

УДК 621.396:681.34

А.А. РОЕНКО, В.В. ЛУКИН, А.А. ЗЕЛЕНСКИЙ

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

ОПРЕДЕЛЕНИЕ ПАРАМЕТРА СДВИГА ВЫБОРКИ ДАННЫХ С СИММЕТРИЧНЫМ НЕГАУССОВЫМ РАСПРЕДЕЛЕНИЕМ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ МЕТОДОВ АДАПТИВНОГО РОБАСТНОГО ОЦЕНИВАНИЯ

Рассмотрены робастные оценки параметра сдвига, принадлежащие трем классам устойчивых оценок. Проведен статистический анализ их свойств для трех видов плотности распределения вероятности с тяжелыми хвостами, симметричных относительно математического ожидания. Предложены два метода адаптации к закону распределения помехи и четыре новых адаптивных робастных оценки параметра сдвига. Проведен их статистический анализ и продемонстрированы их преимущества.

адаптивный метод, робастное оценивание, негауссов закон распределения

Введение

Значительные трудности, возникающие во многих прикладных задачах обработки данных, нередко связаны с тем, что исходная информация (сигнал, изображение) искажена шумами различного типа. Более того, часто статистические характеристики помех либо частично, либо полностью априорно не определены. Для ряда практических приложений известно лишь, что плотность распределения вероятности (ПРВ) шума (ошибок первичных измерений) имеет вид, симметричный относительно параметра сдвига.

Однако, такими свойствами обладает достаточно большое количество законов распределения: гауссов, Коши, Лапласа, показательный, Тихонова и т.д. [1 – 3]. Кроме того, вид основного распределения может быть искажен вследствие присутствия в выборке данных элементов из “загрязняющего распределения” (импульсных помех, выбросов, аномальных измерений), следствием чего могут являться “тяжелые хвосты” итогового распределения помех или измерений [3].

Такая ситуация характерна, например, для систем обработки, основанных на биспектральном анализе, в которых при воздействии гауссова белого

шума на входе системы при малом соотношении сигнал/помеха вид ПРВ шумовой компоненты оценки биспектра становится отличным от нормального [4]. Аналогичные особенности характерны для ПРВ оценок задержки случайных широкополосных сигналов при их взаимно-корреляционной обработке в условиях воздействия интенсивных аддитивных помех [5].

В рассматриваемых приложениях для повышения точности измерений целесообразно использовать робастные методы фильтрации или вторичной обработки данных [3, 6]. В принципе, при заранее известной ПРВ помех или ошибок измерений часто можно выбрать или найти эффективную и несмещенную оценку. Для неизвестных ПРВ существуют итеративные методы и алгоритмы их оценивания и адаптации, однако они работоспособны лишь при достаточно больших размерах (порядка тысяч элементов) имеющейся выборки данных [6].

Как уже отмечалось, закон распределения полагается априорно неизвестным. Кроме того, на практике размер выборки данных, для которой оцениванию подлежит параметр сдвига, может быть порядка десятков.

В связи с этим, *актуальной задачей* является разработка и исследование методов обработки вы-

борок данных ограниченного размера, про которые известно лишь, что их ПРВ имеет симметричный (относительно параметра сдвига), и, скорее всего, негауссов вид. Желательно, чтобы предлагаемые методы были достаточно устойчивыми и адаптивными к степени тяжести хвостов закона распределения.

1. Обзор известных робастных оценок

Как известно, для многих видов ПРВ устойчивые эффективные (т.е. обеспечивающие наименьшую возможную дисперсию) оценки принадлежат к классу L -оценок, т.е. являются линейной комбинацией порядковых статистик для обрабатываемой выборки данных, содержащей n элементов [7]:

$$X_L = \sum_{i=1}^n a_i X_{(i)}, \quad (1)$$

где $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n-1)} \leq X_{(n)}$ – порядковые статистики; $a_i, i \in 1, \dots, n$ – весовые коэффициенты.

Рассмотрим некоторые из L -оценок.

Медианная оценка (*median*), являющаяся оптимальной для распределения Лапласа [7], представляет собой центральный элемент упорядоченной выборки, если n – нечетное, и полусумму $n/2$ -й и $(n/2)+1$ -й порядковых статистик, если n – четное. Другими словами, для нечетного количества порядковых статистик весовой вектор имеет следующий вид:

$$a_i = \begin{cases} 1, & \text{если } i = (n+1)/2; \\ 0, & \text{если } i \neq (n+1)/2. \end{cases} \quad (2)$$

Для четного числа элементов:

$$a_i = \begin{cases} 0,5, & \text{если } i = n/2 \text{ и } i = n/2 + 1; \\ 0, & \text{иначе.} \end{cases} \quad (3)$$

Частным случаем L -оценки является также среднее выборки (*mean*) [7]. Данная оценка является эффективной для гауссова закона распределения и имеет значения весового вектора, равные

$$a_i = \frac{1}{n}, \quad i = 1, \dots, n. \quad (4)$$

Однако, даже незначительные отклонения вида ПРВ от нормального, обусловленные, например, наличием в выборке выбросов, приводят к ухудшению статистических свойств оценок, в результате чего оценка среднего оказывается неоптимальной для данного закона распределения.

