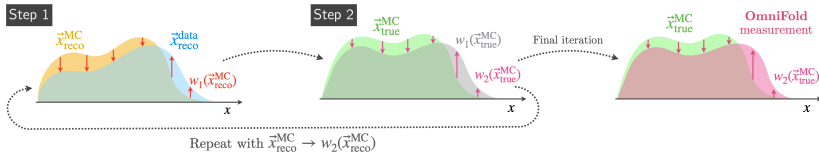


8 Unfolding

❖ correct data for distortions due to the measurement process

e.g. omnifold: a modern unbinned, high-dimensional unfolding method



source: arXiv:1507.09582

■ requirements

- ▶ simulated events that sample the **entire configuration space**
- ▶ realistic detector simulation for the **mapping generator** → **reconstructed level**

■ published information: $\{\text{MC}_{\text{gen}}, w(\text{gen})\}$

- ▶ generator level MC events and per event weights that map to the actual physics
- ▶ which corrected distributions to study can be decided later

► iterative determination of $w(\text{gen})$ by omnifold

■ two types of weight functions for MC data

► $w(\text{gen})$: evaluated using generator (“truth”) level information

► $w(\text{rec})$: evaluated using reconstructed level information

■ iteration, using ML methods to construct weight functions on high-dimensional spaces

► initial condition $w_0(\text{gen}) = 1$

► step 1: $w_n(\text{rec}) \cdot w_n(\text{gen}) \cdot \text{MC}_{\text{rec}} = \text{data}$

► step 2: $w_{n+1}(\text{gen}) \cdot \text{MC}_{\text{true}} = w_n(\text{rec}) w_n(\text{gen}) \cdot \text{MC}_{\text{true}}$

► iteration of steps 1 and 2 until $w_{n+1}(\text{gen}) = w_n(\text{gen})$ i.e. $w_n(\text{rec}) = 1$

classical unfolding methods →

The 1-dim linear unfolding problem

A **true distribution** $f(y)$ is distorted by inefficiencies, systematic shifts and finite resolution. The measurement process is described by a **response function** $R(x, y)$, the distribution of the measured x for every value y . The **observed distribution** $g(x)$ is

$$g(x) = \int_{y_{\min}}^{y_{\max}} dy R(x, y) f(y)$$

- **unfolding: estimate $f(y)$, given $R(x, y)$ and sample of n events drawn from $g(x)$**
 - any difficulty in the 1-dim problem also affects omnifold, but is harder to quantify
 - alternative names for $R(x, y)$:
point-spread function (astronomer), **Green's function** (theorist), **kernel** (mathematician)
 - helpful when dealing with finite statistics: **discretization**

Discretization by histograms with constant bin width

$$a_k = \int_{x_{k-1}}^{x_k} dx g(x) , \quad k = 1, \dots, n_a \quad \text{and} \quad b_i = \int_{y_{i-1}}^{y_i} dy f(y) , \quad i = 1, \dots, n_b$$

- simple intuitive interpretation for coefficients a_k and b_i
- no assumptions about smoothness or curvature of distributions

$$\begin{aligned} a_k &= \int_{x_{k-1}}^{x_k} dx g(x) = \int_{x_{k-1}}^{x_k} dx \int dy R(x, y) f(y) \\ &= \sum_i \int_{x_{k-1}}^{x_k} dx \int_{y_{i-1}}^{y_i} dy R(x, y) f(y) \approx \sum_i \int_{x_{k-1}}^{x_k} dx \int_{y_{i-1}}^{y_i} dy R(x, y) \frac{b_i}{\Delta y} \end{aligned}$$

