УДК 621.391

В.Н. КРАСНИКОВ, А.Б. ЛЕЩЕНКО

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

РЕШЕНИЕ ЗАДАЧИ НАХОЖДЕНИЯ АЛГОРИТМА ОЦЕНИВАНИЯ ПАРАМЕТРОВ ПО ЭМПИРИЧЕСКИМ ДАННЫМ ДЛЯ ПРОИЗВОЛЬНОЙ СИММЕТРИЧНОЙ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Рассмотрена задача построения алгоритма оценивания интересующих исследователя параметров непосредственно по выборочным данным для случая произвольной симметричной функции плотности распределения вероятностей ошибок измерений.

алгоритм оценивания, опытные данные, закон распределения случайных величин, плотность распределения, метод максимального правдоподобия

В случаях математической обработки различного рода эмпирических данных и их анализа, независимо от класса анализируемых задач – технических, экономических, социологических и др. – доверительные оценки как средних значений, так и дисперсий, основаны на гипотезе нормальности закона распределения случайных ошибок измерений и поэтому могут применяться лишь до тех пор, пока результаты эксперимента не противоречат этой гипотезе. Обоснованный выбор распределения, используемого при решении конкретной прикладной задачи, совсем не очевиден. При анализе ошибок измерений, например, их рассматривают как нормально распределенные случайные величины. Теоретическим основанием этому служит центральная предельная теорема теории вероятностей, согласно которой сумма большого числа независимых случайных величин имеет распределение, близкое к нормальному. Это утверждение справедливо при некоторых дополнительных условиях, которые сводятся к требованию, чтобы в состав суммы не входили отдельные слагаемые, явно преобладающие над другими и распределенные не по нормальному закону.

Опуская обсуждение ситуации, когда результаты эксперимента вызывают сомнения относительно закона распределения случайных величин, для решения вопроса о пригодности или непригодности нормального закона распределения надо производить достаточно большое число измерений (использовать достаточно большую выборку данных) и применять критерий соответствия χ^2, который требует значительных расчетов, либо строить гистограммы для оценки соответствия фактического распределения нормальному.

В работе [1] была поставлена и решена задача представления законов распределения случайных величин в виде рядов по системам ортонормированных многочленов, коэффициенты которых определяются через числовые характеристики самих эмпирических данных, что позволяет их использовать в виде моделей распределений.

Поставим задачу отыскания алгоритма оценивания интересующих исследователя параметров непосредственно по выборочным данным для случая произвольной симметричной функции плотности распределения вероятностей ошибок измерений.

Пусть имеется выборка наблюдений $x_1, x_2, \ldots, x_n$, распределение генеральной совокупности которой равно $f(\bar{x},\theta)$, где θ – неизвестный оцениваемый параметр.

Найдем для указанной выборки следующие моменты:

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; S_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - S_1)^2 \; ;$$

$$S_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - S_1)^4 \ ; \ S_{2k-2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - S_1)^{2k-2} \ , \quad (1)$$

для $i = \overline{1,n}$ и k, которое будет определено ниже.

Рассмотрим произвольную функцию плотности распределения вида

$$f_k(x) = \frac{1}{\sqrt{2\pi S_2}} \exp\left(-\frac{(x-S_1)^2}{2S_2}\right) \times$$
$$\times \left[1 + \sum_{i=1}^{k} A_i (x - S_1)^{2i}\right]. \quad (2)$$

Определим коэффициенты A_i, $i = \overline{1,k}$ из (2) так, чтобы

$$\int_{-\infty}^{\infty} f_k(x)\, dx = 1 \ ; \ \int_{-\infty}^{\infty} x f_k(x)\, dx = S_1 \ ;$$

$$\int_{-\infty}^{\infty} (x - S_1)^2 f_k(x)\, dx = S_2 \ ;$$

$$\int_{-\infty}^{\infty} (x - S_1)^{2j} f_k(x) dx = S_{2j}\,,\ j \in [2, k-1]. \quad (3)$$

Соотношения (3) приводят к уравнениям относительно A_i. Прежде всего

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi S_2}} \exp\left(-\frac{(x-S_1)^2}{2S_2}\right)\left[1 + \sum_{i=1}^{k} A_i (x-S_1)^{2i}\right] dx = 1\,,$$

