Incorporating Systematic Uncertainties

Physics 252C - Lecture 12

Prof. John Conway

Simplest case - measurements

- in the Gaussian regime you might simply estimate your α parameter(s) with a simple estimator based on a χ^2 or maximum likelihood
- example: cross section
- is there a point to putting the systematic nuisance parameter into the likelihood or χ^2 ?
- one can simply quote the statistical uncertainty of the result (from the fit) and quote separately the systematic error, estimated from

$$\sigma_\alpha = \left(\frac{\Delta \hat{\alpha}}{\Delta x} \right) \sigma_x$$

add in quadrature or linearly?

- if systematics are uncorrelated, then they add in quadrature
- but if they are 100% correlated, add them <u>linearly</u>
- what about if they are 100% anticorrelated?

Poisson process with background

- "The Manhattan Project" of the CDF Statistics Committee (March 2003)
- the idea was to compare various methods/ideas for incorporating systematic errors on the efficiency and the background
  - n events observed
  - $b \pm \sigma_b$ expected background
  - $\varepsilon \pm \sigma_\varepsilon$ efficiency
- determine the 95% CL upper limit on the signal rate s

poisson process with background

- frequentist approach: regard the values of b and ε as outcomes of (subsidiary) measurements
- there are unknown true values for these parameters
- coverage probability now becomes a function of the true value of s, b, and ε in principle
- some frequentists are comfortable quoting "average coverage"
- but averaging over the possible true values of unknown parameters is rather Bayesian....no?
- typical end point: a hybrid frequentist/Bayes method

Bayesian treatment of problem

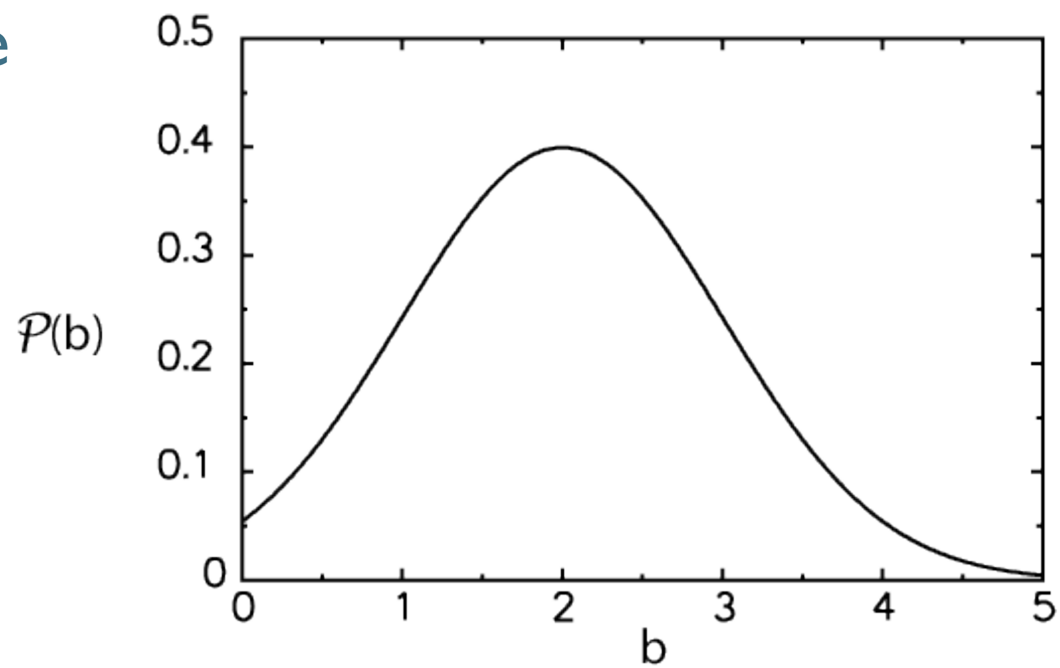
- the goal in the Bayesian approach is a posterior pdf in the parameter of interest: s
- do this by repeated use of Bayes theorem for each uncertainty
- with a background uncertainty only we might have

$$\mathcal{P}(s; n, b, \sigma_b) = \frac{\int \mathcal{L}(n; s, b', \sigma_b)\mathcal{P}(b')\mathcal{P}(s)db'}{\int \mathcal{L}(n; s', b', \sigma_b)\mathcal{P}(b')\mathcal{P}(s)db'ds'}$$