- unfolding becomes a problem of linear algebra

$$a_k = \sum_i R_{ki} b_i \quad \text{or} \quad a = R \cdot b \quad \text{with} \quad R_{ki} = \frac{1}{\Delta y} \int_{x_{k-1}}^{x_k} dx \int_{y_{i-1}}^{y_i} dy R(x, y)$$

a 1-dim toy model for systematic studies

- PDF of true distributions on the interval $0 \leq y \leq 1$

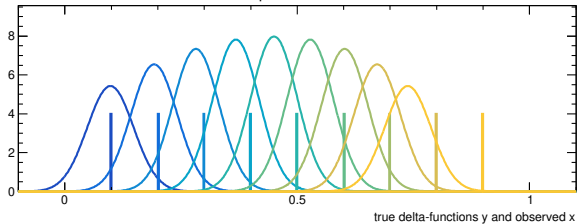
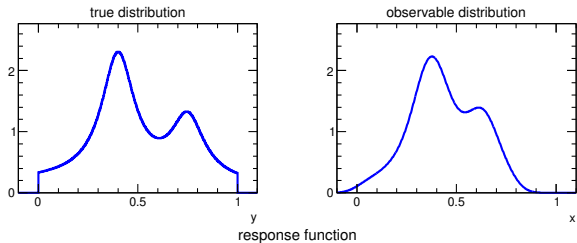
$$b(y) = \frac{20.334}{100 + (10y - 2)^2} + \frac{2.0334}{1 + (10y - 4)^2} + \frac{4.0668}{4 + (20y - 15)^2}$$

- response function

$$R(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (x - y[1 - \beta y])^2\right) \cdot \left(1 - 4\alpha \left(y - \frac{1}{2}\right)^2\right)$$

- ▶ gaussian resolution function (parameter $\sigma = 0.05$)
- ▶ quadratic bias as a function of y (parameter $\beta = 0.1$)
- ▶ parabolic efficiency loss towards phase space limits (parameter $\alpha = 0.5$)

illustration →



Unfolding by matrix inversion

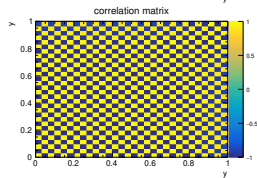
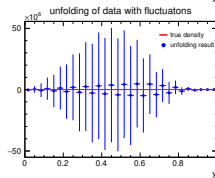
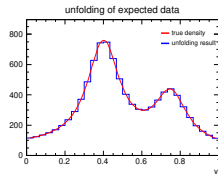
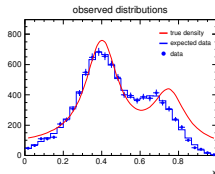
❖ consider the case $n_a = n_b$

$$a = R \cdot b \rightarrow b = R^{-1} \cdot a$$

with covariance matrix

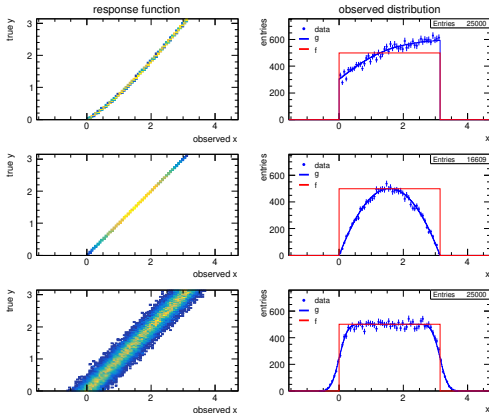
$$C(b) = R^{-1} C(a) R^{-T}$$

- formally correct
- proper covariance matrix
- completely useless



analyze what's going on →

❖ different types of detector effects and a uniform density $f(y) = 1/\pi$ on $[0, \pi]$



biased measurement:

$$R(x, y) = \delta(x - b(y))$$

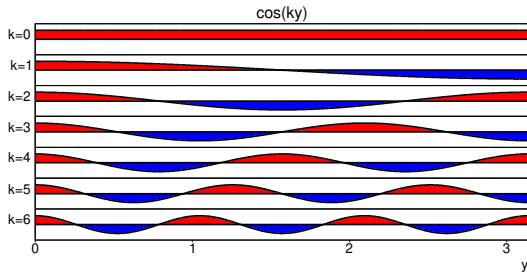
(parabolic) efficiency losses:

$$R(x, y) = \varepsilon(y)\delta(x - y)$$

finite resolution $\sigma = 0.2$:

$$R(x, y) = \frac{e^{-\frac{(x-y)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$$

