

потезу о виде распределений неисключенных систематических погрешностей $\{w(\Delta_{\text{сист}}\lambda_i)\}_{i=1}^N$.

Л и т е р а т у р а

1. BIPM. [Официальный сайт]. <http://www.bipm.org/> (дата обращения 18.02.2014 г.).
2. Report on CIPM key comparison of the second phase multiples and submultiples of the kilogram (CCM.M-K5). [Электронная версия] http://kcdb.bipm.org/appendixb/appbresults/ccm.m-k5/ccm.m-k5_final_report.pdf (дата обращения 18.02.2014 г.).
3. Чуновкина А. Г. Оценивание данных ключевых сличений национальных эталонов. СПб.: НПО «Профессионал», 2009.

4. Цветков Э. И. Анализ результатов сличения двух эталонов // Измерительная техника. 2013. № 4. С. 11—14; Tsvetkov E. I. Analysis of the results of a comparison of two standards // Measurement Techniques. 2013. V. 56. N. 4. P. 368—371.

5. Сулоева Е. С. Возможности имитационного моделирования при сличении двух эталонов // Вестник СЗО МА. 2012. Вып. 27. С. 55—64.

6. Сулоева Е. С. Исследование вероятностей ошибок первого и второго рода от состава априорных знаний и устанавливаемых условий // Вестник СЗО МА. 2012. Вып. 28. С. 48—51.

7. Сулоева Е. С. Особенности принятия решения о годности сличаемых эталонов при стандартном составе априорных знаний // Изв. ВолгГТУ. Серия электроника, измерительная техника, радиотехника и связь. 2013. № 23. С. 78—81.

Дата принятия 18.04.2014 г.

519.24

Анализ свойств выбора оптимального количества интервалов дискретизации области определения плотности вероятности

А. В. ЛАПКО*, В. А. ЛАПКО**

* Институт вычислительного моделирования СО РАН, Красноярск, Россия

** Сибирский государственный аэрокосмический университет им. ак. М. Ф. Решетнева, Красноярск, Россия, e-mail: lapko@icm.krasn.ru

Исследован выбор оптимального количества интервалов дискретизации области определения плотности вероятности в зависимости от коэффициента контрэксцесса. Установлены количественные зависимости между коэффициентами дискретизации и контрэксцесса для симметричных и несимметричных законов распределения. Проведен анализ их особенностей для различных условий использования.

Ключевые слова: плотность вероятности, непараметрическая оценка, дискретизация, коэффициенты эксцесса, контрэксцесса, асимметрии.

The properties of optimal selection of quantity of discretization intervals of domain of probability density have been studied. The quantitative dependences between the coefficients of discretization procedures and antikurtosis coefficients for symmetric and asymmetric distribution laws have been determined. The analysis of their features for different usage conditions has been carried out.

Key words: probability density, nonparametric estimation, discretization procedure, kurtosis coefficient, antikurtosis coefficient, asymmetry coefficient.

Выбор оптимального количества интервалов дискретизации области значений случайной величины x зависит от вида плотности вероятности $p(x)$ [1], поэтому в условиях априорной неопределенности о виде $p(x)$ возникает задача уточнения процедуры дискретизации на основании параметров закона распределения случайной величины.

Пусть имеется выборка $V = (x^i, i = \overline{1, n})$ из n независимых значений одномерной случайной величины x с неизвестной плотностью вероятности $p(x)$. Разобьем область определения $p(x)$ на N непересекающихся интервалов длиной 2β , и сформируем множества случайных величин $X^j, j = \overline{1, N}$. В качестве характеристик X^j примем частоту $\bar{p}^j$ попадания случайной величины x в j -й интервал и его центр z^j . На осно-

ве полученной информации определим массив данных $V_1 = (z^j, \bar{p}^j = \bar{P}^j / (2\beta), j = \overline{1, N})$, составленный из центров z^j интервалов и соответствующих им значений оценок $\bar{p}^j$ плотности вероятности.

