Математичні моделі та методи

УДК 519. 232.2

В.Ю. Дубницкий, А.И. Ходырев

Харьковский учебно-научный институт ГВУЗ «Университет банковского дела», Харьков

ИСПОЛЬЗОВАНИЕ ФУНКЦИЙ НЕАНАЛИТИЧЕСКИХ СРЕДНИХ ВЕЛИЧИН ДЛЯ ОЦЕНКИ СОГЛАСОВАННОСТИ ВЫЧИСЛИТЕЛЬНЫХ ПРОЦЕССОВ

Поставлена задача проверки наличия резко выделяющихся данных в выборке из трёх одинаково распределённых случайных величин. В работе рассмотрены следующие распределения: Вейбулла, гамма, нормальное распределение существенно-положительных величин (полунормальное распределение), обратное распределение Гаусса (распределение Вальда), Лапласа, логистическое распределение, распределение Максвелла, нормальное распределение. По результатам статистического моделирования установлено, что, независимо от исходных распределений, распределение отношения «размах выборки /медиана» для выборок из трёх наблюдений распределено по логлогистическому закону. Приведены способы получения оценок параметров логлогистического распределения по методу моментов и методу максимума правдоподобия, получены выражения для определения величины доверительных интервалов полученных оценок. Определена нижняя граница правосторонней критической области для принятия решения о согласованности результатов вычислений при использовании критерия «размах выборки/медиана выборки».

Ключевые слова: *размах выборки, медиана, распределение Вейбулла, гамма-распределение, нормальное распределение существенно-положительных величин (полунормальное распределение), обратное распределение Гаусса (распределение Вальда), распределение Лапласа, логистическое распределение, распределение Максвелла, нормальное распределение, логарифмически логистическое распределение, метод моментов, метод максимума правдоподобия, оценка параметров логарифмически логистического распределения.*

Введение

До середины XX века даже очень ответственные показатели свойств материалов оценивали по значению средней арифметической, полученной по результатам трёх параллельных испытаний [1, 2]. Ситуация резко изменилась после выхода работы [3], ставшей ныне классической. Проблема малых выборок при оценке свойств материалов и изделий из них приобрела новый смысл после введения понятия принципиально малой выборки [4], то есть такой выборки, которая в данной ситуации не может быть увеличена ни при каких обстоятельствах. Но и при этих условиях речь шла о выборках, содержащих не менее четырех наблюдений. Для различного рода статистических выводов были предложены ядерные методы (парзеновские окна). Современное состояние этих методов изложено в работах [5, 6]. В последние годы интерес к проблеме трёх измерений вновь появился, но уже в связи с новыми обстоятельствами [7].

Анализ литературы. Статистический анализ проблемы трёх измерений, в его современном понимании, был начат в работах [8, 9]. В работе [8] была рассмотрена выборка из трёх наблюдений и изучено распределение следующих статистик: отношение разности двух ближайших измерений к размаху выборки, половины разности двух ближайших наблюдений и среднее из двух ближайших наблюдений. Исходными распределениями генеральной совокупности приняты нормальное, равномерное, треугольное распределения. Для указанных выше статистик определены их математические ожидания. Процентные точки определены только для первой из названных статистик, что существенно ограничивает область применения результатов работы. В работе [9] рассмотрена задача отбраковки сомнительных наблюдений при несколько иных условиях и выборках с объёмом от пяти наблюдений. В рассматриваемой работе принято, что известны параметры исходного нормального распределения, параметры нормального распределения, засоряющего исходное и доля «сорных» наблюдений. При этих допущениях в работе [9] рассчитаны таблицы процентных точек для критериев определения сомнительных наблюдений. В качестве критерия использовано отношение разности наибольшего (наименьшего) значения выборки и оценки математического ожидания этой выборки к своему среднеквадратическому отклонению. К недостаткам такого подхода по нашему мнению следует отнести то, что вследствие неустойчивости оценок параметров к резко выделяющимся значениям эти характеристики могут давать искаженные результаты.

В работе [7] задача оценивания результатов трёх измерений использована для обобщения результатов контроля непрерывных технологических процессов. В этой работе значение каждой из переменных трактуется как результат определения состояния системы, выполненный одним из трёх независимых методов. По мнению авторов работы [7] в простейшей постановке задача о трёх измерениях x_1, x_2, x_3 неизвестной величины x сводится к построению «по возможности более точной оценки $\hat{\delta}$ величины x». В качестве такой оценки в работах [7, 10] рекомендуется применение различных видов средних величин и иных линейных и нелинейных оценок.

