УДК 519.248

Ю.А. ДОЛГОВ

Приднестровский государственный университет им. Т.Г. Шевченко, Тирасполь, Приднестровье

ИССЛЕДОВАНИЕ ИНТЕРВАЛЬНЫХ ОЦЕНОК И МЕТОДОВ ВЫСШЕГО ПОРЯДКА ДЛЯ ВЫБОРОК МАЛОГО ОБЪЕМА

Предлагаются методы определения законов распределения и параметров выборок малого объема, а также методы уменьшения вычислительных ошибок этих параметров. Разработан метод, позволяющий в случае выборок малого объема определять хотя бы приблизительно вид закона распределения генеральной совокупности, из которой взята конкретная выборка, а также в 3-4 раза сокращать размер интервальной оценки выборочной дисперсии, что позволяет в 2-2,5 раз уменьшить долю ложной приемки и ложной браковки за счет уменьшения ошибки вычисления параметров контрольных выборок малого объема.

Ключевые слова: *выборка малого объёма, центральные моменты, точечные и интервальные оценки.*

Постановка задачи

При производстве кристаллов ИМС пооперационный выборочный контроль производится только в пяти (десяти) тестовых ячейках на каждой пластине, где располагается 400-5000 рабочих кристаллов. Такое соотношение величин (контрольная выборка составляет около 0,1% от общего объёма) не подходит ни к одному плану гостированного статистического выборочного контроля и поэтому в производстве используется так называемый граничный метод, недостатком которого является слишком большая доля ложной приемки и ложной браковки. Проделанный анализ показал, что эта доля зависит от объективных и субъективных причин [1], причем последние являются следствием нашего незнания законов распределения контролируемых параметров и достаточно грубых (с большой ошибкой) расчетов параметров выборки малого объема. Уменьшение ошибки вычисления параметров выборок малого объема возможно с помощью метода точечных распределений (МТР) [2]. В настоящей статье остановимся на методах определения законов распределения и параметров выборок малого объема и на методах уменьшения ошибок при расчете этих параметров.

1. Определение точечных оценок центральных методов

Согласно классической теории вероятностей вид закона распределения по выборкам большого объема может быть найден графоаналитическим методом через асимметрию (r_3) и эксцесс (r_4) гистограммы [3] по следующей цепочке формул

$$r_3 = \frac{\mu_3}{\mu_2^{3/2}}; \quad r_4 = \frac{\mu_4}{\mu_2^2}, \qquad (1)$$

где

$$\mu_h = \frac{1}{n-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^h, \ h = 2, 3, 4, \qquad (2)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{N} X_i \qquad (3)$$

и номограмме рис. 1.

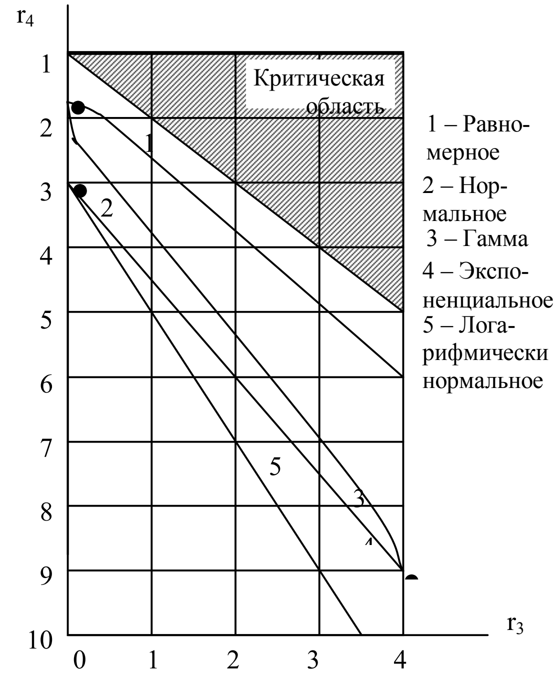


Рис. 1. Номограмма определения закона выборочного распределения

Однако, как упоминалось выше, при выборках малого объема вычисления их параметров имеют слишком большую ошибку. Так, например, по данным [4] относительная ошибка выборочного среднеквадратического отклонения $\sigma(S)/\sigma$ равна $\sqrt{1/(2n-1,4)}$ (при n=5 составляет 0,341 или 34,1%, а при n=10 – 0,232 или23,2%).

