УДК 519.24

**Ю.А. ДОЛГОВ, А.Ю. ДОЛГОВ, М.М. ВАНЯШКИН, А.В. ДЕТКОВА**

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

# ОПРЕДЕЛЕНИЕ ГРАНИЦ И ШИРИНЫ ЯДРА ЭКВИВАЛЕНТНЫХ ВЫБОРОК В МЕТОДЕ ТОЧЕЧНЫХ РАСПРЕДЕЛЕНИЙ

*Предлагаются методы расчета границ эквивалентных выборок и вспомогательного коэффициента, влияющего на ширину ядра в методе точечных распределений для нескольких законов плотности вероятности, необходимые на начальном этапе расчета параметров выборок малого объема (n = 3÷15 элементов) повышенной эффективности. Оба показателя – границы и ширина ядра – зависят от закона распределения и объема выборки. В статье приведена идеология метода точечных распределений, расчетные формулы и результаты исследований. Для всех методов и законов распределения разработано программное обеспечение.*

***Ключевые слова:*** *выборка малого объема, законы плотности вероятности, эквивалентная выборка, новые границы, ширина ядра.*

## Введение

Современное производство изделий электронной техники имеет тенденции к уменьшению объема выборок при контрольных и контрольно-отбраковочных операциях. Все ГОСТЫ, регламентирующие выборочный контроль по количественному или качественному признаку, считают нижней границей контрольной выборки объем, равный 10% от объема контролируемой партии [1, 2].

С другой стороны, современные групповые иерархические технологии производства элементов электронной техники (например, кристаллов интегральных микросхем) предоставляют изготовителю возможность судить обо всех кристаллах на пластине (от 400 до 5000 штук) по 5 (максимум 10) тестовым ячейкам, единственно доступным для промежуточных контрольных измерений. Такой объем является даже не малой, а сверхмалой выборкой, которая, конечно же, обладает весьма низкими прогнозирующими свойствами. Напрашивается вывод о необходимости развития теории обработки выборок малого объема методами, отличными от классических, для различных законов плотности вероятности.

В статье предлагаются методы расчета границ эквивалентных выборок повышенной эффективности, полученные на базе выборок малого объема, и расчета ширины (диапазона покрытия) ядра в методе точечных распределений [3, 4].

## 1. Общие положения

Выборкой малого объема называется выборка, содержащая n = 3÷15 элементов [4]. Для устранения потерь информации при обработке малой выборки необходимо считать каждое измерение центром некоторого виртуального распределения с известным законом [5]. Это позволяет существенно уменьшить интервал неопределенности выборочных оценок, что, в свою очередь, дает возможность значительно снизить объемы выборок, например, при контроле промежуточных операций при производстве промышленной продукции и применить известные статистические методы для ее разбраковки по ходу технологического процесса там, где это ранее было принципиально невозможно.

В основу метода точечных распределений (МТР) положено предварительное числовое определение эмпирической функции распределения

$$f^*(X) = \alpha_0 \cdot f_0(X) + (1-\alpha_0) \cdot f_n^*(X), \qquad (1)$$

где $\alpha_0$ – ценность априорной информации;

$f_0(X)$ – априорная компонента, несущая информацию о форме закона распределения;

$f_n^*(X) = C(\rho)\sum_{i=1}^{n}\mu_i\psi_i(\rho, X)$ – эмпирическая компонента;

$C(\rho)$ – нормирующий множитель;

$\psi_i(\rho, X)$ – некоторое ядро при i-м измерении;

$\rho$ – половина интервала определения ядра;

$\sum_{i-1}^{ò}\mu_i = 1$ – коэффициенты нормировки.

Полная оценка плотности распределения удовлетворяет весьма важным для оценок свойствам состоятельности и несмещенности. Эффективность оценки зависит от формы ядра. Исследованиями установлено, что наиболее простой формой ядра является прямоугольная, а оптимальной – дельтавидная.

Практический алгоритм вычисления эмпирической функции распределения $f^*(X)$ содержит пункты, среди которых имеются:

1) установить предполагаемый класс распределений f(X);

2) определить интервал измерения (a,b) исследуемой величины X в абсолютных единицах;

3) вычислить оптимальное значение половины интервала определения ядра ρ

$$\rho = \rho'(b-a), \tag{2}$$

где ρ' – вспомогательный коэффициент, определяется эмпирически для каждого класса распределения f(X) и объема выборки n.

