АСТРОНОМИЧЕСКИЙ ЖУРНАЛ, 2013, том 90, № 11, с. 959–964

УДК 52-17

О ВЫЧИСЛЕНИИ СРЕДНЕВЗВЕШЕННЫХ ЗНАЧЕНИЙ В АСТРОНОМИИ

З. М. Малкин*

Главная (Пулковская) астрономическая обсерватория Российской академии наук, Санкт-Петербург, Россия

Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

Поступила в редакцию 08.10.2012 г.; принята в печать 04.04.2013 г.

Отмечается, что хотя при вычислении средневзвешенного результата из нескольких определений (что является одной из часто используемых в науке операций) вычисление самого среднего обычно не вызывает вопросов, вычислению его адекватной ошибки не всегда уделяют должного внимания. Применение методов классической статистики к данным реальных наблюдений часто не может быть оправдано, поскольку не выполняются допущения, лежащие в основе этих методов. Типичными примерами служат систематические ошибки усредняемых измерений и недооценка приписываемых им случайных ошибок, используемых для назначения весов. В связи с этим рассматривается новый подход к вычислению ошибки средневзвешенного значения, основанный на комбинации известных методов. Предлагаемый метод позволяет автоматически учитывать как случайные ошибки исходных данных, так и их разброс, что делает его пригодным для определения реалистичной ошибки среднего как для хорошо, так и для плохо согласованных данных наблюдений.

DOI: 10.7868/S0004629913110042

1. ВВЕДЕНИЕ

Вычисление среднего часто применяется в астрономии, как и в других науках, для установления наивероятнейшего значения из двух или нескольких определений, полученных в разных вариантах одного исследования или независимыми группами авторов. При этом одинаково важное значение имеют как само среднее значение определяемой величины, так и его ошибка. Последняя служит не только для оценки статистической достоверности полученного среднего результата. Ошибка определения астрономической или физической величины зачастую имеет решающее значение для проверки теории наблюдениями. Поэтому получению реалистичной ошибки среднего значения должно придаваться не меньшее внимание, чем получению самого среднего значения.

Характерными особенностями рассматриваемой задачи являются малые выборки (начиная от двух значений) и отсутствие достаточной информации об исходных данных, которая позволила бы получить полезные оценки систематических ошибок усредняемых результатов и их корреляционные зависимости. Поэтому применение при их решении стандартных методов, рассчитанных на достаточно представительные выборки, при довольно жестких допущениях о статистических свойствах исходных измерений, оказывается теоретически необоснованным.

Для решения этой задач предложены несколько практических подходов, обычно не имеющих строгого теоретического обоснования, но зато доказавших на практике способность получать реалистичные средние значения и их ошибки. Обзор основных из них приведен в [1]. Однако, поскольку эти методы не применяются в астрономических работах, и нет причин рекомендовать их использование, мы здесь не будем на них останавливаться. Классические процедуры остаются наиболее востребованными в большинстве астрономических работ.

В настоящей работе проведено сравнение двух основных классических подходов к вычислению среднего. Их использование приводит к одинаковым оценкам среднего, но иногда существенно разным оценкам его ошибки в зависимости от согласованности исходных данных. Для преодоления этого недостатка в [2] была предложена комбинированная оценка ошибки среднего, позволяющая учитывать как согласованность усредняемых величин, так и их случайные ошибки. Эффективность нового метода исследована на ряде примеров, взятых из астрономической литературы.

*E-mail: malkin@gao.spb.ru

2. ВЫЧИСЛЕНИЕ СРЕДНЕГО ВЗВЕШЕННОГО И ЕГО ОШИБКИ

При вычислении среднего исходная информация состоит из измерений x_i $(i = 1, \ldots, n)$ со среднеквадратичными ошибками s_i. Задача состоит в получении среднего значения $\bar{x}$ и его ошибки σ. Наиболее часто используется классическое средневзвешенное значение, вычисляемое с весами p_i, обратно пропорциональными дисперсиям s_i^2:

$$p_i = \frac{1}{s_i^2}, \quad p = \sum_{i=1}^{n} p_i, \quad \bar{x} = \frac{\sum_{i=1}^{n} p_i x_i}{p}. \qquad (1)$$