В [7] предложена более устойчивая L -оценка, так называемое трехэлементное среднее (*TRImean*). Усреднению подвергаются порядковые статистики, которые близки к трем квартилям. При четном n отсчеты в упорядоченной выборке, соответствующие верхнему и нижнему квартилям, берутся с весовыми коэффициентами 0,25, а медиана – с весом 0,5:

$$a_i = \begin{cases} 0,25, & \text{если } i_1 = n/4 \text{ и } i_2 = 3n/4; \\ 0,25, & \text{если } i = n/2 \text{ и } i = (n+1)/2; \\ 0, & \text{иначе.} \end{cases} \quad (5)$$

Для нечетного количества отсчетов в выборке

$$a_i = \begin{cases} 0,25, & \text{если } i_1 = (n-1)/4 + 1 \text{ и } i_2 = 3(n-1)/4 + 1; \\ 0,5, & \text{если } i = (n+1)/2; \\ 0, & \text{иначе.} \end{cases} \quad (6)$$

Еще одной оценкой, получаемой при помощи линейной комбинации порядковых статистик, является α -урезанное среднее (α -*trimmed*) [7]. Свойство робастности в данном случае обеспечивается за счет усреднения лишь тех значений, порядковые номера которых попадают в интервал

$$[n/2 - N_{\alpha 1}; n/2 + N_{\alpha 2}],$$

где $N_{\alpha 1}$ и $N_{\alpha 2}$ – параметры усечения:

$$X_{\alpha} = \frac{1}{N_{\alpha 1} + N_{\alpha 2} + 1} \sum_{i=n/2 - N_{\alpha 1}}^{n/2 + N_{\alpha 2}} X_{(i)}. \quad (7)$$

Статистические свойства оценки α -*trimmed* зависят от $N_{\alpha 1}$ и $N_{\alpha 2}$. Поэтому сразу укажем, что ниже рассмотрим оценку α -*trimmed* с $N_{\alpha 1} = N_{\alpha 2} = n/4$.

Кроме оценок, принадлежащих к весьма широкому классу L -оценок, свойством устойчивости для определенных видов ПРВ обладают также некоторые M - и R -оценки [2].

Так, например, мириада выборки (*myriad*), являющаяся оценкой максимального правдоподобия

(M -оценка), оптимальна для распределения Коши [8]. Ее основная особенность – наличие настроечного параметра $k > 0$, при определенном выборе которого мириадная оценка с высокой точностью определяет моду распределения, совпадающую со сдвигом для симметричных плотностей распределения вероятностей.

Формальное определение мириады для заданного k имеет вид [8]:

$$X_{myr} = \text{myriad}\{k; X_1, X_2, \dots, X_n\} = \arg \min_{\theta} \left\{ \sum_{j=1}^n \ln[k^2 + (X_j - \theta)^2] \right\}, \quad (8)$$

где X_j – j -й элемент обрабатываемой выборки данных.

Свойство модальности наблюдается при k , стремящемся к нулю (модальный режим работы мириады). Если же k стремится к бесконечности, то мириадная оценка практически совпадает со средним выборки (линейный режим работы мириады).

Таким образом, мириаду можно использовать и как оценку среднего выборки, и как оценку моды, что достигается выбором соответствующего k . При этом уменьшение настроечного параметра относительно выбранного значения обеспечивает увеличение устойчивости мириадной оценки к присутствию в выборке элементов с распределением, существенно отличающимся от основного. Но в любом случае настроечный коэффициент должен быть согласован с масштабом данных.

В работе [8] показано, что для обеспечения хорошей устойчивости к аномальным измерениям параметр k должен иметь порядок

$$k = (1 \dots 3) \sigma_{очн}, \quad (9)$$

где $\sigma_{очн}$ – среднеквадратическое отклонение (СКО) ошибок измерений, которые полагаются нормальными.

Одним из представителей класса оценок, основанных на ранговых критериях (R -оценки), является оценка Вилкоксона (WC) [2]:

$$X_{WC} = \text{med} \left\{ \frac{X_1 + X_1}{2}, \frac{X_1 + X_2}{2}, \dots, \frac{X_1 + X_n}{2}, \frac{X_2 + X_2}{2}, \frac{X_2 + X_3}{2}, \dots, \frac{X_{n-1} + X_n}{2}, \frac{X_n + X_n}{2} \right\}, \quad (10)$$

т.е. данная оценка представляет собой медиану последовательности размером $\frac{n(n+1)}{2}$, состоящей из элементов исходной выборки и всех возможных попарных средних.

