О ВЫЧИСЛЕНИИ ОШИБКИ СРЕДНЕГО ВЗВЕШЕННОГО

Малкин З.М.[1,2]
[1]Главная (Пулковская) астрономическая обсерватория РАН, Санкт-Петербург, Россия
[2]Астрономический институт им. В.В. Соболева СПбГУ, Санкт-Петербург, Россия

В работе на примере тестовых и реальных данных исследуются два основных классических подхода к вычислению среднего взвешенного значения и его ошибки. Показано, что оба они могут давать неудовлетворительные оценки ошибки среднего при применении к реальным измерениям. Для преодоления обнаруженных проблем предложена комбинированная оценка ошибки среднего, которая дает реалистичную оценку ошибки как для согласованных, так и для несогласованных данных.

Введение

Вычисление средних значений из нескольких определений (измерений) является одной из наиболее частых и важных задач в науке и практике. Во многих случаях 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. Также приходится сталкиваться с не всегда корректно определенными случайными ошибками усредняемых данных. Поэтому применение при их решении стандартных методов, рассчитанных на достаточно представительные выборки, при довольно жестких допущениях о статистических свойствах исходных измерений (например, принадлежность к одной генеральной совокупности), оказывается теоретически необоснованным.

Поэтому было предложено много практических методов решения задачи усреднения наблюдательных данных (что уже характеризует ее сложность). На рис.1 иллюстрируются далеко не все подходы к вычислению среднего значения. Однако в астрономических работах обычно используется классический метод вычисления среднего с весами, хотя это далеко не всегда обосновано, как будет показано ниже.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{x}_w = \frac{\sum_{i=1}^{n} p_i x_i}{p} \qquad \tilde{x} = \frac{\sum_{i=1}^{k} n_i x_i/(s_i^2)}{\sum_{i=1}^{k} n_i/(s_i^2)} \qquad \frac{|\sigma_\xi^2 - \sigma_\xi^{*2}|}{\sigma_\xi^2} : \sqrt{\frac{2}{n-1}} = k$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \sigma_1 = \frac{1}{\sqrt{p}} \qquad w_i = \frac{1}{s_i^2 + s_b^2} \qquad w_i = 1/\sigma_i^2$$

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \sigma_2 = \sigma_1\sqrt{\frac{H}{n-1}}$$

$$\sigma_c = \sqrt{\frac{1}{p}\left(1 + \frac{H}{n-1}\right)} \qquad \sigma_3 = \begin{cases} \sigma_1, & \text{if } H \le \chi^2(Q, n-1) \\ \sigma_2, & \text{if } H > \chi^2(Q, n-1) \end{cases}$$

$$\bar{x} \pm \frac{t_{(1-\alpha/2, n-1)} s}{\sqrt{n}}$$

$$\sigma_2 = \sqrt{\frac{\sum_{i=1}^{n} p_i (x_i - \bar{x}_w)^2}{p\,(n-1)}} \qquad V\hat{a}r(\tilde{x}) = \frac{1}{\sum_{i=1}^{k}\frac{1}{s_i^2}} \qquad V\hat{a}r(\tilde{x}) = \frac{1}{\sum_{i=1}^{k}\frac{1}{s_i^2}}\left(1 + 4\sum_{i=1}^{k}\frac{\hat{w}_i(1-\hat{w}_i)}{n_i - 1}\right)$$

$$\sigma_m = \frac{1.8582}{\sqrt{n-1}} MAD \qquad \sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \tilde{x})^2 + \frac{v_i t_i^2}{1-\gamma_i}}{n+k} \qquad KU = 2\sqrt{s_w^2 + s_b^2} \qquad \bar{x} \pm \frac{t_{(1-\alpha/2, nlab-1)} s}{\sqrt{nlab}}$$

$$V\hat{a}r(\tilde{x}) = \frac{1}{\sum_{i=1}^{k}\left(\frac{n_i-3}{n_i-1}\right)\left(\frac{1}{s_i^2}\right)} \qquad \tilde{x} \pm \Phi^{-1}(\alpha/2)\frac{\sqrt{\sum_{i=1}^{k}(x_i - \tilde{x})^2/(y + t_i^2)^2}}{\sum_{i=1}^{k} 1/(y + t_i^2)} \qquad \hat{w}_i = \frac{\left(\frac{n_i-3}{n_i-1}\right)(1/s_i^2)}{\sum_{j=1}^{k}\left(\frac{n_j-3}{n_j-1}\right)(1/s_j^2)}$$

Рис. 1. Иллюстрация, показывающая выборочную сводку формул, применяемых для обработки реальных данных в соответствии с различными подходами.