Для уменьшения этой ошибки оценки параметров выборок малого объема следует искать по методу точечных распределений (МТР) [2]. Аналогичная цепочка формул с учетом рекомендаций [3] может быть представлена как

$$m_X^* = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} X_j \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]}{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]}; \qquad (4)$$

$$\mu_2^* = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} X_j^2 \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]}{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]} - \left(m_X^*\right)^2; \qquad (5)$$

$$\mu_3^* = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} X_j^3 \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]}{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]} - \qquad (6)$$

$$-3\mu_2^* m_X^* - \left(m_X^*\right)^3;$$

$$\mu_4^* = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} X_j^4 \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]}{\sum_{j=1}^{k}\sum_{i=1}^{n} p_{ji} \exp\left[-4,5\left(\frac{X_j - X_i}{\rho}\right)^2\right]} - \qquad (7)$$

$$-4\mu_3^* m_X^* - 6\mu_2^*\left(m_X^*\right)^2 - \left(m_X^*\right)^4,$$

где X_i – экспериментальные данные;

X_j – середина j-го разряда ($j = \overline{1,k}$), полученного путём деления отрезка (b-a) на k частей (a – начало, b – конец отрезка эквивалентной выборки);

p_{ji}=1 при условии накрывания величиной X_j ядра на основе X_i;

ρ – половина ширины ядра.

Тогда коэффициент асимметрии можно подсчитать по формуле

$$r_3^* = \frac{\mu_3}{\left(\mu_2^*\right)^{3/2}}, \qquad (8)$$

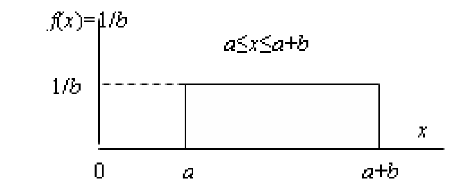
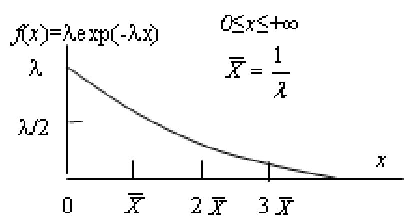
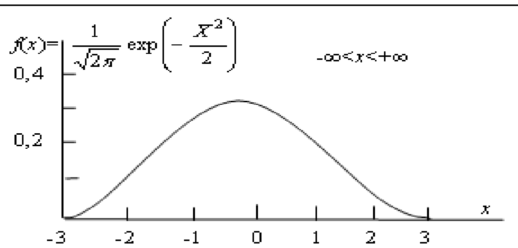
а коэффициент эксцесса – по формуле

$$r_4^* = \frac{\mu_4}{\left(\mu_2^*\right)^2}. \qquad (9)$$

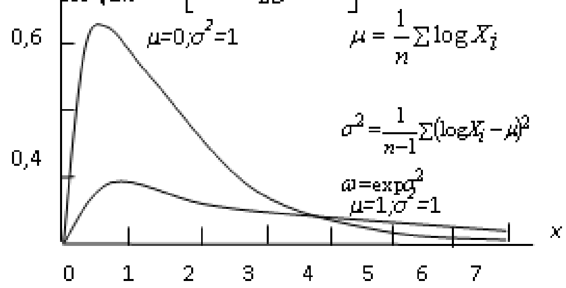
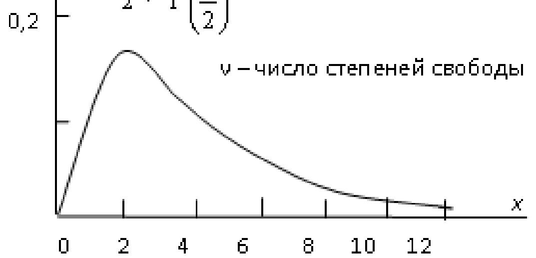
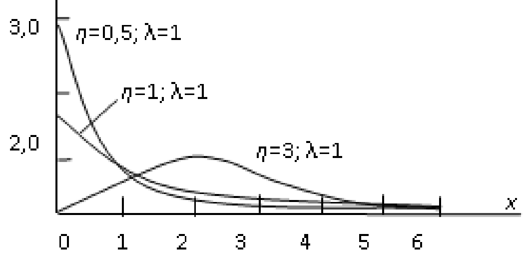
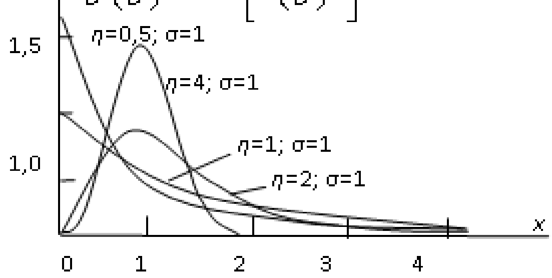
Поскольку вычисления выборочных моментов по формулам (4)-(9) неизбежно являются приблизительными, то для ориентировки в таблице 1 приведены теоретические значения коэффициентов асимметрии и эксцесса [5].

