

Maximum-likelihood fitting of data dominated by Poisson statistical uncertainties

M. R. Stoneking and D. J. Den Hartog

Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706

(Presented on 15 May 1996)

The fitting of data by χ^2 minimization is valid only when the uncertainties in the data are normally distributed. When analyzing spectroscopic or particle counting data at very low signal level (e.g., a Thomson scattering diagnostic), the uncertainties are distributed with a Poisson distribution. We have developed a maximum-likelihood method for fitting data that correctly treats the Poisson statistical character of the uncertainties. This method maximizes the total probability that the observed data are drawn from the assumed fit function using the Poisson probability function to determine the probability for each data point. The algorithm also returns uncertainty estimates for the fit parameters. We compare this method with a χ^2 -minimization routine applied to both simulated and real Thomson scattering data. Differences in the returned fits are greater at low signal level (less than ~ 10 counts per measurement). The maximum-likelihood method is found to be more accurate and robust, returning a narrower distribution of values for the fit parameters with fewer outliers. © 1997 American Institute of Physics. [S0034-6748(97)73401-4]

I. INTRODUCTION

When experimental data are fit to a function that is known to represent the phenomenon under investigation, the standard strategy is to minimize χ^2 with respect to changes in the parameters of the “fit” function. Such a strategy relies on the assumption that the uncertainties in the data are normally distributed. For non-normal uncertainty distributions such as the Poisson distribution, minimizing χ^2 does not maximize the likelihood that the fitted parameters reflect the data. The maximum-likelihood method is described in the standard texts.^{1,2} Direct application of the method using Poisson distributed uncertainties was made previously for a specific x-ray photon counting application.³ However, the general treatment in the case of Poisson distributed uncertainties for an arbitrary fit function [Eq. (2)] and its application to Thomson scattering data from (nonrelativistic) high temperature plasma [Eq. (4)] is new. Published work dealing with Thomson scattering data^{4,5} have pointed out the need for correctly evaluating the measurement uncertainty in the case of low signal photon counting, but those uncertainties were then used in least-squares (or χ^2 minimization) fitting routines, rather than the method described in this article. Our motivation to develop this method arose from the need to accurately and correctly fit Thomson scattering data with very low signal levels (as low as ~ 10 detected photons per channel).

II. DESCRIPTION OF THE MAXIMUM-LIKELIHOOD METHOD

In order to fit data, the function that represents the phenomenon being investigated must be known. Let the function be written as $u(x, a_1, \dots, a_M)$, where x is the independent variable and a_1, \dots, a_M are adjustable parameters to be determined by the fit. For a given set of parameters, one can calculate the probability of measuring the data points, $y_i(x_i)$. The maximum-likelihood fitting method is founded on the assumption that the probability of measuring a set of data

points, y_1, \dots, y_N , given a set of adjustable parameters, a_1, \dots, a_M , is a measure of the likelihood of the parameters given the data, and that maximizing the former also maximizes the latter.² For counting measurements, the probability of measuring a single value y_i is given by the Poisson probability function, $P_i = u^{y_i} \exp(-u)/y_i!$. The total probability of measuring the N values of y_i is the product of the individual probabilities,

$$P = \prod_{i=1}^N \frac{\exp(-u) u^{y_i}}{y_i!}. \quad (1)$$

If the measurement uncertainties were normally distributed the product would be over the Gaussian probabilities with standard deviation σ given by the measurement uncertainties. Maximizing the product of Gaussian probabilities is equivalent to minimizing the χ^2 merit function; χ^2 minimization is therefore an application of the maximum-likelihood method for the case of normally distributed uncertainties.

The total Poisson probability is maximized by setting the partial derivatives (of the natural logarithm) of Eq. (1) with respect to the fit parameters to zero, $\partial/\partial a_j (\ln P) = 0$. The result is a set of M equations the roots of which yield the fit parameters that maximize the Poisson probability,

$$\frac{\partial}{\partial a_j} (\ln P) = \sum_{i=1}^N \frac{\partial u(x_i)}{\partial a_j} \left(1 - \frac{y_i}{u(x_i)} \right) = 0. \quad (2)$$

