Some Thoughts on Chi-Square Expressions

G. Hale (4/30/04)

As most of you know, I have been concerned from early in the Standards CRP about differences in the χ^2 expressions used in EDA and in the other R-matrix and generalized least-squares fitting codes. Recently I had the opportunity to revisit this question during a mini-workshop on the standard cross sections at Los Alamos that benefited from the presence of Don Smith and Ken Hanson to give guidance on statistical matters. Although they were participating mainly to address issues associated with PPP effects, the discussions I had with Ken, especially, helped me to get a fresh perspective on our approach to chi-square data fitting. The result appears to be that the chi-square expression used by EDA is very similar (but not identical) to the others when the normalization parameters are adjusted at each stage to minimize the chi-square for a given set of R-matrix parameters. This is actually the way EDA works, but I had not considered the corresponding expression when making the comparison. Some details follow:

<u>Chi-square Expressions</u>

The χ^2 expression used in EDA is

$$\chi^2_{\text{EDA}} = \sum_i \left[\frac{nX_i(\mathbf{p}) - R_i}{\Delta R_i}\right]^2 + \left[\frac{nS-1}{\Delta S/S}\right]^2, \tag{1}$$

in which R_i are the relative values, and S the scale, for a set of measurements. For simplicity, we consider only one such data set in these expressions, although in general the chi-square expression is a sum of such terms. The experimental (one-sigma) uncertainties for these quantities are denoted as ΔR_i and ΔS, respectively. This form of chi-square implies that the quantities R_i and S are statistically independent, which is a good approximation to the way most scattering measurements are made. Calculated values of the experimental observables in terms of the R-matrix parameters **p** are denoted as $X_i(\mathbf{p})$, and n is an adjustable normalization parameter associated with the experimental scale S. Normally, the experimental value given for S is $1.0\pm\Delta S$, but it is left more general in Eq. (1). This expression also applies to the situation where the measurements are purely relative, in which case the second term of Eq. (1) vanishes, ($\Delta S\to\infty$) and n adjusts the calculation to most closely match the relative measurements, unconstrained by a scale.

The expression in Eq. (1) is exactly quadratic in n:

$$\chi^2_{\text{EDA}} = A(\mathbf{p})n^2 + B(\mathbf{p})n + C, \tag{2}$$

with

$$A(\mathbf{p}) = \sum_i \left[\frac{X_i(\mathbf{p})}{\Delta R_i}\right]^2 + \left[\frac{S}{\Delta S/S}\right]^2,$$

$$B(\mathbf{p}) = -2\left[\sum_i \frac{X_i(\mathbf{p})R_i}{(\Delta R_i)^2} + \frac{S}{(\Delta S/S)^2}\right], \tag{3}$$

$$C = \sum_i \left[\frac{R_i}{\Delta R_i}\right]^2 + \left[\frac{1}{\Delta S/S}\right]^2.$$

This is easily solved for the normalization n_0 that minimizes χ^2 for fixed parameters $\mathbf{p}$,

$$n_0(\mathbf{p}) = -\frac{B(\mathbf{p})}{2A(\mathbf{p})}, \tag{4}$$

by taking $g_n(\mathbf{p}) = \frac{\partial \chi^2_{\mathrm{EDA}}}{\partial n} = 0.$

In terms of the coefficients of Eq. (2), this defines the chi-square function

$$\chi^2_{n_0}(\mathbf{p}) = \frac{B^2(\mathbf{p})}{4A(\mathbf{p})} - \frac{B^2(\mathbf{p})}{2A(\mathbf{p})} + C = C - \frac{B^2(\mathbf{p})}{4A(\mathbf{p})} = \frac{4A(\mathbf{p})C - B^2(\mathbf{p})}{4A(\mathbf{p})}, \tag{5}$$

which is involved in the distribution obtained by "marginalizing" χ^2_{EDA} over normalizations, as will be discussed in the following section. Substituting the expressions in Eqs. (3) gives

$$\chi^2_{n_0}(\mathbf{p}) = \frac{\sum_i \left[\frac{X_i(\mathbf{p})/S - R_i}{\Delta R_i}\right]^2 + \left(\Delta S/S\right)^2 \sum_{ij} \frac{X_i(\mathbf{p})R_j\left[X_i(\mathbf{p})R_j - X_j(\mathbf{p})R_i\right]}{\left(\Delta R_i \Delta R_j\right)^2}}{1 + \left(\Delta S/S\right)^2 \sum_i \left[\frac{X_i(p)}{\Delta R_i}\right]^2}. \tag{6}$$

This can be put into a more reasonable form by defining the dimensionless vectors

$$x_i(\mathbf{p}) = \frac{X_i(\mathbf{p})}{\Delta R_i S} \text{ and } r_i = \frac{R_i}{\Delta R_i}, \tag{7}$$

so that Eq. (6) becomes

$$\chi^2_{n_0}(\mathbf{p}) = \frac{[\mathbf{x}(\mathbf{p}) - \mathbf{r}]^2 + \left(\Delta S/S\right)^2\left[\mathbf{x}^2(\mathbf{p})\mathbf{r}^2 - (\mathbf{x}(\mathbf{p})\cdot\mathbf{r})^2\right]}{1 + \left(\Delta S/S\right)^2 \mathbf{x}^2(\mathbf{p})}. \tag{8}$$

Now we consider the "standard" chi-square expression in which the measurements are considered to be the products $M_i = R_i S$, and the deviations are weighted by the inverse of their covariance matrix $\mathbf{V}_{\mathrm{M}}$,

$$\chi^2 = \sum_{i,j} (X_i(\mathbf{p}) - M_i)(\mathbf{V}_M^{-1})_{ij}(X_j(\mathbf{p}) - M_j). \tag{9}$$