В качестве приближения по эмпирическим данным V_1 плотности вероятности $p(x)$ примем ее непараметрическую оценку [2, 3]:

$$\bar{p}(x) = c^{-1} \sum_{j=1}^N \bar{P}^j \Phi\left[(x - z^j)/c\right], \quad (1)$$

где положительные и нормированные ядерные функции $\Phi[(x - z^j)/c]$ удовлетворяют условию симметричности относительно значений z^j [4, 5].

Коэффициенты размытости c в выражении (1) характеризуют область определения ядерных функций. Их оптимальный выбор осуществляется из условия минимума относительной ошибки аппроксимации

$$\bar{W} = \frac{1}{N} \sum_{t=1}^N \left[\left(\bar{P}^t / (2\beta) \right) - \tilde{p}(z^t) \right]^2,$$

$$\text{где } \tilde{p}(z^t) = c^{-1} \sum_{\substack{j=1 \\ j \neq t}}^N \bar{P}^j \Phi \left[(z^t - z^j) / c \right].$$

Из условия минимума асимптотического выражения среднего квадратического отклонения $\bar{p}(x)$ от $p(x)$ получим процедуру выбора оптимального количества интервалов дискретизации [1]:

$$\bar{N} = k \sqrt{n}, k = \sqrt{\Delta \|p(x)\|^2}, \quad (2)$$

где $\|p(x)\|^2 = \int_{-\infty}^{+\infty} p^2(x) dx$; Δ — длина интервала области определения $p(x)$.

Процедура дискретизации (2) определяется видом восстанавливаемой плотности вероятности, областью ее определения и объемом n исходных статистических данных. Структура зависимости количества $\bar{N}$ интервалов дискретизации от n (2) близка к формуле Хайнхольда—Гаеде [6].

В условиях неполной информации о виде плотности вероятности определение k в выражении (2) возможно оцениванием функционала $\|p(x)\|^2$ либо зависимости k от количественных характеристик закона распределения случайной величины [7, 8].

В данной работе установлена и исследована зависимость между коэффициентами дискретизации и контрэксцесса закона распределения случайной величины.

Влияние вида плотности вероятности на значения оптимального количества интервалов дискретизации. Опре-

делим значения $k = \sqrt{\Delta \|p(x)\|^2}$ для ряда законов распределения случайных величин. Полученные результаты сравним с коэффициентами эксцесса, контрэксцесса и асимметрии [9].

Контрэксцесс определяется выражением

$$\chi = 1 / \sqrt{\eta},$$

где η — параметр эксцесса, определяемый как $\eta = \mu_4 / \sigma^4$, σ — среднее квадратическое отклонение, μ_4 — четвертый центральный момент случайной величины x .

Значение коэффициента асимметрии рассчитывается как отношение центрального момента третьего порядка μ_3 к среднему квадратическому отклонению σ в третьей степени:

$$\gamma_3 = \mu_3 / \sigma^3.$$

При вычислении k , χ , γ_3 границы a , b области определения $p(x)$ находят из условия $\int_a^b p(x) dx = 0,997$, позволяющего

избежать проблемы вычисления k при бесконечной области определения $p(x)$, и создающего объективную основу сравнения коэффициентов k , χ , γ_3 законов распределения.

В качестве примера вычислим k для равномерного закона распределения:

$$p(x) = 1/(b-a) \quad \forall x \in [a, b]. \quad (3)$$

В этих условиях $\Delta = b - a$, а $\|p(x)\|^2 = (b-a)^{-1}$. Отсюда следует, что для случайной величины x с плотностью вероятности (3) $k = 1$ и не зависит от параметров $p(x)$. Установленная закономерность свойственна и другим законам распределения, к которым относятся плотности вероятности типа Гаусса, Лапласа и ряда других (см. табл. 1).