Особо следует рассмотреть используемые в этих работах виды средних величин. Современное состояние теории средних величин изложено в работах [11…13].

Пусть $C = x_1, x_2, ..., x_n$ – действительные числа и $D = f(x_1, x_2, ..., x_n)$.

Если

$$\min f(x_1, x_2, ..., x_n) \le f(x_1, x_2, ..., x_n) \le \le \max(x_1, x_2, ..., x_n); \quad (1)$$

то функцию D называют средней по Коши или слабой средней, определенной на множестве X.

Пусть

$$K = (C, \lambda,)\lambda \ge 0; \quad (2)$$

и

$$D = f(x_1, x_2, ..., x_n) - \quad (3)$$

действительная и непрерывная функции.

Тогда для средней величины по Колмогорову (сильной средней) должны быть выполнены следующие условия:

однородности:

$$f(\lambda x_1, \lambda x_2, ..., \lambda x_n) = = \lambda f(x_1, x_2, ..., x_n); \quad (4)$$

симметрии:

$$f(x_1, x_2, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n) = f(x_i, x_2, ..., x_n, x_{i+1}, ..., x_{i-1}); \quad (5)$$

возрастания:

$$(x_1 \le y_{1,...} x_i \le y_i, ..., x_n \le y_n) \Rightarrow \Rightarrow (f(x_1, x_2, ..., x_n) \le f(y_1, y_2, ..., y_n)). \quad (6)$$

Рассмотрим величину:

$$^m A = \sqrt[m]{\frac{x_1^m + x_2^m ... + x_n^m}{n}}, \quad (7)$$

называемую средней степенной. Очевидно, что при m = 1 получим среднюю арифметическую величину, при m = –1 получим среднюю гармоническую, при m = 2 среднеквадратическую величину.

В работе [11] показано, что:

$$\lim_{m \to 0} {}^m A = \frac{1}{n}\sum_{i=1}^{n} \ln x_i = \left(\prod_{i=1}^{n} x_i\right)^{1/n}. \quad (8)$$

По классификации, принятой в работе [11], эти виды средних величин относят к классу аналитических средних. Выражения для их вычисления приведены в табл. 1.

Таблица 1

Современные представления аналитических средних величин

Вид средней	Дискретные величины	Непрерывные величины
Средняя арифметическая	$A = (x_1 + ... + x_n)/n$	$A = \int_a^b xp(x)dx \Big/ \int_a^b p(x)dx$
Средняя геометрическая	$G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$	$G = \exp\left(\int_a^b \ln xp(x)dx \Big/ \int_a^b p(x)dx\right)$
Средняя гармоническая	$H = n/(1/x_1 + \ldots + 1/x_n)$	$H = \int_a^b p(x)dx \Big/ \int_a^b (p(x)/x)\,dx$

Неаналитическими средними в работе [11] названы медиана выборки, её максимальное и минимальное значения.

Медиану определяют по условию:

$$me = \begin{cases} x_{[(n+1)/2]}, & n = 2k+1; \\ \frac{1}{2}\left(x_{[n/2]} + x_{[(n+1)/2]}\right), & n = 2k. \end{cases} \quad (9)$$

Максимальное и минимальное значение выборки может быть получено из степенной средней в результате предельных переходов, которые приведены в условиях (10) и (11). Для упорядоченной последовательности $x_1 \le x_2 \le ... \le x_n$ максимальный элемент можно определить по условию:

$$x_n = \lim_{m \to \infty} x_n \cdot \left[\frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{x_n}\right]^{1/m}; \quad (10)$$

минимальный элемент можно определить по условию:

$$x_1 = \lim_{m \to \infty} x_1 \left[\frac{1}{n}\sum_{i=1}^{n} \frac{x_i}{x_1}\right]^{1/m}. \quad (11)$$

Следует отметить, что описанные в этих работах методы не могут ответить на вопрос о наличии непротиворечивых результатов в полученных тройках измерений. Методы, ориентированные на отбраковку выделяющихся наблюдений, изложенные в работах [3, 4, 8, 9] требуют больший объем исследуемой выборки, методы, изложенные в работах [7, 10] направлены на получение некоей обобщённой оценки. Таким образом, следует считать, что задача о наличии непротиворечивых данных в выборке из трех наблюдений в доступной авторам литературе в полном объеме не рассмотрена.