In this case, the covariance matrix that corresponds to the same assumptions as those made in the EDA chi-square, namely

$$\begin{aligned} &\mathrm{cov}(R_i, R_j) = (\Delta R_i)^2 \delta_{ij}, \\ &\mathrm{cov}(S,S) = (\Delta S)^2, \\ &\mathrm{cov}(R_i, S) = 0, \end{aligned} \tag{10}$$

is

$$\mathbf{V}_{ij}^{M} = \underbrace{S^2(\Delta R_i)^2\delta_{ij}}_{\text{diagonal piece}} + \underbrace{R_i R_j(\Delta S)^2}_{\text{rank-1 piece}}. \tag{11}$$

In terms of the same dimensionless vectors defined above in Eq. (7), Eq. (9) becomes

$$\chi^2 = (\mathbf{x}(\mathbf{p})-\mathbf{r})^{\mathrm{T}}\left[\mathbf{1}+(\Delta S/S)^2\mathbf{r}\mathbf{r}^{\mathrm{T}}\right]^{-1}(\mathbf{x}(\mathbf{p})-\mathbf{r}). \tag{12}$$

Using a useful identity for inverses involving rank-1 matrices,

$$(\mathbf{1}+\alpha\mathbf{r}\mathbf{r}^{\mathrm{T}})^{-1} = \mathbf{1}-\frac{\alpha}{1+\alpha\mathbf{r}^2}\mathbf{r}\mathbf{r}^{\mathrm{T}}, \tag{13}$$

gives

$$\begin{aligned}\chi^2 &= (\mathbf{x}(\mathbf{p})-\mathbf{r})^{\mathrm{T}}\left[\mathbf{1}-\frac{(\Delta S/S)^2}{1+(\Delta S/S)^2\mathbf{r}^2}\mathbf{r}\mathbf{r}^{\mathrm{T}}\right](\mathbf{x}(\mathbf{p})-\mathbf{r}) \\ &= \frac{\left[\mathbf{x}(\mathbf{p})-\mathbf{r}\right]^2+(\Delta S/S)^2\left[(\mathbf{x}(\mathbf{p})-\mathbf{r})^2\mathbf{r}^2-((\mathbf{x}(\mathbf{p})-\mathbf{r})\cdot\mathbf{r})^2\right]}{1+(\Delta S/S)^2\mathbf{r}^2}.\end{aligned} \tag{14}$$

Finally, the fact that

$$(\mathbf{x}(\mathbf{p})-\mathbf{r})^2\mathbf{r}^2-((\mathbf{x}(\mathbf{p})-\mathbf{r})\cdot\mathbf{r})^2 = \mathbf{x}^2(\mathbf{p})\mathbf{r}^2-(\mathbf{x}(\mathbf{p})\cdot\mathbf{r})^2 \tag{15}$$

makes the numerator of Eq. (14) the same as that of Eq. (8). This means that $\chi^2_{n_0}$ and χ^2 differ only by the interchange of calculated and measured values ($\mathbf{x}(\mathbf{p})\leftrightarrow\mathbf{r}$) in the denominator. Near a solution, these vectors approach each other, so that the difference between the two chi-square values is minimal. Thus, although they may give different results away from the solution point, they are guaranteed to be similar near the chi-square minimum.

<u>Marginalized Distributions</u>

A crucial point raised by Ken (and also discussed in Don's book) is the fact that when a distribution contains "non-essential" parameters, such as the normalization parameter n in χ^2_{EDA} [Eq. (1)], the proper distribution to use for the remaining parameters ($\mathbf{p}$) is "marginalized" (integrated) over n. In this case, it becomes

$$\tilde{P}(\mathbf{p}) = \int_0^\infty \exp[-\tfrac{1}{2}\chi^2_{\mathrm{EDA}}(\mathbf{p},n)]dn. \tag{16}$$

Because of the quadratic dependence of $\tilde{\chi}^2_{\mathrm{EDA}}$ on n expressed by Eq. (2), this integral can be done analytically to obtain

$$\tilde{P}(\mathbf{p}) = \exp\left[-\frac{4A(\mathbf{p})C-B^2(\mathbf{p})}{8A(\mathbf{p})}\right]\sqrt{\frac{\pi}{2A(\mathbf{p})}}\,\mathrm{erfc}\left[\frac{B(\mathbf{p})}{\sqrt{8A(\mathbf{p})}}\right], \tag{17}$$

which, in view of Eq. (5), can be written as

$$\tilde{P}(\mathbf{p}) = \exp\left[-\tfrac{1}{2}\chi^2_{n_0}(\mathbf{p})\right]\sqrt{\frac{\pi}{2A(\mathbf{p})}}\,\mathrm{erfc}\left[\frac{B(\mathbf{p})}{\sqrt{8A(\mathbf{p})}}\right] \tag{18}$$

Near a solution, we expect $B(\mathbf{p}_0)\approx -2A(\mathbf{p}_0)$, with the result that the argument of the complimentary error function is $\sim -\sqrt{A_0/2}$, where $A_0 \equiv A(\mathbf{p}_0)$. Now A_0 is likely to be a rather large number, since the calculated values of the observables should be much larger than their experimental errors. Considering that erfc($-x$) is a constant for $x \geq 2$, it can be neglected in the marginalized probability function, and thus to a good approximation,

$$P_m(\mathbf{p}) = \exp\left[-\tfrac{1}{2}\chi^2_{n_0}(\mathbf{p})\right], \quad (19)$$

and because of the arguments given at the end of the last section, one expects the distribution given by $\exp[-\tfrac{1}{2}\chi^2]$ to be similar near $\mathbf{p}=\mathbf{p}_0$.