Российская академия наук
Институт прикладной астрономии

Сообщения ИПА РАН № 137

З. М. Малкин

О вычислении средневзвешенного значения

Санкт-Петербург
2001

УДК 519.2

З. М. Малкин. О вычислении средневзвешенного значения.

Ключевые слова: Математическая статистика, обработа наблюдений, среднее значение, средневзвешенное значение.

Хотя вычисление взвешенного среднего является широко используемой процедурой в научном анализе, эта задача не имеет до сих пор удовлетворительного однозначного решения. Основной проблемой при ее решении является оценка ошибки средневзвешенного значения, наиболее адекватной исходным данным. В литературе предложены разные подходы к решению этой проблемы, которые сравниваются в настоящей работе путем сопоставления формул и численных тестов. Предлагается комбинированная оценка среднеквадратической ошибки взвешенного среднего, которая, по мнению автора, лишена некоторых недостатков известных оценок. Для упрощения расчетов по обсуждаемым алгоритмам выведены простые аппроксимирующие выражения для некоторых статистических распределений.

Z. M. Malkin. On Computation of Weighted Mean.

Keywords: Statistics, data processing, mean value, weighted mean.

Though computation of weighted mean value of several measurements is a widely used procedure in scientific analysis, this task has still no simple satisfactory solution. The main problem here is a estimation of error of mean most adequate to data. Some approaches described in literature are compared analytically and tested using numerical simulation. Combined estimate of standard deviation of mean providing more stable estimate is proposed. To simplify realization of discussed algorithms simple approximations of some statistics distributions are advanced.

Сообщения Института прикладной астрономии РАН № 137. – Санкт-Петербург, 2001. – 15 с.

Содержание

1. Введение

Вычисление средневзвешенного значения (взвешенного среднего) некоторой величины из нескольких ее определений является распространенной процедурой научного анализа. Типичные задачи такого рода: вычисление среднего значения из нескольких определений, сделанных разными исследователями или по разным сериям наблюдений и т.п.

При этом исходная информация обычно состоит только из нескольких оценок значения определяемой величины x_i и оценок их среднеквадратических ошибок s_i, $i = 1, \dots, n$ (т.е. мы имеем дело с прямыми неравноточными измерениями). Как правило не имеется никакой информации о корреляционных связях между x_i. Задача состоит в нахождении наивероятнейшего среднего значения из измеренных величин $\bar{x}$ и его среднеквадратической ошибки σ.

Поскольку реальное распределение ошибок значений x_i неизвестно, строгого математического решения задачи не существует, за исключением разве что некоторых непараметрических оценок, которые, обычно, оказываются слишком грубыми для практического использования. Теория предлагает два основных метода вычисления средневзвешенного значения. Первый основан на методе максимального правдоподобия, второй — приводит к формуле Бесселя. Оба метода приводят к одинаковым оценкам $\bar{x}$, но к существенно различающимся оценкам σ. Поэтому на практике исследователь сталкивается с проблемой выбора оценки, наиболее адекватной исходным данным.

В литературе (см., например, [1-7]) предлагаются разные подходы к решению этой проблемы. Однако, они практически не используются в астрометрических исследованиях. С другой стороны, как показывает детальный анализ, предлагаемые решения не всегда приводят к удовлетворительным результатам. Поэтому задача этой работы состоит в том, чтобы привлечь внимание к различным методам вычисления средневзвешенного решения, провести их сравнение на различных наборах исходных данных и попытаться предложить для практического использования новый подход к решению рассматриваемой задачи, лишенный некоторых недостатков существующих.

2. Два основных метода вычисления взвешенного среднего

Классическая процедура обработки базируется на предположении, что ошибки измерений x_i распределены нормально с нулевым математическим ожиданием, но, возможно, разными дисперсиями (что считается оправданным на практике на основании центральной предельной теоремы). Для более строгой формулировки этого предположения обозначим истинное значение измеряемой величины как x, ошибки отдельных измерений x_i как ε_i (т.е. $x_i = x + \varepsilon_i$). Тогда наше предположение можно записать как

$$\begin{aligned} E(x_i) &= x \ , \\ E(\varepsilon_i^2) &= s_i^2 \ , \end{aligned} \tag{1}$$

где $E(\ )$ обозначает математическое ожидание.

