Outline

- Introduction on unfolding
- Example unfolding problem
- Unfolding methods
- Comparison
Introduction

- Unfolding: estimate truth distribution from measurement, distorted by
  - detector effects
  - statistical fluctuations
- truth distribution: cross sections or similar quantities
- Unfolding is also referred to as “correction for detector effects”

- Integral equation of 1st kind

\[ \int k(x, y)f(y)dy + \delta(x) = g(x) \]

given observations \( g(x) \)
the kernel \( k(x, y) \)
and fluctuations \( \delta(x) \)
estimate the truth \( f(y) \)

- \( k(x,y) \): detector effects, background, etc
- \( g(x) \) has uncertainties
- \( k(x,y) \) has syst. uncertainties
  \( \rightarrow \) not covered in this talk
Unfolding of binned measurements

- This talk: unfolding of binned (discrete) distributions, where bin-to-bin migrations are described by a matrix equation

\[ \mu_i = \sum \ A_{ij} x_j + b_i \]

- Statistical fluctuations: the observations \( y_i \) are drawn from a Poisson distribution

\[ P(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \]

- Large sample limit: Gaussian distributions

- Correlated bins: multivariate Gaussians

\[ A_{ij} = \frac{N_{ij}^{MC\text{reco},MC\text{truth}}}{N_j^{MC\text{truth}}} \]  
is calculated from MC
Unfolding of binned measurements

- This talk: unfolding of binned (discrete) distributions, where bin-to-bin migrations are described by a matrix equation

\[
(p_i; \mu_i) = e^{-\mu_i} \mu_i y_i
\]

- Statistical fluctuations: the observations \(y_i\) are drawn from a Poisson distribution

\[
\text{(truth+background)} \times \text{detector} \times \text{stat.fluctuations} \rightarrow \text{measurement}
\]

Result: estimator of truth \(\leftarrow\) unfolding algorithm \(\leftarrow\) measurement
Example unfolding problem

- Toy example to illustrate basic properties of unfolding algorithms
- Decay of a heavy particle into two light particles
- Light particles smeared by spatial and energy resolution
- Trigger threshold causes reconstruction inefficiency
- Background important at high $P_T$
- Variable bin size, overflow bin
- Goal: reconstruct $P_T$ distribution

- Two samples of toy events
  - “data” $P_T$ distribution following Landau(6,1.8)
  - “MC” $P_T$ distribution following Landau(5,2)
- Background mainly at high $P_T$
Example unfolding problem

- Toy example to illustrate basic properties of unfolding algorithms
- Decay of a heavy particle into two light particle
- Light particles smeared by spatial and energy resolution
- Trigger threshold causes reconstruction inefficiency
- Background important at high $P_T$
- Variable bin size, overflow bin
- Goal: reconstruct $P_T$ distribution

- Significant migrations at low $P_T$
- Change of bin size leads to change in bin purity
- Efficiency $>95\%$, not important for this study
How to test unfolding results?

- **Tests with real data**
  - Look at (global) correlation coefficients
  - Trivial test: fold back unfolding result and compare to data
    - Unfolding result: $x_j^{\text{unf}}$
    - Fold back and compare to data:
      $$y_i^{\text{data}} \approx \sum_j A_{ij} x_j^{\text{unf}} + b_i$$

- **Test with Monte Carlo**
  - Trivial test: response matrix and MC using the same truth
  - Non-trivial test: use different truth for response matrix and unfold alternative MC (here: "data"): $x_j^{\text{unf}}$
  - Compare to alternative MC truth:
    $$x_j^{\text{truth}} \approx x_j^{\text{unf}}$$

This talk:
- Look at average global correlation coefficients
- Compare folded result with data
- Compare result to “data” truth
- Extract “data” truth parameters using a fit

Quantitative comparison: $\chi^2$

… plus many other things not discussed here, e.g. eigenvalue analysis
Unfolding methods investigated in this talk

- Bin-by-bin correction factors
- Matrix inversion
- Template fit
- Tikhonov regularisation: [Tikhonov 1963]
  implementation: e.g. RUN [Blobel 1984], TUnfold [S.S. 2012]
- IDS method: [Malaescu 2011]
Bin-by-bin correction factors

