

# Statistical considerations

Valerio Bertone

November 8, 2020

## Contents

<b>1</b>	<b>The <math>\chi^2</math> in the presence of correlations</b>	<b>1</b>
<b>2</b>	<b>Additive and multiplicative uncertainties</b>	<b>2</b>
<b>3</b>	<b>Artificial generation of correlated systematics</b>	<b>3</b>
<b>4</b>	<b>Determining the systematic shifts</b>	<b>4</b>
<b>5</b>	<b>Effect of cuts on the <math>\chi^2</math></b>	<b>4</b>
<b>6</b>	<b>Monte Carlo replica generation</b>	<b>6</b>
<b>7</b>	<b>The D’Agostini “bias”</b>	<b>9</b>
7.1	The effect of the $t_0$ prescription . . . . .	14

## 1 The $\chi^2$ in the presence of correlations

Suppose to have an ensemble of  $n$  measurements having the following structure:

$$m_i \pm \sigma_{i,\text{stat}} \pm \sigma_{i,\text{unc}} \pm \sigma_{i,\text{corr}}^{(1)} \pm \cdots \pm \sigma_{i,\text{corr}}^{(k)}, \quad (1)$$

where  $m_i$ , with  $i = 1, \dots, n$ , is the central value of the  $i$ -th measurement,  $\sigma_{i,\text{stat}}$  its (uncorrelated) statistical uncertainty,  $\sigma_{i,\text{unc}}$  its uncorrelated systematic uncertainty<sup>1</sup>, and  $\sigma_{i,\text{corr}}^{(l)}$ , with  $l = 1, \dots, k$ , its correlated systematic uncertainties. With this information at hand, one can construct the full covariance matrix  $V_{ij}$  as follows (see for example Ref. [3]):

$$V_{ij} = (\sigma_{i,\text{stat}}^2 + \sigma_{i,\text{unc}}^2) \delta_{ij} + \sum_{l=1}^k \sigma_{i,\text{corr}}^{(l)} \sigma_{j,\text{corr}}^{(l)}. \quad (2)$$

This is a clearly symmetric matrix. Given a set of predictions  $t_i$  corresponding to the  $n$  measurements of the ensemble, the  $\chi^2$  takes the form:

$$\chi^2 = \sum_{i,j=1}^n (m_i - t_i) V_{ij}^{-1} (m_j - t_j) = \mathbf{y}^T \cdot \mathbf{V}^{-1} \cdot \mathbf{y}, \quad (3)$$

where in the second equality I have used the matricial notation and defined  $y_i = m_i - t_i$ . A convenient way to compute the  $\chi^2$  relies on the Cholesky decomposition of the covariance matrix  $\mathbf{V}$ . In particular, it can be proven that any symmetric and positive definite matrix  $\mathbf{V}$  can be decomposed as:

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T, \quad (4)$$

---

<sup>1</sup>There could be more than one uncorrelated systematic uncertainty. In this case,  $\sigma_{i,\text{unc}}$  is just the square root of the sum in quadrature of all the uncorrelated systematic uncertainties.

where  $\mathbf{L}$  is a lower triangular matrix whose entries are related recursively to those of  $\mathbf{V}$  as follows:

$$\begin{aligned} L_{kk} &= \sqrt{V_{kk} - \sum_{j=1}^{k-1} L_{kj}^2}, \\ L_{ik} &= \frac{1}{L_{kk}} \left( V_{ik} - \sum_{j=1}^{k-1} L_{ij} L_{kj} \right), \quad k < i, \\ L_{ik} &= 0, \quad k > i. \end{aligned} \tag{5}$$

It is then easy to see that the  $\chi^2$  can be written as:

$$\chi^2 = |\mathbf{L}^{-1} \cdot \mathbf{y}|^2. \tag{6}$$

But the vector  $\mathbf{x} \equiv \mathbf{L}^{-1} \cdot \mathbf{y}$  is the solution of the linear system:

$$\mathbf{L} \cdot \mathbf{x} = \mathbf{y}, \tag{7}$$

that can be efficiently solved by forward substitution, so that:

$$\chi^2 = |\mathbf{x}|^2. \tag{8}$$

Following this procedure, one does not need to compute explicitly the inverse of the covariance matrix  $\mathbf{V}$ , simplifying significantly the computation of the  $\chi^2$ .

## 2 Additive and multiplicative uncertainties

The correlated systematic uncertainties  $\sigma_{i,\text{corr}}^{(l)}$  may be either *additive* or *multiplicative*. The nature of the single uncertainties is typically provided by the experiments that release the measurements. A typical example of multiplicative uncertainty is the luminosity uncertainty but there can be others.

Now let us express all the correlated systematic uncertainties  $\sigma_{i,\text{corr}}^{(l)}$  as relative to the associate central value  $m_i$ , so that one defines<sup>2</sup>:

$$\sigma_{i,\text{corr}}^{(l)} \equiv \delta_{i,\text{corr}}^{(l)} m_i \tag{9}$$

and let us also define  $s_i^2 \equiv \sigma_{i,\text{stat}}^2 + \sigma_{i,\text{unc}}^2$  so that Eq. (2) can be rewritten as:

$$V_{ij} = s_i^2 \delta_{ij} + \left( \sum_{l=1}^k \delta_{i,\text{corr}}^{(l)} \delta_{j,\text{corr}}^{(l)} \right) m_i m_j. \tag{10}$$

Now I split the correlated systematic uncertainties into  $k_a$  additive uncertainties and  $k_m$  multiplicative uncertainties, such that  $k_a + k_m = k$ . This way Eq. (10) takes the form:

$$V_{ij} = s_i^2 \delta_{ij} + \left( \sum_{l=1}^{k_a} \delta_{i,\text{add}}^{(l)} \delta_{j,\text{add}}^{(l)} + \sum_{l=1}^{k_m} \delta_{i,\text{mult}}^{(l)} \delta_{j,\text{mult}}^{(l)} \right) m_i m_j. \tag{11}$$

It is believed that this definition of the covariance matrix is problematic in that it results in the so-called D'Agostini bias of the multiplicative uncertainties [2]. A possible solution to this problem is the so-called  $t_0$ -prescription [8], where the experimental central value  $m_i$  in the multiplicative term is replaced by a fixed theoretical predictions  $t_i^{(0)}$ , typically computed in a previous fit in which the “standard” definition of the covariance matrix in Eq. (2) (often referred to as *experimental* definition) is used. Applying the  $t_0$  prescription, the covariance matrix takes the form:

$$V_{ij} = s_i^2 \delta_{ij} + \sum_{l=1}^{k_a} \delta_{i,\text{add}}^{(l)} \delta_{j,\text{add}}^{(l)} m_i m_j + \sum_{l=1}^{k_m} \delta_{i,\text{mult}}^{(l)} \delta_{j,\text{mult}}^{(l)} t_i^{(0)} t_j^{(0)}. \tag{12}$$

---

<sup>2</sup>Note that this redefinition does not change the nature of the uncertainties, additive uncertainties remain additive as well as multiplicative uncertainties remain multiplicative.

### 3 Artificial generation of correlated systematics

In order to implement the definition of the  $\chi^2$  discussed above, it is necessary to have the experimental information in terms of the correlated systematic uncertainties  $\sigma_{i,\text{corr}}^{(l)}$ . This is what the experimental collaborations usually release. However, in some cases this information is given in terms of a covariance matrix. Therefore, one needs to find a workaround to generate correlated systematic uncertainties out of a covariance matrix.

