# Unfolding in particle physics

## a window on solving inverse problems


Francesco Spanò

Royal Holloway University of London


IN2P3/CNRS School of Statistics
Autrans (Grenoble)
28th May - 1st June 2012

# Outline

- **What's Unfolding?**
- Initial unfolding schemes
  - the ML solution: matrix inversion
  - correction coefficients
- Exploring ML:
- Regularized Unfolding & not
  - curvature Tikhonov
  - iterative: Bayes inspired
  - entropy
  - Fully Bayesian unfolding
- Other unfolding schemes
- Two cents of experience and conclusions
  - optimization
  - bias & uncertainty
  - systematics  (& combination)

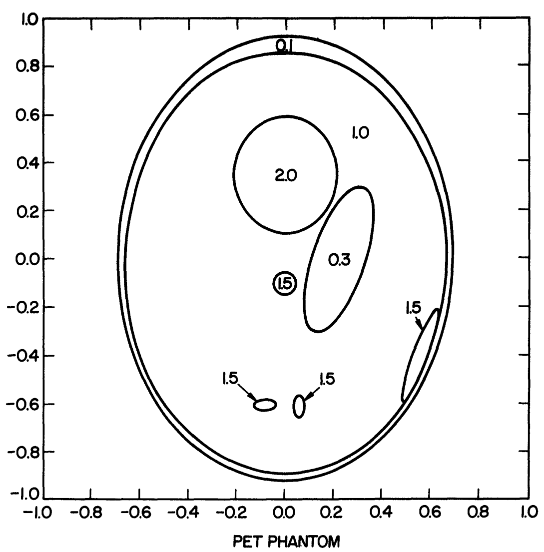
# The “inverse” problem (I)

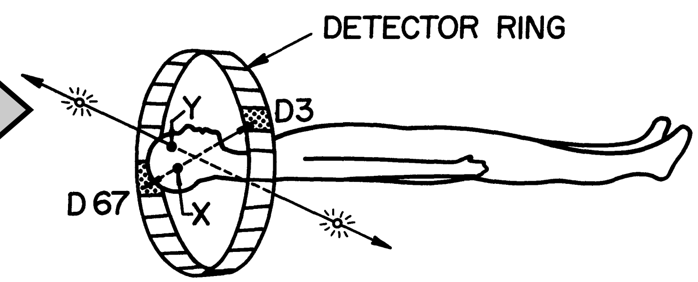
## *Unfolding, Deconvolving, Unsmearing*

