

ОПРЕДЕЛЕНИЕ СРЕДНЕКВАДРАТИЧНОЙ ПОГРЕШНОСТИ И НЕОБХОДИМОГО ЧИСЛА ЧЛЕНОВ РЯДА ИЗМЕРЕНИЙ

И. К. Овчинников, В. Д. Кукуш

Х а р ь к о в

Физические измерения характеризуются тем, что закон распределения погрешностей измерений во многих случаях отличается от нормального [1]. В результате этого при обработке рядов измерений нельзя во всех случаях пользоваться распределениями и оценками, связанными с нормальным законом. Например, при обработке ряда с ограниченным числом членов нельзя пользоваться функцией распределения Стьюдента, когда закон исследуемого распределения неизвестен, так как функция Стьюдента справедлива для выборок с ограниченным числом членов из нормального закона [2].

Широкое использование в науке и технике методов физических измерений повышает требования к результатам, получаемым при обработке измерений. Нельзя считать допустимыми случаи, когда частоты могут принимать значения значительно меньшие, чем ожидаемая вероятность выбранного доверительного интервала, что имеет место при проведении недостаточного числа измерений.

В математической статистике и теории вероятностей [3—5] рекомендуется использовать ряды измерений с минимальным числом членов — порядка 20—30. А в специальной литературе, например по радиотехническим измерениям, минимальное число членов ряда измерений или не оговаривается, или принимается равным 10 [6] и 17 [7]. Необходимо отметить, что обоснований выбора числа членов ряда измерений не приводится.

Ниже рассмотрим обработку ряда измерений с ограниченным числом членов, когда закон исследуемого распределения остается неизвестным.

1. Величина дисперсии при ограниченном числе измерений

Для случайной величины x известны строгие определения [5, 9] математического ожидания $M(x)$ и дисперсии $D(x)$:

$$M(x) = \sum_{i=1}^n x_i p_i \quad (1)$$

$$D(x) = M\{[x - M(x)]^2\}. \quad (2)$$

Так как возможные значения результатов измерений образуют счетное множество, то при определении математического ожидания вместо конечной суммы имеем бесконечный ряд $n \rightarrow \infty$, а после подстановки $p_i = \frac{1}{n} -$

вероятность появления каждого отдельного результата измерения x_i в (1) и (2):

$$M(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

$$D(x) = \frac{1}{n} \sum_{i=1}^n [x_i - M(x)]^2. \quad (4)$$

Мерой разброса случайной величины является среднеквадратичная погрешность $\sigma(x)$, которая связана с дисперсией выражением

$$\sigma^2(x) = D(x). \quad (5)$$

При ограниченном числе измерений n выражение (3) определяет среднее арифметическое значение:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (6)$$

В отличие от постоянной величины — математического ожидания — среднее арифметическое является случайной величиной и имеет закон распределения, близкий к нормальному [3] с дисперсией [8]:

$$D(\bar{x}) = \frac{1}{n} D(x). \quad (7)$$

Величину уклонения среднего арифметического от математического ожидания можно записать в виде

$$\bar{x} - M(x) = t\sigma(\bar{x}), \quad (8)$$

где t — величина уклонения, выраженная в долях среднеквадратичной погрешности $\sigma(x)$, см. (5), (7).

Известно, что замена математического ожидания средним арифметическим приводит к определенной систематической погрешности при определении оценки дисперсии, так называемой смещенной оценке. Величина этой погрешности определяется уклонением x от $M(x)$. И чем больше это уклонение, тем больше величина погрешности.

После подстановки в уравнение (4) величины математического ожидания из формулы (8) после преобразования получим выражение для определения величины дисперсии

$$D_P(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{t^2}{n} D(x). \quad (9)$$

Второй член правой части уравнения (9) определяет поправку в зависимости от выбранной величины уклонения. Так как среднее арифметическое $\bar{x}$ с вероятностью P находится в интервале $[M(x) + t\sigma(\bar{x}); M(x) - t\sigma(\bar{x})]$, то нижний индекс D_P указывает на такую же вероятность того, что выполняется условие

$$D_P(x) \geq D(x). \quad (10)$$

Из выражений (9) и (10) видно, что выбор величины уклонения, равной среднеквадратичной погрешности, приводит к обычно используемой

формуле для определения оценки дисперсии, но при вероятности выполнения неравенства (10) с величиной, равной 0,68.

В тех случаях, когда среднее арифметическое $\bar{x}$ имеет отклонение от математического ожидания $M(x)$, соответствует $t\sigma(\bar{x})$, систематическая погрешность оценки дисперсии равна нулю; можно считать, что оценка дисперсии $D_P(x)$ соответствует дисперсии $D(x)$ исследуемого распределения. Это позволяет заменить $D(x)$ в правой части уравнения (9) на равную или большую величину $D_P(x)$ при сохранении вероятности выполнения условия (10). После такой замены и соответствующего преобразования находим выражение для оценки среднеквадратичной погрешности с определенной вероятностью того, что эта оценка не будет принимать значений меньших, чем среднеквадратичная погрешность исследуемого распределения

$$\sigma_P(x) = \sqrt{\frac{1}{n-t^2} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (11)$$

Величина математического ожидания в соответствии с уравнениями (5), (7), (8), (11) равна

$$M_P(x) = \bar{x} \pm t \frac{\sigma_P(x)}{\sqrt{n}}. \quad (12)$$

Уравнения (11) и (12) с вероятностью P характеризуют случайную величину x .