Given a  $n \times n$  symmetric matrix  $\mathbf{C}$ , it will have  $n$  orthonormal eigenvectors  $\mathbf{x}^{(i)}$ , such that  $\mathbf{x}^{(i)} \cdot \mathbf{x}^{(j)} = \delta_{ij}$ , each of which will have a non-negative eigenvalue  $\lambda_i$  associated:

$$\mathbf{C} \cdot \mathbf{x}^{(i)} = \lambda_i \mathbf{x}^{(i)}, \quad i = 1, \dots, n. \quad (13)$$

If one defines:

$$\sigma_{i,\text{corr}}^{(l)} = \sqrt{\lambda_l} x_i^{(l)}, \quad i, l = 1, \dots, n, \quad (14)$$

one can show that:

$$\sum_{l=1}^n \sigma_{i,\text{corr}}^{(l)} \sigma_{j,\text{corr}}^{(l)} = C_{ij}. \quad (15)$$

To prove this equality, I start from the following matricial relation:

$$\mathbf{C} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{-1}, \quad (16)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix with the eigenvalues  $\lambda_i$  on the diagonal ( $\Lambda_{ij} = \lambda_i \delta_{ij}$ ), while  $\mathbf{Q}$  is a matrix whose columns are the eigenvectors  $\mathbf{x}^{(i)}$  ( $Q_{ij} = x_i^{(j)}$ ). In addition, since in this particular case  $\mathbf{x}^{(i)} \cdot \mathbf{x}^{(j)} = \delta_{ij}$ , this implies that:

$$\mathbf{Q}^T \cdot \mathbf{Q} = \mathbf{I} \quad \Rightarrow \quad \mathbf{Q}^{-1} = \mathbf{Q}^T, \quad (17)$$

so that:

$$\mathbf{C} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^T. \quad (18)$$

It follows that:

$$C_{ij} = \sum_{k,l=1}^n Q_{ik} \Lambda_{kl} Q_{jl} = \sum_{k,l=1}^n x_i^{(k)} \lambda_k \delta_{kl} x_j^{(l)} = \sum_{l=1}^n \lambda_l x_i^{(l)} x_j^{(l)} = \sum_{l=1}^n \sigma_{i,\text{corr}}^{(l)} \sigma_{j,\text{corr}}^{(l)}, \quad (19)$$

as required.

The matrix  $\mathbf{C}$  can be regarded as the correlated contribution to the full covariance matrix  $\mathbf{V}$ . In particular, considering Eqs. (2) and (10), one can write:

$$\mathbf{V} = \mathbf{U} + \mathbf{C}, \quad (20)$$

where  $\mathbf{U}$  is a diagonal matrix of uncorrelated uncertainties:

$$U_{ij} = s_i^2 \delta_{ij}. \quad (21)$$

This defines the matrix  $\mathbf{C}$  as:

$$\mathbf{C} = \mathbf{V} - \mathbf{U}, \quad (22)$$

such that, given a  $n \times n$  covariance matrix  $\mathbf{V}$  along with its uncorrelated contribution  $\mathbf{U}$ , one can generate a set of  $n$  *artificial* correlated systematics according to Eq. (14), where  $\mathbf{C}$  is given in Eq. (22), for each of the  $n$  measurements. This allows one to implement Eq. (12) for the construction of the covariance matrix.

## 4 Determining the systematic shifts

In order to visualise the effect of systematic uncertainties, it is instructive to compute the *systematic shift* generated by the systematic uncertainties. To do so, one needs to write the  $\chi^2$  in terms of the so-called “nuisance parameters”  $\lambda_\alpha$ . One can show that the definition of the  $\chi^2$  in Eq. (3) is equivalent to [3]:

$$\chi^2 = \sum_{i=1}^n \frac{1}{s_i^2} \left( m_i - t_i - \sum_{\alpha=1}^k \lambda_\alpha \sigma_{i,\text{corr}}^{(\alpha)} \right)^2 + \sum_{\alpha=1}^k \lambda_\alpha^2. \quad (23)$$

The optimal value of the nuisance parameters can be computed by minimising the  $\chi^2$  with respect to them, that is imposing:

$$\frac{\partial \chi^2}{\partial \lambda_\beta} = 0. \quad (24)$$

This yields the system:

$$\sum_{\beta=1}^k A_{\alpha\beta} \lambda_\beta = \rho_\alpha, \quad (25)$$

with:

$$A_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{i=1}^n \frac{\sigma_{i,\text{corr}}^{(\alpha)} \sigma_{i,\text{corr}}^{(\beta)}}{s_i^2} \quad \text{and} \quad \rho_\alpha = \sum_{i=1}^n \frac{m_i - t_i}{s_i^2} \sigma_{i,\text{corr}}^{(\alpha)}, \quad (26)$$

that determines the values of  $\lambda_\beta$ . The quantity:

$$d_i = \sum_{\alpha=1}^k \lambda_\alpha \sigma_{i,\text{corr}}^{(\alpha)} \quad (27)$$

in Eq. (23) can be interpreted as a shift caused by the correlated systematic uncertainties. Defining the shifted predictions as:

$$\bar{t}_i = t_i + d_i, \quad (28)$$

the  $\chi^2$  reads:

$$\chi^2 = \sum_{i=1}^n \left( \frac{m_i - \bar{t}_i}{s_i} \right)^2 + \sum_{\alpha=1}^k \lambda_\alpha^2. \quad (29)$$

Therefore, up to a penalty term given by the sum of the square of the nuisance parameters, the  $\chi^2$  takes the form of the uncorrelated definition. In order to achieve a visual assessment of the agreement between data and theory, it appears natural to compare the central experimental values  $m_i$  to the shifted theoretical predictions  $\bar{t}_i$  in units of the uncorrelated uncertainty  $s_i$ .

## 5 Effect of cuts on the $\chi^2$

A relevant question to ask is what is the effect of possible cuts on the input data set on the form of the  $\chi^2$  in the presence of correlations. In order to address this question, I consider a simple data set made of two datapoints equipped with a single uncorrelated uncertainty and a single fully-correlated uncertainty:<sup>3</sup>

$$S : \quad \{m_i \pm \sigma_i^u \pm \sigma_i^c\}, \quad i = 1, 2. \quad (30)$$

The resulting covariance matrix reads:

$$\mathbf{V} = \begin{pmatrix} (\sigma_1^u)^2 + (\sigma_1^c)^2 & \sigma_1^c \sigma_2^c \\ \sigma_1^c \sigma_2^c & (\sigma_2^u)^2 + (\sigma_2^c)^2 \end{pmatrix}, \quad (31)$$

---

<sup>3</sup>Note that one needs to include *also* an uncorrelated uncertainty because otherwise the covariance matrix would be singular. This is a symptom of the fact that if no uncorrelated uncertainties are present, the two points would be totally dependent on each other. This observation will be relevant in the following.

with inverse:

$$\mathbf{V}^{-1} = \frac{1}{(\sigma_1^u \sigma_2^u)^2 + (\sigma_1^c \sigma_2^u)^2 + (\sigma_1^u \sigma_2^c)^2} \begin{pmatrix} (\sigma_2^u)^2 + (\sigma_2^c)^2 & -\sigma_1^c \sigma_2^c \\ -\sigma_1^c \sigma_2^c & (\sigma_1^u)^2 + (\sigma_1^c)^2 \end{pmatrix}. \quad (32)$$