откуда следует

$$\frac{1}{\sqrt{2\pi S_2}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2S_2}\right) dt +$$
$$+ \sum_{i=1}^{k} A_i \frac{1}{\sqrt{2\pi S_2}} \int_{-\infty}^{\infty} t^{2i} \exp\left(-\frac{t^2}{2S_2}\right) dt = 1. \quad (4)$$

Воспользуемся известным значением интеграла [2]

$$\int_{-\infty}^{\infty} t^{2i} \exp\left(-\frac{t^2}{2S_2}\right) dt = (2i-1)!!\, S_2^i \sqrt{2\pi S_2}$$

и получим из (4)

$$\sum_{i=1}^{k} A_i (2i-1)!!\, S_2^{i-1} = 0. \quad (5)$$

Второе соотношение в (3) приводит к тривиальному нулевому тождеству.

Из третьего получаем

$$\frac{t}{\sqrt{2\pi S_2}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{t^2}{2S_2}\right)\left[1 + \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{A_i} t^{2i}\right] dt = S_2\,,$$

откуда

$$\sum_{i=1}^{k} A_i (2i+1)!!\, S_2^{i-1} = 0. \quad (6)$$

Из четвертого соотношения в (3) следует, что

$$\int_{-\infty}^{\infty} t^{2j} \left[1 + \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{A_i} t^{2i}\right] \exp\left(-\frac{t^2}{2S_2}\right) dt =$$
$$= S_{2j} \sqrt{2\pi S_2}\,,\ j = \overline{2, k-1},$$

откуда получаем

$$\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{A_i} [2(i+j)-1]!!\, S_2^{i-1} = \frac{S_{2j} - (2j-1)!!\, S_2^j}{S_2^{j+1}}. \quad (7)$$

Таким образом, неизвестные коэффициенты $A_i, i = \overline{1,k}$ с учетом (5) – (7) удовлетворяют k уравнениям:

$$\sum_{\mathrm{i}=1}^{\mathrm{k}} (2i-1)!!\, S_2^{i-1} \mathrm{A_i} = 0;$$
$$\sum_{\mathrm{i}=1}^{\mathrm{k}} (2i+1)!!\, S_2^{i-1} \mathrm{A_i} = 0,$$
$$\sum_{\mathrm{i}=1}^{\mathrm{k}} (2i+2j-1)!!\, S_2^{i-1} \mathrm{A_i} = \frac{S_{2j} - (2j-1) S_2^j}{S_2^{j+1}}. \quad (8)$$

В системе уравнений (8) удобно ввести новые коэффициенты $B_i = A_i S_2^{i-1}$, благодаря чему определим систему:

$$\sum_{\mathrm{i}=1}^{\mathrm{k}} (2i-1)!!\, B_i = 0; \quad \sum_{\mathrm{i}=1}^{\mathrm{k}} (2i+1)!!\, B_i = 0;$$
$$\sum_{\mathrm{i}=1}^{\mathrm{k}} (2i+2j-1)!!\, B_i = \frac{S_{2j} - (2j-1) S_2^j}{S_2^{j+1}}.$$

Например, при $k = 3$ получим

$$B_1 + 3B_2 + 15B_3 = 0;\ B_1 + 5B_2 + 35B_3 = 0;$$

$$B_1 + 7B_2 + 63B_3 = \gamma;\ \gamma = \frac{S_4 - 3S_2^2}{15 S_2^3}. \quad (9)$$

Из (9) определяем

$$B_1 = \frac{15}{8}\gamma;\ B_2 = -\frac{5}{4}\gamma;\ B_3 = \frac{1}{8}\gamma;$$
$$A_1 = \frac{15}{8}\gamma;\ A_2 = -\frac{5}{4}\cdot\frac{\gamma}{S_2};\ A_3 = \frac{1}{8}\cdot\frac{\gamma}{S_2^2}.$$

В этом месте уместно заметить, что в случае, когда полученная выборка не удовлетворяет нормальному закону распределения, коэффициенты A_i выражаются через эмпирические моменты (1).

