New Correction Factor Errors

Yakov Kulinich

November 1, 2017

In Figure 1, we have truth and reconstructed $\Delta\phi$ distributions on the left-most plot, the response matrix on the center plot, and the correction factors with errors on the right-most. There is also a transfer matrix M_{ij} where $\Delta\phi_{Reco}$ is along the x-axis, along the j-index, and $\Delta\phi_{Truth}$ is along the y-axis, along the i-index.

Define T_i as the total number of entries in the i^{th} bin of the Truth distribution (blue points on left plot), and R_i as the total number of entries in the i^{th} bin of the Reconstructed distribution (red points on left plot).

In terms of the transfer matrix R_j is

$$R_j = \sum_i M_{ij} = M_{ii} + \sum_{i \neq j} M_{ij} \tag{1}$$

The last part is just the diagonal element plus the off-diagonal horizontal elements of the i^{th} bin (on the x-axis).

Similarly, in terms of the transfer matrix, T_i is

$$T_i = \sum_j M_{ij} = M_{ii} + \sum_{j \neq i} M_{ij} \tag{2}$$

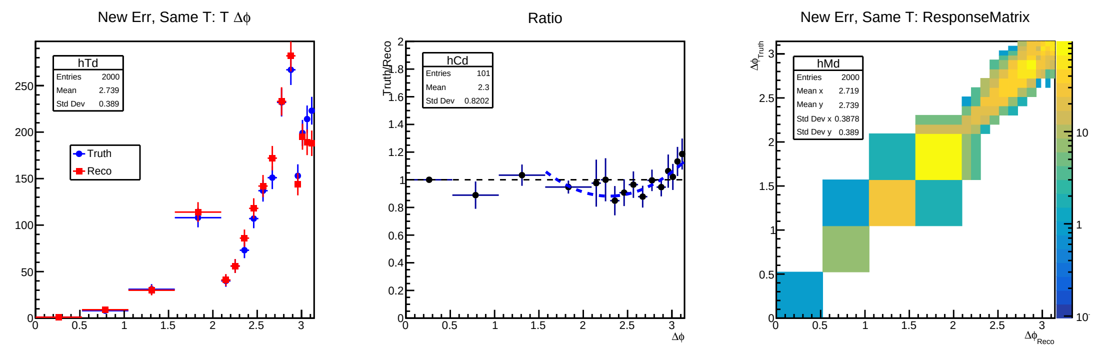

Figure 1: $\Delta\phi$ distributions for truth and reco (left). Response Matrix M_{ij} (center). Correction factors with errors (right).

For some bin i^{th} bin,

$$R_i = T_i - N_{Leaving} + N_{Arriving} = T_i - \sum_{k \neq i} M_{ik} + \sum_{j \neq i} M_{ji} \tag{3}$$

We can express the number leaving and number arriving in terms of off-diagonal row or column elements of M_{ij}, or in terms of T_i, R_i, and diagonal elements of M_{ij}.

$$N_{Leaving} = T_i - M_{ii} \tag{4}$$
$$N_{Arriving} = R_i - M_{ii} \tag{5}$$

Now, T_i is taken as a constant. This means that reconstructed distribution can be different time to time, but the truth distribution stays the same. In the language of a toy MC, this is equivalent to generating one Truth distribution, and smearing it many different times, each time (or for each new "experiment") getting new results.

When T_i is taken as constant, the bin migration of leaving and arriving is different. The distribution of $N_{Leaving}$ is binomial, while $N_{Arriving}$ is Poisson. If T_i is fixed, there is only a certain number of entries that can leave, while the number that arrives depends on, and is a mix of the entries leaving neighboring bins.

In a toy MC [1], for the case where the truth distribution was generated one time, but smearing applied to the reconstructed (case with "fixed" T_i), it is clear from Figure 2 that the migration where entries are leaving is narrower than where the arrive. In the same toy MC, when for every experiment a new truth distribution was used, it is evident that the migration to and from is the same.

Correction factors C_i, which relate T_i and R_i are

$$C_i = \frac{T_i}{R_i} \tag{6}$$

and their respective errors σ_{C_i} are

$$\sigma^2_{C_i} = \frac{C_i^2}{R_i^2}\sigma^2_{R_i} \tag{7}$$

Now since T_i is constant, and the entries leaving a T_i bin follow binomial statistics, while entries arriving are Poisson, we continue from Equation 3. The error in R_i is

$$\sigma^2_{R_i} = \sigma^2_{N_{Leave}} - \sigma^2_{N_{Arrive}} \tag{8}$$
$$\sigma^2_{R_i} = T_i\frac{T_i - M_{ii}}{T_i}(1 - \frac{T_i - M_{ii}}{T_i}) + (R_i - M_{ii}) \tag{9}$$
$$\sigma^2_{R_i} = T_i + R_i - 2M_{ii} - \frac{(T_i - M_{ii})^2}{T_i} \tag{10}$$

[1] A Toy MC with a randomly generated exponential was generated for truth 2000 times, with smearing from ATLAS MC transfer matrix applied to the reconstructed. The experiment was then repeated 10,000 times to get some good statistics on correction factors, their errors, bin migration, etc.

