Unfolding of charged particle multiplicities at LHCb

David Voong

University of Bristol

October 13, 2011

Introduction

- Track multiplicity: Number of tracks reconstructed in an event
- A bench mark measurement for any new energy regime
- Event selection
  - Minimal cuts
  - Multiplicity dominated by soft QCD processes
  - Shown to be difficult to describe with Monte Carlo models
- Excellent measurement for constraining models with real data
- Important for accurately describing background for many analyses

Measurment carried by other experiments

- CMS http://arxiv.org/abs/1011.5531
- ATLAS http://arxiv.org/abs/1012.5104
- LHCb http://arxiv.org/abs/1112.4592(tracks reconstructed from Vertex detector, this analysis unfolds from Long tracks)

LHCb track classification

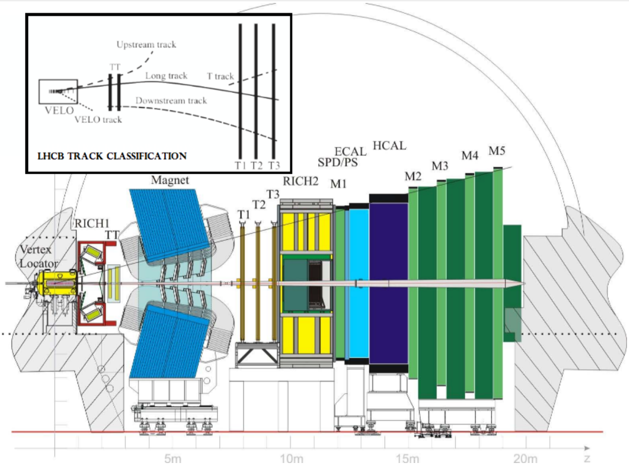


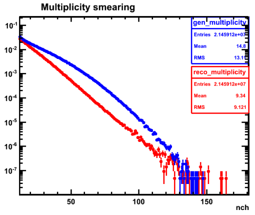
Figure 1: LHCb track classification scheme

Long track benefits

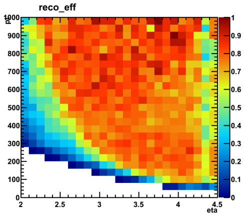
- Momentum information
- PID information

Detector response

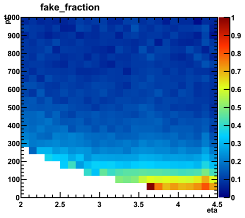
- Lost particles
  - Detector inefficiency
  - reconstruction inefficiency
- Fake particles and background
  - Detector noise
  - interaction of particles with detector material producing particles
  - mis-reconstructed tracks
- Overall effect is a smearing of the true distribution
- Unfolding corrects for these effects to give back the true distribution



(a) MC data, $2.0 <= eta < 4.5$



(b) Reco efficiency



(c) Fake fraction

Response matrix

Smearing of the true distribution is described by the matrix equation

$a = G \cdot b$

$a =$ reconstructed multiplicity distribution (column matrix)
$b =$ true multiplicity distribution (column matrix)
$G =$ response matrix (n by m matrix)

$a_i =$ probability for event to reconstructed i tracks
$b_j =$ probability for event to produce j particles
$G_{ij} =$ probabilty to reconstruct i tracks given the event produced j particles

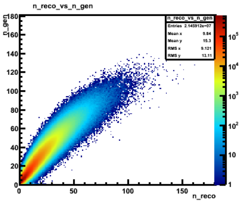
Toy response matrix

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.7 & 0.15 & 0.1 \\ 0.2 & 0.7 & 0.2 \\ 0.1 & 0.15 & 0.7 \end{pmatrix} \cdot \begin{pmatrix} 0.2 \\ 0.6 \\ 0.2 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.5 \\ 0.25 \end{pmatrix}$$

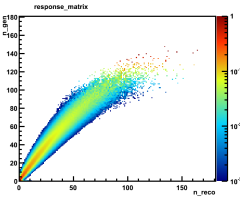
- $a_0 = G_{00} \cdot b_0 + G_{01} \cdot b_1 + G_{02} \cdot b_2$
- $G_{00} + G_{10} + G_{20} = 1 \Rightarrow$ probability conservation