В таком случае применение метода максимального правдоподобия приводит к формулам ([3,4,6]) (индекс при σ здесь и далее указывает на метод вычисления):

$$\begin{aligned} p_i &= \frac{1}{s_i^2} \ , \\ p &= \sum_{i=1}^{n} p_i \ , \\ \bar{x} &= \frac{\sum\limits_{i=1}^{n} p_i x_i}{p} \ , \\ \sigma_1 &= \frac{1}{\sqrt{p}} \ . \end{aligned} \tag{2}$$

Если же значения s_i рассматриваются только как показатель относительной точности измерений при неизвестном масштабном факторе — ошибке единицы веса, применяется другой подход, при котором значения s_i используются только для определения относительных весов x_i, а величина ошибки единицы веса вычисляется, исходя из разброса значений последних. При этом формула для вычисления среднего значения $\bar{x}$ не изменяется, а его среднеквадратическая ошибка σ вычисляется по формулам [1,4,6]:

$$\begin{aligned} H &= \sum_{i=1}^{n} p_i (x_i - \bar{x})^2 \ , \\ \sigma_2 &= \sqrt{\frac{H}{p\,(n-1)}} \ = \ \sigma_1 \sqrt{\frac{H}{n-1}} \ . \end{aligned} \qquad (3)$$

Очевидно, такой подход соответствует оценке средневзвешенного значения методом наименьших квадратов, где каждое значение x_i сопоставляется условному уравнению вида $x_i = x + \varepsilon_i$, а веса условных уравнений p_i определяются по (2).

Таким образом, применение обоих основных подходов (2), (3) приводит к одинаковым оценкам взвешенного среднего $\bar{x}$, но к существенно разным оценкам среднеквадратической ошибки σ. Легко видеть, что оценка σ_1 не учитывает различий в величине x_i, в то время как оценка σ_2 не учитывает абсолютной величины ошибок s_i. Поэтому на практике применение любого из этих подходов может приводить к неудовлетворительным результатам, как показано на численных примерах в разделе 4.

Так, если значения x_i близки между собой, а величины s_i заметно больше расхождений между x_i, более оправданным является применение оценки σ_1. Наоборот, если расхождения между x_i заметно превышают значения s_i, целесообразно применение оценки σ_2. В последнем случае обычно полагают либо наличие систематических ошибок в x_i, либо недооценку s_i.

Разумеется, такой путь выбора оценки ошибки средневзвешенного значения σ, наиболее адекватной характеру исходной информации, нельзя признать удовлетворительным. Кроме того, для практики важно было бы иметь более формализованный критерий, для чего в литературе предложены разные подходы.

3. Практические подходы к разрешению неоднозначности при вычислении ошибки взвешенного среднего

В [1] предлагается следующий подход. При $\sigma_1 > \sigma_2$ расхождения полагаются случайными и можно принять оценку $\sigma = (\sigma_1 + \sigma_2)/2$. При

$\sigma_1 < \sigma_2$ можно предполагать наличие систематических ошибок измерений и следует использовать оценку $\sigma = \sigma_2$.

Для того, чтобы оценить, можно ли считать различие между σ_1 и σ_2 случайным в [1] предлагается использовать отношение

$$k = \frac{|\sigma_1^2 - \sigma_2^2|}{\sigma_1^2} \div \sqrt{\frac{2}{n-1}} \ . \tag{4}$$

При $k > 2$ расхождение между σ_1 и σ_2 считается значимым.

Статистически более определенным выглядит подход, предлагаемый в [5] и поддержанный в [3]. Он базируется на том факте, что при выполнении условий (1) величина H имеет распределение χ^2 с $n-1$ степенями свободы [3,7]. Как можно видеть из 3, величина H растет как с увеличением разброса значений x_i, так и с уменьшением значений s_i, т.е. увеличением весов p_i. Таким образом H является показателем соответствия ошибок s_i, приписанных измерениям x_i, и разброса последних.

Следовательно, если для заданного уровня значимости Q величина H превышает критическое значение $\chi^2(Q, n-1)$, можно предполагать наличие в x_i отскакивающих значений, систематических ошибок или других нарушений предположения (1), а также недооценку значений s_i по сравнению с разбросом значений x_i. В первом случае выбросы просто удаляются с использованием известных критериев (см., например, [7,10]), а при их отсутствии применяется масштабирование s_i с использованием множителя $\sqrt{H/(n-1)}$ [5,3]. Легко видеть, что масштабирующий множитель равен $\sigma_2/\sigma 1$ и при его применении обе оценки совпадут. Таким образом, при описанном подходе ошибка средневзвешенного значения вычисляется как

$$\sigma_3 = \begin{cases} \sigma_1 & \text{при } H \le \chi^2(Q, n-1) \ , \\ \sigma_2 & \text{при } H > \chi^2(Q, n-1) \ . \end{cases} \tag{5}$$

