Journal of the Arkansas Academy of Science

Volume 50 Article 6

1996

Curing a Summing Error That Occurs Automatically When Fitting a Function to Binomial or Poisson Distributed Data

Edwin S. Braithwaite
Cedarville College

Wilfred J. Braithwaite
University of Arkansas at Little Rock

Curing a Summing Error That Occurs Automatically When Fitting a Function to Binomial or Poisson Distributed Data

Edwin S. Braithwaite
Department of Science and Mathematics
Cedarville College
Cedarville, OH 45314

Wilfred J. Braithwaite
Department of Physics and Astronomy
University of Arkansas at Little Rock
Little Rock, AR 72204

Abstract

Without special precautions a sum-rule error occurs automatically when a chi-squared procedure is used to fit a funtion to binomial or Poisson distributed histogram data if the function has at least one linear parameter. Since the square of the variance per channel is equal to the mean population, errors are usually approximated using $\{\sigma_i^2 \cong y_i > 0\}$; this choice for approximating the variance gives a per-channel error weighting of $1/y_i$ that automatically results in a sum-rule error. This sum-rule error consistently and systematically underestimates the total sum of the data points by an amount equal to the value of χ_y^2, resulting in $\sum_i y_i - \sum_i f_i = \chi_y^2$, where $\chi_y^2 \equiv \sum_i (y_i - f_i)^2/y_i$ and $f_i = f(x_i, \{\text{parameters}\})$. In contrast, using $\{\sigma_i^2 \cong f_i > 0\}$ gives the error weighting per channel of $1/f_i$ that automatically results in a less well known sum rule error. This sum-rule error which is only half as large but opposite in sign consistently and systematically overestimates the total sum of the data points by an amount equal to half the value of χ_f^2, that is, it results in $\sum_i y_i - \sum_i f_i = -\frac{\chi_f^2}{2}$, where $\chi_f^2 = \sum_i (y_i - f_i)^2/f_i$. The good news is a combination of error weightings may be constructed which completely eliminates the otherwise automatically cocuring sum-rule error by taking advantage of cancellations occuring between the two sum-rule errors implicit in the two above-mentioned approaches to error-weighting per channel. This fortunitous linear combination of sum-rule error swill combine and cancel if the fitting funtion is a sufficiently viable choice so that $\chi_f^2 \cong \chi_y^2 \cong \nu$ (number of degrees of freedom); consequently a weighted linear combination of these two definitions may be used, $\chi^2 = \frac{1}{3}\chi_y^2 + \frac{2}{3}\chi_f^2$. This choice for $\chi^2 \cong$ is equivalent to choosing an error weighting of $\frac{1}{\sigma^2} = \frac{1}{3y_i} + \frac{2}{3f_i}$, and it essentially eliminates summing errors so that $\sum_i y_i - \sum_i f_i$. An alternate method is presented and proven for $\{\mu_i = f_i\}$ in fitting a function using Maximum Likelihood.

Introduction

Chi-squared fitting is a common method of describing histogram data with an appropriate theoretical function. One advantage of this procedure is the ready availability of error estimation for the fitting parameters using the error matrix procedure (Arndt and MacGregor, 1966; Bevington and Robinson, 1992; Press et al., 1989). A huge amount of binomial or Poisson distributed data is taken every year in the form of histograms, where the error per channel is the square-root of the mean channel population (Boaz, 1983). One reason this body of physical data is so large is that data may now be acquired using automatic data processing techniques under micro-computer control often using multi-parameter analog-to-digital conversion of amplitudes of pulsed analog waveforms or using multiscaling techniques. An example is the construction of histogram data taken in a counting experiment where each "spectrum" is a frequency distribution often analyzed automatically by fitting it with a general theoretical function containing amplitudes and widths for describing a collection of peaks to the populations of all the channels, while also parametrizing the background population underlying these peaks.

The quantum nature of physical interactions requires a microscopic event to either happen or not happen despite the average likelihood of occurrence being represented as a spatially-distributed collection of fractional probabilities. This idea underlies why binomial or Poisson statistics applies to a huge body of physical data.