Calculating the response matrix

- Response matrix calculated from Monte Carlo
  - Dependent only on the detector response simulation
  - Generator model independent (else introduce bias into measurement when unfolding)
- Generate fill 2D histogram of the number of reconstructed tracks vs the number of generated particles
- Apply normalization condtion. Sum of column elements = 1 (Row elements in the case of a histogram)



(d) n reco vs n gen



(e) response matrix

Unfolding

$a = G \cdot b$

- In real data a and G are known, need to solve for b

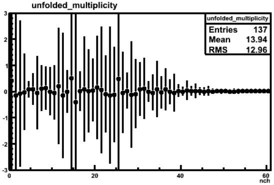
$G^{-1} \cdot a = b$

- Matrix inversion uses method of Singular value decomposition (SVD)

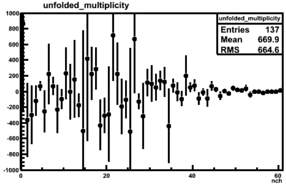
$a = u \cdot W \cdot v^T \cdot b$
$u \cdot u^T = v \cdot v^T = \mathbb{1}$
W is a diagonal matrix with non-zero elements on the diagonal

$b = v \cdot W^{-1} \cdot u^T \cdot a$

Direct matrix inversion



(f) MC cross check



(g) real data

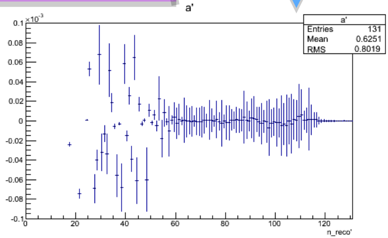
Figure 2: Unfolded multiplicities using a direct matrix inversion

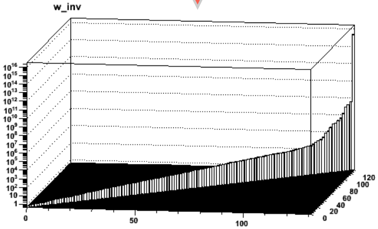
Unphysical behaviour

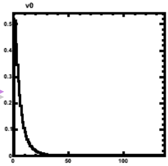
- large errors
- oscillations
- normalization violated

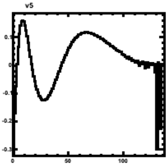
Unfolding instabilities

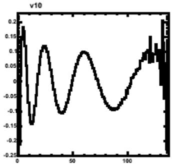
- $b = v \cdot W^{-1} \cdot u^T \cdot a$
- $b = v \cdot W^{-1} a'$
- $b_j = \sum_i v_{ji} \; a'^W_j$

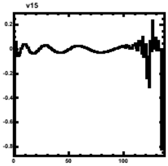












Interpretation

- unfolding interpreted as the expansion of orthogonal functions weighted by unfolding weights of a rotated distribution
- large unfolding weights amplify noise $\longrightarrow$ large error bars in unfolded multiplicity
- higher order functions behave more oscillatory $\longrightarrow$ oscillations in unfolded multiplicity

Regularization

- Apply regularization to remove contributions from higher order terms
- $b = v \cdot R \cdot W^{-1} \cdot u^T \cdot a$
- R is a diagonal matrix, with $R_{ii} = 1, \mathit{if}\ i < R_{cutoff}$
- e.g, $R_{cutoff} = 2$

$$R = \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

Judging the best cut off by eye: MC

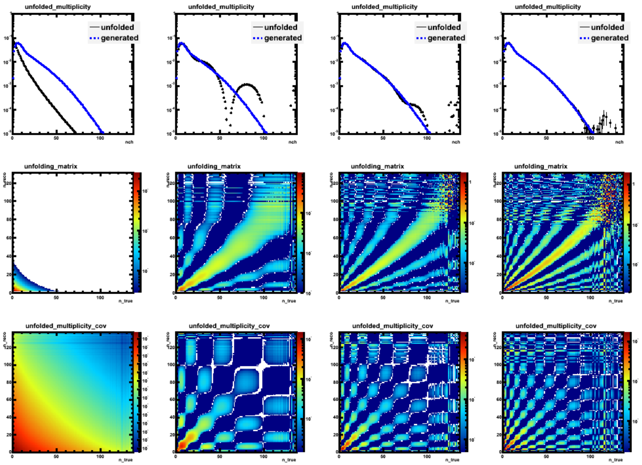


Figure 3: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 1, 10, 20, 30 (from left to right)