Постановка задачи

Определим метод вычислений (М) предназначенный для решения какой-либо задачи в виде:

$$M = \langle S, A.Pr, H \rangle ; \tag{12}$$

где S – математическая модель задачи, A – алгоритм реализации принятой математической задачи, Pr – программная реализация алгоритма, H – аппаратная реализация программы. Методы вычислений будем считать различными, если они отличаются, хотя бы по одному из признаков, названных в условии (12). При выполнении особо ответственных вычислений, как правило, выполняют их диверсификацию. На практике это приводит к выполнению расчетов различными методами и сравнению полученных результатов с целью проверки их непротиворечивости. В рамках данной работы принято, что количество различных методов равно трём.

Цель работы – разработка статистически обоснованного правила принятия решения для оценки непротиворечивости результатов диверсификации вычислительного процесса.

Исходные допущения

Поставленная в работе задача (цель работы) была решена по результатам численного моделирования. Схема проведения эксперимента показана на рис. 1.

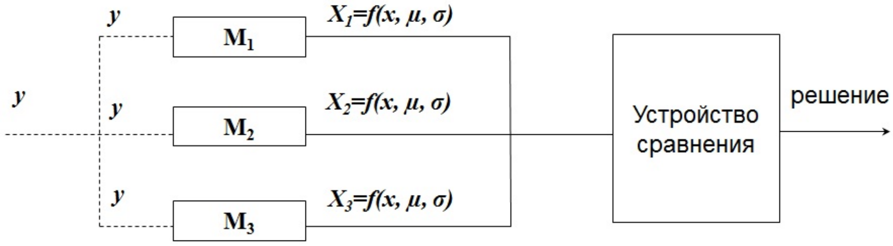

Рис. 1. Схема проведения эксперимента: **y** – входные данные, M_1, M_2 , M_3 – методы расчёта; X_1 , X_2, X_3 – результаты расчёта

Входные данные Y, одинаковые для всех трёх методов, преобразуют в результаты вычислений тремя различными методами. Результаты вычислений, полученных этими методами, принимают одинаково распределёнными случайными величинами с известным законом распределения и его параметрами. Устройство сравнения по разработанному в работе алгоритму вырабатывает решение о наличии (отсутствии) в результатах вычислений резко выделяющихся (противоречивых данных)

Полученные результаты

Сведения об использованных в работе законах распределения, моделирующих результаты вычислений, приведены в табл. 2.

Для всех рассмотренных законов распределения их параметры определены при условии, что среднее значение генерируемой последовательности $\bar{x} = 10$; среднеквадратическое отклонение $s = 1$. Способ решения такой задачи описан в работе [14]. Результаты приведены в табл. 3.

Для проведения численных экспериментов использовали следующую процедуру. Для каждого указанного в табл. 2, 3 закона распределения получали выборку псевдослучайных чисел $X = x_1, x_2, \ldots, x_g$, в нашем случае g = 300. После чего из этой выборки формировали тройки чисел:

$$T = \left\langle (x_1, x_2, x_3)_1 \ldots (x_{g-2}, x_{g-1}, x_g)_n \right\rangle, \quad n = g/3, \quad n = 100. \tag{13}$$

Каждая из этих троек имитировала три независимых результата вычислений. Для каждой тройки вычисляли функцию вида:

$$X_{1i} = \frac{\max(x_{i-1}, x_i, x_{i+1}) - \min(x_{i-1}, x_i, x_{i+1})}{\operatorname{med}(x_{i-1}, x_i, x_{i+1})}, \quad i = 2, \ldots, n-1. \tag{14}$$

Величина X_{1i} –это отношение неаналитических средних, характеризующих данную тройку чисел. Медиана выбрана благодаря её свойству робастности (устойчивости), что обосновано в работе [15].

Таким образом, условие (14) – это функция неаналитических средних величин. В результате применения системы Statgraphics XV.1 установлено, что для всех упомянутых в табл. 2, 3 исходных законов распределения распределение величины, определяемой условием (14), то есть распределение отношения размаха выборки к своей медиане, подчиняется закону логарифмически логистического (логлогистического) распределения. Подробно свойства этого распределения описаны в работе [16], получение оценок параметров этого распределения и их свойства описаны в работе [17]. Плотность этого распределения имеет вид:

Таблица 2

Законы распределения случайных величин, использованных для моделирования результатов вычислений