## 2. Границы эквивалентных выборок

Основой всего метода точечных распределений является определение границ эквивалентных выборок (типа рис.1) – левой – «a» и правой «b».

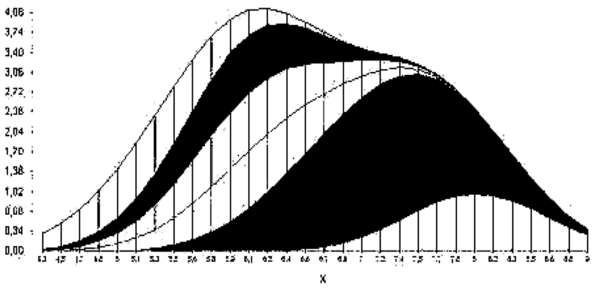


Рис. 1. График суммарного распределения с центрами в экспериментальных точках для выборки X

В работах [3 – 5] основное внимание уделялось нормальному закону распределения и частично экспоненциальному закону и закону Вейбулла. Между тем в справочнике по статистическим распределениям [6] приведены материалы о 24 распределениях, наиболее часто встречающихся в промышленности, медицине, экономике и т.д. Рассмотрим более подробно способы определения границ эквивалентностной выборки (a,b) в зависимости от вида закона распределения f(X) и объема выборки n. Во всех случаях границы (a,b) будут определяться по минимальному уровню значимости q, равному 0,02, (или, что то же самое, по максимальному значению доверительной вероятности $P_{ДОВ}$= 0,98):

$$Ð_{âî â} = F(X) = \int_a^b f(X)\,dX = 0,98\,. \tag{3}$$

Для некоторых законов распределения границы эквивалентных выборок однозначно связаны с параметрами исходных выборок малого объема: средним арифметическим

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{4}$$

и эмпирической дисперсией

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \overline{(X_i - X)^2}\,, \tag{5}$$

где n – объем малой выборки.

Однако для некоторых законов распределения в силу их специфики правые (а иногда и левые) границы зависят от нескольких дополнительных параметров и представляют собой довольно объемные таблицы. В этих случаях таблицы сворачивались в компактные уравнения по методу наименьших квадратов (всего 19 форм), среди которых выбиралось наилучшее, для которой индекс корреляции с исходными данными был наибольшим, а среднеквадратичная ошибка – наименьшей [7]. Результаты исследования представлены в табл. 1.

## 3. Расчет ширины дельтаобразного ядра

Вторым по значению параметром в методе точечных распределений является вспомогательный-коэффициент ρ', определяющий половину ширины ядра (2). В сочетании с ценностью (весом) априорной информации $\alpha_0$ они полностью определяют эмпирическую функцию распределения (1). Коэффициенты $\alpha_0$ и ρ' имеют сложную зависимость (параболу или экспоненту), учитывающую закон распределения, а, кроме того, они взаимно влияют друг на друга. В силу этого можно найти такую минимальную точку на параболе (или установить минимальное значение ординаты экспоненты), абсцисса которой даст нам наилучшие значения $\alpha_{0m}$ и $\rho'_m$. Однако установлено [5], что коэффициенты $\alpha_{0m}$ и $\rho'_m$ могут быть определены только вместе, для чего требуются достаточно трудоемкие исследования при жестко оговоренных условиях.

Поскольку такая работа для распределений, отличных от нормального закона, проводится впервые, то имеет смысл сначала установить характер кривых и порядок величин ρ'. С этой целью на первоначальном этапе ограничимся частным случаем, когда $\alpha_0 = 0$.

