DRAFT 0.1

Glen Cowan
20 December, 2016

Note on rate and shape uncertainties

Consider a set of differential cross sections $d\sigma_i/dy$ where i is a label for a particular final state, e.g., Higgs + n jets with $n = 0, 1, \ldots$ and y here stands for a continuous kinematic variable such as rapidity, $p_{\mathrm{T,H}}$, etc. In practice we will integrate the $d\sigma_i/dy$ over some bins in y to form quantities

$$\sigma_{ij} = \int_j \frac{d\sigma_i}{dy}\, dy \tag{1}$$

where the index j labels a particular bin in the variable y. We would like to express the σ_{ij} as functions of some parameters that can be varied to reflect the uncertainty in the predicted value of σ_{ij}.

This can be done by expressing the differential cross sections as the product of a total cross section and a probability distribution function (pdf) in y:

$$\frac{d\sigma_i}{dy} = \sigma_i f_i(y) \tag{2}$$

where by construction the pdfs $f_i(y)$ are normalized to unity. The uncertainties in the predictions for the total cross sections σ_i and the pdfs $f_i(y)$ may be related but in general are different. We may therefore regard the parametrization to be of the form

$$\frac{d\sigma_i}{dy}(\boldsymbol{\theta}) = \sigma_i(\theta) f_i(y|\theta');, \tag{3}$$

where θ and θ' are sets of nuisance parameters that are not in general disjoint, and $\boldsymbol{\theta}$ is the union of θ and θ'. In fact it is not necessary to consider the factorization in Eq. (3) for what follows.

As a first step, it may be convenient to map the kinematic variable y onto a variable x defined on $0 \leq x \leq 1$. For example, this can be done using

$$x = F_0(y)\ , \tag{4}$$

where $F_0(y)$ is the cumulative distribution of y computed to some given accuracy, e.g., $0th$ order, but in any case let us assume that F_0 is a fixed known function.[1]

An advantage of the transformation (4) is that the pdf of x, $g(x)$, is then to first approximation uniform on $[0, 1]$. To greater accuracy let us suppose can be expressed as a perturbation series in some expansion parameter α,

$$g(x) = g_0(x) + \alpha g_1(x) + \alpha^2 g_2(x) + \ldots\ . \tag{5}$$

[1] This is the procedure used by FT for the B shape function.

Here if the $0th$ order prediction $F_0(y)$ has been used in Eq. (4), then $g_0(x) = 1$.

Let us suppose that the prediction for $g(x)$ is known to, say, $2nd$ order, and the uncertainty arises from the missing functions $g_3(x)$, $g_4(x)$, ... Each of these can be expressed in general as an expansion in some basis functions

$$g_i(x) = \sum_k \theta_k \varphi_k(x) \ . \tag{6}$$

The θ_k thus enter the problem as nuisance parameters. For the basis functions $\varphi_k(x)$ there are many possibilities, such as Bessel functions, Legendre polynomials, etc., (mapped onto the interval $0 \le x \le 1$). One could for example use Bernstein polynomials. The set of $m + 1$ Bernstein basis polynomials of order m are defined as

$$b_{k,m}(x) = \frac{m!}{k!(m-k)!} x^k (1-x)^{m-k} \qquad 0 \le x \le 1 \ . \tag{7}$$

The Bernstein basis polynomials for orders 0 through 5 are shown in Fig. 1.