However a systematic difficulty with fitting a theoretical function to this type of data was indicated in early work (Bevington, 1969) which has shown a serious sum-rule problem arises using the chi-squared technique [area(date) - area(fit) = χ^2] if the measured values of channel population $\{y_i\}$ are used to represent mean channel populations in estimating errors, thus, the standard deviation estimates are taken as $\{\sigma_i^2 \cong y_i > 0\}$. This work showed a consistent and systematic underestimation of the total sum of the data points by an amount equal to the value of $\chi_y^2 \equiv \sum_i (y_i - f_i)^2/y_i$, i.e., $\sum_i y_i - \sum_i f_i = \chi_y^2$, when using a (non-linear) Gaussian and a (linear) background for $f_i \equiv f(x_i, \{\text{parameters}\})$ in an example of using chi-squared fitting of a functional form to histogram data $\{x_i, y_i\}$ when the population of each channel is a Poisson distribution.

Materials and Methods

Consider a function $f_i = f(x_i, \{a_k\})$ in the desciption of an $\{x_i, y_i\}$ histogram of Poisson or Binary distributed data, with each per-channel error given by $\sigma_i \cong \sqrt{y_i}$. If the technique of minimizing chi-squared is used to optimize the parameters $\{a_k\}$ in each f_i, the total sum of the fitting function $\Sigma_i f_i$ systematically fails to agree with $\Sigma_i y_i$ by an amount equal to the un-normalized $\chi^2 \equiv \chi^2_\nu$. Figure 1 proves the underestimation of the sum, $\Sigma_i y_i - \Sigma_i f_i = \chi^2$, and this proof is generalized for any fitting function having at least one linear parameter used to optimize a theoretical function to a histogram of binomial or Poisson distributed date $\{y_i\}$ with standard deviation estimates $\{\sigma_i = \sqrt{y_i}\}$, i.e., since yi is used to estimate each mean channel population.

$\Sigma_i y_i - \Sigma_i f_i = \chi^2$ is proved, when $\{\sigma_i \equiv \sqrt{y_i}\}$ is used in minimizing χ^2:

Consider only the linear parameter set $\{a_k\}$ when using $\{\sigma_i \equiv \sqrt{y_i}\}$ to minimize χ^2, providing a set of equations: $f_i \equiv f(x_i, \{a_k\}) = \Sigma_k a_k g_k(x_i, \alpha)$, with α representing any non-linear parameters.

$$\chi^2 = \Sigma_i \frac{(y_i - f_i)^2}{\sigma_i^2} = \Sigma_i \frac{(y_i - f_i)^2}{(\sqrt{y_i})^2} = \Sigma_i \frac{(y_i - f_i)^2}{y_i} = \Sigma_i (y_i - f_i) - \Sigma_i \frac{(y_i - f_i)}{y_i} f_i.$$

$$\frac{\partial \chi^2}{\partial a_k} = -2\Sigma_i \frac{(y_i - f_i)}{y_i} g_k(x_i, \alpha) = 0, \quad \forall \{a_k\} \text{ parameters. Each term is zero} \Rightarrow$$

$$-\frac{1}{2}\Sigma_k a_k \frac{\partial \chi^2}{\partial a_k} = \Sigma_k a_k \Sigma_i \frac{(y_i - f_i)}{y_i} g_k(x_i, \alpha) = 0.$$

Interchanging summing order above and using $f_i \equiv \Sigma_k a_k g_k(x_i, \alpha)$ gives:

$$\Sigma_i \frac{(y_i - f_i)}{y_i} \Sigma_k a_k g_k(x_i, \alpha) = \Sigma_i \frac{(y_i - f_i)}{y_i} f_i = -\chi^2 + \Sigma_i (y_i - f_i) = 0.$$

Transposing completes the proof: $\Sigma_i (y_i - f_i) = \boxed{\Sigma_i y_i - \Sigma_i f_i = \chi^2}$

Fig. 1. Systematic summing errors are shown for $\{\sigma_i^2 \cong y_i\}$: $\chi^2 = \chi^2_\nu = \Sigma y_i - \Sigma f_i$.

If the function $f_i \equiv f(x_i, \{a_k\})$ provides a good description of the histogram shape, $\chi^2 = \chi^2_\nu \cong \nu = N - p$ (the number of degrees of freedom) where N is the number of data points and p is the number of parameters in the optimization. This underestimation of the total sum results in an <u>average</u> error of only about 1 count per channel. However, there are no guarantees that in any particular channel this error will not substantially exceed 1.