Наименование закона распределения	Функция плотности закона распределения
Распределение Вейбулла	$f(x)=\frac{\alpha}{\beta}x^{\alpha-1}\exp\left\{-\left(x/\beta\right)^{\alpha}\right\}$
Гамма-распределение	$f(x)=\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}\big/\Gamma(\alpha)$
Нормальное распределение существенно-положительных величин (полунормальное распределение)	$f(x)=\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\cdot\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}\right)$
Обратное распределение Гаусса (распределение Вальда)	$f(x)=\frac{1}{x}\cdot\frac{\sqrt{\beta}}{\exp(z/2)}\varphi\left[\sqrt{\beta}\left(\frac{e^{z}-1}{e^{z/2}}\right)\right];\ z=\ln(x/\theta)$
Распределение Лапласа	$f(x)=\frac{\lambda}{2}\exp\left(-\lambda\lvert x-\mu\rvert\right)$
Логистическое распределение	$f(x)=\frac{1}{\sigma}\frac{\exp(z)}{\left[1+\exp(z)\right]^{2}};\ z=\frac{x-\mu}{\sigma}$
Распределение Максвелла	$f(x)=\sqrt{\frac{2}{\pi}}\cdot\frac{(x-\theta)^{2}}{\beta^{3}}\exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\beta}\right)^{2}\right]$
Нормальное распределение	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-(x-\mu)^{2}\big/\left(2\sigma^{2}\right)\right]$

Таблица 3

Численные значения параметров законов распределения случайных величин, использованных для моделирования результатов вычислений

Наименование закона распределения	Численные значения параметров законов распределения случайных величин
Распределение Вейбулла	α=12,15; β=1,043
Гамма-распределение	α=12,15; λ=100
Полунормальное распределение	μ=8,68; σ=1,66
Распределение Вальда	θ=10; β=100
Распределение Лапласа	λ=0,144; μ=10
Логистическое распределение	μ=10; σ=1
Распределение Максвелла	β=1,48; θ=7,64
Нормальное распределение	μ=10; σ=1

$$f(X)=\frac{1}{\sigma X}\cdot\frac{\exp z}{\left[1+\exp z\right]^{2}}, \tag{15}$$

где
$$z=\frac{\ln X-\mu}{\sigma},\quad X>0. \tag{16}$$

Подставив условие (16) в (15) получим выражение:

$$f(X)=\frac{1}{\sigma X}\cdot\frac{\exp\left(\left(\ln X-\mu\right)/\sigma\right)}{\left[1+\exp\left(\left(\ln X-\mu\right)/\sigma\right)\right]^{2}}, \tag{17}$$

которое, после упрощения, примет вид:

$$f(X)=e^{\mu/\sigma}X^{(1-\sigma)/\sigma}\Big/\left(\sigma\left(X^{1/\sigma}+e^{\mu/\sigma}\right)^{2}\right). \tag{18}$$

В условиях (15…18) принято, что μ – параметр масштаба, $-\infty<\mu<\infty$; σ – параметр формы, $0<\sigma<\infty$. В первом приближении оценки $\hat{\mu},\hat{\sigma}$ параметров μ, σ получены методом моментов. При получении этих оценок следует учесть, что, как показано в работе [17], при $\sigma\geq 1$ не существует математического ожидания, при $\sigma\geq 0{,}5$ не существует среднеквадратического отклонения. Используя приведенные в [17] соотношения получим следующие оценки параметров логлогистического закона распределения. медиана:

$$\widehat{\text{Ì å}}_{X}=e^{\mu}; \tag{19}$$

математическое ожидание:

$$M_X=e^{\mu}\Gamma(1+\sigma)\Gamma(1-\sigma); \tag{20}$$

дисперсия:

$$s^{2}=e^{2\mu}\left[\Gamma(1+2\sigma)\Gamma(1-2\sigma)-\left(\Gamma(1+\sigma)\Gamma(1-\sigma)\right)^{2}\right]; \tag{21}$$

коэффициент вариации:

$$v=\frac{s}{m}=\frac{\sqrt{\Gamma(1+2\sigma)\Gamma(1-2\sigma)-[\Gamma(1+\sigma)\Gamma(1-\sigma)]^{2}}}{\Gamma(1+\sigma)\Gamma(1-\sigma)}. \tag{22}$$

Численное значение $\hat{\sigma}$ параметра σ определено из условия (22). Используя свойства гамма-функции, указанные в работе [18]:

$$\Gamma(z+1)=z\Gamma(z),\ \Gamma(z)\approx\sqrt{\frac{2\pi}{z}}\cdot e^{-z}z^{z}\left(1+\frac{1}{12z}\right); \tag{23}$$

