BRAGG, W. H. (1914). *Philos. Mag.* **27**, 881–899 [reproduced in *Acta Cryst.* (1969), A**25**, 3–11].
BRAGG, W. L., DARWIN, C. G. & JAMES, R. W. (1926). *Philos. Mag.* **1**, 897–922.
BRAGG, W. L., JAMES, R. W. & BOSANQUET, C. H. (1921*a*). *Philos. Mag.* **41**, 309–337.
BRAGG, W. L., JAMES, R. W. & BOSANQUET, C. H. (1921*b*). *Philos. Mag.* **42**, 1–17.
BRAGG, W. L. & WEST, J. (1928). *Z. Kristallogr.* **69**, 118–148.
BRIDGMAN, P. W. (1936). *The Nature of Physical Theory*. New York: Dover.
CALVERT, L. D., KILLEAN, R. C. G. & MATHIESON, A. McL. (1976). *Acta Cryst.* A**32**, 648–652.
COCHRAN, W. (1953). *Acta Cryst.* **6**, 260–268.
DARWIN, C. G. (1914). *Philos. Mag.* **27**, 315–333; 675–690.
DARWIN, C. G. (1922). *Philos. Mag.* **43**, 800–829.
GÖTTLICHER, S. (1968). *Acta Cryst.* B**24**, 122–129.
GÖTTLICHER, S. & KIESELBACH, B. (1976). *Acta Cryst.* A**32**, 185–192.
GRANT, D. F., KILLEAN, R. C. G., LAWRENCE, J. L., SENOL, A. & SHARMA, V. C. (1972). Private communication.
HART, M. (1974). *Diffraction Studies of Real Atoms and Real Crystals*, Abstracts, pp. 86–87. Canberra: Australian Academy of Science.
HIRSCH, P. B. & RAMACHANDRAN, G. N. (1950). *Acta Cryst.* **3**, 187–194.
International Tables for X-ray Crystallography (1959). Vol. II, p. 313. Birmingham: Kynoch Press.
KATO, N. (1974). *X-ray Diffraction*, edited by L. V. AZAROFF, p. 336. New York: McGraw-Hill.
LAWRENCE, J. L. & MATHIESON, A. McL. (1977). *Acta Cryst.* A**33**, 388–293.
LEHMANN, M. S. & SCHNEIDER, J. R. (1977). *Acta Cryst.* A**33**, 789–800.
MACKENZIE, J. K. & MATHIESON, A. McL. (1979). *Acta Cryst.* A**35**, 45–50.
MATHIESON, A. McL. (1976). *Nature (London)*, **261**, 306–308; **262**, 236.
MATHIESON, A. McL. (1977*a*). *Acta Cryst.* A**33**, 133–136.
MATHIESON, A. McL. (1977*b*). *Acta Cryst.* A**33**, 610–617.
ROBINSON, B. W. (1933). *Proc. R. Soc. London Ser. A*, **142**, 422–447.
SAKISAKA, Y. (1927). *Jpn. J. Phys.* **4**, 171–181.
SCHNEIDER, J. R. (1974). *J. Appl. Cryst.* **7**, 547–554.
SCHNEIDER, J. R. (1976). *J. Appl. Cryst.* **9**, 394–402.
SCHNEIDER, J. R. (1977). *Acta Cryst.* A**33**, 235–243.
TULINSKY, A., WORTHINGTON, C. R. & PIGNATORO, E. (1959). *Acta Cryst.* **12**, 623–626.
USA NATIONAL RESEARCH COUNCIL (1976). *Status and Future Potential of Crystallography*, p. 22. Washington, DC: National Academy of Sciences.
WEISS, R. J. (1966). *X-ray Determination of Electron Distributions*. Amsterdam: North-Holland.
WILKINS, S. W. (1978) *Proc. R. Soc. London Ser. A*. In the press.
WOOSTER, W. A. & MCDONALD, G. L. (1948). *Acta Cryst.* **1**, 49–54.

