УДК 621.317.08

ОЦЕНИВАНИЕ ГРАНИЦ СЛУЧАЙНОЙ ПОГРЕШНОСТИ РЕЗУЛЬТАТА ИЗМЕРЕНИЯ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

ЗАХАРОВ И.П., ШТЕФАН Н.В., СЕЗОНОВА И.К.

Приводится метод, позволяющий определить границы случайной погрешности для малых выборок, имеющих распределение, отличное от нормального.

Оценивание границ погрешности является важнейшим завершающим этапом процесса измерения. В большинстве литературных источников по этому вопросу рассматривается упрощенная модель, когда закон распределения считается нормальным. На практике распределение измеряемой величины может сильно отличаться от нормального. Это особенно сильно сказывается при обработке выборок ограниченного объема.

При проведении прецизионных и контрольно-поверочных измерений осуществляется апостериорная оценка их погрешности. Наиболее часто за результат измерения принимают среднее арифметическое [1]

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \, , \tag{1}$$

где x_i — результат отдельного наблюдения. Преимущество этой оценки в том, что она единственная, которую можно выразить аналитически и подставлять ее в таком виде в другие соотношения, анализировать их и т.д., а не в том, что она является наиболее эффективной. В этом случае границы случайной погрешности $\bar{x}$ определяются по формуле

$$\varepsilon_{\bar{x}} = \pm t_p \hat{\sigma}_{\bar{x}} \, , \tag{2}$$

здесь $\hat{\sigma}_{\bar{x}}$ — оценка среднего квадратического отклонения результата измерения, выраженного в виде среднего арифметмического; n — число наблюдений; t_p — доверительный коэффициент, зависящий от доверительной вероятности и вида закона распределения.

Если число наблюдений $n > 20...30$, то согласно центральной предельной теореме теории вероятностей $\bar{x}$ распределено по нормальному закону, следовательно, t_p берется для нормального закона распределения.

На практике обычно обрабатываются выборки малого объема. В этом случае закон распределения t_p в выражении (2) не является нормальным, а определяется исходным законом распределения результатов наблюдений и их числом.

Если исходный закон распределения результатов наблюдений нормальный, то для определения границ случайной погрешности $\bar{x}$ рекомендуется пользоваться коэффициентами t_s, которые распределены по закону Стьюдента [2]. Его дифференциальная функция распределения имеет вид

$$s_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} . \tag{3}$$

Это распределение описывает плотность вероятности значений среднего арифметического по выборке из m случайных значений из нормально распределенной генеральной совокупности и получено аналитически как распределение величины

$$T = \sqrt{n}\,\frac{\bar{x} - M_x}{\hat{\sigma}_x} \, , \tag{4}$$

где M_x — математическое ожидание генеральной совокупности.

Очевидно, что для распределений, отличных от нормального, использование распределения Стьюдента некорректно, поскольку величина T не будет распределена по закону Стьюдента. Попытки получить аналитическое выражение для распределений, аналогичных распределению Стьюдента, не увенчались успехом. Поэтому в настоящей статье предлагается численный метод определения доверительных коэффициентов, аналогичных коэффициенту Стьюдента для законов распределения результатов наблюдений, отличных от нормального. Суть предложенного метода заключается в следующем:

1. Моделируется произвольное распределение случайных величин с заранее известными параметрами, представляющее собой генеральную совокупность, которая состоит из m случайных величин ($m \geq 30000$ достаточно велико).

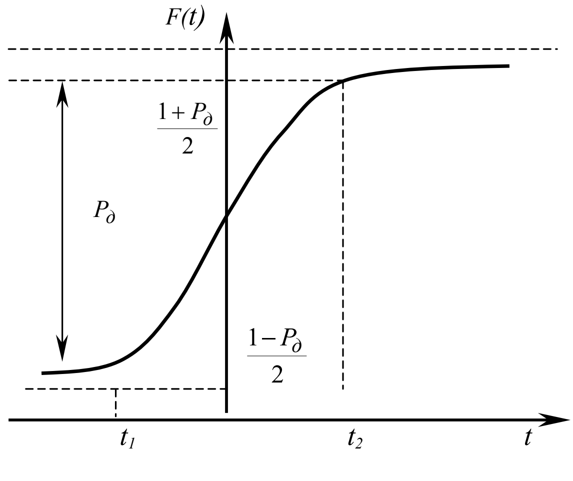
2. Данная генеральная совокупность разделяется на k независимых выборок объемом n таким образом, что $n = \frac{m}{k} = 2,3,4,...$.