In addition, the set of two theoretical predictions  $\{t_i\}$  that allows one to define a column vector of residuals  $\mathbf{r}^T = (r_1, r_2)$  with  $r_i = m_i - t_i$ . The  $\chi^2$  then takes the following explicit expression:

$$\chi^2 = \mathbf{r}^T \cdot \mathbf{V}^{-1} \cdot \mathbf{r} = \frac{r_1^2 [(\sigma_2^u)^2 + (\sigma_2^c)^2] + r_2^2 [(\sigma_1^u)^2 + (\sigma_1^c)^2] - 2r_1 r_2 \sigma_1^c \sigma_2^c}{(\sigma_1^u \sigma_2^u)^2 + (\sigma_1^c \sigma_2^u)^2 + (\sigma_1^u \sigma_2^c)^2}. \quad (33)$$

Now, suppose one wants to exclude the datapoint with  $i = 2$  from the definition of the  $\chi^2$ . The possibly most natural way to proceed is to exclude the point, along with its uncertainties, directly from the set in Eq. (30). This straightforwardly leads to the following  $\chi^2$ :

$$\chi_{\text{cut } 1}^2 = \frac{r_1^2}{(\sigma_1^u)^2 + (\sigma_1^c)^2}. \quad (34)$$

However, this procedure might be questioned in that it artificially modifies the original data set by effectively defining a *new* data set where the point  $i = 2$  is not present. In the absence of correlations this is a sound procedure because each single datapoint can be regarded an independent subset of the original set. However, this is not the case when correlations are present.

Therefore, it is necessary to devise a method that avoids any modification of the data set but yet allows one to prevent specific datapoints to contribute to the  $\chi^2$ . In general, the purpose of excluding datapoints from a fit is to avoid that the model used to compute the predictions is forced outside its application region. One can therefore *enforce* that the model exactly agrees with the datapoints to be excluded. This effectively amounts to set  $t_2 = m_2$ , or equivalently  $r_2 = 0$ , reducing Eq. (33) to:

$$\chi_{\text{cut } 2}^2 = \frac{r_1^2 [(\sigma_2^u)^2 + (\sigma_2^c)^2]}{(\sigma_1^u \sigma_2^u)^2 + (\sigma_1^c \sigma_2^u)^2 + (\sigma_1^u \sigma_2^c)^2}. \quad (35)$$

This has the clear advantage that the experimental information, including correlations, remains totally intact. In order to show that Eq. (35) has the right properties, I first take the limit for  $\sigma_2^c \rightarrow 0$ . This gives:

$$\lim_{\sigma_2^c \rightarrow 0} \chi_{\text{cut } 2}^2 = \frac{r_1^2}{(\sigma_1^c)^2 + (\sigma_1^u)^2}, \quad (36)$$

which reproduces the result of Eq. (34). This was to be expected because no reference to the point  $i = 2$  can remain after the removal of its correlation to the point  $i = 1$ . In addition, it is also very instructive to take the limit  $\sigma_2^u \rightarrow 0$  Eq. (35), which gives:

$$\lim_{\sigma_2^u \rightarrow 0} \chi_{\text{cut } 2}^2 = \frac{r_1^2}{(\sigma_1^u)^2}. \quad (37)$$

In this case, not only any reference to the point  $i = 2$  drops out, but also the dependence of the  $\chi^2$  on the correlation uncertainty of the point  $i = 1$  disappears. Despite at first sight this looks counterintuitive, it seems to be the correct behaviour. To see this, observe that, if the point  $i = 2$  has no uncorrelated uncertainty, its fluctuations are totally driven by the fluctuations of the point  $i = 1$ . Therefore, the point  $i = 2$  is completely dependent on the point  $i = 1$  and thus cannot give any contribution to the  $\chi^2$ . In addition, the correlation uncertainty of the point  $i = 1$  must also be irrelevant because, being  $i = 2$  totally correlated to  $i = 1$ , any correlation of  $i = 1$  to  $i = 2$  translates into a correlation to itself and this cannot contribute to the  $\chi^2$  either. Importantly, if one takes the limit  $\sigma_1^u \rightarrow 0$ , the  $\chi^2$  diverges no matter the value of  $\sigma_1^c$ . This is consistent with setting  $\sigma_1^u = \sigma_2^u$  since the beginning in Eq. (30) which indeed gives a singular  $\chi^2$  because the covariance matrix is singular. Eq. (34) does not produce these features hanging in favour of Eq. (35).

## 6 Monte Carlo replica generation

In this section, I consider the Monte Carlo generation of artificial replicas. The aim is to obtain a formula for the generation of replicas that gives rise to a figure of merit distributed correctly, *i.e.* that follows a  $\chi^2$  distribution with as many degrees of freedom as number of data points. Following Ref. [5], one requires that the  $\chi^2$ , that in matricial form can be expressed as:

$$\chi^2 = \mathbf{y}^T \cdot \mathbf{V}^{-1} \cdot \mathbf{y}, \quad (38)$$

where  $\mathbf{V}$  is the  $n \times n$  covariance matrix, being  $n$  the number of datapoints, and  $\mathbf{y}$  is a column vector with  $n$  entries that can be expressed as:

$$y_j = f_j - m_j, \quad (39)$$

with  $m_j$  a given value (to be thought as the central value) and  $f_j$  its fluctuated value, is distributed like:

$$\chi^2 \sim \sum_{i=1}^n z_i^2 = |\mathbf{z}|^2, \quad (40)$$

where  $z_1, \dots, z_n$  are independent, standard normal random variables, *i.e.* such that:

$$\langle z_i \rangle = 0 \quad \text{and} \quad \langle z_i z_j \rangle = \delta_{ij}, \quad (41)$$

where the symbol  $\langle \dots \rangle$  indicates the average.<sup>4</sup> This immediately tells that:

$$\langle \chi^2 \rangle = n. \quad (44)$$

As already discussed above, Eq. (38) can alternatively be written as:

$$\chi^2 = |\mathbf{L}^{-1} \cdot \mathbf{y}|^2, \quad (45)$$

where  $\mathbf{L}$  is the Cholesky decomposition of the matrix  $\mathbf{V}$ , such that:

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T. \quad (46)$$

Comparing Eqs. (45) and (40), one can infer that the following relation should hold:

$$\mathbf{L}^{-1} \cdot \mathbf{y} \sim \mathbf{z}, \quad (47)$$

which immediately implies:

$$\mathbf{y} \sim \mathbf{L} \cdot \mathbf{z}, \quad (48)$$

or equivalently:

$$f_j \sim m_j + \sum_{i=1}^n L_{ji} z_i. \quad (49)$$

In conclusion, in order for the  $\chi^2$  to have the correct distribution, the fluctuations  $f_j$  around the experimental central values  $m_j$  of a given Monte Carlo replica have to be computed according to Eq. (49).