❖ expansion of the true distribution into orthogonal functions



$$f(y) = \sum_{k=0}^{\infty} a_k \cos(k y)$$

$$\omega_k = \frac{I_g(a_k)}{I_f(a_k)}$$

- ▶ the basis functions $\cos(ky)$ probe $f(y)$ at a resolution $\pi/(k+1)$
- ▶ the information loss due to $R(x, y)$ is measured by the attenuation factors ω_k of the Fisher information

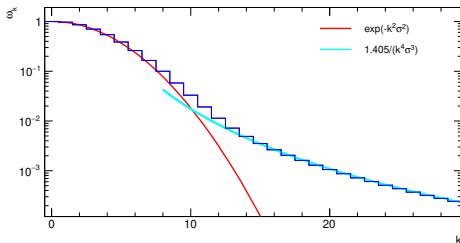
■ no information loss due to bias → full correction by re-calibration

$$\omega_k = 1$$

■ little information loss due to inefficiencies → correction is cheap

$$\omega_0 = 1 \quad \text{and} \quad \omega_k = 1 - \frac{3}{k^2 \pi^2} \quad \text{for} \quad k > 0$$

■ massive loss of information due to finite resolution → full correction impossible



- ▶ exponential attenuation for low-order terms
- ▶ power-law behaviour for high-order terms due to boundaries in y

stable correction methods →

Bin-by-bin correction factors

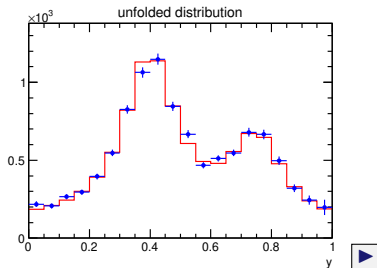
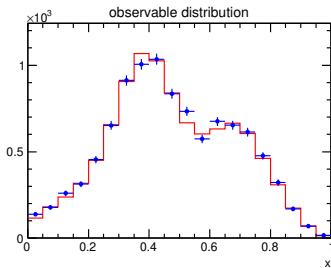
- use the same binning for observed and true distribution
- define bin-dependent correction factors c_k : $b_k = a_k \cdot c_k$
- determination of the correction factors
 - ▶ start with an **assumption** for b_k
 - ▶ **determine** a_k by folding (multiplication) with the response matrix and calculate

$$c_k = \frac{b_k}{a_k} = \frac{b_k}{\sum_{l=1}^{n_b} R_{kl} b_l}$$

- correction factors depend on the assumed distribution b_k , e.g.:
 - ▶ (approximate/expected) true distribution (**unknown**)
 - ▶ uniform distribution (**“objective”**)
 - ▶ measured distribution (**hopefully similar to truth**)
 - ▶ ... and anyhow the correct result is guaranteed only for $b_k = b_k^{\text{true}}$

- ❖ iterate the correction factors to become independent of the assumed b_k
 - Iteration starting from e.g. initial settings $b_k^{(0)} = a_k$:

$$b_k^{(n+1)} = a_k \cdot c_k^{(n+1)} \quad \text{with} \quad c_k^{(n+1)} = \frac{b_k^{(n)}}{\sum_{l=1}^{n_b} R_{kl} b_l^{(n)}}$$



❖ findings

- ❑ correction factors work well if the true distribution is known
 - ▶ no iteration required and no new information from measurement
- ❑ conceptual problems
 - ▶ empty bins are corrected to zero
 - ▶ data from outside physics phase space are ignored
- ❑ iteration removes dependence on unknown distribution, but . . .
 - ▶ naive error propagation $\sigma(b_k) = c_k \cdot \sigma(a_k)$ is evidently wrong
 - ▶ proper error analysis requires bootstrap

- decompose R into efficiencies ε and a smearing matrix S

$$R_{ik} = S_{ik} \cdot \varepsilon_k \quad \text{with} \quad \varepsilon_k = \sum_i R_{ik}$$