- Very simple method:

\[ x_i = \left( y_i - b_i \right) \frac{N_i^{\text{gen}}}{N_i^{\text{rec}}} \]

- \( y_i \) : observed in bin \( i \)
- \( b_i \) : expected background in bin \( i \)
- \( N_i^{\text{gen}} \) : MC truth in bin \( i \)
- \( N_i^{\text{rec}} = \sum_j A_{ij} N_j^{\text{gen}} \) : MC reconstructed in bin \( i \)

Results “looks nice”
No statistical bin-to-bin correlations
but
Method is wrong, fails very basic tests
Unfolding methods investigated in this talk

- Bin-by-bin correction factors
- Matrix inversion
- Template fit
- Tikhonov regularisation: [Tikhonov 1963]
  implementation: e.g. RUN [Blobel 1984], TUnfold [S.S. 2012]
- IDS method: [Malaescu 2011]
Matrix inversion

- If the number of bins is equal on gen and rec level: A is a square matrix
  \[ \text{→ invert it} \]
  folding equation: \( y = Ax + b \)
  invert matrix: \( x = A^{-1}(y - b) \)

Covariance: \( V_{xx} = A^{-1}V_{yy}(A^{-1})^T \)

Correlation coefficients:
\[ \rho_{ij} = \frac{(V_{xx})_{ij}}{\sqrt{(V_{xx})_{ii}(V_{xx})_{jj}}} \]

\( y \): measurements
\( V_{yy} \): covariance matrix of measurements
\( b \): background
\( A \): matrix of migrations

Large bin-to-bin correlations

Unfolded result exhibits bin-to-bin oscillations

Good \( \chi^2 \): no bias

Folded-back result is on the data

\[ P_T \text{(gen)} \]
Template fit

- Choose larger number of reconstructed bins than truth bins → least-square fit
- Idea: use more information → obtain better result?

\[
\chi^2 = (y - b - Ax)^T V_{yy}^{-1} (y - b - Ax)
\]

- $y$: measurements
- $V_{yy}$: covariance matrix of measurements
- $b$: background
- $A$: matrix of migrations
- $A_{ij}$: MC template for truth bin $j$

\[
x = (A^T V_{yy}^{-1} A)^{-1} A^T V_{yy}^{-1} (y - b)
\]

covariance of $x$: $V_{xx} = (A^T V_{yy}^{-1} A)^{-1}$
Template fit

- Choose larger number of reconstructed bins than truth bins → least-square fit
- Idea: use more information → obtain better result
  → Result does not improve much over matrix inversion in this example

New problem: normalisation is not preserved \([N_{\text{data}}=4584, N_{\text{fold}}=4572]\)

Well-known problem with least-square fits to Poisson-distributed data if \(\sqrt{N}\) uncertainties are used

Can be improved by adding a constraint to the fit
Template fit with area constaint

- Template with constraint on the total number of events
- Basic idea: preserve normalisation for the folded-back result by adding the constraint
  \[ \sum (y_i - b_i) = \sum_{i,j} A_{ij} x_j \]
- Technical implementation: see TUnfold documentation
  \[ N_{\text{data}} = N_{\text{fold}} = 4584 \]

\[ \sum (y_i - b_i) = \sum_{i,j} A_{ij} x_j \]

→ Result does not change much over unconstrained template fit, but normalisation is recovered
Tikhonov regularisation

- Basic idea: add terms to the likelihood which damp oscillations in the result.
  \[
  \chi^2 = (y - b - Ax)^T V_{yy}^{-1} (y - b - Ax) + \tau^2 (L(x - x_B))^T L(x - x_B)
  \]
  \[y: \text{measurements}\]
  \[V_{yy}: \text{covariance matrix of measurements}\]
  \[b: \text{background}\]
  \[A: \text{matrix of migrations}\]
  \[x_B: \text{regularisation bias}\]
  \[L: \text{regularisation conditions}\]
  \[\tau: \text{regularisation strength}\]

