How Bad is Good?

A Critical Look at the Fitting of Reflectivity Models using the Reduced Chi-Square Statistic

Charlie Laub and Tonya L. Kuhl
Department of Chemical Engineering and Materials Science
University of California, Davis

Abstract:

The origins of the chi-square statistic that is typically employed as a goodness of fit metric for iterative fitting of reflectivity data will be reviewed. The basis for the rule of thumb that states that a "good fit" is achieved with a reduced chi-square value of 1.0 will be derived and it will be shown that, even for modestly larger errors, the "fit" is not significant in a statistical sense.

We propose a new rule of thumb that states that the reduced chi-square error must be less than 1.5 to achieve a minimum level of significance, e.g. at the 0.1 percent level.

The Chi-Square Statistic

$$\chi^2 = \sum_N \left\{ \frac{1}{\sigma_i^2} \left[y_i - f\left(x_i\right) \right]^2 \right\} \qquad 0 < \chi^2 < \infty$$

N = number of data points
$\sigma^2{}_i$ = "variance", related to the measurement error for y_i
y = independent variable, x = dependent variable
f = assumed relationship between x and y.
y_i = "observed mean"
$f(x_i)$ = "predicted mean"

- Used as a "figure of merit" when fitting a function, f, to data, $\{x_i, y_i\}$ to evaluate the "goodness of fit"

Reflectivity is not a direct technique

- Most reflectivity experiments capture only intensity information
- Moving from inverse space to real space is required in order to obtain information about the sample (e.g. the scattering length density profile)
  - This is commonly performed by iteratively fitting a model to the data
  - Iterative fitting attempts to minimize the chi-square error between the data set and the model-generated reflectivity
  - The model encompasses the SLD profile and mathematical formalism employed to calculate the reflectivity
- When fitting reflectivity data, it is assumed that when the difference between the data and the model approaches zero, the scattering length profile accurately represents the distribution of material inside the sample region (within the limits of non-uniqueness due to lack of phase info)

Statistical Concepts Relevant to Data Fitting

- The "Parent Distribution" – to be determined by taking data
  - has a mean, variance and form that are unknown!
  - Collecting data can be thought of as "taking samples" from the "Parent Distribution" that defines the relationship between the independent and dependent variables
  - The parent distribution is a "Probability Surface"
- The fitting function, f, describes the assumed functional relationship between the independent and dependent variables. f *predicts the mean* for each data point, μ_i.
- The resulting difference between the predicted mean f(x_i) and the observed mean, y_i, is the deviation, Δy_i.
- It is typical to assume that the deviations from the mean will follow some distribution, e.g. normal, poisson, etc.

The Origins of the Chi-Square Statistic

- If the deviations from the mean follow Gaussian statistics, the probability of making any one observation is given by:

$$P_G(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2 \right] \quad \text{where } x - \mu \;\rightarrow\; y_i - f(x_i)$$

- The total probability of obtaining a set of N measurements, $\{x_i, y_i\}$, is equal to the *product* of the probabilities for each data point:

$$P_{\{x,y\}} = \prod_N P_G = \left\{ \prod_N \frac{1}{\sigma\sqrt{2\pi}} \right\} * \left\{ \exp\left[-\tfrac{1}{2} \sum_{i=1}^{N} \left(\frac{y_i - \mathrm{f}(x_i)}{\sigma_i} \right)^2 \right] \right\}$$

- Maximizing the probability is equivalent to minimizing the sum in the exponential term of $P_{\{x,y\}}$, specifically the sum of the deviations, Δy.
- The chi-square statistic is <u>defined by</u> this sum:

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - \mathrm{f}(x_i)}{\sigma_i} \right)^2$$

What is the reduced chi-square error (χ^2/ν) and why should it be equal to 1.0 for a good fit?

- The method of least squares is built on the hypothesis that the optimum description of a set of data is one which minimizes the weighted sum of squares of deviations, Δy, between the data, y_i, and the fitting function f.