2. Оценки дисперсии и математического ожидания

Оценки дисперсии и математического ожидания, вычисленные в соответствии с уравнениями (11) и (12), являются состоятельными и несмещенными. Действительно, в том случае, когда число измерений непрерывно увеличивается и стремится к бесконечности, выполняются следующие условия:

$$n - t^2 \rightarrow n; \quad (13)$$

$$\bar{x} \rightarrow M(x); \quad (14)$$

$$\sqrt{\frac{1}{n} D_P(x)} \rightarrow 0. \quad (15)$$

При условиях (13) и (15) оценка дисперсии $D_P(x)$ приближается к постоянной величине $D(x)$, а оценка математического ожидания $M_P(x)$ — к величине $M(x)$, что и подтверждает состоятельность оценок.

Математическое ожидание величины уклонения среднего арифметического от математического ожидания и квадрата этой величины будут соответственно [8]

$$M[\bar{x} - M(x)] = 0; \quad (16)$$

$$M[\bar{x} - M(x)]^2 = D(x). \quad (17)$$

Исходя из независимости величин t и t^2 , из выражений (8), (16) и (17) находим

$$M(t) = 0; \quad (18)$$

$$M(t^2) = 1. \quad (19)$$

Математическое ожидание оценок $M_P(x)$ и $D_P(x)$ с учетом равенств (18) и (19) показывает, что эти оценки являются несмещенными [3].

3. Вероятность определения оценки среднеквадратичной погрешности и доверительного интервала

Вероятность определения оценки среднеквадратичной погрешности по уравнению (11), т. е. вероятность того, что выполняется условие (10), находится по значению величины t . Так как величина t определяет отклонение $\bar{x} - M(x)$, которое имеет распределение, близкое к нормальному, то каждой величине t соответствует определенное значение вероятности. Например, для $t = 1; 2; 3$ вероятность составляет соответственно 0,68; 0,95; 0,997. Таким образом, обычно используемая расчетная формула для определения оценки среднеквадратичной погрешности

$$\sigma_1(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (20)$$

имеет вероятность 0,68, т. е. треть расчетных значений $\sigma_1(x)$ может иметь значения меньше, чем среднеквадратичная погрешность исследуемого распределения. Это подтверждается и результатами обработки экспериментальных данных.

В табл. 1 приведены результаты определения оценок среднеквадратичных погрешностей 20 рядов измерений периода колебаний напряжения городской сети. Каждый ряд состоит из 15 членов. Сравнение производится с величиной среднеквадратичной погрешности σ , определенной по данным всех рядов измерений (300 членов). Период измерялся цифровым частотомером.

Таблица 1

Оценка среднеквадратичной погрешности	$\sigma_{0,68}$	$\sigma_{0,95}$	$\sigma_{0,997}$
Число выборок $\sigma_p \geq \sigma$	10	18	20
Частость определения σ_p	0,5	0,9	1,0

Как видно из данных табл. 1, частость, определяемая по результатам 20 рядов измерений, хорошо согласуется с ожидаемой вероятностью.

Следует отметить, что отдельные значения оценки среднеквадратичной погрешности, определяемые с вероятностью 0,68, были меньше σ и имели наименьшие значения 0,7 σ . Это показывает, что выбор доверительного интервала 3 $\sigma_{0,68}$ в отдельных случаях может соответствовать интервалу 2 σ .

Найдем вероятность доверительного интервала, определяемого уравнением (12).

В общем случае величину интервала можно представить в виде

$$t_1 \sigma_{t_1}(\bar{x}) = \pm t_1 \sqrt{\frac{1}{n(n-t_2^2)} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (21)$$

где t_1 — величина отклонения, определяющая доверительный интервал;
 t_2 — величина отклонения, использованная для определения оценки среднеквадратичной погрешности.

Необходимость представления доверительного интервала в виде (21) вызвана тем, что оценка дисперсии является случайной величиной с определенной вероятностью ее определения.

Вероятность P , соответствующую интервалу (21), можно найти по теореме умножения вероятностей [3]:

$$P = P_t P_s, \quad (22)$$

где P_t — вероятность интервала, соответствующая отклонению $\pm t_1$;

P_s — условная вероятность определения оценки среднеквадратичной погрешности, вычисленная при условии, что величина отклонения находится в пределах $\pm t_1$.

Величина вероятности P_t определяется значением t при нормальном законе распределения величины $\bar{x}$. При $t = 1, 2, 3$ имеем $P_t = 0,68, 0,95, 0,997$.

