Correlations in Statistical and Systematic Errors

R. Cahn

December 2000

Abstract

These are pedagogical notes devoid of any proper references. We discuss correlations from the point of view of maximum likelihood methods. We show the relationship between the matrix of second derivatives of the likelihood function and the correlation matrix. We address the problem of finding the shifts in one set of variables $\alpha_i, \alpha_j, ..$ induced by changes in another set, $\alpha_a, \alpha_b, ...$ when maximum likelihood is used. We illustrate this with a semi-practical example involving a double gaussian fit to $B\overline{B}$ oscillations. We find a general formula for the expected error in α_i caused by possibly correlated errors in the α_a's.

Let us indicate the log-likelihood by $\mathcal{L}$:

$$\mathcal{L} = \sum_i \log f(x_i, \alpha)$$

where α is a collection of parameters . The components of α may be fixed or may be determined by maximum likelihood. In the vicinity of the maximum,

$$\mathcal{L} = \mathcal{L}_0 + \tfrac{1}{2} \sum_{i,j} \frac{\partial^2 \mathcal{L}}{\partial \alpha_i \partial \alpha_j} \delta\alpha_i \delta\alpha_j + \ldots$$

where the partial derivatives are evaluated at the minimum of the log-likelihood. If we deal with the expectations of the likelihood rather than its value given by a certain dataset, we have similar formulae, with $\mathcal{L} \to < \mathcal{L} >$.

The distribution of fluctuations $\delta\alpha$ is

$$dP = \delta\alpha_1 \ldots \delta\alpha_p (\det M)^{1/2} (2\pi)^{-p/2} e^{-\frac{1}{2} M_{ij} \delta\alpha_i \delta\alpha_j}$$

where

$$M_{ij} = \frac{\partial^2 \langle \mathcal{L} \rangle}{\partial \alpha_i \partial \alpha_j}$$

and where we have fixed the overall normalization. Now consider the correlation

$$\begin{aligned}\langle\delta\alpha_k\delta\alpha_l\rangle &= -\frac{\partial}{\partial M_{kl}}\ln\int\delta\alpha_1\ldots\delta\alpha_p e^{-\frac{1}{2}M_{ij}\delta\alpha_i\delta\alpha_j} \\ &= -\frac{\partial}{\partial M_{kl}}\ln(\det M)^{-1/2} = (M^{-1})_{kl}\end{aligned}$$

where I've used the symmetry of M throughout.

Now we want to connect the correlation defined above with the shift in a parameter caused by fixing some other parameter to an incorrect value.

Suppose we fix $\delta\alpha_0$ to a non-zero value. Now the expected values of the other parameters will no longer be zero, but will be determined by maximizing the log-likelihood:

$$\delta\alpha_0 M_{0j} + \delta\alpha_i M_{ij} = 0; \qquad j \neq 0$$

Now let $\overline{M}$ be the matrix without the zeroth row and column. Then we can find $\delta\alpha_k$ as

$$\delta\alpha_k = -\delta\alpha_0 M_{0j}(\overline{M}^{-1})_{jk}$$

It suffices to find $\overline{M}^{-1}$. We assert that

$$(\overline{M}^{-1})_{jk} = (M^{-1})_{jk} - \frac{(M^{-1})_{j0}(M^{-1})_{0k}}{(M^{-1})_{00}}$$

This can be confirmed using

$$\begin{gathered}(M^{-1})_{jk}\overline{M}_{kl} + (M^{-1})_{j0}M_{0l} = \delta_{jl} \\ (M^{-1})_{0k}\overline{M}_{kl} + (M^{-1})_{00}M_{0l} = 0\end{gathered}$$

Combining these results we find

$$\begin{aligned}\delta\alpha_k &= -\delta\alpha_0 M_{0j}[(M^{-1})_{jk} - \frac{(M^{-1})_{j0}(M^{-1})_{0k}}{(M^{-1})_{00}}] \\ &= \delta\alpha_0\frac{(M^{-1})_{0k}}{(M^{-1})_{00}}\end{aligned}$$

If we introduce the expected uncertainties in each of the α_is

$$\sigma_i^2 = (M^{-1})_{ii}$$

we can write

$$\frac{\delta\alpha_k}{\sigma_k} = \frac{\langle\delta\alpha_k\delta\alpha_0\rangle}{\sigma_k\sigma_0}\frac{\delta\alpha_0}{\sigma_0}$$

Now let us generalize this by asking what are the values of $\delta\alpha_i, \delta\alpha_j, \ldots$ that maximize the log-likelihood subject to the constraint that $\delta\alpha_a$, $a = a, b, \ldots$ are fixed. Then we have

$$\delta\alpha_a M_{aj} + \delta\alpha_i M_{ij} = 0;$$

where the j are the indices of the parameters free to vary while the a are the indices of the fixed parameters.

