ORNL/CP-96092

EFFICIENT USE OF BAYES' EQUATIONS FOR FITTING NEUTRON-INDUCED CROSS SECTION AND INTEGRAL DATA[a]

N. M. Larson
Oak Ridge National Laboratory
Building 6010, MS 6370
Post Office Box 2008
Oak Ridge, TN 37831-6370
(423)574-4659
nml@ornl.gov

L. C. Leal
Oak Ridge National Laboratory
Building 6010, MS 6370
Post Office Box 2008
Oak Ridge, TN 37831-6370
(423)574-5281
e5a@ornl.gov

H. Derrien
Oak Ridge National Laboratory
Building 6010, MS 6370
Post Office Box 2008
Oak Ridge, TN 37831-6370

ABSTRACT

Analysis of differential and integral neutron data requires the use of a fitting procedure such as Bayes' method, used in the analysis code SAMMY. In this paper, techniques for more efficient use of Bayes' method are described. A reformulation of Bayes' equations permits truly simultaneous fitting to multiple measurements without overhead costs required for calculating theoretical values simultaneously. Several techniques for properly including data covariances are also described.

I. INTRODUCTION

Many quasi-independent processes are involved in the effort to determine resonance parameters which adequately describe a data set consisting of various differential and/or integral measurements. First, each individual measurement undergoes the data reduction process, which is the transformation from raw data into a recognizable physical quantity such as cross section. Second, a theoretical description (e.g., via R-matrix resonance theory) is used to model that physical quantity. Additional computations are needed to adequately describe specific experimental conditions (e.g., finite temperature, resolution broadening). Finally, physically reasonable values for all relevant parameters are determined by using some kind of fitting procedure; the hope is that those values will provide the best representation for all measurements included in the data set. In this paper we concentrate on the final task, the fitting procedure, and suggest a variety of ways in which current procedures can be enhanced in order to provide more efficient and/or more correct results.

II. DERIVATION OF BAYES' EQUATIONS

The fitting procedures used in analysis codes are, almost universally, based on Bayes' theorem, which can be written

$$p(P|DB) \propto p(P|B)\, p(D|PB)\ , \qquad (1)$$

in which $p(a|b)$ is the probability that a is true, given that b is true. The quantity P represents the parameters whose values are to be determined, D represents the experimental data to be analyzed, and B represents all other relevant information. Thus $p(P|B)$ is the prior probability density function (pdf) for the parameters, and $p(P|DB)$ is the posterior pdf. The quantity $p(D|PB)$ is often referred to as the likelihood function; it is the pdf for observing the data D, given that parameters P are correct.

Bayes' equations (sometimes called generalized least-squares equations) can be derived from Bayes' theorem using three basic assumptions: (1) The prior pdf is a joint normal distribution. (2) The likelihood function is a joint normal distribution. (3) The true value (of the theoretical description for the data) is a linear function of the parameters. A derivation of Bayes' equations is presented in the SAMMY users' manual;[1] here we merely state one version of the results:

[a]Research sponsored by the Office of Environmental Management, U.S. Department of Energy, under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.

DISCLAIMER

$$P' = P + M' Y \,, \tag{2}$$

$$M' = (M^{-1} + W)^{-1} \,, \tag{3}$$

$$Y = G^t V^{-1} (D - T) \,, \tag{4}$$

and

$$W = G^t V^{-1} G \,. \tag{5}$$

In these equations, P represents the initial (prior) parameter values, M the initial parameter covariance matrix, D the data, V the data covariance matrix, T the theory, and G the partial derivatives of T with respect to P (i.e., the sensitivity matrix). Primes indicate updated (posterior) values and covariance matrix. The quantities Y and W are defined by Eqs. (4) and (5).

In this form, the relationship with the more familiar least-squares equations is readily seen. In the limit of no prior knowledge, the initial covariance matrix M is diagonal, with each diagonal element having infinite value; thus, the inverse of M is zero. The equation for M' then reduces to $M' = W^{-1}$, which is the usual least-squares description.

III. USE OF BAYES' EQUATIONS

Many analysis codes (e.g., REFIT[2]) use the least-squares description for the fitting process, thus implicitly assuming zero prior knowledge of the parameter values. In this case all data *must* be included simultaneously within any analysis, since it is not possible to communicate results from one analysis to another. However, when Bayes' equations are used directly (as in SAMMY), the parameter covariance matrix from one analysis can be used as input to a subsequent analysis, thus providing an accurate fit to both measurements. Nevertheless, because the third assumption of the derivation (the linearity hypothesis) is only approximately correct, difficulties are sometimes encountered: Changes in the order in which measurements are analyzed may cause changes in the final results. The inclusion of anomalous data may skew results, and may make it impossible to converge on one particular set of parameter values. In general, the existence of second-order effects requires that caution be used during the fitting procedure.