Ошибка среднего вычисляется как

$$\sigma_1 = \frac{1}{\sqrt{p}}. \qquad (2)$$

Легко видеть, что σ_1 зависит только от ошибок входных данных s_i и не зависит от самих значений усредняемых величин x_i, в частности от их разброса. Это является характерной особенностью метода, поскольку он основан на предположении, что все x_i принадлежат к одной генеральной совокупности, а значит различия между ними признаются случайными. Как показано ниже на модельных и реальных примерах, оценка σ_1 нередко получается явно заниженной и неадекватной входным данным, что, собственно, и послужило стимулом к этой работе.

При другом подходе, основанном на методе наименьших квадратов, среднее значение и его ошибка находятся из решения системы условных уравнений вида $x_i = \bar{x} + \varepsilon_i$ с весами p_i. Решение дает то же самое значение $\bar{x}$, что и в предыдущем случае, но с другой оценкой σ:

$$\sigma_2 = \sqrt{\frac{\sum_{i=1}^{n} p_i (x_i - \bar{x})^2}{p(n-1)}}. \qquad (3)$$

Отсюда следует, что σ_2 зависит от разброса исходных данных и от соотношения их ошибок, но не от их абсолютной величины. Иными словами, при масштабировании s_i на одинаковый множитель величина σ_2 не изменится (σ_1 при этом изменится пропорционально этому множителю). На практике σ_2 обычно применяется для вычисления средних без учета весов, т.е. при p_i=1.

Для полноты заметим, что иногда ошибка среднего вычисляется просто как средняя ошибка входных значений (см., например, [3]). Однако такой способ вряд ли можно считать статистически обоснованным и практически полезным.

Таким образом, два основных метода классической статистики дают одинаковые значения средневзвешенного значения, но разные оценки его ошибки, причем ни одна из них не использует всю входную информацию. Легко показать, что $\sigma_1 = \sigma_2$ для согласованных данных, степень чего определяется по известному критерию нормированного χ^2:

$$\frac{\chi^2}{\mathrm{dof}} = \frac{1}{n-1} \sum_{i=1}^{n} p_i (x_i - \bar{x})^2. \qquad (4)$$

Если χ^2/dof существенно больше единицы, это говорит о наличии в исходных данных систематических ошибок (смещений) либо о недооценке ошибок s_i. Для согласованных данных, т.е. таких, для которых приписанные им случайные ошибки соответствуют разбросу, должно быть $\chi^2/\mathrm{dof} = 1$. Поскольку

$$\sigma_2 = \sqrt{\frac{\chi^2}{\mathrm{dof}}} \sigma_1, \qquad (5)$$

то $\sigma_1 = \sigma_2$ при $\chi^2/\mathrm{dof} = 1$. Это условие редко выполняется на практике. Обычно наблюдается нарушение этого критерия в ту или иную сторону, что говорит о несогласованности данных или просто о слишком малом объеме выборки.

Иногда применяется следующий прием искусственного приведения исходных данных к согласованному виду (см., например, [4]). Сначала вычисляется классическая ошибка среднего σ_1, затем χ^2/dof. Если последняя величина больше единицы, ошибки исходных данных масштабируются на множитель $\sqrt{\chi^2/\mathrm{dof}}$, в результате чего достигается выполнение равенства $\chi^2/\mathrm{dof} = 1$. Очевидно, что после этой операции оценка ошибки σ_1 становится равной σ_2. Представляется, что такой метод (фактически, подгонка исходных данных под заданный критерий) не дает общего решения проблемы. То же самое можно сказать про похожий способ, примененный, например, в [5, 6], где выполнение критерия $\chi^2/\mathrm{dof} = 1$ достигается сложением в квадратуре исходных ошибок s_i с подобранной для выполнения условия согласия аддитивной ошибкой.

Различные подходы к выбору одного из двух значений σ_1 или σ_2, предложенные в обзорах и монографиях [7—10], рассмотрены в работах [1, 2]. Все они требуют для принятия решения произвольного выбора некоторого граничного параметра, обычно доверительной вероятности. При этом небольшие изменения этого граничного параметра иногда приводят к “переключению” между σ_1 и σ_2, т.е. к скачку в окончательной ошибке результата [1, 2]. По этой причине, по мнению автора, эти методы не могут быть рекомендованы к использованию в астрономических задачах.