---

<sup>4</sup>The relevant identities are:

$$\langle z_i^{2n-1} \rangle = 0 \quad \text{and} \quad \langle z_i^{2n} \rangle = (2n-1)!! \quad n = 1, 2, 3, \dots, \quad (42)$$

where  $(2n-1)!! = 1 \cdot 3 \cdot \dots \cdot (2n-1)$  is the odd semi-factorial. When considering combinations of two random variables the result is:

$$\langle z_i^n z_j^m \rangle = \begin{cases} \langle z_i^{n+m} \rangle & i = j \\ \langle z_i^n \rangle \langle z_j^m \rangle & i \neq j \end{cases} \quad (43)$$

I now show that Eq. (49) reproduces the statistical properties of the original ensemble of data. To do so, I compute average and correlations of Eq. (49) using Eq. (41). The average gives:

$$\langle f_j \rangle = m_j + \sum_{i=1}^n L_{ji} \langle z_i \rangle = m_j, \quad (50)$$

as expected. The correlation instead gives:

$$\begin{aligned} \langle f_j f_k \rangle &= \left\langle \left( m_j + \sum_{i=1}^n L_{ji} z_i \right) \left( m_k + \sum_{i'=1}^n L_{ki'} z_{i'} \right) \right\rangle \\ &= m_j m_k + \sum_{i,i'=1}^n L_{ji'} L_{ki'} \langle z_i z_{i'} \rangle = \langle f_j \rangle \langle f_k \rangle + V_{jk}, \end{aligned} \quad (51)$$

again as expected. These are fundamental requirements of a sound Monte Carlo ensemble (see also Eq. (10) of Ref. [8]).

Eq. (49) should be compared to the more “standard” Monte-Carlo replica generation formulas such as those given in Refs. [6, 7]. I now specifically consider Eq. (20) of Ref. [7] that I report here in terms of relative uncertainties  $\delta$ ’s matching our notation:

$$\tilde{f}_j \sim m_j \prod_{m=1}^{k_m} \left( 1 + \delta_{\text{mult},j}^{(m)} x_m \right) \left( 1 + \delta_{\text{unc},j} z_j + \sum_{a=1}^{k_a} \delta_{\text{add},j}^{(a)} y_a \right), \quad (52)$$

where  $z_j$ ,  $y_a$ , and  $x_m$  are independent random variables. I now expand the product in the front up to second order in  $x$

$$\prod_{m=1}^{k_m} \left( 1 + \delta_{\text{mult},j}^{(m)} x_m \right) = 1 + \sum_{m=1}^{k_m} \delta_{\text{mult},j}^{(m)} x_m + \sum_{m=1}^{k_m} \sum_{\substack{m'=1 \\ m' \neq m}}^{k_m} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},j}^{(m')} x_m x_{m'} + \text{trilinear}, \quad (53)$$

where the “trilinear” corrections correspond to combinations of the kind  $x_i x_j x_k$ , with  $i \neq j \neq k \neq i$ , such that their average vanishes:  $\langle x_i x_j x_k \rangle = \langle x_i \rangle \langle x_j \rangle \langle x_k \rangle = 0$ . This allows one to expand Eq. (52) as:

$$\begin{aligned} \tilde{f}_j &\sim m_j \left( 1 + \delta_{\text{unc},j} z_j + \sum_{a=1}^{k_a} \delta_{\text{add},j}^{(a)} y_a + \sum_{m=1}^{k_m} \delta_{\text{mult},j}^{(m)} x_m \right. \\ &+ \delta_{\text{unc},j} \sum_{m=1}^{k_m} \delta_{\text{mult},j}^{(m)} z_j x_m + \sum_{a=1}^{k_a} \sum_{m=1}^{k_m} \delta_{\text{add},j}^{(a)} \delta_{\text{mult},j}^{(m)} y_a x_m \\ &\left. + \sum_{m=1}^{k_m} \sum_{\substack{m'=1 \\ m' \neq m}}^{k_m} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},j}^{(m')} x_m x_{m'} \right) + \text{trilinear}. \end{aligned} \quad (54)$$

Since Eq. (54) involves only terms that have at most one power of the same random variable, the average of  $\tilde{f}_j$  automatically yields:

$$\langle \tilde{f}_j \rangle = m_j, \quad (55)$$

in agreement with the expectation. I now compute the correlations making use of the combination

rules of random variables. This gives:

$$\begin{aligned}
\langle \tilde{f}_j \tilde{f}_k \rangle &\sim m_j m_k \left( 1 + \delta_{\text{unc},j}^2 \delta_{jk} + \sum_{a=1}^{k_a} \delta_{\text{add},j}^{(a)} \delta_{\text{add},k}^{(a)} + \sum_{m=1}^{k_m} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},k}^{(m)} \right. \\
&+ \delta_{\text{unc},j}^2 \sum_{m=1}^{k_m} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},k}^{(m)} + \sum_{a=1}^{k_a} \sum_{m=1}^{k_m} \delta_{\text{add},j}^{(a)} \delta_{\text{add},k}^{(a)} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},k}^{(m)} \\
&+ \left. \sum_{m=1}^{k_m} \sum_{\substack{m'=1 \\ m' \neq m}}^{k_m} \delta_{\text{mult},j}^{(m)} \delta_{\text{mult},k}^{(m)} \delta_{\text{mult},j}^{(m')} \delta_{\text{mult},k}^{(m')} + \mathcal{O}(\delta^6) \right) \\
&= \langle \tilde{f}_j \rangle \langle \tilde{f}_k \rangle + V_{jk} + \mathcal{O}(\sigma_{\text{mult}}^2 s^2, \sigma_{\text{mult}}^2 \sigma_{\text{add}}^2).
\end{aligned} \tag{56}$$

In conclusion, Eq. (52) reproduces exactly the central values of the original ensemble (see Eq. (55)) while it reproduces correlations only up to corrections proportional to the fourth power of the typical uncertainty. Interestingly, these corrections are all proportional to  $\sigma_{\text{mult}}^2$ . This means that, in the absence of multiplicative uncertainties, somewhat expectedly Eq. (52) *exactly* reproduces the correlations. On the other hand, if multiplicative uncertainties are large these correction become potentially sizeable preventing Eq. (52) from correctly reproducing the correlations of the original ensemble.

The case of Eq. (13) of Ref. [6] is more involved because of the presence of the square-root sign for the relative multiplicative uncertainties. This case can be handled again by expanding to second order the factor depending on multiplicative uncertainties with the only difference that, due to the expansion:

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \mathcal{O}(x^3), \tag{57}$$

the square root produces a factor 1/2 in front of the linear terms and a factor  $-1/8$  in front of the quadratic terms. This is expected to reduce the impact of the corrections terms in Eq. (56).