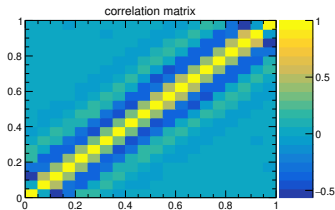
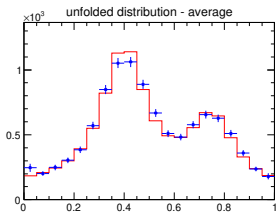
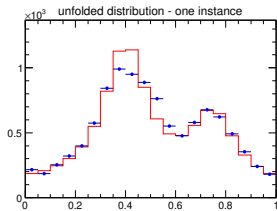
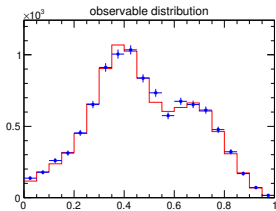
- interpret S as conditional probabilities $S_{ik} = p(\text{obs}_i | \text{true}_k)$, which implies

$$p(\text{obs}_i, \text{true}_k) = p(\text{obs}_i | \text{true}_k) p(\text{true}_k) = p(\text{true}_k | \text{obs}_i) p(\text{obs}_i)$$

- use Bayes' theorem to construct an inverse smearing matrix

$$H_{ki} = p(\text{true}_k | \text{obs}_i) = \frac{p(\text{obs}_i | \text{true}_k) p(\text{true}_k)}{p(\text{obs}_i)} = \frac{S_{ik} P_k}{(S \cdot P)_i}$$

- ▶ assume e.g. uniform prior probabilities and apply H iteratively m times
- ▶ finally apply efficiency corrections and do error analysis via bootstrap



► bayesian unfolding is ...

- mathematically sound approach, explicitly use positivity of probabilities
- can move measurements from unphysical regions into allowed phase space
- no matrix inversion required, i.e. works also for non-square response matrices
- iteration removes the dependence on the initial estimates for the prior
 - too few iterations: result correlated with initial values
 - too many iterations: result becomes unstable
- the unfolding method is defined by the algorithm and the number of iteration steps

Visualizing the properties of unfolding methods

❖ unfold a 2-dim problem presented by an image T

$$M = R \ T \ R^T$$

- R distorts the image by applying the response independently to all columns
- R^T applies the response independently to all rows of the distorted image
- unfolding inverts the process
 - ▶ first unfold all rows then unfold all columns
 - ▶ always start with a uniform prior
 - ▶ always use the same number of iterations

- compare results for infinite statistics and for 10^7 events
- study bin-by-bin correction factors and the Lucy-Richardson method

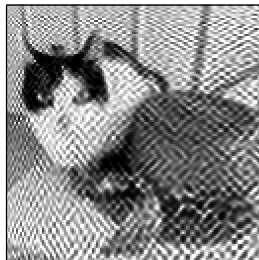
original



data



unfolding result



Lucy-Richardson, IBU with prior=1
100 iterations



Findings

- efficiency corrections always work
- bias correction requires a matrix-based method, as e.g. Lucy-Richardson
- full smearing corrections require infinite statistics
 - ▶ finite statistics lead to artefacts
 - ▶ artefacts are tolerable for a partial correction

unfolding corrects for bias and efficiency losses
smearing can only be partially corrected

formal description →

Regularisation matrix and posterior response matrix

- the estimate b for the true distribution is a function F of the measurements a

$$b = F(a) \quad \text{or} \quad b_k = F_k(a_1, \dots, a_{n_a}) \quad \text{for} \quad k = 1, \dots, n_b$$

- error propagation to calculate the covariance matrix $C(b)$

$$C(b) = M C M^T \quad \text{with Jacobian} \quad M_{ki} = \frac{db_k}{da_i}$$

- ▶ the Jacobian may be a by-product of the unfolding algorithm
- ▶ it can always be estimated by numerical differentiation
- ▶ perfect unfolding for $n_a \geq n_b$ is obtained for $M_{\text{LSQ}} = (R^T C^{-1} R)^{-1} R^T C^{-1}$

combinations of R and $M \rightarrow$

❖ the regularisation matrix Q

- convolving b with R gives values $\hat{a}$ that should be consistent with the data a

$$\hat{a} = R b$$

- sensitivity of re-folded values to variations in the measurements

$$\frac{d\hat{a}_i}{da_j} = \sum_k \frac{d\hat{a}_i}{db_k} \frac{db_k}{da_j} = \sum_k R_{ik} M_{kj} \quad \text{or} \quad Q = RM$$