Для практического применения желательно иметь оценку ошибки среднего, позволяющую учитывать как ошибки исходных усредняемых значений s_i, так и разброс самих значений x_i

без введения элементов субъективного выбора. Удовлетворяющий этим требованиям и простой в использовании метод вычисления σ предложен в [2]. Он базируется на следующих соображениях.

Представим каждое исходное значение в виде $x_i = x + \varepsilon_i + \varepsilon_{0i}$, где x — истинное значение измеряемой величины, $\varepsilon_i \in N(0, s_i)$ — случайная ошибка, а $\varepsilon_{0i} \in N(0, \sigma_0)$ — систематическая ошибка x_i. Систематический характер ошибки ε_{0i} означает, она смещает x_i, но не смещает s_i.

Тогда мы можем записать для математического ожидания каждого из исходных значений $\mathcal{E}(x_i) = x + \varepsilon_{0i}$ $(i = 1, \ldots, n)$. Рассматривая их как набор n условных уравнений и решая его методом наименьших квадратов, получим оценку σ_0, которая, очевидно, совпадает с σ_2, которую можно рассматривать как аддитивную ошибку по отношению к σ_1. Таким образом, мы приходим к комбинированной оценке ошибки средней взвешенной величины:

$$\sigma_c = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{\frac{1}{p}\left(1 + \frac{\chi^2}{\mathrm{dof}}\right)}. \qquad (6)$$

Такое квадратичное суммирование случайной и систематической ошибок является общепринятым в астрономической практике [11] (см. также цитируемые там работы).

Конечно, приведенное выше обоснование предлагаемого подхода к вычислению ошибки среднего не является строго статистически обоснованным, но то же самое справедливо, пожалуй, для всех практически используемых методов решения рассматриваемой задачи. В то же время оценка σ_c, в отличие от других, сочетает в себе несколько привлекательных качеств: она позволяет автоматически учитывать все входную информацию, включая случайные ошибки исходных данных и их разброс, а также она очень проста в применении. В следующем разделе будут приведены примеры ее использования в сравнении с классическими оценками σ с привлечением тестовых и реальных данных.

3. ТЕСТОВЫЕ И РЕАЛЬНЫЕ ПРИМЕРЫ

Для сравнения трех методов вычисления ошибки среднего, описанных в предыдущем разделе, рассмотрим сначала тестовый пример (рисунок). В нем вычисляется среднее значение и его ошибка для трех наборов данных, имеющих одинаковые значения усредняемых величин, но разные их ошибки. В каждом следующем наборе данных ошибки в 3 раза больше, чем в предыдущем. Таким образом, соотношение ошибок сохраняется, но их масштаб увеличивается от случая A к случаю B.

Приведенные на рисунке результаты показывают, что σ_1 растет пропорционально s_i, как и следует из (1) и (2). Однако для случая A эта оценка выглядит явно заниженной, поскольку не учитывает разброс данных, существенно превышающий их случайную ошибку (т.е. данные не согласованы). Поэтому оценка σ_c для этого случая определяется, практически, величиной σ_2, зависящей как раз от разброса, как следует из (6).

Поскольку сами усредняемые значения одинаковы во всех трех наборах данных, так же как соотношение ошибок, оценка σ_2 получилась одна и та же во всех трех случаях, как следует из (3). Но это также кажется неудовлетворительным, поскольку естественно ожидать, что ошибка среднего должна расти с ошибками исходных данных. Так, в случае B мы наблюдаем ситуацию, противоположную случаю A. В этом случае уже σ_2 выглядит явно заниженной.

Оценка σ_c выглядит наиболее реалистичной во всех трех случаях, поскольку она отражает как разброс данных, так и рост σ_2 ошибок от случая a к случаю c.

В таблице приведены несколько примеров реальных результатов вычисления средневзвешенных значений, взятых из астрономических публикаций. Надо заметить, что авторские ошибки среднего обычно соответствуют σ_1.

