SLAC-TN-74-2
February 1974

The Mean of Several Quotients of Two Measured Variables - Applications in Electron Scattering Experiments

Arie Bodek

Massachusetts Institute of Technology
and Laboratory for Nuclear Science
Cambridge, Massachusetts 02139*

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. The case of poisson statistics is discussed in detail, along with specific applications in electron scattering experiments.

* Work supported in part by the U. S. Atomic Energy Commission under contract No. AT-(11-1)-3069.

(The first part of this SLAC technical note is available also as SLAC-PUB-1372)

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 σ_D is the cross section for the scattering of a projectile particle (π^+, e^-, ν, etc.) from a deuterium nucleus and σ_H is the cross section for the 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 σ_D, σ_H, and z by S_D, S_H, and S_z respectively (in general S will denote a standard deviation); the mean of σ_D, σ_H, and z by D, H, and Z respectively (in general capital letters will denote a mean); and the fractional standard deviations

$$\Delta_D = S_D/D, \quad \Delta_H = S_H/H, \quad \text{and} \quad \Delta_z = S_z/Z$$

(in general Δ will denote a fractional standard deviation).

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

$$\Delta_z^2 = \Delta_H^2 + \Delta_D^2 \tag{1}$$

$$\Delta = \Delta_H \text{ and } \Delta_D = C\Delta$$

$$\Delta_z^2 = \Delta^2 (1 + C^2) \tag{2}$$

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.

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

$$\frac{1}{\overline{S}^2} = \frac{1}{S_2^2} + \frac{1}{S_2^2} \tag{4}$$

Case B is the case of C << 1, 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.

$$\frac{1}{\overline{Z}} = \frac{(1/Z_1)/\left(\Delta_1/Z_1\right)^2 + (1/Z_2)/\left(\Delta_2/Z_2\right)^2}{1/\left(\Delta_1/Z_1\right)^2 + 1/\left(\Delta_2/Z_2\right)^2} \tag{5}$$

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

Remember that Δ 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 σ_D/σ_H = z or to σ_H/σ_D = 1/z. The weighted geometrical mean has this nice quality

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

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

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 σ_H and σ_D are normally distributed, and σ_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 [1]

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

From this distribution it follows that the variable Yp is normally distributed with zero mean and unit variance [1].

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

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.

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

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 Δ'^2. Expanding we get

$$Y_p^2 = \frac{\Delta'^2}{\Delta^2(C^2+1)}\left[1 - \frac{2}{C^2+1}\Delta' + \frac{4 - C^2 - 1}{(C^2+1)^2}\Delta'^2 \ldots\right] \quad \text{(Exact)}$$

$$Y_A^2 = \frac{\Delta'^2}{\Delta^2(C^2+1)} \quad \text{(Arithmetic)}$$

$$Y_H^2 = \frac{\Delta'^2}{\Delta^2(C^2+1)}\left[1 - 2\Delta' + 3\Delta'^2 \ldots\right] \quad \text{(Harmonic)}$$

$$Y_G^2 = \frac{\Delta'^2}{\Delta^2(C^2+1)}\left[1 - \Delta' + \frac{11}{12}\Delta'^2 \ldots\right] \quad \text{(Geometric)}$$

We can see from the above expansions that for $c >> 1$, $Y_p \to Y_A$, for $c << 1$, $Y_p \to Y_H$. For $c = 1$, $Y_p \to Y_G$ 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 $C = 2$, $\quad Y_p^2 \cong (Y_A^2 + Y_G^2)/2$

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

So as long as the $1/2\,\Delta_H < \Delta_D < 4\,\Delta_H$, 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 σ_D/σ_H, the error in σ_D/σ_H is minimized if the time is divided such as to make

$$\Delta_D^2 = \sqrt{\frac{\sigma_H}{\sigma_D}}\,\Delta_H^2, \qquad \text{i.e.,} \qquad C^4 = \frac{\sigma_H}{\sigma_D}.$$

(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 σ_D and σ_H are designed such as to minimize the error in σ_D/σ_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 Δ 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[2] 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.[2]