Acta Cryst. (1979). A**35**, 57–60

A Comparison of the Least-Squares and Maximum-Likelihood Estimators for Counts of Radiation Quanta which follow a Poisson Distribution

BY P. F. PRICE

Institut Max von Laue–Paul Langevin, BP 156*X, Grenoble, France*
and School of Physics, University of Melbourne, Parkville, Victoria 3052, *Australia**

(*Received* 21 *October* 1977; *accepted* 25 *July* 1978)

Abstract

The weighted least-squares method cannot correctly be used when measurements have errors given by counting statistics. The usual procedure results in bias in the values and errors in the calculated variances of the parameters. The maximum-likelihood method requires only a minor change in the least-squares equations and is generally thought to have more desirable properties for its estimates.

Introduction

Suppose a functional relationship exists between two measurable quantities y_0 and x given by

$$y_0 = f(\theta_0, x), \tag{1}$$

where the θ_0 are the true, but unknown, parameters, and suppose we have a list of measurements $y(x)$ of $y_0(x)$ at various values of the independent variable x. We assume the x's to be measured precisely but the y's to be imprecise.

$$\mathbf{y} = \mathbf{y}_0 + \boldsymbol{\varepsilon},\dagger \tag{2}$$

* Present address.

† Vector and matrix notation used throughout.

0567-7394/79/010057-04$01.00

where $\boldsymbol{\varepsilon}$ are the errors in the measurements. If we know the (proportional) variance matrix, $\mathbf{V}$, of the measurements,

$$\mathbf{V} = E[\{\mathbf{y} - E(\mathbf{y})\}\ \{\mathbf{y} - E(\mathbf{y})\}^t], \tag{3}$$

where $E[u]$ refers to the expectation value of u, and the superscript t refers to the transpose of a matrix, then we can evaluate the (weighted) least-squares (LS) estimates, $\hat{\boldsymbol{\theta}}_{LS}$, of the parameters $\boldsymbol{\theta}_0$. These are obtained by minimizing

$$S = (\mathbf{y} - \hat{\mathbf{y}})^t\ \mathbf{W}(\mathbf{y} - \hat{\mathbf{y}}) \tag{4}$$

with $\mathbf{W} = \mathbf{V}^{-1}$ and $\hat{\mathbf{y}}(\mathbf{x}) = f(\hat{\boldsymbol{\theta}},\mathbf{x})$.

If $f(\boldsymbol{\theta},x)$ is linear in $\boldsymbol{\theta}$

$$\hat{\mathbf{y}} = \mathbf{T}\hat{\boldsymbol{\theta}}, \tag{5}$$

with $\mathbf{T}$ the design matrix

$$\mathbf{T}_{kl} = \partial f(\boldsymbol{\theta},x_k)/\partial\theta_l, \tag{6}$$

then the $\hat{\boldsymbol{\theta}}_{LS}$ are given by

$$\hat{\boldsymbol{\theta}}_{LS} = (\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1}\ \mathbf{T}^t\mathbf{W}\mathbf{y}. \tag{7}$$

The $\hat{\boldsymbol{\theta}}_{LS}$ have some desirable properties as estimates of $\boldsymbol{\theta}_0$ (Hamilton, 1964; Bard, 1974). These properties only hold exactly if $f(\boldsymbol{\theta},x)$ is linear in $\boldsymbol{\theta}$ and if the errors have zero mean, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$. {See Price (1979) for a discussion of parameter bias when f is non-linear or when $E[\boldsymbol{\varepsilon}] \neq \mathbf{0}$.}

In this case $\mathbf{y}_0 = \mathbf{T}\boldsymbol{\theta}_0$ and the substitution of (2) in (7) gives

$$\hat{\boldsymbol{\theta}}_{LS} = \boldsymbol{\theta}_0 + (\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1}\ \mathbf{T}^t\mathbf{W}\boldsymbol{\varepsilon}. \tag{8}$$

Now if $\mathbf{W}$ (and $\mathbf{T}$) is a constant matrix (*i.e.* would not change if the experiment were to be repeated indefinitely), as it is if $\mathbf{W} = \mathbf{V}^{-1}$, then

$$E[\hat{\boldsymbol{\theta}}_{LS}] = \boldsymbol{\theta}_0, \tag{9}$$

i.e. $\hat{\boldsymbol{\theta}}_{LS}$ is an unbiased estimate of $\boldsymbol{\theta}_0$.

