DRAFT 0.5

Glen Cowan
28 July 2022

Systematic uncertainties in unfolding

1 Introduction

Consider the basic unfolding problem where one observes n_i events, $i = 1, \ldots, N$, in N bins of a histogram. Suppose these are assumed to be independent and to each follow a Poisson distribution with expectation value $E[n_i] \equiv \nu_i$ given by

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i \ , \tag{1}$$

where μ_j, $j = 1, \ldots, M$, are the expected numbers of entries in the "true" histogram and β_i is the expected contribution from background processes. For the moment, suppose that the uncertainty in the β_i can be neglected and we will focus on the uncertainty in the response matrix,

$$R_{ij} = P(\text{event found in bin } i | \text{ true value in bin } j) \ , \tag{2}$$

defined here to include all effects of acceptance, efficiency, migration, etc.

The goal is to estimate the parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$. Using the Poisson model for the n_i, the likelihood function is

$$L(\boldsymbol{\mu}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} \ , \tag{3}$$

where the ν_i depend on the parameters $\boldsymbol{\mu}$ through Eq. (1). Maximizing $L(\boldsymbol{\mu})$ gives the maximum-likelihood (ML) estimators or alternatively one may maximize a linear combination of $\ln L(\boldsymbol{\mu})$ and a regularization function to obtain regularized estimators. A more complete description of the formalism and notation of unfolding is given in Ch. 11 of Ref. [1].

2 Including nuisance parameters in the model

Suppose the response matrix R is not fully known but can be expressed as a function of some vector of nuisance parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$. The likelihood function thus includes the nuisance parameters $\boldsymbol{\theta}$ through the response matrix. The correlations between the estimators of the parameters of interest $\boldsymbol{\mu}$ and the nuisance parameters $\boldsymbol{\theta}$ have the usual effect of inflating the statistical errors on the $\hat{\mu}_i$, and in this way the systematic uncertainties are incorporated into the final result.

Several approaches can be used to obtain the full likelihood function $L(\boldsymbol{\mu}, \boldsymbol{\theta})$. Suppose the θ_k are defined such that $\theta_k = 0$ gives the nominal response matrix, and a variation of θ_k

of $\pm\sigma_{u_k}$ thus gives a one-sigma variation of R. That is, the best estimates u_k of the θ_k are treated as measured quantities with "observed" values $u_k = 0$. Suppose here we take the u_k to be independent and Gaussian distributed with means θ_k and standard deviations σ_{u_k}. In this way the likelihood function becomes

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \left[\frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} \right] \prod_{k=1}^{K} \left[\frac{1}{\sigma_{u_k}} \varphi \left(\frac{u_k - \theta_k}{\sigma_{u_k}} \right) \right] , \tag{4}$$

where ν_i is now a function of both $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$, and φ is the standard (zero mean and unit variance) Gaussian pdf. To find the ML estimators of the μ_i, the likelihood (4) is maximized using the n_i from the observed histogram and $u_i = 0$.

Usually the response matrix $R_{ij}(\boldsymbol{\theta})$ is not available as a parametric function of the nuisance parameters. Instead one has a nominal Monte Carlo model that can be used to determine a matrix $R_{ij}^{(0)}$, and then one can make some number of variations to this model either by generating new Monte Carlo data or by reweighting the events from the nominal sample. Suppose there are $2K$ variations corresponding to changes of plus and minus one standard deviation of each nuisance parameter. In principle each MC sample can be used to determine a response matrix $R_{ij}^{(l)}$ with $l = 0, \ldots, 2K$, and from these one can parametrize the matrix $R_{ij}(\boldsymbol{\theta})$ using Machine-Learning algorithms such as neural networks (see, e.g., [2]).

The Machine-Learning approach to parametrising the response matrix poses many challenges and therefore in practice one may consider simpler procedures that entail certain approximations. If the variations of the response matrix within the one-sigma range of the θ_k are not too nonlinear we can expand $R(\boldsymbol{\theta})$ to first order in a Taylor series about $\boldsymbol{\theta} = 0$ and use this in the expression for $\nu_i(\boldsymbol{\mu}, \boldsymbol{\theta})$ to obtain

$$\begin{aligned} \nu_i(\boldsymbol{\mu}, \boldsymbol{\theta}) &\approx \sum_{j=1}^{M} \left(R_{ij}(0) + \sum_{k=1}^{K} \left. \frac{\partial R_{ij}}{\partial \theta_k} \right|_{\boldsymbol{\theta}=0} \theta_k \right) \mu_j + \beta_i \\ &= \nu_i(\boldsymbol{\mu}, 0) + \sum_{k=1}^{K} \left(\sum_{j=1}^{M} \left. \frac{\partial R_{ij}}{\partial \theta_k} \right|_{\boldsymbol{\theta}=0} \mu_j \right) \theta_k \\ &= \nu_i(\boldsymbol{\mu}, 0) + \sum_{k=1}^{K} \delta\nu_i^{(k)}(\boldsymbol{\mu}) \frac{\theta_k}{\sigma_{u_k}} , \end{aligned} \tag{5}$$

where we have defined

$$\delta\nu_i^{(k)}(\boldsymbol{\mu}) = \sigma_{u_k} \sum_{j=1}^{M} \left. \frac{\partial R_{ij}}{\partial \theta_k} \right|_{\boldsymbol{\theta}=0} \mu_j \, . \tag{6}$$

