STATISTICAL ERRORS FROM BIN-BY-BIN UNFOLDING

bsection Terminology: From the MC you count the number of truth-reconstructed pairs and populate a *transfer matrix*, **M** which is in terms of number of counts. In what follows, the first matrix index is for the reconstructed bins and the second index is for truth bins. You also have the *truth spectrum,* **T**, that you used to generate the transfer matrix. The truth spectrum includes all truth jets, not just the ones that were reconstructed, and thus contains information that is not in **M**. The *reconstructed spectrum*, **R**, is simply the projection of the transfer matrix:

$$R_i = \sum_j M_{ij} = M_{ii} + \sum_{j \neq i} M_{ij} \tag{1}$$

The correction factors are then:

$$C_i = \frac{T_i}{R_i} \tag{2}$$

The statistical fluctuations in the correction factors come from the statistical uncertainty in the R_i values, which are the number of jets in a given reconstructed bin **given** a certain number of truth jets (i.e. T_i). The uncertainty is

$$\delta C_i^2 = \frac{T_i^2}{R_i^4} \delta R_i^2 \tag{3}$$

R_i has two different types of uncertainty. Using the form RHS of Eq. 1,

$$\delta C_i^2 = \frac{T_i^2}{R_i^4} \left(\delta M_{ii}^2 + \sum_{j \neq i} \delta M_{ij}^2 \right) \tag{4}$$

the first term has binomial fluctuations ($N = T_i$ and $p = R_i/T_i$)

$$\delta M_{ii}^2 = M_{ii} \left(1 - \frac{M_{ii}}{T_i} \right) \tag{5}$$

while the second term is taken to be independent of T_i, if the errors are Poisson ($\delta M_{ij}^2 = M_{ij}$), the result is

$$\delta C_i^2 = \frac{T_i^2}{R_i^4} \left(M_{ii} \left(1 - \frac{M_{ii}}{T_i} \right) + \sum_{j \neq i} \delta M_{ij} \right) = \frac{T_i^2}{R_i^3} \left(1 - \frac{M_{ii}^2}{T_i R_i} \right) \tag{6}$$

If you neglected to consider the binomial nature of the diagonal elements of the transfer matrix, you would overestimate the uncertainty by

$$\Delta[\delta C_i^2] = \frac{T_i M_{ii}^2}{R_i^4}\,, \qquad \frac{\Delta[\delta C_i^2]}{\delta C_i^2} = \frac{1}{\frac{R_i T_i}{M_{ii}^2} - 1}\,, \tag{7}$$

which can become large in bins where there is not significant migration and where the efficiency is high ($M_{ii} \sim R_i \sim T_i$).