## The Mean of Several Quotients of Two Measured Variables

Arie Bodek Massachusetts Institute of Technology and Laboratory for Nuclear Science Cambridge, Massachusetts 02139†

 $\operatorname{and}$ 

Stanford Linear Accelerator Center Stanford, California 94305

## Abstract

It is shown that the weighted geometrical mean is a more correct estimate of the mean of several quotients of two measured variables when the variables are normally distributed and have comparable errors.

<sup>†</sup> Work supported in part by the U. S. Atomic Energy Commission under contract No. AT-(11-1)-3069.

(Submitted for publication)

The results of many experiments in physics are often presented in the form of derived quantities rather than in the form of the experimentally measured quantities. Therefore, it is not generally true that the distributions of the derived quantities are represented by the standard normal distribution. However, in many cases one wishes to find the mean of several results extracted from several experiments, or to obtain a fit to the derived quantities as a function of some other physical variable. The standard procedure that is followed is to weight the data by inverse of the squares of the quoted standard deviations when taking the mean or performing a minimum chi-squared fit to the data. The above procedure is in general only correct when the distribution of the derived quantity can be well approximated by the standard normal distribution. When the normal distribution gives a poor representation of the true distribution, one must go back to the directly measured quantities and perform a maximum likelihood analysis on the directly measured experimental data from the several experiments that one is trying to compare. This approach is often tedious because standard available computer programs cannot be used. In addition, the detailed information which is needed in order to perform such an analysis is in general not available in all the experimental papers. However, if one can find a transformation that transforms the given, non-normal distribution into a distribution which is more nearly normal, then the averaging, or minimum chi-squared fitting, can be done on the transformed variables weighted by the inverses of their errors, and a reverse transformation done at the end. Standard available computer programs can be used to fit the transformed variables.

A common example is p = 1/s, where s is a measured quantity which is normally distributed. When we wish to find the mean of several determinations of p, or when we wish to fit the functional dependence of p on some other physical variable, it is best to find the mean of s or the functional dependence of s, and then take the inverse. In general,  $\bar{p}$ , the arithmetic mean of p, will not be the inverse of  $\bar{s}$ , the arithmetic mean of s;  $\bar{p}' = 1/s$  will be the correct mean.

We now investigate the distribution of the quotient of two poisson distributed quantities. As an example, we will work with the ratio  $z = {}^{\sigma}D/\sigma_{H}$ , where  $\sigma_{D}$  is the cross section for the scattering of a projectile particle  $(\pi^{+}, e^{-}, \nu, \text{ etc.})$  from a deuterium nucleus and  $\sigma_{H}$  is the cross section for the

1

scattering of the particle from a hydrogen nucleus. Since cross sections are measured by counting the number of scattered particles, the measured cross sections are samples from poisson distributions. If the cross section determinations are based on a number of scattered particles greater than about 30, then the distributions are well approximated by normal distributions from which the negative tail portions have been truncated. Let us define the standard deviation of  $\sigma_D$ ,  $\sigma_H$ , and z by  $S_D$ ,  $S_H$ , and  $S_z$  respectively (in general S will denote a standard deviation); the mean of  $\sigma_D$ ,  $\sigma_H$ , and z by D, H, and Z respectively (in general capital letters will denote a mean); and the fractional standard deviations

 $\Delta_{D} = {}^{S}D/D$ ,  $\Delta_{H} = {}^{S}H/H$ , and  $\Delta_{z} = {}^{S}z/Z$ 

(in general  $\Delta$  will denote a fractional standard deviation).

Using the standard rules for the propagation of errors we obtain the fractional error in z.

$$\Delta_{z}^{2} = \Delta_{H}^{2} + \Delta_{D}^{2}$$
(1)  

$$\Delta = \Delta_{H} \text{ and } \Delta_{D} = C \Delta_{H}$$

$$\Delta_{z}^{2} = \Delta^{2} (1 + C^{2})$$
(2)

Before we investigate the distribution of z in detail, we write down the expressions for obtaining the average of two measurements of z for three special cases. We use common sense arguments to obtain those expressions, but later we will show that they follow from the actual distribution of z.