$C >> 1$ poisson-arithmetic:

$$\bar{Z} = \frac{Z_1^2/S_1^2 + Z_2^2/S_2^2}{Z_1/S_1^2 + Z_2/S_2^2} \tag{9}$$

$$\frac{\bar{Z}}{\bar{S}^2} = \frac{Z_1}{S_1^2} + \frac{Z_2}{S_2^2} \tag{10}$$

C < < 1 poisson-harmonic:

$$\bar{Z} = \frac{Z_1/\Delta_1^2 + Z_2/\Delta_2^2}{1/\Delta_1^2 + 1/\Delta_2^2} \tag{11}$$

$$\frac{\bar{Z}}{\bar{\Delta}^2} = \frac{Z_1}{\Delta_1^2} + \frac{Z_2}{\Delta_2^2} \tag{12}$$

TABLE 1

Experimental Conditions			Mean of 1000 "Experiments"			
H_2 Counts	D_2 Counts	C	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 ± 0.0011	0.9982 ±0.0011	0.9998 ± 0.0011
900	3600	0.5	0.9994 ±0.0012	0.9961 ± 0.0012	0.9982 ±0.0012	1.0006 ± 0.0012
3600	900	2.0	0.9990 ±0.0012	0.9982 ± 0.0012	1.0002 ± 0.0012	1.0023 ± 0.0012
90,000	900	10.0	0.9990 ±0.0011	0.9990 ± 0.0011	1.0006 ± 0.0011	1.0023 ± 0,0011

REFERENCES (Part I only)

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

   M. G. Kendall and A. Stuart, The Advanced Theory of Statistics (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.), Handbook of Physics (McGraw-Hill, New York, 1958) p. 1-151.

2. A. Bodek, "The mean of several quotients of two measured variables--Applications in electron scattering experiments", SLAC Technical Note (1974), (Part II of this technical note).

PART II : APPLICATIONS IN ELECTRON SCATTERING EXPERIMENTS

A. More on the ratio distribution

P(z), the distribution of the ratio $z = \sigma_D/\sigma_H$ when C = 1 (i.e. equal fractional errors in the measured hydrogen and deuterium cross sections), is shown in figure 1. The percent deviations from P(z) of the standard normal distribution A(z) (used for arithmetic means), of the function G(z), which the normal distribution of ln(z) (used for geometric means), and of the function H(z), which is the normal distribution of 1/z (used for harmonic means), are shown in figure 2. We show the case of H = D = 1.00, $\bar{z}$ = D/H = 1.00, $S_H = S_D = 0.033$, $S_z = 0.042$ and $\Delta_z = S_z/\bar{z} = 0.042$. As can be seen in figure 2, G(z) is an excellent approximation to the exact distribution P(z) for that case.

$$P(z) = \frac{H S_D^2 + D S_H^2 z}{\sqrt{2\pi}\,(S_D^2 + S_H^2 z^2)^{3/2}} \exp\left[-\frac{(D - H z)^2}{2\,(S_D^2 + S_H^2 z^2)}\right]$$

$$A(z) = \frac{1}{\sqrt{2\pi}\, S_z} \exp\left[\frac{(z - \bar{z})^2}{2 S_z^2}\right]$$

$$G(z) = \frac{1}{\sqrt{2\pi}\, z \Delta_z} \exp\left[-\frac{[\ln(z/\bar{z})]^2}{2\Delta_z^2}\right]$$

$$H(z) = \frac{\bar{z}}{\sqrt{2\pi}\, \Delta_z z^2} \exp\left[-\frac{(\bar{z}/z - 1)^2}{2\Delta_z^2}\right]$$

