IV.D. TREATMENT OF DATA COVARIANCE MATRICES

If we accept the assumption that the use of Bayes' equations (or least squares) is an appropriate choice for the data-fitting procedure, then there are two possible methods for propagating the uncertainty on a data-reduction parameter (e.g., normalization) through the analysis process. The first is to treat the parameter as a variable to be fitted, having a prior uncertainty associated with the initial value of that parameter. The second is to treat the value of the parameter as fixed, but include the uncertainty in the data covariance matrix (DCM).

The expectation is that if both methods are correct, then the two methods should yield identical results (at least in the absence of complicating features such as nonlinearities). If results are not identical, then one or the other of the methods must be flawed in some fashion.

In this section, we present a description of the DCM that works equally well for fitting raw data and for fitting reduced data, giving consistent and reliable results. Prior uncertainties are reflected in the final results, both for varied parameters and for those fixed parameters that are used to determine the DCM.

A derivation of this description of the DCM is outlined here; details are shown in Section IV.D.1. See also [NL04a], [NL04b], and the SAMMY workshop notes.

We begin with the assumption that correct results are obtained by treating a data-reduction parameter as a variable to be fitted. This is a reasonable assumption because this treatment uses a diagonal DCM, and there is little or no argument regarding appropriate treatment for diagonal DCMs. The data being analyzed are the "raw" measurements (e.g., measurements of a quantity related to the cross section) plus other measurements related to the value of the data-reduction parameter.

Simple algebra is then used to re-express the equations in terms of "reduced" data (e.g., data that have been converted from raw counts to cross section by dividing by the normalization parameter). Comparison of the resulting equations to the usual version of Bayes' equations for fitting to reduced data indicates the need for modification of the usual version, with regard to the definition of the DCM. In particular, it is necessary to define the DCM in terms of the theoretical values, not in terms of the measured values, for the quantity to be fitted (e.g., the cross section).

The resulting formulae for Bayes' equations can be written as follows for the M+W method:

$$\begin{aligned} P' - P &= M'Y & M' &= \left(W + M^{-1} \right)^{-1} \\ Y &= G^t V^{-1} (D - T) & W &= G^t V^{-1} G \end{aligned} \quad , \tag{IV D.1}$$

in which V is shown to have the form

$$V = v + g\, m\, g^t \quad . \tag{IV D.2}$$

In these equations, the data D may be either raw or reduced, and the DCM V is calculated at the true (i.e., theoretical) value of the cross sections. The quantity v represents the portion of V that derives from the statistics of the data-taking process, and $g\,m\,g^t$ represents the "common" or "systematic" portion that derives from corrections made to the raw data (e.g., division by an overall normalization factor). In this formulation, the systematic portion must be calculated from theory rather than measurement.

The only difference between this approach and the "usual" approach is in the definition of the partial derivatives (sometimes denoted "sensitivities") g: in the usual approach, these are evaluated at the experimental value. Here, these are evaluated at the theoretical value. (For example, the derivative associated with division by an overall normalization factor would be the cross section divided by the normalization. In the usual approach, the experimental cross section would be used for this derivative; in this approach, the theoretical cross section would be used.)

Not only does this approach allow us to obtain the same results (to first order) as would be obtained by fitting raw data, but we are also able to avoid numerical problems, logical inconsistencies, and some of the issues related to discrepant data.

A simple example may serve to illustrate the difference between the usual DCM and the version advocated here. Suppose two measurements were taken of the same quantity, one yielding 10000 ± 100 counts and the other 12100 ± 110. The covariance matrix associated with the raw data is therefore

$$\begin{bmatrix} (100)^2 & 0 \\ 0 & (121)^2 \end{bmatrix} . \qquad \text{(IV D.3)}$$

After being normalized by a factor of 100, the reduced data are $d_1 = 100$ and $d_2 = 121$. Generating the covariance matrix for the reduced data in the usual fashion requires the following steps: First, write the reduced data in terms of the raw data, and then take small increments of both sides of the equation, giving

$$\begin{aligned} d_i &= r_i / n \\ \delta d_i &= \delta r_i / n - (\delta n / n)(r_i / n) \quad . \end{aligned} \qquad \text{(IV D.4)}$$