It is interesting to note there is no sum-rule error and $\Sigma_i y_i = \Sigma_i f_i$ must hold for a simple least-squares fit which ignores the role of errors by setting σ_i=1 for each channel, giving $\chi^2 = \Sigma_i (y_i - f_i)^2$, provided the function, f_i, includes an overall constant term (a_1) to be optimized. The proof of $\Sigma_i y_i - \Sigma_i f_i$ is as follows: $\frac{\partial \chi^2}{\partial a1} = 2\Sigma_i (y_i - f_i) = 0$, giving $\Sigma_i (y_i - f_i) = 0$. Unfortunately this cure of the sum-rule error is of little importance since errors may not be ignored in any serious data analysis procedure.

Several remedial efforts have been suggested for minimizing the effects of the sum-rule error (Bevington, 1969; Bevington and Robinson, 1992). These remediations include smoothing histogram data using a convolution which damps the channel-to-channel flutter in $\{y_i\}$ artificially reducing the size of χ^2_ν, thus reducing the systematic summing error, $\Sigma_i y_i - \Sigma_i f_i = \chi^2_\nu$. Among other adverse effects, smoothing degrades the resolution of the experimental peaks in a greater or lesser proportion depending on the remediation being sought.

A second approach for remediation suggests a coarser binning of histogram data to reduce the number of degrees of freedom, ν, thus reducing the size of the systematic summing error $\Sigma_i y_i - \Sigma_i f_i = \chi^2_\nu \cong \nu$. Unfortunately, in providing fewer histogram bins for describing each experimental peak this approach also degrades the resolution of the experimental peaks in a greater or lesser proportion depending on the remediation being sought. A third approach suggests fitting backgrounds in regions far removed from any peaking followed by fitting the peaks while holding background parameters constant. Although a careful examination of the role of peaks versus background is generally a good idea when possible, it is not a general approach consequently it has limited applicability in curing the sum-rule problem.

One would not expect summing errors to play an important role when the number of counts in every histogram channel is quite large. However, counting statistics only improve as the square root of each mean channel population; many researchers carrying out difficult experiments or those committed to production line analysis work may not have the luxury of acquiring a large number of counts per channel in their experimental spectra.

Though many experimenters use chi-squared fitting procedures to represent their data, many may be unaware of the automatic onset of a sum-rule difficulty associated with this type of analysis, and possibly a substantial number may be more concerned with determining other fitting parameters than the populations of the peaks. However, when there are uncorrected systematic errors in a fitting method, these difficulties may have adverse consequences in determining the size of other fitting parameters, not just those associated with the strength of each peak. Thus, it would be better to fix the fundamental problem of systematic errors in the sum rule, rather than try to patch up consequences in some ad hoc fashion.

Results and Discussion

The cure to the sum-rule problem discovered by Bevington (1969) of a systematic error in (underestimating) the sum may be found by re-defining the χ^2. This cure is possible because, by remarkable contrast to Bevington's earlier work, if errors are estimated using $\{\sigma_i^2 \cong f_i > 0\}$ instead of $\{\sigma_i^2 \cong y_i > 0\}$, the chi-squared method is found to consistently and systematically overestimate the total sum of data points by half the value of $\chi_f^2 = \Sigma_i (y_i - f_i)^2/f_i$ (Braithwaite, 1974).

Underestimation (in Fig. 1) and overestimation (in Fig. 2) of the total sum are proven for the choices of error estimates $\{\sigma_i^2 \cong y_i > 0\}$ and $\{\sigma_i^2 \cong f_i > 0\}$, respectively, using only the normal equations for the linear parameters. Non-linear parameters only affect the systematic summing errors because of their influence on the numerical size of χ^2 and thus on the viability of the choice of optimizing function.

$\Sigma_i y_i - \Sigma_i f_i = -\frac{\chi^2}{2}$ is proved, when $\{\sigma_i \equiv \sqrt{f_i}\}$ is used in minimizing χ^2:

Consider only the linear parameter set $\{a_k\}$ when using $\{\sigma_i \equiv \sqrt{f_i}\}$ to minimize χ^2, providing a set of equations: $f_i \equiv f(x_i, \{a_k\}) = \Sigma_k a_k g_k(x_i, \alpha)$, with α representing any non-linear parameters.

