ISSN 1063-7729, Astronomy Reports, 2013, Vol. 57, No. 11, pp. 882–887.

On the Calculation of Mean-Weighted Value in Astronomy

Z. M. Malkin*

Central (Pulkovo) Astronomical Observatory, Russian Academy of Sciences, St. Petersburg, Russia
St. Petersburg State University, St. Petersburg, Russia

Received October 8, 2012; in final form, April 4, 2013

Abstract—Although the calculation of the weighted mean of several individual values (one of the most frequently used operations in scientific analysis) is straightforward, the calculation of the corresponding uncertainty does not always receive the attention it requires. The application of methods of classical statistics to real observational data is often not justified, since the assumptions lying at the basis of these methods are not satisfied. The presence of systematic uncertainties in the averaged measurements and underestimation of the corresponding random errors used to define the weights are typical examples. A new approach to calculating the uncertainty of weighted mean based on a combination of known methods is considered. The proposed method makes it possible to automatically take into account both the random errors and scatter of the input data, making the method suitable for determining realistic errors for mean values in the case of observational data displaying both good and poor consistency.

DOI: 10.1134/S1063772913110048

1. INTRODUCTION

As in other sciences, mean values are calculated in astronomy to establish the most probable value based on two or more estimates obtained in multiple measurements during a single study or several independent studies. Both the mean value itself and its uncertainty are equally important. The latter serves not only to estimate the statistical reliability of the derived mean value, but often plays a deciding role in verifying theories based on observations. Therefore, the derivation of realistic errors for a mean value should receive no less attention than the calculation of the mean value itself.

The calculation of a mean and its uncertainty is usually carried out using small samples (as few as two values), and there is often an absence of sufficient information about the input data to enable useful estimation of any systematic errors in the averaged data or correlational dependences between them. Therefore, the application of standard methods assuming the use of fairly representative samples and imposing strict requirements on the statistical properties of the input measurements is theoretically unjustified.

Several practical approaches to solving this problem have been proposed, usually without a rigorous theoretical basis, but demonstrating in practice the ability to obtain realistic mean values and their uncertainties. A review of the most important of these approaches is presented in [1]. However, since these methods are not applied in astronomical studies and there is no basis for recommending their use, we will not consider them further here. Classical procedures remain the most in demand for astronomical analyzes.

We have compared the two main classical approaches to calculating a mean value. Their application leads to the same estimate of the mean, but sometimes appreciable differences in the estimates of its uncertainty, depending on the consistency of the input data. To remove this inadequacy, a combined estimate of the uncertainty of a mean was proposed in [2], which makes it possible to take into account both the consistency of the averaged data and their random errors. The effectiveness of this new method is investigated using a number of examples taken from the astronomical literature.

2. CALCULATION OF A WEIGHTED MEAN AND ITS UNCERTAINTY

The input data for the calculation of a mean value is a set of measurements x_i $(i = 1, \ldots, n)$ with their rms uncertainties s_i. The task at hand is to obtain the mean $\bar{x}$ and its error σ. Most often, the classical weighted mean is used, calculated with weights p_i that are inversely proportional to the dispersions s_i^2:

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

The error in the mean is calculated as

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

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

It is easy to see that σ_1 depends only on the uncertainties in the input data s_i and not on the values of the averaged quantities x_i; in particular, it does not depend on their scatter. This is characteristic of this method, since it is based on the assumption that all the x_i values belong to the same parent population, so that any differences between them are purely random. As we show below using model and real examples, the uncertainty σ_1 is often clearly underestimated and inadequate to fully describe the input data; this has essentially provided the stimulus leading to the current study.

In another approach based on the least-squares method, the mean and its uncertainty are found by solving a system of conditional equations of the form $x_i = \bar{x} + \varepsilon_i$ with the weights p_i. This solution yields the same value for $\bar{x}$ as the previous method, but with a different error σ:

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

It follows that σ_2 depends on the scatter of the input data and their relative errors, but not on their absolute values. In other words, if s_i is scaled by some factor, σ_2 remains unchanged (σ_1 varies proportional to this scaling factor). In practice, σ_2 is usually used to calculate mean values without applying weights, i.e., with p_i=1.

For completeness, we note that the uncertainty of a mean is sometimes calculated simply as the mean of the input errors (see, e.g., [3]). However, this method is not statistically justified, and so not of practical use.