Оценка Вилкоксона способна устранять до $(0,25 \dots 0,3)n$ односторонних выбросов (импульсных помех), т.е. измерений, существенно отличающихся от истинного значения в меньшую, либо в большую сторону. Однако, недостатком данной оценки является большая сложность вычислений, связанных прежде всего со значительным объемом сортируемых данных.

Одним из способов уменьшения необходимых вычислительных затрат является применение так называемой оценки Ходжеса-Лемана (HL) [2], которая обладает характеристиками, мало отличающимися от статистических характеристик WC . При четном n оценка Ходжеса-Лемана имеет вид

$$X_{HL} = \text{med} \left\{ X_1; X_2; \dots; X_n; \frac{X_1 + X_n}{2}; \frac{X_2 + X_{n-1}}{2}; \dots; \frac{X_{n/2} + X_{(n+2)/2}}{2} \right\}. \quad (11)$$

Отметим, что при симметричности ПРВ относительно математического ожидания (сдвига, моды) все приведенные выше оценки (α -урезанное среднее при условии одинаковых $N_{\alpha 1}$ и $N_{\alpha 2}$) являются несмещенными.

2. Анализ эффективности рассмотренных оценок

Проведем предварительный анализ свойств описанных выше оценок при помощи численного моделирования на ЭВМ. Полученные данные ниже представлены для трех видов ПРВ, относящихся к рассматриваемому классу симметричных негауссовых.

Первое ПРВ, вернее, подкласс ПРВ, основывается на модели составного распределения [2, 7]:

$$\rho_1(x) = (1 - P_0)p_{осн}(x) + P_0p_{загр}(x), \quad (12)$$

где $p_{осн}(x)$ – ПРВ основного распределения; $p_{загр}(x)$ – ПРВ загрязняющего распределения; P_0 – параметр загрязняющего распределения, характеризующий вероятность появления в выборке элементов, принадлежащих $p_{загр}(x)$, причем для СКО $p_{осн}(x)$ и $p_{загр}(x)$ выполняется условие

$$\sigma_{загр} \gg \sigma_{осн}.$$

В качестве основной ПРВ было использовано нормальное распределение с нулевым математическим ожиданием (МО) и единичной дисперсией ($\sigma_{осн}^2 = 1$). Для моделирования отсчетов, представляющих собой выбросы, использовалась гауссова ПРВ с нулевым МО и СКО $\sigma_{загр} = 5$. При моделировании вероятность P_0 в (12) изменялась с шагом 0,04 от 0 до 0,4. Таким образом, исследования были проведены для достаточно широкого подкласса ПРВ $\rho_1(x)$.

В [9] нами показано, что ПРВ произведения двух $\rho_2(x)$ и трех $\rho_3(x)$ нормальных независимых случайных величин с нулевыми МО имеют негауссов характер и обладают тяжелыми хвостами. При этом, если дисперсии этих величин равны единице, то для $\rho_2(x)$ и $\rho_3(x)$ дисперсии также равны единице. В связи с этим было принято решение проанализировать статистические свойства описанных выше оценок и для ПРВ $\rho_2(x)$ и $\rho_3(x)$.

Для всех трех ПРВ формировались 4000 реализаций выборок с $n = 64$. К ним применялись перечисленные выше оценки и для каждой из оценок рассчитывалась дисперсия, определяемая следующим образом:

$$\sigma_{ВЫХ}^2 = \sum_{l=1}^{4000} \left[\hat{X}_l - m_{\hat{X}} \right]^2, \quad (13)$$

где \hat{X}_l – оценка, полученная для l -й реализации; $m_{\hat{X}}$ – математическое ожидание оценок. Для мириады выборки k было фиксированным:

$$k = 2\sigma_{осн} = 2.$$

Для ПРВ $\rho_1(x)$ полученные данные представлены в табл. 1 для различных значений P_0 . Результаты, полученные для выборок случайных величин с ПРВ $\rho_2(x)$ и $\rho_3(x)$ представлены в табл. 2.

Анализ значений дисперсии оценок, приведенных в обеих таблицах, показывает, что в зависимости от параметров и вида ПРВ минимальной дисперсией может обладать та или иная оценка.

Для случая составного распределения (12) при очень малом P_0 наилучшие результаты, как и ожидалось, дает применение обычного усреднения. Медиана выборки характеризуется худшей точностью по сравнению с большинством рассматриваемых оценок на интервале значений P_0 от 0 до 0,28. При большем P_0 точность данной оценки улучшается, и при $P_0 = 0,4$ сравнивается с α -урезанным средним и оказывается лучше, чем для всех остальных оценок (*WC*, *HL*, *TRImean*), кроме мириады с $k = 2$.