Вспомогательный коэффициент ρ' определяется эмпирически для каждого класса распределения f(x) и объема выборки п. При вычислении ρ' необходимо выполнить последовательно следующие операции:

1. Задается плотность f(x) распределения случайной величины X;
2. Задается дельтообразная функция $D(p, x, x_t)$;
3. Задается объем выборки п;
4. Принимается начальное значение переменной $a_o$= 0;

Таблица 1

Формулы для вычисления границ эквивалентных выборок

| Распределение | Функция f(x) | Параметры | a | b |
|---|---|---|---|---|
| Нормальное | $\frac{1}{S\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{X-\bar{X}}{S}\right)^2\right\}$ | $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$<br>$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}$ | $\bar{X}-2,4S;$ | $\bar{X}+2,4S$ |
| Экспоненциальное | $\frac{1}{\bar{X}}\exp\left(-\frac{X}{\bar{X}}\right)$ | $\hat{\lambda}=\frac{1}{\bar{X}}$<br>$S=\bar{X}$ | 0 | $3,912\bar{X}$ |
| Вейбулла | $\frac{CX^{C-1}}{g^C}\exp\left[-\left(\frac{X}{g}\right)^C\right]$ | $\hat{g}=\left[\frac{1}{n}\sum_{i=1}^{n}X_i^C\right]^{\frac{1}{C}}$ ,<br>$\hat{C}=n\times$<br>$\times\left[\frac{1}{g^C}\sum_{i=1}^{n}X_i^C\log X_i-\sum_{i=1}^{n}\log X_i\right]^{-1}$ | 0 | $g\sqrt[C]{3,912}$ |
| Стьюдента | $\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{u^2}{n}\right)^{-\frac{n+1}{2}}$ | - | $\bar{X}-Z_nS$ | $\bar{X}+Z_nS$ ,<br>где $Z_n=1,75+\frac{11,1}{n}$ |
| Хи-квадрат | $\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}X^{\frac{n}{2}-1}\cdot e^{-\frac{1}{2}X}$ | - | 0 | 1,53n+5,95<br>или табличное<br>$\chi^2_{0,025}$ |
| Логнормальное | $\frac{1}{XS\sqrt{2\pi}}\exp\left\{-\frac{\left[\log\left(\frac{X}{m}\right)^2\right]}{2S^2}\right\}$ | $\hat{m}=\exp\left(\hat{\mu}-\frac{\hat{\sigma}^2}{n}\right)$<br>$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\log X_i$<br>$\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(\log X_i-\hat{\mu}\right)^2$ | 0 | -7,26+4,726$\sigma^2$+<br>+9636m |
| Гамма | $\begin{cases}\frac{\lambda^\eta}{\tilde{A}(\eta)}X^{\eta-1}e^{-\lambda X};X\geq 0;\lambda>0;\eta<0\\ 0-\text{âî ñòàëüíû õñëó÷àÿõ}\end{cases}$<br>Если выборка смещена относительно нуля по оси абсцисс на величину θ, то гамма –распределение имеет место на интервале от θ до ∞ | $\hat{\lambda}=\frac{\bar{X}}{S^2}$<br>$\hat{\eta}=\hat{\lambda}\cdot\bar{X}=\frac{\bar{X}^2}{S^2}$ | 0 | 8,25-7,194λ+<br>+3,195η+1,3 $\lambda^2$-<br>-0,774λη |
| Парето | $CX^{-C-1}$ | $M[X]=\frac{C}{C-1};\quad C>1;$<br>$D[X]=\frac{C}{C-2}-\left(\frac{C}{C-1}\right)^2;C>2$ | 1 | $\sqrt[C]{50}$ |
| Равномерное | $\frac{1}{b-a}$ | - | a | b |

5. Датчик случайных чисел с заданной f(x) генерирует выборку, содержащую п реализаций:

$x_1, x_2, \dots, x_i, \dots x_n$

6. Принимается начальное значение переменной $\rho = \rho'_0$

7. По выборке строится оценка плотности распределения для заданных $a_0$ и $\rho=\rho'(b-a)$.

С точки зрения организации программы практически оказалось удобнее сначала строить оценку функции распределения F*(x). Это обусловлено тем, что в результате вычислительного процесса получается оценка, состоящая из ряда дискретных значений. Значения оценки определены в точках оси абсцисс: $x_1, x_2, \dots, x_i, \dots x$. Для построения оценки F*(x) необходимо установить интервал (a, b) изменения случайной величины X.

Для сохранения равенства F*(b) =1 было введено нормирование во всех точках интервала (a,b):

$$F^*(X_j) = \sum_{j=1}^{j} f^*(X_j) \Big/ \sum_{j=1}^{n} f_l^*(X_j); \qquad (7)$$

где n – число точек разбиения (a, b);

j – текущий номер точки разбиения (a, b).