- ▶ Q measures the damping of fluctuations by the unfolding algorithm
 - $d\hat{a}_i/da_i = 1$: no regularisation
 - $d\hat{a}_i/da_i = 0$: measurement is ignored
- ▶ $\text{Tr}Q$ counts the number of the measurements effectively used to get b

$$N_{\text{eff}} = \text{Tr}(RM)$$

❖ the posterior response matrix P

■ sensitivity of the unfolding result to the true distribution

$$\frac{db_k}{db_l^{\text{true}}} = \sum_i \frac{db_k}{da_i} \frac{da_i}{db_l^{\text{true}}} = \sum_i M_{ki} R_{il} \quad \text{or} \quad P = MR$$

- ▶ P is a response matrix between truth b^{true} and the unfolding result b
- ▶ perfect unfolding corresponds to $db_k/db_l^{\text{true}} = \delta_{kl}$
- ▶ perfect unfolding is achieved for $M = M_{\text{LSQ}}$
- ▶ the traces of P and Q are identical

$$N_{\text{eff}} = \text{Tr}(MR) = \text{Tr}(RM) \leq \min(n_a, n_b)$$

implementation of these concepts →

Diagonalizing the inverse problem $a = R b$

- Fisher information matrix about b for gaussian or Poisson fluctuations of a :

$$I = R^T C^{-1} R$$

- eigenvectors of I – arranged as columns of a matrix V :

$$I V = V S \quad \text{with} \quad V V^T = V^T V = 1$$

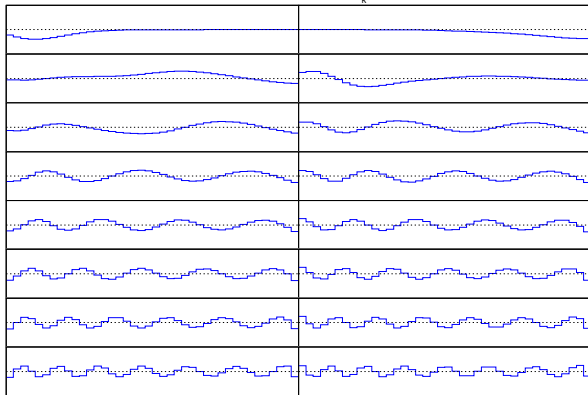
- ▶ S is the diagonal matrix of the eigenvalues of $I = V S V^T$
- ▶ eigenvalues are assumed to be sorted in decreasing order
- expansion of b in the basis given by the eigenvectors of I (“Fisher basis”):

$$b = V \beta \quad \Leftrightarrow \quad \beta = V^T b$$

toy problem →

► orthonormal eigenvectors of the information matrix

illustration of basis functions $v_k, k=0 \dots 15$



❖ least-squares estimate of β

$$\chi^2 = (a - R V \beta)^T C^{-1} (a - R V \beta) = \min$$

■ result

$$\beta = S^{-1} \alpha \quad \text{with} \quad \alpha = V^T R^T C^{-1} a$$

- ▶ α is the result of a linear transformation of a
- ▶ the covariance matrices α and β are diagonal

$$C_\alpha = S \quad \text{and} \quad C_\beta = S^{-1}$$

- the coefficients β are **unbiased**, **uncorrelated** and have **minimum variance**
 - ▶ unbiased because linear function of unbiased measurements a
 - ▶ uncorrelated because S is a diagonal matrix
 - ▶ minimum variance because of C^{-1} in the least-squares estimate

findings →

Distribution of information over the expansion coefficients

- the β_i contribute uncorrelated independent information about b
- the contribution of β_i to the total variance of the unfolding result b is

$$\frac{1}{S_{ii}} \quad \text{with} \quad \sum_i \frac{1}{S_{ii}} = \text{Tr } C_b$$