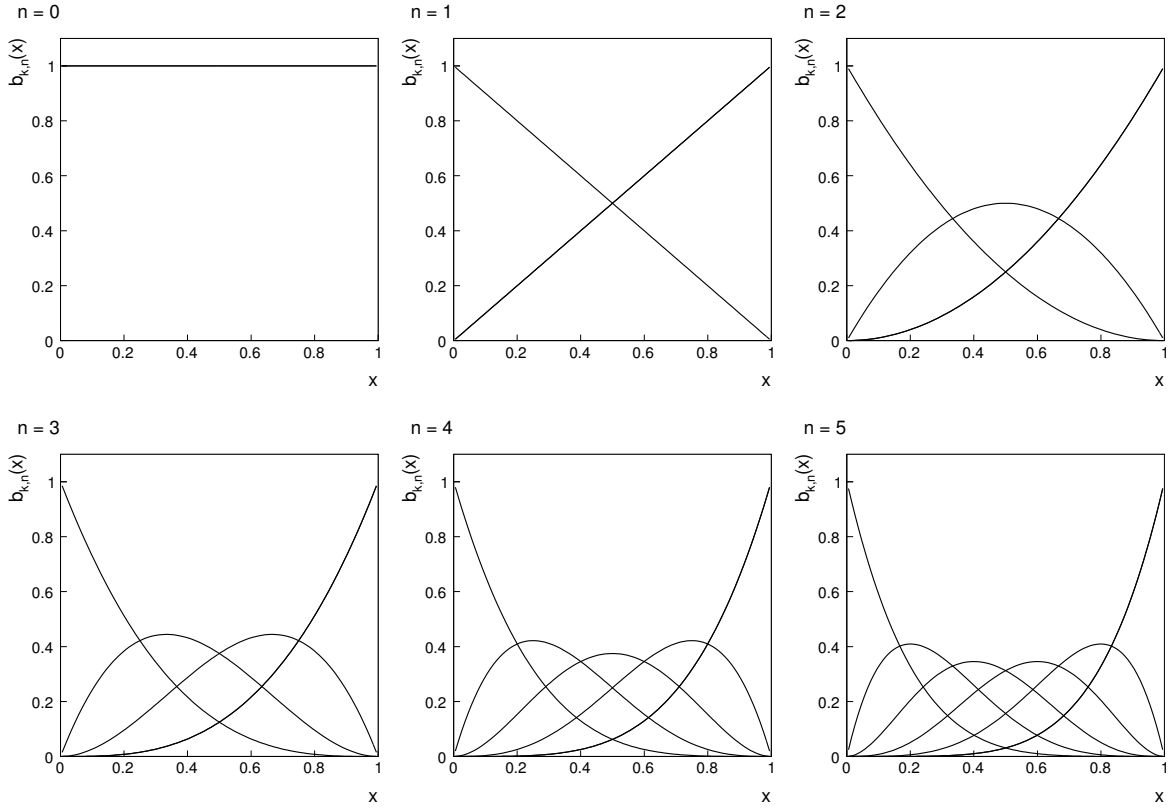

Figure 1: Bernstein basis polynomials of different orders n.

For $\theta_k = 1$, $k = 0, \ldots, m$ in Eq. (6), one has $g_i(x) = 1$, so it is easy to identify the point in parameter space that corresponds to no modification.

An important property of Bernstein polynomials is that a basis polynomial of a given order $m - 1$ can always be written in terms of those of order m:

$$b_{k,m-1}(x) = \frac{m-k}{m} b_{k,m}(x) + \frac{k+1}{m} b_{k+1,m}(x) \ . \tag{8}$$

This means that the Bernstein polynomials defined using basis functions of successively increasing order form a nested family. That is, the model of order m contains as a special case the model of order $m-1$. This will be important in constructing the likelihood ratio test to determine whether it is necessary to increase the number of parameters in the model.

To determine the appropriate number of basis functions one needs to consider the amount of structure expected in the function $g_i(x)$. For Bernstein basis polynomials of order m, for example, the peak of $b_{k,m}$ is at k/m, i.e., the peaks are separating by a spacing of $1/m$. So if one can specify a rough length Δx below which one does not expect significant structure, then one should choose an order $m \approx 1/\Delta x$.

To use the parametrized prediction in an analysis, one needs to specify what parameters will be added to the model and how they should be constrained. Suppose one considers only one additional function $g_i(x)$ corresponding to the first missing order in the expansion. One would then have a corresponding set of coefficients $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_m)$.

In a Bayesian analysis, one could then specify a prior pdf $\pi(\boldsymbol{\theta})$ that represents one's degree of belief as to where these values lie. As a default it would probably be peaked about $\boldsymbol{\theta} = 0$, but this is not necessarily the case. The components may or not be correlated; this is subjective input that must be supplied by the analyst.

In a frequentist analysis, one must take the nominal values for $\boldsymbol{\theta}$ and treat them as measurements, $\tilde{\boldsymbol{\theta}}$. These will have a sampling distribution, which could be, e.g., a Gaussian,

$$p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left[-\tfrac{1}{2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T V^{-1} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})\right] \; . \tag{9}$$

Here $V_{ij} = \mathrm{cov}[\tilde{\theta}_i, \tilde{\theta}_j]$ is the covariance matrix of the "measurements" of the nuisance parameters, which must be supplied by the analyst. The nominal values of $\tilde{\boldsymbol{\theta}}$ would normally be take to be zero.

Alternatively one may consider that the measured values $\tilde{\lambda}_i$ of $\lambda_i = \ln \theta_i$ follow a Gaussian distribution, in which case the $\tilde{\theta}_i = e^{\tilde{\lambda}_i}$ follow a log-normal distribution. One should further consider the possibility that the tails of the distribution fall off less quickly than those of a Gaussian, e.g., using a Student's t distribution.

Once the function $g(x)$ and thus also the differential cross section $d\sigma_i/dy$ are known as a function of the vector of nuisance parameters $\boldsymbol{\theta}$ and the nature of the constraints has been specified, then these can be used in an analysis with any desired function of the $d\sigma_i/dy$, such as the cross section for a particular subprocess within any given bins of y.