IV.A. DERIVATION OF BAYES' EQUATIONS

A derivation of Bayes' equations is given in this section. It should be noted that this is not the only possible derivation; alternatives can be found in [JM80] and [AG73].

In Section IV.A.1, results obtained in this section are used as the starting point for defining the three forms of Bayes' equations used in SAMMY. A discussion of the chi-squared values is given in Section IV.A.2, and the iteration scheme that compensates for non-linearity is presented in Section IV.A.3.

Our derivation begins with Bayes' theorem, which may be written in the form

$$p\left(P \mid DX\right) = p\left(P \mid X\right) p\left(D \mid PX\right) \tag{IV A.1}$$

where

- P represents the parameters of the (extended) R-matrix theory, and D represents the experimental data to be analyzed.
- X represents "background" or "prior" information such as the data from which prior knowledge of the parameters P was derived. X is assumed to be independent of D.
- $p(P|DX)$ is the probability for the value of the parameters, conditional upon the new data D, and is what we seek. It is conventional to call $p(P|DX)$ the posterior probability. Since P represents several parameters, $p(P|DX)$ is a joint probability density function (joint pdf). The expectation values of P times $p(P|DX)$ are taken as the new estimates for the parameters; the associated covariance matrix gives us a measure of how well the parameters are determined and of the interdependencies of those determinations.
- $p(D|PX)$ is the probability density function for observing the data D given that the parameters P are correct. It is a function of the parameters P of the model and is proportional to the likelihood function of the data D.
- $p(P|X)$ is the joint pdf for the values of the parameters P of the model, prior to consideration of the new data D; it is known as the prior joint pdf. The expectation values of P times $p(P|X)$ are the prior estimates for the values of the parameters; the associated covariance matrix gives a measure of how well the parameters are known before consideration of the new data.

Let $P = \{P_k\}$ for $k = 1$ to K be the set of all parameters of the theoretical model to be considered. The joint pdf $p(P|X)$ is assumed to be a joint normal pdf having as expectation value the vector $\overline{P}$ and the covariance matrix M. Under this assumption, the pdf may be written

$$p\left(P \mid X\right) \propto \exp\left[-\frac{1}{2}\left(P-\overline{P}\right)^t M^{-1}\left(P-\overline{P}\right)\right] \ , \tag{IV A.2}$$

where superscript t denotes the transpose.

The experimental data are represented by a data vector D whose components D_i are the L data points. The experimental conditions are assumed to be such that the data D (i.e., the D_i's) have a joint normal distribution with mean $\langle D \rangle = T = T(P)$ and covariance matrix V. The likelihood function is then

$$p(D \mid PX) \sim \exp\left[-\frac{1}{2}(D-T)^t V^{-1}(D-T) \right] \quad . \tag{IV A.3}$$

Here T represents theory (i.e., calculated values of cross section or transmission plus corrections), and the covariance matrix V represents not only the experimental "errors" of the data but also any theoretical "errors" resulting from approximations used in calculating T. Obviously V need not be diagonal.

Combining Eqs. (IV A.1), (IV A.2), and (IV A.3) gives an expression for the pdf of P after consideration of new data D [i.e., for $p(P|DX)$], expressed in terms of the "true" value T. What is needed, however, is an expressions for $p(P|DX)$ expressed in terms of the parameters P. This is obtained formally by considering T a function of P, performing a Taylor expansion about $\overline{P}$ [the expectation value of $p(P|X)$], and keeping only the linear terms:

$$T(P) = \overline{T} + G(P - \overline{P}) + \ldots \quad , \tag{IV A.4}$$

where $\overline{T} = T(\overline{P})$. The elements of G, often denoted the "sensitivity matrix," are the partial derivatives of T_n with respect to the parameters P_k , evaluated at $P = \overline{P}$:

$$G_{nk} = \left. \frac{\partial T_n}{\partial P_k} \right|_{P=\overline{P}} \quad \text{for} \begin{cases} n = 1 \text{ to } L \\ k = 1 \text{ to } K \end{cases} . \tag{IV A.5}$$

Because T is a vector of dimension L (equal to the number of data points), and P is a vector of dimension K (equal to the number of parameters), the sensitivity matrix G has dimension $L \times K$.

Substituting Eq. (IV A.4) into Eq. (IV A.3) and using Eq. (IV A.2), we obtain for the posterior joint pdf [Eq. (IV A.1)]

$$\begin{aligned} p(P \mid DX) \sim \exp\Big[-\frac{1}{2}(P-\overline{P})^t M^{-1}(P-\overline{P}) + \\ \left(D-\overline{T}-G(P-\overline{P})\right)^t V^{-1}\left(D-\overline{T}-G(P-\overline{P})\right)\Big] \quad . \end{aligned} \tag{IV A.6}$$

Because of the three basic assumptions we have made, that is,

i. the prior joint pdf is a joint normal, Eq. (IV A.2),
ii. the likelihood function is a joint normal, Eq. (IV A.3), and
iii. the true value is a linear function of the parameters, Eq. (IV A.4) with (IV A.5),

it follows that the posterior joint pdf is also a joint normal. Denoting its expectation value by $\bar{P}'$ and its covariance matrix by M', we may write

$$p\left(P \mid DX\right) \sim \exp\left[-\frac{1}{2}\left\{\left(P-\bar{P}'\right)^t \left(M'\right)^{-1} \left(P-\bar{P}'\right)\right\}\right] \quad . \tag{IV A.7}$$