3. Для каждой выборки определяется величина $T(n)$ по формуле (4).

4. Очевидно, что совокупность величин $T(n)$ для одинаковых значений n представляет собой их распределения.

5. Для каждого полученного распределения $T(n)$ строим эмпирическую интегральную функцию распределения $F(T)$.

6. Для заданных доверительных вероятностей P_∂ определяем значения коэффициентов t_1, t_2 (рисунок).



$$t_1 = F^{-1}\left(\frac{1-P_\partial}{2}\right), \qquad (5)$$

$$t_2 = F^{-1}\left(\frac{1+P_\partial}{2}\right). \qquad (6)$$

Очевидно, что для симметричных законов распределения $t_1 = -t_2$, поэтому коэффициент $t(n, P_\partial)$ определяется как

$$t(n, P_\partial) = \frac{t_2 - t_1}{2}. \qquad (7)$$

Для реализации предложенной методики используем встроенный генератор случайных чисел Microsoft Mathcard 7.0, с помощью которого можно получить равномерный и нормальный законы распределения случайной величины x_i. Остальные распределения были получены как результат трансформации равномерного закона при нелинейном преобразовании, функция которого задавалась аналитически:

$$x_i = F^{-1}(y_i), \qquad (8)$$

где x_i – случайная величина, имеющая заданное распределение; y_i – случайная величина, распределенная равномерно на интервале (0,1); $F^{-1}(y)$ – функция, обратная к интегральной функции заданного распределения.

Проверка методики осуществлялась на нормальном законе распределения. Результаты расчета коэффициентов Стьюдента t_s для различных доверительных вероятностей приведены в табл. 1-3 (в табл. 1 – для доверительной вероятности 0,9, в табл. 2 – для 0,95, в табл. 3 – для 0,99). В табл. 1-3 t_s - табличное значение коэффициента Стьюдента, $t_{s\ расч.}$ – его рассчитанное значение.

Из табл. 1-3 видно, что погрешность метода составляет не более 4,5%.

Таблица 1

Число наблюдений n	t_s	$t_{s\ расч.}$	Погрешность, %
3	2,90	2,92	-0,7
5	2,14	2,13	0,6
7	1,87	1,94	-3,5
9	1,84	1,86	-1,2
11	1,79	1,81	-1,1
13	1,78	1,78	0,1
15	1,78	1,76	1,1
17	1,76	1,75	0,6
19	1,75	1,73	1,2
21	1,74	1,72	1,2
23	1,74	1,72	1,2
25	1,73	1,71	1,2
27	1,73	1,71	1,2

Таблица 2

Число наблюдений n	t_s	$t_{s\ расч.}$	Погрешность, %
3	4,2	4,3	-3,2
5	2,85	2,78	2,4
7	2,36	2,45	-3,6
9	2,28	2,31	-1,3
11	2,16	2,23	-3,1
13	2,20	2,18	0,8
15	2,19	2,15	1,9
17	2,16	2,12	1,9
19	2,11	2,1	0,5
21	2,07	2,08	-0,5
23	2,04	2,07	-1,4
25	2,03	2,06	-1,5
27	2,00	2,06	-2,9

Таблица 3

Число наблюдений n	t_s	$t_{s\ расч.}$	Погрешность, %
3	9,74	9,93	-1,89
5	4,80	4,6	4,30
7	3,59	3,71	-3,21
9	3,33	3,36	-0,86
11	3,24	3,16	2,65
13	2,94	3,05	-3,56
15	2,94	2,98	-1,34
17	2,93	2,92	0,47
19	2,84	2,88	-1,37
21	2,82	2,85	-1,05
23	2,81	2,82	-0,35
25	2,80	2,8	0,00
27	2,78	2,78	0,00

Результаты вычисления коэффициентов Стьюдента для других симметричных распределений представлены в табл. 4-7 (в табл. 4 – для распределения арксинус, в табл. 5 – для равномерного, в табл. 6 – для треугольного, в табл. 7 – для Лапласа).

Из табл. 4-7 видно, что при $n \geq 12$ все законы стремятся к закону распределения Стьюдента.