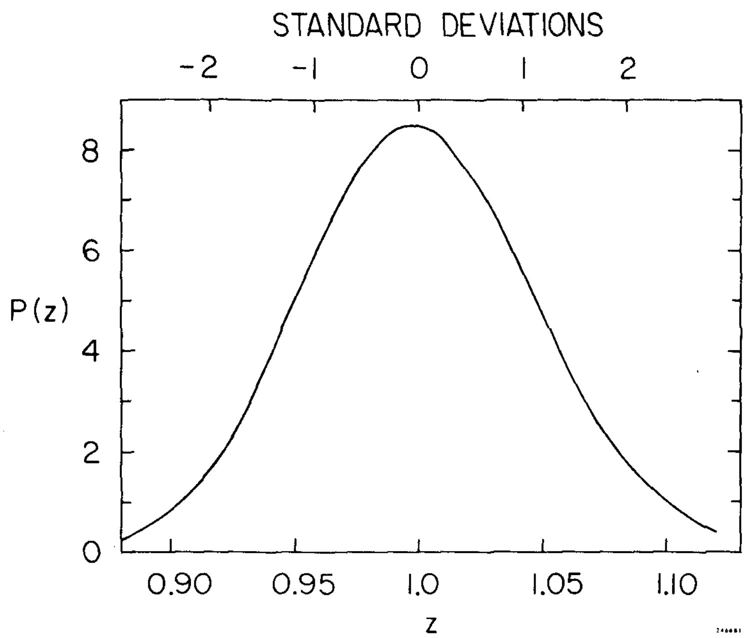

Figure 1

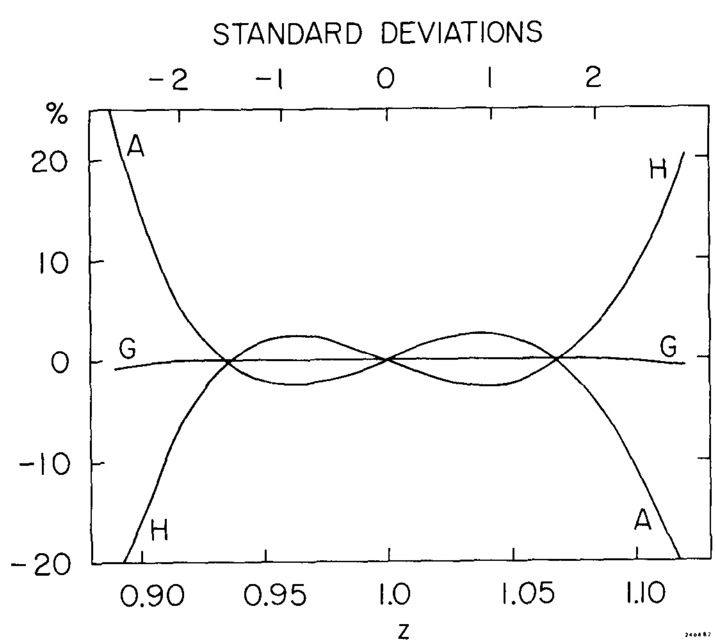

Figure 2

B. Poisson Distributions

There are two common difficulties that are encountered when the determination of the mean of several experimentally determined quantities is required.

1. The quantities are not normally distributed.
2. The estimated standard deviations are poor estimates of the true standard deviations of the distributions.

The ratio distribution represented an example of the first difficulty. The case of poisson statistics can be considered as representative either of the first or of the second difficulty. We take the case where we want to combine two measurements of the cross section σ .

$$\sigma_1 = N_1/Q_1 \pm \sqrt{N_1}/Q_1 \quad , \quad \sigma_2 = N_2/Q_2 \pm \sqrt{N_2}/Q_2$$

where N denotes the number of scattered particles detected and Q denotes the incident flux and other correction factors that convert N to a cross section (we will call Q collectively the "charge").