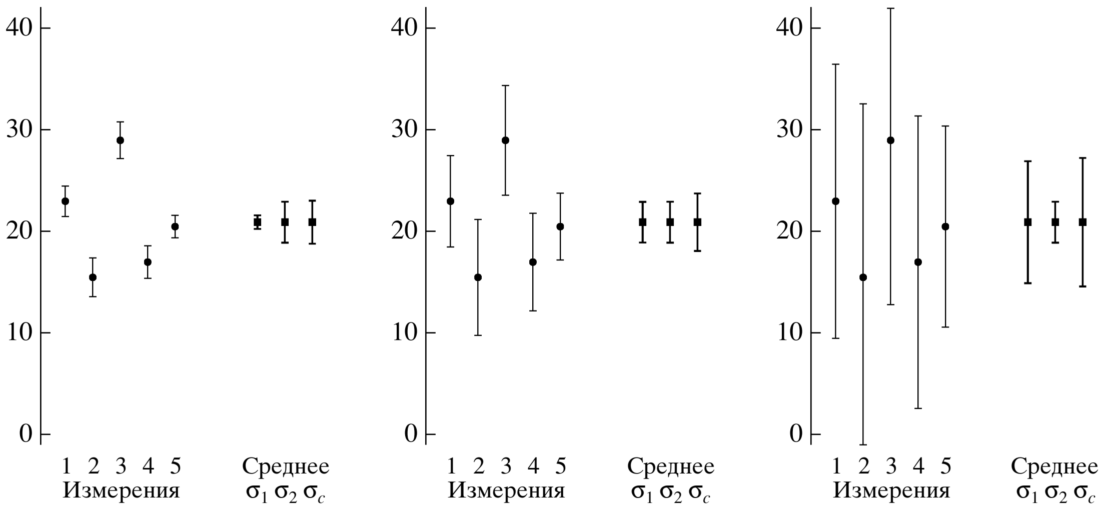
В первом примере разброс усредняемых значений (σ_2) существенно больше случайной ошибки (σ_1), что соответствует $\chi^2/\mathrm{dof} > 1$. В результате авторская оценка ошибки среднего выглядит недооцененной более чем в 1.5 раза по сравнению с комбинированной оценкой σ_c. Второй пример иллюстрирует это еще лучше — разброс исходных данных почти в 2 раза больше случайной ошибки, поэтому авторская оценка ошибки среднего занижена вдвое. Такая же ситуация в примере 6.

В то же время в примере 3 случайная ошибка оказалась вдвое больше разброса, данные согласованы, и поэтому комбинированная оценка близка к авторской. В этом случае $\chi^2/\mathrm{dof} < 1$, и применение авторами оценки σ_1 практически не приводит к недооценке ошибки среднего. То же самое можно сказать про пример 4.

В примере 5 авторская величина ошибки среднего выглядит недооцененной в 3.5 раза, поскольку не учитывает большой разброс исходных данных. Здесь также наглядно видна еще одна особенность оценки σ_1 — она всегда меньше минимальной ошибки входных величин, как следует из (1), (2). Таким образом, одно исходное значение с недооцененной ошибкой автоматически приводит к недооценке ошибки среднего независимо от качества остальных усредняемых данных.

Можно привести много других примеров вычисления средневзвешенных оценок как из цитированных выше, так и из других работ, к которым

Случай А	Случай Б	Случай В
$x_1 = 23.0 \pm 1.4$	$x_1 = 23.0 \pm 4.2$	$x_1 = 23.0 \pm 12.6$
$x_2 = 15.5 \pm 1.7$	$x_2 = 15.5 \pm 5.1$	$x_2 = 15.5 \pm 15.3$
$x_3 = 29.0 \pm 1.4$	$x_3 = 29.0 \pm 4.2$	$x_3 = 29.0 \pm 12.6$
$x_4 = 17.0 \pm 1.6$	$x_4 = 17.0 \pm 4.8$	$x_4 = 17.0 \pm 14.4$
$x_5 = 20.5 \pm 1.0$	$x_5 = 20.5 \pm 3.0$	$x_5 = 20.5 \pm 9.0$
$\bar{x} = 21.41$	$\bar{x} = 21.41$	$\bar{x} = 21.41$
$\sigma_1 = 0.60$	$\sigma_1 = 1.81$	$\sigma_1 = 5.42$
$\sigma_2 = 2.15$	$\sigma_2 = 2.15$	$\sigma_2 = 2.15$
$\sigma_c = 2.24$	$\sigma_c = 2.81$	$\sigma_c = 5.83$



Пример вычисления взвешенного среднего для трех наборов данных, состоящих из 5 одних и тех же усредняемых величин с разными ошибками.

