интерквантильной с фильтрацией (ИКФ).

Объединив описанные выше решения, сформулируем процедуру получения автоматической интерквантильной оценки (АИК).



1. К исходной выборке  $X$  применяется ИК оценка с  $s = [0,1 \cdot N]_2$  и определяется  $t_{\text{est}}^{\text{init}}$ .

2. На основании  $t_{\text{est}}^{\text{init}}$  определяется область аппроксимации  $[0,1(t_{\text{est}}^{\text{init}} + s/2); 1,9(t_{\text{est}}^{\text{init}} + s/2)]$  и затем уточненный номер квантиля  $t_{\text{est}}^{\text{I}}$ .

3. Квантиль  $t_{\text{est}}^{\text{I}}$  используется для определения квазиоптимального интерквантильного расстояния  $s^{\text{qopt}} = [1,444 \cdot t_{\text{est}}^{\text{I}} - 0,067N]_2$ , при этом, если  $s^{\text{qopt}} > N - 1$ , то  $s^{\text{qopt}} = [N - 1]_2$ .

4. Пункты 1-3 повторяются с  $s = s^{\text{qopt}}$ , находится итоговый квантиль  $t_{\text{est}}^{\text{II}}$  и итоговая оценка  $X(t_{\text{est}}^{\text{II}})$ .

Результаты моделирования для ИКФ (при оптимальном интерквантильном расстоянии) (индекс IQRF) и АИК (индекс AIQRF) оценок для различных рассматриваемых вариантов параметров модели (1) представлены в табл. 7.

Как видно из табл. 7, автоматическая оценка для всех рассматриваемых вариантов обеспечивает практически такие же результаты, как и ИКФ оценка при оптимальном выборе интерквантильного расстояния. При этом даже для самых неблагоприятных вариантов (5 и 6) величина относительной приведенной ошибки не превосходит 4%, а для остальных вариантов оказывается не выше 1,5%.

## Заключение

Задача определения параметра сдвига существенно несимметричных односторонних распределений с тяжелым «хвостом» требует применения оценок с очень высокой робастностью.

Относительно простая квантильная оценка с фиксированным  $n\%$  обеспечивает приемлемые результаты лишь для очень узкого диапазона изменения параметров распределения (в частности  $p$ ).

Более сложная мириадная оценка обеспечивает достаточно точные результаты в широком диапазоне изменения параметров распределения, но только

при условии оптимального выбора настроечного параметра  $k/\sigma_{m \text{opt}}$ , величина которого, в свою очередь, также зависит от параметров распределения. Практическое решение в виде комбинированной квантильно-мириадной оценки обеспечивает приемлемую точность, но применимо лишь для частной задачи определения характеристик помех, где между параметрами  $m$  и  $\sigma_m^2$  существует взаимосвязь.

Более эффективным и универсальным оказался подход на основе использования квантильных оценок с изменяемым номером квантиля. Использование для определения необходимого номера квантиля бутстреп подхода (скорректированная бутстреп оценка) оказалось, в целом, более эффективным по сравнению с квантильной и мириадной оценками, но все же в некоторых ситуациях недостаточно точным. Комбинирование бутстреп и мириадной оценки позволило улучшить точность, но ограничило область применения такой оценки.

Наилучшим по точности и наиболее универсальным оказался подход на основе определения необходимого номера квантиля по минимуму интерквантильного расстояния (ИКФ и АИК оценки). АИК оценка применима для всего рассматриваемого диапазона изменения параметров распределения, не использует каких-либо взаимосвязей между этими параметрами, при этом обеспечивает точность (величину относительной приведенной ошибки) для наиболее неблагоприятных ситуаций не хуже 4%, а в большинстве ситуаций – не хуже 1,5%.

## Литература

1. Schowengerdt, R. A. *Remote Sensing: Models and Methods for Image Processing [Text]* / R. A. Schowengerdt. – NY (USA) : Academic Press, 2006. – 560 p.
2. *Noise Identification and Estimation of its Statistical Parameters by Using Unsupervised Variational Classification [Text]* / B. Vozel, K. Chehdi, L. Klaine, V. Lukin, S. Abramov // *IEEE ICASSP : Proc.* 2006. – Toulouse, France, 2006. – V. 2. – P. 841-844.
3. Huber, P. J. *Robust Statistics [Text]* / P. J. Huber. – NY(USA) : John Wiley & Sons, 2004. – 308 p.
4. Astola, J. *Fundamentals of nonlinear digital fil-*

Таблица 7

Численные критерии точности ИКФ и АИК оценок для различных вариантов параметров модели (1)

Вариант	ИКФ оценка (при оптимальном $s$ )				АИК оценка			
	$\Delta_{\text{IQRF}}$	$\theta_{\text{IQRF}}^2$	$\varepsilon_{\text{IQRF}}$	$\varepsilon_{\text{IQRF}}^{\%}$	$\Delta_{\text{AIQRF}}$	$\theta_{\text{AIQRF}}^2$	$\varepsilon_{\text{AIQRF}}$	$\varepsilon_{\text{AIQRF}}^{\%}$
1	0,04	1,44	1,44	1,20	-0,13	1,59	1,61	1,27
2	-0,40	14,78	14,95	1,29	-0,50	16,55	16,80	1,37
3	-0,17	2,14	2,17	1,47	-0,18	2,24	2,28	1,51
4	-0,30	11,39	11,48	1,69	-0,39	11,81	11,95	1,73
5	-0,13	90,55	90,56	3,17	-1,60	95,18	97,75	3,30
6	-0,85	259,13	259,86	4,03	-0,97	253,41	254,36	3,99
7	-0,31	32,88	32,98	1,44	-0,48	34,36	34,59	1,47

tering [Text] / J. Astola, P. Kuosmanen. – Boca Raton (USA) : CRC Press LLC, 1997. – 276 p.