Case A is the case of C > > 1, i.e., when the error in the hydrogen cross section in the denominator is much smaller than the error of the deuterium cross section in the numerator. We then expect the ratio z to be normally distributed because we are effectively dividing a normally distributed variable by a constant. In that case the average of two determinations of z is just the weighted arithmetic mean of the two values.

$$\overline{Z} = \frac{Z_1 / S_1^2 + Z_2 / S_2^2}{1 / S_1^2 + 1 / S_2^2}$$
(3)

$$\frac{1}{\overline{s}^2} = \frac{1}{s_2^2} + \frac{1}{s_2^2} \tag{4}$$

Case B is the case of C << l, i.e., when the error in the deuterium cross section in the numerator is much smaller than the error of the hydrogen cross section in the denominator. This is the inverse of case A and therefore we expect 1/z to be normally distributed. In that case we should take the harmonic mean of z for two z determinations.

$$\frac{1}{\overline{Z}} = \frac{(1/Z_1) / (^{\Delta}1/Z_1)^2 + (1/Z_2) / (^{\Delta}2/Z_2)^2}{(^{\Delta}1/Z_1)^2 + (^{\Delta}2/Z_2)^2}$$
(5)

$$\left(\frac{\overline{Z}}{\overline{\Delta}}\right)^2 = \left(\frac{Z_1}{\Delta_1}\right)^2 + \left(\frac{Z_2}{\Delta_2}\right)^2$$
(6)

Remember that  $\Delta$  denotes fractional error and not absolute error.

Case C is the case of C = 1, i.e., when the fractional error in the hydrogen cross section is the same as the fractional error in the deuterium cross section. Because of this symmetry we expect our expression for obtaining the average of two z measurements to yield the same  $\overline{Z}$  whether we apply it to  $\sigma_D/\sigma_H = z$  or to  $\sigma_H/\sigma_D = 1/z$ . The weighted geometrical mean has this nice quality

$$\ln (\overline{Z}) = \frac{\ln Z_1 / \Delta_1^2 + \ln Z_2 / \Delta_2^2}{1 / \Delta_1^2 + 1 / \Delta_2^2}$$
(7)

$$1/\overline{\Delta}^2 = 1/\Delta_1^2 + 1/\Delta_2^2$$
 (8)

The above expression is a result of the assumption that  $\ln z$  is normally distributed when C = 1.

We look now at the exact distribution for z and show that it indeed becomes normal in z for C > > 1, normal in 1/Z for C < < 1, and normal in ln z for C = 1.

In the case when  $\sigma_H$  and  $\sigma_D$  are normally distributed, and  $\sigma_H$  is assumed to be practically always positive (this condition is satisfied in our case since the poisson distribution is never negative) the distribution of the quotient becomes <sup>1</sup>

$$P(z) = \frac{1}{\sqrt{2\pi}} \quad \frac{HS_D^2 - DS_H^2 z}{(S_D^2 + S_H^2 z^2)^{3/2}} \quad \exp \left[ -\frac{1}{2} \frac{(D - Hz)^2}{(S_D^2 + S_H^2 z^2)} \right]$$

From this distribution it follows that the variable Yp is normally distributed with zero mean and unit variance  $^{1}$ .

$$Y_p^2 = \frac{(1 - z/Z)^2}{C^2 \Delta + \Delta^2 (z/Z)^2}$$
;  $Z = D/H$ 

If z is normally distributed (Case A), then it follows that the variable  $Y_A$  is normally distributed with zero mean and unit variance.

$$Y_A^2 = \frac{(1 - z/Z)^2}{\Delta^2 (C^2 + 1)}$$

If 1/z is normally distributed (Case B), then it follows that the variable  $Y_{H}$  is normally distributed with zero mean and unit variance.