Таблица 1

Теоретические значения коэффициентов асимметрии (r_3) и эксцесса (r_4) некоторых распределений

Распределение	Вид и форма	r_3	r_4
Равномерное		0	1,8
Экспоненциальное		2	9
Нормальное		0	3

Окончание табл.1

Распределение	Вид и форма	r_3	r_4
Логарифмически нормальное	$f(x)=\frac{1}{\sigma X\sqrt{2\pi}}\exp\left[-\frac{(\log X-\mu)^2}{2\sigma^2}\right]$ 	$(\omega+2)\sqrt{\omega-1}$	$\omega^4+2\omega^3+$ $+3\omega^2-3$
Хи-квадрат	$f(X)=\frac{X^{(\nu-2)/2}\exp\left(\frac{X}{2}\right)}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)}$ 	$2\sqrt{\frac{2}{\nu}}$	$3+\frac{12}{\nu}$
Гамма	$f(X)=\frac{\lambda^{\eta}}{\Gamma(\eta)}X^{\eta-1}e^{-\lambda X}$ 	$\frac{2}{\sqrt{\eta}}$	$\frac{3(\eta+2)}{\eta}$
Вуйбулла	$f(X)=\frac{\eta}{\sigma}\left(\frac{X}{\sigma}\right)^{\eta-1}\exp\left[-\left(\frac{X}{\sigma}\right)^{\eta}\right]$ 	*	**

* $$r_3=\left\{\Gamma\left(1+\frac{3}{\eta}\right)-3\Gamma\left(1+\frac{2}{\eta}\right)\Gamma\left(1+\frac{1}{\eta}\right)+2\left[\Gamma\left(1+\frac{1}{\eta}\right)\right]^3\right\}\Bigg/\left\{\Gamma\left(1+\frac{2}{\eta}\right)-\left[\Gamma\left(1+\frac{1}{\eta}\right)\right]^2\right\}^{3/2};$$

** $$r_4=\left\{\Gamma\left(1+\frac{4}{\eta}\right)-4\Gamma\left(1+\frac{3}{\eta}\right)\Gamma\left(1+\frac{1}{\eta}\right)+6\Gamma\left(1+\frac{2}{\eta}\right)\left[\Gamma\left(1+\frac{1}{\eta}\right)\right]^2-3\left[\Gamma\left(1+\frac{1}{\eta}\right)\right]^4\right\}\Bigg/\left\{\Gamma\left(1+\frac{2}{\eta}\right)-\left[\Gamma\left(1+\frac{1}{\eta}\right)\right]^2\right\}^4.$$

Пример 1. Пусть в результате измерений контрольной выборки в некотором технологическом процессе получены значения (в условных единицах) X: 0, 5, 11, 17, 23, 28, 33, 39.

Определить вид распределения.

Решение. С помощью метода точечных распределений (МТР) найдём выборочные моменты

$m_X^* = 19,62$;

$\mu_2^* = 11,3917$;

$\mu_3^* = 4,0451$, откуда $r_3^* = 0,0027 \to 0$;

$\mu_4^* = 28260,1901$, откуда $r_4^* = 1,6781 \to 1,8$.