$$\chi^2 = \Sigma_i \frac{(y_i - f_i)^2}{\sigma_i^2} = \Sigma_i \frac{(y_i - f_i)^2}{(\sqrt{f_i})^2} = \Sigma_i \frac{(y_i - f_i)^2}{f_i} = \Sigma_i \frac{(y_i - f_i)}{f_i} y_i - \Sigma_i (y_i - f_i).$$

$\frac{\partial \chi^2}{\partial a_k} = -\left(2\Sigma_i \frac{(y_i - f_i)}{f_i} + \Sigma_i \left[\frac{y_i - f_i}{f_i}\right]^2\right) g_k(x_i, \alpha) = 0, \ \forall \{a_k\}$ parameters. Each term is zero $\Rightarrow -\frac{1}{2}\Sigma_k a_k \frac{\partial \chi^2}{\partial a_k} = \Sigma_k a_k \left(\Sigma_i \frac{(y_i - f_i)}{f_i} + \frac{1}{2}\Sigma_i \left[\frac{y_i - f_i}{f_i}\right]^2\right) g_k(x_i, \alpha) = 0.$

Interchanging summing order above and using $f_i \equiv \Sigma_k a_k g_k(x_i, \alpha)$ gives:

$$\Sigma_i \frac{(y_i - f_i)}{f_i} f_i + \frac{1}{2}\Sigma_i \left[\frac{y_i - f_i}{f_i}\right]^2 f_i = \Sigma_i y_i - \Sigma_i f_i + \frac{1}{2}\chi^2 = 0.$$

Transposing completes the proof: $\Sigma_i (y_i - f_i) = \boxed{\Sigma_i y_i - \Sigma_i f_i = -\frac{\chi^2}{2}}$

Fig. 2. Systematic summing errors are shown for $\{\sigma_i^2 \cong f_i\}$ giving $\Sigma y_i - \Sigma f_i = \frac{-\chi_f^2}{2}$.

If the fitting function is a viable choice it will result in $\chi_y^2 \cong \chi_f^2$, so these two versions of χ^2, although of very similar size, will yield different signs and sizes in their associated systematic summing errors. Cancellation of these sum-rule errors by using the new weighting, $\chi^2 = \frac{1}{3}\chi_y^2 + \frac{2}{3}\chi_f^2$, in formulating and minimizing of χ^2 should essentially eliminate these summing errors. This weighting is completely equivalent to replacing the error weighting terms $\frac{1}{\sigma^2}$ by $\frac{1}{3y_i} + \frac{2}{3f_i}$ in the routine computing χ^2. Although $\frac{1}{3}$ vs $\frac{2}{3}$ should provide an adequate weighting in essentially eliminating sum-rule errors, one might view this choice as providing a good first guess for the new weighting.

Figure 3 proves an alternate method of curing the sum-rule errors for Poisson-distributed data if a Maximum Likelihood formulation with $\{\mu_i = f_i\}$ is used instead of the error-weighted χ^2 formulation. The only change needed in the numerical approach for optimizing the parameter search is to replace the routine calculating $\chi^2 = \Sigma_i \frac{1}{\sigma_i^2}(y_i - f_i)^2$ by routine (of the same name) which calculates $\Sigma_i [f_i - y_i \log(f_i)]$ in place of χ^2 for the same fitting function.

$\Sigma_i y_i = \Sigma_i f_i$ is proved in fitting data by Maximum Likelihood for $\{\mu_i \equiv f_i\}$ when the population of each channel follows a Poisson distribution

Consider only linear parameters $\{a_k\}$, and $\{\mu_i \equiv f_i\}$ in the Maximum Likelihood function: $P = \Pi_i \frac{f_i^{y_i}}{y_i!} e^{-f_i}$, for $f_i \equiv f(x_i, \{a_k\}) = \Sigma_k a_k g_k(x_i, \alpha)$, with α representing non-linear parameters. Maximizing P is equivalent to minimizing $F \equiv -L \equiv -\log(P) = \Sigma_i [f_i - y_i \log(f_i)] + \text{constant}$, using $\frac{\partial F}{\partial a_k} = 0$.