We now apply this method for the fit function applicable to Thomson scattering data reduction. For this diagnostic method, monochromatic radiation is Thomson scattered off the electrons in a high-temperature plasma.⁶ The scattered spectrum of radiation has a Doppler spread determined by the electron velocity distribution. Scattered light is collected in wavelength bands close to the wavelength of the incident light, and fit to the spectrum expected from a Maxwellian electron distribution. Relativistic effects can be neglected for temperatures less than ~ 1 keV,⁶ and in that case the fit function is a Gaussian with two parameters,

$$u(x_i) = a_1 \exp\left(\frac{-x_i^2}{2a_2^2}\right) \Delta x_i, \quad (3)$$

where x is the wavelength shift ($\Delta\lambda$) and Δx is the channel width (in wavelength units). The Gaussian width, a_2 , is related to the electron temperature by $a_2^2 = 2\lambda_0^2 T_e / mc^2$ (for 90° scattering). The electron density is proportional to the area under the Gaussian, $a_1 a_2 \propto n_e$. It is important to insert the channel width (and any channel to channel calibration factors) into the fitting function [Eq. (3)] rather than modifying the raw data. If the raw photon count data are modified, the Poisson statistical character of the errors is distorted.

Applying the maximum-likelihood method to the fit function in Eq. (3) yields the equations

$$\sum_{i=1}^N (u(x_i) - y_i) = 0, \quad \sum_{i=1}^N x_i^2 (u(x_i) - y_i) = 0. \quad (4)$$

We employ the Newton–Raphson method to find the roots of Eq. (4). Since there are only two fit parameters, the Jacobian matrix is 2×2 , can be immediately inverted. For more than two fit parameters, a numerical matrix inversion is required.

An estimate of the uncertainty in the fit parameters is made by direct application of the error propagation equation for a set of N independent measurements,

$$\sigma_a^2 = \sum_{i=1}^N \sigma_{y_i}^2 \left(\frac{\partial a}{\partial y_i} \right)^2, \quad (5)$$

where $\sigma_{y_i} = \sqrt{y_i}$ are the Poisson measurement uncertainties for each data point. The derivative in Eq. (5) is evaluated using a five-point numerical derivative. The fit parameters are determined for four “perturbed” values of y in addition to the best-fit value in order to evaluate the derivative in Eq. (5).

III. TESTING THE FITTING ALGORITHM WITH SIMULATED DATA

Simulated Thomson scattering data were generated for five wavelength channels by drawing random deviates that were Poisson distributed about a mean given by Eq. (3). The values of the fit parameters returned by the maximum-likelihood (Poisson) code are compared with the simulation input values and with the fits returned by a χ^2 -minimizing code. As expected, differences between the two codes, and between the code and the simulation inputs are smaller at high signal intensity (larger simulation values for a_1). Figure 1(a) shows the difference (%) in the returned values for the electron temperature from the Poisson and χ^2 codes. The range of the difference is less than 5% at values of a_1 greater than 20. The difference range between the fit parameters and the actual value (350 eV) is larger, about 20% at high signal [Fig. 1(b)]. This reflects the rapid convergence of the Poisson probability distribution toward the normal probability distribution with increased mean values. The less rapid convergence of the fits toward the simulation is understood, since the uncertainty (i.e., random fluctuation) in each data point is significant beyond the signal level where the two probability distributions converge.

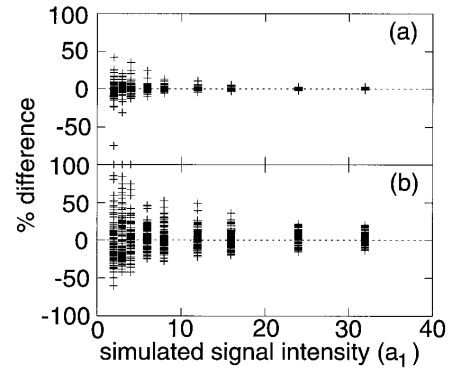


FIG. 1. (a) The difference between the temperature determined by the maximum-likelihood method and χ^2 -minimization method for simulated data, (b) the difference between maximum-likelihood method and the input value for temperature (350 eV).