We now use the maximum likelihood method[1] to find the best estimate for σ, σ_*, and its error $\Delta\sigma_*$ based on the results of the above two measurements. The best value is obtained by maximizing the likelihood $\mathcal{L}$ or maximizing its logrithm w.

$$\mathcal{L}(\sigma) = \left[\frac{(\sigma Q_1)^{N_1}}{N_1!} e^{-\sigma Q_1} \right] \left[\frac{(\sigma Q_2)^{N_2}}{N_2!} e^{-\sigma Q_2} \right]$$

$$w = \ln \mathcal{L} = N_1 \ln \sigma - \sigma Q_1 + N_2 \ln \sigma - \sigma Q_2 + N_1 \ln Q_1 + N_2 \ln Q_2 - \ln (N_1! N_2!)$$

$$\frac{\partial w}{\partial \sigma} = 0 \implies \qquad \sigma_* = \frac{N_1 + N_2}{Q_1 + Q_2}$$

$$\Delta\sigma_* = \left[-\frac{\partial^2 w}{\partial \sigma^2} \right]^{-1/2}_{\sigma = \sigma_*} = \frac{\sqrt{N_1 + N_2}}{Q_1 + Q_2}$$

The above result is equivalent to weighting the measured cross sections by their corresponding charges, i.e.

$$\sigma_* = (\sigma_1 Q_1 + \sigma_2 Q_2) / (Q_1 + Q_2)$$

However, if we use the standard gaussian formula and weight the cross sections by the inverse of the square of their errors we get a different value σ_*'.

$$\sigma_*' = \frac{Q_1 + Q_2}{Q_1^2/N_1 + Q_2^2/N_2} , \quad \Delta\sigma_*' = 1/\left(Q_1^2/N_1 + Q_2^2/N_2 \right)^{1/2}$$

In the special case $Q_1 = Q_2$, $N_1 = N(1 + \Delta')$ and $N_2 = (1 - \Delta')$ the two methods yield:

$$\sigma_* = N/Q \pm \sqrt{N}/(\sqrt{2}\,Q) \qquad \text{Poisson (correct)}$$

$$\sigma_*' = \frac{N}{Q}\left(1 - \Delta'^2\right) \pm \frac{\sqrt{N(1-\Delta'^2)}}{\sqrt{2}\,Q} \qquad \text{Gaussian (incorrect)}$$

On the average $\Delta'^2 = 1/N$, so weighting the cross sections by the inverse of the square of their errors underestimates the mean by a fraction of about 1/N. It is because cross sections based on small N are weighted more due to their small estimated errors. Even if we combine a large number of cross sections each based on about N counts we will find that the mean calculated using the gaussian formula will be incorrect and low by a fraction 1/N. We combined the results of 10,000 monte carlo experiments each measuring the same unit cross section on the basis of about 100 events. We weighted each experiment by its charge and obtained a mean of 1.0018 ± 0.0010 . On the other hand, weighting each experiment by the square of the inverse of its statistical error yielded a mean of 0.9917 ± 0.0010. For N = 100 we have 1/N = 1%, and the results of the gaussian combining is indeed 1% low, which is 10 standard deviations low in this particular example.

One way of understanding this difference is to realize that the standard gaussian formula requires that we weight each experiment by the inverse of the square of the true standard deviation. $\sqrt{N}/Q$ is a poor approximation to the true standard deviation unless N is close to N_e where N_e is the expected number of counts; $N_e = \sigma Q$ where σ is the exact value of the cross section. Therefore, the true standard deviation is $\sqrt{\sigma/Q}$. So weighting by the charge is equivalent to weighting by a factor which is proportional to the square of the inverse of the true standard deviation. We may also take the point of view that the poisson distribution is only approximately normal and requires special treatment, just like the distribution for the ratio.

The correct mean can still be calculated even when the charge of each experiment is not given, when we realize that $Q_1 = \sigma_1/(\Delta\sigma_1)^2$ and $N_1 = \sigma_1^2/(\Delta\sigma_1)^2$. Therefore, the correct mean for poisson ditributed variables is (we shall call it the poisson-arithmetic mean),

$$\sigma_x = \frac{\sigma_1^2/(\Delta\sigma_1)^2 + \sigma_2^2/(\Delta\sigma_2)^2}{\sigma_1/(\Delta\sigma_1)^2 + \sigma_2/(\Delta\sigma_2)^2}$$

and the correct error is

$$(\Delta\sigma_x)^2 = \frac{\sigma_1^2/(\Delta\sigma_1)^2 + \sigma_2^2/(\Delta\sigma_2)^2}{[\sigma_1/(\Delta\sigma_1)^2 + \sigma_2/(\Delta\sigma_2)^2]^2}$$

the expression for the error can be rewritten as

$$\frac{\sigma_x}{(\Delta\sigma_x)^2} = \frac{\sigma_1}{(\Delta\sigma_1)^2} + \frac{\sigma_2}{(\Delta\sigma_2)^2}$$