*“True”*

*Reco/ Measured*

Vardi, Shepp, Kaufman JSTOR,V80, N389 (1985) pp.8



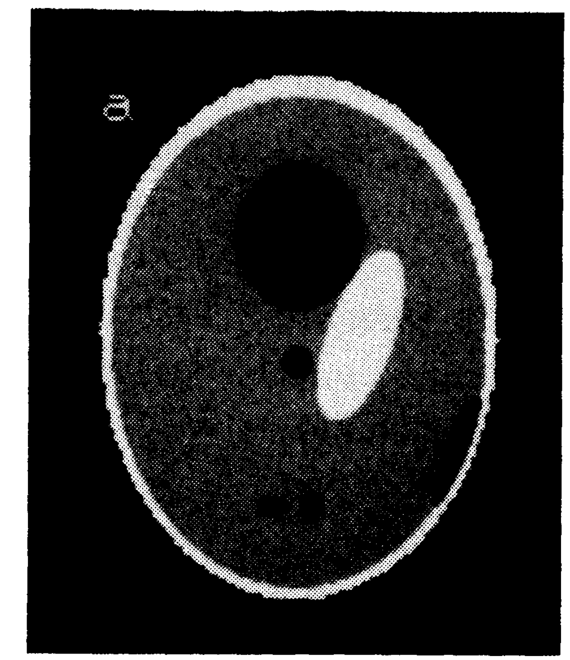
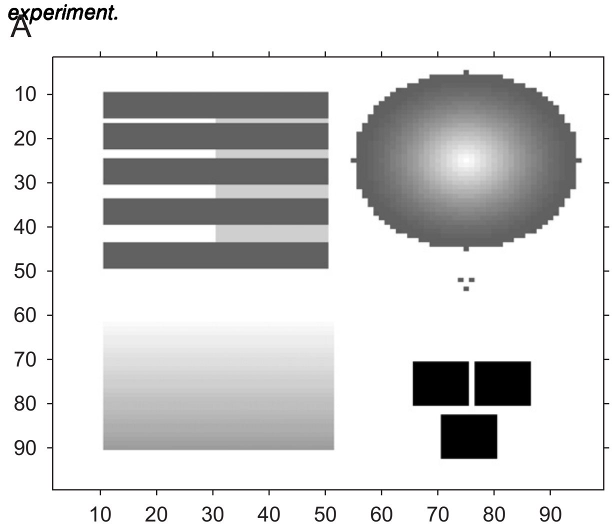
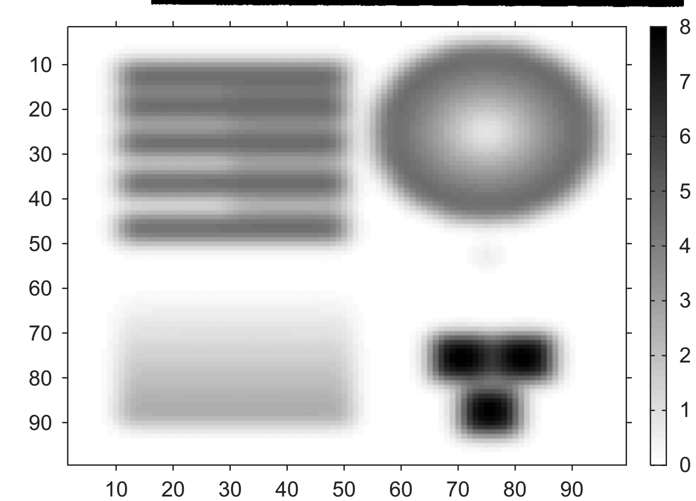




*Figure 2. The phantom used in the computer simulation of the PET experiment.*

O.Helene et al, Nucl Instr Meth. A, 580 (2007) pp. 1466-1473




Gaussian blurring



*how do we “go back”, invert the procedure?*

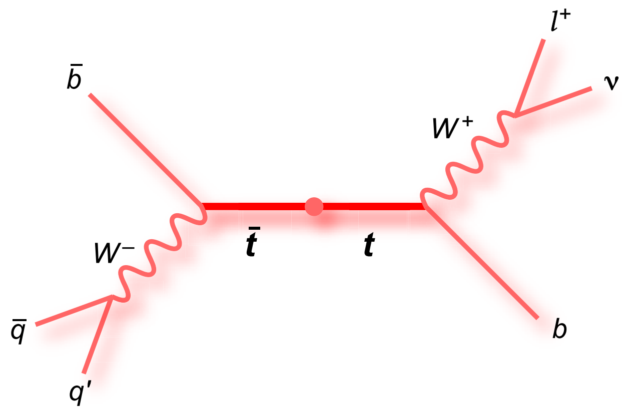
*What does “going back “ mean?*

?

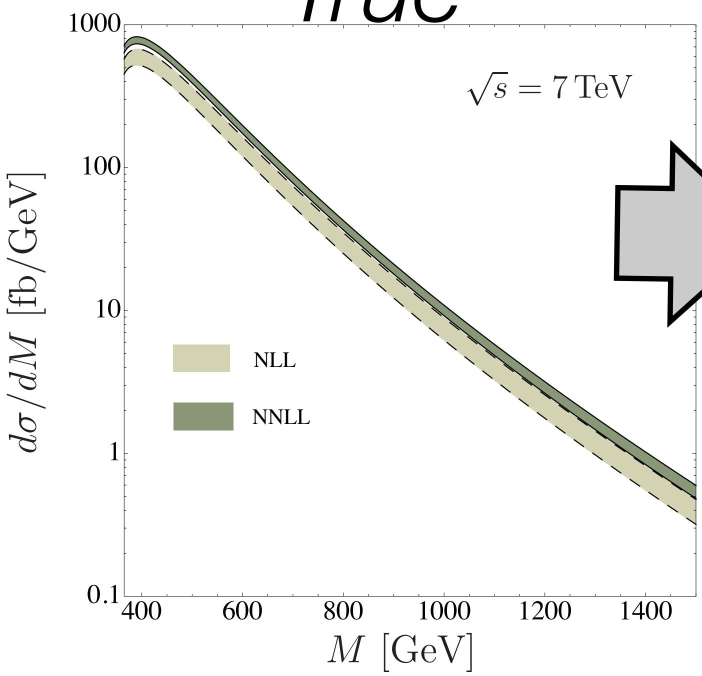
# The "inverse" problem (II)