получим, что:

$$v=\frac{\sqrt{2\sigma\,\Gamma(2\sigma)\Gamma(1-2\sigma)-[\sigma\,\Gamma(\sigma)\Gamma(1-\sigma)]^{2}}}{\sigma\,\Gamma(\sigma)\Gamma(1-\sigma)}. \tag{24}$$

Принимая во внимание указанные выше ограничения на область допустимых значений параметра формы ($\sigma < 0{,}5$), величина $\hat{\sigma}$ – корень уравнения (24) – определена методом половинного деления [19]. Используя условие (20) оценка $\hat{\mu}$ параметра масштаба μ определена в виде:

$$\hat{\mu} = \ln\left(m \Big/ \left(\hat{\sigma}\,\Gamma\left(\hat{\sigma}\right)\Gamma\left(1-\hat{\sigma}\right)\right)\right). \qquad (25)$$

Оценки, полученные по методу моментов, использованы в качестве начальных в процедуре получения оценок по методу максимума правдоподобия [20]. С учётом условия (17) функция правдоподобия для логлогистической плотности вероятности примет вид:

$$L(x;\mu,\sigma) = \prod_{j=1}^{n} \frac{e^{\mu/\sigma} X_j^{(1-\sigma)/\sigma}}{\sigma\left(X_j^{1/\sigma} + e^{\mu/\sigma}\right)^2}\,. \qquad (26)$$

Логарифм функции правдоподобия примет вид:

$$\operatorname{Ln} L(x;\mu,\sigma) = n \cdot \mu / \sigma +$$

$$+\frac{1-\sigma}{\sigma}\sum_{j=1}^{n} \ln x_j - n\ln\sigma - 2\sum_{j=1}^{n} \ln\left(x_j^{1/\sigma} + e^{\mu/\sigma}\right). \qquad (27)$$

Исходя из условий использования метода максимума правдоподобия, потребуем выполнение условий:

$$\frac{\partial \ln L}{\partial \mu} = 0\,, \quad \frac{\partial \ln L}{\partial \sigma} = 0\,. \qquad (28)$$

Выполняя условия (28) получим систему уравнений:

$$\begin{cases} \dfrac{n}{\sigma} - 2e^{\mu/\sigma}\displaystyle\sum_{j=1}^{n}\left[\sigma\left(x_1^{1/\sigma} + e^{\mu/\sigma}\right)\right]^{-1} = 0; \\ \dfrac{2}{\sigma^2}\displaystyle\sum_{j=1}^{n}\frac{x_j^{1/\sigma}\ln x_j + \mu e^{\mu/\sigma}}{x_j^{1/\sigma} + e^{\mu/\sigma}} - \\ \quad -\left(\dfrac{n\mu}{\sigma^2} + \dfrac{1}{\sigma^2}\displaystyle\sum_{j=1}^{n}\ln x_j + \dfrac{n}{\sigma}\right) = 0. \end{cases} \qquad (29)$$

Принимая оценки параметров $\hat{\sigma}, \hat{\mu}$, полученные методом моментов, в качестве начальных и решая систему (29) получим значения $\tilde{\sigma}, \tilde{\mu}$ - оценки параметров логлогистического закона распределения, полученные по методу максимума правдоподобия. В табл. 4 приведены численные значения полученных оценок.

Результатом данного этапа работы следует считать установление того факта, что для распределений, перечисленных в табл. 4, распределение отношения размаха выборки к медиане для выборки из трёх наблюдений не зависит от исходных распределений. Полученный результат позволил получить значения правосторонних процентных точек распределения отношения вида (14), которые приведены в табл. 5.

Таблица 4

Параметры логлогистического закона распределения
отношения «размах выборки / медиана выборки»

Вид закона распределения	Выборочные параметры закона распределения и его числовые характеристики				
	Среднее значение, $\overline{X}_1$	Среднеквадратическое отклонение, s	Медиана, $\widehat{\text{Ì å}}_X$	Параметр масштаба, μ	Параметр формы, σ
Распределение Вейбулла	1,576	0,828	1,421	0,3506	0,145
Гамма-распределение	2,588	1,831	2,156	0,768	0,253
Полунормальное распределение	12,450	31,079	12,450	2,522	0,387
Распределение Вальда	5,875	8,228	4,153	1,423	0,353
Распределение Лапласа	2,141	0,864	1,986	0,686	0,140
Логистическое распределение	2,372	2,493	1,965	0,675	0,199
Распределение Максвелла	4,024	6,185	2,997	1,097	0,286
Нормальное распределение	2,288	1,163	2,018	0,702	0,221

Таблица 5

Нижняя граница правосторонней критической области
критерия «размах выборки / медиана выборки»