- core likelihood: Poisson
- $\mathcal{P}$(b) is the prior for b (including σ_b)
- $\mathcal{P}$(s) is the prior for s

Prior for background

- suppose your background is b = 2 ± 1 events
- what does this really <u>mean</u>?
- the error represents some sort of 68% CL region
- are we gaussian-distributed? what about near 0 ?
- we are saying that there is some probability density for vanishing background...
- probably this is because we had four events passing cuts



Prior for background

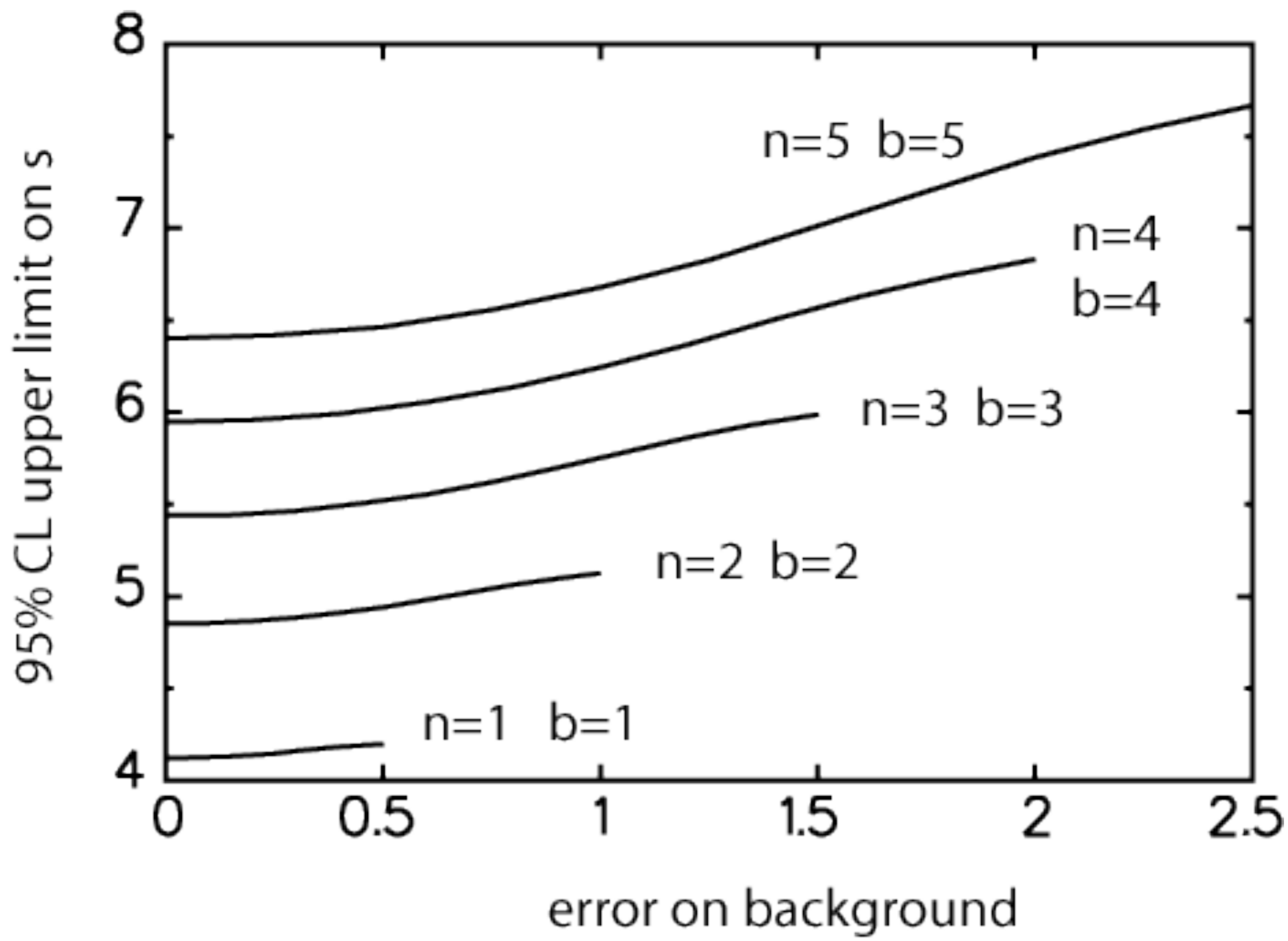
- truncated gaussian seems to be a very popular choice
- note that the normalization of this is uncertain
- if we really mean 68% CL for the interval 1-3 then we are not properly normalized
- other choices are justifiable, and used:
  - gamma distribution (sum of exponential dists.)
  - log normal (product of normal dists.)
  - Poisson prior (a la Lecture 8 on Bayes)?
- choice tends not to matter much for background