При значениях вероятности появления аномальных измерений, лежащих в интервале от 0,04 до 0,08, наилучшими среди рассматриваемых оказываются оценки Вилкоксона и Ходжеса-Лемана, которые, кроме того, до значения P_0 , равного 0,3, обеспечивают меньшие значения дисперсии по сравнению с медианой выборки. Применение мириадной оценки с $k = 2$ позволяет обеспечить наименьшую дисперсию на интервале значений P_0 от 6 до 40%, в то время, как при использовании α -урезанного среднего для составной ПРВ ($P_0 \in [0,2; 0,4]$) удается получить наилучший результат среди рассматриваемых L и R -оценок.

При обработке выборок случайных величин с ПРВ $\rho_2(x)$ и $\rho_3(x)$ (данные в табл. 2) наилучшей среди рассматриваемых оказывается медианная

оценка, причем в первом случае (для $\rho_2(x)$) выигрыш в выходной дисперсии относительно значения наилучшей из других оценок достигает двух, а во втором ($\rho_3(x)$) – четырех раз.

Таблица 1

Дисперсии рассмотренных оценок для ПРВ $\rho_1(x)$ при различных P_0

P_0	<i>Mean</i>	<i>Median</i>	α - <i>trimmed</i>	<i>Myriad</i>	<i>TRImean</i>	<i>WC</i>	<i>HL</i>
0	0,0155	0,0243	0,0198	0,0169	0,0190	0,0165	0,0168
0,04	0,0299	0,0256	0,0211	0,0187	0,0197	0,0182	0,0185
0,08	0,0459	0,0278	0,0234	0,0205	0,0228	0,0214	0,0217
0,12	0,0597	0,0300	0,0256	0,0220	0,0242	0,0242	0,0246
0,16	0,0746	0,0320	0,0272	0,0236	0,0276	0,0266	0,0268
0,20	0,0909	0,0352	0,0312	0,0264	0,0299	0,0320	0,0319
0,24	0,1041	0,0374	0,0334	0,0285	0,0330	0,0354	0,0354
0,28	0,1180	0,0398	0,0362	0,0325	0,0388	0,0400	0,0391
0,32	0,1355	0,0465	0,0421	0,0368	0,0430	0,0476	0,0461
0,36	0,1527	0,0506	0,0477	0,0400	0,0500	0,0552	0,0532
0,40	0,1650	0,0535	0,0535	0,0453	0,0575	0,0631	0,0603

Таблица 2

Дисперсии рассмотренных оценок при обработке выборок случайных величин с ПРВ $\rho_2(x)$ и $\rho_3(x)$

	<i>Mean</i>	<i>Median</i>	α - <i>trimmed</i>	<i>Myriad</i>	<i>TRImean</i>	<i>WC</i>	<i>HL</i>
ρ_2	0,0160	0,0031	0,0063	0,0088	0,0067	0,0068	0,0065
ρ_3	0,0159	0,0005	0,0019	0,0051	0,0025	0,0022	0,0020

Оценки α -урезанного среднего, Ходжеса-Лемана, Вилкоксона и трехэлементного среднего характеризуются примерно равными дисперсиями, которые меньше, чем для мириады и среднего выборки. Для случайных величин с ПРВ $\rho_2(x)$ и $\rho_3(x)$ точность оценки *mean*, как и следовало ожидать, значительно хуже, чем для всех остальных рассмотренных выше робастных оценок.

Таким образом, в условиях априорной неопределенности о ПРВ случайных величин весьма сложно

не только синтезировать алгоритм оптимального оценивания, но и выбрать такую робастную оценку, которая обеспечивала бы дисперсию, близкую к минимально достижимой. Поэтому желательно разработать такие методы, которые бы сочетали в себе достоинства приведенных выше оценок, т.е. адаптировались к виду закона распределения, и в зависимости от его параметров выбирали наилучшую (или, по крайней мере, достаточно хорошую) для данного конкретного случая оценку параметра сдвига.

3. Способы адаптации к ПРВ помехи

Из анализа данных, приведенных в табл. 1 и 2, видно, что по мере утяжеления хвостов ПРВ на первые позиции по обеспечиваемой точности выходят оценки, которые обладают более высокой устойчивостью, что хорошо согласуется с известными теоретическими положениями.

Поэтому одним из возможных методов адаптации к виду ПРВ представляется использование какого-либо параметра, характеризующего степень "тяжести хвостов" закона распределения. При этом желательно, чтобы данный параметр однозначно указывал, какую оценку выбирать для обработки выборки данных с обеспечением приемлемо высокой точности.

В принципе, такой подход не является абсолютно новым. В частности, в [10] описана адаптивная α -урезанная оценка среднего (ААТМЕ). Данный метод основан на расчете параметра α для рассматриваемой выборки данных. Оптимальное значение параметра усечения определяется следующим выражением:

$$N_\alpha = \{ |1 - \beta| (n-1) / 2 \}, \quad (14)$$

где $\{ \dots \}$ – операция округления до целого числа.