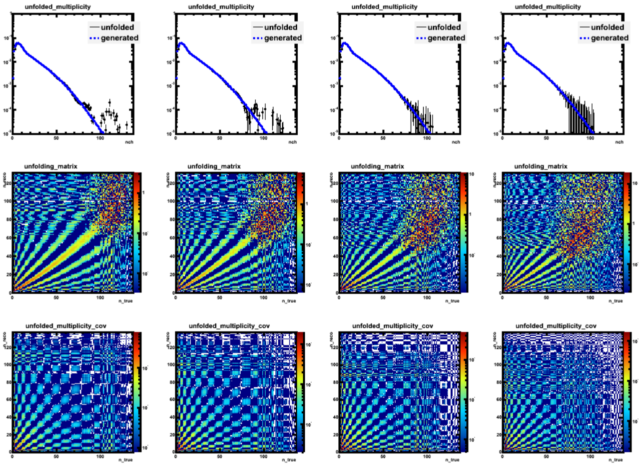


Figure 4: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 40, 50, 60, 70 (from left to right)

Judging the best cut off by eye: MC

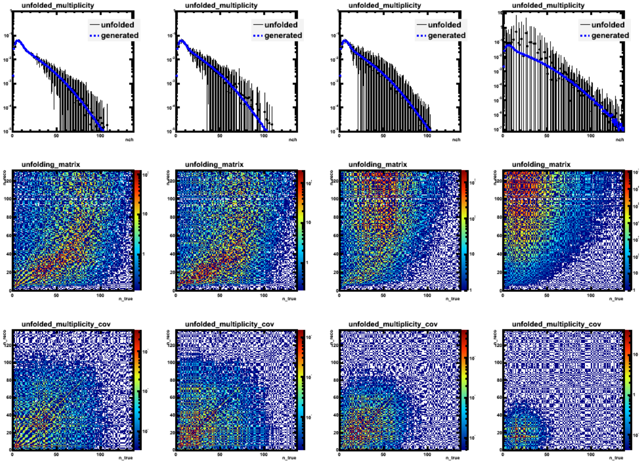


Figure 5: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 100, 110, 120, 130 (from left to right)

Judging the best cut off by eye: Real data

- Less events in real data
- Unfolding method tuned to MC data
- Expect noise amplication to turn on at lower orders in unfolding expansion

Judging the best cut off by eye: Real data

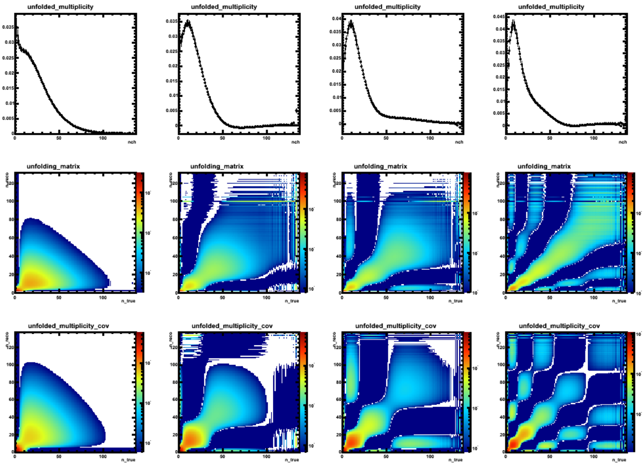


Figure 6: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 2, 4, 6, 8 (from left to right)

