DRAFT 0.0

Glen Cowan
18 May 2020

Discussion on including (Bayesian) systematics

This note is to supplement discussions with Sean, Aaron and the Exotics Group on how to include systematic uncertainties into the clockwork gravity analysis as described in Aaron's talk [1]. Let us suppose that distribution of a kinematic variable m (here the dielectron mass) is specified by a model to have some form $f(m|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$ is a vector of nuisance parameters. In practice the distribution may be known in the form of a histogram but that will not alter significantly the discussion on how to include systematic uncertainties.

Consider first the two main ways to treat the systematic uncertainties: Bayesian and frequentist. In the frequentist approach, we suppose we have a set of values $\tilde{\boldsymbol{\theta}}$ that enter on the same footing as measurements (in many cases they may in fact be control measurements). They are characterized by a sampling distribution $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$. If the components $\tilde{\theta}_i$ are independent, this could be for example a product of Gaussians for $\tilde{\theta}_i$ centred about the hypothesized parameter value θ_i. Here the $\tilde{\theta}_i$ are random variables and the nuisance parameters θ_i are unknown constants. A probability is only associated with the $\tilde{\theta}_i$. So as an example the sampling distribution of $\tilde{\boldsymbol{\theta}}$ might be

$$p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{\tilde{\theta}_i}} e^{-(\tilde{\theta}_i - \theta_i)^2 / 2\sigma^2_{\tilde{\theta}_i}} \ . \tag{1}$$

This function is used together with the rest of the data to construct likelihood function for the parameters of interest, say, $\boldsymbol{\mu} = (k, m_5)$, and the nuisance parameters $\boldsymbol{\theta}$. The nuisance parameters are then eliminated by profiling [2, 3].

In the Bayesian approach, one associates a probability with the parameters $\boldsymbol{\theta}$ and a prior pdf $\pi(\boldsymbol{\theta})$ must be assigned that quantifies the analyst's degree of belief about where the true value of $\boldsymbol{\theta}$ lies. A common method for assigning this prior would be to use the same set of control measurements $\tilde{\boldsymbol{\theta}}$ as was used in the frequent case. Here the initial prior (sometimes called the ur-prior, i.e., before any control measurements) could be a constant, $\pi_0(\boldsymbol{\theta}) = \text{const.}$, and then the prior $\pi(\boldsymbol{\theta})$ is given by the probability of $\boldsymbol{\theta}$ given the control measurements as found from Bayes' theorem,

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}) = \frac{p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})}{\int p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \ . \tag{2}$$

For the special case of a constant $\pi_0(\boldsymbol{\theta})$ and Gaussian $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$, one finds that the prior $\pi(\boldsymbol{\theta})$ is actually equal to $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$. This equality does not hold in general and even in this special case one must keep in mind that the prior $\pi(\boldsymbol{\theta})$ is a pdf for $\boldsymbol{\theta}$, whereas $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ is a pdf for $\tilde{\boldsymbol{\theta}}$.

My understanding from the slides by Aaron [1] is that the analysis follows the Bayesian approach, at least for purposes of the systematic uncertainties. The kinematic distribution $f(m)$ is known at specific points in the space of the nuisance parameters $\boldsymbol{\theta}$, namely, at their estimated values $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$, which gives the nominal distribution, and with one component, say θ_i set to $\theta_i = \tilde{\theta}_i \pm \sigma_{\tilde{\theta}_i}$, and all other components still set to their nominal values. So for N nuisance parameters one has $2N + 1$ versions of $f(m)$, the nominal one $f_0(m) = f(m|\tilde{\boldsymbol{\theta}})$ and

$$f_{\pm j}(m) = f(m|\tilde{\theta}_1, \ldots, \tilde{\theta}_j \pm \sigma_{\tilde{\theta}_j}, \ldots, \tilde{\theta}_N) \;, \quad j = 1, \ldots, N \;. \tag{3}$$

This is then used to find $f(m|\boldsymbol{\theta})$ for arbitrary $\boldsymbol{\theta}$ with (I think) the recipe

$$f(m|\boldsymbol{\theta}) = f_0(m) + \sum_{j=1}^{N} \frac{\theta_j}{\sigma_{\tilde{\theta}_j}} \frac{(f_{+j}(m) - f_{-j}(m))}{2} \;, \tag{4}$$

where the factor of 2 in the denominator accounts for the fact that the variation is symmetrized by taking the difference between the "up 1-σ" and "down 1-σ" distributions. This is similar (or perhaps identical for some choice of options) to what is done in `HistFactory` [4].

For arbitrary fixed m, one has therefore a nominal prediction $f(m|\boldsymbol{\theta})$ that depends on $\boldsymbol{\theta}$. And if we treat $\boldsymbol{\theta}$ as a (Bayesian) random variable then f also has a pdf given by

$$p(f|m) = \int \delta(f - f(m|\boldsymbol{\theta}))\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \;, \tag{5}$$

where here the f with no arguments is now a random variable, whereas the function $f(m|\boldsymbol{\theta})$ is treated as a label for the distribution found from Eq. (4) at given values of m and $\boldsymbol{\theta}$.

The integral of Eq. (5) can be computed numerically by sampling $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$, using this value to evaluate $f(m|\boldsymbol{\theta})$ and then recording the resulting $f(m)$ (Aaron refers to this as a "toy distribution"). For example, at fixed m one can imagine the resulting distribution of f values will have some roughly bell-shaped distribution with a certain expectation value $E[f|m]$ and variance $V[f|m]$ (or standard deviation $\sigma_f(m) = \sqrt{V[f|m]}$).

In a similar way for any two fixed values m and m' one could determine the covariance $\text{cov}[f(m), f(m')]$, and thus the mean function $E[f|m]$ and covariance $\text{cov}[f(m), f(m')]$ can be used to define a Gaussian Process [5]. This could in principle be used in subsequent stages of the analysis to incorporate the systematic uncertainties.

The mean function $E[f|m]$ is not necessarily identical to the nominal distribution $f(m|\tilde{\boldsymbol{\theta}})$ but intuitively one would expect the two to be very close. One can therefore define $E[f|m] \pm \sigma_f(m)$ as giving the nominal distribution with a one-sigma error band. That is, the width of the band represents the posterior standard deviation of f given $\tilde{\boldsymbol{\theta}}$.

It is not completely clear whether this is indeed what is being described by Aaron in Ref. [1] – apologies if I've misunderstood. What one needs to clarify at this point is how the information thus obtained is used in the rest of the analysis. It is not obvious how the nominal distribution and an error band (or even a covariance function) can be used to search for a signal or to set limits on the parameters of the signal model.

In contrast, in the frequentist approach the nominal distribution and error band are not needed. Rather, one requires the distribution $f(m|\boldsymbol{\theta})$ for an arbitrary point in parameter space, found using an interpolation procedure either as described above or as implemented, for example, in `HistFactory` [4]. Once this is found the frequentist analysis can be done with the profile likelihood ratio as described in Refs. [2, 3].

References

[1] Aaron White, *Dilepton Background Systematics*, Presentation at ATLAS PC Meeting, 6 March 2020.

[2] G. Cowan, *Ideas on how to include systematics in a search* (draft note), www.pp.rhul.ac.uk/~cowan/stat/notes/test_with_systematics.pdf, 2020.

[3] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, Eur. Phys. J. C 71 (2011) 1554.

[4] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata and Wouter Verkerke, *HistFactory: A tool for creating statistical models for use withRooFit and RooStats*, CERN-OPEN-2012-016.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, the MIT Press, 2006; www.gaussianprocess.org/gpml/ .