The following scheme is proposed as an alternative to either of the two techniques listed above. First, it appears that all measurement-dependence in Eqs. (2-5) is summarized in Y and W, which are strictly additive with respect to independent measurements:

$$Y = Y_1 + Y_2 + Y_3 + \ldots \,, \tag{6}$$

and

$$W = W_1 + W_2 + W_3 + \ldots \,, \tag{7}$$

where the subscripts refer to individual measurements. It therefore follows that Y_i and W_i may be generated independently, kept in temporary storage, and added together when needed. Use of this technique therefore eliminates the overhead costs associated with simultaneously calculating theoretical values for all measurements and with storing and manipulating large arrays. Conversely, use of this technique also eliminates the need to solve Bayes' equations sequentially for each measurement, thus reducing the effects caused by the breakdown of the linearity assumption.

This scheme will soon be implemented into SAMMY, permitting truly simultaneous fitting for all data from many measurements.

A. Implementation in SAMMY

Implementation of the solution scheme for Bayes' equations, as outlined above, within the SAMMY analysis code will require major restructuring both internally within the code and also externally, in the input to the code. Because such restructuring will affect SAMMY users more directly than previous modifications to the code, the author is requesting feedback from users in advance of the proposed changes. A simplified description of the restructured input is presented in the following paragraph; SAMMY users are urged to review the detailed description in the appendix to this paper, and contact the author (at *nml@ornl.gov*) with their comments.

The "new" input to SAMMY will consist, essentially, of four different kinds of files. The first is a control or command file, which will provide specific instructions as to what operations SAMMY is to carry out. The second is an R-matrix parameter file (either in ENDF format or in the same format as portions of SAMMY's current PARameter file and INPut file), which contains all the information relevant to the R-matrix theory for one particular nuclide (masses, quantum numbers, spin-group definitions,

resonance energies and widths, radii, etc.). The third is a data file, containing energy, cross section, and uncertainty values for a particular experiment; the available formats will likely be the same as those for SAMMY's current DATa file. Associated directly with the data file is a data description file, which provides details about that experiment: what type of data these are, which nuclides occur in what abundances, what data-reduction processes are to be included in the analysis, how the data covariance matrix is to be described. Any given SAMMY run can include several of all types of files except the control file. As is now the case when SAMMY is used to determine appropriate parameter values and covariances, the output from a SAMMY run will include updated versions of the input files (i.e., for the R-matrix parameter files and for the data description files, assuming resonance parameters and data-reduction parameters were flagged to be varied in this run).

IV. DATA COVARIANCE MATRICES

Although raw data are truly independent (and thus have zero values for off-diagonal data covariances), the data-reduction process described in the introduction gives rise to non-zero off-diagonal data covariances. Unfortunately, these covariances are usually ignored (even though most analysts acknowledge their importance), simply because it is not practical to either generate, store, or use very large matrices. (A differential measurement may contain hundreds of thousands of data points, for which the covariance matrix is very large indeed.) A variety of options is available in SAMMY to circumvent this problem while properly taking into account the data-reduction parameters and their uncertainties.

A. Data Reduction as Part of Theoretical Calculation

Some of the more simple data-reduction processes (such as background subtraction and normalization) can be applied in reverse to the theoretical calculations. Values, uncertainties, and covariances for the data-reduction parameters are then determined in the fitting procedure along with values, uncertainties, and covariances for resonance parameters. Mathematically this is equivalent to applying the normalization and backgrounds to the data and then incorporating the data covariance matrix into the fitting procedure, yet it does not require storage or inversion of the large (off-diagonal) data covariance matrix.

The latest release of the SAMMY code contains several new options for energy-dependent backgrounds.

B. Implicit Data Covariance

The general form for the data covariance matrix can be written

$$V^{ij} = v^i \delta^{ij} + \sum_{kl} X_k^i \, w_{kl} \, X_l^j \qquad (8)$$

where V is the data covariance, v represents the statistical uncertainties (and is therefore diagonal), X is the sensitivity matrix (partial derivative of the reduced data with respect to data-reduction parameters), and w is the covariance matrix for the data-reduction parameters (often diagonal). [Superscripts are used in Eq. (8) to indicate data indices, subscripts indicate parameter indices; in the following equations, indices are suppressed.] The inverse of the matrix in Eq. (8) can be written as

$$\begin{aligned} V^{-1} &= (v + X w X^t)^{-1} \\ &= v^{-1} - v^{-1} X (w^{-1} + X^t v^{-1} X)^{-1} X^t v^{-1} \end{aligned} \qquad (9)$$