$$Y_{\rm H}^2 = \frac{(1 - Z/z)^2}{\Delta^2 (C^2 + 1)}$$

and if  $\ln z$  is normally distributed (Case C), then it follows that the variable  $Y_{G}$  is normally distributed with zero mean and unit variance.

$$Y_{G}^{2} = \frac{(\ln z/Z)^{2}}{\Delta^{2}(C^{2} + 1)}$$

We now investigate the above four distributions near their means. Let  $z/Z = 1 + \Delta'$ . We expand  $Y_p^2$ ,  $Y_A^2$ ,  $Y_H^2$ , and  $Y_G^2$  and keep terms to order  ${\Delta'}^2$ . Expanding we get

$$Y_{p}^{2} = \frac{\Delta'^{2}}{\Delta^{2}(C^{2}+1)} \left[1 - \frac{2}{C^{2}+1} \Delta' + \frac{4 - C^{2}-1}{(C^{2}+1)^{2}} \Delta'^{2} \dots \right]$$
(Exact)

$$Y_{A}^{2} = \frac{\Delta r^{2}}{\Delta^{2} (C^{2} + 1)}$$
 (Arithmetic)

$$Y_{\rm H}^2 = \frac{{\Delta'}^2}{{\Delta^2}({\rm C}^2 + 1)} \left[ 1 - 2\,{\Delta'} + 3\,{\Delta'}^2 \dots \right]$$
 (Harmonic)

$$Y_{G}^{2} = \frac{\Delta'^{2}}{\Delta^{2}(C^{2}+1)} \left[1 - \Delta' + \frac{11}{12} \Delta'^{2} \dots \right]$$
 (Geometric)

We can see from the above expansions that for c >> 1,  $Y_p \rightarrow Y_A$ , for c << 1,  $Y_p \rightarrow Y_H$ . For c = 1,  $Y_p \rightarrow Y_G$  up to terms of order  ${\Delta'}^2/2$ . This means that when the fractional errors of the two variables in a quotient are the same, the resulting distribution of the ratio is well approximated by a distribution in which the logarithm of z is normally distributed.

Note

For C = 2, 
$$Y_p^2 \cong (Y_A^2 + Y_G^2)/2$$
  
For C = 1/2,  $Y_p^2 \cong (Y_H^2 + Y_G^2)/2$ 

So as long as the  $1/2 \Delta_{\rm H} < \Delta_{\rm D} < 4 \Delta_{\rm H}$ , the geometric mean of several determinations of z will provide a better estimate of z than the arithmetic mean of z or the harmonic mean of z. In experiments designed to measure  $\sigma_{\rm D}/\sigma_{\rm H}$ , the error in  $\sigma_{\rm D}/\sigma_{\rm H}$  is minimized if the time is divided such as to make

$$\Delta_{\rm D}^2 = \sqrt{\frac{\sigma_{\rm H}}{\sigma_{\rm D}}} \Delta_{\rm H}^2, \qquad \text{i.e.,} \qquad {\rm C}^4 = \frac{\sigma_{\rm H}}{\sigma_{\rm D}}$$

(As long as other corrections to the cross sections are small and the beam conditions for each target and the number of nuclei in each target are the same). In general, it is probably best to calculate the three different means and compare them. The geometrical mean will be the closest to the true mean as long as  $16 < \sigma_D / \sigma_H < 1/16$  and the measurements of  $\sigma_D$  and  $\sigma_H$  are designed such as to minimize the error in  $\sigma_D / \sigma_H$ .

In order to get an idea of how different the three means can be, we conducted a monte-carlo experiment. We assumed that  $\sigma_D/\sigma_H = 1$  and the distribution of each was normal with unit mean and 0.033 standard deviation (corresponding to 900 counts). We sampled 1000 such ratios. We compare the three different means to the true mean which is the arithmetic mean of the 1000 deuterium cross sections divided by the arithmetic mean of the 1000 hydrogen cross sections. The results are shown in Table 1. The geometrical mean is indeed the closest to the true mean. We also show results for other running conditions. In general, arithmetic mean < geometrical mean < harmonic mean, and the various means will differ by a fraction of about  $2\Delta^2$  where  $\Delta$  is representative of the fractional error of the input data points.