Substituting Eqs. (IV A.2), (IV A.3), and (IV A.7) into (IV A.1) gives

$$\begin{aligned}\exp\left[-\frac{1}{2}\left\{\left(P-\bar{P}'\right)^t \left(M'\right)^{-1} \left(P-\bar{P}'\right)\right\}\right] &\sim \exp\left[-\frac{1}{2}\left(P-\bar{P}\right)^t M^{-1} \left(P-\bar{P}\right)\right] \\ &\times \exp\left[-\frac{1}{2}\left(D-T\right)^t V^{-1} \left(D-T\right)\right] \quad ,\end{aligned} \tag{IV A.8}$$

which is valid up to a normalization constant. To simplify this expression, we substitute Eq. (IV A.4) for *T*, obtaining

$$\begin{aligned}\exp\left[-\frac{1}{2}\left\{\left(P-\bar{P}'\right)^t \left(M'\right)^{-1} \left(P-\bar{P}'\right)\right\}\right] &\sim \exp\left[-\frac{1}{2}\left(P-\bar{P}\right)^t M^{-1} \left(P-\bar{P}\right)\right] \\ \times \exp\left[-\frac{1}{2}\left(D-\bar{T}-G\left(P-\bar{P}\right)\right)^t V^{-1} \left(D-\bar{T}-G\left(P-\bar{P}\right)\right)\right] &\quad .\end{aligned} \tag{IV A.9}$$

In this form, the theoretical function *T* is treated as strictly linear with respect to the parameters. (In Section.IV.A.3, this formula will be generalized for nonlinear *T*.) Rearranging the exponents gives

$$\begin{aligned}\exp\Bigg[-\frac{1}{2}\Big\{&\left(P-\bar{P}'\right)^t \left(M'\right)^{-1} \left(P-\bar{P}'\right) \\ &-\left(P-\bar{P}\right)^t \left(M^{-1}+G^t V^{-1} G\right)\left(P-\bar{P}\right) \\ &+\left(P-\bar{P}\right)^t G^t V^{-1}\left(D-\bar{T}\right)+\left(D-\bar{T}\right)^t V^{-1} G\left(P-\bar{P}\right) \\ &-\left(D-\bar{T}\right)^t V^{-1}\left(D-\bar{T}\right)\Big\}\Bigg] \quad = \quad \text{constant},\end{aligned} \tag{IV A.10}$$

where the value of the constant depends upon the unspecified normalizations. After setting

$$P-\bar{P}' = P-\bar{P}+\bar{P}-\bar{P}' \tag{IV A.11}$$

and rearranging, Eq. (IV A.10) becomes

$$\exp\left[-\frac{1}{2}\left\{\left(P-\bar{P}\right)^t\left(M'\right)^{-1}\left(P-\bar{P}\right)+\left(\bar{P}-\bar{P}'\right)^t\left(M'\right)^{-1}\left(P-\bar{P}\right)\right.\right.$$
$$+\left(P-\bar{P}\right)^t\left(M'\right)^{-1}\left(\bar{P}-\bar{P}'\right)+\left(\bar{P}-\bar{P}'\right)^t\left(M'\right)^{-1}\left(\bar{P}-\bar{P}'\right)$$
$$-\left(P-\bar{P}\right)^t\left(M^{-1}+G^tV^{-1}G\right)\left(P-\bar{P}\right) \quad \text{(IV A.12)}$$
$$+\left(P-\bar{P}\right)^t G^t V^{-1}\left(D-\bar{T}\right)+\left(D-\bar{T}\right)^t V^{-1}G\left(P-\bar{P}\right)$$
$$\left.\left.-\left(D-\bar{T}\right)^t V^{-1}\left(D-\bar{T}\right)\right\}\right] = \text{constant,}$$

or

$$\exp\left[-\frac{1}{2}\left\{\left(P-\bar{P}\right)^t\left[\left(M'\right)^{-1}-\left(M^{-1}+G^tV^{-1}G\right)\right]\left(P-\bar{P}\right)\right.\right.$$
$$+\left(P-\bar{P}\right)^t\left[\left(M'\right)^{-1}\left(\bar{P}-\bar{P}'\right)+G^tV^{-1}\left(D-\bar{T}\right)\right]$$
$$+\left[\left(\bar{P}-\bar{P}'\right)^t\left(M'\right)^{-1}+\left(D-\bar{T}\right)^tV^{-1}G\right]\left(P-\bar{P}\right) \quad \text{(IV A.13)}$$
$$\left.\left.+\left(\bar{P}-\bar{P}'\right)^t\left(M'\right)^{-1}\left(\bar{P}-\bar{P}'\right)-\left(D-\bar{T}\right)^tV^{-1}\left(D-\bar{T}\right)\right\}\right] = \text{constant.}$$

In order for Eq. (IV A.13) to be valid for all P, the exponent must be constant. For this to happen, the terms linear and quadratic in $\left(P-\bar{P}\right)$ must separately be equal to zero. This gives us two equations relating the primed (posterior) values to the unprimed (prior) values:

$$\left(M'\right)^{-1}-\left(M^{-1}+G^tV^{-1}G\right)=0 \quad \text{(IV A.14)}$$

and

$$\left(M'\right)^{-1}\left(\bar{P}-\bar{P}'\right)+G^tV^{-1}\left(D-\bar{T}\right)=0 \quad , \quad \text{(IV A.15)}$$

which can be rewritten in the form

$$\left(M'\right)^{-1}=M^{-1}+G^tV^{-1}G \quad \text{(IV A.16)}$$

and

$$\bar{P}'=\bar{P}+M'G^tV^{-1}\left(D-\bar{T}\right) \quad . \quad \text{(IV A.17)}$$

Equations (IV A.16) and (IV A.17) are denoted “Bayes’ Equations,” sometimes called “generalized least squares” for reasons which shall be discussed later. Three distinct but equivalent versions of these equations are used within the SAMMY code, as described in the next section of this manual (Section IV.A.1).