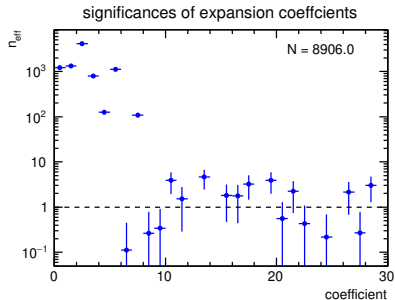
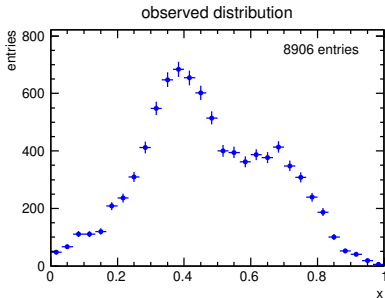
- effectively n_i events are used to determine the coefficient β_i

$$n_i^{\text{eff}} = \frac{\beta_i^2}{\sigma^2(\beta_i)} = \beta_i^2 S_{ii} \quad \text{with} \quad \sum_i n_i^{\text{eff}} \approx N \quad \text{the total number of events recorded}$$

(follows from $a^T C^{-1} a = N$ for Poisson distributed bin entries)

→ β_i with small S_{ii} are not measured but dominate the variance

example →



- ▶ 10^4 events generated – 8906 events accepted
- ▶ need 30 coefficients to fully specify the true distribution in 30 bins
- ▶ only $O(10)$ coefficients can be measured – the others must be guessed

simplest solution →

Truncation by cutoff-matrix D

❖ keep only the leading m coefficients and set the others to zero

$$b_0 = V \beta \rightarrow b = V D \beta \quad \text{with} \quad D = \begin{pmatrix} 1_m & 0 \\ 0 & 0 \end{pmatrix} \quad (m\text{-dim unit matrix})$$

■ relation between b and b_0 : posterior response P

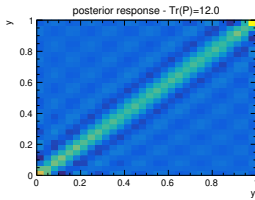
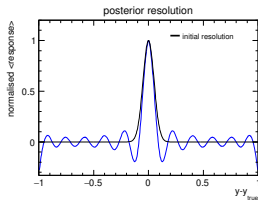
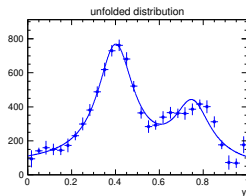
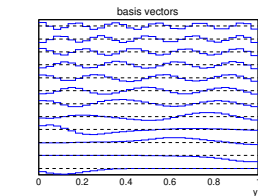
$$b = V D (V^T V) \beta = P b_0 \quad \text{with} \quad P = V D V^T \quad \text{and} \quad \text{Tr}(P) = m$$

■ note: perfect unfolding $P = 1$ requires $D = 1$

■ covariance matrices C_{b_0} and C_b

$$C_{b_0} = I^{-1} \quad \text{and} \quad C_b = P I^{-1} P^T$$

Numerical example



- using $m = 12$ terms
- correction of
 - ▶ efficiency losses
 - ▶ measurement biases
- resolution
 - ▶ marginal improvements
- visible artefacts due to fluctuations in data

Summary

- unfolding is a notoriously under-constrained problem
 - ▶ a small number of expansion coefficients contains all available information
 - ▶ efficiency losses and bias are usually corrected
 - ▶ smearing effects persist
 - ▶ attempts to get close to full corrections lead to artefacts
- a huge literature exists about methods to produce estimates of the true distribution
 - ▶ all more or less work, with differences that are not really understood
- criteria to compare results from different methods:
 - ▶ consistency of the result with the data
 - ▶ statistical uncertainties of the result: covariance matrix C_b
 - ▶ residual distortions as quantified by means of the posterior response P