As mentioned earlier, the above analysis can also be applied when performing minimum chi-squared fits to the data. It can also be applied to results which are derived from ratios. For example, the neutron to proton cross section ratio is approximately  $(\sigma_D/\sigma_H) - 1$ . Therefore, one must add 1 before taking the mean of the logarithm and subtract 1 at the end. Further study of the ratio distribution and applications of this analysis to specific problems can be found in Ref. 2.

Our derivations were based on the assumption that H and D were normally distributed. Therefore, when we took the means of our 1000 "experiments" we weighted them equally. In the case of poisson statistics when  $\sqrt{N}$  is taken as the error we may have the case that samples from the same distribution are not weighted equally. In that case it can be shown<sup>2</sup> that for C = 1 the geometric mean is still correct. However, for C > > 1 and for C << 1 we should use the poisson-arithmetic and poisson-harmonic means respectively.<sup>2</sup>

C > > 1 poisson arithmetic:

$$\overline{Z} = \frac{Z_1^2 / S_1^2 + Z_2^2 / S_2^2}{Z_1 / S_1^2 + Z_2 / S_2^2}$$
(9)

$$\frac{\overline{Z}}{s^2} = \frac{Z_1}{s_1^2} + \frac{Z_2}{s_2^2}$$
(10)

## C < < 1 poisson-harmonic:

$$\overline{Z} = \frac{Z_1 / \Delta_1^2 + Z_2 / \Delta_2^2}{1 / \Delta_1^2 + 1 / \Delta_2^2}$$
(11)

$$\frac{\overline{Z}}{\overline{\Delta}^2} = \frac{Z_1}{\Delta_1^2} + \frac{Z_2}{\Delta_2^2}$$
(12)

7

Experimental Conditions			Mean of 1000 "Experiments"			
H <sub>2</sub> Counts	D <sub>2</sub> Counts	С	True	Arithmetic	Geometric	Harmonic
900	900	1.0	0.9990 ±0.0015	0.9956 ± 0.0015	0.9990 ± 0.0015	1.0023 ±0.0015
900	90,000	0.1	0.9998 ±0.0011	$0.9965 \pm 0.0011$	0.9982 ±0.0011	0.9998 ±0.0011
900	3600	0.5	0.9994 ±0.0012	$0.9961 \pm 0.0012$	0.9982 ±0.0012	1.0006 ±0.0012
3600	900	2.0	0.9990 ±0.0012	$0.9982 \pm 0.0012$	1.0002 ±0.0012	1.0023 ±0.0012
90,000	900	10.0	$0.9990 \pm 0.0011$	$0.9990 \pm 0.0011$	İ.0006 ±0.0011	$1.0023 \pm 0.0011$

TABLE 1

8

## REFERENCES

 W. T. Eadie, D. Dryard, F. E. James, M. Roos, B. Sadoulet, <u>Statistical</u> <u>Methods in Experimental Physics</u> (North-Holland Publishing Company, Amsterdam, 1971) p. 26.

M. G. Kendall and A. Stuart, <u>The Advanced Theory of Statistics</u> (Charles Griffin, London, 1963) vol. 1, p. 271.

R. C. Geary, "The frequency distribution of the quotient of two normal variables", J. Royal Statist. Soc. 93 (1930) 442.

E. U. Condon and H. Odishaw (eds.), <u>Handbook of Physics (McGraw-Hill,</u> New York, 1958) p. 1-151.

A. Bodek, "The mean of several quotients of two measured variables--Applications in electron scattering experiments", SLAC Technical Note (1974), in process.

2.