*Unfolding, Deconvolving, Unsmearing*

*Reco/ Measured*

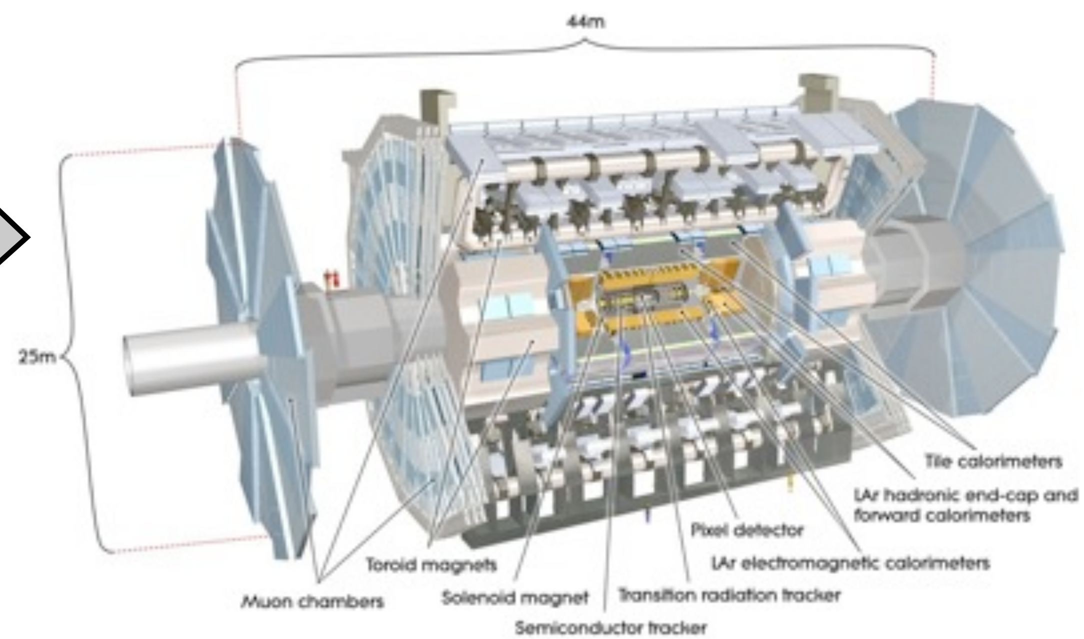



*"True"*

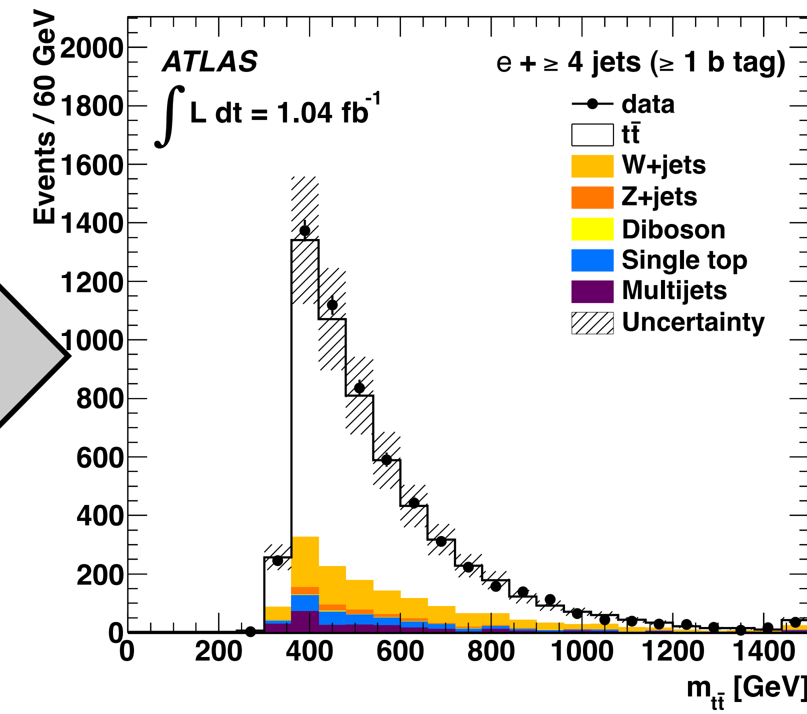



***$m_{tt}^{true}$***

***ATLAS***




accepted by Eur Phys J C
arxiv:1203.4211[hep-ex]




***$m_{tt}^{reco}$***

*how do we "go back", invert the procedure?*
*What does "going back " mean?*

?

# Stating the “inverse” problem

## *Unfolding, Deconvolving, Unsmearing*

- Estimate the prob distribution function for a (random) variable y

*Fredholm equation of 1st type*

“True” f(y) → Measurement $g(s) = \int_\Omega K(s,y)\, f(y)dy$ → Measured/ observed g(s)