It is instructive to work out whether Eq. (49) can be directly related to the Monte-Carlo replicas generation formulas given in Refs. [6, 7]. The aim is to find a closed-form expression for the Cholesky-decomposition matrix  $\mathbf{L}$  of the covariance matrix  $\mathbf{V}$  in terms of the single uncertainties. Through Eq. (49), this will finally allow one to derive a formula of the same kind of those given in Refs. [6, 7]. The basic observation is that the covariance matrix can be written as:

$$\mathbf{V} = \mathbf{D} + \sum_{l=1}^k \mathbf{V}^{(l)} = \sum_{l=0}^k \mathbf{L}^{(l)} \cdot \mathbf{L}^{(l)T}, \tag{58}$$

with:

$$L_{ij}^{(l)} = \begin{cases} s_i \delta_{ij} & l = 0, \\ \sigma_i^{(l)} \delta_{j1} & 1 \leq l \leq k. \end{cases} \tag{59}$$

On the other hand, one also knows that:

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T, \tag{60}$$

therefore:

$$\mathbf{L} \cdot \mathbf{L}^T = \sum_{l=0}^k \mathbf{L}^{(l)} \cdot \mathbf{L}^{(l)T}. \tag{61}$$

The question is to express  $\mathbf{L}$  in terms of all the  $\mathbf{L}^{(l)}$ . Unfortunately though this equation has no general solution. However, a step forward can be done by exploiting the fact that the matrices  $\mathbf{V}^{(l)}$  are positive semidefinite (*i.e.* they have zero determinant) and thus their Cholesky decomposition is not unique. In particular, provided that  $k \leq n$ , one can choose:

$$\tilde{L}_{ij}^{(l)} = \sigma_i^{(l)} \delta_{jl}, \quad 1 \leq l \leq k \leq n, \tag{62}$$



such that:

$$\left( \sum_{l=1}^k \tilde{\mathbf{L}}^{(l)} \right) \cdot \left( \sum_{l'=1}^k \tilde{\mathbf{L}}^{(l')} \right) = \sum_{l=1}^k \mathbf{V}^{(l)}. \quad (63)$$

However, choosing:

$$\tilde{\mathbf{L}} = \mathbf{L}^{(0)} + \sum_{l=1}^k \tilde{\mathbf{L}}^{(l)}, \quad (64)$$

and multiplying it for its transpose would give:

$$\tilde{\mathbf{L}} \cdot \tilde{\mathbf{L}}^T = \mathbf{V} + \sum_{l=1}^k \left( \mathbf{L}^{(0)} \cdot \tilde{\mathbf{L}}^{(l)T} + \tilde{\mathbf{L}}^{(l)} \cdot \mathbf{L}^{(0)T} \right), \quad (65)$$

that gives back the covariance matrix plus a spurious term. Writing the expression above in terms of single entries, one finds:

$$\sum_{m=1}^n \tilde{L}_{im} \tilde{L}_{jm} = V_{ij} + s_i \sigma_j^{(i)} + s_j \sigma_i^{(j)}. \quad (66)$$

The peculiarity here is that the upper index of  $\sigma_j^{(i)}$  and  $\sigma_i^{(j)}$ , since  $k \leq n$  (and in fact usually  $k \ll n$ , *i.e.* the number of correlated uncertainties is typically much smaller than the number of data points), exceeds the allowed range. To fix this problem, I just assume  $\sigma_j^{(i)} = 0$  for  $i > k$ . This finally gives:

$$\sum_{m=1}^n \tilde{L}_{im} \tilde{L}_{jm} = V_{ij}, \quad \text{for } i, j > k, \quad (67)$$

that means that, by adopting the Cholesky decomposition in Eq. (62), part of the covariance matrix (usually the largest part) is exactly recovered. Therefore,  $\tilde{\mathbf{L}}$  Eq. (64) seems to be a generally good approximator to the real Cholesky decomposition of the covariance matrix  $\mathbf{V}$ . Plugging this equation into Eq. (49) immediately gives:

$$f_j \sim m_j + z_j s_j + \sum_{l=1}^k z_l \sigma_{j,\text{corr}}^{(l)}. \quad (68)$$

It is interesting to focus on similarities and differences with the formulas present in the literature. I will specifically consider Eq. (13) of Ref. [6] and Eq. (20) of Ref. [7].

Let us start by considering the case in which there are only uncorrelated and correlated additive uncertainties. In this case, Refs. [6, 7] are in mutual agreement and only in partial agreement with our Eq. (68), the difference being that the random numbers used to generate the fluctuation associated to the  $k$  correlated uncertainties are equal to the first  $k$  of the  $n$  used for the uncorrelated uncertainties (remember the requirement  $k \leq n$ ). Of course, if  $k \ll n$ , as it often happens, the agreement between Eq. (68) and Refs. [6, 7] is asymptotically good.

When considering also correlated multiplicative uncertainties there are differences across formulas of Refs. [6, 7] and Eq. (68). While Eq. (68) treats multiplicative uncertainties on the same footing as additive uncertainties (consistently with the construction of the covariance matrix in Eq. (2)), Refs. [6, 7] introduce a multiplicative factor. In addition, Ref. [6] distinguishes between relative and absolute multiplicative uncertainties.

## 7 The D’Agostini “bias”

The main topic of this section is the so-called D’Agostini “bias” (DAB) originally presented in Ref. [2]. Quoting from the abstract of that paper: “Best fits to data which are affected by systematic uncertainties on the normalisation factor have the tendency to produce curves lower than expected [...]”. This bias can become unacceptable if the normalisation error is large, or a large number of data points

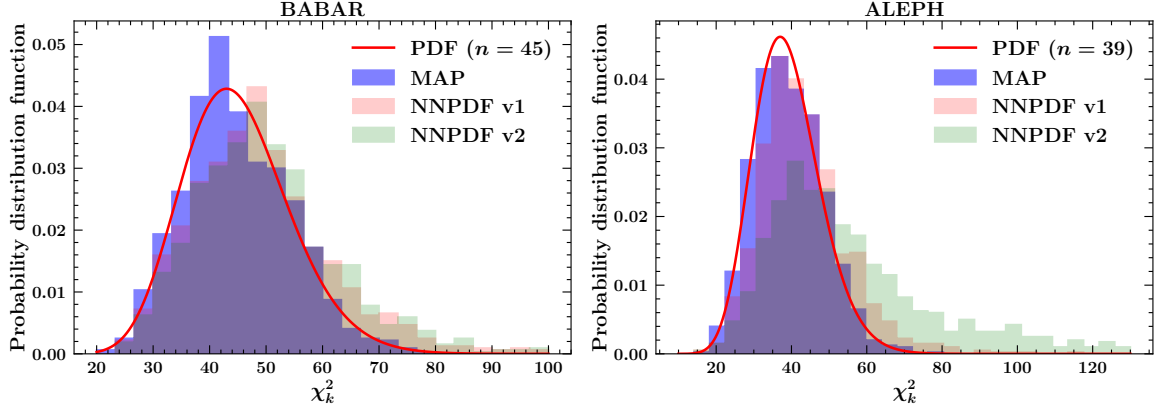


Figure 1: Distribution of the  $\chi^2$ 's of the Monte Carlo replica w.r.t. to the corresponding central values for BABAR (left) and ALEPH (right) experiments compared to the probability distribution function with to the appropriate number of degrees of freedom given by the number of data points. The blue histograms are produced using Eq. (49), the red histogram is produced using Eq. (13) of Ref. [6] where multiplicative uncertainties are treated as relative, while the green histogram is produced using Eq. (20) of Ref. [7].

are used". In the following, I will summarise the arguments behind this conclusion. I will then move to giving my own view on this question arguing that the DAB is in fact not a bias but rather a feature of correlation uncertainties in general and not of multiplicative uncertainties in particular. This implies that the distinction between additive and normalisation uncertainties is effectively immaterial. Specifically, what really triggers the DAB is *not* the fact the uncertainty is on the multiplication factor but rather that it typically gives raise to correlated uncertainties that are different in *absolute* size. To show this, below I will show that the DAB takes place also in the presence of additive uncertainties that are different in absolute size. Conversely, I will show that the DAB does not takes place when considering multiplicative uncertainties and the central values are identical. The bottom line is that any difference in absolute size between correlated uncertainties causes a deviation from the expectation according to which the best fit value should be given by the weighted average of the experimental central values weighted by the inverse of the squared uncorrelated uncertainties. This deviation from the expectation is the origin of the DAB.