Параметр β при этом рассчитывается следующим образом:

$$\beta = \sigma^2 / (\sigma^2 + \sigma_{II}^2); \quad (15)$$

$$\sigma^2 = \begin{cases} \sigma_X^2 - \sigma_{II}^2, & \text{если } \sigma_X^2 \geq \sigma_{II}^2; \\ 0, & \text{если } \sigma_X^2 < \sigma_{II}^2, \end{cases} \quad (16)$$

где σ_X^2 – дисперсия обрабатываемой выборки данных; σ_{II}^2 – дисперсия основного распределения (задается априорно и в данном случае при моделировании полагалась равной единице).

Однако, во-первых, такой вариант адаптации не является единственно возможным. Во-вторых, проведенный выше анализ показывает, что многие робастные оценки не хуже, чем α -урезанная оценка

среднего. Поэтому ниже нами предложен ряд других адаптивных робастных оценок. При их синтезе учитывались аналогии с разработкой локально-адаптивных фильтров с жестким и мягким переключением [11].

В случае адаптивного оценивания с жестким переключением алгоритм в обобщенном виде описывается выражением

$$\hat{X} = \begin{cases} \hat{X}_1, & \text{если } \xi \in [\psi_0, \psi_1); \\ \hat{X}_2, & \text{если } \xi \in [\psi_1, \psi_2); \\ \dots & \dots \\ \hat{X}_L, & \text{если } \xi \in [\psi_{L-1}, \psi_L), \end{cases} \quad (17)$$

где ξ – параметр адаптации, ψ_l – l -й порог, $l = 1, \dots, L$; \hat{X}_l , $l = 1, \dots, L+1$ – оценка, полученная при использовании l -й методики робастного оценивания. При этом порог ψ_0 может быть равен $-\infty$, а $\psi_L = \infty$.

Очевидно, что, во-первых, адаптивная оценка ААТМЕ относится к классу оценок (17), причем в качестве параметра адаптации ξ выступает β , а в качестве \hat{X}_l – оценки α -урезанного среднего с различными N_α . Во-вторых, свойства адаптивных робастных оценок (17) зависят от многих параметров и факторов – выбора пороговых значений ψ_l , $l = 1, \dots, L$, выбора параметра адаптации ξ и его статистических характеристик, а также выбора L и методик робастного оценивания.

Одним из параметров, частично характеризующим негауссову ПРВ, является коэффициент эксцесса ПРВ:

$$K = \frac{\mu_4}{\sigma^4} - 3; \quad \mu_4 = \int_{-\infty}^{\infty} (x - m_x)^4 p(x) dx, \quad (18)$$

μ_4 – центральный момент четвертого порядка; σ , $p(x)$ – СКО и ПРВ рассматриваемой выборки.

Как известно, коэффициент эксцесса характеризует сглаженность кривой около МО и для нормальной плотности вероятности K равен нулю [3]. Положительное значение указывает на то, что вид ПРВ

в окрестности максимума имеет более высокую и более острую вершину, чем гауссов закон распределения. Наоборот, отрицательное значение коэффициента эксцесса указывает на более низкий и более плоский характер вершины по сравнению с нормальной кривой.

Однако, параметр K (18) не позволяет однозначно определить степень “тяжести хвостов” распределения, так как одному значению данного параметра соответствует несколько видов ПРВ [6]. Проиллюстрируем это для полученных результатов численного моделирования. Например, значение коэффициента эксцесса, равное 5, соответствует составным законам распределения (12) при $P_0 \approx 0,04$ и $P_0 \approx 0,27$ (рис. 1). В то же время, как видно из табл. 1, для $P_0 \approx 0,04$ наименьшую дисперсию обеспечивает оценка Вилкоксона, а для $P_0 \approx 0,27$ наибольшей точностью среди рассмотренных оценок обладает мириада с $k = 2$. Кстати, для случайных величин с ПРВ $\rho_2(x)$ и $\rho_3(x)$ значения параметра K соответственно равны 3,62 и 8,58.

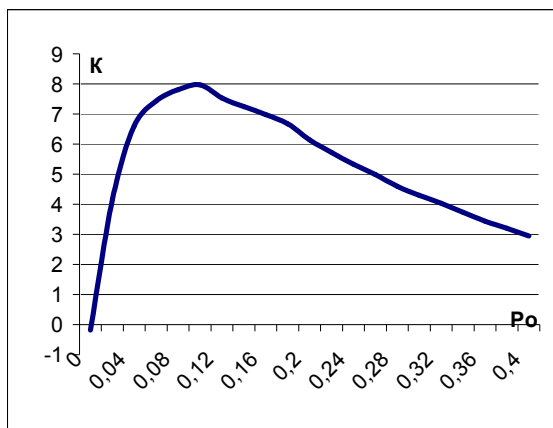


Рис. 1. Зависимость усредненных значений параметра K от P_0 для ПРВ (12)

Одним из способов решения данной проблемы является применение так называемого коэффициента эксцесса на базе процентилей ($ПКЭ$) [7], который является более робастной величиной и позволяет, как монотонно спадающая функция для распределения вида (12), однозначно определять степень “тя-

жести хвостов” (рис. 2). Для ПРВ $\rho_2(x)$ и $\rho_3(x)$ средние значения $ПКЭ$ соответственно равны 0,165 и 0,121, т.е. распределениям с более тяжелыми хвостами соответствуют меньшие средние значения $ПКЭ$.