По номограмме рис. 1 с большой долей вероятности можно сделать предположение, что выборочное распределение стремится к равномерному виду.

2. Определение интервальных оценок центральных моментов

В отличие от генеральных совокупностей, параметры которых имеют только однозначно неслучайные величины, выборки характеризуются оценками генеральных параметров, которые имеют при некоторой доверительной вероятности минимальные и максимальные значения, внутри которых находятся и сами генеральные параметры, причем в большинстве случаев расположенные несимметрично (интервальные оценки).

Среди всех начальных и центральных методов главенствующую роль по частоте использования имеют средние арифметические (3) и выборочные дисперсии (2 при h= 2).

Проблема интервальных оценок для выборок большого объема хорошо исследована (например, [6]):

$$\bar{X} - t_{\text{табл}}(q;\nu)\sqrt{\frac{S^2}{n}} < M[X] < \bar{X} + t_{\text{табл}}; \quad (10)$$

$$nS^2 / \chi_2^2 < \sigma^2 < nS^2 / \chi_1^2 , \quad (11)$$

где n – объем выборки; $t_{\text{табл}}(q;\nu)$ – табличный критерий Стьюдента при q – уровне значимости и ν=n-1 – числе степеней свободы; $\chi_1^2\left(1-q/2;\nu\right)$ и $\chi_2^2\left(q/2;\nu\right)$ – хи-квадрат распределение при $q/2$ и $1-q/2$ уровнях значимости и ν – числе степеней свободы; $S^2 = \mu_2$ – выборочная дисперсия.

Проблема интервальных оценок для выборок малого объема исследована не так хорошо и оставляет простор для дискуссий. Понятно, что использую формулы (10) и (11) для выборок малого объема мы получим чрезмерно большие интервалы (пример 2). Для их уменьшения в работе [7] предлагается воспользоваться следующими выражениями, базирующимся на методе точечных распределений:

– для нормального и других симметричных распределений

$$m_X^* - U_{\frac{1-\beta}{2}}\sqrt{\mu_2^*} < M[X] < m_X^* + U_{\frac{1+\beta}{2}}\sqrt{\mu_2^*} ; \quad (12)$$

– для экспоненциального и аналогичных ему распределений

$$U_{\frac{1-\beta}{2}} \cdot m_X^* < M[X] < U_{\frac{1+\beta}{2}} \cdot m_X^* ; \quad (13)$$

– для распределения Вейбулла и аналогичных ему

$$U_{\frac{1-\beta}{2}} \cdot \frac{m_X^*}{0,886} < M[X] < U_{\frac{1+\beta}{2}} \cdot \frac{m_X^*}{0,886}, \quad (14)$$

– для выборочной дисперсии в случае нормального, экспоненциального и подобных распределений

$$\frac{\mu_2^*}{U_{\frac{1+\beta}{2}}} < M[\mu_2] < \frac{\mu_2^*}{U_{\frac{1-\beta}{2}}} ; \quad (15)$$

– для распределения Вейбулла и аналогичных ему

$$\frac{0,214\mu_2^*}{U_{\frac{1+\beta}{2}}} < M[\mu_2] < \frac{0,214\mu_2^*}{U_{\frac{1-\beta}{2}}} , \quad (16)$$

где β - доверительная вероятность, а U_2 – соответствующая квантиль.

Однако, на наш взгляд, существует ещё одна возможность получить интервальные оценки повышенной точности для выборок малого объема. Дело в том, что согласно основополагающей концепции метода точечных распределений, количество информации до и после всех преобразований не должно изменяться, а , следовательнго, можно записать равенство

$$D_n(\beta)\sqrt{n_Э} = D(\beta)\sqrt{n} ,$$

откуда

$$n_Э = n\left[\frac{D(\beta)}{D_n(\beta)}\right]^2 , \quad (17)$$

где D(β) – статистика Колмогорова при доверительной вероятности β [8]; $D_n(\beta)$ – аналогичная ей статистика МТР [7]; n – объем экспериментальной выборки малого объема; $n_э$ – эквивалентный объем выборки, обеспечивающий ту же точность D(β) при расчете по классическим формулам.

