BAYESIAN STATISTICAL METHODS FOR PARTON ANALYSES

GLEN COWAN
Physics Department, Royal Holloway, University of London,
Egham, Surrey TW20 0EX, UK,
E-mail: g.cowan@rhul.ac.uk

The uncertainties in predictions for LHC observables are often dominated by systematic effects that are difficult to quantify in the traditional frequentist statistical framework. Uncertainties related to parton densities are an important example. Difficulties with the frequentist approach to this problem are examined and the Bayesian alternative is explored.

1. Introduction

To predict a cross section that can be measured at a hadron collider such as the LHC, one computes the convolution of a parton level cross section with parton density functions (PDFs). Uncertainties can thus stem from the limited order of the perturbatively computed parton-level cross section and also from the imperfect modelling of non-perturbative physics through the PDFs. Furthermore the parameters entering into the prediction are determined by fits to data that themselves have imperfectly understood systematics and which are not in all cases mutually consistent.

Most previous analyses of PDFs have been done using frequentist statistical methods. In this framework one does not speak of the probability of a parameter; these rather are treated as constants whose values must be estimated. The PDF parameters are often determined by least-squares fits using data from deep-inelastic scattering and other processes. One constructs a global χ^2 whose minimum, $\chi^2_{\rm min}$, determines the fitted parameter values. The rule from frequentist statistics to obtain the standard deviations of the fitted parameters is to vary the parameters until one finds $\chi^2 = \chi^2_{\rm min} + 1$. This recipe, however, often results in unrealistically small errors in predicted cross sections.

The apparent failure of the '$\chi^2_{\rm min} + 1$' rule takes place because, in addition to the statistical errors, one can have model uncertainties and systematics that are not fully taken into account. In order to report realistic estimates for uncertainties, several groups producing PDF fits have chosen to allow the χ^2 to increase from its minimum by substantially greater amounts, such as 50 or 100. This results in reasonable error estimates but it is an *ad hoc* recipe obtained by extension of a frequentist statistical method to a problem for which it was not designed. The Bayesian statistical approach offers a more transparent means to incorporate systematic uncertainties into predicted cross sections.

2. The Bayesian approach

In Bayesian statistics, a probability can be associated not only with data but also with a hypothesis, e.g,. a hypothesized parameter value. In this case the probability is interpreted as a degree of belief about where the parameter's true value lies.

Suppose the experiments we consider provide us with a set of data $\vec{y}$. The probability to obtain these data will be given by a joint probability density function $f(\vec{y}|\vec{\theta})$, where $\vec{\theta}$ is a set of parameters. In general we can write for the expectation value of the ith measurement $E[y_i] = \mu(x_i;\theta) + b_i$. Here $\mu(x;\vec{\theta})$ is the prediction of our model as a function of a control variable x and b_i is a potential bias.

If we evaluate the joint probability density $f(\vec{y}|\vec{\theta})$ with the data actually obtained and regard it as a function of $\vec{\theta}$, then this is the likelihood function $L(\vec{y}|\vec{\theta})$. The probability for the parameters $\vec{\theta}$ given the data $\vec{y}$ is obtained using Bayes' theorem as

$$p(\vec{\theta}|\vec{y}) = \frac{L(\vec{y}|\vec{\theta})\pi(\vec{\theta})}{\int L(\vec{y}|\vec{\theta})\pi(\vec{\theta})\, d\vec{\theta}} \propto L(\vec{y}|\vec{\theta})\pi(\vec{\theta}) \ . \tag{1}$$

Here $\pi(\vec{\theta})$ is the prior probability for $\vec{\theta}$, which reflects our degree of belief about the parameter values before consideration of the data $\vec{y}$.

Often experimental data $\vec{y} = (y_1, \ldots, y_n)$ are reported together with an $n \times n$ covariance matrix $V_{\rm stat}$, which reflects their statistical errors, and also with a separate matrix $V_{\rm sys}$ for the systematic uncertainties. In a frequentist least-squares fit one would estimate the parameters $\vec{\theta}$ from the minimum of $\chi^2(\vec{\theta}) = (\vec{y} - \vec{\mu}(\vec{\theta}))^T V^{-1} (\vec{y} - \vec{\mu}(\vec{\theta}))$. For the covariance matrix V, one could use only $V_{\rm stat}$ but in order to include the systematic errors one can also take the sum $V = V_{\rm stat} + V_{\rm sys}$. The minimum of $\chi^2(\vec{\theta})$ gives the parameter estimates and the '$\chi^2_{\rm min} + 1$' rule gives their errors (covariances).