The problem now is to invert $\overline{M}$ where this is the matrix restriced to the j indices. Once this is done, we have

$$\delta\alpha_k = -\delta\alpha_a M_{aj}\overline{M}^{-1}_{jk}$$

Now

$$\begin{aligned}(M^{-1})_{jk}\overline{M}_{kl} + (M^{-1})_{ja}M_{al} = \delta_{jl}\\ (M^{-1})_{ak}\overline{M}_{kl} + (M^{-1})_{ab}M_{bl} = 0\end{aligned}$$

It follows that

$$M_{cl} = -(\tilde{M}^{-1})^{-1}_{cb}M^{-1}_{bj}\overline{M}_{jl}$$

where $\tilde{M}^{-1}$ is the part of the full matrix M^{-1} that is labeled just by the $a, b, \ldots$ indices. In particular, $(\tilde{M}^{-1})^{-1} \neq M$. Inserting this into Eq. 1

$$(M^{-1})_{jk}\overline{M}_{kl} - (M^{-1})_{ja}(\tilde{M}^{-1})^{-1}_{ab}M^{-1}_{bk}\overline{M}_{kl} = \delta_{jl}$$

from which we see that

$$\overline{M}^{-1}_{jk} = M^{-1}_{jk} - M^{-1}_{ja}(\tilde{M}^{-1})^{-1}_{ab}M^{-1}_{bk}$$

Now just substitute this back in:

$$\begin{aligned}\delta\alpha_k &= -\delta\alpha_a M_{aj}[M^{-1}_{jk} - M^{-1}_{jc}(\tilde{M}^{-1})^{-1}_{cb}M^{-1}_{bk}] \\ &= \delta\alpha_a M_{ab}M^{-1}_{bk} + \delta\alpha_a(\tilde{M}^{-1})^{-1}_{ab}M^{-1}_{bk} - \delta\alpha_a M_{ab}M^{-1}_{bk} \\ &= \delta\alpha_a(\tilde{M}^{-1})^{-1}_{ab}M^{-1}_{bk}\end{aligned}$$

We can write this as

$$\frac{\delta\alpha_k}{\sigma_k} = \frac{\delta\alpha_a}{\sigma_a}(\sigma_a\langle\delta\alpha_a\delta\alpha_b\rangle^{-1}\sigma_b)\frac{\langle\delta\alpha_b\delta\alpha_k\rangle}{\sigma_b\sigma_k}$$

where we must remember that $\langle\delta\alpha_a\delta\alpha_b\rangle^{-1}$ refers to the inverse in the space of the fixed parameters only. If we indicate the correlation matrix by ρ:

$$\rho_{mn} = \frac{\langle\delta\alpha_m\delta\alpha_n\rangle}{\sigma_m\sigma_n}$$

and the correlation matrix restricted to $a, b, ...$ by $\tilde{\rho}$ we have simply

$$\frac{\delta\alpha_k}{\sigma_k} = \frac{\delta\alpha_a}{\sigma_a}\tilde{\rho}^{-1}_{ab}\rho_{bk}$$

where the implicit sum is over a and b.

As a concrete example consider $B\overline{B}$ oscillations in which we seek a fit only to parameters of the resolution, fixing at their correct values the lifetime, Δm_d, and the dilution D. To make things relatistic, 10% of the events are taken to be charged B, though no other background is included. Using a double Gaussian to fit the data which are taken to be in fact given by a 80% narrow Gaussian of width $\Sigma_N = 100$ μm and 20% a wide Gaussian of width $\Sigma_W = 300$ μm, we consider three fits. In one the two widths and the fraction f in the narrow Gaussian are all taken to be free. In a second the fraction is frozen at its true value. Finally, both the wide Gaussian's width and the fraction are fixed at the true values.

Asymptotically the elements of the correlation matrix are proportional to $1/N$ where N is the number of events. Using Minuit we find the correlation matrices are given, dropping the $1/N$, by

$$\begin{aligned}\rho^{(1)} &= (1) \\ \rho^{(2)} &= \begin{pmatrix} 1 & -0.51 \\ -0.51 & 1 \end{pmatrix}\end{aligned}$$

$$\rho^{(3)} = \begin{pmatrix} 1 & 0.78 & 0.90 \\ 0.78 & 1 & 0.95 \\ 0.90 & 0.95 & 1 \end{pmatrix}$$