In most experimental situations, the final cross sections quoted have had several correction factors applied. Similarly, the quoted errors have been increased to account for the errors in the various correction factors. Common examples are empty targets subtractions, positron and other background subtractions, efficiency corrections, radiative corrections and other similar corrections. If the main error comes from counting statistics, then the above expression for the poisson-arithmetic mean should still be used because it still accounts for the basic problem that small N lead to small errors and large N lead to large errors in runs for which the amount of charge is the same and therefore the expected number of counts is the same.

We now consider the problem of fitting the measured cross sections to a known functional form with undetermined coefficients. The standard procedure is to minimize χ^2. Minimizing χ^2 is equivalent to maximizing the likelihood only if the distributions are gaussian. Obtaining the mean of several cross sections is equivalent to a one parameter fit. As we have shown, a different expression than the one derived from the χ^2 minimization must be used in the case of poisson statistics.

In the case of linear fits of more than one parameter the standard χ^2 gaussian formula leads to a set of linear equations. The fitting function Y is a linear sum of known functions $F_i(X_a)$ with undetermined coefficients C_i. We have P cross section measurements at P values of X_a, and K coefficients C_i to be determined.

$$\chi^2 = \sum_{a=1}^{P} \frac{\left[\sigma_a - \sum_{b=1}^{K} C_b F_b(X_a)\right]^2}{(\Delta\sigma_a)^2}, \quad Y = \sum_{b=1}^{K} C_b F_b(X_a)$$

We obtain the following set of K equations (one for each C_i) by setting $\frac{\partial \chi^2}{\partial C_i} = 0$.

$$\sum_{a=1}^{P} \frac{\left[\sigma_a - \sum_{b=1}^{K} C_b F_b(X_a)\right]}{(\Delta\sigma_a)^2} F_i(X_a) = 0 \qquad 2\text{-}1$$

The K equations in 2-1 are the standard least squares fit equations. We shall now show that in the case of poisson statistics a similar set of equations should be used, but unfortunately it is a set of non-linear equations. However, the non-linear equations can easily be solved by using the linear gaussian set of equations to obtain an initial solution and iterating (probably only once).

Using $N_a = \sigma_a^2/(\Delta\sigma_a)^2$ and $Q_a = \frac{\sigma_a}{(\Delta\sigma_a)^2}$ we write the likelihood function $\mathcal{L}$ for the case of poisson statistics.

$$\mathcal{L} = \prod_{a=1}^{P} \frac{\left[Y(X_a)\, Q_a\right]^{N_a}}{N_a!} e^{-Y(X_a)\, Q_a}$$

$$W = \ln \mathcal{L} = \sum_{a=1}^{P} \left[N_a \ln Y(X_a) - Y(X_a)\, Q_a\right] + \text{Const}$$

to maximize the likelihood we set $\frac{\partial W}{\partial C_i} = 0$.

$$\sum_{a=1}^{\ell} \left\{ \frac{N_a F_i(X_a)}{\sum_{b=1}^{K} c_b F_b(X_a)} - F_i(X_b) Q_a \right\} = 0$$

Substituting for N_a and Q_a we obtain a set of K non-linear equations (one for each C_i)

$$\sum_{a=1}^{\ell} \frac{\left[\sigma_a - \sum_{b=1}^{K} c_b F_b(X_a) \right] F_i(X_a)}{(\Delta\sigma_a)^2 \cdot \left\{ \frac{\sum_{b=1}^{k} c_b F_b(X_a)}{\sigma_a} \right\}} = 0 \qquad 2\text{-}2$$

$\Leftarrow \{Y(X_a)/\sigma_a\}$

The set of equations in 2-2 is approximated by the set of equations in 2-1 only when $\{Y(X_a)/\sigma_a\} = 1$ and can be set to 1.0 in the denominators of the equations in 2-2. Equation 2-2 can be solved by using the gaussian 2-1 set of equations to obtain an initial solution for $Y(X_a)$ and iterate using the set of equations in 2-1 again, but each time modifying the input errors by factors $\sqrt{Y(X_a)/\sigma_a}$. This way, the standard least squares computer routines can still be used with only the input errors slightly modified.