The following Bayesian analysis will give essentially the same result as a least-squares fit with $V = V_{\rm stat} + V_{\rm sys}$. We can take

$$L(\vec{y}|\vec{\theta},\vec{b}) \propto \exp\left[-\tfrac{1}{2}(\vec{y}-\vec{\mu}(\vec{\theta})-\vec{b})^T V_{\rm stat}^{-1}(\vec{y}-\vec{\mu}(\vec{\theta})-\vec{b})\right] , \qquad (2)$$

$$\pi_b(\vec{b}) \propto \exp\left[-\tfrac{1}{2}\vec{b}^T V_{\rm sys}^{-1}\vec{b}\right] , \qquad \pi_\theta(\vec{\theta}) = \text{const.} , \qquad (3)$$

$$p(\vec{\theta},\vec{b}|\vec{y}) \propto L(\vec{y}|\vec{\theta},\vec{b})\pi_\theta(\vec{\theta})\pi_b(\vec{b}) , \qquad (4)$$

where in (4), Bayes' theorem is used to obtain the joint probability for the parameters of interest, $\vec{\theta}$, and also the biases $\vec{b}$. To obtain the probability for $\vec{\theta}$ we integrate (marginalize) over $\vec{b}$,

$$p(\vec{\theta}|\vec{y}) = \int p(\vec{\theta},\vec{b}|\vec{y})\, d\vec{b} . \qquad (5)$$

The mode of $p(\vec{\theta}|\vec{y})$ will be at the same position as the least-squares estimates, and its covariance will be the same as obtained from the $\chi^2_{\rm min}+1$ rule. Similar approaches have been investigated by [1,2].

3. The error on the error

If one stays with the prior probabilities used above, the Bayesian and least-squares approaches deliver the same result. The advantage of the Bayesian framework is that it allows one to refine the assessment of the systematic uncertainties as expressed through the prior probabilities.

For example, the least-squares fit including systematic errors is equivalent to the assumption of a Gaussian prior for the biases. A more realistic prior would take into account the experimenters own uncertainty in assigning the systematic error, i.e., the 'error on the error'. Suppose, for example, that the ith measurement is characterized by a reported systematic uncertainty $\sigma_i^{\rm sys}$ and an unreported factor s_i, such that the prior for the bias b_i is

$$\pi_b(b_i) = \int \frac{1}{\sqrt{2\pi}\sigma_i^{\rm sys}} \exp\left[-\frac{1}{2}\frac{b_i^2}{(s_i\sigma_i^{\rm sys})^2}\right] \pi_s(s_i)\, ds_i . \qquad (6)$$

Here the 'error on the error' is encapsulated in the prior for the factor s, $\pi_s(s)$. For this we can take whatever function is deemed appropriate. For some types of systematic error it could be close to the ideal case of a delta function centred about unity. Many reported systematics are, however, at best rough guesses, and one could easily imagine a function $\pi_s(s)$ with a mean of unity but a standard deviation of, say, 0.5 or more. We have studied using a Gamma distribution for $\pi_s(s)$, which results in substantially longer tails for the prior $\pi_b(b)$ than those of the Gaussian. Related studies using an inverse Gamma distribution can be found in [3,4].

Using a prior for the biases with tails longer than those of a Gaussian results in a reduced sensitivity to outliers, which arise when an experimenter overlooks an important source of systematic uncertainty in the estimated error of a measurement. Furthermore the width of the posterior distribution, which effectively tells one the uncertainty on the parameter of interest, becomes coupled to the internal consistency of the data used. In contrast, high value of $\chi^2_{\min}$ does not lead to small values of the errors obtained from the $\chi^2_{\min} + 1$ rule.

The method can be generalized to cover a wide variety of model uncertainties by including prior probabilities for an enlarged set of model parameters. These additional parameters could represent, for example, the limited flexibility of the parameterization of PDFs at low Q^2, missing higher order terms in the perturbative parts of the prediction, etc.

4. Conclusions

In the Bayesian statistical approach one encapsulates systematic uncertainties in prior probabilities for an enlarged set of model parameters. By using computational tools such as Markov Chain Monte Carlo, one can obtain the predicted uncertainties for quantities of interest. An effort is underway to apply this approach to observables for the LHC.

Acknowledgements

Many thanks to my collaborator Clare Quarman and to the conference organizers for an extremely interesting and fruitful meeting.

References

1. S.I. Alekhin, Eur. Phys. J. C10 (1999) 395.
2. W. Giele, S. Keller and D. Kosower, *Parton Distribution Function Uncertainties*, hep-ph/0104052.
3. G. D'Agostini, *Sceptical combination of experimental results; General considerations and application to ε'/ε*, hep-ex/9910036.
4. V. Dose and W. von der Linden, *Outlier tolerant parameter estimation*, in XVIII Workshop on Maximum Entropy and Bayesian Methods, Kluwer, 1999.