$\frac{\partial F}{\partial a_k} = \Sigma_i \left(1 - \frac{y_i}{f_i}\right) g_k(x_i, \alpha) = 0, \ \forall \{a_k\}$ parameters. Each term is zero $\Rightarrow$

$\Sigma_k a_k \frac{\partial F}{\partial a_k} = \Sigma_k a_k \Sigma_i \left(1 - \frac{y_i}{f_i}\right) g_k(x_i, \alpha) = 0.$ Interchanging summing order and using $f_i \equiv \Sigma_k a_k g_k(x_i, \alpha) \Rightarrow \Sigma_i \left(1 - \frac{y_i}{f_i}\right) \Sigma_k a_k g_k(x_i, \alpha) = \Sigma_i (f_i - y_i) =$

$\Sigma_i f_i - \Sigma_i y_i = 0.$ Transposing completes the proof: $\boxed{\Sigma_i y_i = \Sigma_i f_i}$.

Fig. 3. Systematic summing erros are cured with a maximum likelihood formulation using $\{\mu_i \cong f_i\}$.

Figure 4 shows a second-order expansion describing χ^2 in terms of all the search parameters $\{\alpha_k\}$, both linear and non-linear, near the minimum of χ^2. The first derivative of this expression is taken for each parameter, and each resulting expression is set to zero to provide N simultaneous linear equations for a set of candidate steps $\{\Delta\alpha_k\}$ to be taken toward the global minimum. Figure 4 shows these N equations as a single matrix equation. Since an absolute minimum in χ^2 is needed for every parameter in $\{\alpha_k\}$, the second derivative must be greater than zero for each diagonal-element term of the square matrix, called the curvature matrix.

Derivatives shown above may be found numerically, and since cross partials commute, only half of the off-diagonal elements in Fig. 4 need to be calculated. As noted earlier, each diagonal element in Fig. 4 must be positive to assure an absolute minimum in χ^2 occurs for every parameter in $\{\alpha_k\}$. If each diagonal element is positive and an inverse matrix is found to exist, then a set of candidate changes $\{\Delta\alpha_k\}$ in the parameters may be computed by inverting the matrix equation and using the inverse matrix to multiply the column vector on the right which is comprised of first partials of χ^2 with respect to each fitting parameter.

Using all parameters $\{\alpha_k\}$ to find absolute minimum in χ^2

The expansion: $\chi^2 = \chi_o^2 + \Sigma_k \left[\frac{\partial\chi^2}{\partial\alpha_k}\right]_o \Delta\alpha_k + \frac{1}{2}\Sigma_k\Sigma_m \left[\frac{\partial^2\chi^2}{\partial\alpha_k\partial\alpha_m}\right]_o \Delta\alpha_k \Delta\alpha_m$

is used to describe χ^2 for all search parameters $\{\alpha_k\}$ near a χ^2 minimum.

For χ^2 to be an absolute minimum in $\{\alpha_k\}$, two requirements must be satisfied: $\frac{\partial\chi^2}{\partial\alpha_k} = 0$ and $\left[\frac{\partial^2\chi^2}{\partial\alpha_k^2}\right]_o > 0$, for every parameter in $\{\alpha_k\}$. Using $\frac{\partial\chi^2}{\partial\alpha_k} = 0$ in the quadratic expansion above results in the matrix equation below:

$$\begin{bmatrix} \left[\frac{\partial^2\chi^2}{\partial\alpha_1^2}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_1\partial\alpha_2}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_1\partial\alpha_3}\right]_o & \cdots & \left[\frac{\partial^2\chi^2}{\partial\alpha_1\partial\alpha_N}\right]_o \\ \left[\frac{\partial^2\chi^2}{\partial\alpha_2\partial\alpha_1}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_2^2}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_2\partial\alpha_3}\right]_o & \cdots & \left[\frac{\partial^2\chi^2}{\partial\alpha_2\partial\alpha_N}\right]_o \\ \left[\frac{\partial^2\chi^2}{\partial\alpha_3\partial\alpha_1}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_3\partial\alpha_2}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_3^2}\right]_o & \cdots & \left[\frac{\partial^2\chi^2}{\partial\alpha_3\partial\alpha_N}\right]_o \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \left[\frac{\partial^2\chi^2}{\partial\alpha_N\partial\alpha_1}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_N\partial\alpha_2}\right]_o & \left[\frac{\partial^2\chi^2}{\partial\alpha_N\partial\alpha_3}\right]_o & \cdots & \left[\frac{\partial^2\chi^2}{\partial\alpha_N^2}\right]_o \end{bmatrix} \begin{bmatrix} \Delta\alpha_1 \\ \Delta\alpha_2 \\ \Delta\alpha_3 \\ \vdots \\ \Delta\alpha_N \end{bmatrix} = - \begin{bmatrix} \left[\frac{\partial\chi^2}{\partial\alpha_1}\right]_o \\ \left[\frac{\partial\chi^2}{\partial\alpha_2}\right]_o \\ \left[\frac{\partial\chi^2}{\partial\alpha_3}\right]_o \\ \vdots \\ \left[\frac{\partial\chi^2}{\partial\alpha_N}\right]_o \end{bmatrix}$$

This NxN matrix is inverted and used to solve for a candidate set of $\{\Delta\alpha_k\}$.

