СИЭТ5-99

СУЧАСНІ ІНФОРМАЦІЙНІ ТА ЕНЕРГОЗБЕРІГАЮЧІ ТЕХНОЛОГІЇ ЖИТТЄЗАБЕЗПЕЧЕННЯ ЛЮДИНИ

ЗБІРНИК НАУКОВИХ ПРАЦЬ

ВИПУСК №5

Офіційні спонсори

W.J. Export - Import, INC.
КИЇВСЬКЕ ПРЕДСТАВНИЦТВО

Укрнафтагазбанк

УКР ЦСМ

Спеціальне видання міжнародного науково-технічного журналу ВОТТП

Спілка переробників зерна України

Міжнародний благодійний приватний фонд Сергія Сітька

УДК 681.3.01 : 629.5 Жуков Ю.Д., Гордеев Б.Н., Негометянов Ю.Б.
Украинский государственный морской
технический университет, Николаев

ПАРАМЕТРИЧЕСКИЙ МЕТОД ОЦЕНИВАНИЯ СИГНАЛОВ НА ФОНЕ НЕГАУССОВОЙ ПОМЕХИ

Задача оценивания сигналов на фоне негауссовой помехи часто встречается на практике. Например, для рефлектометрических систем типа «САДКО» качество оценивания параметров слабых трендов на фоне сигнальных выбросов определяет возможность выделения сигналов от слабых неоднородностей [1]. Цель настоящей работы - синтез метода оценивания, пригодного для выделения трендов на фоне негауссовой помехи.

Метод наименьших квадратов (МНК), предложенный К. Гауссом в 1813 году, актуален для нормально распределенных случайных величин - например, полученных суммированием, характерным для линейной фильтрации. Однако он неустойчив по отношению к таким факторам, как:

1) наличие промахов и искажений в исходных данных;

2) несоответствие закона распределения случайных величин гауссовой модели (в первую очередь - по асимметрии и эксцессу),

в чем можно убедиться путем поочередного отбрасывания точек, начиная с наиболее удаленных, и оценки производимого при этом воздействия.

Необходимость автоматического исключения абсурдных решений, получаемых по МНК вследствие неоднородности статистики, наличия промахов и искажений, привела к созданию робастных методов обработки [2,3]. Крайним случаем подобных оценок, часто используемым в качестве начального приближения других робастных методов, являются медианные оценки, оптимальные при наличии интенсивных импульсных помех. Медианная оценка является оценкой максимального правдоподобия (М-оценкой) для распределения Лапласа (двойного экспоненциального) [2,3], что определяет ее коренное отличие от усредненной, соответствующей гауссовому (нормальному) закону распределения помехи. Усеченные и виндзорированные оценки, получаемые отбрасыванием крайних по величине отсчетов выборки, реализуют компромисс между медианой и средним в непараметрическом варианте. К параметрическим вариантам компромисса можно отнести метод с коррекцией «хвостов» распределения по закону Лапласа [3], а также метод аппроксимации закона распределения на классе симметричных экспоненциальных распределений, к которому принадлежат распределения Гаусса и Лапласа [2]. Таким образом, робастные методы предполагают коррекцию исходных данных либо применяемой математической модели в целях достижения устойчивости оценок.

В настоящей работе описывается адаптивный параметрический метод оценивания, использующий М-оценки параметров случайной величины $y(x)$, определяемой дифференциальным уравнением для класса распределений Пирсона [4]:

$$y' = \frac{x-a}{b_0 + b_1 x + b_2 x^2} y \tag{1}$$

где

a, b_0, b_1, b_2 - вещественные параметры модели.