- Regularisation bias \(x_B\): set to zero or to MC truth

- Regularisation conditions \(L\): set to unity matrix [or mimic second derivatives, “curvature”]

- Regularisation strength \(\tau\): “small” number
  \[\tau \ll 1/\sigma\]
  where \(\sigma\sim\text{uncertainty after unfolding}\)

\(\tau \ll 1/\sigma\)

\(\chi^2\) = measurement residual

\(V_{yy}\) = measurement covariance matrix

\(b\) = background

\(A\) = migration matrix

\(x_B\) = regularisation bias

\(L\) = regularisation condition matrix

\(\tau\) = regularisation strength

In addition, apply area constraint to preserve normalisation.
Tikhonov regularisation (e.g. TUnfold)

- Basic idea: add terms to the likelihood which damp oscillations in the result.
- This is working well: no oscillations, moderate correlations and uncertainties
- Basic tests look reasonable
- Question: objective to choose $\tau$
Choice of the regularisation parameter $\tau$

- Eigenvalue analysis (SVD)
  \rightarrow not discussed in this talk

- Scan of parameter $\tau$
  - L-curve scan
  - Scan of global correlation coefficients

- Other data driven methods (e.g. compare stat and syst errors, define convergence criteria) \rightarrow not discussed in this talk
L-curve scan

- Algorithm is often used in medical image processing

  for each $\tau$ repeat the unfolding:

  $$\chi^2 = (y - b - Ax)^T V_{yy}^{-1} (y - b - Ax)$$

  $$+ \tau^2 (L(x - x_B))^T L(x - x_B)$$

  $$\equiv L_x + \tau^2 L_y$$

  study parametric plot of: $\log L_x$ vs $\log L_y$

- Parametric plot is “L-shaped”
  $\rightarrow$ kink (largest curvature) defines $\tau$

  For a review, see: [P. C. Hansen 2000]
Scan of global correlation coefficients

- Global correlation coefficient (bin i)
  \[ \rho_i = \sqrt{1 - \frac{1}{(V_{xx})_{ii}(V_{xx}^{-1})_{ii}}} \]
  
  \(V_{xx}\) : result's covariance matrix

- Take average of all \(\rho_i\) and study dependence on \(\tau\) → choose point with smallest avg(\(\rho_i\))
  (idea by V. Blobel/DESY)

- Comparison to L-curve scan: stronger regulatisation, more bias, smaller uncertainties & correlations
Unfolding methods investigated in this talk

- Bin-by-bin correction factors
- Matrix inversion
- Template fit
- Tikhonov regularisation: [Tikhonov 1963]
  implementation: e.g. RUN [Blobel 1984], TUnfold [S.S. 2012]
- IDS method: [Malaescu 2011]
Iterative method

\[ x_j^{(N+1)} = x_j^{(N)} \frac{\sum_i A_{ij}}{\epsilon_j} \frac{y_i}{\sum_k A_{ik} x_k^{(N)}} \]

- Re-invented by D'Agostini 1995 as "Iterative Bayesian unfolding"

Note: efficiency is absorbed in a redefinition of \( A, x \) in the original works: \( x' = \epsilon x \) and \( A' = A/\epsilon \)

- Mathematical properties (Shepp/Vardi 1982 and Mülthei/Schorr 1987)
  - Ultimately converges to a maximum of the (Poisson) Likelihood
    \[ \rightarrow \text{like matrix inversion but with all } x \geq 0 \]
  - Convergence is very slow

- Use in HEP:
  - Stop after \( N \) iterations \( \rightarrow \) result will be "smooth" [regularized] but is biased to the start value

Regularisation strength:
Tikhonov: \( \tau \) ↔ Iterative: \( N_{\text{iter}} \)
Iterative method with background