Fig. 4. A second-order expansion procedure is used to develop equations for finding the minimum in χ^2 for all parameters $\{\alpha_k\}$ both linear and nonlinear.

If any diagonal elements in Fig. 4 are found to be zero or negative, the initial parameter choice is not suffuciently close to a global minimum. Methods to remediate this problem are available in the literature (Bevington and Robinson, 1992; Marquardt, 1963; Daniel 1971). One possible remediation is to change the size and/or sign of zero or negative diagonal elements to make all diagonal elements greater than zero and then take fractionally smaller steps than prescribed by the set of candidate changes $\{\Delta\alpha_k\}$ which are calculated after matrix inversion.

Since the goal was to minimize chi-squared, numerical values were calculated for each of these derivative, symmetrically around each expansion point to provide a good representation of the quadratic expansion of χ^2 near and at its global minimum. Figure 5 outlines the procedure used for calculating both the first and second derivatives numerically. This figure shows a slightly more elaborate method of calculating the second derivatives (Abrahamovich and Stegun, 1984). The simpler methods of calculating derivatives gave adequate results: after several iterations $\Delta\chi^2$ was reductin by ~10^{-15}, with the corresponding candidate step sizes also approaching zero, $\{\Delta\alpha_k \sim 10^{-15}\}$. Also the inverse matrix routine was tested; the product of the matrix and its inverse gave 1 for each diagonal element and either zero or numbers about 10^{-15} for the off-diagonal elements when using 64-bit double precision for the real variables in all the routines including the numerical derivatives.

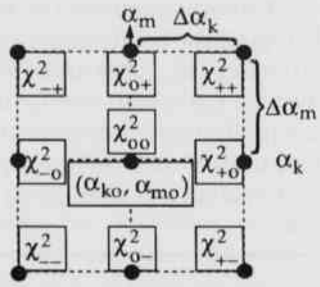
Procedures for Calculating Derivatives Numerically

Each derivative is calculated symmetrically around the expansion point to provide a good representation of the quadratic expansion of χ^2 near and at its minimum.

$$\left[\frac{\partial\chi^2}{\partial\alpha_k}\right]_o \cong \frac{\chi^2_{+o} - \chi^2_{-o}}{2\Delta\alpha_k}$$

$$\left[\frac{\partial^2\chi^2}{\partial\alpha_k^2}\right]_o \cong \frac{\chi^2_{+o} + \chi^2_{-o} - 2\chi^2_{oo}}{(\Delta\alpha_k)^2}$$

$$\left[\frac{\partial^2\chi^2}{\partial\alpha_k\partial\alpha_m}\right]_o \cong \frac{\chi^2_{++} + \chi^2_{--} - \chi^2_{+-} - \chi^2_{-+}}{4(\Delta\alpha_k)(\Delta\alpha_m)}$$



- A slightly more elaborate method uses the following:

$$\left[\frac{\partial^2\chi^2}{\partial\alpha_k^2}\right]_o \cong \frac{\chi^2_{++} - 2\chi^2_{o+} + \chi^2_{-+} + \chi^2_{+o} - 2\chi^2_{oo} + \chi^2_{-o} + \chi^2_{+-} - 2\chi^2_{o-} + \chi^2_{--}}{3(\Delta\alpha_k)^2}$$

$$\left[\frac{\partial^2\chi^2}{\partial\alpha_k\partial\alpha_m}\right]_o \cong \frac{\chi^2_{++} + 2\chi^2_{oo} + \chi^2_{--} - \chi^2_{+o} - \chi^2_{-o} - \chi^2_{o+} - \chi^2_{o-}}{2(\Delta\alpha_k)(\Delta\alpha_m)}$$

Fig. 5. Procedures in calculating numerical derivatives to fit all parameters $\{\alpha_k\}$.