In order to present the argument, Ref. [2] considers a data set made of two data points with the respective uncorrelated uncertainty,  $x_i \pm \sigma_i$  with  $i = 1, 2$ . In addition, it is assumed that  $x_i$  measure the same quantity and thus they can be fitted with a single constant  $\mu$ . Two cases are treated. In the first case the two data points are also affected by a single offset (additive) correlated uncertainty  $\sigma_c$ . In the second case, instead, the data points are affected by a normalisation correlated uncertainty whose size is proportional to the central values through the factor  $\delta_f$ . Given the simplicity of the problem, it is possible in both cases to analytically compute the best value of  $\mu$  and its uncertainty  $\delta\mu$  by minimising the  $\chi^2$  and computing its second derivative at the minimum. As usual the  $\chi^2$  is defined as:

$$\chi^2 = \sum_{i,j=1}^2 y_i V_{ij}^{-1} y_j = \mathbf{y}^T \cdot \mathbf{V}^{-1} \cdot \mathbf{y}, \quad (69)$$

where  $y_i = \mu - x_i$  and  $V_{ij}^{-1}$  are the entries of the inverse of the covariance matrix.

Let us start with the offset case in which the covariance matrix takes the form:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_2^2 + \sigma_c^2 \end{pmatrix}, \quad (70)$$

whose inverse is:

$$\mathbf{V}^{-1} = \frac{1}{D} \begin{pmatrix} \sigma_2^2 + \sigma_c^2 & -\sigma_c^2 \\ -\sigma_c^2 & \sigma_1^2 + \sigma_c^2 \end{pmatrix}, \quad (71)$$

with  $D = \sigma_1^2 \sigma_2^2 + (\sigma_1^2 + \sigma_2^2) \sigma_c^2$ . With this I can straightforwardly compute the  $\chi^2$ :

$$\chi^2 = \frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 + (y_1 - y_2)^2 \sigma_c^2}{D}, \quad (72)$$

that can be minimised w.r.t. to  $\mu$  by requiring that:

$$\left. \frac{d\chi^2}{d\mu} \right|_{\mu=\bar{\mu}} = 2 \frac{\sigma_2^2 y_1 + \sigma_1^2 y_2}{D} \Big|_{\mu=\bar{\mu}} = 0, \quad (73)$$

where  $\bar{\mu}$  is the best fit value of  $\mu$ . Notice the cancellation in the numerator of the correlated uncertainty  $\sigma_c$ . This “accidental” cancellation in the offset case, caused by the fact that  $\sigma_c$  is the same for both  $x_1$  and  $x_2$ , can be regarded as the (misleading) evidence that the DAB only takes place in the normalisation case. The solution to Eq. (73) is:

$$\bar{\mu} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \quad (74)$$

that fulfils the intuitive expectation that the best value of  $\mu$  should be a weighted average of the central values weighted by the inverse of the uncorrelated uncertainties squared. One can also compute the uncertainty of  $\bar{\mu}$ ,  $\delta\mu$ , by simply computing the inverse of the second derivative of the  $\chi^2$  at  $\mu = \bar{\mu}$ :<sup>5</sup>

$$\delta\mu^2 = \left[ \frac{1}{2} \frac{d^2\chi^2}{d\mu^2} \right]_{\mu=\bar{\mu}}^{-1} = \frac{\sigma_1^2 \sigma_2^2 + (\sigma_1^2 + \sigma_2^2) \sigma_c^2}{\sigma_2^2 + \sigma_1^2} = \frac{1 + \frac{\sigma_c^2}{\sigma_1^2} + \frac{\sigma_c^2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}. \quad (75)$$

I now move to considering the normalisation case in which the covariance matrix reads:

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 + \delta_f^2 x_1^2 & \delta_f^2 x_1 x_2 \\ \delta_f^2 x_1 x_2 & \sigma_2^2 + \delta_f^2 x_2^2 \end{pmatrix}, \quad (76)$$

whose inverse is:

$$\mathbf{V}^{-1} = \frac{1}{D} \begin{pmatrix} \sigma_2^2 + \delta_f^2 x_2^2 & -\delta_f^2 x_1 x_2 \\ -\delta_f^2 x_1 x_2 & \sigma_1^2 + \delta_f^2 x_1^2 \end{pmatrix}, \quad (77)$$

with  $D = \sigma_1^2 \sigma_2^2 + \delta_f^2 (\sigma_1^2 x_2^2 + \sigma_2^2 x_1^2)$ . The resulting  $\chi^2$  is:

$$\chi^2 = \frac{1}{D} [\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 + \delta_f^2 (x_2 y_1 - x_1 y_2)^2]. \quad (78)$$

The minimum is computed as:

$$\left. \frac{d\chi^2}{d\mu} \right|_{\mu=\bar{\mu}} = 2 \frac{\sigma_2^2 y_1 + \sigma_1^2 y_2 + \delta_f^2 (x_2 - x_1)(x_2 y_1 - x_1 y_2)}{D} \Big|_{\mu=\bar{\mu}} = 0. \quad (79)$$

This time, differently from the offset case, the correlated uncertainties in the denominator do not cancel and the result of the equation above is:

$$\bar{\mu} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2 + (\delta_f x_1 - \delta_f x_2)^2} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \left( \frac{\delta_f x_1 - \delta_f x_2}{\sigma_1 \sigma_2} \right)^2}, \quad (80)$$

while:

$$\delta\mu^2 = \left[ \frac{1}{2} \frac{d^2\chi^2}{d\mu^2} \right]_{\mu=\bar{\mu}}^{-1} = \frac{\sigma_1^2 \sigma_2^2 + \delta_f^2 (\sigma_1^2 x_2^2 + \sigma_2^2 x_1^2)}{\sigma_2^2 + \sigma_1^2 + \delta_f^2 (x_1 - x_2)^2} = \frac{1 + \frac{\delta_f^2 x_1^2}{\sigma_1^2} + \frac{\delta_f^2 x_2^2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \left( \frac{\delta_f x_1 - \delta_f x_2}{\sigma_1 \sigma_2} \right)^2}. \quad (81)$$

---

<sup>5</sup>Actually, the second derivative of the  $\chi^2$  does not depend on  $\mu$  and thus it is not necessary to set  $\mu = \bar{\mu}$ .

According to Ref. [2], the term proportional to  $\delta_f$  in the denominator of the r.h.s. of Eq. (80) is a source of bias in that it deviates from the intuitive expectation of Eq. (74). Being this additional term positive, it tends to diminish the best fit value.