Also the variance of the estimates $\hat{\boldsymbol{\theta}}_{LS}$ is given by

$$\text{var}\,(\hat{\boldsymbol{\theta}}_{LS}) = E[\{\hat{\boldsymbol{\theta}}_{LS} - E[\hat{\boldsymbol{\theta}}_{LS}]\}\ \{\hat{\boldsymbol{\theta}}_{LS} - E[\hat{\boldsymbol{\theta}}_{LS}]\}^t]$$

$$\text{var}\,(\hat{\boldsymbol{\theta}}_{LS}) = (\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1}\ \mathbf{T}^t\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{T}(\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1} \tag{10}$$

(assuming the constancy of $\mathbf{W}$ and $\mathbf{T}$), and if $\mathbf{V} = \sigma^2\,\mathbf{W}^{-1}$

$$\text{var}\,(\hat{\boldsymbol{\theta}}_{LS}) = \sigma^2(\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1} = (\mathbf{T}^t\mathbf{V}^{-1}\mathbf{T})^{-1}. \tag{11}$$

These are well known results and it can be shown that when $\mathbf{W}$ is proportional to $\mathbf{V}^{-1}$ (Hamilton, 1964; Bard, 1974) the $\hat{\boldsymbol{\theta}}_{LS}$ have minimum variance amongst all linear, unbiased estimates of $\boldsymbol{\theta}_0$.

If $\mathbf{W}$ is a constant matrix but not proportional to $\mathbf{V}^{-1}$, (8) and (9) still result in $\hat{\boldsymbol{\theta}}$ being an unbiased estimate of $\boldsymbol{\theta}_0$, but the variance of $\hat{\boldsymbol{\theta}}$ is given by (10) rather than (11). It can be shown (Tukey, 1975) that if

$$A \leq \frac{\mathbf{W}_{ij}}{\mathbf{V}_{ij}^{-1}} \leq rA \tag{12}$$

then the variance of the least-squares estimate of any linear combination of the parameters is

$$\leq \frac{(r+1)^2}{4r} \tag{13}$$

times as large as if the proper weights were used. This is small comfort however, as unless $\mathbf{V}$ is known we cannot calculate the true variance (10) or the smaller variance (11), and we do not know the relationship between the true variance and the inverse of the normal equations matrix, $(\mathbf{T}^t\mathbf{W}\mathbf{T})^{-1}$.

Quantum counting statistics

We now consider the application of the method of least squares to counts of radiation quanta, such as X-ray photons or neutrons. Observed count rates are theoretically distributed (neglecting instrumental effects) as a Poisson distribution

$$p(y) = [\exp(-y_0)\ y_0^y]/y!, \tag{14}$$

where $p(y)$ is the probability distribution function for y. As is well known this distribution has all moments equal to y_0. Thus a least-squares refinement on the measurements $\mathbf{y}$ requires knowledge of the variance of each y and hence of each y_0. Strictly speaking then, we cannot perform least squares when the errors in the measurements obey counting statistics.

