DRAFT 0.6

Glen Cowan
21 May 2019

Notes on Correlated Errors

This note describes how to prepare the ingredients needed to combine measurements in a manner that treats correlated systematic uncertainties into account. The basic picture is described in Sec. 1 and Sec. 3 provides further details on how this can be implemented.

1 Basic picture

As discussed in Ref. [1], the phrase "correlated systematics" is often taken to mean the situation where a nuisance parameter affects multiple measurements in a coherent way. Suppose, for example, that the expectation values $E[y_i]$ of measured quantities y_i with $i = 1, \ldots, L$ are functions $\varphi_i(\boldsymbol{\mu}, \boldsymbol{\theta})$ of parameters of interest $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ and nuisance parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. Suppose further that the nuisance parameters are defined such that for $\boldsymbol{\theta} = 0$ the y_i are unbiased measurements of the nominal model $\varphi_i(\boldsymbol{\mu})$. Expanding φ_i to first order in $\boldsymbol{\theta}$ therefore gives

$$E[y_i] = \varphi_i(\boldsymbol{\mu}, \boldsymbol{\theta}) \approx \varphi_i(\boldsymbol{\mu}) + \sum_{j=1}^{N} R_{ij}\theta_j \ , \tag{1}$$

where the factors $R_{ij} = \partial\varphi_i / \partial\theta_j|_{\boldsymbol{\theta}=0}$ determine how much θ_j biases the measurement y_i.

Suppose that the R_{ij} are known, either from symmetry (e.g., a particular θ_j could be known to contribute equally to all of the y_i) or they are determined using a Monte Carlo simulation. As before suppose one has a set of independent Gaussian-distributed control measurements u_j used to constrain the nuisance parameters, with mean values θ_j and standard deviations σ_{u_j}. One can define the total bias of measurement y_i as

$$b_i = \sum_{j=1}^{N} R_{ij}\theta_j \ . \tag{2}$$

and an estimator for b_i is

$$\hat{b}_i = \sum_{j=1}^{N} R_{ij}u_j \ . \tag{3}$$

These estimators of the biases are correlated. As the control measurements are assumed independent, and therefore $\mathrm{cov}[u_k, u_l] = V[u_k]\delta_{kl}$, the covariance of the bias estimators is

$$U_{ij} = \mathrm{cov}[\hat{b}_i, \hat{b}_j] = \sum_{k=1}^{N} R_{ik}R_{jk}V[u_k] \ . \tag{4}$$

It is in the sense described here that the proposed model is capable of treating correlated systematic uncertainties. That is, although the control measurements u_i are independent they result in a nondiagonal covariance for the estimated biases of the measurements.

The overall scale of the R_{ij} for a given j can be absorbed into the definition of θ_j, and the corresponding uncertainty is thus reflected in the standard deviation σ_{u_j} of its estimate. Ratios of the R_{ij}, e.g., R_{ij}/R_{kj}, reflect the relative influence of θ_j on y_i and y_k. If these ratios are uncertain and thus should not be treated as fixed constants, one can introduce further nuisance parameters, which are constrained with yet more control measurements. To the extent that the assumed variances of these control measurements are themselves uncertain one could treat them as adjustable parameters with gamma-distributed estimates, just as for other nuisance parameters in the model.

If one is given the $L \times L$ covariance matrix U it is possible provided certain conditions are satisfied to find the N variances $V[u_i]$ and thus construct the model entirely in terms of the independent measurements u_i. For example, suppose

$$y_1 \sim \mathrm{Gauss}\left(\mu + \theta_1 + \theta_3, \sigma_{y_1}\right) \ , \tag{5}$$

$$y_2 \sim \mathrm{Gauss}\left(\mu + \theta_2 + \theta_3, \sigma_{y_2}\right) \ . \tag{6}$$

In this case matrix R is

$$R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \tag{7}$$

The covariance matrix U of the bias estimates are

$$U_{11} = V[\hat{b}_1] = V[u_1] + V[u_3] \ , \tag{8}$$

$$U_{22} = V[\hat{b}_2] = V[u_2] + V[u_3] \ , \tag{9}$$

$$U_{12} = U_{21} = \mathrm{cov}[\hat{b}_1, \hat{b}_2] = \mathrm{cov}[u_1 + u_3, u_2 + u_3] = V[u_3] \ . \tag{10}$$

These equations can be solved for

$$V[u_1] = U_{11} - U_{12} \ , \tag{11}$$

$$V[u_2] = U_{22} - U_{12} \ , \tag{12}$$

$$V[u_3] = U_{12} \ . \tag{13}$$

That is, given a systematic covariance matrix U and the information on what nuisance parameters are common to what measurements, it can be possible to solve for the variances $V[u_j]$ of an independent set of control measurements u_j.