$ПКЭ$ определяется следующим образом

$$K_P = \frac{Q}{P_{90} - P_{10}}, \quad (19)$$

где $Q = \frac{1}{2}(Q_3 - Q_1)$ – полуинтерквартильный интервал; P_{90} и P_{10} – соответственно 90-й и 10-й процентиля [7].

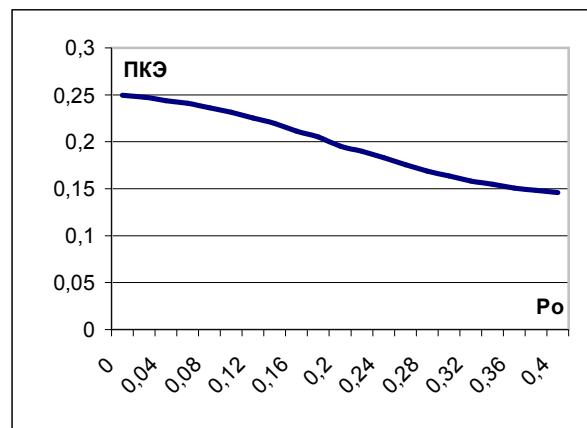


Рис. 2. Зависимость усредненных значений $ПКЭ$ от P_0 для ПРВ (12)

Таким образом, для синтезирования адаптивной оценки (17) необходимо знать пороговые значения $ПКЭ$, при которых одна из оценок \hat{X}_l начинает обладать меньшей дисперсией, чем другая.

Для этого необходимо заранее выбрать те оценки, которые будут использоваться, и их количество. Ниже рассмотрено два достаточно простых варианта с $L = 1$ или 2.

Один из исследованных методов адаптивного робастного оценивания параметра сдвига можно представить следующим образом (*adaptive1*):

$$X_{ADAPT1} = \begin{cases} X_{MED}, & K_p \in (0; \psi_0); \\ X_\alpha, & K_p \in [\psi_0; \psi_1); \\ X_{MEAN}, & K_p \in [\psi_1; +\infty), \end{cases} \quad (20)$$

где X_{MEAN} – среднее выборки; X_{MED} – медиана

выборки; X_α – оценка α -урезанного среднего выборки.

При этом был учтен тот факт (см. предыдущий раздел), что для малых значений P_0 наименьшую дисперсию среди используемых трех оценок обеспечивает среднее выборки. α -урезанное среднее характеризуется лучшей точностью при $P_0 \leq 0,4$. При больших значениях P_0 лучшие результаты обеспечивает применение медианы.

Пороговые значения ψ_0 и ψ_1 в (20) могут быть установлены экспериментальным путем, и для составного распределения (12) $\psi_0 = 0,19$ и $\psi_1 = 0,2485$.

Основой для еще одной адаптивной оценки (*adaptive2*) служат медиана выборки и оценка Ходжеса-Лемана. Адаптация к виду закона распределения в данном методе также осуществляется на основе анализа ПКЭ

$$X_{ADAPT2} = \begin{cases} X_{MED}, & K_P \in (0; \psi_0); \\ X_{HL}, & K_P \in [\psi_0; +\infty), \end{cases} \quad (21)$$

где X_{HL} – оценка Ходжеса-Лемана.

Рекомендуемый порог: $\psi_0 = 0,21$.

При этом учитывается, что для $P_0 \leq 0,32$ оценка HL характеризуется лучшей точностью по сравнению с медианой выборки, и, кроме того, обеспечивает меньшую дисперсию, чем оценка α -урезанного среднего, при $P_0 \leq 0,2$.

Существуют и другие методы адаптации, учитывающие масштаб данных и, в неявном виде, степень тяжести хвостов ПРВ. Один из известных методов робастного оценивания масштаба данных основан на расчете медианы модулей остаточных разностей (*ММОР*):

$$\begin{aligned} \text{ММОР} &= \\ &= \text{med} \{ |X_i - \text{med}(X_1, X_2, \dots, X_n)|, i = 1, \dots, n \}, \end{aligned} \quad (22)$$

где med – медиана выборки данных.