Для несимметричных распределений, очевидно, доверительный интервал следует определять симметричным по вероятности:

$$\left(\bar{x} - t_{p1}\hat{\sigma}_{\bar{x}}; \bar{x} + t_{p2}\hat{\sigma}_{\bar{x}}\right). \qquad (9)$$

Таким образом, предложенный численный метод позволяет рассчитывать доверительные коэффициенты практически для любых распределений при малом числе наблюдений, если закон распределения известен.

В литературе [3] при малых объемах выборки в качестве характеристики рассеяния рекомендуется

Таблица 4

Число наблюдений n	$P_\partial = 0{,}9$	$P_\partial = 0{,}95$	$P_\partial = 0{,}99$
3	7,50	9,73	26,87
5	2,25	3,46	8,27
7	1,98	2,61	5,22
9	1,91	2,43	3,87
11	1,89	2,30	3,32
13	1,82	2,37	3,26
15	1,79	2,27	3,21
17	1,79	2,15	3,12
19	1,77	2,12	3,11
21	1,74	2,08	2,98
23	1,76	2,07	2,94
25	1,75	2,06	2,94
27	1,75	2,05	2,80

Таблица 5

Число наблюдений n	$P_\partial = 0{,}9$	$P_\partial = 0{,}95$	$P_\partial = 0{,}99$
3	3,82	5,52	14,38
5	2,21	3,12	6,07
7	1,93	2,45	4,30
9	1,90	2,39	3,61
11	1,87	2,25	3,33
13	1,76	2,25	3,26
15	1,77	2,20	3,14
17	1,78	2,14	2,92
19	1,76	2,13	2,94
21	1,76	2,07	2,94
23	1,71	2,02	2,82
25	1,72	2,00	2,91
27	1,71	2,06	2,78

Таблица 6

Число наблюдений n	$P_\partial = 0{,}9$	$P_\partial = 0{,}95$	$P_\partial = 0{,}99$
3	2,93	4,37	10,08
5	2,16	2,73	4,90
7	1,91	2,34	3,74
9	1,85	2,30	3,34
11	1,78	2,20	3,25
13	1,78	2,21	2,98
15	1,78	2,18	2,95
17	1,77	2,12	2,94
19	1,76	2,12	2,87
21	1,74	2,07	2,85
23	1,72	2,03	2,84
25	1,71	2,02	2,82
27	1,72	2,02	2,78

Таблица 7

Число наблюдений n	$P_\partial = 0{,}9$	$P_\partial = 0{,}95$	$P_\partial = 0{,}99$
3	2,45	3,45	3,45
5	2,01	2,51	2,51
7	1,86	2,30	2,30
9	1,83	2,20	2,20
11	1,77	2,14	2,14
13	1,78	2,13	2,13
15	1,77	2,13	2,13
17	1,76	2,10	2,10
19	1,75	2,09	2,09
21	1,75	2,06	2,06
23	1,74	2,04	2,04
25	1,74	2,00	2,00
27	1,70	1,98	1,98

применять среднее абсолютное отклонение (САО), оценка которого определяется по формуле

$$C\hat{A}O = \frac{\sum_{i=1}^{n} |x_i - x|}{n} . \qquad (10)$$

Такая оценка рассеяния по сравнению с оценкой среднеквадратичного отклонения обладает меньшей дисперсией при малом числе наблюдений n и является, естественно, более эффективной, кроме того, она менее чувствительна к наличию грубых погрешностей и промахов. Это дает возможность для оценки рассеяния в этом случае применять оценку САО, тогда граница доверительного интервала для среднего арифметического будет иметь вид:

$$\bar{x} \pm k_P \cdot C\hat{A}O . \qquad (11)$$

Очевидно, что коэффициент k_P зависит от числа наблюдений и доверительной вероятности P_∂. Распределение его можно найти как распределение величины

$$K = \frac{\bar{x} - M_x}{C\hat{A}O} . \qquad (12)$$

Воспользовавшись предложенным выше методом, получим коэффициенты k_p для различных распределений и доверительных вероятностей. Результаты расчета приведены в табл. 8-12 (в табл. 8 – для распределения арксинус, в табл. 9 – для равномерного, в табл. 10 – для треугольного, в табл. 11 – для нормального, в табл. 12 – для Лапласа).