Thus, the two well grounded methods in classical statistics considered here yield the same value for the weighted mean, but different estimates of its error; one of these methods does not use all the available input information. It is easy to show that $\sigma_1 = \sigma_2$ for consistent input data, with the degree of consistency determined by the known criterion for normalizing χ^2:

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

If χ^2/dof appreciably exceeds unity, this indicates the presence of systematic errors (offsets) in the input data, or alternatively underestimation of the input uncertainties s_i. For self-consistent data, i.e., when the assigned random errors correspond to the scatter in the data, we should find that $\chi^2/\text{dof} = 1$. Since

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

$\sigma_1 = \sigma_2$ when $\chi^2/\text{dof} = 1$. This condition is rarey satisfied in practice. A violation of this criterion in one direction or the other is usually observed, indicating a lack of consistency in the input data, or simply an inadequately small sample size.

The following method for artificially bringing the input data into a consistent form is sometimes used (see, e.g., [4]). First, the classical error of the mean σ_1 is calculated, then χ^2/dof. If this latter quantity exceeds unity, the uncertainties in the input data are scaled by the factor $\sqrt{\chi^2/\text{dof}}$, with the result that the equality $\chi^2/\text{dof} = 1$ is achieved. It is obvious that, after this operation, the error estimates σ_1 become equal to σ_2. This approach (essentially, fitting the input data using the specified criterion) does not yield a general solution to this problem. The same is true of the similar method applied, for example, in [5, 6], where the criterion $\chi^2/\text{dof} = 1$ is achieved by adding an appropriately selected error to the input errors s_i in quadrature.

Various approaches to choosing one of the two values σ_1 or σ_2 proposed in [7–10] are considered in [1, 2]. These all require an arbitrary choice of some boundary parameter, usually a confidence level. Further, small variations of this boundary parameter sometimes lead to "switching" between σ_1 and σ_2, i.e., to a jump in the final error of the resulting mean [1, 2]. For this reason, we believe that these methods should not be recommended in application to astronomical data.

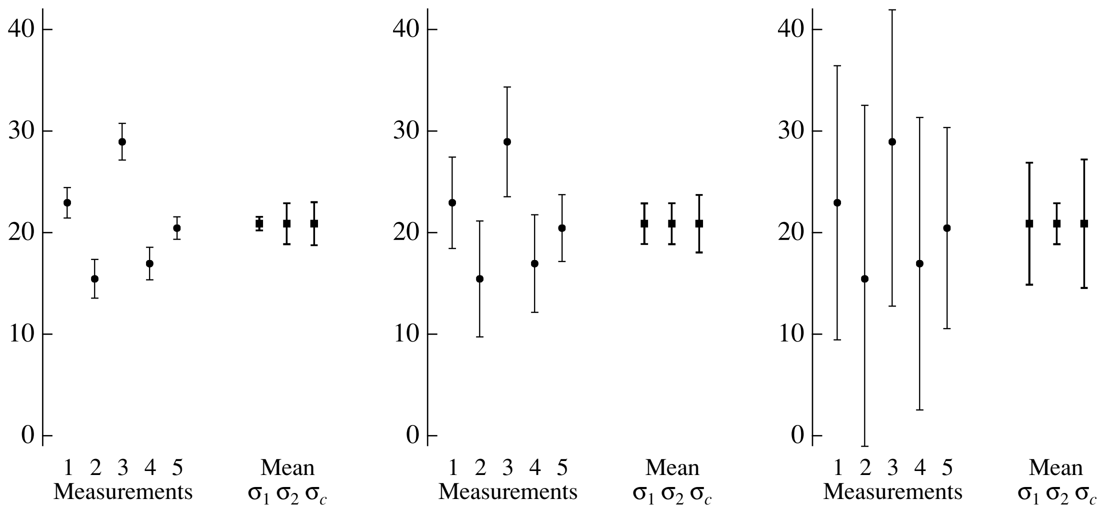
For practical applications, it would be desirable to have an estimate of the error of the mean that takes into account both the uncertainties in the input averaged values s_i and the scatter of these values x_i, without introducing subjectively chosen elements. A method for calculating σ that satisfies these requirements and is simple to apply is proposed in [2]. This method is based on the following considerations.

We represent each input value in the form $x_i = x + \varepsilon_i + \varepsilon_{0i}$, where x is the true value of the measured quantity, $\varepsilon_i \in N(0, s_i)$ is the random error in x_i, and $\varepsilon_{0i} \in N(0, \sigma_0)$ is the systematic error in x_i. The systematic character of ε_{0i} means that it shifts the value of x_i, but not s_i.