I will now show that the additional term in the denominator of Eq. (80) is not a specific feature of normalisation uncertainties. To show this, I consider the same data set of two data points but where the data point  $x_1$  is affected by the additive uncertainty  $\sigma_{c,1}$  and the point  $x_2$  by the additive uncertainty  $\sigma_{c,2}$ . According to the same procedure followed above the final result for best-fit value and uncertainty of  $\mu$  is:

$$\bar{\mu} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \left(\frac{\sigma_{c,1} - \sigma_{c,2}}{\sigma_1 \sigma_2}\right)^2}, \quad (82)$$

and:

$$\delta\mu^2 = \frac{1 + \frac{\sigma_{c,1}^2}{\sigma_1^2} + \frac{\sigma_{c,2}^2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \left(\frac{\sigma_{c,1} - \sigma_{c,2}}{\sigma_1 \sigma_2}\right)^2}. \quad (83)$$

These equations are (somewhat expectedly) equal to Eqs. (80) and (81) with the only difference that  $\delta_f x_i$  is replaced by  $\sigma_{c,i}$ . Therefore, the DAB shows up also when additive uncertainties are considered as long as they are different in absolute value. Therefore, the DAB is not a prerogative of normalisation uncertainties only. As expected, the equations above reduce to Eqs. (74) and (75) when  $\sigma_{c,1} = \sigma_{c,2} = \sigma_c$ .

Another important aspect of Eqs. (80) and (81) is that the term that causes the “bias” is proportional to the difference  $x_1 - x_2$ . Therefore, if the central values happen to be the same that term vanishes and thus no bias is present. This observation, along with the fact that the DAB is potentially present also when additive uncertainties are considered, leads to a simple conclusion: the DAB is present in all cases whenever there are correlated uncertainties different in absolute value. This is trivially the case when one introduces different correlated additive uncertainties,  $\sigma_{c,1}$  and  $\sigma_{c,2}$  with  $\sigma_{c,1} \neq \sigma_{c,2}$ . But this also the case in the presence of a normalisation uncertainties when central values are different ( $x_1 \neq x_2$ ). On the contrary, if  $\sigma_{c,1} = \sigma_{c,2}$  in the additive case or  $x_1 = x_2$  in the multiplicative case, no “bias” is present. The conclusion is that the very distinction between additive and multiplicative uncertainties is immaterial.

Now that I have established that the DAB is not a prerogative of normalisation uncertainties only, I will argue that this is in fact not a bias but a manifestation of large correlated uncertainties in general that indicates an intrinsic inconsistency of the data set being considered. I start by observing that, in the presence of correlations, a visual comparison between data and best fit value is meaningful only when including the systematic shifts introduced in Sect. 4. The task is particularly simple due to the small number of points and the fact that there one single correlated uncertainty. In particular, there is a single nuisance parameter that in the normalisation case is given by:

$$\lambda = \frac{\frac{\delta_f x_1^2}{\sigma_1^2} + \frac{\delta_f x_2^2}{\sigma_2^2}}{1 + \frac{\delta_f^2 x_1^2}{\sigma_1^2} + \frac{\delta_f^2 x_2^2}{\sigma_2^2}} - \bar{\mu} \frac{\frac{\delta_f x_1}{\sigma_1^2} + \frac{\delta_f x_2}{\sigma_2^2}}{1 + \frac{\delta_f^2 x_1^2}{\sigma_1^2} + \frac{\delta_f^2 x_2^2}{\sigma_2^2}}. \quad (84)$$

It is the quantity:

$$\hat{\mu}_i = \bar{\mu} + \lambda \delta_f x_i, \quad (85)$$

that can be directly compared to the data-point central values. Unfortunately, the analytic form of this quantity is not very informative but to see the effect of the systematic shifts I consider the very same example given in the introduction of Ref. [2], *i.e.* a data set of two measurements:  $8.0 \pm 2\%$  and  $8.5 \pm 2\%$  having a common 10% normalisation uncertainty.

Fig. 2 shows the result of minimising the  $\chi^2$  to determine the best fit value of the parameter  $\mu$ . The solid error bars display the uncorrelated uncertainties while the dashed error bars display the sum in quadrature of correlated and uncorrelated uncertainties. The best fit value  $\bar{\mu}$ , Eq. (74), is represented by the horizontal red line. As argued in Fig. 2, this line lies below both the central experimental values ( $\bar{\mu} \simeq 7.87$ ) contradicting the common expectation that best fit value should be somewhere between

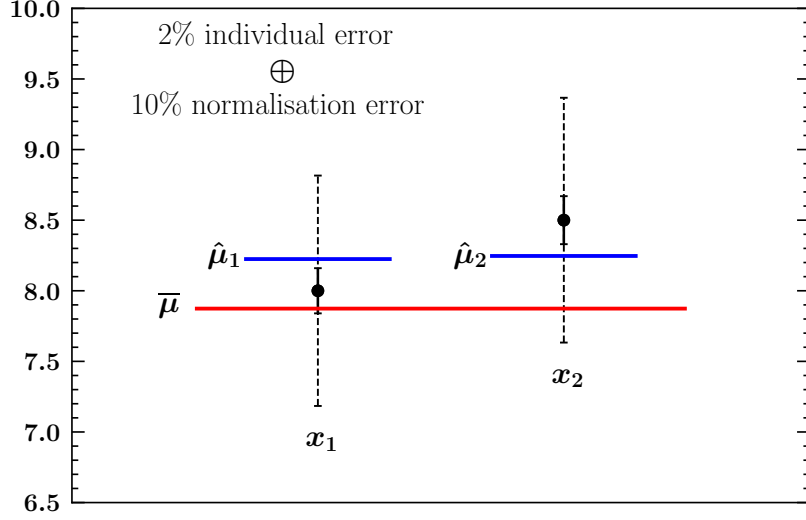


Figure 2: Numerical example taken from Ref. [2] of a data set of two points with central value 8.0 and 8.5 (black points) each affected by a 2% uncorrelated uncertainty (solid error bars) and a common 10% normalisation uncertainty (the dashed error bars indicate the sum in quadrature of correlated and uncorrelated uncertainties). The horizontal red line corresponds to the best fit value given in Eq. (74) while the horizontal blue lines correspond to the shifted values according to Eq. (85).

the experimental central values. However, as explained above, a visual comparison in the presence of (large) correlations is meaningful only when including the systematic shifts. The inclusion of the shifts yields the two horizontal blue lines in Fig. 2. The shifted values are  $\hat{\mu}_1 \simeq 8.22$  and  $\hat{\mu}_2 \simeq 8.25$  that indeed fulfil the common expectation. It is worth noticing that the best fit  $\chi^2$  per degree of freedom is around 4.4, signifying that the measurements, that are assumed to refer to the same quantity, are in strong tension. This is a first suggestion that there is no problem with the standard inclusion of correlated normalisation uncertainties in the covariance matrix while the problem seems to be in the data set.

One may ask what is the meaning of a value of  $\bar{\mu}$  that can be smaller than both  $x_1$  and  $x_2$  or in general far from both of them. To answer this question I first observe that, according to Eq. (74),  $\bar{\mu}$  tends to zero as the uncorrelated uncertainties  $\sigma_1$  and  $\sigma_2$  tend to zero. This can be seen more explicitly by setting  $\sigma_1 = \sigma_2 = \varepsilon$  and expanding the  $\chi^2$  around  $\varepsilon = 0$ :

$$\chi^2 \rightarrow \frac{\mu^2(x_1 - x_2)^2}{\varepsilon^2(x_1^2 + x_2^2)} + \frac{[\mu(x_1 + x_2) - (x_1^2 + x_2^2)]^2}{\delta_f^2(x_1^2 + x_2^2)^2} + \mathcal{O}(\varepsilon^2). \quad (86)$$

Therefore, if  $x_1 \neq x_2$ , for  $\varepsilon \rightarrow 0$  the dominant term is the first one and the minimum of the  $\chi^2$  in this limit is clearly at  $\mu = 0$ . If  $x_1 = x_2 = x$ , instead, the first term drops and the best fit value of the parameter  $\mu$  is determined by the second term that has its minimum at  $\mu = x$ .