In the final matrix the order of the rows and colums is Σ_N, Σ_W, f At the same time we find for the expected uncertainty with only the narrow width being free

$$\sigma(\Sigma_N) = \frac{770\ \mu\text{m}}{\sqrt{N}}$$

With both widths free we find

$$\begin{aligned} \sigma(\Sigma_N) &= \frac{900\ \mu\text{m}}{\sqrt{N}} \\ \sigma(\Sigma_W) &= \frac{2270\ \mu\text{m}}{\sqrt{N}} \end{aligned}$$

and with all three variables free

$$\begin{aligned} \sigma(\Sigma_N) &= \frac{2050\ \mu\text{m}}{\sqrt{N}} \\ \sigma(\Sigma_W) &= \frac{7050\ \mu\text{m}}{\sqrt{N}} \\ \sigma(f) &= \frac{12.8}{\sqrt{N}} \end{aligned}$$

Using these results we can address questions of systematic errors. For example, suppose we are doing a fit leaving free only the variable Σ_N. Suppose that we've got the value $f = 0.8$ correct but have incorrectly taken $\Sigma_W = 400\ \mu$m. This will lead to a shift in Σ_N, which we calculate from $\rho^{(2)}$:

$$\begin{aligned} \frac{\delta\Sigma_N}{900} &= -0.51\frac{100\ \mu\text{m}}{2270} \\ \delta\Sigma_N &= -20\mu\text{m} \end{aligned}$$

Using the correlation matrix $\rho^{(3)}$ we can find the shift in Σ_N arising from simultaneous errors in Σ_W and f. As a check, let's calculate the error due to a shift of 100 μm in Σ_W and a zero shift in f, i.e. exactly the same circumstance as above, but now looked as a three-variable problem.

We need to calculate the inverse of the correlation matrix restricted to the variables Σ_W and f:

$$\begin{aligned} \tilde{\rho} &= \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix} \\ \tilde{\rho}^{-1} &= \begin{pmatrix} 10.25 & -9.74 \\ -9.74 & 10.25 \end{pmatrix} \end{aligned}$$

We need to insert these elements into

$$\frac{\delta\Sigma_N}{\sigma(\Sigma_N)} = \frac{\delta\Sigma_W}{\sigma(\Sigma_W)} \left(\tilde{\rho}^{-1}_{WW} \rho^{(3)}_{WN} + \tilde{\rho}^{-1}_{Wf} \rho^{(3)}_{fN} \right)$$

where we have dropped the contribution from $\delta f = 0$. Explicitly

$$\begin{aligned} \frac{\delta\Sigma_N}{2050} &= \frac{100\ \mu\mathrm{m}}{7050} (10.25 \cdot 0.78 - 9.74 \cdot 0.90) \\ \delta\Sigma_N &= -22\ \mu\mathrm{m} \end{aligned}$$

The agreement is adequate, since errors were introduced both by rounding off this calculate and in the Minuit calculation of the correlation matrices.

Suppose we treat both Σ_N and f as free parameters and ask how they are modified by a shift of 100 μm in Σ_W. Here we have the simple calculations:

$$\begin{aligned} \frac{\delta\Sigma_N}{\sigma_N} &= \rho_{NW} \frac{\delta\Sigma_W}{\sigma_W} \\ \frac{\delta\Sigma_N}{2050} &= 0.78 \frac{100\ \mu\mathrm{m}}{7050} \\ \delta\Sigma_N &= 23\ \mu\mathrm{m} \\ \frac{\delta f}{12.8} &= 0.95 \cdot \frac{100}{7050} \\ \delta f &= 0.17 \end{aligned}$$

Note that increasing the wide Gaussian led to an increase in the narrow Gaussian, but also to a big depletion in the wide Gaussian's population. In fact, we see that our shift of 100 μm is too large to be considered a small perturbation since the shift in f is very large.

An important point here is that a change in Σ_W of course led to a change in f. If we simply used $\rho^{(3)}$ and accumulated the effects of the shifts in Σ_W

and f, the resulting shifts in these variables would not be what we intended. This is the reason that when we postulate simultaneous shifts in more than one parameter the formula we use is somewhat complicated: it contains the quantity $\tilde{\rho}^{-1}$.

In the event that we are given the correlations of the systematic errors in the parameters $\delta\alpha_a, \delta\alpha_b, \ldots$, we compute the resulting error in α_k as

$$\frac{\langle \delta\alpha_k^2 \rangle}{\sigma_k^2} = \frac{\langle \delta\alpha_a \delta\alpha_c \rangle}{\sigma_a \sigma_c} \tilde{\rho}_{ab}^{-1} \rho_{bk} \tilde{\rho}_{cd}^{-1} \rho_{dk}$$

If the systematic errors are uncorrelated,

$$\langle \delta\alpha_a \delta\alpha_c \rangle = \Delta_a^2 \delta_{ac}$$

then

$$\frac{\langle \delta\alpha_k^2 \rangle}{\sigma_k^2} = \frac{\Delta_a^2}{\sigma_a^2} (\tilde{\rho}_{ab}^{-1} \rho_{bk})^2$$

where there are sums over a and b.

Even though the systematic errors are assumed to be uncorrelated, correlations between the variables are properly included.