Prior for signal

- the "default" choice tends to be to use a flat (uniform) prior in the signal rate
    - this is an improper prior - formally it is non-normalizable
    - must cut off to perform integral!
    - if you use a flat prior you must investigate the effect of any cutoff on your limit
- other choices (beyond the scope here; project?):
    - 1/s (a.k.a. Jeffreys); esentially flat in log(s)
    - $1/\sqrt{s}$

Limits versus background uncertainty

- clearly the dependence of the limit on the signal is fairly weakly dependent on the error on the background:



Uncertainty on efficiency

$$\mathcal{P}(s;n,\epsilon,\sigma_\epsilon) = \frac{\int \mathcal{L}(n;s,\epsilon',\sigma_\epsilon)\mathcal{P}(\epsilon')\mathcal{P}(s)d\epsilon'}{\int \mathcal{L}(n;s',\epsilon',\sigma_\epsilon)\mathcal{P}(\epsilon')\mathcal{P}(s)d\epsilon' ds'}$$

- Bayesian treatment is the same...with the same considerations regarding priors for ε
- "default" choice might be flat prior for s, and a gaussian for ε
- if we do that there is a problem: denominator integral diverges logarithmically
- consider n=0, b=0, ε=1

$$\int\int e^{-\epsilon' s'} e^{-(\epsilon'-1)^2/2\sigma_\epsilon^2} d\epsilon' ds' \propto \int \frac{1}{s'} ds'$$

solutions to the log divergence problem

- the problem is the flat prior coupled with a prior for the efficiency which has non-zero density at zero efficiency
- “do we really believe our efficiency might be zero?”
- must change one prior or the other
- change flat to flat with cutoff?
- change from gaussian prior to log normal or other?
- this is a real problem!
- crops up in different forms
- less of a problem, or not a problem at all, for multi-bin likelihoods

the status so far

- simplest problem: Poisson process with background, single channel counting
  - no solution apparent from frequentist paradigm
  - most straightforward solution in Bayesian case has logarithmic divergence
  - should we event attempt a more complicated problem?
- as it turns out, more complex problems are more amenable...

marginalization

- the procedure whereby we apply Bayes Theorem and remove nuisance parameters by integrating them away is called marginalization
- this is an extremely common approach to incorporation of systematic errors in likelihoods
- it is used even in the "frequentist" or "frequentist inspired" methods we will study later
- in effect, when you marginalize away a nuisance parameter you are averaging the likelihood with respect to that parameter
- can perform marginalization by Monte Carlo!

likelihoods with nuisance parameters

- let's focus on the numerator of our Bayesian expression, and write a general form:

$$\mathcal{L}(\bar{x};\bar{\alpha},\bar{\beta}) = \mathcal{P}(x;\bar{\alpha},\bar{\beta}')\mathcal{P}(\beta_1')\mathcal{P}(\beta_2')...$$

- as an example let's consider the multi-bin spectrum of Poisson-distributed values

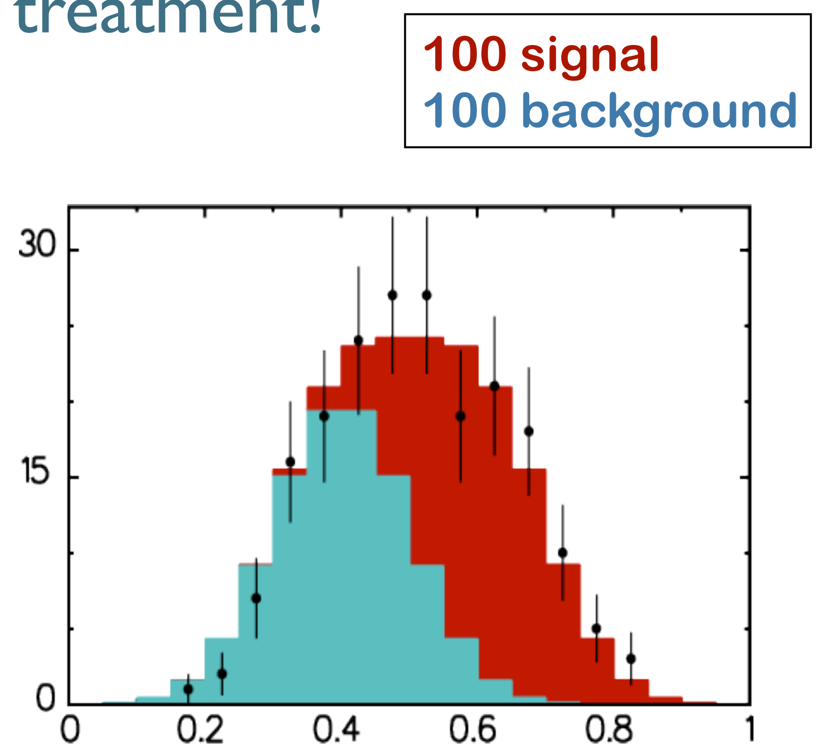
- let's say that we represent the priors for the nuisance parameters by Gaussians:

$$\mathcal{L}(\bar{x};\bar{\alpha},\bar{\beta}) = \prod_{i=1}^{n} \frac{\mu_i^{n_i} e^{-\mu_i}}{n!} \times G(\beta_1', \beta_1, \sigma_{\beta_1}) \times G(\beta_2', \beta_2, \sigma_{\beta_2}) \times ...$$

- here the μ_i are functions of the α's and β's

profile likelihood

- it is of course tempting to maximize the likelihood with respect to the nuisance parameters rather than marginalize (average over the nuisance parameters)
- this is called <u>profiling</u>
- profiling is not a proper Bayesian treatment!
- profiling gives very similar results to marginalization for typical Poisson spectrum problems
- example: double gaussian fit



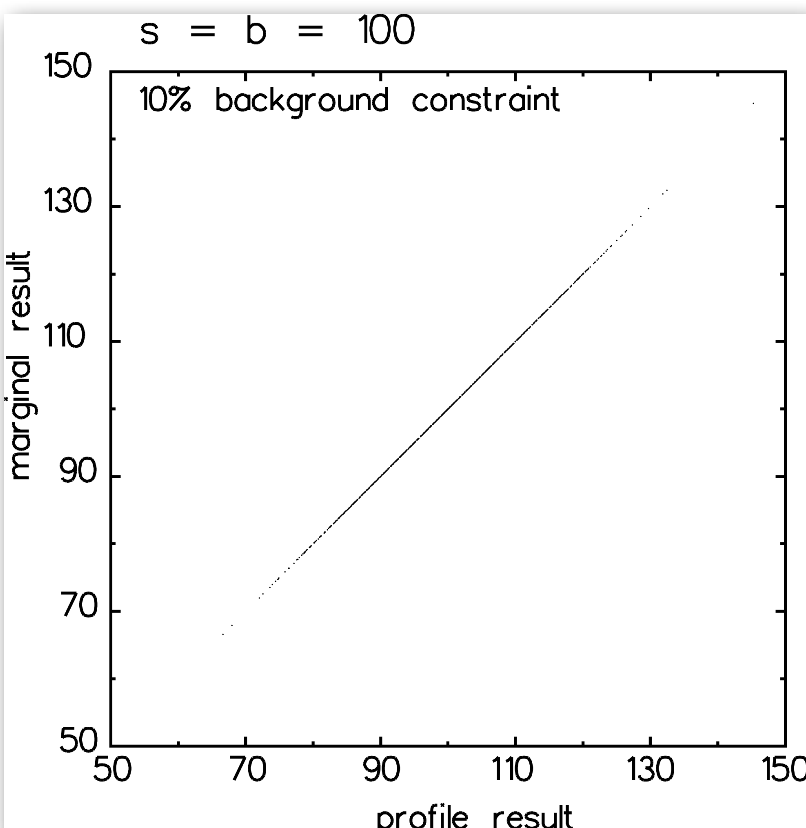
profile likelihood: minimization