Таблица 1

Значения коэффициентов эксцесса, контрэксцесса и k для симметричных законов распределения случайных величин

Вид распределения	Коэффициенты		
	эксцесса	контрэксцесса	k
Нормальный	-0,173	0,595	1,294
Лапласа	1,788	0,457	1,704
Равномерный	-1,195	0,744	0,997
Параболический	-0,864	0,684	1,070
Кубическая парабола	-0,618	0,648	1,145
Треугольный	-0,614	0,647	1,123

Если плотность вероятности случайной величины x меняет свой вид при изменении ее параметров, то данная тенденция сопровождается изменением k . Например, k для логнормального закона распределения

$$p(x) = (2\pi)^{-1/2} \sigma x \exp [-(\ln x - \mu)/(2\sigma^2)]$$

не зависит от параметра μ , но его значения определяются значениями параметра σ . С ростом σ в интервале [1—2] коэффициент k возрастает с 2,37 до 7,97.

Отмеченная особенность справедлива и для распределения Эрланга:

$$p(x) = (m\alpha)^m x^{m-1} \exp(-\alpha mx) / \Gamma(m),$$

где параметр $\alpha > 0$, m — натуральное число, $\Gamma(m)$ — гамма-функция. Для данного закона распределения k не зависит от α , а его значения определяются параметром m .

Симметричные плотности распределения $p(x)$ для $x \in [M(x) - a, M(x) + a]$ и их нормированные фрагменты $p_1(x) \forall [M(x), M(x) + a]$ характеризуются одним и тем же значением k . Здесь $M(x)$ — математическое ожидание случайной величины x . Например, плотности вероятности треугольного и линейного вида определяются значением $k = 1,12$; законам же распределения Лапласа и экспоненциального вида соответствует $k = 1,7$.

Оценивание k по значениям коэффициента контрэксцесса. На основании информации $V_1 = (k^i, \chi^i, i=1, n_1)$, $n_1 = 17$, представленной в табл. 2, восстановим зависимость

$k = \varphi_1(\chi)$ для несимметричных законов распределения. Из анализа информации табл. 1, 2 следует, что свойства коэффициентов k , χ являются близкими. Поэтому возникает задача оценивания k по χ в приведенных выше условиях.

Таблица 2

Значения коэффициентов эксцесса, контрэксцесса и k для несимметричных законов распределения случайных величин

№ п/п	Вид распределения	Коэффициенты			
		эксцесса	контрэксцесса	асимметрии	k
1	Логнормальный $\sigma = 1$	11,913	0,259	2,951	2,378
2	Логнормальный $\sigma = 1,25$	20,448	0,206	3,883	3,217
3	Логнормальный $\sigma = 1,5$	31,651	0,170	4,867	4,506
4	Логнормальный $\sigma = 1,75$	40,278	0,152	5,564	6,158
5	Логнормальный $\sigma = 2$	41,582	0,150	5,748	7,977
6	Экспоненциальный	3,139	0,404	1,662	1,704
7	Линейный	-0,627	0,649	0,562	1,123
8	Нормированный фрагмент параболического $x \geq M(x)$	-0,847	0,681	0,379	1,070
9	Нормированный фрагмент нормального $x \geq M(x)$	0,360	0,546	0,898	1,294
10	Нормированный фрагмент кубической параболы $x \geq M(x)$	-0,490	0,631	0,586	1,145
11	Эрланга $m = 5$	0,459	0,538	0,759	1,349
12	Эрланга $m = 3$	0,916	0,505	0,977	1,362
13	Эрланга $m = 2$	1,486	0,472	1,192	1,415
14	Хи-квадрат с числом степеней свободы $m_1 = 8$	0,630	0,525	0,848	1,349
15	Хи-квадрат $m_1 = 6$	0,917	0,505	0,977	1,363
16	Хи-квадрат $m_1 = 4$	1,486	0,472	1,192	1,415
17	Хи-квадрат $m_1 = 2$	3,139	0,404	1,662	1,704

