1 The χ^2 in the presence of correlations

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

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

where m_i, with $i = 1, \ldots, n$, 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[1], and $\sigma^{(l)}_{i,\text{corr}}$, with $l = 1, \ldots, k$, 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]):

$$V_{ij} = \left(\sigma^2_{i,\text{stat}} + \sigma^2_{i,\text{unc}}\right) \delta_{ij} + \sum_{l=1}^{k} \sigma^{(l)}_{i,\text{corr}} \sigma^{(l)}_{j,\text{corr}} \,. \tag{2}$$

This is a clearly symmetric matrix. Given a set of predictions t_i corresponding to the n measurements of the ensamble, the χ^2 takes the form:

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

where in the second equality we have used the matricial notation and defined $y_i = m_i - t_i$. A convenient way to compute the χ^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:

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T \,, \tag{4}$$

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

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

It is then easy to see that the χ^2 can be written as:

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

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 χ^2.

[1] 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.

2 Additive and multiplicative uncertainties

The correlated systematic uncertainties $\sigma^{(l)}_{i,\mathrm{corr}}$ 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 $\sigma^{(l)}_{i,\mathrm{corr}}$ as relative to the associate central value m_i, so that we define[2]:

$$\sigma^{(l)}_{i,\mathrm{corr}} \equiv \delta^{(l)}_{i,\mathrm{corr}} m_i \tag{9}$$

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

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

Now we 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:

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

It is well known 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 [1], 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:

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

3 Artificial generation of correlated systematics

In order to implement the definition of the χ^2 discussed above, it is necessary to have the experimental information in terms of the correlated systematic uncertainties $\sigma^{(l)}_{i,\mathrm{corr}}$. 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 λ_i associated:

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

[2]Note that this redefinition does not change the nature of the uncertainties, additive uncertainties remain additive as well as multiplicative uncertainties remain multiplicative.

If we define:

$$\sigma^{(l)}_{i,\mathrm{corr}} = \sqrt{\lambda_l} x^{(l)}_i \,, \quad i,l = 1,\ldots,n \,, \tag{14}$$

one can show that:

$$\sum_{l=1}^{n} \sigma^{(l)}_{i,\mathrm{corr}} \sigma^{(l)}_{j,\mathrm{corr}} = C_{ij} \,. \tag{15}$$

To prove this equality we start from the following matricial relation:

$$\mathbf{C} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{-1} \,, \tag{16}$$

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

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

so that:

$$\mathbf{C} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^T \,. \tag{18}$$

It follows that:

$$C_{ij} = \sum_{k,l=1}^{n} Q_{ik} \Lambda_{kl} Q_{jl} = \sum_{k,l=1}^{n} x^{(k)}_i \lambda_k \delta_{kl} x^{(l)}_j = \sum_{l=1}^{n} \lambda_l x^{(l)}_i x^{(l)}_j = \sum_{l=1}^{n} \sigma^{(l)}_{i,\mathrm{corr}} \sigma^{(l)}_{j,\mathrm{corr}} \,, \tag{19}$$

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:

$$\mathbf{V} = \mathbf{U} + \mathbf{C} \,, \tag{20}$$

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

$$U_{ij} = s_i^2 \delta_{ij} \,. \tag{21}$$

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

$$\mathbf{C} = \mathbf{V} - \mathbf{U} \,, \tag{22}$$

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 us 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, we need to write the χ^2 in terms of the so-called "nuisance parameters" λ_α. One can show that the definition of the χ^2 in Eq. (3) is equivalent to [3]:

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

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

$$\frac{\partial \chi^2}{\partial \lambda_\beta} = 0\,. \tag{24}$$

This yields the system:

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

with:

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

that determines the values of λ_β. The quantity:

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

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

$$\bar{t}_i = t_i + d_i\,, \tag{28}$$

the χ^2 reads:

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

Therefore, up to a penalty term given by the sum of the square of the nuisance parameters, the χ^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 χ^2

A relevant question to ask is what is the effect of possible cuts on the input dataset on the form of the χ^2 in the presence of correlations. In order to address this question, we consider a simple dataset made of two datapoints equipped with a single uncorrelated uncertainty and a single fully-correlated uncertainty:[3]

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

The resulting covariance matrix reads:

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

[3]Note that we need to include *also* an uncorrelated uncertainty because otherwise the covariance matrix would be singular. This 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, we a have a set of two theoretical predictions $\{t_i\}$ that allows us to define a column vector of residuals $\mathbf{r}^{\mathrm{T}} = (r_1,\ r_2)$ with $r_i = m_i - t_i$. The χ^2 then takes the following explicit expression:

$$\chi^2 = \mathbf{r}^{\mathrm{T}} \cdot \mathbf{V}^{-1} \cdot \mathbf{r} = \frac{r_1^2\left[(\sigma_2^u)^2 + (\sigma_2^c)^2\right] + r_2^2\left[(\sigma_1^u)^2 + (\sigma_1^c)^2\right] - 2r_1r_2\sigma_1^c\sigma_2^c}{(\sigma_1^u\sigma_2^u)^2 + (\sigma_1^c\sigma_2^u)^2 + (\sigma_1^u\sigma_2^c)^2} . \quad (33)$$

Now, suppose we want to exclude the datapoint with $i = 2$ from the definition of the χ^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 χ^2:

