Unfolding in ALICE

Jan Fiete Grosse-Oetringhaus, CERN for the ALICE collaboration

PHYSTAT 2011
CERN, January 2011
Content

• Unfolding methods used in ALICE

• Applications of unfolding in ALICE
  – Observations, Issues and remarks
Unfolding Methods

- $\chi^2$-Minimization with regularization (only 1D)
  \[
  \chi^2(T^*) = \sum_m \left( \frac{M_m^* - \sum_t R_{mt} T_t^*}{e_m} \right)^2 + \beta R(T^*)
  \]
  \[
  R(T) = \sum_t (a_t)^2
  \]
  \[
  \beta \text{ weight}
  \]

- Available regularizations
  - Mostly used: linear and constant
    - $a_t = \frac{T_t'}{\sqrt{T_t}} = \frac{T_t - T_{t-1}}{\sqrt{T_t}}$ (prefer constant)
    - $a_t = \frac{T_t''}{\sqrt{T_t}} = \frac{T_{t-1} + 2T_t - T_{t+1}}{\sqrt{T_t}}$ (prefer linear least curvature)
  - $a_t = \frac{\hat{T}_t''}{\sqrt{\hat{T}_t}} = \ln T_t$ (prefer exp)
  - $R(T) = \sum_t T_t \ln \frac{T_t}{\varepsilon_t}$ (reduced cross-entropy)

- Implementation uses Minuit, Migrad
  - Uncertainties from Minuit
Unfolding Methods (2)

• Iterative Bayesian Unfolding (1D + nD)

\[
\tilde{R}_{tm} = \frac{R_{mt}P_t}{\sum_{t'} R_{mt}P_{t'}} \quad U_t = \sum_m \tilde{R}_{tm} M_m
\]

d‘Agostini

• Optionally smoothing can be applied

\[
\hat{U}_t = (1-\alpha)U_t + \frac{\alpha}{3} (U_{t-1} + U_t + U_{t+1})
\]

• Uncertainty on unfolded distribution by randomization of input spectrum
  – Poisson distribution used per bin
Unfolding Methods (3)

- Unfolding methods part of the software framework
  - Main interface function (for 1D case):
    Unfold(TH2* correlation, TH1* efficiency, TH1* measured, TH1* initialConditions, TH1* result)
  - nD case uses THnSparse

- Functions to evaluate bias \( b_t \) due to regularization

\[
b_t = \sum_m \frac{\partial T_t}{\partial M_m} ((RT)_m - M_m)
\]

- Derivate calculated numerically

\[
\frac{\partial T_t}{\partial M_m} = \frac{1}{6d} \left[ 8 \left( f \left( \frac{d}{2} \right) - f \left( -\frac{d}{2} \right) \right) - (f(d) - f(-d)) \right]
\]

\[
f(x) = T_t(M | M_m = M_m + x\sqrt{M_m})
\]

(see e.g. Cowan)
Example: Evaluate Regularization Weight with MC

- Cooking down the consistency of the unfolded solution with the MC to one number is tricky
  - Different regions, different qualitative shape
Applications

- Multiplicity distribution
  - Efficiency ~ 70% (inactive modules in pixel detector due to cooling issues) → far off the diagonal
  - Significant spread in response matrix → wide correlations in unfolded spectrum

- $p_T$ spectrum
  - Shift is 2-4% for 20-30 GeV
  - Unfolding needed, response matrix not very wide
Applications (2)

- Jet spectrum
  - Wide response matrix $\sigma \sim 20\%$
  - Corrections to full (charged+neutral) jet $\rightarrow$ significant shift because all neutrals missing (1/3)
  - Caveat: MC simulations usually done in $p_T,\text{hard}$ bins
    - Jet yield at low $p_T \gg$ MC at low $p_T$
Example Issue

- Observation: nice hump in unfolded distribution (which got people immediately excited)
- Usual cross-checks $\rightarrow$ independent of
  - Regularization scheme
  - Unfolding method
- Forcing the distribution to be exponential
  - Moves the structure to the residuals
  - However, in fewer bins and (visually) much less significant
- Uncertainties of unfolded spectrum and wide response matrix can enhance fluctuations in the data significantly
- Slope change visible in measured data $\rightarrow$ transition gets spread out by regularization
Systematic Uncertainties

- From the unfolding procedure
  - Bias
  - Some fluctuations (1-5%) remain
  - Use MC to get a feeling
- Uncertainties are correlated between the bins
- Due to the response matrix
  - Uncertainties on the response matrix "propagate through" the unfolding into the unfolded distribution
  - Create different response matrices resembling the uncertainty
    - E.g. detector efficiency uncertainty of 1% $\rightarrow$ Create three response matrices with -1%, 0, +1% efficiency w.r.t. nominal
    - Use difference in unfolded distribution as systematic uncertainty $\rightarrow$ tricky because overlaid by "typical" fluctuations, small effects cannot be disentangled
Summary

- ALICE uses unfolding (up to now) for multiplicity distribution, $p_T$ spectrum and jet spectrum
- The ALICE software framework provides
  - $\chi^2$-minimization with regularization
  - Iterative Bayesian unfolding
- Unfolding usually requires a lot of tuning
- Systematic uncertainties sometimes tricky
- Uncertainties on unfolded distributions are always understood (by others) as single bin uncertainties
  - Can we invent a nice way to visualize that this is not the case?

Thanks for help with preparing this talk to Christian Klein-Boesing, Jan Rak

Unfolding tutorial (introductory slides, exercises and solutions):
www.cern.ch/jgrosseo/permanent/unfolding
(prepared for a Helmholtz power week)
Feel free to use it!