УДК 519.24

ВОССТАНОВЛЕНИЕ ДВУМЕРНЫХ ПЛОТНОСТЕЙ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ РЯДАМИ ОРТОНОРМИРОВАННЫХ ПОЛИНОМОВ

к. т. н. Лысенко Н. А.
ГВУЗ «Днепропетровский национальный университет им. О. Гончара», г. Днепропетровск

Аналитические описания статистических закономерностей случайных параметров объектов или явлений в виде законов распределения вероятностей измерений необходимы при проектировании технологий неразрушающего контроля и технической диагностики. Задача восстановления плотностей распределения вероятностей не является новой [4-6,8] и на сегодняшний день разработаны разнообразные методы для ее решения. Среди непараметрических методов оценивания выделяют сплайны [5], ядерные [4,6] и проекционные [8] оценки. Метод оценивания на основе рядов ортонормированных полиномов относится проекционным оценкам и включает в себя набор различных видов полиномов. В [2] рассмотрен данный подход к восстановлению одномерных законов распределения по экспериментальным данным. В [3] исследованы ошибки восстановления, установлено влияние размера выборки измерений, вида ортонормированного полинома и количества членов ряда на величину ошибок, предложено правило для ограничения количества членов ряда. Однако, объект контроля может характеризоваться несколькими параметрами, поэтому практическую значимость имеет разработка модели многомерного закона распределения вероятностей на основе выборок экспериментальных данных, возможно, связанных между собой корреляционными связями. В данной работе рассматривается модель двумерного закона распределения вероятностей, как частный случай многомерной модели.

Постановка задачи

Пусть $|x_1| = |x_{11}x_{12}..x_{1n}|$ и $|x_2| = |x_{21}x_{22}..x_{2n}|$ - выборки измерений коррелированных случайных величин с неизвестным законом распределения вероятностей. Требуется восстановить двумерную плотность распределения вероятностей рядами ортонормированных полиномов.

Согласно [3], оценка одномерного закона распределения вероятностей на основе рядов ортонормированных полиномов имеет вид:

$$W^*(x) = A\sum_{i=0}^{N} b_i^* \sqrt{q(x)}Q_i(x),$$

где $Q_i(x)$ - ортонормированный полином; i - порядок полинома; $q(x)$ - весовая функция; коэффициенты b_i^* вычисляются по формуле

$$b_i^* = \frac{1}{n}\sum_{k=1}^{n} \sqrt{q(x_k)}Q_i(x_k);$$

A - нормирующий множитель, определяется из условия равенства единице площади под кривой $W^*(x)$

$$A = \left(\sum_{i=0}^{N} b_i^* q_i\right)^{-1}, \; q_i = \int_{-\infty}^{\infty} \sqrt{q(x)}Q_i(x)dx.$$

Рассмотрим два способа восстановления двумерного закона распределения вероятностей рядами ортонормированных полиномов.

Основная часть

Поскольку общей развернутой теории ортогональных многочленов от нескольких переменных не существует [7], модель двумерного закона распределения можно представить, как произведение двух одномерных моделей:

$$W^*(x_1, x_2) = A_1 A_2 \sqrt{q_1(x_1)q_2(x_2)} \sum_{i=0}^{N_1} b_{1i}^* Q_1(i, x_1) \sum_{j=0}^{N_2} b_{2j}^* Q_2(j, x_2) \quad (1)$$

Эффективность восстановления двумерных плотностей распределения вероятностей на основе данной модели (1) проверялась путем проведения вычислительных экспериментов. В данной статье приведены результаты восстановления нормального закона распределения с использованием ортонормированных полиномов Эрмита. Ошибки восстановления вычислялись по формуле

$$\delta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(W^*(x_1, x_2) - W(x_1, x_2)\right)^2 dx_1 dx_2, \quad (2)$$

где $W(x_1, x_2)$ - двумерный нормальный закон распределения, с параметрами a_1, a_2, σ_1, σ_2 и r. В ходе вычислительного эксперимента генерировались выборки двумерных нормальных случайных величин с этими параметрами и вычислялись ошибки восстановления. Средние значения ошибок восстановления по результатам 100 опытов приведены в табл. 1.

Таблица 1

Средние значения ошибок восстановления двумерных законов распределения вероятностей (способ 1)

n \ r	**0.1**	**0.5**	**0.9**
25	0,0065	0,014	0,087
50	0,0031	0,0096	0,086
100	0,0018	0,0085	0,084

Из табл. 1 видно, что с ростом значений коэффициента корреляции значения средних ошибок восстановления увеличиваются на порядок. Стремительный рост ошибок показывает, что данная модель не учитывает взаимное влияние выборок.