reconstruction limitations
non uniform efficiency
resolution effects
(*smearing*)

- Due to the transformation, in general y and s can belong to multi-dimensional spaces with different dimensions

**Goal is to find f(y) : statistical estimation problem**

***efficiency, consistency, unbiasedness***

if have theory prediction g(y,a), fold it with K(y,x) and compare/extract parameters

if **no parametrized prediction exists,** unfolding means finding f(x) from g(x)

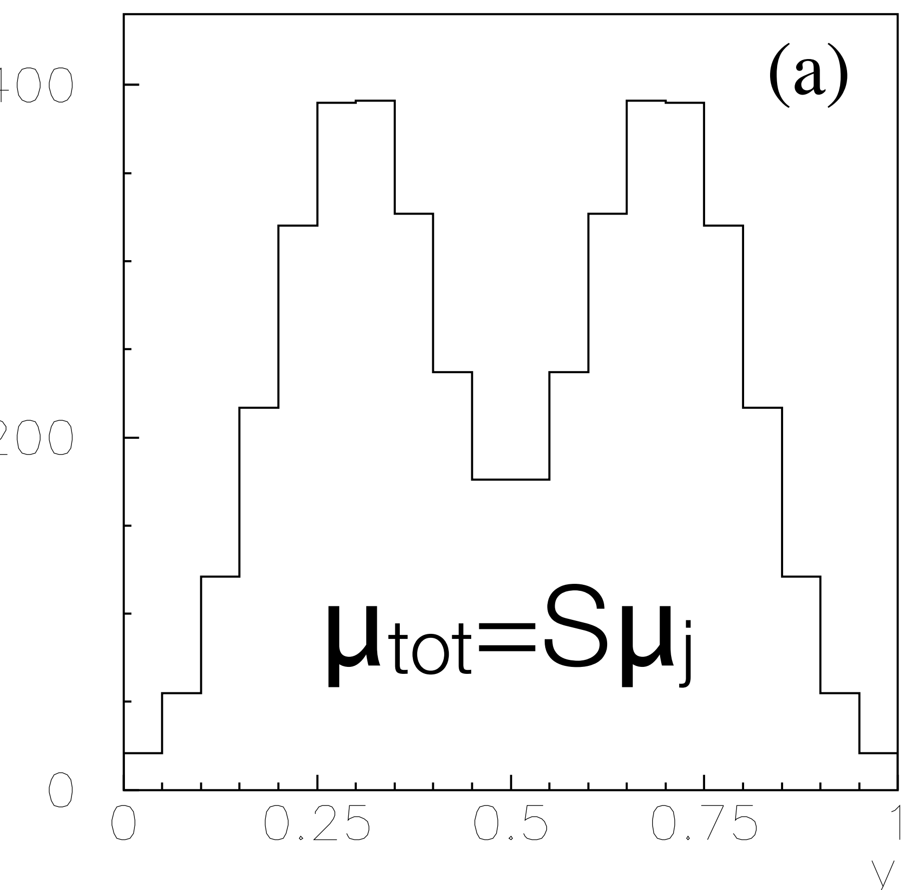
# Stating the inverse problem: continuous to discrete