In a recent publication Bevington and Robinson (1992) show a fit to a 60-point histogram with a 6-parameter fitting function. Their chi-squared analysis uses three linear parameters to describe a quadratic background and three parameters to describe the single peak: a Lorentzian with a variable area, position and width. The maximum peak in the data histogram contained 81 counts, and the smallest histogram populations were 6 and 1 in channel 1 and 2, respectively.

Figure 6(a) shows the result of calculations using Bevington and Robinson's published (1992) optimization parameters: giving $\chi^2 = \chi^2_y = 59$, with N - p = 60 - 6 = 54. The sum of data over 60 channels is $\sum_i y_i = 2000$ with the sum of their fitting function over 60 channels $\sum_i f_i = 1940$, discrepant by roughly $\chi^2_y = 59$. Their quoted fitting parameter (#4) gives an area of 276±44, lying roughly in between their sum $\sum_i (f_i - b_i) = 261$ and their sum $\sum_i (y_i - b_i) = 321$.

In contrast, Fig. 6(b) shows results from an optimization using the newly-prescribed chi-squared weighting: $\chi^2 = \frac{1}{3}\chi^2_y + \frac{2}{3}\chi^2_f \rightarrow 60$, which is completely equivalent to replacing the error-weighting terms $\frac{1}{\sigma_i^2}$ by $\frac{1}{3y_i} + \frac{2}{3f_i}$ in analyzing

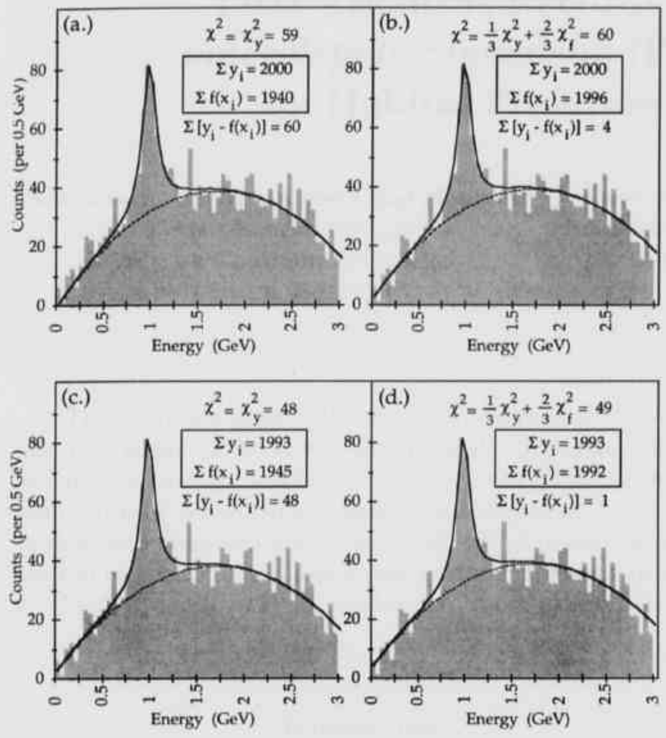


Fig. 6. A Lorentzian peak with a second-degree polynomial backgound is fitted to data, shown in gray. Systematic summing errors are seen for $\chi^2 = \chi_y^2$, in (a.) and (c.). Modifying the weighting of χ^2 in (b.) and (d.) cures this problem. (a.) and (b.) use all data; (c.) and (d.) skip the first 2 data points.

the same $\{x_i,y_i\}$ histogram data set, with N - p = 60 - 6 = 54. The sum of data over 60 channels is $\sum_i y_i$ = 2000 as compared with the sum of the new fitting function over 60 channels $\sum_i f_i$ = 1996, which is discrepant by only 4. The fitting parameter (#4) gives an area of 260±42, in good agreement with $\sum_i (f_i - b_i)$ = 252 and $\sum_i (y_i - b_i)$ = 256.

Figures 6(c) and 6(d) differ from Figs. 6(a) and 6(b) in that histogram data channels 1 and 2 are ignored in Figs. 6(c) and 6(d). Comparing these figures shows the importance of eliminating data channels containing little information but having χ^2 weightings sufficiently large to unduly distort the chi-squared fit.