However, for comparison with the maximum-likelihood (ML) equations derived below, we write down the LS equations assuming we know the variances $\mathbf{y}_0$. The $\hat{\boldsymbol{\theta}}_{LS}$ are the solutions to $\partial S/\partial\theta_i = 0$ for all i, *i.e.*

$$\sum_k \frac{1}{y_{0k}}\left[(y_k - \hat{y}_{LS,k})\frac{\partial f_k}{\partial\theta_i}\right]_{\hat{\boldsymbol{\theta}}_{LS}} = 0 \quad \text{for all } i. \tag{15}$$

Here f_k means $f(\boldsymbol{\theta},x_k)$ and we have dropped the matrix notation. Now if $f(\boldsymbol{\theta},x)$ is non-linear in $\boldsymbol{\theta}$ we would probably approximate the variance matrix of $\hat{\boldsymbol{\theta}}_{LS}$ by the inverse of either the Gauss or Newton curvature (normal equations) matrices, *i.e.*

$$\text{var}\,(\hat{\boldsymbol{\theta}}_{LS}) \simeq \boldsymbol{\alpha}^{-1} \tag{16}$$

with

$$\boldsymbol{\alpha}_{ij}^G = \sum_k \frac{1}{y_{0,k}}\left[\frac{\partial f_k}{\partial\theta_i}\frac{\partial f_k}{\partial\theta_j}\right]_{\hat{\boldsymbol{\theta}}_{LS}} \tag{17}$$

or

$$\boldsymbol{\alpha}_{ij}^N = \sum_k \frac{1}{y_{0,k}}\left[\frac{\partial f_k}{\partial\theta_i}\cdot\frac{\partial f_k}{\partial\theta_j} - (y_k - \hat{y}_{LS,k})\frac{\partial^2 f_k}{\partial\theta_i\,\partial\theta_j}\right]_{\hat{\boldsymbol{\theta}}_{LS}}. \tag{18}$$

If f is linear then $\alpha^G = \alpha^N = \mathbf{T}^t\mathbf{W}\mathbf{T}$ and (16) is an equality.

It is not unusual for measurements subject to counting-statistics errors to be analysed by least squares where the variances of the counts are taken to be the *observed count*, y, rather than the expectation value of this, y_0. This results in biased parameter estimates [$\mathbf{W}$ in (8) is not a constant matrix]. This has recently been noted again, by Wilson (1976). The bias is in such a direction that the fitted curve y is 'pushed low', since counts which are lower than their expectation value have higher weight than counts which are higher than their mean. In fact if we make enough measurements (*i.e.* as our sample size n goes to ∞) we will certainly obtain a number of measurements of a zero count, which will force the fitted curve to be identically zero. Bevington (1969) claims that the area under the curve is underestimated by an amount approximately equal to the residual, S.

Maximum-likelihood estimates of θ_0

If the functional form of the probability distribution function $p(y)$ is known,

$$p(y) = p(y,y_0,\varphi_0), \tag{19}$$

where y_0 and possibly some additional parameters φ_0 (*e.g.* the variance) are unknown, and if we have the expression (1) for y_0, then the ML estimates for θ_0 and φ_0 can be computed (Bard, 1974; Kendall & Stuart, 1967). If the measurements, y_k, are independent we maximize the likelihood function

$$L(\mathbf{y},\boldsymbol{\theta},\boldsymbol{\varphi}) = \prod_k p[y_k, f(\boldsymbol{\theta},x_k),\boldsymbol{\varphi}] \tag{20}$$

resulting in estimates $\hat{\boldsymbol{\theta}}_{ML}$, $\hat{\boldsymbol{\varphi}}_{ML}$ for the parameters.

ML estimates $\hat{\boldsymbol{\theta}}_{ML}$ also have some desirable properties (Kendall & Stuart, 1967; Bard, 1974) which generally seem to be more favoured by the statisticians than those of the LS estimates. In general the $\hat{\boldsymbol{\theta}}_{ML}$ are non-linear estimates, which allows their variance to be less than that of the $\hat{\boldsymbol{\theta}}_{LS}$. They are *asymptotically* (*i.e.* as the sample size, or number of measurements, n, goes to infinity) unbiased; for finite sample sizes their bias is of the order n^{-1}. Theoretical lower bounds to the variance of an estimator (the minimum variance bound, MVB) can be derived and the ML estimates asymptotically achieve this MVB.