Таким образом, литературные источники рекомендуют применение одной из двух оценок σ, описанных выше, в зависимости от некоторых критериев, из которых предложенный в [5] выглядит более обоснованным. Однако на практике нередки случаи, когда оценки σ_1 и σ_2 различаются весьма значительно, иногда в несколько раз. Особенно сильно эта неопределенность сказывается в случае обработки малой выборки исходных данных. Рассмотрим в следующем разделе некоторые численные примеры, на которых можно проверить устойчивость рассматриваемых оценок.

4. Численные тесты

В табл. 1 приведены результаты вычислений средневзвешенных значений из двух величин x_i со среднеквадратическими ошибками s_i по формулам (2), (3), (5). При вычислении σ_3 использовался уровень значимости $Q = 99\%$, что соответствует значению критерия $\chi^2(0.99, 1) = 6.63$. В последней колонке приведена оценка, вычисленная по (6) и описанная в следующем разделе.

Пример 1 является классическим для демонстрации нереальности оценки σ_2 при равных значениях x_i. Примеры 2–15 показывают поведение различных оценок σ при изменении s_i, а примеры 16–23 — при изменении x_i.

Из примеров 1, 6, 8 видно, что оценка ошибки средневзвешенного значения σ_1 не зависит от самих усредняемых величин, в частности от их разброса, а определяется только значениями их ошибок (что, конечно, очевидно и из формул (2), но численные примеры более наглядны). Ясно, что такая оценка явно занижена, например, для тестов 8–12.

Примеры 2–7, 8–15 так же наглядно демонстрируют, что оценка ошибки средневзвешенного значения σ_2 не зависит от самих усредняемых величин, а только от их разброса, что также не вполне удовлетворительно, например для тестов 7, 14 и 15, где эта оценка явно занижена.

Оценка σ_3 обеспечивает более реальную оценку средневзвешенного значения, однако и она обладает рядом нежелательных особенностей. Во первых, для определенных наборов исходных данных величина σ_3 не зависит от ошибок усредняемых величин, что кажется неестественным (примеры 2–3 и 8–11, т.е. там, где $\sigma_3 = \sigma_2$). Другим неприятным свойством оценки σ_3 являются резкие скачки, которые легко видеть в изменении значения σ_3 при переходе от примера 3 к примеру 4, от примера 11 к примеру 12 и от примера 19 к примеру 20.

Примеры 16–23 показывают, что все три оценки σ_1, σ_2, σ_3 ведут себя неудовлетворительно также при изменении значений x_i и неизменных значениях s_i.

Легко показать также, что значительные скачки в оценке ошибки взвешенного среднего (фактически, скачки при выборе одной из двух оценок σ_1 или σ_2) могут наблюдаться даже для одного и того же набора усредняемых значений $\{x_i,\ s_i\}$, но при задании разных значений доверительной вероятности Q.

Таблица 1. Численные примеры вычисления взвешенного среднего.

No	x_1	x_2	$s_{1,2}$	$\bar{x}$	H	σ_1	σ_2	σ_3	σ_4
1	1.0	1.0	0.5	1.0	0.00	0.354	0.000	0.354	0.354
2	1.0	2.0	0.1	1.5	50.00	0.071	0.500	0.500	0.505
3			0.2		12.50	0.141	0.500	0.500	0.520
4			0.3		5.56	0.212	0.500	0.212	0.543
5			0.5		2.00	0.354	0.500	0.354	0.612
6			1.0		0.50	0.707	0.500	0.707	0.866
7			2.0		0.12	1.414	0.500	1.414	1.500
8	10.0	20.0	0.1	15.0	5000.00	0.071	5.000	5.000	5.000
9			0.5		200.00	0.354	5.000	5.000	5.012
10			1.0		50.00	0.707	5.000	5.000	5.050
11			2.0		12.50	1.414	5.000	5.000	5.196
12			3.0		5.56	2.121	5.000	2.121	5.431
13			5.0		2.00	3.536	5.000	3.536	6.124
14			10.0		0.50	7.071	5.000	7.071	8.660
15			20.0		0.12	14.142	5.000	14.142	15.000
16	10.0	10.0	1.0	10.0	0.00	0.707	0.000	0.707	0.707
17	10.0	11.0		10.5	0.50	0.707	0.500	0.707	0.866
18	10.0	12.0		11.0	2.00	0.707	1.000	0.707	1.225
19	10.0	13.0		11.5	4.50	0.707	1.500	0.707	1.658
20	10.0	14.0		12.0	8.00	0.707	2.000	2.000	2.121
21	10.0	15.0		12.5	12.50	0.707	2.500	2.500	2.598
22	10.0	16.0		13.0	18.00	0.707	3.000	3.000	3.082
23	10.0	17.0		13.5	24.50	0.707	3.500	3.500	3.571