We can then write for the mathematical expectation of each of the input values $\mathcal{E}(x_i) = x + \varepsilon_{0i}$ ($i = 1, \ldots, n$). Treating these as a set of n conditional equations and solving this system using the least-squares method, we obtain the estimate σ_0, which clearly corresponds to an additive error in relation to σ_1. We thus arrive at the combined estimate of the error of a weighted mean:

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

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



Example of calculating a weighted mean for three sets of data consisting of the same five quantities with different errors.

This quadratic addition of the random and systematic errors is generally accepted in astronomical problems [11] (see also references therein).

Of course, the basis for this approach to calculating the error of a mean is not rigorously justified statistically, but the same is perhaps true for all methods aimed at solving this problem in practice. At the same time, the error estimate σ_c, in contrast to others, has several attractive qualities: it automatically takes into account all available input information, including both the random errors and the scatter of the input data, and it very simple to apply. In the following section, we present examples of applying this approach to test and real data, compared to the use of the two classical estimates of σ.

3. TEST AND REAL EXAMPLES

We first consider a test example to compare the three methods for calculating the error of a mean described in the previous section (figure). Namely, we calculated the mean and its error for three sets of data with the same values for the quantities to be averaged but different errors for these quantities. In each subsequent dataset, the errors are a factor of three larger than in the previous dataset. Thus, the ratio of the errors is preserved, but they are increasingly scaled upward from Case A to Case C. The results presented in the figure show that σ_1 grows in proportion to s_i, as follows from (1) and (2). However, this error seems obviously underestimated in Case A, since the scatter of the data, which appreciably exceeds the random errors (i.e., the data are not consistent), has not been taken into account. Therefore, the estimate σ_c for this case is essentially determined by σ_2, which depends on the scatter, as follows from (6).

Since the values of the averaged quantities and the ratios of their errors are the same for all three datasets, the estimate σ_2 is the same in all three cases, as follows from (3). However, this is also unsatisfactory, since it is natural to expect that the error of the mean should grow with the errors of the input data.

Examples of calculating weighted means from the astronomical literature

Example	Ref.	Parameter	Data	Result from Ref.	σ_1	σ_2	σ_c
1	[12]	Oort constant 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]	$H - K$ color index of an asteroid (911) Agamemnon	0.05 ± 0.05	$0.12 \pm \mathbf{0.02}$	0.018	0.034	**0.039**
			0.15 ± 0.02				
3	[14]	Lower limit for the magnitude of the gradient of the rotational velocity of the Galaxy $\partial V_\odot/\partial z$	22.0 ± 4.1	$20.1 \pm \mathbf{2.9}$	2.86	1.90	**3.44**
			18.2 ± 4.0				
4	[15]	Slope of the luminosity function of galaxies α	-1.73 ± 0.20	$-1.87 \pm \mathbf{0.13}$	0.132	0.091	**0.160**
			-2.01 ± 0.21				
			-1.91 ± 0.32				
5	[16]	Isotopic ratio ^{12}CN/^{13}CN in molecular clouds	67.0 ± 28.0	$67.5 \pm \mathbf{1.0}$	1.06	3.58	**3.74**
			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]	Ellipticity of galaxy clusters	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				

In Case C, we see the situation that is opposite to Case A, so that σ_2 now appears underestimated.

The estimate σ_c appears to be most realistic in all three cases, since it reflects both the scatter in the data and the growth in the errors from Case A to Case C.

The Table presents several examples of real results for calculations of weighted means taken from astronomical publications. Note that the errors of the mean quoted in the original papers usually correspond to σ_1.

In Example 1, the scatter of the averaged values (σ_2) is appreciably larger than the random errors (σ_1), which corresponds to $\chi^2/\mathrm{dof} > 1$. As a result, the published error of the mean appears to be underestimated by more than a factor of 1.5 compared to the combined estimate σ_c. Example 2 illustrates this even better—the scatter in the input data is nearly twice the random errors, so that the published error of the mean is probably underestimated by a factor of two. The same situation arises in Example 6.

In contrast, in Example 3, the random errors are twice the scatter in the data, so that the data are self-consistent, and the combined error estimate is close to the published value. In this case, $\chi^2/\mathrm{dof} < 1$, and the authors' use of σ_1 essentially does not lead to underestimation of the error of the mean. The same is true for Example 4.

In Example 5, the published error of the mean appears to be underestimated by a factor of 3.5, since it does not take into account the large scatter in the input data. Here, another property of σ_1 is also clearly visible—it is always smaller than the minimum uncertainty in the input data, as follows from (1), (2). Thus, one input data point with an underestimated uncertainty automatically leads to underestimation of the error of the mean, independent of the quality of the remaining averaged data.