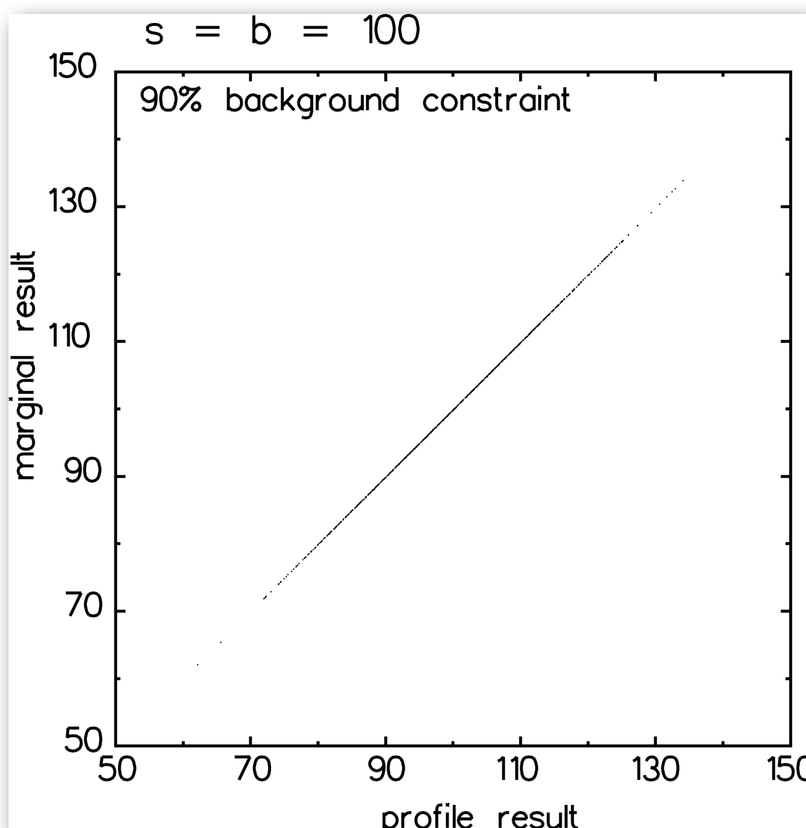
$$-\ln\mathcal{L}(\bar{x};\bar{\alpha},\bar{\beta}) = \sum_{i=1}^{n}(\mu_i - n_i \ln \mu_i) + \frac{(\beta_1' - \beta_1)^2}{2\sigma_1^2} + \frac{(\beta_2' - \beta_2)^2}{2\sigma_2^2} + \ldots$$

- minimize -lnL; the gaussian terms become additive "penalty" terms
- gaussian terms pull the nuisance parameters to their central values
- data pull the nuisance parameters away from their central values via the μ_i
- how do the results compare with marginalization?

compare profiling with marginalization

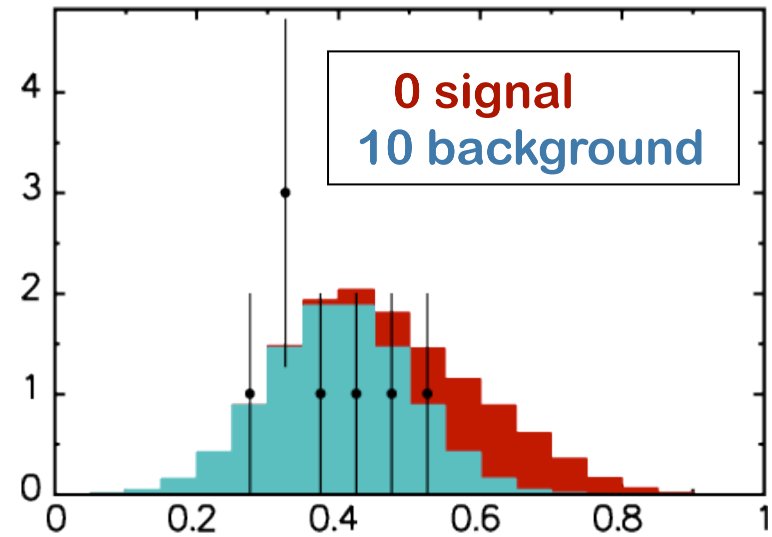
- plot marginalization result compared with profiling result for two different constraint levels for background
- two methods give basically identical results