5. Blind evaluation of noise variance in images using myriad operation [Text] / S. Abramov, V. Lukin, A. Zelensky, J. Astola // Image Processing: Algorithms and Systems : Proc. 2002. – San Jose, USA, 2002. – V. 4667. – P. 192-203.

6. On required accuracy of mixed noise parameter estimation for image enhancement via denoising [Text] / V. Abramova, S. Abramov, V. Lukin, K. Egiazarian, J. Astola // EURASIP JIVP. – 2014. – V. 2014, № 1. – P. 1-16.

7. Nonlinear Signal and Image Processing: Theory, Methods, and Applications [Text] / Ed. by K. E. Barner, G. R. Arce. – Boca Raton (USA) : CRC Press LLC, 2003. – 560 p.

8. Efron, B. An Introduction to the Bootstrap [Text] / B Efron., R. J. Tibshirani. – Boca Raton (USA) : Chapman and Hall/CRC, 1994. – 456 p.

### References

1. Schowengerdt, R. A. Remote Sensing: Models and Methods for Image Processing. NY (USA), Academic Press Publ., 2006. 560 p.

2. Vozel, B., Chehdi, K., Klaine, L., Lukin, V.,

Abramov, S. Noise Identification and Estimation of its Statistical Parameters by Using Unsupervised Variational Classification. Proc. IEEE ICASSP. Toulouse, France, 2006, vol. 2, pp. 841-844.

3. Huber, P. J. Robust Statistics. NY(USA), John Wiley & Sons Publ., 2004. 308 p.

4. Astola, J., Kuosmanen, P. Fundamentals of nonlinear digital filtering. Boca Raton (USA), CRC Press LLC Publ., 1997. 276 p.

5. Abramov, S., Lukin, V., Zelensky, A., Astola, J. Blind evaluation of noise variance in images using myriad operation. Proc. Image Processing: Algorithms and Systems, San Jose, USA, 2002, vol. 4667, pp. 192-203.

6. Abramova, V., Abramov, S., Lukin, V., Egiazarian, K., Astola, J. On required accuracy of mixed noise parameter estimation for image enhancement via denoising. EURASIP JIVP, 2014, vol. 2014, no. 1, pp. 1-16.

7. Nonlinear Signal and Image Processing: Theory, Methods, and Applications. Ed. by Barner, K. E., Arce, G. R. Boca Raton (USA), CRC Press LLC Publ., 2003. 560 p.

8. Efron, B., Tibshirani, R. J. An Introduction to the Bootstrap. Boca Raton (USA), Chapman and Hall/CRC Publ., 1994. 456 p.

Поступила в редакцію 22.02.2017, рассмотрена на редколлегии 15.03.2017

### СТІЙКЕ ОЦІНЮВАННЯ ПАРАМЕТРУ ЗСУВУ РОЗПОДІЛІВ У ЗАДАЧАХ АВТОМАТИЧНОЇ ОБРОБКИ ЗОБРАЖЕНЬ

С. К. Абрамов

Розглянуто задачу визначення параметру зсуву закону розподілу вибірок даних, які характерні для задач автоматичної обробки зображень. Показано, що закон розподілу таких вибірок має виразний несиметричний характер, що зумовлено присутністю важкого однобічного «хвоста». Застосування відомих стійких оцінок параметру зсуву в таких умовах не забезпечує бажаної точності. У зв'язку з цим запропоновано декілька нових підходів, що базуються на використанні міриадної та квантильної оцінок. Показано, що запропоновані рішення забезпечують високу точність й стійкість. Надано рекомендації щодо їх застосування на практиці.

**Ключові слова:** стійке оцінювання, параметр зсуву, автоматична обробка зображень.

### ROBUST LOCATION PARAMETER ESTIMATION OF DISTRIBUTIONS IN AUTOMATIC IMAGE PROCESSING TASKS

S. K. Abramov

The task of location parameter estimation of samples distribution processed in automatic image processing tasks is considered. It is shown that the distribution law of such samples is clearly non-symmetric due to one-side heavy tail presence. Applying of known robust location parameter estimators in such conditions does not provide desirable accuracy. Therefore some new approaches based on using myriad and quantile estimators are proposed. It is shown, that the proposed solutions provide high accuracy and robustness. Recommendations about applying the proposed techniques are given.

**Key words:** robust estimation, location parameter of distribution, automatic image processing.

**Абрамов Сергей Клавдиевич** – канд. техн. наук, доцент, докторант кафедри приёма, передачи и обработки сигналов, Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: s.abramov@rtsla.khai.edu.

**Abramov Sergey Klavdievich** – Candidate of Technical Science, Assistant Professor of Dept. of Receivers, Transmitters and Signal Processing, National Aerospace University named after N. Ye. Zhukovsky “KhAI”, Kharkov, Ukraine, e-mail: s.abramov@rtsla.khai.edu.