Averaging Data with Correlations

Orin Dahl
April 1, 1996

I. Introduction

The averaging and fitting procedures in RPP at present ignore correlations in the input data. However, in some cases such correlation can be very important.

For example, the CP violation parameter ϕ_{+-}, the $K_L^0 - K_S^0$ mass difference Δm, and the K_S^0 mean life τ_s are all correlated. Some measurements from this example are:

1. ADLER 92B
   $\phi_{+-} = 42.3 \pm 4.6$
2. CULLEN 70
   $\Delta m = 0.542 \pm 0.006$
3. GIBBONS 93
   $\phi_{+-} = 42.21 \pm 0.9 + 189(\Delta m - 0.5286) - 460(\tau_s - 0.8922)$
4. CARITHERS 75
   $\phi_{+-} = 45.5 \pm 2.8 + 244(\Delta m - 0.5348)$
5. GEWENIGER 74B
   $\tau_s = 0.8937 \pm 0.0048$

where ϕ_{+-} is in units of degrees
Δm is in units of $10^{10}\hbar^{-1}$
τ_s is in units of $10^{10}s$

We may easily include the correlations in a χ^2 fit by writing

$$\chi^2 = \sum_i (f - m^i) W^i (f - m^i)$$

and then minimizing χ^2 with respect to the vector f. Here
m^i = measurement vector for experiment i
W^i= weight matrix (inverse of error matrix) for experiment i
f = fitted vector.

If an experiment does not measure some of the parameters, the weight for that parameter is zero and the "measurement" value is irrelevant. In our example, GEWENIGER 74B is

$$m^5 = \begin{pmatrix} \varphi_{+-} \\ \Delta m \\ \tau_s \end{pmatrix} = \begin{pmatrix} * \\ * \\ 0.8937 \end{pmatrix} \qquad W^5 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/0.0048^2 \end{pmatrix}.$$

(The Quantities marked $*$ are irrelevant.)

However, it is not obvious how to write the weight matrix for experiments such as GIBBONS 93 or CARITHERS 75, since the measurement of φ_{+-} in these experiments is not independent, but is dependent upon the values of Δm and τ_s.

II. Weight matrix for dependent measurements

Dependent measurements such as GIBBONS 93 and CARITHERS 75 are normally written in the form

$$f_1 = m_1 \pm \sigma + \sum_{k=2}^{n} A_k (f_k - m_k) \tag{2.1}$$

where m_k are the measurements of n variables,
f_k are the adjusted values of n variables.
$A_k = \partial m_1/\partial m_k$ $k = 2, n$ are known constants, and
σ is the error in m_1 with the remaining variables held constant.

We wish to determine the weight matrix for this type of measurement. The quantity that is really measured is

$$x_1 = m_1 - \sum_{k=2}^{n} A_k m_k. \tag{2.2}$$

Note that σ is also the error in x_1

$$\langle \delta x_1^2 \rangle = \sigma^2. \tag{2.3}$$

For convenience, let us define A_1 to be

$$A_1 \equiv -1. \tag{2.4}$$

Then we may write x_1 as

$$x_1 = -\sum_{k=1}^{n} A_k m_k. \tag{2.5}$$

Now consider a set of n independent orthogonal variables x_k $k = 1, n$ where x_1 is the variable just defined, and $x_2 \ldots x_n$ are all unmeasured variables and are functions of the several m_k's. Let the weight matrix for this set of variables be U.
We must have

$$U_{ij} = 0 \qquad i \neq j \tag{2.6}$$

since the variables are orthogonal and uncorrelated. Also

$$U_{kk} = 0 \qquad k = 2, n \tag{2.7}$$

since these x_k's are unmeasured. Thus

$$U = \begin{pmatrix} 1/\sigma^2 & 0 & 0 & \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \\ & \vdots & & \end{pmatrix} \tag{2.8}$$

or

$$U_{k\ell} = \frac{1}{\sigma^2} \delta_{k1} \delta_{1\ell}. \tag{2.9}$$

Now transform U back to the original $m_2, m_3 \cdots$ coordinates. Let the weight matrix in the m_1 coordinates be W. The coordinate transformation from $m_1, m_2, m_3 \cdots$ to $x_1, x_2, x_3 \cdots$ is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix} = R \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} \partial x_1/\partial m_1 & \partial x_1/\partial m_2 & \partial x_1/\partial m_3 & \\ \partial x_2/\partial m_1 & \partial x_2/\partial m_2 & \partial x_2/\partial m_3 & \cdots \\ \partial x_3/\partial m_1 & \partial x_3/\partial m_2 & \partial x_3/\partial m_3 & \\ & \vdots & & \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \end{pmatrix} \tag{2.10}$$

or

$$x_1 = R_{ij} m_j \tag{2.11}$$

with

$$R_{ij} = \partial x_i / \partial m_j. \tag{2.12}$$

Thus

$$R_{1j} = -A_j. \tag{2.13}$$

Since we have not specified the exact form of x_k $k = 2, n$, we cannot write down the exact form of R_{ij} $i = 2, n$. As we shall see, this does not matter; these terms get multiplied by a weight of zero.