Таким образом, предложенный численный метод позволяет рассчитать доверительные коэффициенты для различных распределений, которые могут быть использованы при расчете доверительных границ случайной погрешности с ограниченным объемом выборки.

Таблица 8

Число наблюдений n	$P_{\partial} = 0{,}9$	$P_{\partial} = 0{,}95$	$P_{\partial} = 0{,}99$
3	3,48	6,63	20,66
5	1,32	2,06	4,87
7	0,98	1,34	2,60
9	0,80	1,07	1,73
11	0,70	0,86	1,39
13	0,61	0,78	1,09
15	0,56	0,72	1,06
17	0,52	0,65	0,99
19	0,49	0,60	0,94
21	0,45	0,56	0,80
23	0,43	0,53	0,76
25	0,41	0,49	0,74
27	0,41	0,49	0,64

Таблица 9

Число наблюдений n	$P_{\partial} = 0{,}9$	$P_{\partial} = 0{,}95$	$P_{\partial} = 0{,}99$
3	2,64	4,42	11,39
5	1,33	1,84	3,63
7	0,98	1,29	2,13
9	0,80	1,02	1,54
11	0,71	0,91	1,34
13	0,61	0,78	1,14
15	0,56	0,68	1,00
17	0,53	0,67	0,91
19	0,49	0,60	0,89
21	0,46	0,55	0,81
23	0,42	0,54	0,72
25	0,40	0,46	0,68
27	0,38	0,46	0,66

Таблица 10

Число наблюдений n	$P_{\partial} = 0{,}9$	$P_{\partial} = 0{,}95$	$P_{\partial} = 0{,}99$
3	2,08	3,34	7,87
5	1,26	1,62	2,65
7	0,94	1,15	1,83
9	0,78	0,98	1,33
11	0,68	0,80	1,29
13	0,62	0,74	1,01
15	0,56	0,68	0,88
17	0,52	0,63	0,81
19	0,48	0,59	0,83
21	0,46	0,55	0,72
23	0,45	0,53	0,72
25	0,43	0,49	0,72
27	0,40	0,49	0,65

Таблица 11

Число наблюдений n	$P_{\partial} = 0{,}9$	$P_{\partial} = 0{,}95$	$P_{\partial} = 0{,}99$
3	2,28	3,50	7,51
5	1,27	1,67	2,89
7	0,93	1,19	1,79
9	0,80	1,00	1,51
11	0,69	0,85	1,26
13	0,64	0,80	1,05
15	0,59	0,71	0,91
17	0,53	0,64	0,85
19	0,52	0,63	0,83
21	0,48	0,57	0,72
23	0,46	0,54	0,72
25	0,44	0,53	0,70
27	0,42	0,50	0,63

Таблица 12

Число наблюдений n	$P_{\partial} = 0{,}9$	$P_{\partial} = 0{,}95$	$P_{\partial} = 0{,}99$
3	1,55	2,65	5,94
5	1,18	1,50	2,40
7	0,94	1,15	1,73
9	0,83	0,99	1,36
11	0,73	0,87	1,16
13	0,66	0,78	1,00
15	0,63	0,75	0,97
17	0,58	0,68	0,93
19	0,56	0,64	0,87
21	0,50	0,62	0,79
23	0,50	0,58	0,74
25	0,48	0,57	0,72
27	0,47	0,56	0,67

Литература: 1. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений. Л.: Энергоатомиздат. Ленингр. отд-ние, 1991. 304 с. 2.*Крамер Г.* Математические методы статистики. М.: Мир, 1976. 648 с. 3. *Закс Л.* Статистическое оценивание. М.: Статистика, 1976. 598 с.

Поступила в редколлегию 02.08.2001

Рецензент: д-р физ.-мат. наук, проф. Яковлев С.В.

Захаров Игорь Петрович, канд. техн. наук, доцент кафедры МИТ ХНУРЭ. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 40-93-31, 27-60-17.

Штефан Наталья Владимировна, аспирант кафедры МИТ ХНУРЭ. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 40-93-31, 27-60-17.

Сезонова Ирина Константиновна, канд. техн. наук, доцент кафедры информатики НУВД. Научные интересы: разработка информационно-поисковых систем, моделирование интеллектуальных систем. Увлечение: музыка. Адрес: Украина, 61080, Харьков, пр. 50 лет СССР, 27, тел. 50-30-80.