применимы сделанные выводы. В целом можно сказать, что некритическое использование классической оценки σ_1 для ошибки средневзвешенного значения часто приводит к существенно заниженной величине этой ошибки. Напротив, комбинированная оценка ошибки среднего σ_c во всех случаях выглядит наиболее реалистичной и соответствующей исходным данным без введения в процесс вычислений элементов произвольного выбора или искусственного изменения исходных данных.

4. ЗАКЛЮЧЕНИЕ

Хотя вычисление средневзвешенного значения является часто используемой процедурой в астрономических работах, не все возникающие при этом проблемы решены удовлетворительным для практики образом. В частности, нельзя считать оправданным формальное использование методов классической статистики к реальным данным наблюдений, которые часто имеют систематические ошибки и недооцененные случайные. Наиболее часто используемая оценка σ_1 во многих случаях дает явно заниженное значение ошибки среднего. Применяемые в ряде работ различные эмпирические приемы искусственного масштабирования ошибок исходных величин для достижения выполнения критерия согласия $\chi^2/\mathrm{dof} = 1$ полностью не решают проблему.

В настоящей работе проведено сравнение двух классических методов вычисления ошибки взвешенного среднего, основанных на использовании весов исходных величин, обратно пропорциональных их среднеквадратичным ошибкам (σ_1), и на учете отклонения исходных величин от среднего методом наименьших квадратов (σ_2). Ни одна из рассмотренных оценок ошибки среднего не учитывает всей входной информации, что приводит к недооценке вклада в ошибку среднего или случайных ошибок входных величин или разброса их значений. Чтобы решить эту проблему, предложена комбинированная оценка σ_c, эффективность

Примеры вычисления средневзвешенных значений в астрономической литературе

№ п/п	Работа	Параметр	Данные	Авторский результат	σ_1	σ_2	σ_c
1	[12]	Постоянная Орта A	15.0 ± 0.8	$14.2 \pm \mathbf{0.5}$	0.44	0.65	**0.79**
			14.4 ± 1.2				
			11.3 ± 1.1				
			14.8 ± 0.8				
			14.5 ± 1.5				
2	[13]	Показатель цвета $H - K$ астероида	0.05 ± 0.05	$0.12 \pm \mathbf{0.02}$	0.018	0.034	**0.039**
		(911) Agamemnon	0.15 ± 0.02				
3	[14]	Нижний предел модуля градиента	22.0 ± 4.1	$20.1 \pm \mathbf{2.9}$	2.86	1.90	**3.44**
		скорости вращения Галактики $\partial V_\odot/\partial z$	18.2 ± 4.0				
4	[15]	Наклон функции светимости	-1.73 ± 0.20	$-1.87 \pm \mathbf{0.13}$	0.132	0.091	**0.160**
		галактик α	-2.01 ± 0.21				
			-1.91 ± 0.32				
5	[16]	Соотношение изотопов ^{12}CN/^{13}CN	67.0 ± 28.0	$67.5 \pm \mathbf{1.0}$	1.06	3.58	**3.74**
		в молекулярных облаках	54.0 ± 15.1				
			48.8 ± 19.5				
			62.1 ± 22.2				
			85.0 ± 3.3				
			64.9 ± 35.6				
			70.7 ± 3.6				
			36.3 ± 3.5				
			68.7 ± 1.3				
			133.6 ± 33.0				
			68.4 ± 4.9				
6	[17]	Эллиптичность скоплений галактик	0.37 ± 0.05	$0.27 \pm \mathbf{0.03}$	0.028	0.047	**0.055**
			0.27 ± 0.07				
			0.24 ± 0.08				
			0.26 ± 0.06				
			0.09 ± 0.07				

которой проверена на модельных и практических примерах. В результате тестирования показано, что предлагаемая оценка позволяет получить реалистичную ошибку среднего как для хорошо, так и для плохо согласованных усредняемых данных. Важно, что предлагаемый метод также хорошо работает для малых выборок, начиная с 2–3 значений, что часто встречается на практике.