Второй способ заключается в декорреляции случайных величин $|x_1|$ и $|x_2|$ путем преобразования [1]

$$y_1 = (x_1 - a_1)\cos\alpha + (x_2 - a_2)\sin\alpha\ ,$$
$$y_2 = -(x_1 - a_1)\sin\alpha + (x_2 - a_2)\cos\alpha\ ,$$

где $tg2\alpha = \dfrac{2r\sigma_1\sigma_2}{\sigma_1^{\ 2} - \sigma_2^{\ 2}}$, а r - коэффициент корреляции. Двумерный закон распределения вероятностей восстанавливается рядами ортонормированных полиномов по некоррелированным выборкам $|y_1|$ и $|y_2|$ в соответствии с формулой (1). При переходе к новым случайным величинам неизвестные параметры заменяются их оценками: выборочным коэффициентом корреляции r^*, средними значениями $\bar{y}_1$, $\bar{y}_2$ и выборочными дисперсиями D^*_{Y1}, D^*_{Y2}. Ошибки восстановления законов распределения вероятностей предлагаемым способом оценивались по результатам вычислительного эксперимента. Эксперимент повторялся 100 раз для каждого конкретного сочетания параметров и определялись ошибки восстановления. Затем вычислялись средние значения ошибок. Результаты исследования приведены в табл. 2. Анализируя данные табл. 2, отметим, что с усилением корреляционных связей ошибки по-прежнему растут, но значительно медленнее, чем в первом случае.

Таблица 2

Средние значения ошибок восстановления двумерных законов распределения вероятностей (способ 2)

n \ r	**0.1**	**0.5**	**0.9**
25	0.0071	0.0094	0.019
50	0.0029	0.004	0.0079
100	0.0019	0.0022	0.0043

На рис. 1 показаны средние ошибки восстановления для различных значений коэффициента корреляции ($r = 0.1,\ 0.5,\ 0.9$) при использовании двух рассмотренных способов восстановления на примере коротких выборок ($n = 25$).

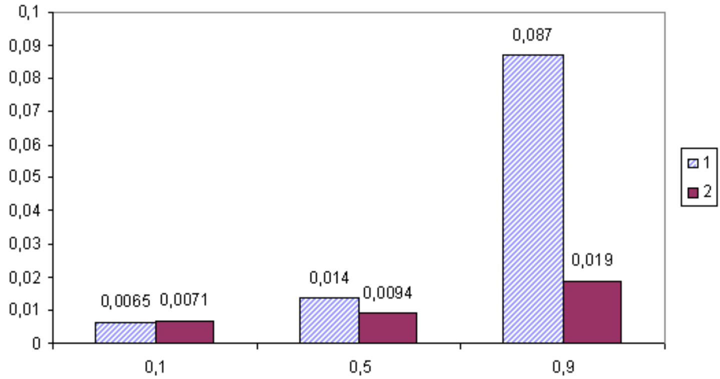

Рис. 1. Средние значения ошибок восстановления, полученные (1) – первым и (2) – вторым способом при использовании коротких выборок ($n = 25$)

Таким образом, добавление процедуры декореляции выборок исходных данных к двумерной модели плотности распределения (1) позволяет существенно снизить ошибки восстановления и тем самым расширить границы применимости данной модели.

Выводы

В ходе исследования предложены два способа решения задачи восстановления двумерных законов распределения вероятностей:

1) двумерная модель плотности распределения вероятностей на основе рядов ортонормированных полиномов, демонстрирует высокую точность оценивания в случае независимости выборок исходных данных, однако, при наличии сильной корреляции ($r \geq 0{,}5$) ошибки восстановления резко возрастают;
2) способ, использующий эту же модель, но добавляющий в алгоритм оценивания процедуру декорреляции исходных данных, оказался более эффективным в условиях коррелированности выборок; усложнение алгоритма восстановления компенсируется значительным снижением ошибок восстановления.

Восстановленные двумерные плотности распределения вероятностей могут быть использованы для построения информационно-измерительных технологий неразрушающего контроля и технической диагностики.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240с.
2. Красников В.Н., Лещенко А.Б. Моделирование законов распределения случайных величин по эмпирическим данным // Радіоелектронні і комп'ютерні системи. – 2005. – №4(12) – С. 108-111.
3. Малайчук В.П., Петренко А.Н., Лысенко Н.А. Восстановление законов распределения вероятностей рядами ортонормированных полиномов // Системні технології. Регіональний міжвузівський збірник наукових праць. – 2005. – Випуск 3 (38) – С. 73-88.
4. Parzen E. On estimation of a probability density function and mode // Annals Math. Statist. – 1962. - Vol.33, №3. – P.1065-1076.
5. Приставка А.Ф., Райко О.В. Гистосплайны. – Днепропетровск: Изд-во ДГУ, 1991. – 136с.
6. Rosenblatt M. Remarks on some non-parametric estimates of a density function // Annals Math. Statist. – 1956. – Vol.27, №3. – P. 832-837.
7. Осадчий В. И., Зюзько Т.И., Орлов В.П. Применение многомерных рядов Эрмита для представления информации о распределении физических полей объектов // Электронный журнал «Труды МАИ». – http: // **www.mai.ru**/science/trudy/.
8. Ченцов Н.Н. Оценка неизвестной плотности распределения по наблюдениям // Доклады АН СССР. – 1962. – Т. 147, №1. – С. 45-48.