

УДК 519.248

Ю.А. СТОЛЯРЕНКО

*Приднестровский государственный университет им. Т.Г. Шевченко,
Тирасполь, Молдова***МЕТОД ТОЧЕЧНЫХ РАСПРЕДЕЛЕНИЙ**

Существует большое количество отраслей человеческой деятельности (промышленность, экономика, медицина, образование и т.д.), где получение большого количества информации для статистической обработки невозможно по различным причинам. Например, производство кристаллов интегральных микросхем. О качестве кристаллов пластины необходимо судить по показаниям 5 или 10 тестовых ячеек. При существующих методах браковки в итоге получается очень большой процент брака. Следовательно, для повышения процента выхода годных изделий без дополнительных капиталовложений необходима разработка методов обработки выборок малого объема.

Ключевые слова: пассивные данные, выборка малого объема, эквивалентная выборка.

Введение

В трудах Гаскарова Д.В. и Шаповалова В.И. [1] была высказана идея о том, чтобы избежать потери информации при обработке статистических данных классическими методами необходимо рассматривать каждое измерение, обрабатываемой пассивной выборки малого объема, как центр распределения некоторой виртуальной выборки с соответствующими параметрами.

В дальнейшем к получившемуся набору данных (обработанных по соответствующему алгоритму) можно будет применить методы классической статистической обработки данных.

1. Разработка метода точечных распределений

Эта идея была взята за основу при разработке метода, позволяющего обработать выборку малого объема и получить так называемую виртуальную или эквивалентную выборку большого объема, для того, чтобы появилась возможность продолжить обработку методами классической статистики.

Метод получил название «Метод точечных распределений» (МТР) и реализует следующий алгоритм [2]:

1. Установить предполагаемый класс распределений $f(x)$ (например, нормальный закон).
2. Определить интервал изменения (a,b) контролируемой величины X в абсолютных единицах (а.е.).
3. Вычислить оптимальное значение ρ в а.е.:

$$\rho = \rho' \cdot (b - a), \text{ [а.е.]} \quad (1)$$

вспомогательный коэффициент ρ' определяем эмпирически для каждого класса распределения $f(x)$ и объема выборки n .

4. Расположить элементы выборки

$$X_1, \dots, X_i, \dots, X_n$$

в порядке возрастания их значений.

5. Для каждого элемента выборки X_i , определить верхнюю X_{iB} и нижнюю X_{iH} границы интервала определения i -ой дельтаобразной функции:

$$X_{iH} = \begin{cases} a, & \text{если } X_i - \rho \leq a; \\ X_i - \rho, & \text{в остальных случаях;} \end{cases}$$

$$X_{iB} = \begin{cases} b, & \text{если } X_i + \rho \geq b; \\ X_i + \rho, & \text{в остальных случаях.} \end{cases}$$

6. Для каждого элемента выборки X_i определить корректировочный коэффициент A_i , который равен:

- для нормального закона распределения

$$A_i \equiv 1;$$

- для экспоненциального распределения

$$A_i = \begin{cases} 1 - \frac{X_i - \rho - a}{\rho}, & \text{если } X_{iH} = a; \\ 1, & \text{в остальных случаях.} \end{cases}$$

7. Интервал (a, b) разбить на k интервалов дискретности ($k = 20 - 30$). При этом величина X_j – верхняя граница j -го интервала дискретности

$$X_0 = a; X_k = b,$$

а

$$X'_j = a + (2j - 1) \cdot \frac{b - a}{2k} -$$

центр j -го интервала дискретности.

8. Вычислить значение нормированной плотности $f^*(X'_j)$ на интервал с центром в точке X_j :

$$f^*(X_j) = \frac{k}{b-a} \cdot \frac{f'(X'_j)}{\sum_{j=1}^n f'(X'_j)}; \quad (2)$$

где $f'(X'_j)$ – значения ненормированной оценки плотности в центре j -го интервала дискретности. Например, для дельтаобразной формы ядра:

$$f'(X) = \alpha \cdot f_0(X) + (1-\alpha) \frac{3A_i}{\rho\sqrt{2\pi}} \times \sum_{i=1}^n p_{ij} \cdot \exp\left[-4,5\left(\frac{X'_i - X_j}{\rho}\right)^2\right], \quad [1/a.e.], \quad (3)$$

где p_{ij} – условие «накрывания» интервалом задания $\pm\rho$ i -й дельтаобразной функции центра j -го интервала дискретности:

$$p_{ij} = \begin{cases} 1, & \text{если } X_{iH} \leq X'_j \leq X_{iB}; \\ 0, & \text{в остальных случаях.} \end{cases} \quad (4)$$

Вопрос об априорной компоненте $f_0(X)$ входящей в выражение (3) ненормированной оценки плотности в центре j -го интервала $f(X_j)$ может быть решен по-разному. В простейшем случае можно считать, что априорная компонента обратно пропорциональна интервалу определения величины

$$f_0(X) = \frac{1}{b-a}.$$

В более сложных случаях для нормального распределения

$$f_0(X) = \frac{1}{\sigma_X\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{X - m_X}{\sigma_X}\right)^2\right],$$

где математическое ожидание m_X и СКО σ_X оценивается по достаточно большому количеству выборок.