- Background could be subtracted from the data
- Or: background could be added to the folded MC in the denominator. This guarantees the desired property $x \geq 0$
- D'Agostini suggests to include the background normalisation as extra bin $x_{n+1}$. This also guarantees $x \geq 0$ but results in an extra parameter $\rightarrow$ make sure to then include a background control bin in the set of measurement bins

$$x_j^{(N+1)} = x_j^{(N)} \sum_i A_{ij} \frac{y_i - b_i}{\sum_k A_{ik} x_k^{(N)}}$$

efficiency: $\epsilon_j = \sum_i A_{ij}$

start values: $x_j^{(-1)}$ [e.g. MC truth]

OR

$$x_j^{(N+1)} = x_j^{(N)} \sum_i A_{ij} \frac{y_i}{\sum_k A_{ik} x_k^{(N)} + b_i}$$

efficiency: $\epsilon_j = \sum_i A_{ij}$

start values: $x_j^{(-1)}$ [e.g. MC truth]
Evaluation of the covariance matrix

- Matrix inversion methods (with or without Tikhonov regularisation): covariance matrix is calculated analytically

- Iterative methods: non-linear, covariance matrix calculation in general has to be done by other means

- Replica method [used in this talk]
  - Apply statistical fluctuations on the data histogram
    \[ \rightarrow \text{N replicas of the data} \]
  - Repeat the unfolding for each replica
  - Covariance is estimated from RMS of the results

- Bootstrap method:
  similar idea, but based on events
  \[ \rightarrow \text{test complete analysis chain} \]
Iterative method: $0^{\text{th}}$ iteration

\[ x^{(N+1)}_j = x^{(N)}_j \sum_i \frac{A_{ij}}{\epsilon_j} \sum_k A_{ik} x^{(N)}_k + b_i \]

efficiency: \( \epsilon_j = \sum_i A_{ij} \)

start values \( x^{(-1)}_j \) set to MC truth

- $0^{\text{th}}$ iteration: “Bayesian unfolding” from 1995 D'Agostini paper
- Result “looks nice”, very small uncertainties, but fails all tests

→ the method has to be iterated

All correlation coefficients are positive → this is "smearing", not "unfolding"
Iterative method: $1^{\text{st}}$ iteration

\[ x_j^{(N+1)} = x_j^{(N)} \sum_i A_{ij} \frac{y_i}{\epsilon_j} \sum_k A_{ik} x_k^{(N)} + b_i \]

- Convergence rate is expected to grow quadratically with the number of bins [Mülthei/Schorr 1987]

- Look at $1^{\text{st}}$ iteration
  - Neighboring bins have positive correlation (expect: negative)
  - Shape not described
  - Folded-back different from data
  \[ \rightarrow \] have to iterate further
Iterative method: 10\textsuperscript{th} iteration

\[ x_j^{(N+1)} = x_j^{(N)} \sum_i \frac{A_{ij}}{\epsilon_j} \sum_k A_{ik} x_k^{(N)} + b_i \]

- Convergence rate is expected to grow quadratically with the number of bins [Mülthei/Schorr 1987]
- Look at 10\textsuperscript{th} iteration
  - Similar to Tikhonov with strong regularisation
Iterative method: 100\textsuperscript{th} iteration

\[ x_j^{(N+1)} = x_j^{(N)} \sum_i \frac{A_{ij}}{\epsilon_j} \sum_k A_{ik} x_k^{(N)} + b_i \]

- Convergence rate is expected to grow quadratically with the number of bins [Mülthei/Schorr 1987]
- Look at 100\textsuperscript{th} iteration
  - Similar to Tikhonov with weak regularisation
Iterative method: $1000^{\text{th}}$ iteration

\[ x_j^{(N+1)} = x_j^{(N)} \sum_i \frac{A_{ij}}{\epsilon_j} \sum_k A_{ik} x_k^{(N)} + b_i \]

- Convergence rate is expected to grow quadratically with the number of bins [Mülthei/Schorr 1987]
- Look at $1000^{\text{th}}$ iteration
  - Similar to matrix inversion, but all guaranteed to be $x \geq 0$
  - Objective to choose number of iterations? Scan of correlation?
IDS method by B. Malaescu