При вычислении среднего большое значение имеет вычисление его адекватной ошибки. При наличии проблем в исходных данных, упомянутых выше, трудно говорить о строго статистически обоснованной ошибке среднего. Поэтому часто говорят о получении просто реалистичной ошибки, соответствующей согласию усредняемых величин. В астрономической литературе нередки примеры неудачного решения этой задачи. Настоящая работа посвящена вопросам вычисления адекватной ошибки средневзвешенного значения в рамках классических процедур.

Вычисление среднего взвешенного и его ошибки

Рассмотрим задачу вычисления среднего взвешенного из группы измерений. Пусть мы имеем n исходных значений x_i, $i = 1, ..., n$ с ошибками s_i. Требуется найти их средневзвешенное значение $\bar{x}$ и его ошибку σ. Тогда для классической оценки средневзвешенного значения имеем известные и наиболее часто применяемые формулы:

$$p_i = \frac{1}{s_i^2}, \quad p = \sum p_i, \quad \bar{x} = \frac{\sum p_i x_i}{p}, \quad \sigma_1 = \frac{1}{\sqrt{p}} \quad , \tag{1}$$

где p_i – веса усредняемых значений. При применении этой оценки ошибки среднего к реальным данным мы встречаемся со следующими двумя проблемами:

– σ_1 зависит только от ошибок исходных данных, рассогласование самих исходных значений x_i на ошибку среднего не влияет;

– одно исходное значение с недооцененной ошибкой автоматически приводит к недооценке ошибки среднего, поскольку, как видно из (1), ошибка среднего всегда меньше любой из ошибок исходных величин.

Как следствие, практическое применение формул (1) может приводить к неудовлетворительным результатам.

При другом подходе можно вычислить средневзвешенное значение с применением метода наименьших квадратов к условным уравнениям вида $x_i = \bar{x} + \varepsilon_i$ с теми же весами p_i, что и раньше. В этом случае мы получим то же значение среднего $\bar{x}$, но другое значение его ошибки:

$$\sigma_2 = \sqrt{\frac{\sum p_i (x_i - \bar{x})^2}{p(n-1)}}. \tag{2}$$

Из (2) видно, что оценка σ_2 зависит только от разброса исходных данных x_i, и от относительной величины их ошибок s_i, но она инвариантна по отношению к абсолютной величине ошибок s_i (т.е. при их масштабировании на постоянный множитель ни среднее значение, ни его ошибка не меняются; ошибка σ_1 в этом случае изменится на этот множитель). Этот метод обычно используется для вычисления невзвешенного среднего, т.е. при p_i.= 1.

Таким образом, два основных метода классической статистики дают одинаковые значения среднего, но разные значения его ошибки. При этом обе оценки ошибки среднего, как σ_1, так и σ_2, не используют всю входную информацию, содержащуюся во входных данных. Легко показать [4], что $\sigma_1 = \sigma_2$ для согласованных данных, причем степень согласованности определяется по известному критерию нормированного χ^2:

$$\frac{\chi^2}{dof} = \frac{1}{n-1} \sum p_i (x_i - \bar{x})^2, \tag{3}$$

где dof – число степеней свободы, в нашем случае $n–1$. Для согласованных данных, для которых случайные ошибки соответствуют разбросу, должно быть $\chi^2/dof = 1$. Это эквивалентно равенству оценок σ_1 и σ_2. Если величина χ^2/dof существенно больше единицы, это говорит о наличии в исходных данных систематических ошибок либо о недооценке ошибок исходных данных s_i.