If $p(y_k)$ are normal distributions with known variances σ_k^2 then it is easy to show that $\hat{\boldsymbol{\theta}}_{ML} = \hat{\boldsymbol{\theta}}_{LS}$. When the σ_k^2 are totally unknown the method cannot be used as there are more unknown parameters (the $\boldsymbol{\theta}_0$ and the σ_k's) than observations.

When $p(y_k)$ is a Poisson distribution as given by (14) the ML equations become

$$\log L(\mathbf{y},\boldsymbol{\theta}) = \sum_k [-f_k(\boldsymbol{\theta}) + y_k \log f_k(\boldsymbol{\theta}) - \log y_k!]; \tag{21}$$

equating $\partial(\log L)/\partial\boldsymbol{\theta} = \mathbf{0}$, we find that the $\hat{\boldsymbol{\theta}}_{ML}$ are solutions to

$$\sum_k \frac{1}{\hat{y}_{ML,k}} \left[(y_k - \hat{y}_{ML,k}) \frac{\partial f_k}{\partial \theta_i}\right]_{\hat{\boldsymbol{\theta}}_{ML}} = 0 \quad \text{for all } i. \tag{22}$$

The variance matrix $\mathbf{D}_{ML} = \mathrm{var}(\hat{\boldsymbol{\theta}}_{ML})$ of these estimates $\hat{\boldsymbol{\theta}}_{ML}$ cannot be simply calculated for finite sample sizes. However, it is known (Kendall & Stuart, 1967) that asymptotically it is given by

$$(\mathbf{D}_{ML}^{\text{asympt.}})_{ij}^{-1} = E\left[-\frac{\partial^2 \log L}{\partial\theta_i\,\partial\theta_j}\right], \tag{23}$$

which is easily evaluated for the Poisson distribution as

$$(\mathbf{D}_{ML}^{\text{asympt.}})_{ij}^{-1} = \sum_k \frac{1}{y_{0,k}} \left[\frac{\partial f_k}{\partial\theta_i}\frac{\partial f_k}{\partial\theta_j}\right]_{\boldsymbol{\theta}_0}. \tag{24}$$

Expression (22) is seen to be obtained from (15) by the replacement of the true variance $y_{0,k}$ by the calculated variance $\hat{y}_k$. Since y_0 is unknown, expressions (17), (18) and (24) are uncalculable. Inspection of these indicates, however, that for large sample sizes the variance of the $\hat{\boldsymbol{\theta}}_{ML}$ should be approximately given by the inverse of the matrices $\boldsymbol{\alpha}^{GML}$ or $\boldsymbol{\alpha}^{NML}$ where

$$\alpha_{ij}^{GML} = \sum_k \frac{1}{\hat{y}_{ML,k}} \left[\frac{\partial f_k}{\partial\theta_i}\frac{\partial f_k}{\partial\theta_j}\right]_{\hat{\boldsymbol{\theta}}_{ML}} \tag{17'}$$

and

$$\alpha_{ij}^{NML} = \sum_k \frac{1}{\hat{y}_{ML,k}} \left[\frac{\partial f_k}{\partial\theta_i}\frac{\partial f_k}{\partial\theta_j} - (y_k - \hat{y}_{ML,k})\frac{\partial^2 f_k}{\partial\theta_i\,\partial\theta_j}\right]_{\hat{\boldsymbol{\theta}}_{ML}}. \tag{18'}$$

These expressions are obtained from (17) and (18) by replacing $y_{0,k}$ by $\hat{y}_{ML,k}$ in the weights, and by evaluating the fitted values $\hat{\mathbf{y}}$ and their derivatives at the $\hat{\boldsymbol{\theta}}_{ML}$ rather than at the $\hat{\boldsymbol{\theta}}_{LS}$.