Combining Eq. (9) with Eq. (4) for Y and (5) for W gives

$$\begin{aligned} Y &= G^t V^{-1} (D - T) \\ &= G^t v^{-1} (D - T) \\ &\quad - G^t v^{-1} X (w^{-1} + X^t v^{-1} X)^{-1} \\ &\qquad\qquad X^t v^{-1} (D - T) \, , \end{aligned} \qquad (10)$$

and

$$\begin{aligned} W &= G^t V^{-1} G \\ &= G^t v^{-1} G \\ &\quad - G^t v^{-1} X (w^{-1} + X^t v^{-1} X)^{-1} \\ &\qquad\qquad X^t v^{-1} G \, . \end{aligned} \qquad (11)$$

Thus only terms like $G^t v^{-1} G$, $X^t v^{-1} G$, and $X^t v^{-1} X$ (and their transposes) are needed; it is never necessary to calculate V^{-1} directly. It *is* necessary to calculate the inverse of $(w^{-1} + X^t v^{-1} X)$, but this matrix has dimensions equal to the number of data-reduction parameters (which is a relatively small number). Thus the necessity to store and invert very large arrays is again eliminated.

This scheme is operational in version M of SAMMY, which is available from the Radiation Safety Information Computational Center [3].

V. CONCLUSION

Bayes' method is the fitting procedure used in the analysis code SAMMY to determine those parameter values, uncertainties, and covariance matrix which give the best fit to all available data. A restructuring has been proposed for the SAMMY code and the SAMMY input, in order to more efficiently utilize the strengths of Bayes' method; one beneficial side-effect of the restructuring will be a more logical organization for the input to the code. Methodologies have been described for incorporating data covariance information into the analysis process; these methodologies are available in the current version of SAMMY, and will of course be included in any restructured version as well.

APPENDIX

Readers who are familiar with the formats of SAMMY input are requested to read this appendix carefully and provide the authors with feedback regarding specific details of the proposed changes. Please send comments, both positive and negative, to nml@ornl.gov.

Do not expect these changes to happen immediately. This is a major undertaking, so the work itself will require substantial time. In addition, these improvements are not yet funded, and that too will take time. If these improvements are important to your work, you may wish to include SAMMY development work in your next funding request.

A. Input

As always, the author will attempt to operate within the "grandfather clause" -- that is, to ensure that "old" input will still work even with the new version of SAMMY, to as great an extent as possible. With the restructuring, however, virtually all input files will require reorganization.

The current SAMMY input stream is essentially a list of file names corresponding to SAMMY INPut, PARameter, DATa, COVariance files. The restructured input stream would include the names of the COMmand file, PARameter files, first EXPerimental file, first DATa file, first DCV (Data CoVariance, optional), second EXPerimental file, second DATa file, etc.

The new COMmand file will be similar to the command lines from the current INPut file, except it will include neither data-specific information nor resonance-parameter-specific information. DO NOT SOLVE BAYES' EQUATIONS is one of the few commands that apply here.

The new PARameter file will be a combination of pieces of the current INPut and PARameter files. In particular, the spin-group information (quantum numbers) will be here, as will the resonance parameters and anything else directly related to the R-matrix (scattering radius, e.g.). Anything related to the measured data will not be here (see below). The new PARameter file would correspond to one nuclide only; there will be one such file for each nuclide. Consequently there will need to be some identifier specifying which nuclide this is; possibly this will follow ENDF conventions. Certainly the use of ENDF rather than SAMMY-style PARameter files will still be permitted; unfortunately ENDF formats are not sufficiently general to cover all contingencies, so SAMMY cannot use them exclusively.

The EXPeriment file will look very much like a combination of pieces of the current INPut file and latter portions of the PARameter file. It will contain descriptive information about the data set; that is, it will give Doppler- and resolution-broadening parameters, backgrounds and normalization, multiple-scattering information, nuclide abundances, etc. In other words, the EXP file will contain all relevant information about the experiment which produced the data. Probably the energy range will be specified here, and not in the input stream (after the file name, an option permitted in the current version of SAMMY). The option of including more than one energy range for the same data will be disallowed; data file names can simply be repeated if a user wishes to break one measurement into more than one energy range, for example if there are a very large number of data points.

The DATa file is identical to today's DATa file. There are at least four different options in the latest version of SAMMY: original (3 data points per line, in

3(F15.1,F15.1,F7.1) format), CSISRS (one data point per line in 3F11.1 format), in ODF file (binary), in TWENTY format (3F20.1, for those times when many significant digits are required). An option to include the ENDF format may also be added here; the author welcomes your suggestions for any additional formats.

The Data CoVariance file should be significantly changed from the current format, which is exceedingly awkward to use. [See Section C. below.] Users' suggestions would be appreciated here.

B. Output Files

It is possible to have "flagged" parameters related either to the theoretical cross sections (i.e. from the theory) or to the measured data (i.e. from the experiment). Prior values for the parameters will be in the PAR or EXP files, respectively, along with the flag that announces whether it is to be varied or not. SAMMY output will therefore need to include a new version of all the input files in which a parameter was flagged; logistically this could be awkward to implement, especially since the output covariance matrix will include connections between all the various PAR and DAT files.