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

However, this procedure might be questioned in that it artificially modifies the original dataset by effectively defining a *new* dataset where the point $i = 2$ is not present. In the absence of correlations this is a sound procedure because each single datapoint can be regarder 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 dataset but yet allows one to prevent specific datapoints to contribute to the χ^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:

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

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

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

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 $\sigma_2^u \to 0$ Eq. (35), which gives:

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

In this case, not only any reference to the point $i = 2$ drops out, but also the dependence of the χ^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, we 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 χ^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 χ^2 either. Importantly, if we take the limit $\sigma_1^u \to 0$, the χ^2 diverges no matter the value of σ_1^c. This is consistent with setting $\sigma_1^u = \sigma_2^u$ since the beginning in Eq. (30) which indeed gives a singular χ^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 we consider the Monte Carlo generation of articial 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 χ^2 distribution with as many degrees of freedom as number of data points. Following Ref. [5], we require that the χ^2, that in matricial form can be expressed as:

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

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:

$$y_j = f_j - m_j\,, \tag{39}$$

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

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

where $z_1, \ldots, z_n$ are independent, standard normal random variables. As already discussed above, Eq. (38) can alternatively be written as:

$$\chi^2 = \left|\mathbf{L}^{-1} \cdot \mathbf{y}\right|^2\,, \tag{41}$$

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

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T\,. \tag{42}$$

Comparing Eqs. (41) and (40), one can infer that the following relation should hold:

$$\mathbf{L}^{-1} \cdot \mathbf{y} \sim \mathbf{z}\,, \tag{43}$$

which immediately implies:

$$\mathbf{y} \sim \mathbf{L} \cdot \mathbf{z}\,, \tag{44}$$

or equivalently:

$$f_j \sim m_j + \sum_{i=1}^{n} L_{ji} z_i\,. \tag{45}$$

In conclusion, in order for the χ^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. (45).

It is instructive to work out how Eq. (45) relates to more "standard" approaches to the generation of Monte Carlo replicas such as those 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. (45), this will finally allow us 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:

$$\mathbf{V} = \mathbf{D} + \sum_{l=1}^{k} \mathbf{V}^{(l)} = \sum_{l=0}^{k} \mathbf{L}^{(l)} \cdot \mathbf{L}^{(l)T} \,, \tag{46}$$

with:

$$L_{ij}^{(l)} = \begin{cases} s_i \delta_{ij} & l = 0 \,, \\ \sigma_i^{(l)} \delta_{j1} & 1 \leq l \leq k \,. \end{cases} \tag{47}$$

On the other hand, we also know that:

$$\mathbf{V} = \mathbf{L} \cdot \mathbf{L}^T \,, \tag{48}$$

therefore:

$$\mathbf{L} \cdot \mathbf{L}^T = \sum_{l=0}^{k} \mathbf{L}^{(l)} \cdot \mathbf{L}^{(l)T} \,. \tag{49}$$

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 be 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:

$$\widetilde{L}_{ij}^{(l)} = \sigma_i^{(l)} \delta_{jl} \,, \quad 1 \leq l \leq k \leq n \,, \tag{50}$$

such that:

$$\left(\sum_{l=1}^{k} \widetilde{\mathbf{L}}^{(l)} \right) \cdot \left(\sum_{l'=1}^{k} \widetilde{\mathbf{L}}^{(l')} \right) = \sum_{l=1}^{k} \mathbf{V}^{(l)} \,. \tag{51}$$

However, choosing:

$$\widetilde{\mathbf{L}} = \mathbf{L}^{(0)} + \sum_{l=1}^{k} \widetilde{\mathbf{L}}^{(l)} \,, \tag{52}$$

and multiplying it for its transpose would give:

$$\widetilde{\mathbf{L}} \cdot \widetilde{\mathbf{L}}^T = \mathbf{V} + \sum_{l=1}^{k} \left(\mathbf{L}^{(0)} \cdot \widetilde{\mathbf{L}}^{(l)T} + \widetilde{\mathbf{L}}^{(l)} \cdot \mathbf{L}^{(0)T} \right) \,, \tag{53}$$

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

$$\sum_{m=1}^{n} \widetilde{L}_{im} \widetilde{L}_{jm} = V_{ij} + s_i \sigma_j^{(i)} + s_j \sigma_i^{(j)} \,. \tag{54}$$

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 we just assume $\sigma_j^{(i)} = 0$ for $i > k$. This finally gives:

$$\sum_{m=1}^{n} \widetilde{L}_{im}\widetilde{L}_{jm} = V_{ij}\,, \quad \text{for } i,j > k\,, \tag{55}$$

that means that, by adopting the Cholesky decomposition in Eq. (50), part of the covariance matrix (usually the largest part) is exactly recovered. Therefore, $\widetilde{\mathbf{L}}$ Eq. (52) seems to be a generally good approximator to the real Cholesky decomposition of the covariance matrix $\mathbf{V}$. Plugging this equation into Eq. (45) immediately gives:

$$f_j \sim m_j + z_j s_j + \sum_{l=1}^{k} z_l \sigma_{j,\text{corr}}^{(l)}\,. \tag{56}$$

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

We 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. (56), the difference being that the random numbers used to generate the fluctuation associated to the k correlated uncertainties are not totally independent from those used for the correlated uncertainties are in fact equal to the first k of the n used for the uncorrelated uncertainties (remember that we required $k \leq n$).

When also considering correlated multiplicative uncertainties there are differences across formulas of Refs. [6, 7] and Eq. (56).

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]].