Using this first-order approximation, the likelihood function becomes

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \left[\frac{(\nu_i(\boldsymbol{\mu}, 0) + \Delta\nu_i(\boldsymbol{\mu}, \boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\nu_i(\boldsymbol{\mu}, 0) + \Delta\nu_i(\boldsymbol{\mu}, \boldsymbol{\theta}))} \right] \prod_{k=1}^{K} \left[\frac{1}{\sigma_{u_k}} \varphi \left(\frac{u_k - \theta_k}{\sigma_{u_k}} \right) \right] , \tag{7}$$

where

$$\Delta\nu_i(\boldsymbol{\mu},\boldsymbol{\theta}) = \sum_{k=1}^{K} \delta\nu_i^{(k)}(\boldsymbol{\mu}) \frac{\theta_k}{\sigma_{u_k}} \ . \tag{8}$$

To find the $\delta\nu_i(\boldsymbol{\mu})$ from Eq. (6), we need to differentiate the response matrix with respect to the nuisance parameters; this requires $R(\boldsymbol{\theta})$, and we will suppose that this is not readily available. An approximate procedure is to use the $2K+1$ Monte Carlo models (nominal plus $2K$ variations) to estimate the corresponding expectation values of the observed distribution, which we will write as $\nu_i^{(l)}$, $l = 0, \ldots, 2K$. It is important to note that these numbers are constants not depending on $\boldsymbol{\mu}$ or $\boldsymbol{\theta}$, in contrast to the $\nu_i(\boldsymbol{\mu},\boldsymbol{\theta})$ from Eq. (1). In practice the values are estimates from finite MC samples and thus have statistical errors; here we will assume that these errors can be neglected.

The approximation considered here is to replace the terms $\delta\nu_i^{(k)}(\boldsymbol{\mu})$ in Eq. (8) by corresponding terms estimated from the nominal Monte Carlo model and the $2K$ variations. To do this we can define

$$\delta\tilde{\nu}_i^{(k)}(\boldsymbol{\theta}) = \begin{cases} \nu_i^{(k,+)} - \nu_i^{(0)} & \theta_k \geq 0 \ , \\ \nu_i^{(0)} - \nu_i^{(k,-)} & \theta_k < 0 \ , \end{cases} \tag{9}$$

where $\nu_i^{(k,+)}$ and $\nu_i^{(k,-)}$ refer to the values of $\nu_i^{(l)}$ found from the modified Monte Carlo samples that correspond to plus and minus one-sigma variations of the nuisance parameter θ_k and $\nu_i^{(0)}$ is from the nominal model. The terms $\delta\tilde{\nu}_i^{(k)}(\boldsymbol{\theta})$ are then used to define

$$\Delta\tilde{\nu}_i(\boldsymbol{\theta}) = \sum_{k=1}^{K} \delta\tilde{\nu}_i^{(k)}(\boldsymbol{\theta}) \frac{\theta_k}{\sigma_{u_k}} \ . \tag{10}$$

By using these $\Delta\tilde{\nu}_i(\boldsymbol{\theta})$ in the likelihood function (7) instead of $\Delta\nu_i(\boldsymbol{\mu},\boldsymbol{\theta})$, one no longer needs to find the response matrix explicitly in terms of the nuisance parameters. As it is much easier to determine the folded distributions $\nu_i^{(l)}$ from the MC samples corresponding to the different variations, than it is to find the response matrices $R_{ij}^{(l)}$, this method represents an important simplification.

There are two approximations used in this approach: First, the variation of the response matrix R is assumed to be sufficiently linear in the nuisance parameters so that the first-order expansion is justified. Second, we are ignoring the dependence of $\Delta\nu_i(\boldsymbol{\theta})$ on the parameters $\boldsymbol{\mu}$, which would be present were one to use the definitions of $\Delta\nu_i$ through Eqs. (6) and (8). The validity of both of these approximations should be checked when this method is used.