C. The ratio of two poisson variables

As we have shown in the previous section, the mean of several cross sections should be computed using the poisson-arithmetic mean. In the case of several ratios of two poisson variables we still use the geometric mean to find the mean of the ratios when C = 1 (equal fractional errors in the quantities in the numerator and denominator). However, for $C \gg 1$ and for $C \ll 1$ we should use the poisson-arithmetic or the poisson-harmonic means respectively. (C was defined in part I as the ratio of the fractional errors of the quantities in the numerator and denominator). We provide a summary of the various means on the next page. In general all means will be within the fraction $\pm 2\Delta'^2$ of each other, where Δ' is representative of the fractional error of the experiments that are being combined.

Summarizing

C = 1 (Geometric mean)

$$\ln \bar{Z} = \frac{\ln Z_1/\Delta_1^2 + \ln Z_2/\Delta_2^2}{1/\Delta_1^2 + 1/\Delta_2^2}$$

$$\frac{1}{\bar{\Delta}^2} = \frac{1}{\Delta_1^2} + \frac{1}{\Delta_2^2}$$

C>>1 (Poisson-Arithmetic mean)

$$\bar{Z} = \frac{Z_1^2/S_1^2 + Z_2^2/S_2^2}{Z_1/S_1^2 + Z_2/S_2^2}$$

$$\frac{\bar{Z}}{\bar{S}^2} = \frac{Z_1}{S_1^2} + \frac{Z_2}{S_2^2}$$

C<<1 (Poisson-Harmonic mean)

$$\bar{Z} = \frac{Z_1/\Delta_1^2 + Z_2/\Delta_2^2}{1/\Delta_1^2 + 1/\Delta_2^2}$$

$$\frac{\bar{Z}}{\bar{\Delta}^2} = \frac{Z_1}{\Delta_1^2} + \frac{Z_2}{\Delta_2^2}$$

Arithmetic mean (not to be used for poisson distributions)

$$\bar{Z} = \frac{Z_1/S_1^2 + Z_2/S_2^2}{1/S_1^2 + 1/S_2^2}$$

$$\frac{1}{\bar{S}^2} = \frac{1}{S_1^2} + \frac{1}{S_2^2}$$

Harmonic mean (not to be used for poisson distributions)

$$\bar{Z} = \frac{Z_1^2/\Delta_1^2 + Z_2^2/\Delta_2^2}{Z_1/\Delta_1^2 + Z_2/\Delta_2^2}$$

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

D. The neutron to proton cross section ratio for deep-inelastic electron scattering.

The neutron to proton cross section ratio is extracted from the ratio of hydrogen and deuterium cross sections. If we neglect deuteron corrections then the ratio r is: $r = \sigma n / \sigma p = \sigma D / \sigma H - 1.0$

Now ln ($\sigma D/\sigma H$) is approximately normally distributed. This is because $1.4 \lesssim \sigma D/\sigma H \lesssim 2$ (Ref. 2). In experiments designed to minimize the error in $\sigma D/\sigma H$ the running times will be planned such as to make $1.4 \lesssim C^2 \lesssim 2$. This is well within the region where the geometric mean is applicable. All averaging and fitting should be done on $y_i = \ln (r_i + 1)$ and $\Delta y_i = \Delta r_i / (r_i + 1)$. If we wish to fit r_i to $r = a + b x + c x^2$, we should use a non linear fitting program, input arrays of y_i, Δy_i and x_i and $Y_{fit} = \ln (a + bx + cx^2 + 1)$, where a,b,c are to be determined.