This can be interpreted as follows: as  $\sigma_1$  and  $\sigma_2$  tend to zero, the covariance matrix tends to become singular signifying that  $x_1$  and  $x_2$  are no longer independent. However, if  $x_1 = x_2 = x$  the singularity cancels. This is due to the fact that, without uncorrelated uncertainties, the number of points of the data set effectively reduces to one. As a matter of fact, in this case  $\varepsilon$  can safely be set to zero and the  $\chi^2$  becomes:

$$\chi^2 = \left( \frac{\mu - x}{\delta_f x} \right)^2, \quad (87)$$

that is evidently the  $\chi^2$  associated to one single data point with central value  $x$  and uncertainty  $\delta_f x$ .<sup>6</sup> As a consequence, the best fit value of the parameters  $\mu$  is simply given by  $x$ , as one would expect.

<sup>6</sup>Note that in the case of one-point data set there is no meaning in talking about correlated uncertainties. Therefore,  $\delta_f x$  plays the role of an uncorrelated uncertainty.

If  $x_1 \neq x_2$ , instead, the singularity of the covariance matrix does not cancel and the first term in Eq. (86) dominates in the limit  $\varepsilon \rightarrow 0$ . This is a situation in which two totally dependent measurements of the same quantity have different central values. It is thus to be expected that the  $\chi^2$  becomes arbitrarily large because the two data points are inconsistent and thus impossible to be simultaneously fitted. In this situation, the parameter  $\mu$  attempts to minimise the  $\chi^2$  by tending to zero. More specifically, when correlations dominate over uncorrelated uncertainties, the  $\chi^2$  is proportional to (the square of) the cross difference  $\delta_f x_1 y_2 - \delta_f x_2 y_1$ ,<sup>7</sup> with  $y_i = \mu - x_i$ . Therefore, if  $x_1 \neq x_2$ , this difference is minimal when  $y_i \simeq x_i$  which is equivalent to  $\mu \simeq 0$ . It is also interesting to observe that the first term in Eq. (86) does not depend of  $\delta_f$ . Therefore, the size of the  $\chi^2$  is entirely driven by the difference  $x_1 - x_2$ , no matter the precise value of normalisation uncertainty.

In conclusion, it appears that data sets dominated by normalisation (or equivalently widely spread correlated) uncertainties carry an intrinsic degree of inconsistency that, when attempting to fit them, manifests into the DAB. Therefore, from this point of view, the DAB is not a flaw of the fitting procedure but rather an indication that the data set is intrinsically inconsistent. It thus appears counterproductive, if not wrong, to adjust the definition of the  $\chi^2$  to accommodate an inconsistency of the data set.

## 7.1 The effect of the $t_0$ prescription

In this section, I consider the effect of the iterative  $t_0$  prescription on the best fit value  $\bar{\mu}$ . The  $t_0$  prescription essentially replaces the normalisation uncertainties  $\delta_f x_1$  and  $\delta_f x_2$  with one single correlated uncertainties  $\delta_f t_0$ , where  $t_0$  is a “guess” for  $\mu$ . The net effect of this replacement is that the terms  $\delta_f x_1$  and  $\delta_f x_2$  in the r.h.s. of both Eqs. (80) and (81) are replaced by  $\delta_f t_0$  and thus cancel, reducing to Eqs.(74) and (75). But this should not be surprising because the  $t_0$  prescription is essentially replacing two correlated uncertainties that are different in value with two correlated uncertainties that are instead *equal*. This makes the normalisation case artificially equal to the offset case in which the DAB is not present.

Now, what happens if one iterates the  $t_0$ ? Let us indicate with  $\bar{\mu}_i$  and  $\delta\mu_i^2$  best fit value and variance at the  $i$ -th iteration. As explained above, the zero-th iteration gives:

$$\bar{\mu}_0 = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \quad (88)$$

and:

$$\delta\mu_0^2 = \frac{1 + \frac{\delta_f^2 t_0^2}{\sigma_1^2} + \frac{\delta_f^2 t_0^2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}. \quad (89)$$

Since  $\bar{\mu}_0$  does not depend on  $t_0$ , all following iterations will give the exact same result, *i.e.*  $\bar{\mu}_i = \bar{\mu}_0 \forall i > 0$ . The stabilisation happens only one step further for the uncertainty for which one finds  $\delta\mu_i^2 = \delta\mu_1^2 \forall i > 1$ , with:

$$\delta\mu_1^2 = \frac{1 + \frac{\delta_f^2 \bar{\mu}_0^2}{\sigma_1^2} + \frac{\delta_f^2 \bar{\mu}_0^2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} + \delta_f^2 \bar{\mu}_0^2. \quad (90)$$

As above, I set  $\sigma_1 = \sigma_2 = \varepsilon$  getting:

$$\bar{\mu}_0 = \frac{x_1 + x_2}{2}, \quad (91)$$

and:

$$\delta\mu_1^2 = \frac{\varepsilon^2}{2} + \frac{\delta_f^2 (x_1 + x_2)^2}{4}. \quad (92)$$

Now, I plug  $\bar{\mu}_0$  into the original definition of the  $\chi^2$ . The result is:

$$\chi^2 = \left( \frac{x_1 - x_2}{2} \right)^2 \left[ \frac{1}{\varepsilon^2 + \delta_f^2 (x_1^2 + x_2^2)} + \frac{1}{\varepsilon^2} \right]. \quad (93)$$

<sup>7</sup>In general, this can be seen as  $\sigma_{c,1} y_2 - \sigma_{c,2} y_1$ .

## References

- [1] R. D. Ball *et al.* [NNPDF Collaboration], JHEP **1005** (2010) 075 doi:10.1007/JHEP05(2010)075 [arXiv:0912.2276 [hep-ph]].
- [2] G. D’Agostini, Nucl. Instrum. Meth. A **346** (1994) 306. doi:10.1016/0168-9002(94)90719-6
- [3] R. D. Ball *et al.*, JHEP **1304** (2013) 125 doi:10.1007/JHEP04(2013)125 [arXiv:1211.5142 [hep-ph]].
- [4] R. Boughezal, A. Guffanti, F. Petriello and M. Ubiali, JHEP **1707** (2017) 130 doi:10.1007/JHEP07(2017)130 [arXiv:1705.00343 [hep-ph]].
- [5] Wikipedia contributors, Chi-square distribution — Wikipedia, The Free Encyclopedia, 2020.
- [6] R. D. Ball *et al.* [NNPDF], Nucl. Phys. B **809** (2009), 1-63 [erratum: Nucl. Phys. B **816** (2009), 293] doi:10.1016/j.nuclphysb.2008.09.037 [arXiv:0808.1231 [hep-ph]].
- [7] R. D. Ball *et al.* [NNPDF], JHEP **04** (2015), 040 doi:10.1007/JHEP04(2015)040 [arXiv:1410.8849 [hep-ph]].
- [8] R. D. Ball *et al.* [NNPDF], JHEP **05** (2010), 075 doi:10.1007/JHEP05(2010)075 [arXiv:0912.2276 [hep-ph]].