Figure 6(c) shows the result of calculations using Bevington and Robinson's published (1992) optimization parameters: giving $\chi^2 = \chi_y^2$ = 48 with N - p = 58 - 6 = 52. The sum of data over 58 channels is $\sum_i y_i$ = 1993 with the sum of their fitting function over 58 channels $\sum_i y_i$ = 1945, now discrepant by χ_y^2 = 48. Fitting parameter (#4) gives an area of 269 (276 before) with $\sum_i (f_i - b_i)$ = 257 (261 before) and $\sum_i (y_i - b_i)$ = 305 (321 before). Figure 6(d) shows results of an optimization using $\chi^2 = \frac{1}{3}\chi_y^2 + \frac{2}{3}\chi_f^2 \rightarrow 49$ in analyzing the same $\{x_i,y_i\}$ histogram data set with N - p = 58 - 6 = 52. The sum of data over 58 channels is $\sum_i y_i$ = 1993 and the sum over 58 channels of $\sum_i f_i$ = 1992, discrepant by only 1. Fitting parameter (#4) gives an area of 258 (260 before) with $\sum_i (f_i - y_i)$ = 247 (252 before) and $\sum_i (y_i - b_i)$ = 247 (256 before).

Thus we have presented and proved a method for curing systematic sum-rule errors which automatically arise when using an error-weighted chi-squared fitting of a function with at least one linear fitting parameter to binomial or Poisson distributed histogram data. Further, an example histogram of data was fitted by a Lorentzian on a polynomial background. This example demonstrated the systematic summing error discovered by Bevington (1969) as well as showing an adequate elimination of the summing errors by using the presently proposed prescription for approximating the error-weightings with $\frac{1}{\sigma_i^2} = \frac{1}{3y_i} + \frac{2}{3f_i}$. In addition, an alternate method of curing the sum-rule errors for Poisson-distributed data was presented and proven if a Maximum Likelihood formulation with $\{\mu_i = f_i\}$ is used to replace the error-weighted χ^2 formulation. The only change needed in the numerical approach for optimizing the parameter search is to replace the routine calculating $\chi^2 = \sum_i \frac{1}{\sigma_i^2} (y_i - f_i)_2$ by a routine (of same name) which calculates $\sum_i [f_i - y_i \log(f_i)]$ in place of χ^2 for the same fitting function.

ACKNOWLEDGMENTS.—Support was provided by the U.S. Department of Energy through Grant DE-FG05-92ER40753] and by the Department of Science and Mathematics of Cedarville College. In addition, the authors wish to acknowledge helpful discussions with Dr. Donald C. Wold.

Literature Cited

Abrahamovich, M. and **I.M. Stegun.** 1984. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. Not corrected since its second printing in 1964. (Supt. of Documents, U.S. Govt Printing Office, Washington, D.C., 20402).

Arndt, R.A. and **M.H. MacGregor.** 1966. Chi-Squared Minimization Techniques. Methods in Computational Physics, Vol. 6: Nuclear Physics. Edited by B. Alder, S. Fernbach and M. Rotenberg. (Academic Press, New York), 253-296.

Bevington, P.R. and **D.K. Robinson.** 1992. Data Reduction and Error Analysis for the Physical Sciences. (McGraw Hill, Inc., New York), 115-167; 168-179.

Bevington, P.R. 1969. Data Reduction and Error Analysis for the Physical Sciences. (McGraw Hill, Inc., New York) 249; 134-163; 204-254.

Boaz, M.L. 1983. Mathematical Methods in the Physical Sciences (2nd Ed.). Ch. 16, Sec. 10. Applications to

Experimental Measurements (John Wiley & Sons, New York) 731-735.

Braithwaite, W.J. 1974. Systematic Summing Errors in Linear and Non-Linear Optimization of Poisson Distributed Data. ***Bull. Am. Phys. Soc. 19,*** 1078.

Daniel, J.W. 1971. The Approximate Minimization of Functionals. Prentice-Hall Series in Automatic Computation, Edited by G. Forsythe (Prentice-Hall, Inc., New Jersey) 159-179.

Marquardt, D.W. 1963. An Algorithm for Least Squares Estimation of Nonlinear Parameters. J. Soc. Ind. Appl. Math, vol. II, no 2, 431-441.

Press, W.H., B.P. Flannery, S.A. Teukolsky and **W.T. Vetterling.** 1989. NUMBERICAL RECEIPES The Art of Scientific Computing (Fortran Version), Ch. 10: Minimization and Maximization of Functions. (Cambridge Univ. Press, New York), 274-326; 498-546.