From Eqs. (11)-(13) one can see that the covariance matrix that emerges from this model has certain properties that go beyond its minimal requirement of being positive semi-definite. Since all of the variances $V[u_i]$ must be non-negative, one must have $U_{11} \geq U_{12}$ and also $U_{22} \geq U_{12}$. If the elements of U are assigned using Eqs. (8)-(10), then these inequalities are satisfied by construction.

Suppose, on the other hand, one were to start by writing down the the matrix U as $U_{ij} = \delta_{ij}\sigma_i^2 + (1 - \delta_{ij})\rho\sigma_i\sigma_j$, and then choose "by hand" values for ρ, σ_1 and σ_2. If, for example, $\rho = 1$ and $\sigma_1 \neq \sigma_2$, then Eqs. (11) and (12) say that one of $V[u_1]$ or $V[u_2]$ will be assigned a negative value. So that covariance matrix U could not have come from the model described above.

2 Special cases

In this section several specific cases are investigated, including that of a fully correlated uncertainty in Sec. 2.1 and two-point systematics in Sec. 2.2.

2.1 Fully correlated uncertainty

An interesting special case is that of L independent Gaussian measurements $\mathbf{y} = (y_1, \ldots, y_L)$ with expectation values

$$E[y_i] = \mu + \theta \,, \qquad i = 1, \ldots, L \,, \tag{14}$$

where μ is the parameter of interest and the single bias parameter θ is common to all of the measurements. In the notation of Sec. 1 this corresponds to having $\varphi(\mu) = \mu$, i.e., the fit function corresponds to estimating a common mean with $R_{ij} = 1$ for all $i = 1, \ldots, L$ and for $j = 1$ only, since there is only $N = 1$ nuisance parameter θ. We also have a single indendent measurement u with mean θ and standard deviation σ_u.

If σ_u is known then the log-likelihood function is (cf. Eq. (53) of Ref. [1]),

$$\ln L(\mu, \theta) = -\frac{1}{2} \sum_{i=1}^{L} \frac{(y_i - \mu - \theta)^2}{\sigma_{y_i}^2} - \frac{1}{2} \frac{(u - \theta)^2}{\sigma_u^2} \,. \tag{15}$$

Or if one treats the variance σ_u^2 as a free parameter with a gamma-distributed estimate v, then the profile likelihood is found to be (see Ref. [1], Eq. (55)),

$$\ln L'(\mu, \theta) = -\frac{1}{2} \sum_{i=1}^{L} \frac{(y_i - \mu - \theta)^2}{\sigma_{y_i}^2} - \frac{1}{2} \left(1 + \frac{1}{2r^2} \right) \ln \left[1 + 2r^2 \frac{(u - \theta)^2}{v} \right] \,. \tag{16}$$

Here r is the relative "error-on-the-error" parameter defined by Eq. (9) of Ref. [1]. In the limit $r \to 0$, Eq. (16) reduces to Eq. (15) with the replacement $v \to \sigma_u 2$.

Assuming for the moment that an appropriate value of r has been chosen for the error-on-the-error parameter, we can determine the estimators for μ and θ by setting the corresponding derivatives of $\ln L'$

$$\frac{\partial \ln L'}{\partial \mu} = \sum_{i=1}^{N} \frac{y_i - \mu - \theta}{\sigma_{y_i}^2} \,, \tag{17}$$

$$\frac{\partial \ln L'}{\partial \theta} = \sum_{i=1}^{N} \frac{y_i - \mu - \theta}{\sigma_{y_i}^2} + \frac{(1 + 2r^2)(u - \theta)}{\sigma_u^2 + 2r^2 (u - \theta)^2} \,, \tag{18}$$

