Likelihood for Multisource Spectrum

- very common analysis: use binned spectrum in one or more dimensions to gain statistical sensitivity
- typically several backgrounds contribute
- focus here is how to represent all types of systematic uncertainties in likelihood as nuisance parameters
- example problem: t’ search from CDF
- methods here are beginning to be used in CMS, and could easily be incorporated into RooStats
- useful for limits, discovery, and measurements of new physics processes
Example: $t'$ Search in CDF

- use 3D binned likelihood in $H_T$, $M_{\text{reco}}$, $N_{\text{jet}}/\chi^2$

$$H_T \geq 5 \text{ jet}, \chi^2 < 8$$
$$\geq 5 \text{ jet}, \chi^2 < 8$$
$$4 \text{ jet}, \chi^2 > 8$$
$$4 \text{ jet}, \chi^2 < 8$$

$M_{\text{reco}}$
Example: $t'$ Search in CDF

- main backgrounds: $t\bar{t}$, $W+\text{jets}$, QCD
Systematic Uncertainties

- integrated luminosity
- background normalizations
- lepton ID data/MC ratios
- jet energy scale
- initial/final state radiation
- $Q^2$ scale
- Monte Carlo statistics

Types of nuisance parameters:
- multiplicative
- template morphing

“Barlow-Beeston light”
Multiplicative nuisance parameters

- likelihood is Poisson at its core, with bin contents coming from simple sums over sources
  - parameter(s) of interest: $\alpha$
  - nuisance parameters: $\beta$
- nuisance parameters are gaussian-constrained:

$$- \ln L(\bar{x}; \bar{\alpha}, \bar{\beta}) = \sum_{i=1}^{n} (\mu_i - n_i \ln \mu_i) + \frac{(\beta'_1 - \beta_1)^2}{2\sigma_1^2} + \frac{(\beta'_2 - \beta_2)^2}{2\sigma_2^2} + ...$$

$$\mu_i = L\sigma_1 \beta_1 \beta_2 \epsilon_{1i} + L\sigma_2 \beta_2 \beta_3 \epsilon_{2i} + ...$$
Pitfalls of multiplicative parameters

- must avoid normalization of any source going negative...MINUIT bounds or
- must avoid bins with just one source vanishing or likelihood will suddenly jump
- especially true of signal source, near zero signal
Template “morphing”

- energy scales in particular can cause coherent shifts in shape (and normalization!) of template
- typically evaluate template shapes at three values of systematic effect corresponding to \(-1\sigma, 0, +1\sigma\) shifts in the base effect
- can use more evaluations, or other points
- define morphing nuisance parameter which allows one to continuously “morph” shape as parameter changes

\[
\mu_i = \mu_i^0 + f M(\mu_i^-, \mu_i^0, \mu_i^+ )
\]
Template morphing

- sophisticated “horizontal” treatment* of Read works in 1D only
- use “vertical morphing” for t’ analysis in 3D
- interpolate quadratically within range
- extrapolate linearly outside range
- avoids “kink” at zero; no symmetrization needed
- multiple morphing parameters work well

to treat bin statistical uncertainties, Barlow and Beeston suggested* introducing a nuisance parameter $\beta_i$ for each source $i$ in each bin

implemented Newton-type iterative method to solve simultaneous nonlinear equations for $\beta_i$ which maximize likelihood

each bin is independent, so this can be done inside our MINUIT fit for profile likelihood

problem: MINUIT fails due to small jumps in $\beta_i$ as expected number in source changes

*Comp. Phys Comm. 77 (1993) 219
MC Statistics: Barlow-Beeston “lite”

- only one overall B-B parameter $\beta$ is needed per bin since the source uncertainties are independent!
- constrain $\beta$ to 1 within total statistical uncertainty in bin (add in quadrature)
- can solve analytically for $\beta_i$ in bin $i$:

$$
\frac{-\ln L_i}{\partial \beta_i} = 0 \implies \beta_i^2 + (\mu_i \sigma_i^2 - 1) \beta_i - n_i \sigma_i^2 = 0
$$

- this does represent fairly the effect of the overall statistical uncertainty in each bin
- works for non-MC sources
- method works well! MINUIT is stable!
Profiling vs. Marginalization

In binned likelihood fits to multisource spectra, the result of using a profile treatment of nuisance parameters gives nearly identical results to marginalization.
Summary

- three basic types of nuisance parameters are sufficient for almost all binned Poisson likelihood problems:
  - normalization
  - morphing (correlated shape and normalization)
  - statistical uncertainties
- in the context of a binned multisource spectrum likelihood, have shown straightforward, stable methods to implement these
- works for profiled or marginalized likelihood
- could/should be standardized within RooStats