Аналогично гауссовой ПРВ, где с вероятностью 0,997 можно считать, что случайная величина, рас-

пределенная по данному закону, находится в интервале $[m - 3\sigma; m + 3\sigma]$ (где m – МО, а σ – СКО), так и в данном методе усреднению предлагается подвергать только те значения в выборке, которые попадают в интервал $[\text{med} - 3 \cdot \text{ММОР}; \text{med} + 3 \cdot \text{ММОР}]$:

$$X_{\text{ММОР}} = \frac{1}{N} \sum_{i=1}^n X_{(i)} \delta_i, \quad (23)$$

где

$$\delta_i = \begin{cases} 1, & \text{если } X_{(i)} \in [\text{med} - 3 \cdot \text{ММОР}; \text{med} + 3 \cdot \text{ММОР}]; \\ 0, & \text{если } X_{(i)} \notin [\text{med} - 3 \cdot \text{ММОР}; \text{med} + 3 \cdot \text{ММОР}], \end{cases} \quad (24)$$

$$N = \sum_{i=1}^n \delta_i; \quad (25)$$

n – общее число отчетов.

Предложенная оценка (23) – (25) (*AdММОР*) адаптируется к масштабу данных, а заодно и к степени “тяжести хвостов” распределения, поскольку грубые погрешности (выбросы) не участвуют в формировании значения оценки (23), так как они обычно находятся за пределами указанного выше интервала.

Адаптивной оценкой может являться также и мириада в случае, когда параметр k рассчитывается для каждой выборки (*AdMyr*), адаптируясь к масштабу данных следующим образом:

$$k = \gamma \text{ММОР}, \quad (26)$$

где γ – коэффициент пропорциональности, который при моделировании выбирался равным 3. При этом учитывался тот факт, что для гауссовой ПРВ *ММОР* и σ связаны соотношением $\sigma \approx 1,5 \text{ММОР}$ [7], и тогда для мириадной оценки для обеспечения $k \approx 2\sigma$ выражение (26) справедливо при $\gamma = 3$.

4. Анализ эффективности адаптивных методов оценивания при помощи численного моделирования

Статистические характеристики предложенных методов адаптивного робастного оценивания были исследованы путем численного моделирования для

трех видов плотности распределения вероятностей, указанных выше, с применением методики, описанной в разделе 2. Результаты, полученные для ПРВ (12), показаны на рис. 3, а в табл. 3 приведены значения дисперсий для разных оценок для ПРВ $\rho_2(x)$ и $\rho_3(x)$.

Анализ полученных данных (рис. 3) показывает, что все предложенные методы оценивания обеспечивают меньшую дисперсию, чем известная оценка ААТМЕ.

При малых P_0 применение адаптивной оценки (20) на основе ПКЭ (*adaptive1*) приводит к худшей точности, чем мириада с расчетом параметра k отдельно для каждой выборки (*AdMyr*), метод адаптивного оценивания *adaptive2* и оценка (23) на основе ММОР.

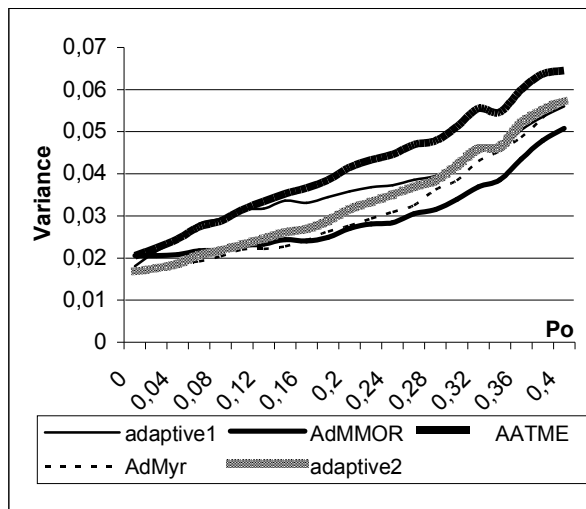


Рис. 3. Зависимость дисперсии адаптивных оценок от вероятности P_0 для составной ПРВ (12)

Специально проведенный анализ показал, что для оценки *adaptive1* это обусловлено следующим. Параметр K_p , как, впрочем, и другие параметры адаптации (β , ММОР), является случайной величиной. Но для адаптивных оценок (20) и (21) пороги устанавливаются с учетом средних значений K_p . Поэтому, например, при $P_0 = 0,04$ с достаточно большой вероятностью значения K_p для конкретных выборок данных могут превысить $\psi_1 = 0,2485$, и

тогда в соответствии с (20) находится среднее данной выборки, которое чувствительно к возможному наличию выбросов. Это, в результате, увеличивает дисперсию для оценки *adaptive1* и делает ее большей, чем для оценки *adaptive2* (21), для которой обе используемые оценки (Ходжеса-Лемана и медиана) обладают устойчивостью к выбросам.

При $P_0 > 0,3$ оценки *adaptive1*, *AdMyr* и *adaptive2* показывают практически одинаковые результаты. Наименьшие значения дисперсии при $P_0 > 16\%$ обеспечивает метод (23) на основе ММОР.