Judging the best cut off by eye: Real data

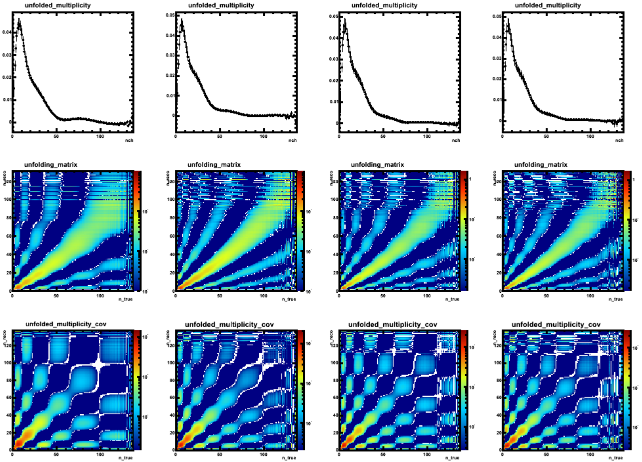


Figure 7: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 10, 12, 14, 16 (from left to right)

Judging the best cut off by eye: Real data

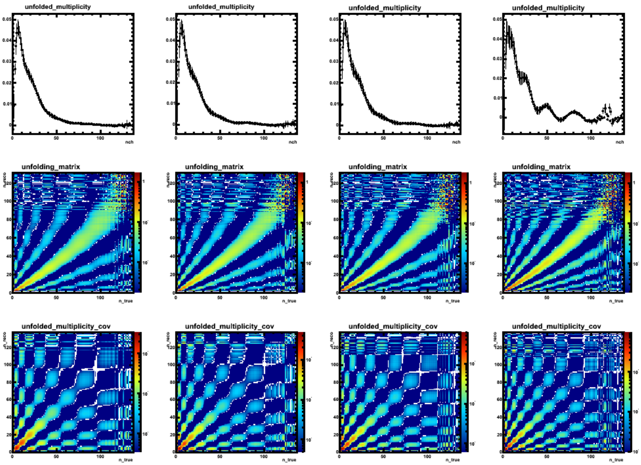


Figure 8: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 18, 20, 22, 24 (from left to right)

Summary and conclusion

- An generator independent unfolding method has been developed
- Technique is sensitive to statistical errors which result in unphysical behaviour
- A regularization scheme is used to correct for this

Final words

- A working unfolding method is in place
- Future improvements,
  - more sophisticated regularization scheme (Method of Reduced Cross-Entropy)
  - Alternative unfolding method (cross check), Parameterization method, no matrix inversion
  - Parameterization of the response matrix, smoothing particularly significant in low statistics region
- Results to be used to apply constraints to Monte Carlo generator models
- Put us in the best possible position to keep making world's best measurements and increase sensitivity to new Physics

BACKUP

Parameterization method

- No matrix inversion
- Numerical trial and error method
- Define a parameterization with which to fit the true distribution
- True distribution is not known $\therefore$ parameterization must be generic enough to represent any shape

Parameterizations

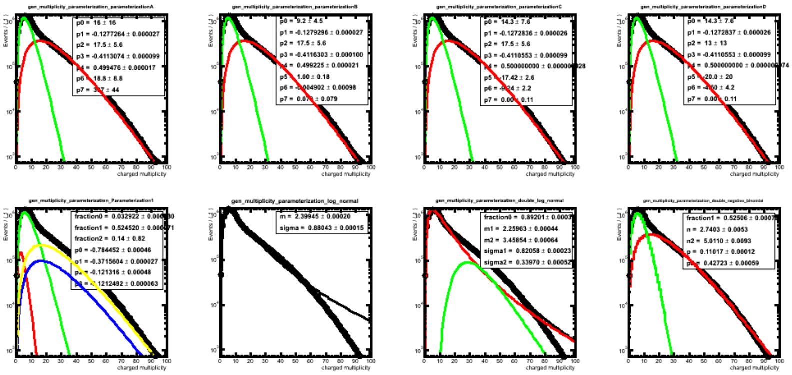
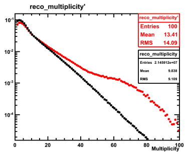