- **IDS**: Iterative Dynamically Stabilized unfolding
- Based on iterative improvements of the matrix of (truth,reco) MC events
- Mathematics not discussed here in detail
- Method is using
  - Significance of data vs (iterated) MC in each bin
  - adjustment of the normalisation in each step
  - also includes a bin-by-bin correction-like contribution
- Method converges to the same result as the standard iterative method
- Speed of convergence is expected to be improved
- The bin-by-bin contributions may lead to reduced correlation coefficients
Iterative methods: scan of avg(ρᵢ)

- Regularisation strength has to be chosen (τ for Tikhonov ↔ N_{iter} here)
- Try scan of global correlation coefficients
  [reminder: this yielded strong regularisation for Tikhonov method]
- Iterative minimum [N=20] is similar in amplitude to Tikhonov case
- IDS minimum [N=3] is much lower than other methods → scan of correlations is not expected to give optimal results for this method

![Graph showing comparison of different methods]
Comparison of results and truth

- Comparison (1): $\chi^2$ test data against unfolded results
- Comparison (2): fit [known] parameterisation of data truth
Comparison (1) $\chi^2$ wrt “data” truth

- Test $\chi^2$ of unfolded results against “data” truth
- For real analyses, such tests can be done by unfolding alternative truth models

<table>
<thead>
<tr>
<th>Method</th>
<th>$X^2 / N_{D.F.}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tikhonov L-curve</td>
<td>1.75</td>
</tr>
<tr>
<td>Tikhonov min((\rho_i))</td>
<td>6.30</td>
</tr>
<tr>
<td>bin-by-bin</td>
<td>4.24</td>
</tr>
<tr>
<td>iterative, N=20 min((\rho_i))</td>
<td>1.12</td>
</tr>
<tr>
<td>IDS, N=3 min((\rho_i))</td>
<td>9.88</td>
</tr>
<tr>
<td>IDS, N=11</td>
<td>0.97</td>
</tr>
</tbody>
</table>

- For the example studied, iterative+min(\(\rho_i\)) performs best
- IDS does not work with the min(\(\rho_i\)) condition, N>10 seems appropriate
Comparison(2) wrt data truth parameters

- Fit results by the analytic function used to generate the truth:
  \[ \text{Landau}(\mu,\sigma) \]
- Only the width \( \sigma \) is shown here (more difficult to fit)

<table>
<thead>
<tr>
<th>Method</th>
<th>fit of width ( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tikhonov L-curve</td>
<td>1.858 ± 0.057</td>
</tr>
<tr>
<td>Tikhonov min(avg(( \rho_i )))</td>
<td>1.965 ± 0.049</td>
</tr>
<tr>
<td>bin-by-bin</td>
<td>2.064 ± 0.046</td>
</tr>
<tr>
<td>iterative, N=20 min(avg(( \rho_i )))</td>
<td>1.906 ± 0.071</td>
</tr>
<tr>
<td>IDS, N=3 min(avg(( \rho_i )))</td>
<td>2.268 ± 0.034</td>
</tr>
<tr>
<td>IDS, N=11 truth</td>
<td>1.915 ± 0.050</td>
</tr>
</tbody>
</table>

- For this test Tikhonov with L-curve is doing better than the iterative method
Selection of other unfolding methods

- **SVD [Hoecker et al, 1995]**
  - Equivalent to matrix inversion with Tikhonov regularisation, parameter $\tau$ from Eigenvalue analysis

- **Shape-constrained unfolding [Kuusela, Panaretos 2015]**

- **Improved D'Agostini [2010]**

- **Fully Bayesian [Choudalakis 2012]**

Plus many other methods
Please apologize for not listing them
Summary

- Unfolding: get measurements independent of the detector response
- Alternative: publish folding matrix with the result
- Many methods exist, only a few have been compared in this talk
- Big unfolding families investigated in this talk:
  - Matrix inversion + Tikhonov regularisation (parameter $\tau$)
  - Iterative methods + truncation after $N_{\text{iter}}$ steps
- Main question: how to choose the regularisation strength. Objectives studied in this talk: L-curve and scan of global correlation coefficients
- Tikhonov: L-curve scan is favored. Iterative: correlation scan seems to work
- Danger to obtain biased results if regularisation is too strong