The method can be extended easily to cover uncertainties in the expected numbers of background events β_i. Suppose these can be expressed in terms of a set of nuisance parameters as $\beta_i(\boldsymbol{\theta})$. As above, suppose the best estimates of the nuisance parameters are $u_k = 0$, and the u_k are treated as Gaussian distributed random variables with standard deviations σ_{u_k}. Then these parameters will enter into the likelihood function through the β_i and one also includes the corresponding Gaussian terms for the u_k as in Eq. (4).

If the background predictions are available from a nominal model $\beta_i^{(0)}$ and also from some variations $\beta_i^{(k,+)}$ and $\beta_i^{(k,-)}$ corresponding to plus and minus one standard deviation of θ_k, then these can be used in the same way as for the terms $\nu_i^{(k,+)}$ and $\nu_i^{(k,-)}$ in Eq. (9) to compute the corresponding $\delta\tilde{\nu}_i^{(k)}$. These can then be included in the likelihood through their contribution to $\Delta\tilde{\nu}_i$ in Eq. (10) together with the corresponding Gaussian terms for the u_k.

3 Propagating the systematic uncertainties

Once the likelihood function $L(\boldsymbol{\mu}, \boldsymbol{\theta})$ is available with its full dependence on the parameters of interest $\boldsymbol{\mu}$ and the nuisance parameters $\boldsymbol{\theta}$, there are several ways to obtain parameter estimates and their uncertainties.

The usual procedure in the frequentist approach is to maximize the likelihood with respect to all of the parameters, yielding estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\theta}}$. Letting $\boldsymbol{\lambda} = (\mu_1, \ldots, \mu_M, \theta_1, \ldots, \theta_K)$ be the $M + K$-dimensional vector of all the parameters, the inverse of the covariance matrix $U_{ij} = \mathrm{cov}[\hat{\lambda}_i, \hat{\lambda}_j]$ can be estimated from the Hessian matrix

$$U_{ij}^{-1} = - \left. \frac{\partial^2 \ln L}{\partial \lambda_i \, \partial \lambda_j} \right|_{\hat{\boldsymbol{\lambda}}} \; . \tag{11}$$

The covariance matrix for $\hat{\boldsymbol{\mu}}$ is obtained by first inverting U^{-1} and then extracting the submatrix corresponding to the parameters of interest.

In the case of regularized unfolding, however, the estimators are constructed by maximimizing a linear combination of the log-likelihood and a regularization function $S(\boldsymbol{\mu})$,

$$\Phi(\boldsymbol{\mu}, \boldsymbol{\theta}) = \ln L(\boldsymbol{\mu}, \boldsymbol{\theta}) + \tau S(\boldsymbol{\mu}) \; , \tag{12}$$

where τ is the regularization parameter. In this case it is no longer true that the inverse Hessian of $\ln L(\boldsymbol{\mu})$ or $\Phi(\boldsymbol{\mu})$ gives the covariance matrix of the parameters. Approximate formulae are given for this case in Ref. [3] and Ref. [1] Sec. 11.6, but the approximations used are not valid for very strong or very weak regularization.

A more precise determination of the covariance matrix of the estimators can be obtained using a toy Monte Carlo simulation. The two approaches below should be equivalent, but the second method with Bayesian treatment of the systematics should be much easier to implement.

3.1 Covariance with fully frequentist treatment

Suppose the estimators are constructed, e.g., by maximizing $\Phi(\boldsymbol{\mu}, \boldsymbol{\theta})$ or by some other method. In any case, one finds a set of estimators $\hat{\boldsymbol{\mu}}(\mathbf{n}, \mathbf{u})$ that are functions of the data, which include both the primary measurements $\mathbf{n}$ as well as the control measurements $\mathbf{u}$. The covariances of the estimators are

$$\mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = E\left[\hat{\mu}_i \hat{\mu}_j\right] - E\left[\hat{\mu}_i\right] E\left[\hat{\mu}_j\right] \; . \tag{13}$$

If the expectation values are taken only with respect to the primary data $\mathbf{n}$ while holding the control measurements fixed to their nominal values, then one obtains the statistical covariance matrix, e.g., using

$$E_{\mathbf{n}}[\hat{\mu}_i] = \sum_{\mathbf{n}} \hat{\mu}_i(\mathbf{n}, \mathbf{u}) P(\mathbf{n}|\boldsymbol{\mu}, \mathbf{u}) \; . \tag{14}$$

Or if one fixes $\mathbf{n}$ to the observed values and computes the expectation value with respect to the control measurements $\mathbf{u}$, then one finds the systematic covariances,

$$E_{\mathbf{u}}[\hat{\mu}_i] = \int \hat{\mu}_i(\mathbf{n}, \mathbf{u}) f(\mathbf{u}|\boldsymbol{\theta})\, d\mathbf{u}\ . \tag{15}$$