Ignore for the moment the fact that $\langle \delta x_i \delta x_j \rangle$ is infinite and simply think of these terms as approaching infinity as a limit. Then we have

$$U_{ij}^{-1} = \langle \delta x_i \delta x_j \rangle \tag{2.14}$$

and

$$W_{ij}^{-1} = \langle \delta m_i \delta m_j \rangle. \tag{2.15}$$

Then

$$U^{-1} = R W^{-1} \tilde{R} \tag{2.16}$$

$$U = \tilde{R}^{-1} W R^{-1} \tag{2.17}$$

$$W = \tilde{R} U R \tag{2.18}$$

Ok, we now see that in our case we have

$$W_{ij} = R_{ki} U_{k\ell} R_{\ell j} \tag{2.19}$$

but since

$$U_{k\ell} = \frac{1}{\sigma^2} R_{1k} R_{1\ell} \tag{2.20}$$

we get

$$W_{ij} = \frac{1}{\sigma^2} R_{1i} R_{1j} \tag{2.21}$$

or

$$W_{ij} = \frac{1}{\sigma^2} Ai Aj \tag{2.22}$$

i.e.,

$$W = \frac{1}{\sigma^2} \begin{pmatrix} A_1^2 & A_1A_2 & A_1A_3 & \\ A_2A_1 & A_2^2 & A_2A_3 & \cdots \\ A_3A_1 & A_3A_2 & A_3^2 & \\ & \vdots & & \end{pmatrix} \tag{2.23}$$

with $A_1 = -1$. The matrix W is, of course, singular. Thus, in our example GIBBONS 93 is

$$m^3 = \begin{pmatrix} \varphi_{+-} \\ \Delta m \\ \tau_s \end{pmatrix} = \begin{pmatrix} 42.21 \\ 0.5286 \\ 0.8922 \end{pmatrix} \quad W^3 = \frac{1}{0.9^2} \begin{pmatrix} 1 & -189 & 460 \\ -189 & 189^2 & 189 * (-460) \\ 460 & 189 * (-460) & (-460)^2 \end{pmatrix}$$

and CARITHERS 75 is

$$m^4 = \begin{pmatrix} \varphi_{+-} \\ \Delta m \\ \tau_s \end{pmatrix} = \begin{pmatrix} 45.5 \\ 0.5348 \\ * \end{pmatrix} \quad W^4 = \frac{1}{2.8^2} \begin{pmatrix} 1 & -224 & 0 \\ -224 & 224^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

III. Forming the averages

To form the averages we use a standard weighted least-squares procedure with correlations. We are given

m^i = measurement vector for experiment i
W^i = weight matrix (inverse of error matrix) for experiment i

We wish to find
f = fitted vector
E = error matrix for the fitted vector f_i
We shall use matrix notation for the individual vectors with explicit sums over the experiments i. Now we have

$$\chi^2 = \sum_i (f - m^i) W^i (f - m^i). \tag{3.1}$$

To find the vector f, we minimize χ^2 with respect to f. We have

$$0 = \frac{1}{2} \partial\chi^2/\partial f = \sum_i W^i (f - m^i). \tag{3.2}$$

Define

$$G = \sum_i W^i \tag{3.3}$$

$$p = \sum_i W^i m^i; \tag{3.4}$$

then equation (3.2) becomes

$$Gf - p = 0 \tag{3.5}$$

or

$$f = G^{-1} p. \tag{3.6}$$

We also want the error matrix E where

$$E = \langle \delta f \delta f \rangle. \tag{3.7}$$

The measured error matrix for experiment i is given by

$$\langle \delta m^i \delta m^i \rangle = W^{i^{-1}}. \tag{3.8}$$

Let us define

$$D^i = \partial f / \partial m^i. \tag{3.9}$$

Then

$$E = \sum_i D^i W^{i^{-1}} \tilde{D}^i \tag{3.10}$$

where the sum is over the experiments.

From equations (3.4) and (3.6), we see that

$$D^i = G^{-1} W^i \tag{3.11}$$

and thus

$$E = G^{-1} \left(\sum_i W^i W^{i^{-1}} W^i \right) G^{-1}. \tag{3.12}$$

But since

$$\sum_i W^i W^{i^{-1}} W^i = \sum_i W^i = G \tag{3.13}$$

we have

$$E = G^{-1}. \tag{3.14}$$

Thus, we have our procedure. We form G and p using equations (3.3) and (3.4), then find f and E using equations (3.6) and (3.14), and finally form χ^2 using equation (3.1).

IV. Constrained fits

We may also impose constraints between the variables when we form the averages. In this case, we use a standard weighted least squares procedure and introduce Lagrangian multipliers to satisfy the constraints.

As in the previous section, we are given

m^i = measurement vector for experiment i

W^i = weight matrix (inverse of error matrix)for experiment i.

We wish to find

f = fitted vector

E = error matrix for the fitted vector f.

We also have some constraints b (b is a vector) for which we want $b(f) = 0$. We shall assume that the constraints are linear and write

$$b = Bf + b_0 \tag{4.1}$$

where B and b_0 are constants.