5. Альтернативная оценка ошибки взвешенного среднего

Тестовые примеры, приведенные в предыдущем разделе, показывают, что рекомендуемые в литературе оценки σ_1, σ_2, σ_3 обладают рядом особенностей, которые при их практическом применении могут приводить к неудовлетворительных результатам. Чтобы избежать такой ситуации, можно предложить комбинированную оценку

$$\sigma_4 = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{\frac{1}{p}\left(1 + \frac{H}{n-1}\right)}\ . \tag{6}$$

Численные тесты, приведенные в предыдущем разделе (последняя колонка табл. 1), показывают, что σ_4 обеспечивает вполне реальную оценку ошибки средневзвешенного значения без упомянутых проблем. В частности, при ее применении не наблюдается скачков при изменении исходных данных или применении разных статистических критериев.

К сожалению, автор не может предложить строгого статистического обоснования оценки 6. Самые общие соображения по этому поводу могут быть следующими. Предположим, что каждое значение x_i получено, в свою очередь, усреднением n_i измерений (что, как правило, и происходит на практике). Представим каждое "первичное" измерение в виде $x_{ij} = x + \varepsilon_{i0} + \varepsilon_{ij}$, где i – номер серии измерений, j – номер измерения в серии, ε_{i0} – систематическая ошибка i-ой серии измерений, постоянная для всех измерений данной серии, ε_{ij} – случайная ошибка j-го измерения в i-ой серии.

Положим, что ошибки ε_{i0}, ε_{ij} распределены нормально: $\varepsilon_{i0} \in N(0, \sigma_0^2)$, $\varepsilon_{ij} \in N(0, n_i\sigma_i^2)$. Здесь σ_0 рассматривается как характеристика разброса возможных значений систематических ошибок измерений.

Тогда для известных (заданных) значений x_i и σ_i имеем:

$$
\begin{aligned}
x_i &= \frac{\sum\limits_{j=1}^{n_i} x_{ij}}{n_i} \ , \\
\sigma_i &= \sqrt{\frac{\sum\limits_{j=1}^{n_i} (x_{ij} - x_i)^2}{n_i\,(n_i - 1)}} \ , \\
E(x_i) &= x + \varepsilon_{i0} \ .
\end{aligned}
\qquad (7)
$$

В первом приближении можно сделать достаточно реалистическое предположение, что систематическая ошибка σ_{0i}, постоянная для i-ой серии измерений, искажает значение x_i, но не сказывается на величине σ_i. Очевидно, величина σ_{0i} неизвестна, иначе она была бы учтена в x_i.

Последнее уравнение в (7) можно рассматривать как условное уравнение МНК, совокупность которых для всех i позволят составить и решить систему нормальных уравнений для нахождения x и σ_0, которая, очевидно, совпадает с оценкой σ_2. Далее это значение можно рассматривать как аддитивную составляющую ошибки взвешенного среднего σ, наряду с оценкой σ_1, полученной в предположении отсутствия систематических ошибок измерений. Таким образом мы приходим к выражению (6).

6. Аппроксимация статистических распределений

Практическое применение некоторых описанных выше методов связано с применением ряда статистических критериев. При этом применение их табулированных значений практически крайне неудобно, и предпочтение обычно отдается аппроксимации статистических функций. Аппроксимирующие выражения для основных функций распределения можно найти, например, в [8,9]. Однако, формулы приведенные в этих и аналогичных изданиях зачастую слишком сложны и, как правило, справедливы для больших значений числа степеней свободы, что делает их бесполезными для многих практических приложений. В то же время, можно предложить более простые формулы, обеспечивающие достаточную точность аппроксимации для всех необходимых на практике случаев.