При этом моменты распределения нулевого и высших порядков m_n, если они существуют, связаны с параметрами модели соотношениями:

$$-am_n + nb_0 m_{n-1} + (n+1)b_1 m_n + (n+2)b_2 m_{n+1} = -m_{n+1}.$$

Подставляя $n = 0,1,2,3$, легко получить:

$$\begin{cases} -am_0 + b_1 m_0 + 2b_2 m_1 = -m_1; \\ -am_1 + b_0 m_0 + 2b_1 m_1 + 3b_2 m_2 = -m_2; \\ -am_2 + 2b_0 m_1 + 3b_1 m_2 + 4b_2 m_3 = -m_3; \\ -am_3 + 3b_0 m_2 + 4b_1 m_3 + 5b_2 m_4 = -m_4. \end{cases}$$

Если величина y стандартизована (т.е. проведены вычитание среднего и деление на среднеквадратичное отклонение σ), то ее моменты равны $m_1 = 0$, $m_2 = m_0 = 1$, $m_3 = \gamma_1$, $m_4 = \gamma_2$, а система уравнений имеет вид:

$$\begin{cases} b_1 = a; \\ b_0 + 3b_2 = -1; \\ 2b_1 + 4b_2\gamma_1 = -\gamma_1; \\ 3b_0 + 3b_1\gamma_1 + 5b_2\gamma_2 = -\gamma_2. \end{cases}$$

Подставляя решения полученной системы линейных уравнений в (1), можно получить уравнение вида

$$\frac{d\ln y}{dx} = \frac{(10\gamma_2 - 12\gamma_1^2 - 18)x - \gamma_1(9-\gamma_2)}{-4\gamma_2 + 3\gamma_1^2 + \gamma_1(9-\gamma_2)x + (6 - 2\gamma_2 + 3\gamma_1^2)x^2} \tag{2}$$

Третий нормированный момент γ_1 характеризует асимметрию закона распределения (для симметричных законов $\gamma_1 = 0$), а четвертый γ_2 - асимптотическое поведение («хвосты») этого закона. При этом для равномерного, нормального и двойного экспоненциального распределений $\gamma_2 = 1.8; 3; 6$ соответственно.

Таким образом, множество распределений Пирсона содержит как симметричные, так и асимметричные законы распределения, что является явным преимуществом перед вышеуказанными подходами.

Для анализа возможностей модели при анализе симметричных распределений подставим в (2) $\gamma_1 = 0$, $\gamma_2 = \varepsilon + 3$, где ε- коэффициент эксцесса:

$$\frac{d\,ln\,y}{dx} = -\frac{(5\varepsilon+6)x}{6+2\varepsilon+\varepsilon x^2} \qquad (3)$$

Интегрированием (3) легко получить:

$$ln\,y = -\frac{5\varepsilon+6}{2\varepsilon}ln(6+2\varepsilon+\varepsilon x^2),$$

$$y \sim (6+2\varepsilon+\varepsilon\cdot x^2)^{-2.5-3/\varepsilon}. \qquad (4)$$

Отметим, что формулы (3) и (4) позволяют восстановить при $\varepsilon = 0$ нормальный и при $\varepsilon = -1.2$ равномерный законы распределения.

При $\varepsilon > 0$ область определения плотности вероятности (4) неограничена. Например, для $\varepsilon = 3$ вместо ожидаемого двойного экспоненциального имеет место распределение вида $y \sim (4+x^2)^{3.5}$.

При $-3 < \varepsilon < 0$ функция плотности вероятности (4) существует для $|x| < \sqrt{2+6/\varepsilon}$, т.е. в ограниченном симметричном интервале. В случае $-1.2 > \varepsilon > 0$ функция плотности вероятности убывает к краям интервала, при $-3 < \varepsilon < -1.2$ - возрастает. Так, для $\varepsilon = -1$ $y \sim \sqrt{4-x^2}$, т.е. функция плотности вероятности представляет собой полуокружность. Близкое к этому значение эксцесса ($\varepsilon = -1.038$) имеют экспоненциальный закон распределения ($ln\,y = -|x|^7$) погрешностей градуировки шкал аналоговых вычислительных приборов, изготовленных в ряде стран Восточной Европы [2], а также оценки ($\gamma_1 = 0; -0.9 > \varepsilon > -1.3$) для аналогового выхода рефлектометрических систем типа «САДКО».