Вид закона распределения	Уровни доверительной вероятности α				
	0,05	0,025	0,01	0,005	0,001
Распределение Вейбулла	2,177	0,025	0,01	3,061	3,868
Гамма-распределение	4,541	5,447	6,895	8,227	12,375
полунормальное распределение	38,908	51,393	73,702	96,566	180,32
распределение Вальда	11,742	15,136	21,029	26,906	47,556
Распределение Лапласа	2,999	3,316	3,778	4,166	5,223
Логистическое распределение	3,530	4,073	4,903	5,634	7,767
Распределение Максвелла	6,956	8,545	11,152	13,619	21,605
Нормальное распределение	3,868	4,534	5,571	6,501	9,286

Эти результаты позволили сформулировать следующее правило принятия решения. Если величина критерия «размах выборки/медиана выборки» для соответствующего закона распределения исходных данных и принятого уровня доверительной вероятности α меньше табличного значения, то итог вычислений не содержит резко выделяющихся результатов.

Для определения доверительных интервалов оценок параметров логлогистического распределения, полученных методом максимума правдоподобия в соответствии с работой [20], следует определить элементы информационной матрицы Фишера:

$$\begin{pmatrix} \mathrm{Var}(\mu) & \mathrm{Cov}(\mu,\sigma) \\ \mathrm{Cov}(\mu,\sigma) & \mathrm{Var}(\sigma) \end{pmatrix} =$$

$$= \left.\begin{pmatrix} -\dfrac{\partial^2 L}{\partial \mu^2} & -\dfrac{\partial^2 L}{\partial \mu \partial \sigma} \\ -\dfrac{\partial^2 L}{\partial \mu \partial \sigma} & -\dfrac{\partial^2 L}{\partial \mu^2} \end{pmatrix}^{-1}\right|_{\mu=\tilde{\mu},\sigma=\tilde{\sigma}}, \quad (30)$$

Элементы матрицы (30) примут вид:

$$\frac{\partial^2 L}{\partial \mu^2} = -2\sum_{j=1}^{n} \frac{\sigma^{\mu/\sigma} x_j^{1/\sigma}}{\sigma^2 \left(x_j^{1/\sigma} + e^{\mu/\sigma}\right)^2}, \quad (31)$$

$$\mu = \tilde{\mu},\ \sigma = \tilde{\sigma};$$

$$\frac{\partial^2 L}{\partial \sigma^2} = \frac{2n\mu}{\sigma^3} + \frac{2}{\sigma^3}\sum_{j=1}^{n} \ln x_j + \frac{n}{\sigma^2} - 2H; \quad (32)$$

$$\mu = \tilde{\mu},\ \sigma = \tilde{\sigma};$$

$$H = \frac{A + BC + 2\mu\sigma e^{2\mu/\sigma}}{Q};\ A = \sum_{j=1}^{n} \sigma\, x_j^{2/\sigma} \ln x_j; \quad (33)$$

$$B = e^{\mu/\sigma} \sum x_j^{1/\sigma}, \quad \mu = \tilde{\mu},\ \sigma = \tilde{\sigma};$$

$$C =$$

$$= \left\{ \sum_{j=1}^{n} \left[\left(\ln x_j^2 \right) + 2(\sigma - \mu) \ln x_j + \mu(\mu + 2\sigma) \right] \right\}, \quad (34)$$

$$\mu = \tilde{\mu},\ \sigma = \tilde{\sigma};$$

$$Q = \sigma^4 \sum_{j=1}^{n} \left(x_j^{1/\sigma} + e^{\mu/\sigma} \right)^2,\ \mu = \tilde{\mu},\ \sigma = \tilde{\sigma}; \quad (35)$$

$$\frac{\partial^2 L}{\partial \mu \partial \sigma} = (-1) \times$$

$$\times \frac{n\sigma \sum_{j=1}^{n} x_i^{2/\sigma} + T + \sigma \cdot e^{2\mu/\sigma}(n-2)}{\sigma^3 \left(\sum_{j=1}^{n} x_j^{1/\sigma} + e^{\mu/\sigma} \right)^2}, \quad (36)$$

$$\mu = \tilde{\mu}, \quad \sigma = \tilde{\sigma};$$

$$T = 2e^{\mu/\sigma} \sum_{j=1}^{n} x_j^{1/\sigma} \cdot \left(\sum_{j=1}^{n} \ln x_j + n\sigma - \mu - \sigma \right), \quad (37)$$

$$\mu = \tilde{\mu}, \quad \sigma = \tilde{\sigma}.$$

Тогда верхнюю и нижнюю границу доверительных интервалов для полученных оценок логлогистического распределения можно определить по условию:

$$\mu = \tilde{\mu} \pm K_\alpha \sqrt{\mathrm{Var}(\tilde{\mu})},\ \ \ln \sigma = \ln \tilde{\sigma} \pm \frac{K_\alpha \sqrt{\mathrm{Var}(\tilde{\sigma})}}{\tilde{\sigma}}. \quad (38)$$

Значения коэффициенты K_α для различных уровней доверительной вероятности α приведены в табл. 6.