Для приближенного вычисления квантилей χ^2-распределения $\chi^2(Q,\nu)$, где Q – уровень значимости в процентах, ν – число степеней свободы, можно применить выражение ([9]):

$$\chi^2(Q,\nu) = \nu \left(1 - t + \Psi(p)\sqrt{t}\right)^3 \ , \tag{8}$$

где $t = 2/9\nu$, $p = 1 - 0.01\,Q$, $\Psi(p)$ – функция обратная функции нормального распределения $N(0,1)$. Для $0.95 < p < 0.99999$ можно предложить апроксимирующую формулу:

$$\Psi(p) = -0.795617 - 0.014665\,q + 1.576687\,\sqrt{q - 0.513114} \ , \tag{9}$$

где $q = -\ln(1-p)$, которая имеет ошибку апроксимации $1.5 \cdot 10^{-4}$, что примерно втрое лучше точности формулы, предложенной в [9] и обеспечивающей аппроксимацию с ошибкой до $5 \cdot 10^{-4}$. Другие простые формулы для приближенного вычисления некоторых статистических критериев с достаточной для практических нужд точностью, которые могут быть полезны при решении как рассматриваемой в настоящей работе, так и других задач, можно найти в [10,11].

7. Заключение

Несмотря на то, что вычисление средневзвешенного значения является широко распространенной задачей, она не имеет однозначного, пригодного для практического применения решения даже для простейшего случая некоррелированных ошибок измерений.

Проведенное в настоящей работе сравнение различных методов вычисления взвешенного среднего и их тестирование на численных примерах показало, что предлагаемые в литературе и используемые на практике оценки ошибки средневзвешенного значения не обеспечивают решения задачи. При определенном сочетании исходных данных (усредняемых измерений) такие оценки (2), (3), (5) могут быть резко заниженными, испытывать скачки при небольших изменениях в исходных данных и т.д.

Предложенная автором для практического применения оценка (6) лишена указанных недостатков и она обеспечивает реальную величину ошибки взвешенного среднего. К сожалению, автор не может предложить строгого статистического обоснования этой оценки, что, впрочем, характерно и для ряда других подходов к решению рассматриваемой задачи.

Конечно, настоящая работа не претендует на окончательное решение этой важной для практики проблемы. В частности, приведенные выше оценки средневзвешенного значения и его ошибки могут оказаться несостоятельными при наличии корреляционных связей между измерениями [6]. Там же рассмотрен новый подход к гарантированной характеристике точности. Однако при практическом решении задачи определения средневзвешенного значения информация о корреляциях обычно отсутствует, что вынуждает нас считать предположение об их отсутствии истинным. Применение гарантированных оценок [6], при отсутствии данных о корреляции исходных измерений, также приводит к произволу в определении точности средневзвешенных значений, однако позволяет оценить влияние возможных корреляционных связей.

Список литературы

[1] Агекян Т. А. Основы теории ошибок для астрономов и физиков. М.: Наука, 1972.

[2] Большаков В. Д., Гайдаев П. А. Теория математической обработки геодезических измерений. М.: Недра, 1977.

[3] Брандт З. Статистические методы анализа наблюдений. М.: Мир, 1975.

[4] Bevington P. R. Data reduction and error analysis for the physical sciences. McGraw–Hill Book Company, USA, 1969.

[5] Rosenfeld A. H., Barbero-Galtieri A., Podolski W. J., Price L. R., Soding P., Wohl C. G. Data on Particles and Resonant States. Rev. Mod. Phys., 1967, **39**, No 1, 1–51.

[6] Эльясберг П. Е. Измерительная информация: сколько ее нужно? как ее обрабатывать? М.: Наука, 1983.

[7] Ватутин В. А., Телевинова Т. М., Чистяков В. П. Вероятностные методы в физических исследованиях. М.: Наука, 1985.

[8] Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1983.

[9] Справочник по специальным функциям. М.: Наука, 1979.

[10] Малкин З. М. Об исключении резко выделяющихся измерений. Астрон. Цирк., 1993, № 1555, 33–34.

[11] Малкин З. М. Простое вычисление квантилей t-распределения. Астрон. Цирк., 1993, № 1554, 43.

З. М. Малкин
О вычислении средневзвешенного значения.

Оригинал-макет подготовлен с помощью системы LaTeX

Подписано к печати 21.01.2001 Формат 60 × 90/16. Офсетная печать. Печ.л. 1
Уч.-изд.л. 1 Тираж 100 Заказ 30 бесплатно

Отпечатано в типографии ПИЯФ РАН
(188350 Ленинградская обл., г. Гатчина, Орлова роща).

Институт прикладной астрономии РАН, 197110, С.-Петербург, Ждановская ул., 8.