Далее, для функции плотности распределения вероятностей $f_3(x_i,\vec{\theta})$, где $\vec{\theta}$ – некоторая совокупность неизвестных параметров (в общем случае), получаем

$$f_3(x_i,\vec{\theta}) = \frac{1}{\sqrt{2\pi S_2}} \exp\left[-\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^2 / 2S_2\right] \times$$

$$\times\left\{1+\frac{\gamma}{8}\left[15\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^2 - 10\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^4 / S_2 + \right.\right.$$

$$\left.\left.+\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^6 / S_2^2\right]\right\}$$

при ограничении $S_4 < \frac{36}{7} S_2^2$.

Последнее необходимо, чтобы $f_3(x) > 0$.

Теперь пользуясь предписанием метода максимального правдоподобия [3] получим алгоритм оценивания, который базируется на распределении случайных величин, отвечающем имеющейся выборке. Логарифм функционала правдоподобия $\Psi_n(\vec{x}/\vec{\theta})$ для рассмотренного нами случая имеет вид

$$\Psi_n(\vec{x}/\vec{\theta}) = \sum_{i=1}^{n} \ln\frac{1}{\sqrt{2\pi S_2}} - \frac{\sum_{i=1}^{n}\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^2}{2S_2} +$$

$$+\sum_{i=1}^{n}\ln\left\{1+\frac{\gamma}{8}\left[15\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^2 - 10\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^4 / S_2 + \right.\right.$$

$$\left.\left.+\left(x_i - \sum_{j=1}^{m} a_{ij}\theta_j\right)^6 / S_2^2\right]\right\}. \qquad (10)$$

Для упрощения (10) можно последнее слагаемое разложить в ряд

$$\ln\left\{1+\frac{\gamma}{8}\left(15X_i^2 - \frac{10}{S_2}X_i^4 + \frac{1}{S_2^2}X_i^6\right)\right\} \approx$$

$$\approx \frac{\gamma}{8}\left(15X_i^2 - \frac{10}{S_2}X_i^4 + \frac{1}{S_2^2}X_i^6\right),$$

где $X_i = x_i - \sum_{j=1}^{m} a_{ij}\theta_j$.

Тогда (10) можно переписать в виде

$$\Psi_n(\vec{x}/\vec{\theta}) = \sum_{i=1}^{n} \ln\frac{1}{\sqrt{2\pi S_2}} - \frac{\sum_{i=1}^{n} X_i^2}{2S_2} +$$

$$+\frac{\gamma}{8}\sum_{i=1}^{n}\left(15X_i^2 - \frac{10}{S_2}X_i^4 + \frac{1}{S_2^2}X_i^6\right)$$

и после дифференцирования по неизвестным параметрам $\vec{\theta}$ получим систему уравнений

$$\frac{\partial}{\partial\theta_\nu}\Psi_n(\vec{x}/\vec{\theta}) = -\frac{\sum_{i=1}^{n} X_i a_{i\nu}}{S_2} +$$

$$+\frac{\gamma}{8}\sum_{i=1}^{n}\left(30X_i a_{i\nu} - \frac{40}{S_2}X_i^3 a_{i\nu} + \frac{6}{S_2^2}X_i^5 a_{i\nu}\right) = 0,$$

$$\nu = \overline{1,m}.$$

Решая данную систему уравнений, например, методом Ньютона [4], определяем искомые неизвестные параметры.

Заключение

Достоинство представленного в данной работе подхода к вопросу оценки параметров состоит в том, что искомые параметры находятся из системы уравнений, полученных без привлечения гипотетических законов распределения вероятностей данных измерений (конечно, с учетом оговоренного ограничения на симметрию распределения).

Литература

1. Красников В.Н., Лещенко А.Б. Моделирование законов распределения случайных величин по эмпирическим данным // Радиоэлектронные и компьютерные системи. – 2005. – № 4 (12). – С. 108-111.

2. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм , рядов и произведений. – М.: Наука, 1971. – 1101 с.

3. Мудров В.И., Кушко В.Л. Методы обработки измерений. – М.: Радио связь, 1983. – 303 с.

4. Самарский А.А. Введение в численные методы. – М.: Наука, 1982. – 271 с.

Поступила в редакцию 1.06.2006

Рецензент: д-р физ.-мат наук, проф. А.Г. Николаев, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.