The true test of the maximum-likelihood method with Poisson uncertainties comes at low signal intensities where the difference between the Poisson distribution and the normal distribution is greatest. Figure 2 is a scatter plot of values for the electron temperature returned by the two codes plotted against each other for approximately 3000 fits to simulated data. The amplitude parameter a_1 was held fixed at a value of 2.0 that simulates the low end of our experimentally realized signal range (about 50 total counts in 5 channels), and the electron temperature was fixed at 350 eV. All of the scatter is a result of the random (Poisson) fluctuations; no additional noise sources are simulated. The scatter in the temperature values are not independent but tend to lie along the “line of agreement.” However, there is greater scatter vertically than horizontally, indicating that the distribution of returned values for the maximum-likelihood code is narrower with fewer outliers than the distribution from the χ^2 -minimizing code.

The distribution of returned temperature values from the Poisson code is asymmetric having a wing on the high-temperature side (Fig. 3) with a mean value 11% higher than the simulated temperature. The χ^2 code returns a mean temperature that is 17% higher than the simulated value with a similarly shaped distribution. At higher signal levels, the distributions become symmetric and the means approach the actual value (for both codes).

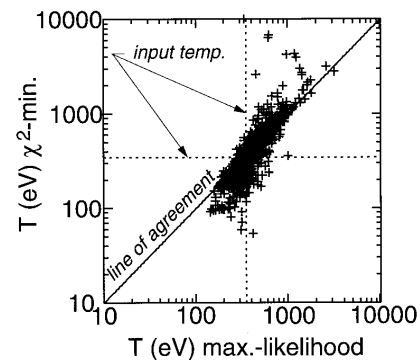


FIG. 2. Fit values for temperature from maximum-likelihood and χ^2 -minimization methods plotted against each other (simulation input value is 350 eV).

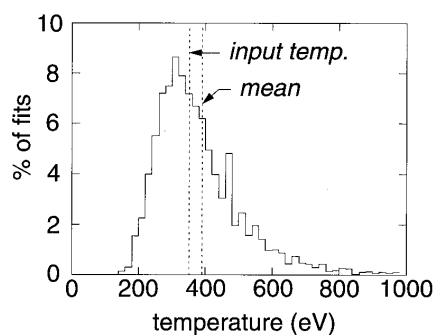


FIG. 3. The distribution of values for the temperature returned by the maximum-likelihood fitting routine is asymmetric.

The other parameter of interest in Thomson scattering data analysis is the area under the Gaussian spectrum (area = $a_1 a_2 \sqrt{2\pi}$) which is proportional to electron density. For the same data set shown in Fig. 2, the distribution of returned values of area is symmetric, with mean values close to that of the simulation. The maximum-likelihood fit value for the area is larger than the simulated value by 1.5%, and the χ^2 value is 5% low.

IV. TESTING THE FITTING ALGORITHM WITH REAL DATA

Thomson scattering data was acquired in the Madison Symmetric Torus (MST), a reversed field pinch plasma.^{7,8} Three plasma density values (~ 30 spectra for each case) were obtained to get Thomson scattering signal variation of $\sim 3\times$. The plasma current was held fixed at ~ 340 kA.

The trends observed in the simulations are also seen in the fitting of real data. At low plasma density (hence low Thomson scattering signal), the maximum-likelihood method

is more robust returning a narrower distribution of values for the electron temperature with fewer outliers. Since the parent distribution is unknown in the experiment and there may be significant discharge to discharge fluctuations in the parameters of the parent distribution, it is more difficult to assess the differences between the fitting routines with real data than it is with simulated data, hence our reliance on simulations for the critical tests of the fitting algorithms.

To estimate the values for the parent distribution at each density, we summed the Thomson scattering data for all the discharges and fit the accumulated spectrum. The fit parameters from the two fitting routines for this “ensemble” spectrum are in close agreement. Table I is a summary of the comparison of the two fitting algorithms using real data. Row 3 of Table I shows the temperature for the ensembled spectra. For comparison with the simulation results of the previous section, the value of the fit parameter a_1 (divided by the number of shots in the ensemble), for the low density data set is 1.4 (see Fig. 1). For the moderate and high density cases, $a_1=3.4$ and 4.6, respectively. Row 4 gives the mean (standard deviation) of the approximately 30 temperature values returned by the maximum-likelihood method. Row 5 presents the mean (standard deviation) for fits returned by the χ^2 routine. The number of detected photons (quantum efficiency is about 4%) in each channel are shown in rows 6 through 10. At these signal levels counting statistics dominate the measurement uncertainties for single shot spectra. At the lowest density, the maximum-likelihood fitting routine returns a distribution of temperature values with mean closer to the ensemble value and with a smaller standard deviation than the χ^2 routine. At high density, the two codes return temperature values with very similar mean and standard deviation. The trends agree with the simulation results and confirm the conclusion that for low signal level the maximum-

TABLE I. Summary of the comparison of the maximum-likelihood and χ^2 -minimization methods using real Thomson scattering data from MST.