Squaring and taking expectation values then gives

$$\begin{aligned} V_{ij} &= \langle \delta d_i \delta d_j \rangle \\ &= \langle \delta r_i \delta r_j \rangle / n^2 + (\langle \delta n\, \delta n \rangle / n^2)(r_i / n)(r_j / n) \\ &\qquad + \langle \delta r_i \delta n \rangle / n^2 (r_i / n) + \langle \delta n\, \delta r_j \rangle / n^2 (r_j / n) \\ &= \langle \delta r_i \delta r_j \rangle / n^2 + (\langle \delta n\, \delta n \rangle / n^2)(r_i / n)(r_j / n) \\ &= \delta_{ij}\, \Delta^2 r_i / n^2 + (\Delta^2 n / n^2)\, d_i\, d_j \quad . \end{aligned} \qquad \text{(IV D.5)}$$

(This derivation assumes that $\langle \delta r_i \delta n \rangle = \langle \delta n \, \delta r_j \rangle = 0$, that is, that *r* and *n* are uncorrelated, a reasonable assumption.) If $\Delta n = 0.5$, then the usual DCM for the reduced data has the form

$$V = \begin{bmatrix} \Delta^2 r_1 / n^2 + \left(\Delta^2 n / n^2\right) d_1^2 & \left(\Delta^2 n / n^2\right) d_1 d_2 \\ \left(\Delta^2 n / n^2\right) d_1 d_2 & \Delta^2 r_2 / n^2 + \left(\Delta^2 n / n^2\right) d_2^2 \end{bmatrix}$$
$$= \begin{bmatrix} (1)^2 + (0.005)^2 (100)^2 & (0.005)^2 (100)(121) \\ (0.005)^2 (100)(121) & (1.21)^2 + (0.005)^2 (121)^2 \end{bmatrix} = \begin{bmatrix} 1.25 & 0.325 \\ 0.325 & 1.5125 \end{bmatrix} . \quad \text{(IV D.6)}$$

For this simple example, Bayes' equations can be used to calculate the average of these two values; simply set *t* = *P* (theory equal to parameter) so that $G^t = \begin{bmatrix} 1 & 1 \end{bmatrix}$. The solution is *t* = 109.20 ± 0.92.

In contrast, the DCM using the more rigorous approach advocated here would have the form

$$V = \begin{bmatrix} \Delta^2 r_1 / n^2 + \left(\Delta^2 n / n^2\right) t^2 & \left(\Delta^2 n / n^2\right) t^2 \\ \left(\Delta^2 n / n^2\right) t^2 & \Delta^2 r_2 / n^2 + \left(\Delta^2 n / n^2\right) t^2 \end{bmatrix}$$
$$= \begin{bmatrix} (1)^2 + (0.005)^2 t^2 & (0.005)^2 t^2 \\ (0.005)^2 t^2 & (1.21)^2 + (0.005)^2 t^2 \end{bmatrix} , \quad \text{(IV D.7)}$$

in which *t* is the as-yet-unknown theoretical value for the quantity being measured. Clearly, in this form the solution of Bayes' equations is somewhat more complicated, but not unduly so. The solution is *t* = 108.52 ± 0.94, which in this example is only slightly different from the "usual" solution. In cases where the data are markedly discrepant, the differences can be significant.

In SAMMY, any parameter for which derivatives are available (i.e., any parameter that is capable of being fitted) can be treated as a contributor to the DCM. Such parameters are denoted "propagated uncertainty parameters," or PUPs. In Section IV.D.2, application of the PUP method within SAMMY is described. A list is given of those parameters that may be treated as PUPs.

Section IV.D.3 deals with methods of generating and inverting the DCM. In particular, the "implicit data covariance" (IDC) method is introduced, which has many advantages over explicitly generating, storing, and inverting the full DCM.

Even though the method is known to be flawed,* it is nevertheless possible to generate the DCM externally in the usual fashion [e.g., as described in Eqs. (IV D.4) and (IV D.5) for normalization] and provide this as input to SAMMY. Providing the DCM in this fashion is far preferable to another common practice, totally ignoring the off-diagonal elements of the DCM.

Many SAMMY test cases give examples of the various treatments of DCM. See Section XII.B for a list of test cases.

* These flaws include inconsistency with the fit-to-raw-data method; excessive computer time and memory requirements for explicitly creating, storing, and inverting the full DCM; potential problems with discrepant data; and possible loss of accuracy and associated numerical difficulties if using ASCII formats to communicate values for the matrix elements.