Иногда применяется следующий прием искусственного приведения усредняемых данных к согласованному виду. Сначала вычисляются σ_1 и χ^2/dof. Если последняя величина больше единицы, исходные ошибки s_i умножаются на множитель $\sqrt{(\chi^2/dof)}$, в результате чего достигается выполнение равенства $\chi^2/dof = 1$ и, соответственно, равенство ошибок σ_1 и 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, где выполнение критерия $\chi^2/dof = 1$ 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 работах литературе, рассмотрены в [2, 4]. 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. в скачке в окончательной ошибке результата. По этим причинам, 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. Удовлетворяющая этим требованиям комбинированная оценка предложена в [2]:

$$\sigma_c = \sqrt{\sigma_1^2 + \sigma_2^2}\,. \tag{4}$$

Тестирование оценок ошибки среднего

Для сравнения трех методов вычисления ошибки среднего, описанных в предыдущем разделе, рассмотрим сначала тестовый пример (рис. 1). 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 три раза больше, чем в предыдущем. Таким образом, соотношение ошибок сохраняется, но их масштаб увеличивается от случая *a* к случаю *c*.

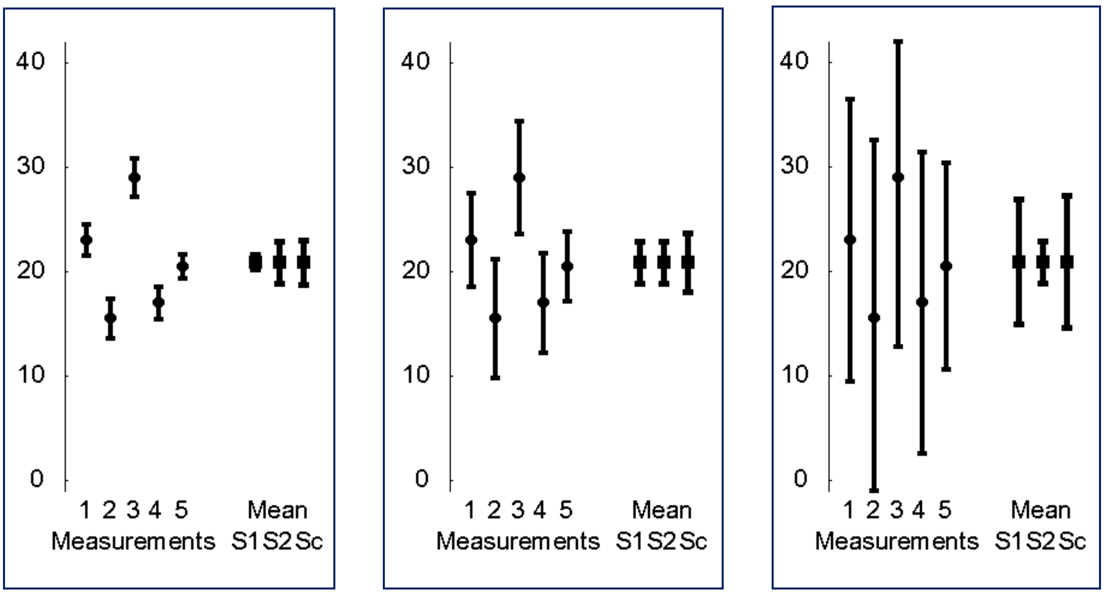

Рис. 2. Пример вычисления взвешенного среднего для трех наборов данных, состоящих из пяти одних и тех же усредняемых величин с разными ошибками. В каждом следующем случае ошибки увеличены в три раза по сравнению с предыдущим.

Приведенные на рис. 2 результаты показывают, что σ_1 растет пропорционально s_i, как и следует из (1). 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 исходных данных, как следует из (2).

Поскольку сами усредняемые значения одинаковы во всех трех наборах данных, также как соотношение ошибок, оценка σ_2 получилась одна и та же во всех трех случаях, как следует из (2). Но это также кажется неудовлетворительным, поскольку естественно ожидать, что ошибка среднего должна расти с ошибками исходных данных. Так в случае *c* мы наблюдаем ситуацию противоположную случаю *a*, когда случайные ошибки существенно превышают разброс (т.е. данные согласованы). 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 ошибок от случая *a* к случаю *c*.