- The sum of squares of deviations is characterized by the "estimated variance of the fit", s^2, which is an estimate of the variance of the parent distribution, σ^2.
- The ratio of s^2/σ^2 can be estimated by χ^2/ν, where $\nu = N - p - 1$, N is the number of observations and p is the number of fitting parameters. χ^2/ν is called the <u>reduced chi-square statistic</u>.

- If the fitting function accurately predicts the means of the parent distribution, then the estimated variance, s^2, should agree well with the variance of the parent distribution, σ^2, and their ratio should be close to one.
- This explains the origin of the rule of thumb for chi-square fitting that states that a "good fit" is achieved when the reduced chi-square equals one.

Assigning Significance to the reduced chi-square statistic

- The reduced chi-square statistic is just a 2-parameter distribution!
- A more useful distribution is obtained if we integrate it from x= χ^2/ν to x=infinity and tabulate the result as a function of χ^2/ν and ν. This new distribution (here called $Q(\chi^2/\nu,\nu)$) describes the probability that χ^2/ν for a set of deviations obtained by <u>*randomly sampling N observations*</u> from a normal distribution would <u>*exceed*</u> the value for χ^2/ν that was obtained by the deviations obtained by fitting f to our N data points, $\{x_i,y_i\}$.
- $Q(\chi^2/\nu,\nu)$ is a function that can be calculated numerically and has been tabulated (consult appendix `C-4` of Bevington, for example).
- $Q(\chi^2/\nu,\nu)$ is a probability distribution, therefore: $0 < Q < 1$.

How to statistically evaluate the goodness-of-fit.

- It's simple!
  - If Q is approximately 0.5, the deviations between f and {y} agree with what you would expect statistically, meaning that the variance in the data is approximated well by the variance of the fit because, at Q=0.5, the reduced chi square value is near 1 which means that s^2 and σ^2 are approximately equal.

  - If Q is "small" then the fit is "poor" because a set of random samples from the parent distribution have a higher probability of giving rise to a reduced chi-square value that is equal to or less than the value obtained by fitting of f to y. E.g. if Q=0.01, there is a 99% probability that random deviations explain the deviations in the data better than the predictor of the mean, f.

  - If Q is very close to 1 because χ^2/ν is very near zero, then most likely the estimate of the uncertainties in the data, $\{\sigma_i\}$, is too large.

Q: The probability that the reduced chi-square value obtained by randomly sampling N observations from a Gaussian distribution <u>is larger than</u> the reduced chi-square value obtained via fitting a function to a data set having ν degrees of freedom (ν=100 is typical).

ν	0.9	0.5	0.1	0.01	0.001
10	0.487	**0.934**	1.599	2.321	2.959
20	0.622	**0.967**	1.421	1.878	2.266
30	0.687	**0.978**	1.342	1.696	1.990
40	0.726	**0.983**	1.295	1.592	1.835
50	0.754	**0.987**	1.263	1.523	1.733
60	0.774	**0.989**	1.240	1.473	1.660
70	0.790	**0.990**	1.222	1.435	1.605
80	0.803	**0.992**	1.207	1.404	1.560
90	0.814	**0.993**	1.195	1.379	1.525
100	0.824	**0.993**	1.185	1.358	1.494
140	0.850	**0.995**	1.156	1.299	1.410
200	0.874	**0.997**	1.130	1.247	1.338

Probability level for the hypothesis that the fit describes the deviations better than a random sample

the reduced chi-square value that was obtained by fitting f to {x,y}.

Insight Gained by Using the Reduced Chi-Square Statistic

- The reduced chi-square statistic simultaneously measures *both*:
  - The deviations between the data and the mean of the parent distribution that occur because there are less than an infinite number of samples/observations.
  - The discrepancy between the mean of the parent distribution and the mean predicted by the fitting function or "model", f.
- If the model is chosen poorly such that f cannot describe relationship correctly, the values for the chi-squared statistic will be high. For example:
  - Fitting a smooth SLD profile with too few discrete layers
  - Using an approximate model (e.g. Born approximation, etc.)
- To be significant at a relatively low probability level of 0.001, the reduced chi-square error must be less than about 1.5.