to zero. Solving for μ and θ gives the estimators

$$\hat{\mu} = \frac{\sum_{i=1}^{N} y_i / \sigma_{y_i}^2}{\sum_{i=1}^{N} 1 / \sigma_{y_i}^2} - u \tag{19}$$

$$\hat{\theta} = u \,, \tag{20}$$

where for the actual measurement one would take $u = 0$. The standard deviation of $\hat{\theta}$ is σ_u and for that of $\hat{\mu}$ one finds

$$\sigma_{\hat{\mu}} = \left[\frac{1}{\sum_{i=1}^{N} 1/\sigma_{y_i}^2} + \sigma_u^2 \right]^{1/2} . \tag{21}$$

Superficially this appears to say that the standard deviation $\sigma_{\hat{\mu}}$ is independent of r. But of course we don't know the exact value of σ_u^2. And although v is initially regarded as the estimate of σ_u^2, the maximum-likelihood estimator for σ_u^2 found from the full likelihood is

$$\widehat{\sigma_u^2} = \frac{v}{1 + 2r^2} . \tag{22}$$

so that the estimate of the standard deviation of $\hat{\mu}$ is

$$\hat{\sigma}_{\hat{\mu}} = \left[\frac{1}{\sum_{i=1}^{N} 1/\sigma_{y_i}^2} + \frac{v}{1 + 2r^2} \right]^{1/2} . \tag{23}$$

Thus we find that $\hat{\mu}$ and its true standard deviation are independent of r, but that the maximum likelihood estimate of $\sigma_{\hat{\mu}}$ decreases for increasing r.

2.2 Possible Ansatz for two-point systematics

Another interesting example is where one has L measurements y_i, $i = 1, \ldots, L$ of a parameter μ, whose expectation values are modeled as

$$E[y_i] = \mu + C_i \theta_0 + S_i \theta_i . \tag{24}$$

That is, there is one parameter, θ_0, that is a common contribution to the bias of all of the measurements, scaled in general by a known constant C_i. Each measurement then contains a separate bias contribution through the parameter θ_i, scaled in general by a known constant S_i.

There are thus $N = L + 1$ nuisance parameters $(\theta_0, \theta_1, \ldots, \theta_L)$, and as before we suppose that each is estimated by an independent Gaussian distributed value u_i with standard deviation σ_{u_i}. The matrix R_{ij} as defined through Eq. (1) (but with the index j starting at zero) is therefore

$$R_{ij} = \begin{cases} C_i & i = 1, \ldots, L, \; j = 0, \\ S_i \delta_{ij} & i = 1, \ldots, L, \; j = 1, \ldots, L. \end{cases} \tag{25}$$

The bias of the ith measusrement is therefore $b_i = C_i \theta_0 + S_i \theta_i$, which can be estimated using

$$\hat{b}_i = C_i u_0 + S_i u_i . \tag{26}$$

The bias estimators have covariance

$$\mathrm{cov}[\hat{b}_i, \hat{b}_j] = C_i C_j \sigma_{u_0}^2 + S_i S_j \delta_{ij} \sigma_{u_i}^2 . \tag{27}$$

In this example we will suppose that the constants C_i and S_i are known (often they would be equal to unity), and the required ingredients are estimates of the $N = L + 1$ standard deviations $\sigma_{u_0}, \ldots, \sigma_{u_L}$ and the corresponding error-on-error parameters r_i.

Let us suppose that the nuisance parameters $\theta_0, \ldots, \theta_L$ are introduced into the model to account for a particular class of uncertainty. As an example, one could consider the hadronization uncertainty involved in an estimate of the strong coupling constant α_{s} from distributions of event-shape variables (thrust, maximum jet mass, etc.). Suppose that the only way to roughly estimate the size of the uncertainty is to use two different hadronization models to correct for the effect (e.g., Pythia and Herwig).

Let us suppose following Ref. [1] (cf. Eqs. (12)-(14)), that the best estimate for the result is taken using the average of the two correction procedures and that the difference between the two provides information on the corresponding uncertainty. More precisely, the standard deviation of the bias estimate is taken as the difference divided by $\sqrt{2}$. Using this procedure (or whatever alternative is deemed appropriate) we will suppose that the analyst can assign meaningful values to the standard deviations of the bias estimates $\sigma_{\hat{b}_i}$ for all of the measurements, i.e., $i = 1, \ldots, L$.