- Operatively: measurements are limited in number and resolution → converted to histograms: discretization

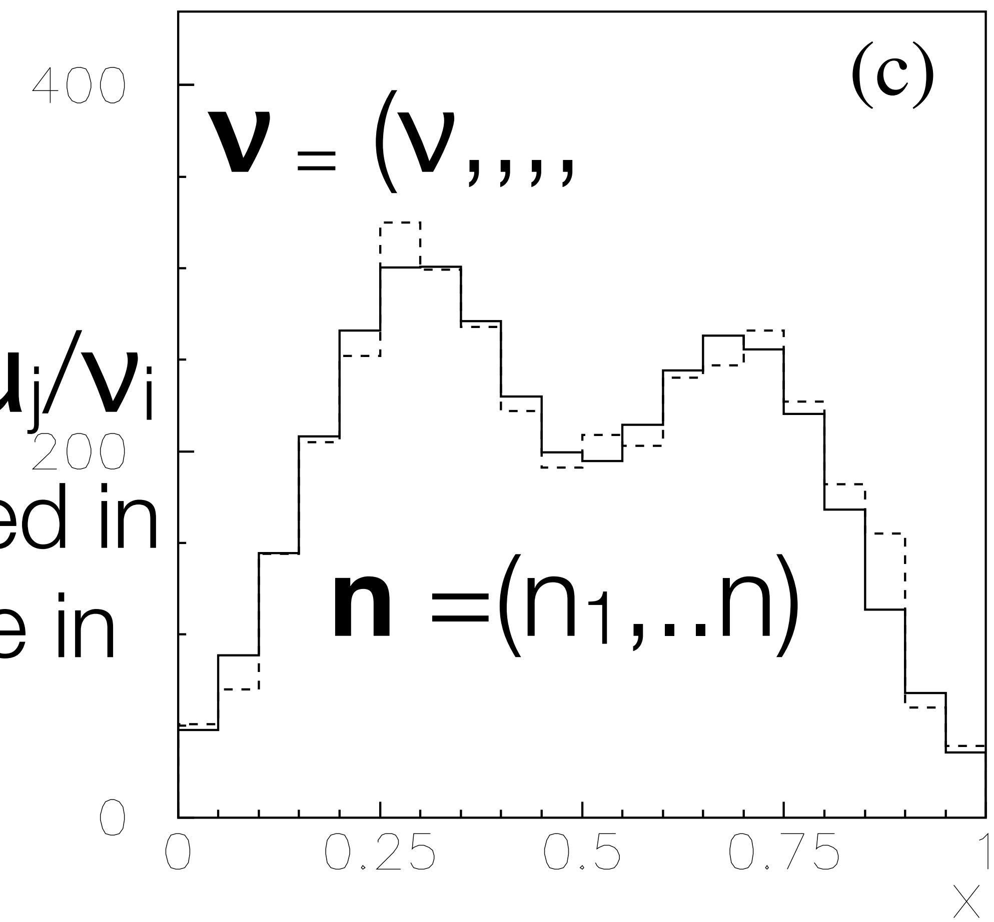
figures from G Cowan, A survey of unfolding methods for Part Phys ,PHYSTAT2002

$$g(s) = \int_\Omega K(s,y)\, f(y)\, dy$$

true




expected/observed




$$K(s,y) \rightarrow R(i,j) = \mu_j/\nu_i$$

P(observed in bin i |true in bin j)