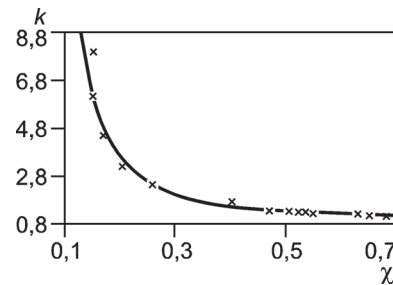
По результатам вычислительных экспериментов установим вид зависимости (см. рисунок)

$$\bar{k} = \bar{\varphi}_1(\chi) = \lambda_0 \exp(\chi^{-\lambda_1}) + \lambda_2. \quad (4)$$

Из условия минимума эмпирической ошибки аппроксимации

$$\bar{W}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \left| [k^i - \bar{\varphi}_1(\chi^i)] / k^i \right|$$

определим оптимальные параметры $\lambda_0 = 0,16$, $\lambda_1 = 0,67$, $\lambda_2 = 0,58$ зависимости (4). Этим параметрам соответствует $\bar{W}_1 = 0,047$.



Зависимость коэффициента k в процедуре дискретизации от контрэксцесса χ для несимметричных законов распределения: значения данных табл. 2

Приняв за основу аппроксимацию типа (4), оценим зависимость $k = \varphi_2(\chi)$ для симметричных законов распределения. При оптимальных параметрах $\bar{\lambda}_0 = 0,24$, $\bar{\lambda}_1 = 0,8$, $\bar{\lambda}_2 = 0,15$ средняя относительная погрешность аппроксимации $\bar{W}_2 = 0,013$.

Установленная зависимость (4) является устойчивой к изменению количества плотностей вероятности, используемых при ее оценивании. Исключим из табл. 2 информацию о плотностях вероятности № 11—17. Тогда при оценивании зависимости (4) ее параметры принимают значения $\lambda_0 = 0,14$, $\lambda_1 = 0,69$, $\lambda_2 = 0,59$. Им соответствует относительная погрешность аппроксимации $\bar{W}_1 = 0,053$.

С ростом χ значения k в (4) уменьшаются. Отметим, что χ косвенно характеризует островершинность $p(x)$ в области ее модального значения. Большим значениям χ свойственны более плоские вершины $p(x)$. В этих условиях уменьшение k сопровождается снижением количества $\bar{N}$ интервалов дискретизации области определения $p(x)$.

Если $\chi > 0,5$, то k из (2) принимает значения в интервале $[1,07; 1,36]$. Тогда, при числе наблюдений $n = 100$ случайной величины x , количество интервалов дискретизации $\bar{N} \in [10; 14]$. Этим условиям соответствуют плотности вероятности № 7—12, 14, 15 (см. табл. 2). Причем увеличение коэффициента асимметрии сопровождается ростом количества $\bar{N}$ интервалов дискретизации области определения $p(x)$. Данный вывод подтверждается анализом информации табл. 1.

Для симметричных законов распределения ($\gamma_3 = 0$) $k \in [1; 1,3]$. Исключение составляет лишь закон распределения Лапласа, для которого $k = 1,7$, $\chi = 0,457$. С уменьшением коэффициента контрэксцесса и одновременным увеличением коэффициента асимметрии γ_3 условия оценивания закона распределения усложняются. Поэтому возникает необходимость повышения количества $\bar{N}$ интервалов дискретизации области значений случайной величины.

Выводы. Коэффициент k процедуры оптимального выбора количества интервалов дискретизации определяется нелинейным функционалом от плотности вероятности случайной величины. Для симметричных плотностей вероятности его значения являются постоянными. Это характерно, например, для законов распределения Гаусса, Лапласа, равномерного и треугольного. Данная закономерность свойственна также для несимметричных плотностей вероятности, вид которых не зависит от изменения параметров закона распределения (например, линейная и экспоненциальная

плотности вероятности). Между свойствами коэффициентов k и контрэксцесса χ существует нелинейная зависимость, позволяющая по χ оценивать коэффициент k процедуры дискретизации.