Тогда ненормированная оценка плотности для нормального закона запишется в виде:

$$f'(X'_j) = \frac{1}{\sqrt{2\pi}} \left\{ \frac{\alpha}{\sigma_X} \cdot \exp\left[-\frac{1}{2}\left(\frac{X'_j - X_i}{\sigma_X}\right)^2\right] + \frac{3(1-\alpha)}{\rho} \sum_{i=1}^n p_{ij} \exp\left[-4,5\left(\frac{X'_j - X_i}{\rho}\right)^2\right] \right\} = \frac{B_j}{\sqrt{2\pi}}, \quad (5)$$

где символом B_j обозначено выражение в фигурных скобках.

Обработав выборку малого объема с помощью рассмотренного алгоритма получаем виртуальную выборку, пригодную для обработки классическими методами.

2. Развитие метода точечных распределений

Приведенный выше алгоритм метода точечных распределений рассматривался в разрезе нормально-го закона распределения.

Но, статистические данные, получаемые в различных отраслях, имеют различные виды законов распределения.

В связи с этим, появляются задачи: определение интервала изменения (a , b) для каждого вида распределения; определение ширины ядра и алгоритма его определения; определения вероятностей появления случайной величины в интервале дискретности.

Также важными задачами являются:

- отсеивание грубых промахов в выборках малого объема;
- определение вида закона распределения малой выборки.

Заключение

Для качественного проведения обработки выборок малого объема необходимо наличие специализированного программного обеспечения и, соответствующих алгоритмов для его написания.

Необходимо разработать МТР для следующих видов законов распределения:

- χ^2 распределение;
- распределение Парето;
- гамма-распределение;
- логарифмически нормальное распределение;
- распределение Стьюдента;
- равномерное распределение [3].

Литература

1. Гаскаров, Д.В., Малая выборка [Текст] / Д.В. Гаскаров, В.И. Шаповалов. – М.: Статистика, 1987. – 248 с.
2. Столяренко, Ю.А. Контроль кристаллов интегральных схем на основе статистического моделирования методом точечных распределений [Текст]/Ю.А. Столяренко. – Дисс. на соиск. уч. степ. канд. ... техн. наук. – М. ГУП «СПУРТ», 2006. – 192 с.
3. Митропольский, А.К. Техника статистических вычислений [Текст] / А.К. Митропольский. – 2-е изд., перераб. и доп. – М.: Наука, 1971. – 576 с.

Поступила в редакцію 19.03.2012

Рецензент: д-р техн. наук, проф. В.И. Хаханов, Харьковский национальный университет радиоэлектроники, Харьков, Украина.

МЕТОД ТОЧКОВИХ РОЗПОДІЛІВ

Ю.О. Столяренко

Існує велика кількість галузей діяльності людини (промисловість, економіка, медицина, освіта, тощо.), де отримання великої кількості інформації для статистичної обробки неможливе через різні причини. Наприклад, виробництво інтегральних схем. Про якість кристалів пластини необхідно судити за показаннями 5 або 10 тестових клітинок. При існуючих методах бракування в результаті виходить дуже великий відсоток браку. Отже, для підвищення відсотка виходу придатних виробів без додаткових капіталовкладень необхідно розробка методів обробки вибірок малого обсягу.

Ключові слова: пасивні дані, невеликі вибірки, еквівалентні вибірки.

METHOD OF POINT ALLOCATIONS

Y.A. Stolyarenko

There are many branches of human activity (industry, economics, medicine, education, etc.), where a large amount of information for the statistical processing is not possible for various reasons. For example, the production of integrated circuits. On the quality of crystals to stiny must be judged according to 5 or 10 test cells. The existing methods of rejection boundaries in a very large percentage of marriage. Consequently, for the rising rate of out good products without additional capital investment not-required an elaboration of methods of processing small amount of samples.

Key words: passive data, small sample, equivalent sample.

Столяренко Юлія Александровна – канд. техн. наук, доцент кафедри «Информационные системы и автоматизированное управление технологическими процессами» ПГУ им. Т.Г. Шевченко, Молдова, ПМР, e-mail: mlcc@mail.ru.