Density (10^{19} m^{-3})	0.5	1.3	1.8
# of discharges	31	27	41
Ensemble fit temp. (eV)	379	289	210
Max.-likelihood fit mean (eV) (std. dev.)	405 (164)	307 (121)	223 (78)
χ^2 -minimization fit mean (eV) (std. dev.)	428 (205)	310 (137)	218 (76)
Avg. counts in channel #1 (background)	4.6 (1.4)	10.6 (2.6)	13.7 (2.9)
Avg. counts in channel #2 (background)	12.0 (2.6)	27.7 (4.6)	33.2 (6.5)
Avg. counts in channel #3 (background)	9.6 (1.4)	19.5 (3.1)	22.6 (4.3)
Avg. counts in channel #4 (background)	5.8 (1.7)	11.8 (4.1)	11.1 (6.5)
Avg. counts in channel #5 (background)	5.5 (2.6)	10.4 (5.3)	13.2 (9.6)

likelihood fitting routine is more robust, returning fewer outliers, a narrower distribution of fit parameter values and a mean closer to the parent distribution value. At higher signal level, the two routines return similar fit parameters and are equally robust.

To rigorously include the background plasma light present in the experiment, the fit function [Eq. (3)] should be modified to include background counts as

$$u(x_i) = a_1 \exp\left(\frac{-x_i^2}{2a_2^2}\right) \Delta x_i + a_3 f_b(x_i) \Delta x_i, \quad (6)$$

where the shape of the background spectrum, $f_b(x_i)$, is determined from the accumulated background counts over many similar discharges. The fit parameter a_3 , is determined by maximizing the Poisson probability given the background counts from a single shot. Using Eq. (2), $a_3 = \Sigma y_i^b / \Sigma f_b(x_i) \Delta x_i$, where y_i^b are the measured background counts in each channel (spectroscopic surveys indicate that background light is dominated by impurity line radiation, rather than continuum bremsstrahlung, and constitutes about 25–30% of the measured counts as shown in Table I, rows 6–10). This modification to the method described in section II was attempted using both simulated and real data. With simulated data, the results were similar to the results obtained without including background light. Using real data, we found that the inclusion of background light in the manner of Eq. (6) produced less reliable fits. This presumably results from shot to shot changes in the shape of the background spectrum. Instead of implementing Eq. (6) to deal with background light, it was found to be more satisfactory

(though not rigorously correct) to subtract the background counts before implementing the fitting method as described in Sec. II. The relatively small background light level compared with the signal minimizes the distortion to the Poisson statistical character of the data. The numbers in Table I were obtained by subtracting background counts before applying the fitting routine. Background counts are sampled 125 ns before and 125 ns after the laser pulse in each of the five wavelength channels, so the background is accurately measured if fluctuation power is dominated by frequencies below ~ 4 MHz.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of David Ruppert to the development of this analysis technique. This work was supported by the U.S. Department of Energy.

¹P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physics Sciences*, 2nd ed. (McGraw-Hill, New York, 1992).

²W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN, The Art of Scientific Computing*, 2nd ed. (Cambridge University, Cambridge, 1992).

³J. H. M. M. Schmitt and T. Maccacaro, *Astrophys. J.* **310**, 334 (1986).

⁴G. W. Hart, F. M. Levinton, and D. H. McNeill, *Rev. Sci. Instrum.* **57**, 2218 (1986).

⁵M. Mattioli and R. Papoular, *Plasma Phys.* **17**, 165 (1975).

⁶J. Sheffield, *Plasma Scattering of Electromagnetic Radiation* (Academic, New York, 1975).

⁷R. N. Dexter, D. W. Kerst, T. W. Lovell, S. C. Prager, and J. C. Spratt, *Fusion Tech.* **19**, 131 (1991).

⁸D. J. Den Hartog and M. Cekic, *Meas. Sci. Technol.* **5**, 1115 (1994).