There are many other examples of calculations of weighted means, both in the studies cited above and in other works, to which our conclusions can be applied. Overall, the uncritical use of the classical estimate σ_1 for the error of weighted mean often leads to appreciable underestimation of this error. On the contrary, in all cases, the combined error estimate σ_c yields results that are more realistic and correspond better to the input data, without the introduction of arbitrarily chosen elements or artificial adjustment of the input data.

4. CONCLUSION

Although weighted means are often calculated in astronomical studies, not all the problems that arise in this case have been solved in a way that is satisfactory in practice. In particular, the formal application of methods of classical statistics to real observational data, which often have systematic errors and/or underestimated random errors, is not justified. In many cases, the most frequently used estimate σ_1 yields clearly underestimated errors for the calculated mean value. Various empirical techniques aimed at scaling the input errors in order to satisfy the criterion $\chi^2/\mathrm{dof} = 1$, as has been done in a number of studies, do not fully solve this problem.

Here, we have compared two classical methods for calculating the error of a weighted mean, based on assigning weights to the input data that are inversely proportional to their rms errors (σ_1) and taking into account the deviations of the input data from their mean based on a least-squares approach (σ_2). Neither of these makes use of all the input information, leading to underestimation of the contribution to the error of the mean of either the random errors or the scatter of the input data. A combined error estimate σ_c has been proposed to solve this problem, whose effectiveness we have tested using model and real examples. This testing indicates that the use of σ_c makes it possible to obtain realistic errors for weighted means, for input data that display both good and poor consistency. It is important that this proposed method also works well for small samples, even those containing only two or three values, as is often the case in practice.

Unfortunately, researchers do not always approach the determination of the error of a mean quantity with a critical eye. It is important to bear in mind that calculation of a small formal error using the standard formulas for a weighted mean (1) and (2) when the scatter of the input data is large compared to their random errors testifies to either underestimation of the formal input errors or the presence of systematic differences between the input data. As a rule, it is difficult to estimate the latter (otherwise, they would have been taken into account in the published results). Therefore, applying a classical approach to determining the error of a mean quantity is often not justified statistically. The proposed approach automatically takes into account both the errors and the scatter of the input data, as has been confirmed in real studies, such as [18–20].

REFERENCES

1. Z. Malkin, arXiv:1110.6639 [physics.data-an] (2011).
2. Z. M. Malkin, Soobshch. Inst. Prikl. Astron. RAN, No. 137 (2001).
3. L. Rizzi, E. V. Held, I. Saviane, et al., Mon. Not. R. Astron. Soc. **380**, 1255 (2007); arXiv:0707.0521 (2007).
4. A. B. Kovačević, Mon. Not. R. Astron. Soc. **419**, 2725 (2012); arXiv:1109.6455 (2011).
5. M. T. Murphy, J. K. Webb, and V. V. Flambaum, Mon. Not. R. Astron. Soc. **384**, 1053 (2008); arXiv:astro-ph/0612407 (2006).
6. K. Ando, T. Nagayama, T. Omodaka, et al., Publ. Astron. Soc. Jpn. **63**, 45 (2011); arXiv:1012.5715 (2012).
7. A. H. Rosenfeld, A. Barbaro-Galtieri, W. J. Podolsky, et al., Rev. Mod. Phys. **39**, 1 (1967).

8. T. A. Agekyan, *Principles of Error Theory for Astronomers and Physicists,* 2nd ed. (Nauka, Moscow, 1972) [in Russian].
9. S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers,* 3rd ed. (Springer, New York, 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. Z. M. Malkin, Astron. Rep. **57**, 128 (2013).
12. J. Klačka, 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. V. V. Vityazev and A. S. Tsvetkov, Astron. Lett. **38**, 411 (2012).
15. R. J. Bouwens, G. D. Illingworth, P. A. Oesch, et al., Astrophys. J. Lett. **752**, L5 (2012); 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); arXiv:1010.1798 (2010).
18. J. Sokolova and Z. Malkin, Astron. Astrophys. **474**, 665 (2007).
19. Z. Malkin, in *Measuring the Future,* Ed. by A. Finkelstein and D. Behrend (Nauka, St. Petersburg, 2008), p. 256; arXiv:0911.3124 (2009).
20. Z. M. Malkin, Astron. Rep. **55**, 810 (2011).

Translated by D. Gabuzda