К сожалению, авторы не всегда критически подходят к определению ошибки среднего значения. Всегда надо иметь в виду, что малая формальная ошибка, вычисленная по стандартным формулам среднего взвешенного (1) и (2) при большом по сравнению с этой ошибкой разбросе исходных данных свидетельствует о недооценке их формальных ошибок или наличии в этих данных систематических различий. Последние, как правило, не поддаются оценке (в противном случае они были бы учтены в публикуемых результатах). Поэтому применение классического подхода к определению ошибки среднего значения зачастую является статистически необоснованным. Предлагаемый способ автоматически учитывает как ошибки исходных данных, так и их разброс, что было подтверждено в реальных исследованиях — например, [18–20].

СПИСОК ЛИТЕРАТУРЫ

1. Z. Malkin, e-Print arXiv:1110.6639 [physics.data-an] (2011).
2. З. М. Малкин, Сообщ. Ин-та прикл. астрон. РАН № 137 (2001).
3. L. Rizzi, E. V. Held, I. Saviane, *et al.*, Monthly Not. Roy. Astron. Soc. **380**, 1255 (2007); e-Print arXiv:0707.0521 (2007).
4. A. B. Kovačević, Monthly Not. Roy. Astron. Soc. **419**, 2725 (2012); e-Print arXiv:1109.6455 (2011).
5. M. T. Murphy, J. K. Webb, and V. V. Flambaum, Monthly Not. Roy. Astron. Soc. **384**, 1053 (2008); e-Print arXiv:astro-ph/0612407 (2006).
6. K. Ando, T. Nagayama, T. Omodaka, *et al.*, Publ. Astron. Soc. Japan **63**, 45 (2011); e-Print arXiv:1012.5715 (2012).
7. A. H. Rosenfeld, A. Barbaro-Galtieri, W. J. Podolsky, *et al.*, Rev. Mod. Phys. **39**, 1 (1967).
8. Т. А. Агекян, *Основы теории ошибок для астрономов и физиков*, 2-е изд. (М.: Наука, 1972).
9. S. Brandt, *Data analysis: statistical and computational methods for scientists and engineers, 3rd ed.* (N.Y.: Springer, 1999).
10. W. Bich, M. Cox, T. Estler, *et al.*, *Proposed guidelines for the evaluation of the key comparison data*, http://www.bipm.org/cc/CCAUV/Allowed/3/CCAUV02-36.pdf.
11. З. М. Малкин, Астрон. журн. 90 (2013, в печати).
12. J. Klačka, e-Print arXiv:0912.3112 [astro-ph.GA] (2009).
13. D. W. Smith, P. E. Johnson, W. L. Buckingham, and R. W. Shorthill, Icarus **99**, 485 (1992).
14. В. В. Витязев, А. С. Цветков, Письма в Астрон. журн. 38, 467 (2012).
15. R. J. Bouwens, G. D. Illingworth, P. A. Oesch, *et al.*, Astrophys. J. (Letters) **752**, L5 (2012); e-Print arXiv:1105.2038 (2011).
16. A. M. Ritchey, S. R. Federman, and D. L. Lambert, Astrophys. J. **728**, 36 (2011); e-Print arXiv:1012.1296 (2012).
17. J. Sayers, S. R. Golwala, S. Ameglio, and E. Pierpaoli, Astrophys. J. **728**, 39 (2011); e-Print arXiv:1010.1798 (2010).
18. J. Sokolova and Z. Malkin, Astron. and Astrophys. **474**, 665 (2007).
19. Z. Malkin, in: *Measuring the Future*, Proc. Fifth IVS General Meeting, St.-Petersburg, Russia, March 2–6, 2008, eds A. Finkelstein, D. Behrend (St.–Petersburg: Nauka, 2008), p. 256; e-print arXiv:0911.3124 (2009).
20. З. М. Малкин, Астрон. журн. **88**, 880 (2011).