Таблица 6.

Коэффициенты K_α

для уровня доверительной вероятности α

α	0,95	0,975	0,99	0,995	0,999
К α	1,64	1,96	2,32	2,57	3,01

Выводы

1. Поставлена задача проверки наличия резко выделяющихся данных в выборке из трёх одинаково распределённых случайных величин.
2. В работе рассмотрены следующие распределения: Вейбулла, гамма, нормальное распределение существенно-положительных величин (полунормальное распределение), обратное распределение Гаусса (распределение Вальда), Лапласа, логистическое распределение, Максвелла, нормальное. Все рассмотренные распределения моделировали при одинаковом среднем значении и среднеквадратическом отклонении.
3. Экспериментально, по результатам статистического моделирования, установлено, что независимо от исходных распределений распределение отношения «размах/медиана» для выборок из трёх наблюдений распределено по логлогистическому закону.
4. Приведены способы получения оценок параметров логлогистического распределения по методу моментов и методу максимума правдоподобия, получены выражения для определения величины доверительных интервалов полученных оценок.
5. Определена нижняя граница правосторонней критической области для принятия решения о согласованности результатов вычислений при использовании критерия «размах выборки/медиана выборки».

Список литературы

1. *Скрамтаев Б.Г. Строительные материалы [Текст] / Б.Г. Скрамтаев, Н.А. Герасимов, Г.Г. Мудров. – М.: Стройиздат Наркомтяжстроя, 1940. – 560 с.*
2. *Строительная индустрия. Справочное руководство по гражданскому и промышленному строительст-*

ву в семнадцати томах [Текст] / Гл. ред. В.И. Вельман. // Строительные материалы. Часть первая. / Под ред. Н.А. Попова. – Москва – Ленинград. – 1934. – 534 с.

3. Налимов В.В. Применение математической статистики при анализе вещества [Текст] / В.В. Налимов. – М. : ФИЗМАТГИЗ, 1960.- 430 с.

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

5. Расин Дж. Непараметрическая эконометрика: вводный курс [Текст] / Дж. Расин // Квантиль. – 2008. – № 4. – С. 7-76.

6. Садыхов Г.С. Модели и методы оценки остаточного ресурса изделий радиоэлектроники [Текст] / Г.С. Садыхов, В.П. Савченко, Н.И. Сидняев. – М.: МГТУ им. Н.Э. Баумана,2015. – 382 с.

7. Мироновский Л.А. Алгоритмы оценивания результатов трёх измерений [Текст] / Л.А. Мироновский. – СПб.: Профессионал, 2010. – 192 с.

8. Либлейн Ю. Ближайшие друг к другу два из трёх измерений [Текст] / Ю. Либлейн // Введение в теорию порядковых статистик. Ред. А.Я. Боярский. – М.: Статистика, 1970. – С. 122-127.

9. Диксон У. Отбраковка сомнительных наблюдений [Текст] / У.Диксон // Введение в теорию порядковых статистик. Ред. А.Я. Боярский / М. : Статистика, 1970. – С. 274-307.

10. Мироновский Л.А. Оценивание результатов измерений по малым выборкам [Текст] / Л.А. Мироновский, В.А. Слаев // Информационно-управляющие системы. – 2011. – № 3. – С. 69-78.

11. Джини К. Средние величины [Текст] / К. Джини. – М.: СТАТИСТИКА, 1970. – 443 с.

12. Колмогоров А.Н. Об определении среднего [Текст] / А.Н. Колмогоров // Избранные труды. Математика и механика. – М. : Наука, 1985. – С. 136 – 138.

13. Калинин С.И. Средние величины степенного типа. Неравенства коши и Ки Фана [Текст] / С.И. Калинин. – Киров : ВГГУ, 2002. – 361 с.

14. Дубницкий В.Ю. Решение в явном виде обратной задачи моделирования непрерывной случайной величины [Текст] / В.Ю. Дубницкий, И.Г. Скорикова // Системи обробки інформації. – Х.: ХУПС, 2015. – Вип. 1 (126). – С. 106- 110.