If the constraints are not linear, we must iterate. We make a guess as to the values of f, expand the constraints b about those values, and solve to find new (hopefully better) values of f. Then we repeat the process starting with the new values of f.

An obvious first guess is the result of an unconstrained average — or of a constrained fit with the nonlinear constraints ignored.

We shall use matrix notation for the individual vectors and explicitly write the sums over the experiments. We have

$$\chi^2 = \sum_i (f - m^i) W^i (f - m^i) \tag{4.2}$$

which we wish to minimize with respect to the vector f. To include the constraints, we introduce Lagrangian multipliers α and write

$$M = \chi^2 + 2b\alpha. \tag{4.3}$$

We then minimize M with respect to f and α. We find

$$0 = \frac{1}{2}\partial M/\partial f = \sum_i W^i (f - m^i) + \tilde{B}\alpha \tag{4.4}$$

$$0 = \frac{1}{2}\partial M/\partial \alpha = Bf + b_0. \tag{4.5}$$

Define

$$G \equiv \sum_i W^i \tag{4.6}$$

and

$$p \equiv \sum_i W^i m^i \tag{4.7}$$

and equation (4.4) becomes

$$Gf - p + \tilde{B}\alpha = 0. \tag{4.8}$$

Define

$$q \equiv G^{-1} p \tag{4.9}$$

and multiply equation (4.8) by G^{-1}. We then get

$$f = q - G^{-1}\tilde{B}\alpha. \tag{4.10}$$

We multiply equation (4.10) by B to get

$$Bf = Bq - BG^{-1}\tilde{B}\alpha \tag{4.11}$$

then define

$$H \equiv BG^{-1}\tilde{B} \tag{4.12}$$

and substitute equations (4.12) and (4.5) into equation (4.11) to get

$$\alpha = H^{-1}(b_0 + B_q). \tag{4.13}$$

Thus, we have our solution. We use equation (4.13) to solve for α and then substitute it into equations (4.10) to find f. We also wish to get the fitted error matrix E where

$$E \equiv \langle \delta f \delta f \rangle. \tag{4.14}$$

We also have that the measured error matrix for experiment i is

$$\langle \delta m^i \delta m^i \rangle = W^{i^{-1}}. \tag{4.15}$$

Let us define

$$D^i = \partial f / \partial m^i. \tag{4.16}$$

Then

$$E = \sum_i D^i W^{i^{-1}} \tilde{D}^i \tag{4.17}$$

where the sum is over the experiments. Now from equations (4.10) and (4.9)

$$D^i = G^{-1} \partial p / \partial m^i - G^{-1} \tilde{B} \partial \alpha / \partial m^i \tag{4.18}$$

and from equation (4.13)

$$\partial \alpha / \partial m^i = H^{-1} B G^{-1} \partial p / \partial m^i. \tag{4.19}$$

Since from equation (4.7)

$$\partial p / \partial m^i = W^i \tag{4.20}$$

we have that

$$D^i = (1 - G^{-1} \tilde{B} H^{-1} B) G^{-1} W^i \tag{4.21}$$

and

$$E = \left(1 - G^{-1} \tilde{B} H^{-1} B\right) G^{-1} \left(\sum_i W^i W^{i^{-1}} W^i \right) G^{-1} \left(1 - \tilde{B} H^{-1} B G^{-1}\right). \tag{4.22}$$

Since

$$\sum_i W^i W^{i^{-1}} W^i = \sum_i W^i = G \tag{4.23}$$

we have

$$E = (1 - G^{-1} \tilde{B} H^{-1} B) G^{-1} (1 - \tilde{B} H^{-1} B G^{-1}) \tag{4.24}$$

or

$$E = G^{-1} - G^{-1} \tilde{B} H^{-1} B G^{-1}. \tag{4.25}$$

We finally use equation (4.2) to find χ^2.

V. Summary

We have

m^i = measurement vector for experiment i
W^i = weight matrix for experiment i.

We wish to find

f = fitted vector

E = error matrix for the fitted vector f

χ^2 for the average.

We may also have some constraints of the form

$$b = Bf + b_0 \tag{5.1}$$

where B and b_0 are constants.

First form

$$G = \sum_i W^i \tag{5.2}$$

$$p = \sum_i W^i m^i. \tag{5.3}$$

Then invert G to get

$$G^{-1}$$

and

$$q = G^{-1} p. \tag{5.4}$$

If we have no constraints

$$\begin{aligned} f &= q \\ E &= G^{-1}. \end{aligned} \tag{5.5}$$

If we do have some constraints, we form

$$H = BG^{-1}\tilde{B} \tag{5.6}$$

$$\alpha = H^{-1}(b_0 + Bq), \tag{5.7}$$

and the fitted vector and error are

$$f = q - G^{-1}\tilde{B}\alpha \tag{5.8}$$

$$E = G^{-1} - G^{-1}\tilde{B}H^{-1}BG. \tag{5.9}$$

In either case, we form χ^2 by

$$\chi^2 = \sum_i (f - m^i) W^i (f - m^i). \tag{5.10}$$