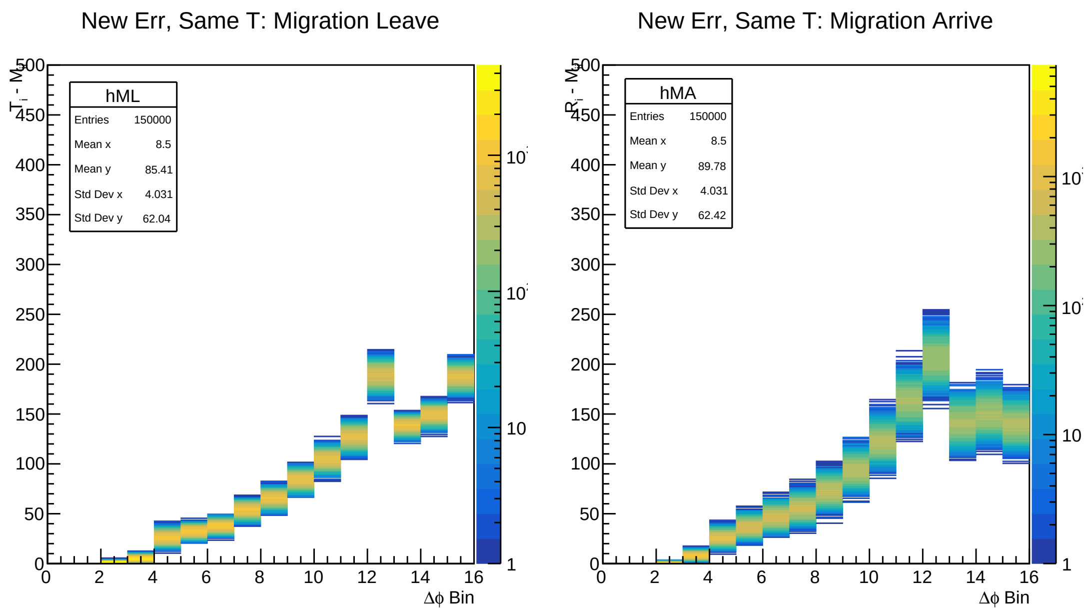

Figure 2: For the case where for every experiment a the same generated truth distribution but differently smeared reconstructed distribution, histogram of migration between $\Delta\phi$ bins (x and y axes) for entries arriving (right) and entries leaving (left).Migration where entries leave has a binomial (narrower) distribution, while entries arriving is Poisson.

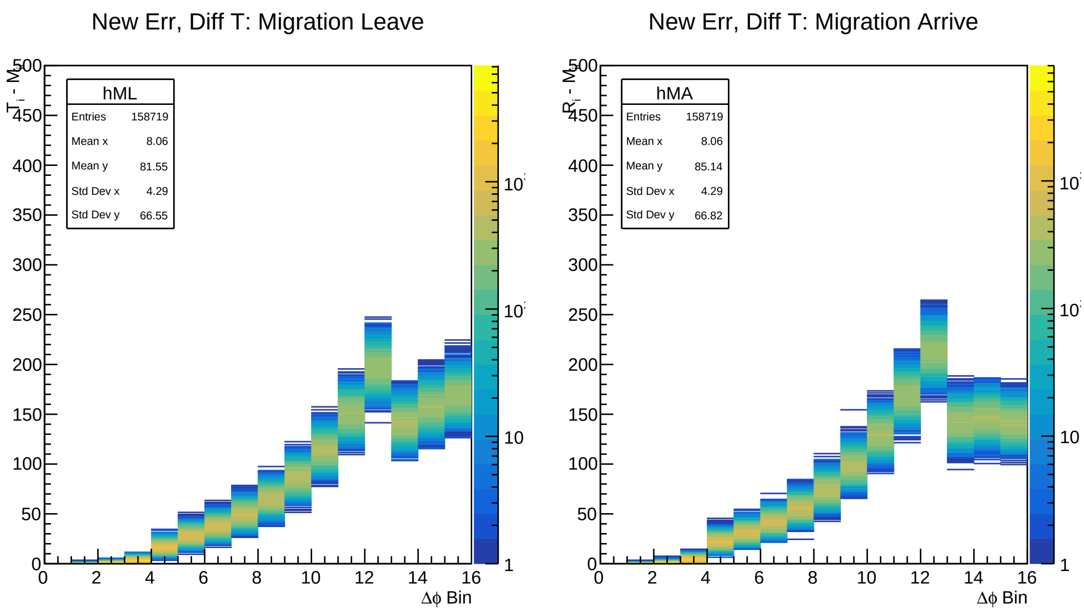

Figure 3: For the case where for every experiment a new truth distribution is generated and the reconstructed is smeared from that, histogram of migration between $\Delta\phi$ bins (x and y axes) for entries arriving (right) and entries leaving (left). Both migrations have Poisson distributions.

From this, plugging into Equation 7, the error on the correction factor is

$$\sigma_{C_i}^2 = \frac{T_i^2}{R_i^2}\Big(T_i + R_i - 2M_{ii} - \frac{(T_i - M_{ii})^2}{T_i}\Big) \tag{11}$$

$$\sigma_{C_i}^2 = \frac{T_i^2}{R_i^3}\Big(1 - \frac{M_{ii}^2}{T_i R_i}\Big) \tag{12}$$

Which is the same result as in bbb_derivation.pdf. Again, fixed T_i was assumed in this derivation.