8.По построенной оценке F*(x) вычисляются оценка f*(x) и величина $H_N$ по формуле:

$$H_N = \int_U f^*(x) \cdot \ln \frac{f^*(x)}{f(x)} dx; \qquad (8)$$

где U=1,2,...k отрезков.

9. Действия по п. 7-8 повторяются для переменной ρ', меняющейся от начального значения $\rho'_0$ до конечного значения $\rho'_k$ с постоянным шагом Δ ρ';

10. Действия по п. 5-9 повторяются достаточно большое число раз L;

11. Вычисляются оценки математического ожидания $\overline{H}_N$ и функции распределения $F(H_N)$ величины $H_N$:

$$\overline{H}_N = \frac{1}{L} \cdot \sum_{l=1}^{L} H_{Nl}, \qquad (9)$$

где $H_{NL}$ – реализация, полученная при выполнении пункта 8.

Оценка $F(H_N)$ строится методом гистограмм, так как значение L достаточно велико.

В результате выполнения вышеперечисленных операторов получается оценка функции для конкретных f(x), n и $\alpha_0$. И в построенной гистограмме (рис. 2) значением коэффициента ρ' является ярко выраженный минимум, т.е. при данных условиях существует оптимальное значение переменной $\rho'=\rho'_m$.

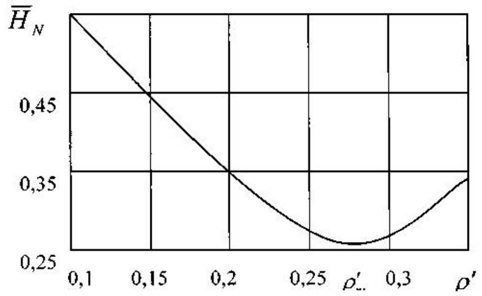


Рис. 2. Оценка зависимости величины $\overline{H}_N$ от переменной ρ'

Подобная работа была проведена для всех законов распределения, упомянутых в таблице 1и для каждого объема малой (исходной) выборки в диапазоне $n=\overline{3,15}$. Для удобства дальнейшего использования полученные числа (по 13 на каждый закон) сворачивались в формулы по уже упомянутой методике [7]. Выбор наилучшего из возможных 19 уравнений осуществлялся по наибольшему индексу корреляции и наименьшему среднеквадратическому отклонению одновременно. Результаты этой работы представлены в табл. 2.

Таблица 2

Зависимости величины ρ'от вида закона распределения для дельтаобразной формы ядра

| Закон распределения | Уравнение регрессии $\hat{\rho}'$ | Коридор существования |
|---|---|---|
| Нормальный | $\hat{\rho}' = \dfrac{0,3435n}{n-0,7066}$ | ± 0, 0034 |
| Экспоненциальный | $\hat{\rho}' = 0,3788 - 0,00065n^2 + 1,56347e^{-n}$ | ± 0,0028 |
| Вейбулла | $\hat{\rho}' = 0,8406 - 0,3196 \ln n + 0,0674(\ln n)^2$ | ± 0,0058 |
| Стьюдента | $\hat{\rho}' = 0,4772 + 0.0012n - 0,00004n^2$ | ± 0,0087 |
| Хи-квадрат | $\hat{\rho}' = 0,5279 - 0,0045n + 0,000085n^2$ | ± 0,0059 |
| Логнормальный | $\hat{\rho}' = 0,5522 + 0,00624n + 0,000257n^2$ | ± 0,0090 |
| Равномерный | $\hat{\rho}' = 0,3586 - 0,000616n - 0,000045n^2$ | ± 0,0087 |

## Заключение

1. По условию F(x)=0, 98 найдены границы эквивалентных (виртуальных) выборок малого объема для девяти законов плотности вероятности.

2. Для плавающих границ (зависящих от объема выборки) найдены расчетные формулы.

3. Для семи законов плотности вероятности найдены уравнения регрессии для вспомогательного коэффициента $\hat{\rho}'$, определяющего ширину дельтаобразного ядра (в первом приближении).

4. Для уточнения параметров регрессии $\hat{\rho}'$ и распространения метода на другие законы распределения, а также для выявления подобных регрессий в случае ядра прямоугольной формы, работу в этом направлении следует продолжить.

