

Statistical uncertainties in inclusive jet analyses
at detector level – why they are not \sqrt{N}
(with N = total number of jet counts)

Olaf Behnke (DESY)

LHC-EW WG: Jets and EW bosons, Jun 13, 2018

The problem:

- Number of selected pp collision events n (per integrated lumi) follow Poisson distribution with expectation value λ :
 $P(n; \lambda) = e^{-\lambda} \lambda^n / n!$
 $\langle n \rangle = \lambda; \quad V(n) = \langle n^2 \rangle - \langle n \rangle^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$
- Estimator of λ using max. Likelihood plus gaussian approximation for uncertainty:
 $\hat{\lambda} = n; \quad \hat{V}(\hat{\lambda}) = n$
- Example measurement:
 99 ± 10 events
- What if we count always two jets per event for inclusive jet cross section (imagine physics \equiv Leading order process)
 198 ± 14 is **incorrect!**
 198 ± 20 is correct

Inclusive Jet Cross Sections

- Notation: n_1, n_2, \dots, n_m denote the observed number of events with one, two and m jets.
- Each number follows its own statistically independent Poisson distribution with expectation values $\lambda_1, \lambda_2, \dots, \lambda_m$

- Total number of jets n_j :

$$n_j = n_1 + 2n_2 + \dots + m n_m$$

$$\langle n_j \rangle = \lambda_1 + 2\lambda_2 + \dots + m \lambda_m$$

$$V(n_j) = \langle n_j^2 \rangle - \langle n_j \rangle^2 = \lambda_1 + 4\lambda_2 + \dots + m^2 \lambda_m$$

- Estimator of inclusive jet number \mathcal{N} :

$$\hat{\mathcal{N}} = n_j$$

$$\hat{V}(\hat{\mathcal{N}}) = n_1 + 4n_2 + \dots + m^2 n_m$$

Differential Jet Cross Sections, two p_T bins and ≤ 2 jets/event

- n_1^1 := observed #events with one jet in first p_T bin
- n_2^1 := observed #events with one jet in second p_T bin
- n_{11}^2 := observed #events with two jets in first p_T bin
- n_{22}^2 := observed #events with two jets in second p_T bin
- n_{12}^2 := observed #events with one jet in each p_T bin
- Total number of jets in the two bins n_{j1} and n_{j2} :
 $\langle n_{j1} \rangle = \lambda_1^1 + \lambda_{12}^2 + 2\lambda_{11}^2$; $\langle n_{j2} \rangle = \lambda_2^1 + \lambda_{12}^2 + 2\lambda_{22}^2$
 $V(n_{j1}) = \lambda_1^1 + \lambda_{12}^2 + 4\lambda_{11}^2$; $V(n_{j2}) = \lambda_2^1 + \lambda_{12}^2 + 4\lambda_{22}^2$
 $V(n_{j1}, n_{j2}) = \lambda_{12}^2$
- Estimators of inclusive jet numbers:
 $\widehat{\mathcal{N}}_1 = n_1^1 + n_{12}^2 + 2n_{11}^2$; $\widehat{\mathcal{N}}_2 = n_2^1 + n_{12}^2 + 2n_{22}^2$
 $\widehat{V}(\widehat{\mathcal{N}}_1) = n_1^1 + n_{12}^2 + 4n_{11}^2$; $\widehat{V}(\widehat{\mathcal{N}}_2) = n_2^1 + n_{12}^2 + 4n_{22}^2$
 $\widehat{V}(\widehat{\mathcal{N}}_1, \widehat{\mathcal{N}}_2) = n_{12}^2$
- Formulae can be easily generalised to any # of bins and jet multiplicities

- The “multi-count/event problem” occurs **NOT** if one uses one observable per event, e.g. the p_T of the leading jet.
- The detector level covariance matrix $\mathbf{V}_{\mathcal{N}}$ with statistical correlations (Page 4) can be used directly in the least square unfolding with TUnfold:

$$\chi^2 = [\mathcal{N} - \mathbf{K}\mathbf{x}]^t \mathbf{V}_{\mathcal{N}}^{-1} [\mathcal{N} - \mathbf{K}\mathbf{x}]$$

with (example inclusive jet cross sections vs p_T):

\mathcal{N} = vector of jet counts at detector level in different p_T bins

$\mathbf{V}_{\mathcal{N}}$ = Estimated variance of the jet counts

\mathbf{x} = vector of jet xsecs at true level in different p_T bins

\mathbf{K} = Detector response matrix.