При коэффициенте контрэксцесса $\chi > 0,5$ значения коэффициента k процедуры дискретизации изменяются в относительно малом интервале. В подобных условиях количество интервалов дискретизации области определения плотностей вероятности отличаются незначительно. С уменьшением χ условия восстановления плотности вероятности усложняются, особенно при больших значениях коэффициента асимметрии. При этом значение k и количество интервалов дискретизации увеличиваются.

Полученные результаты имеют важное значение в задачах оценивания плотностей вероятности и идентификации законов распределения случайных величин.

Л и т е р а т у р а

1. Лапко А. В., Лапко В. А. Оптимальный выбор количества интервалов дискретизации области изменения одномерной случайной величины при оценивании плотности вероятности // Измерительная техника. 2013. № 7. С. 24—27; Lapko A. V., Lapko V. A. Optimal selection of the number of sampling intervals in domain of variation of a one-dimensional random variable in estimation of the probability density // Measurement Techniques. 2013. V. 56. N. 7. P. 763—767.

2. Лапко А. В., Лапко В. А. Непараметрические методики анализа множеств случайных величин // Автометрия. 2003. Т. 39. № 1. С. 54—61.

3. Lapko A. V., Lapko V. A. Non-parametric Analysis Techniques of Random Values sets // Opt., Instrument. Data Proces. 2003. Т. 39, № 1. P. 44-50.

4. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistic. 1962. Т. 33, № 3. P. 1065—1076.

5. Епанечников В. А. Непараметрическая оценка многомерной плотности вероятности // Теория вероятности и ее применения. 1969. Т. 14. № 1. С. 156—161.

6. Heinhold I., Gaede K. W. Ingenieur statistic. München: Wien, Springer Verlag, 1964.

7. Лапко А. В., Лапко В. А., Егорочкин И. А. Непараметрические оценки смеси плотностей вероятности и их применение в задаче распознавания образов // Системы управления и информационные технологии. 2009. Т. 35. № 1. С. 60—64.

8. Лапко А. В., Лапко В. А. Анализ дисперсии среднеквадратической ошибки аппроксимации непараметрической оценки плотности вероятности ядерного типа // Информатика и системы управления. 2012. Т. 33. № 3. С. 132—139.

9. Гмурман В. Е. Теория вероятностей и математическая статистика. М.: Высшая школа, 1999.

Дата принятия 22.04.2014 г.

ИЗМЕРЕНИЯ В ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ

004.056:061.68.50

Метод помехоустойчивого кодирования информации каналов передачи данных телеизмерительных информационных систем с коррекцией ошибок в двух байтах информации

А. А. ПАВЛОВ*, А. Н. ЦАРЬКОВ*, П. А. ПАВЛОВ*, Д. Е. СОРОКИН*, А. В. ГУСЕВ*, К. В. ГУСЕВ**, М. И. ЛАСЯК**

* Институт инженерной физики, Серпухов, Россия, e-mail: pavlov_iif@mail.ru

** Институт информатизации образования РАО, Москва, Россия

Предложен метод повышения помехоустойчивости каналов передачи данных телеизмерительных информационных систем путем коррекции ошибок в двух байтах информации при минимальных временных затратах.

Ключевые слова: корректирующие алгебраический и циклические коды, одиночные и двойные байты ошибок.

A method of noise immunity increase for telemetric information systems data transmission channels at the expense of errors correction in two bytes of information at minimal time costs.

Key words: correcting algebraic code, correcting cyclic code, single and double bytes error.

Информационно-измерительная система (ИИС) является совокупностью функционально объединенных измерительных, вычислительных и других вспомогательных технических средств для получения информации [1, 2]. Отличительная особенность телеизмерительных информационных

систем (ТИИС) — длина канала связи. Канал связи является наименее надежной частью этих систем, поэтому обеспечение надежности и достоверности передачи данных для ТИИС является важной и ответственной задачей [1, 2].