Таким образом, в рамках симметричного пирсоновского класса нормальный закон распределения ($\varepsilon = 0$) является пограничным случаем между финитными распределениями ($\varepsilon < 0$), при которых часть данных исключается, и распределениями с тяжелыми хвостами ($\varepsilon > 0$), при которых промахи получают ненулевые веса, что указывает на большую гибкость этого класса по сравнению с экспоненциальной моделью.

М-оценка вектора параметров $\vec{\alpha}$ для $x = h(i,\vec{\alpha}); y = f(x,\varepsilon)$ на выборке $x_i (i = 0 \ldots n-1)$ максимизирует отношение правдоподобия

$$R(\vec{\alpha}) = \prod_i \frac{f(x_i,\gamma_1,\gamma_2)}{f(x_i - h(i,\vec{\alpha}),\gamma_1,\gamma_2)}$$

или его логарифм

$$L(\vec{\alpha}) = \sum_i (ln\,f(x_i,\gamma_1,\gamma_2) - ln\,f(x_i - h(i,\vec{\alpha}),\gamma_1,\gamma_2)),$$

что достигается решением системы уравнений вида

$$\sum_i \frac{d\,ln\,f(x_i - h(i,\vec{\alpha}),\gamma_1,\gamma_2)}{dx}\cdot\frac{\partial h(i,\vec{\alpha})}{\partial\alpha_k} = 0, \qquad (5)$$

где α_k - компоненты вектора $\vec{\alpha}$. При этом параметры γ_1, γ_2 могут быть оценены либо путем решения системы (5) совместно с уравнениями

$$\sum_i \frac{d \ln f(x_i - h(i,\vec{\alpha}), \gamma_1, \gamma_2)}{d\gamma_s} = 0, \text{ где } s = 1,2,$$

либо прямой оценкой в рамках итерационной процедуры.

Так, при оценке среднего система (5) сводится к уравнению вида

$$\sum_i \frac{x_i - a}{(6+2\varepsilon)\sigma^2 + \varepsilon(x_i - a)^2} = 0. \qquad (6)$$

Для выборки $\{x_i\}$={0.96; 1.01; 0.97; 1.02; 1.04; 1.00; 10.52},содержащей выброс, при оцененных по выборке значениях $\sigma = 3.33$ и $\varepsilon = 3.027$ решение уравнения (6) дает значение $a = 1.53$, что явно лучше среднего арифметического $\bar{x} = 2.36$. Это дает возможность устранять промахи без привлечения других робастных оценок, т.е. с учетом реальной формы закона распределения. Изменением в формуле (5) функции $h(i,\vec{\alpha})$, описывающей форму тренда, можно получить систему уравнений для тренда требуемого порядка.

Проведенные оценки и обоснования показывают, что предложенный метод оценивания параметров случайных сигналов на базе множества распределений Пирсона обладает большой гибкостью и пригоден для оценки трендов сигнала с учетом искажений сигнальной информации и возможности грубых промахов в исходной информации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Жуков Ю.Д., Гордеев Б.Н., Негометянов Ю.Б., Логвиненко Ю.И. Компенсация трендов отраженного сигнала в судовых полиметрических системах: // Автоматизация судовых технических средств: науч.-техн. сб. –1998. –Вып. 3. Одесса: ОГМА. – 260 с.

2. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений.–2-е изд., перераб. и доп. –Л.: Энергоатомиздат. Ленингр. отд-ние, 1991. – 304 с.: ил.

3. Устойчивые статистические методы оценки данных/Под ред. Р. Л. Лонера, Г. Н. Уилкинсона/Пер. с англ.–М.: Машиностроение, 1984. –232 с.

4. K. Pearson, Philos. Trans. Roy. Soc. London, Ser. A, 185, 71–110 (1894).