Некоторые данные расчетов по формуле (17) приведены в табл. 2.

Таблица 2

Реальные n и эквивалентные $n_э$ объемы выборок

n	3	4	5	6	7	8	9	10
D/D_n	1,863	1,891	1,876	1,821	1,791	1,780	1,741	1,690
$n_э$	10	14	17	20	23	25	27	29
$\sigma(S)/\sigma$, %	46,6	38,9	34,1	30,7	28,2	26,2	24,5	23,2
$\sigma\left(\sqrt{\mu_2^*}\right)\Big/\sigma$, %	23,2	19,4	17,5	16,1	15,0	14,3	13,8	13,3

Тогда, опираясь на эквивалентные объемы выборок $n_э$, можно использовать для расчета интервальных оценок моментов классические формулы (10) и (11).

$$m_X^* - t_{табл}(q;\nu)\sqrt{\frac{\mu_2^*}{n_Э}} < M[X] < \quad (18)$$

$$< m_X^* + t_{табл}(q;\nu)\sqrt{\frac{\mu_2^*}{n_Э}};$$

$$\frac{n_Э \mu_2^*}{\chi_2^2} < \sigma^2 < \frac{n_Э \mu_2^*}{\chi_1^2}, \quad (19)$$

где $\nu = n_Э - 1$;

$$\chi_1^2\left(1 - {}^q\!/_2;\nu = n_Э - 1\right);$$

$$\chi_2^2\left({}^q\!/_2;\nu = n_Э - 1\right).$$

Произведём сравнение всех методов расчета интервальной оценки выборочной дисперсии с точки зрения их эффективности.

Пример 2. Пусть в результате измерений контрольной выборки в некотором технологическом процессе получены значения величин контролируемого параметра, распределенного по нормальному закону, X: 5,73; 6,40; 7,70; 8,07; 8,47.

Найти интервальные оценки выборочной дисперсии всеми возможными способами.

Решение. Объем выборки n=5, а её параметры вычисленные по формулам (3) и (2), равны

$$\bar{X} = 7,274;$$

$$S^2 = \mu_2 = 1,3483.$$

Те же вычисления МТР по формулам (4) и (5) равны

$$m_X^* = 7,286; \quad \mu_2^* = 1,7370.$$

Тогда согласно формуле (11) и таблицам χ^2[8]

$$\frac{5 \cdot 1,3483}{10,14} < \sigma^2 < \frac{5 \cdot 1,3483}{0,484}$$

или $0,605 < \sigma^2 < 13,929$,

согласно формуле (15)

$$\frac{1,737}{1,960} < \sigma^2 < \frac{1,737}{0,125}$$

или $0,886 < \sigma^2 < 13,896$,

согласно формуле (19)

$$\frac{17 \cdot 1,737}{28,85} < \sigma^2 < \frac{17 \cdot 1,737}{6,908}$$

или $1,024 < \sigma^2 < 4,275$.

Для наглядности представим полученные результаты в графическом виде (рис. 2).

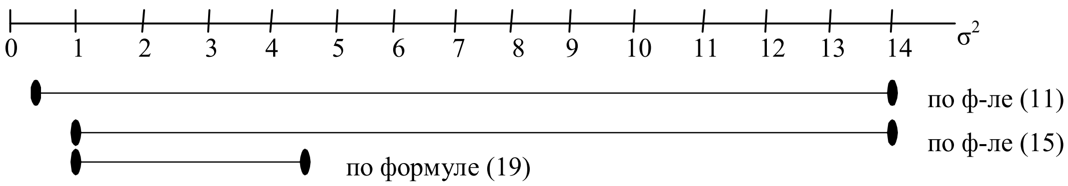


Рис. 2. Величины интервальных оценок выборочных дисперсий, вычисленных по разным формулам.

Выводы

Разработан метод, позволяющий в случае выборок малого объема (n=3÷16) определять хотя бы приблизительно вид закона распределения генеральной совокупности, из которой взята конкретная выборка, а также в 3-4 раза сокращать размер интервальной оценки выборочной дисперсии, что позволяет в 2-2,5 раз уменьшить долю ложной приемки и ложной браковки за счет уменьшения ошибки вычисления параметров контрольных выборок малого объема.

