DRAFT 0.1

Glen Cowan
17 July 2017

Decomposition of fit error into statistical and systematic components

On 17 July 2017 the Higgs WG had a discussion on the decomposition of a total fit error (e.g., in the Higgs mass) into statistical and systematic components [1]. This note sketches out a possible way of approaching this problem.

1 Formalism for the total error

Suppose we measure a set of numbers $\mathbf{x}$ whose probability model has a parameter of interest μ and a vector of nuisance parameters $\boldsymbol{\theta}$. These appear in a likelihood function $L_0(\mu, \boldsymbol{\theta}) = P(\mathbf{x}|\mu, \boldsymbol{\theta})$, where the subscript 0 here refers to the absence of the constraint terms described below.

Some of the components of $\boldsymbol{\theta}$ con be constrained by auxiliary measurements $\mathbf{y}$, whose probability depends on the nuisance parameters, i.e., $P(\mathbf{y}|\boldsymbol{\theta}) \equiv L_{\rm aux}(\boldsymbol{\theta})$. For example, for a component θ_i there could be a measured value $y_i = \tilde{\theta}_i$, described by a sampling distribution $p(\tilde{\theta}_i|\theta_i) \equiv L_i(\theta_i)$. In some cases the auxiliary measurement could be based on the same data sample as that used for the main measurement; that is, it could use an independent subset of the events, but this subset will also scale with luminosity together with the subset used for the main measurement. In other cases the auxiliary measurement could be a completely separate experiment; in still others y_i (or $\tilde{\theta}_i$) could simply represent the nominal value, e.g., of a theory parameter, and we treat this formally as if it were a measurement. Regardless of the interpretation, the auxiliary measurements $\mathbf{y}$ are bundled together with the primary measurements $\mathbf{x}$ (here assumed independent) to give the joint probability

$$P(\mathbf{x}, \mathbf{y}|\mu, \boldsymbol{\theta}) = P(\mathbf{x}|\mu, \boldsymbol{\theta}) P(\mathbf{y}|\boldsymbol{\theta}) \ , \tag{1}$$

or written in terms of the corresponding likelihood terms,

$$L(\mu, \boldsymbol{\theta}) = L_0(\mu, \boldsymbol{\theta}) L_{\rm aux}(\boldsymbol{\theta}) \ . \tag{2}$$

By maximizing $L(\mu, \boldsymbol{\theta})$ simultaneously over all of the parameters one obtains an estimator $\hat{\mu}$ for the parameter of interest whose variance reflects the full statistical error, here called $\sigma_{\rm tot}$.

2 Decomposition into statistical and systematic components

Suppose we want to know how $\sigma_{\rm tot}$ will vary if we had a data sample of a different size, i.e., if we had a different luminosity. One may define a parameter

$$\lambda = \frac{\mathcal{L}}{\mathcal{L}_0} \tag{3}$$

as the ratio of the hypothetical to real luminosities. Then the likelihood function can be rewritten, at least approximately, as a function of λ. To do this one must simply scale all quantities related to the number of events in the main data set with λ. This is done both for the expected numbers of events (which are functions of the parameters μ and $\boldsymbol{\theta}$) as well as for the observed numbers of events. This may leave a noninteger number of events, but for now let us assume that the log-likelihood function remains well defined under this change (for Poisson terms this should not be a problem as it is similar to using an Asimov data set). If the likelihood contains any terms that contain directly a statistical error σ_i that is expected to scale with luminosity in the usual way, then this should be replaced by $\sigma_i/\sqrt{\lambda}$. In this way one obtains a new likelihood function

$$L(\mu, \boldsymbol{\theta}; \lambda) \; , \tag{4}$$

which contains the additional parameter λ. Note that here λ is not an adjustable parameter; rather, one fixes its value and fits μ and $\boldsymbol{\theta}$. Note also that it is up to the analyst to decide how the likelihood will change with luminosity; in some cases there could be terms where the scaling is not entirely obvious and some assumptions must be introduced.

For each choice of λ one can carry out the full fit and thus determine $\sigma_{\rm tot}(\lambda)$. One can then define what we mean by a "statistical error" as the component of $\sigma_{\rm tot}$ that scales inversely as the square root of the luminosity. That is, we can make the approximate ansatz

$$\sigma_{\rm tot}(\lambda) = \sqrt{\frac{\sigma_{\rm stat}^2}{\lambda} + \sigma_{\rm sys}^2} \; , \tag{5}$$

and if we have values of $\sigma_{\rm tot}(\lambda)$ determined at several values of λ (0.5, 1., 1.5,...) then these can be used to fit $\sigma_{\rm stat}$ and $\sigma_{\rm sys}$.

More specifically, suppose we plot $\sigma_{\rm tot}^2$ versus λ^{-1}; this should be a straight line with slope $\sigma_{\rm stat}^2$ and intercept $\sigma_{\rm sys}^2$. It is doubtful that the full dependence will be linear, but we can make this approximation in the neighborhood of $\lambda = 1$ and use this to define $\sigma_{\rm stat}$ and $\sigma_{\rm sys}$.

References

[1] Eric Feng et al., `indico.cern.ch/event/654020/`, CERN, 17 July, 2017.