3 What is needed in practice

What one needs in practice is a general procedure for constructing the likelihood function for parameters of interest $\boldsymbol{\mu}$ and some set of nuisance parameters. Suppose one is given the probability (density) $P(\mathbf{y}|\boldsymbol{\mu})$, e.g., a multivariate Gaussian with a given covariance matrix, which encodes the statistical errors in the primary measurements $\mathbf{y} = (y_1, \ldots, y_L)$. Let us suppose further that the L measurements are also characterized by an $L \times L$ systematic covariance matrix T, interpreted as relating to potential additive biases in the y_i. We will suppose that this matrix can be written as the sum of two terms, $T = U + W$, where the part U can be related to control measurements of nuisance parameters and W is whatever is left over. For example, the expectation value of y_i may be modeled as

$$E[y_i] = \mu + \beta_i + \eta_i \ , \tag{28}$$

where μ is the parameter of interest and β_i and η_i are two different contributions to the bias. The term β_i can be explicitly related to uncertainties connected to control measurements, i.e., we take

$$\beta_i = \sum_{j=1}^{N} R_{ij} \theta_j \ , \tag{29}$$

where as in Sec. 1 the R_{ij} are known factors. Similar to above we can estimate β_i with

$$\hat{\beta}_i = \sum_{j=1}^{N} R_{ij} u_j \ , \tag{30}$$

where $u_j \sim \mathrm{Gauss}(\theta_j, \sigma_{u_j})$ are independent control measurements with variances $V[u_j] = \sigma_{u_j}^2$. Because there can be θ_j that contribute to the same y_i, the estimators $\hat{\beta}_i$ are correlated with covariance matrix

$$U_{ij} = \mathrm{cov}[\hat{\beta}_i, \hat{\beta}_j] = \sum_{k=1}^{N} R_{ik} R_{jk} V[u_k] \ . \tag{31}$$

In general it may be that the information about the origin of some of the systematic uncertainties is incomplete, i.e., the remaining parts of the bias $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ are not constrained by given control measurements. Rather, the only available information about the η_i comes from the total systematic covariance matrix

$$T = U + W \ . \tag{32}$$

Here U is given by Eq. (31) and the remaining part $W = T - U$ is defined to be whatever is left over. If W is not a positive-definite matrix, then the information supplied is not consistent. One then needs to go back and ensure that one begins with a consistent set of inputs.

We suppose that there are control measurements $\mathbf{z} = (z_1, \ldots, z_N)$ Gaussian distributed about $\boldsymbol{\eta}$ with covariance matrix W. As with the u_j, in the real experiment the z_i would be taken as zero (or in general their best estimates). The likelihood function can be written as

$$\begin{aligned} L(\mu, \boldsymbol{\theta}, \boldsymbol{\eta}) &= P(\mathbf{y}|\mu, \boldsymbol{\theta}, \boldsymbol{\eta}) \\ &\times \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{u_j}} e^{-(u_j - \theta_j)^2 / 2\sigma_{u_j}^2} \\ &\times \frac{1}{(2\pi)^{N/2} |W|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{z} - \boldsymbol{\eta})^T W^{-1} (\mathbf{z} - \boldsymbol{\eta}) \right] \ . \end{aligned} \tag{33}$$

This likelihood function can be generalized to include "errors-on-errors" according to the procedure of Ref. [1] by treating the σ_{u_j} as nuisance parameters. But it is not possible to do this for the portion of the systematic uncertainty attributed to $W = T - U$, since its origin has not been documented, i.e., it is not connected with well-defined control measurements. Nevertheless the procedure above allows one to treat the errors on at least some of the systematic errors. One cannot possibly treat errors on the errors unless one has some information on their origin and therefore the procedure described above may be as good as can be achieved in practice.

References

[1] Glen Cowan, *Statistical Models with Uncertain Error Parameters*, Eur. Phys. J. C (2019) 79:133; arXiv:1809.05778.