In general the presence of $\hat{\mathbf{y}}$ in the weights results in the ML procedure being non-linear, even when $f(\boldsymbol{\theta},x)$ is linear in $\boldsymbol{\theta}$. A simplification occurs in the case when we are seeking to estimate a constant background, when

$$y_0 = f(\theta_0,x) = \theta_0.$$

In either method the weights are constant and thus (15) and (22) result in

$$\hat{\theta}_{\mathrm{LS}} = \hat{\theta}_{\mathrm{ML}} = \left(\sum_k y_k\right)/n$$

with variance

$$\mathrm{var}\,(\hat{\theta}) = \theta^0/n$$

and estimated variance given by (17′) or (18′) as

$$\alpha^{-1} = \hat{\theta}/n.$$

Conclusions

The weighted LS estimator, which possesses some desirable properties, cannot properly be used when the errors in the measurements follow a Poisson distribution. An attempt to do so by replacing the true variances $y_{0,k}$ by y_k (*i.e.* the 'true' mean count rate by the observed count rate) results in a bias which will primarily affect a scale factor. The ML estimator can be used and requires only a minor change (using $\hat{y}_k$, the calculated count rate, as the variance) in the LS equations. ML estimators also have some desirable properties which generally seem to be preferred to those of LS.

Wilson (1976) suggests using $\frac{1}{3}(y_k + 2\hat{y}_k)$ as the variance of y_k. We recommend the use of the ML method with its well studied properties.

The author acknowledges helpful discussions with M. S. Lehmann.

Additional note: Comments made to the author since the submission of this paper suggest the worth of clarifying exactly when these results can be used in crystallographic refinements.

We have assumed (equations 2 and 14) that the observed variables have errors given by the Poisson distribution. Such is the case when the observed variables are 'raw' count rates, and the model allows for all sources of counts received by the detector, including background (see Sabine & Clarke, 1977). There are a number of other refinement situations (*e.g.* a background refinement) where the method can be used.

Structure refinements from single-crystal data usually have as observed variables net intensities (gross intensities less background) at best and thus do not have errors given by the Poisson distribution. Wilson (1978) has given the error distribution of a variable which is the difference between two Poisson-distributed variables (see also Wilson, 1979). We have not proved in this paper that the ML estimation procedure for this case simply requires the exchange of calculated intensities for observed intensities in the LS weighting schemes. Nevertheless, we have pointed out (see *Introduction*) that the use of observed intensities in LS weighting schemes results in a non-constant weight matrix and thence to parameter bias. Similarly we have established that we do not know (from a single observed intensity) an unbiased estimate of the reciprocal variance, and thus we cannot properly use weighted least squares. The relative complexity of Wilson's expressions (which themselves may not adequately describe the true error distribution of our intensities) suggests that a formal answer to this problem may be both difficult to obtain and complex to implement. In lieu of the formal answer being obtained the author suggests a cautious use of the answer suggested by this paper – *i.e.* the replacement of calculated for observed variables in weighting schemes.

The probable insensitivity of the result to small changes in calculated intensities suggests that computing time could be reduced by only calculating the weights (from the calculated intensities) once or twice near the beginning of refinement and once or twice near the end.

References

BARD, Y. (1974). *Nonlinear Parameter Estimation*. New York: Academic Press.

BEVINGTON, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill.

HAMILTON, W. C. (1964). *Statistics in Physical Science*. New York: Ronald.

KENDALL, M. G. & STUART, A. (1967). *The Advanced Theory of Statistics*, Vol. II, 2nd ed. London: Griffin.

PRICE, P. F. (1979). *Acta Cryst.* A. Submitted.

SABINE, T. M. & CLARKE, P. J. (1977). *J. Appl. Cryst.* **10**, 277–280.

TUKEY, J. W. (1975). In *Applied Statistics*, edited by R. P. GUPTA. Amsterdam: North-Holland.

WILSON, A. J. C. (1976). *Acta Cryst.* A**32**, 994–996.

WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 474–475.

WILSON, A. J. C. (1979). *Acta Cryst.* A. Submitted.