15. Хьюбер Дж. Робастность в статистике [Текст] / Дж. Хьюбер. – М.: Мир, 1984. – 304 с.

16. Джонсон Н.Л. Одномерные непрерывные распределения [Текст] / Н.Л. Джонсон, С.Коц, Н. Балакришнан. – М.: БИНОМ. Лаборатория знаний, 2010-2012. – 600 с.

17. Life Data Analysis Referenses. The Loglogistic Distribution. P.246- 250 [Электронный ресурс] – Режим доступа: http://www.synthesisplatform.net/references/Life_Data_Analysis_Reference.pdf.

18. Кузнецов Д.С. Специальные функции [Текст] / Д. С. Кузнецов. – М. : Высшая школа, 1962. – 246 с.

19. Поршнев С.В. Численные методы на базе Mathcad [Текст] / С.В. Поршнев, И.В. Беленков. – Санкт-Петербург.: БХВ-Петербург, 2005. – 464 с.

20. Кендалл М. Статистические выводы и связи [Текст] / М. Кендалл, А. Стьюарт. – М.: НАУКА, 1973. – 900 с.

Поступила в редколлегию 27.05.2016

Рецензент: д-р физ.-мат. наук, проф. М.В. Новожилова, Харьковский национальный университет строительства и архитектуры, Харьков.

ВИКОРИСТАННЯ ФУНКЦІЙ НЕАНАЛІТИЧНИХ СЕРЕДНІХ ВЕЛИЧИН ДЛЯ ОЦІНКИ УЗГОДЖЕНОСТІ ОБЧИСЛЮВАЛЬНИХ ПРОЦЕСІВ

В.Ю. Дубницький, О.І. Ходирєв

Сформульовано задачу перевірки наявності даних, що різко виділяються у вибірці з трьох однаково розподілених випадкових величин. У роботі розглянуто наступні розподіли: Вейбула, гамма, нормальний розподіл суттєво-додатних величин (напівнормальний розподіл), зворотний розподіл Гауса (розподіл Вальда), Лапласа, логістичний розподіл, Максвела, нормальний. За результатами статистичного моделювання встановлено, що незалежно від початкових розподілів розподіл відношення «розмах /медіана» для вибірок з трьох спостережень розподілений за логлогістичним законом. Наведено способи визначення оцінок параметрів логлогістичного розподілу за методом моментів і методом максимуму правдоподібності, отримано вирази для визначення величин довірчих інтервалів отриманих оцінок. Визначено нижню межу правосторонньої критичної області для прийняття рішення про узгодженість результатів обчислень при використанні критерію «розмах вибірки / медіана вибірки».

Ключові слова: *розмах вибірки, медіана, розподіл Вейбулла, гамма-розподіл, нормальний розподіл суттєво-позитивних величин (напівнормальний розподіл), зворотне розподіл Гауса (розподіл Вальда), розподіл Лапласа, логістичне розподіл, розподіл Максвелла, нормальний розподіл, логарифмічно логістичний розподіл, метод моментів, метод максимуму правдоподібності, оцінка параметрів логарифмічно логістичного розподілу.*

APPLICATION OF NON-ANALYTYICAL MEAN VALUE FUNCTIONS FOR CONSISTENCY EVALUATION OF COMPUTATION PROCESSES

V.Yu. Dubnitskiy, A.I. Khodyrev

A problem was formulated of checking availability of data that sharply stand out in a sample among three identically distributed random values. The work dwells on such distributions: Weibull, gamma, normal distribution of substantially positive values (half-normal distribution), inverse Hauss distribution (Wald distribution), Laplace distribution, logistic distribution, Maxwell distribution, normal distribution. The results of statistical modeling showed that irrespective of initial distributions, the distribution of span/median ratio for samples from three observations always followed loglogistic law. Several techniques specified for evaluation of loglogistic parameters using method of moments and method of maximum likelihood, expressions found for determination of confidence interval values for obtained evaluations. Bottom limit was obtained of right-hand critical region for making decision on consistency of calculation results, using sample span/median criterion.

Keywords: *sweep sample, the median, the Weibull distribution, the gamma distribution, the normal distribution significantly positive values (seminormal distribution), the inverse Gaussian distribution (distribution-division Wald), Laplace distribution, logistic distribution, Maxwell distribution, the normal distribution, log-logistic distribution, method of moments, maximum likelihood method, parameter estimation log-logistic distribution.*