Литература

1. Dolgov, Y.A. Stochastic Check for Control of Electronic Wares Quality [Текст] / Y.A. Dolgov, A.Y. Dolgov // Trans. of 10-th Intern. Symp. of Apllied Stochastic Models and Data Analysis. – 12-15 June

2001. – V. 1/2. – Univer.Technologic de Compiegne. – France. – P. 387 – 390.

2. Долгов, Ю.А. Метод повышения точности вычисления параметров выборки малого объема (метод точечных распределений) [Текст] / Ю.А. Долгов, А.Ю. Долгов, Ю.А. Столяренко // Вестник ПГУ-2010. – Юб. вып.– С. 232 – 242.

3. Хан, Г. Статистические модели в инженерных задачах [Текст] / Г. Хан, С. Шапиро. – М.: Мир, 1969. – 396 с.

4. Шор, Я.Б. Статистические методы анализа и контроля качества и надежности [Текст] / Я.Б. Шор. – М.: Сов. радио, 1962. – 553 с.

5. Хастингс, Н. Справочник по статистическим распределениям [Текст] / Н. Хастингс, Дж. Пикок. – М.: Статистика, 1980 – 95 с.

6. Митропольский, А.К. Техника статистических вычислений [Текст] / А.К. Митропольский. – 2-е изд., перераб. и доп. – М.: Наука, 1971. – 576 с.

7. Гаскаров, Д.В. Малая выборка [Текст] / Д.В. Гаскаров, В.И. Шаповалов. – М.: Статистика, 1978. – 248 с.

8. Большев, Л.Н. Таблицы математической статистики [Текст] / Л.Н. Большев, Н.В. Смирнов. – 3-е изд.– М.: Наука, 1983. – 416 с.

Поступила в редакцию 2.03.2012

Рецензент: д-р техн. наук, проф. Р.Б. Дунец, Национальный университет «Львовская политехника», Львов, Украина

ДОСЛІДЖЕННЯ ІНТЕРВАЛЬНИХ ОЦІНОК І МЕТОДІВ ВИЩОГО ПОРЯДКУ ДЛЯ ВИБІРОК МАЛОГО ОБ'ЄМУ

Ю.О. Долгов

Пропонуються методи визначення законів розподілу і параметрів вибірок малого об'єму, а також методи зменшення обчислювальних помилок цих параметрів. Розроблений метод, що дозволяє у разі вибірок малого об'єму визначати хоч би приблизно вид закону розподілу генеральної сукупності, з якої узята конкретна вибірка, а також в 3-4 рази скорочувати розмір інтервальної оцінки вибіркової дисперсії, що дозволяє в 2-2,5 разів зменшити частку помилкового приймання і помилкового бракування за рахунок зменшення помилки обчислення параметрів контрольних вибірок малого об'єму.

Ключові слова: вибірка малого об'єму, центральні моменти, точкові і інтервальні оцінки.

THE INVESTIGATION OF INTERVAL ASSESSMENTS AND METHODS OF THE MAXIMUM ORDER FOR SMALL SIZE SAMPLES

Y.A. Dolgov

There are offered methods of definition of laws of distribution and parametres of small size samples, and also methods of reduction of computing errors of these parametres. A method, allowing in the case of selections of small volume to determine even approximately type of law of distributing of general aggregate which a concrete selection is taken from, is developed, and also in 3-4 times to abbreviate the size of interval estimation of selective dispersion, that allows in 2-2,5 times to decrease the stake of false formal acceptance and false rejection due to diminishing of error of calculation of parameters of control selections of small volume.

Key words: small size sample, the central moments, dot and interval estimations.

Долгов Юрий Александрович – д-р техн. наук, профессор, зав. кафедрой «Информационных технологий и автоматизированного управления производственными процессами» Приднестровского государственного университета им. Т.Г. Шевченко (ПГУ), Тирасполь, Приднестровье, e-mail: dolax@ mail333.com.