При обработке выборок случайных величин, имеющих ПРВ $\rho_2(x)$ и $\rho_3(x)$, обладающим наиболее высокой точностью оказался метод на основе оценки Ходжеса-Лемана и медианы (*adaptive2*, табл. 3). Вместе с тем, в рассматриваемых ситуациях все адаптивные оценки, включая *adaptive2*, характеризуются несколько большей дисперсией, чем медиана (данные в табл. 2). Причина этого заключается в упомянутых выше особенностях статистических свойств (случайного характера) параметра K_p .

Практически одинаковые результаты обеспечиваются при использовании оценок ААТМЕ и *AdMMOR*. Несколько хуже точности для оценок *adaptive1* и *AdMyr*.

Систематизируя результаты моделирования для всех трех плотностей распределения вероятностей, можно сделать вывод о том, что наилучшими среди рассмотренных пяти (одной известной и четырех предложенных) адаптивных оценок представляются *adaptive2* и *AdMMOR*.

Таблица 3

Зависимость дисперсии рассмотренных оценок при обработке случайных процессов $\rho_2(x)$ и $\rho_3(x)$

	<i>AdMyr</i>	<i>AdMMOR</i>	<i>adaptive1</i>	ААТМЕ	<i>adaptive2</i>
ρ_2	0,0066	0,0061	0,0068	0,0058	0,0048
ρ_3	0,0020	0,0016	0,0016	0,0016	0,0009

Заключення

Для ряда негауссовых плотностей распределения вероятностей проведен сравнительный анализ робастных оценок, принадлежащих различным классам (L , R и M).

Предложены два новых метода адаптации к виду закона распределения. На их основе получено четыре новых робастных оценки параметра сдвига, которые могут применяться при отсутствии четких сведений о виде и параметрах распределений. При этом предложенные оценки обладают точностными характеристиками, близкими к предельно достижимым.

В будущем планируется исследовать свойства предложенных оценок для других негауссовых законов распределения, симметричных относительно параметра сдвига.

Литература

1. Левин Б.Р. Теоретические основы статистической радиотехники. Кн. 1. – М.: Сов. радио, 1978. – 552 с.
2. Хьюбер Дж.П. Робастность в статистике: Пер. с англ. – М.: Мир, 1984. – 304 с.
3. Теория вероятностей: Учеб. для вузов / А.В. Печинкин, О.И. Тескин, Г.М. Цветкова и др.; Под ред. В.С. Зарубина и А.П. Крищенко. – М.: МГТУ им. Н.Э. Баумана, 2001. – 456 с.
4. Тоцкий А.В., Астола Я., Егиазарян К.О., Зеленский А.А., Курбатов И.В., Лукин В.В. Восстановление сигналов по оценкам биспектров в присутствии гауссовых и негауссовых помех // Зарубежная радиоэлектроника. – 2002. – № 11. – С. 44 – 58.
5. Доля А.М., Зеленський А.А., Лукин В.В. Оцінка затримки випадкових широкопосмугових сигналів при

малому співвідношенні сигнал/шум // Мережі і системи телекомунікації на залізничному транспорті. – 1999. – Вип. 35. – С. 87 – 95.

6. Шелухин О.И. Негауссовские процессы в радиотехнике. – М.: Радио и связь, 1999. – 287 с.

7. Suoranta R. Amplitude domain approach to digital filtering. Theory and applications of L-filters: Thesis ... doctor of technology: 10.11.95. – Espoo, Technical research centre of Finland, VTT, 1995. – 199 p.

8. Абрамов С.К., Зеленский А.А., Лукин В.В. Определение моды выборки данных с существенно несимметричным законом распределения на основе использования мириадной оценки и порядковых статистик // Радіотехнічні системи та засоби літальних апаратів. – Х. Нац. аерокосм. ун-т «ХАІ». – 2002. – № 29. – С. 74 – 83.

9. Totsky A.V., Roenko A.A., Lukin V.V., Zelensky A.A., Astola J., Egiazarian K. Combined bispectrum-median reconstruction of 1-D signal waveform // The 2004 International TICSP Workshop on Spectral Methods and Multirate Signal Processing (SMMSP 2004). – Vienna, Austria. – 2004. – P. 87 – 93.

10. Taguchi A. Adaptive α -trimmed mean filters with excellent detail-preserving // Proceedings of ICASSP. – 1994. – Vol. 5 – P. 61 – 64.

11. Lukin V.V., Ponomarenko N.N., Astola J.T., Saarinen K.P. Algorithms of Image Nonlinear Adaptive Filtering Using Fragment Recognition by Expert System // Proceedings of I@T/SPIE Symposium on Electronic Imaging: Science and Technology. – San Jose, CA, USA. – Feb. 1996. – SPIE Vol. 2662. – P. 179 – 190.

Поступила в редакцію 1.04.05

Рецензент: д-р техн. наук, проф. В.К. Волосюк, Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Харьков.