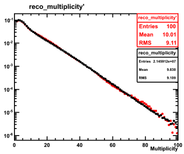
Figure 9: From top left to bottom right, Parameterization a, b, c, d, Voong, Log normal, Double log normal, Double negative binomial

Parameterization method

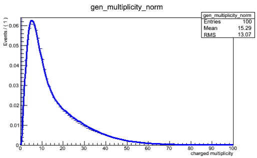
1. Make an estimate of the parameters which describe the true distribution
2. Smear it with the response matrix
3. Perform a chi2 fit between the smeared distribution (*reco_multiplicity′*) and the observed reconstructed distribution
4. Propogate the parameters of the fit to the true distribution



(a) 2



(b) 3



(c) 4

Judging the best cut off by eye: Real data

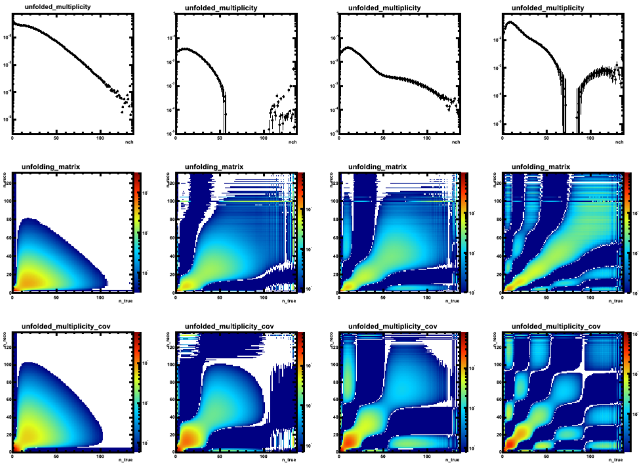


Figure 10: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 2, 4, 6, 8 (from left to right)

Judging the best cut off by eye: Real data

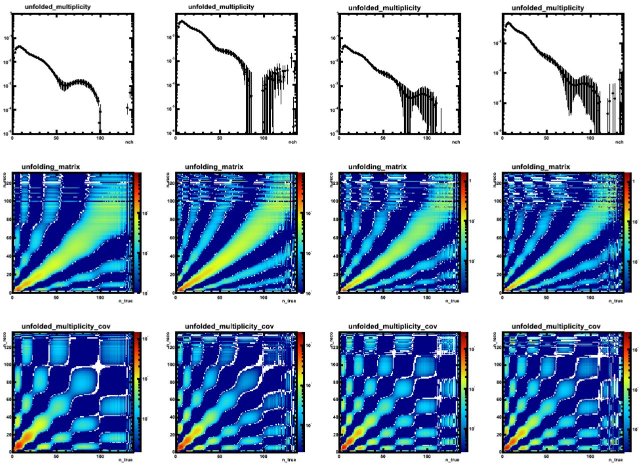


Figure 11: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 10, 12, 14, 16 (from left to right)

Judging the best cut off by eye: Real data

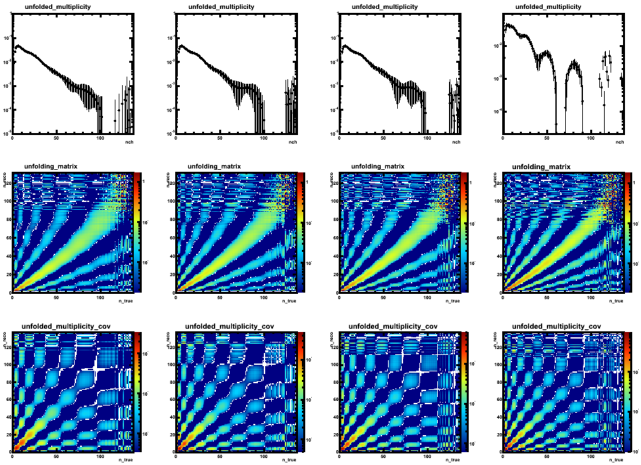


Figure 12: Unfolded multiplicity (top), unfolding matrix (middle), covariance matrix of unfolded multiplicity (bottom) for regularization cut off 18, 20, 22, 24 (from left to right)