В табл. 1 приведено несколько примеров реальных результатов вычисления средневзвешенных значений, взятых из астрономических публикаций. Надо заметить, что авторские ошибки среднего обычно соответствуют σ_1. В таблице приведены следующие данные: определение постоянной Оорта A [3], измерение показателя цвета H–K астероида (911) [6], определение соотношения изотопов ^{12}CN/^{13}CN в молекулярных облаках [5], оценка нижнего предела вертикального градиента угловой скорости вращения Галактики $\partial V_\Theta/\partial z$ [1].

Таблица 1. Примеры вычисления среднего из астрономической литературы.

№	Работа	Исходные данные	Авторский результат	σ_1	σ_2	σ_c
1	[3]	15.0 ± 0.8 14.4 ± 1.2 11.3 ± 1.1 14.8 ± 0.8 14.5 ± 1.5	14.2 ± 0.5	0.44	0.65	0.79
2	[6]	0.05 ± 0.05 0.15 ± 0.02	0.12 ± 0.02	.018	.034	.039
3	[5]	67.0 ± 28.0 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	67.5 ± 1.0	1.0	3.6	3.7
		133.6 ± 33.0 68.4 ± 4.9				
4	[1]	22.0 ± 4.1 18.2 ± 4.0	20.1 ± 2.9	2.9	1.9	3.4

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

В примере 3 авторская величина ошибки среднего выглядит недооцененной в 3.5 раза, поскольку не учитывает большой разброс исходных данных. Здесь также наглядно видна особенность оценки σ_1, отмеченная выше, – она всегда меньше минимальной ошибки входных величин, что способствует получению неудовлетворительного результата в данном случае.

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

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

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 $\chi^2/dof = 1$, на наш взгляд, не решают проблему.

В настоящей работе проведено сравнение двух классических методов вычисления ошибки взвешенного среднего, основанных на использовании весов исходных величин обратно пропорциональных их среднеквадратическим 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 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. К тому же она проста в использовании. Важно, что предлагаемый метод также хорошо работает для малых выборок, начиная с двух 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) 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. Напротив, предлагаемый способ автоматически учитывает как ошибки исходных данных, так и их разброс.

Более подробное изложение этой работы приведено в [4].

Литература

1. *Витязев В.В., Цветков А.С.* Кинематика звезд северного и южного галактических полушарий. Письма в Астрон. журн., 2012, т. 38, 467–485.
2. *Малкин З.М.* О вычислении средневзвешенного значения. Сообщ. ИПА РАН, 2001, № 137.
3. *Klačka J.,* Galactic tide. arXiv:0912.3112, 2009.
4. *Malkin Z.* On computation of a common mean. arXiv:1110.6639, 2011.
5. *Ritchey A.M., Federman, S.R., Lambert D.L.* Interstellar CN and CH^+ in Diffuse Molecular Clouds: $^{12}C/^{13}C$ Ratios and CN Excitation. AJ, 2011, v. 728, 36.
6. *Smith D.W., Johnson P.E., Buckingham W.L., Shorthill R.W.* JHK photometry of selected Trojan and Hilda asteroids Icarus, 1992, v. 99, 485–488.

ON COMPUTATION OF THE ERROR OF THE WEIGHTED MEAN

Malkin Z. M.[1,2]

[1]*Central Astronomical Observatory at Pulkovo of RAS, St. Petersburg, Russia*

[2]*Sobolev Astronomical Institute, St. Petersburg State University, St. Petersburg, Russia*

Combining several measurements of the same physical quantity is one of the most general and important tasks in science and practical applications. Small samples, biased input estimates, not always adequate reported uncertainties, and unknown error distribution make a rigorous solution very difficult, if not impossible. For this reason, many methods to compute a common mean and its uncertainty were proposed, each with own advantages and shortcomings. Most of them are variants of the weighted average (WA) approach with different strategies to compute WA and its standard deviation. In this paper, these two methods in most widely used modifications are compared using simulated and real data. To overcome some problems of known approaches to compute the WA uncertainty, a new combined estimate has been proposed. It has been shown that the proposed method can help to obtain more robust and realistic estimate suitable for both consistent and discrepant measurements.