Or the expectation values can be taken with respect to both $\mathbf{n}$ and $\mathbf{u}$ to obtain the full covariance,

$$E_{\mathbf{n},\mathbf{u}}[\hat{\mu}_i] = \sum_{\mathbf{n}} \int \hat{\mu}_i(\mathbf{n}, \mathbf{u})\, P(\mathbf{n}|\boldsymbol{\mu}, \boldsymbol{\theta})\, f(\mathbf{u}|\boldsymbol{\theta})\, d\mathbf{u}\ . \tag{16}$$

In all cases the expectation values can be found by Monte Carlo sampling of $\mathbf{n}$ and/or $\mathbf{u}$ as appropriate and computing the corresponding averages. In this way the covariance matrix is a function of the assumed values of $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ used to generate the MC data. The estimated covariance matrix is found simply by using the estimates or some other nominal values of $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ for these.

The basic steps to find the systematic covariance matrix with the frequentist procedure are therefore

1. Compute $\nu_i(\boldsymbol{\mu}, \boldsymbol{\theta}),\ i = 1, \ldots, N$ with some nominal values of $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$. Treat these ν_i values as an Asimov data set, i.e., with no Poisson fluctuations, corresponding to the parameter point $(\boldsymbol{\mu}, \boldsymbol{\theta})$.
2. Sample $\mathbf{u} \sim f(\mathbf{u}|\boldsymbol{\theta})$.
3. Apply the unfolding algorithm with the nuisance parameters $\boldsymbol{\theta}$ set equal to $\mathbf{u}$ in the likelihood to find a set of estimates $\hat{\boldsymbol{\mu}}$. Record the values in an n-tuple.
4. Repeat the procedure many times and compute the covariance by estimating the expectation values below with the corresponding averages:

$$\mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = E\left[\hat{\mu}_i \hat{\mu}_j\right] - E\left[\hat{\mu}_i\right] E\left[\hat{\mu}_j\right]\ . \tag{17}$$

This will give the systematic covariance matrix. Alternatively, one can sample toy data $\mathbf{n} \sim \mathrm{Poisson}(\boldsymbol{\nu})$ as well as sampling $\mathbf{u} \sim f(\mathbf{u}|\boldsymbol{\theta})$; in this case one obtains the full covariance matrix.

3.2 Covariance using Bayesian treatment of systematics

The Bayesian treatment of the systematic uncertainties should lead to results that are either identical or similar to the frequentist method above. Here the control measurements $\mathbf{u}$ are treated as fixed and one computes the expectation values with respect to $\boldsymbol{\theta}$ using the prior

$$\pi(\boldsymbol{\theta}) \propto f(\mathbf{u}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})\ . \tag{18}$$

Here the ur-prior $\pi_0(\boldsymbol{\theta})$ reflects the prior before the control measurements, usually taken as a constant. With a constant ur-prior and a Gaussian for $f(\mathbf{u}|\boldsymbol{\theta})$ one obtains

$$\pi(\boldsymbol{\theta}) = \prod_{k=1}^{K} \left[\frac{1}{\sigma_{u_k}} \varphi\left(\frac{u_k - \theta_k}{\sigma_{u_k}}\right)\right] \tag{19}$$

The basic steps are therefore

1. Sample $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

2. Compute $\nu_i(\boldsymbol{\theta}),\ i = 1, \ldots, N$ with the sampled value of $\boldsymbol{\theta}$

3. Treat the $\boldsymbol{\nu}$ values as an Asimov data set, i.e., with no Poisson fluctuations, corresponding to the nuisance parameter point $\boldsymbol{\theta}$.

4. Apply the unfolding algorithm using the nominal response matrix R to find a set of estimates $\hat{\boldsymbol{\mu}}$. Record the values in an n-tuple.

5. Repeat the procedure many times and compute the covariance by estimating the expectation values below with the corresponding averages:

$$\mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = E\left[\hat{\mu}_i \hat{\mu}_j\right] - E\left[\hat{\mu}_i\right] E\left[\hat{\mu}_j\right] \ . \tag{20}$$

The covariances obtained by this procedure reflect only the systematic uncertainties through the prior $\pi(\boldsymbol{\theta})$, not the statistical uncertainties. If in step (3) one were to generate a data set using $\mathbf{n} \sim \mathrm{Poisson}(\boldsymbol{\nu})$ then the covariances include both the statistical and systematic uncertainties.

References

[1] G. Cowan, *Statistical Data Analysis*, Oxford University Press, 1998.

[2] A. Bozson et al., *ML methods for response matrix parametrisation*, presented at ATLAS Statistics Forum, 1 December 2016, `https://indico.cern.ch/event/590264/`.

[3] P. Verschuuren, *Unfolding with systematic uncertainties*, 2020.