The mean can be wrong by as much as $2\Delta'^2$ where Δ' is the fractional error of the individual $\sigma D/\sigma H$ ratios to be combined. Therefore, taking the wrong mean can lead to a large systematic error in the case where many measurements of poor statistics are to be averaged. When $\sigma D/\sigma H \cong 1.4$, averaging several determinations of the ratio r when the determinations are obtained from $\sigma D/\sigma H$ measurements with errors of 10% can result in an error of 0.028 if the arithmetic mean is used instead of the geometrical mean. This is important if enough runs are combined such as to make the combined error comparable to this number. It is clear that when only two or three runs are combined the systematic shift will be small in comparison to the resulting statistical errors.

E. The ratio of scalar to transverse virtual photoabsorption cross sections.

The electroproduction cross sections can be written in terms of cross sections for absorption of scalar and transverse virtual photons σs and σt.

For electrons of incident E, scattering angle θ , and scattered energy E' the cross section can be written as

$$\sigma \equiv \frac{d^2\sigma}{d\Omega dE'} = \Gamma\ (\sigma_t + \epsilon\sigma_s)$$

$$\Gamma = \frac{\alpha}{4\pi^2}\frac{2kE'}{q^2 E(1-\epsilon)}\ ,\quad k = \frac{W^2-M^2}{2M}$$

$$q^2 = 4EE'\sin^2\frac{\theta}{2}\ ,\quad W^2 = M^2 + 2M\nu - q^2,\quad \nu = E - E'$$

$$0 \leq \epsilon \leq 1 \qquad \epsilon = \frac{1}{1+2\tan^2\frac{\theta}{2}(1+\nu^2/q^2)}$$

where M is the nucleon mass.

Γ is the virtual photon flux and ϵ is the polarization of the virtual photon. The quantity R = σ_s/σ_t is of current theoretical interest. R is extracted from two or more cross section measurements at different values of $\epsilon(\theta)$ and fixed values of W and q^2. We take as an example the case of two cross sections.

$$\sigma_1 = \Gamma_1\ \sigma_t\ (1+\epsilon_1 R)$$
$$\sigma_2 = \Gamma_2\ \sigma_t\ (1+\epsilon_2 R)$$

$$\frac{\sigma_1}{\sigma_2} = \frac{\Gamma_1}{\Gamma_2}\left(\frac{1+\epsilon_1 R}{1+\epsilon_2 R}\right)$$

Now R is usually a small quantity (It's average value is about 0.16, Ref. 3) So we expand the above expression,

$$\left(\frac{\sigma_1}{\sigma_2}\right)\left(\frac{\Gamma_2}{\Gamma_1}\right) \cong 1 + (\epsilon_1 - \epsilon_2) R$$

Usually, the ϵ difference $\Delta\epsilon = \epsilon_1 - \epsilon_2$ is about 0.35 (Ref.3). Also, σ_1 is usually measured more accurately since the cross sections at the small angles (large ϵ) are larger. Typically, the ratio of the errors of σ_1 to the errors in σ_2 is about 1/2. This mean that the situation is somewhere close to half way between the case where ln (1 + 0.35 R) is gaussian distributed and the case where 1/(1 + 0.35 R) is gaussian. It is clear that in no case do we have the situation where the distribution of R itself is approximated by a gaussian. It is clear now that all least square fits ought to be done on the variable ln (1 + 0.35 R) to reduce systematic errors.

References (Part II only)

1. J. Orear, Notes on Statistics for Physicists, UCRL-8017 (Aug. 1958).

2. A. Bodek, M. Breidenbach, D.L. Dubin, J.E. Elias, J. I. Friedman, H. W. Kendall, J.S. Poucher, E.M. Riordan, M.R. Sogard, D.H.Coward, Phys. Rev. Lett. 30, 1087 (1973).

   A. Bodek, Ph. D. Thesis, Massachusetts Institute of Technology (1972), available as LNS-COO-3069-116.

3. E. M. Riordan, Ph.D. Thesis, Massachusetts Institute of Technology, 1973, available as LNS-COO-3069-176.