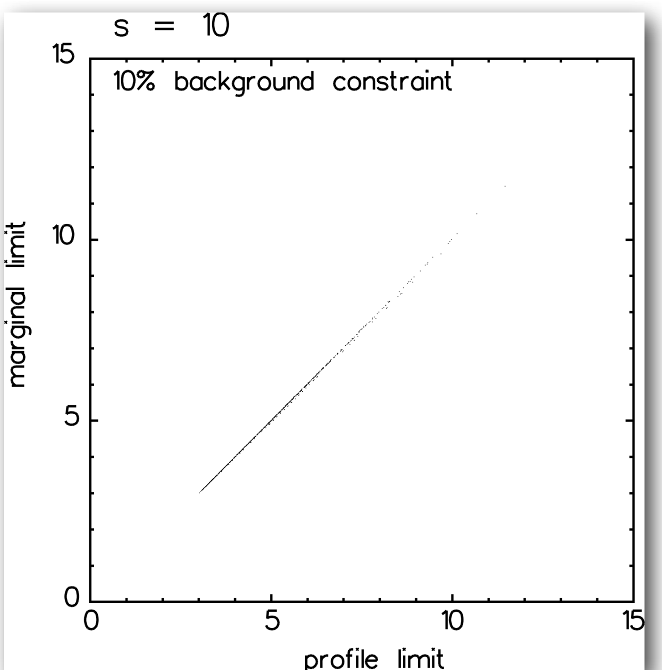


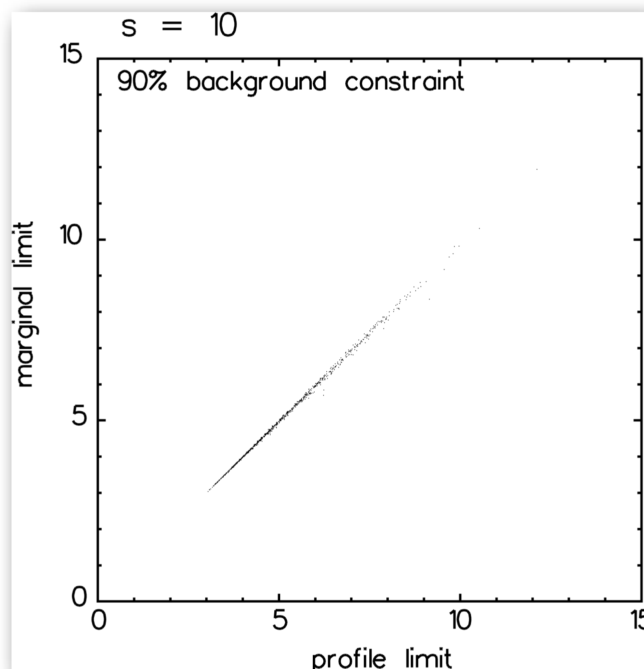


limits comparing profiling/marginalization

- in this case we generate background only and calculate 95% CL upper limit on signal
- again results are practically identical







marginalization/profiling

- are the methods in fact equivalent?
- no: but empirically they give very similar results
- the big advantage of profiling over marginalization is computation speed
- I use it lately in my own analyses
- normalize your nuisance priors!
- for truncated Gaussian, for example, need to correct for lost portion
- don’t lose the 2 in the denominators!

nuisance parameters in likelihoods

$$-\ln \mathcal{L}(\bar{x};\bar{\alpha},\bar{\beta}) = \sum_{i=1}^{n}(\mu_i - n_i \ln \mu_i) + \frac{(\beta_1' - \beta_1)^2}{2\sigma_1^2} + \frac{(\beta_2' - \beta_2)^2}{2\sigma_2^2} + \ldots$$

- let's look in a bit more detail at how to write likelihoods with nuisance parameters
- the bin contents are $\mu_i = L\sigma_1\epsilon_{1i} + L\sigma_2\epsilon_{2i} + \ldots$
- here, the value of L, and some σ's are nuisance parameters
- the ε are in effect efficiency histograms, a.k.a. templates
- can have arbitrary additional multiplicative factors:

$$\mu_i = L\sigma_1\beta_1\beta_2\epsilon_{1i} + L\sigma_2\beta_2\beta_3\epsilon_{2i} + \ldots$$

correlations!

template “morphing”

- due to systematic uncertainties, the values of the efficiency templates may change in a coherent (that is correlated) way
- “shape” and “normalization” errors can be taken into account simultaneously
- create “+1σ” and “-1σ” (or other) templates representing effect of change (energy scale, efficiency, Q^2, ...)
- can introduce “morphing” parameter f_j with nominal value 0, and gaussian constraint of 1:

$$\epsilon_{ji} = \epsilon_{ji}^0 + f_j \frac{\epsilon_{ji}^+ - \epsilon_{ji}^-}{2}$$

template "morphing"

$$\epsilon_{ji} = \epsilon_{ji}^0 + f_j \frac{\epsilon_{ji}^+ - \epsilon_{ji}^-}{2}$$

- this symmetrizes the effect of the systematic uncertainty...do we want that?
- what if the bin is really going through a maximum at $f_j = 0$?
- solution: quad/linear morphing
- ensures that templates are adhered to
- avoids polynomials going crazy for $|f| > 1$