Условная вероятность $P_s = 1$, если $t_2 \geq t_1$, так как в этом случае при определении оценки среднеквадратичной погрешности учитывается отклонение в пределах $\pm t_1$. С другой стороны, при $t_2 < t_1$, условная вероятность $P_s < 1$, так как при определении оценки среднеквадратичной погрешности не учитываются отклонения, превышающие величину t_2 . Такие значения условной вероятности приводят к тому, например, что интервалы $2\sigma_{0,68}$ или $3\sigma_{0,68}$ не обеспечивают ожидаемой вероятности 0,95 или 0,997. Наиболее оптимальным следует считать условие

$$t_1 = t_2, \quad (23)$$

когда доверительный интервал и оценка среднеквадратичной погрешности определяются при одинаковых величинах отклонения.

Сравнение результатов 15 рядов измерений по 40 членов с результатами, полученными при обработке соответствующих трех рядов измерений по 200 членов, показывает, что частота доверительного интервала $3\sigma_{0,68}$ составляет 0,67 при ожидаемой вероятности 0,997, так как пять рядов измерений дают доверительные интервалы меньшие, чем доверительные интервалы, определенные по результатам 200 измерений. Ряды измерений получены в соответствии с последовательностью данных, приведенных в работе [5].

4. Выбор числа членов ряда измерений

Уравнение (11) позволяет установить зависимость между вероятностью P определения оценки среднеквадратичной погрешности $\sigma_P(x)$ и числом членов n ряда измерений.

Рассмотрим величину

$$a = \frac{\sigma_P(x) - \sigma_0(x)}{\sigma_0(x)}, \quad (24)$$

где

$$\sigma_0(x) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (25)$$

Из уравнения (24) после подстановки выражения (25) получим

$$a = \frac{\sqrt{n} - \sqrt{n - t^2}}{\sqrt{n - t^2}}. \quad (26)$$

Необходимое число измерений можно выбрать при различных значениях величины a .

Нижний предел для числа измерений выбираем при $a = 2$, откуда

$$n_n \approx 1,3t^2. \quad (27)$$

Среднее число измерений выбираем при $a = 0,5$, откуда

$$n_{cp} \approx 2t^2. \quad (28)$$

Верхний предел числа измерений выбираем при $a = 0,1$, откуда

$$n_v \approx 6t^2. \quad (29)$$

В табл. 2 приводится число необходимых измерений, вычисленных в соответствии с выражениями (27) — (29) с некоторыми приближениями при различной величине надежности.

Таблица 2

P	t^2	n_n	n_{cp}	n_v
0,68	1	2	2	6
0,95	4	6	8	24
0,997	9	12	18	54

Из табл. 2 видно, что в зависимости от величины вероятности необходимо выбирать определенное число измерений. При тех измерениях, где не требуется высокой точности определения оценки дисперсии, можно пользоваться небольшим числом измерений n_n или n_{cp} . При измерениях с большой точностью необходимо использовать ряды с числом измерений n_v .

ВЫВОДЫ

1. Вычисление величины оценки среднеквадратичной погрешности для характеристики разброса значений случайной величины x должно производиться с определенной вероятностью того, что полученная величина не будет принимать значений меньших, чем среднеквадратичная погрешность исследуемого распределения.

2. Расчетная формула (11) может быть использована для получения оценки дисперсии с определенной вероятностью. Эта оценка является состоятельной и несмещенной. Формула (11) подтверждает совпадение частоты при обработке экспериментальных данных с ожидаемой вероятностью.

3. Необходимые требования по точности измерений и величине вероятности определения оценки дисперсии позволяют найти необходимое число членов ряда измерений.

4. Полученные результаты справедливы при любом законе распределения исследуемой случайной величины.

ЛИТЕРАТУРА

1. Б. Е. Рабинович. Методика суммирования частных погрешностей в области радиотехнических измерений. Сб. «Вопросы радиоэлектроники», «Радиоизмерительная техника», серия VI, вып. 4, 1961.

2. И. В. Дунин-Барковский, Н. В. Смирнов. Теория вероятностей и математическая статистика в технике (общая часть). Гостехиздат, 1955.

3. Е. С. Вентцель. Теория вероятностей. Изд-во «Наука», 1964.

4. И. Г. Венецкий, Г. С. Кильдишев. Основы математической статистики, Госстатиздат, 1963.
 5. А. М. Длин. Математическая статистика в технике. Изд-во «Советская наука», 1958.
 6. Г. А. Ремез. Курс основных радиотехнических измерений. Изд-во «Связь», 1966.
 7. С. М. Верник, В. Г. Савенко. Измерения в проводной связи. Изд-во «Связь», 1966.
 8. Н. В. Смирнов, И. В. Дунин-Барковский. Курс теории вероятностей и математической статистики для технических приложений. Изд-во «Наука», 1965.
 9. Р. С. Гутер, Б. В. Овчинский. Элементы численного анализа и математической обработки результатов опыта. Физматгиз, 1962.
-