(A) One possibility is for SAMMY to have a standard set of names as in the current version. Output files would be, e.g., SAM1.PAR, SAM2.PAR, SAM3.PAR,... SAM1.EXT, SAM2.EXT, ... A disadvantage is that this may result in confusion as to which output file corresponds to which input file; the user may have to look inside each file to be sure. (Alternatively, SAMMY could print a listing of output file names.)

(B) Another possibility is to require that the input PARameter and EXPerimental file names have "standard" extensions PAR and EXT. SAMMY could then provide an output file with the same name but a different extension (say, PPP and EEE). There are at least two drawbacks to this option: (1) It may not be possible to implement this in a machine-independent fashion. (2) There may be a danger of overwriting existing files.

(C) Yet another possibility is for the user to provide file names in the input stream, either (1) two in one line (but that becomes awkward when long path-names are included), (2) output file name on a new line following input name for every file (but then the input stream would require modification to do a "no Bayes" calculation, or (3) a separate listing of output names. In either of the last two cases the listing would give output file names for ALL corresponding input files, whether or not any variable is flagged in the file, but only those required would actually be created.

The author's preference is for option (C3), but consensus will rule. Any specific suggestions in this regard should be sent to the author as soon as possible. If you can suggest a better alternative to those listed above, please do so.

A similar question (with similar options) relates to naming conventions for SAMMY-generated "plot files" or ODF files. [The plot files contain both "prior" and "posterior" calculated cross sections, as well as experimental data.]

C. Other Considerations

The input of uncertainties (and correlations, when needed) is not done in a consistent manner in the current version of SAMMY. Since other aspects of the input are being changed, now would appear to be an appropriate time for improving this as well.

1. Experimental data. The current formats for input of experimental data also provide uncertainties; these formats are generally adequate for situations when the data covariance matrix is diagonal. (If this is not the case in your experience, please contact the author.) The situation for off-diagonal correlations/covariances is quite different. Currently the only real option for specifying the experimental covariance matrix (see pages 127-128 of the SAMMY manual) can be frustrating to use: It requires the DCV file to contain covariances only for those data points actually used in the calculation; hence a change in the energy range requires a new DCV file. One possible improvement is to have a file in the current format, with all covariances specified, with the same ordering as that used in the DATa file (and SAMMY extracting the relevant pieces as needed); however, for more than a few data points this file would be very large indeed, so another option would be needed. Another possibility is for the file to contain correlations times 100 (as is done for the SAMMY output parameters) rather than covariances. If you have used data covariances, please contact the author with your comments.

2. Resonance parameters. Currently there are a variety of options for inputting prior uncertainties for resonance parameters. The easiest is to use defaults; the user may specify a value for FUDGE, and the uncertainty on a parameter is set to FUDGE times the value of that parameter (with some exceptions, as described in the SAMMY manual). Most likely that option will be retained in some fashion, but feedback on specifics is requested: Should there be one FUDGE for all variables, from all PAR and EXP files? Or should each PAR and EXP file provide its own value for FUDGE?

With most options recently added to SAMMY, initial parameter values and uncertainties are specified together (either on the same line or on consecutive lines). That is *not* true for the resonance parameters. An option to permit this might be added; each line (or lines, for more than three particle channels) would be followed by a line(s) specifying the initial uncertainties (in the same location as the values in the previous line). Because this will effectively double the size of the parameter file, it would be an option rather than a new default (so there would be a command line at the beginning of the new PAR file specifying its use).

3. Data-related parameters. Currently, initial parameter values and uncertainties are specified together (either on the same line or on consecutive lines); this will remain the case with the restructuring.

ACKNOWLEDGMENTS

This research was performed at the Oak Ridge National Laboratory, managed by Lockheed Martin Energy Research Corp., under contract DE-AC05-96OR22464 with the U. S. Department of Energy.

REFERENCES

1. N. M. Larson, *Updated User Guide for SAMMY: Multilevel R-matrix Fits to Neutron Data Using Bayes' Equations*, ORNL/TM-9179/R3, Martin Marietta Energy Research, Inc., Oak Ridge National Laboratory, Oak Ridge, TN (Sept. 1996). R4 is expected to be published during 1998.

2. M. C. Moxon, *REFIT, A Least Squares Fitting Program for Resonance Analysis of Neutron Transmission and Capture Data*, AEA-Intec-0470 (April 1991).

3. Radiation Safety Information Computational Center (RSICC) [formerly, Radiation Shielding Information Center (RSIC)], P.O. Box 2008, Oak Ridge, TN 37831-6362 USA. See also the web site http://www-rsicc.ornl.gov/rsic.html .