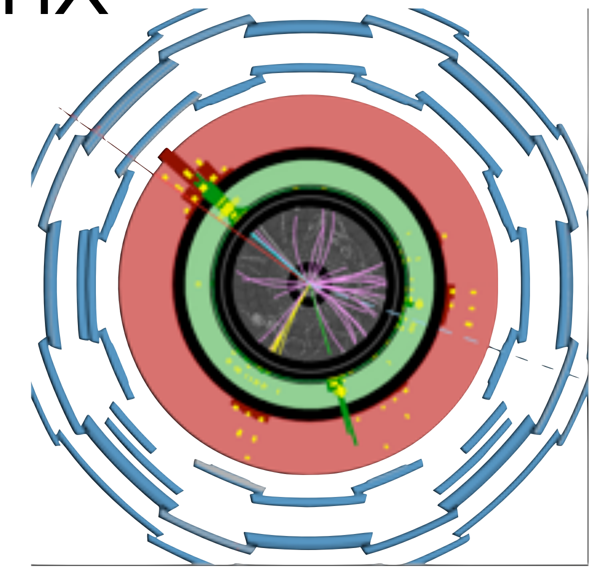
$$f(y) \rightarrow \mu_j = \int_{y_{j-1}}^{y_j} f(y)\,dy$$

ν, μ: constants
n random

$$g(s) \rightarrow \nu_i = \int_{s_{i-1}}^{s_i} g(s)\,ds$$

$$E[n_i] = \nu_i = \sum_j R(i,j)\, \mu_j$$

# Ingredients: from transfer function to matrix



- The general definition of the transfer matrix

$$R(i,j) = \frac{\int \int K(s,y)\, f(y)dy\, ds}{\int f(y)dy}$$

- If we have the Kernel from the problem we solve it directly
- R is usually obtained from
  - detailed simulation of the measuring apparatus when many effects are included: MC events are generated with fSim , our best guess of f and mapped to gSim the resulting best guess of g
  - test measurements, for instance exposing calorimeter to particle beam of well known fixed energy x=x0 implies we measure $\delta(y-y_0)$ so the measurement gives R directly

$$\int_a^b K(s,y)\, \delta(y-y_0)dy= K(s,y_0)$$

R is in general a rectangular MxN matrix

# Additional Ingredients: efficiency and backgrounds

- Some interesting events are not observed due to detection inefficiency. The efficiency of detection/acceptance be included in the estimate of the response matrix.

$$\sum_{j=1} R(i,j) = \sum_{j=1} P(\text{observed in bin } i | \text{ true value in bin } j) =$$
$$= P(\text{observed anywhere} | \text{ true value in bin } j) = \boldsymbol{\varepsilon}_j$$

- Some observed events are due to backgrounds and they modify the observed distribution

$$g(s) = \int_\Omega K(s,y)\, f(y) dy + b(y)$$

$$\boldsymbol{\beta}_i = \int b(y) dy$$

$$E[\mathbf{n}_i\,] = \boldsymbol{\nu}_i = \sum_j R(i,j)\, \boldsymbol{\mu}_j + \boldsymbol{\beta}_i$$

$\boldsymbol{\beta}_i$ is expected N of bkg events in ***OBSERVED*** distribution

# The first step: maximum likelihood solution

log=$\log_e$

- Given the problem $\boldsymbol{\nu} = R\,\boldsymbol{\mu} + \boldsymbol{\beta}$ one can consider if the inverse of R exists and then provide the solution as

$$\boldsymbol{\mu} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta})$$

- Suppose the data are independent Poisson observation

$$P(n_i \mid \nu_i) = \nu_i^{n_i} \frac{e^{-\nu_i}}{n_i!}$$

- The log likelihood is

$$\log L(\boldsymbol{\mu}) = \sum_{j=1}^{N} (n_i \log \nu_i - \nu_i - \log n_i!)$$