## Литература

*1. Наулер Л. Статистические методы контроля качества продукции / Л.Наулер, Дж. Хауэлл, Б.Голд и др.- М.: Изд-во стандартов, 1984. – 104 с.*

*2. Статистический приемочный контроль по количественному признаку. Планы контроля: ГОСТ 20736-75 (Ст. СЭВ 1672-79). – М.: Изд-во стандартов, 1982. – 120с.*

*3. Долгов А.Ю. Повышение эффективности статистических методов контроля и управления технологическими процессами изготовления микросхем / А.Ю. Долгов // Дис. на соиск. уч. степ. канд. техн. наук. – М.: МГАПИ, 2000. – 217 с.*

*4. Столяренко Ю.А. Контроль кристаллов интегральных микросхем на основе статистического моделирования методом точечных распределений / Ю.А. Столяренко // Дис. на соиск. уч. степ. канд. техн. наук. – М.: ГУП НПЦ «Спурт», 2006. – 191 с.*

*5. Гаскаров Д.В. Малая выборка / Д.В. Гаскаров, В.И. Шаповалов. – М.: Статистика, 1978. – 248 с.*

*6. Хастингс Н. Справочник по статистическим распределениям / Пер. с англ. / Н. Хастингс, Дж. Пинок. – М.: Статистика, 1980. – 96 с.*

*7. Долгов Ю.А. Статистическое моделирование: Учеб. для ВУЗов / Ю.А. Долгов. – Тирасполь: РИО ПГУ, 2002. – 280 с.*

*Поступила в редакцию 14.01.2010*

**Рецензент:** д-р.техн. наук, проф. А.А. Мельник, Национальный технический университет "Львовская политехника", Львов.

### ВИЗНАЧЕННЯ КОРДОНІВ І ШИРІНИ ЯДРА ЕКВІВАЛЕНТНИХ ВИБІРОК В МЕТОДІ ТОЧКОВИХ РОЗПОДІЛІВ

***Ю.О. Долгов, О.Ю. Долгов, М.М. Ваняшкін, Г.В. Деткова***

Пропонуються методи розрахунку межи еквівалентних вибірок та допоміжного коефіцієнту, що впливає на ширину ядра в методі крапкових розподілень для декількох законів щільності ймовірності, які необхідні на початковому етапі розрахунку параметрів вибірок малого обсягу (n = 3÷5 елементів) підвищеної ефективності. Оби два показника (межи та ширини ядра) залежіть від закону розподілення і обсягу вибірки. В статті приведено ідеологію метода крапкових розподілень, розрахункові формули та результати досліджень. Для всіх методів і законів розподілення розроблено програмне забезпечення.

**Ключові слова**: вибірка малого обсягу, закони щільності ймовірності, еквівалентна вибірка, нові межи, ширина ядра.

### DEFINITION OF BOUNDARIES AND KERNEL WIDTH OF EQUIVALENT SAMPLES IN POINT-DISTRIBUTION METHOD

***Y.A. Dolgov, A.Y. Dolgov, M.M. Vanyashkin, A.V. Detkova***

There is offered some methods for calculation of equivalent sample boundaries and auxiliary coefficient too, which have influence on kernel width in point-distribution method for several density function laws, those it are necessary on first stage of calculation of small size sample parameters (n=3-15 elements) of higher effectiveness. The both indexes – the boundaries and kernel width – are depended on distribution law and sample size. In this article it is brought out an ideology of point- distribution method, some calculate formulas and investigative results. Its are wrought computer programs for all methods and distribution laws.

**Key words:** small size sample, density probability, equivalent sample, new boundaries, kernel width.

**Долгов Юрий Александрович –** д-р техн. наук, проф., зав. кафедрой информационных технологий и автоматизированного управления производственными процессами, ПГУ им. Т.Г.Шевченко, Тирасполь, Приднестровье, Молдова, e-mail dolgov@spsu.ru.

**Долгов Алексей Юрьевич** – канд. техн. наук, доцент кафедры ИТУ ПГУ, Тирасполь, Приднестровье, Молдова, e-mail dolgov@spsu.ru.

**Ваняшкин Михаил Михайлович** – аспирант ИТУ ПГУ, Тирасполь, Приднестровье, Молдова.

**Деткова Анна Васильевна** - аспирант ИТУ ПГУ, Тирасполь, Приднестровье, Молдова.