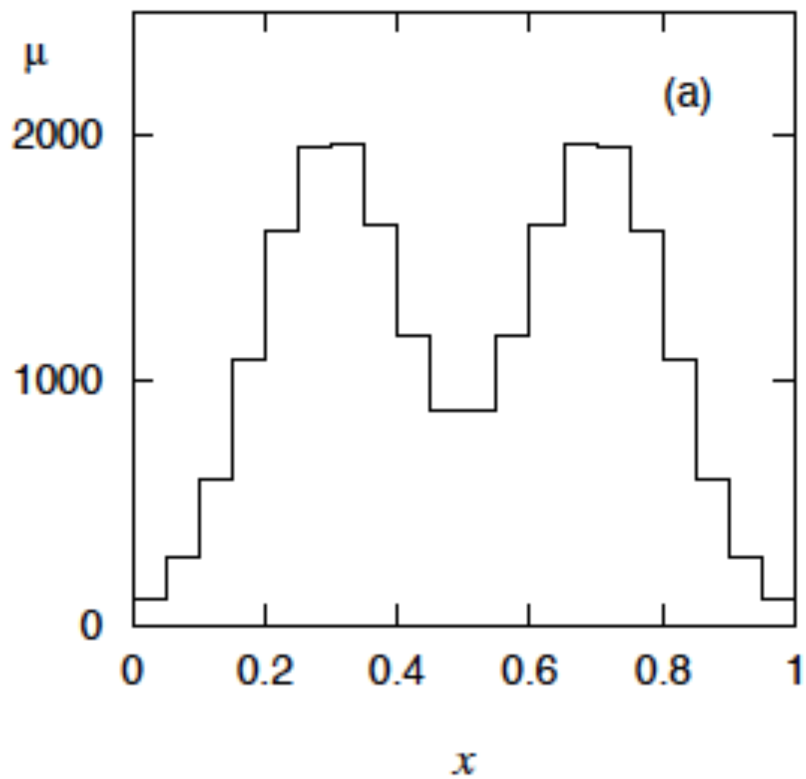
- The maximum likelihood (ML) estimator is ($d\log L/d\boldsymbol{\mu}=0$)

$$\boldsymbol{\nu}_{ML} = \mathbf{n} \longrightarrow \boldsymbol{\mu}_{ML} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta})$$

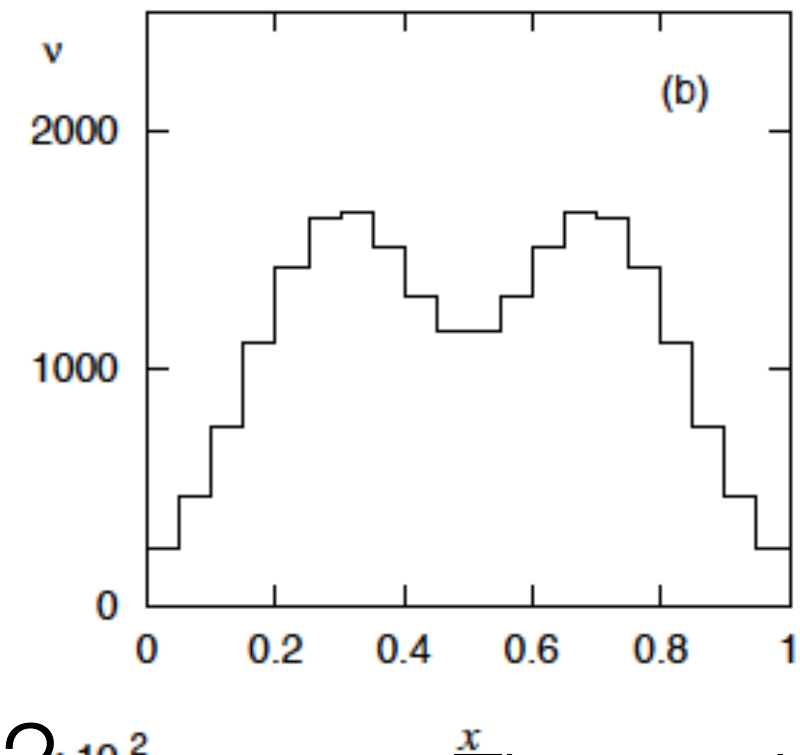
# ML solution:does it work?

G Cowan, A survey of unfolding methods for Part Phys ,PHYSTAT2002

True




Expected



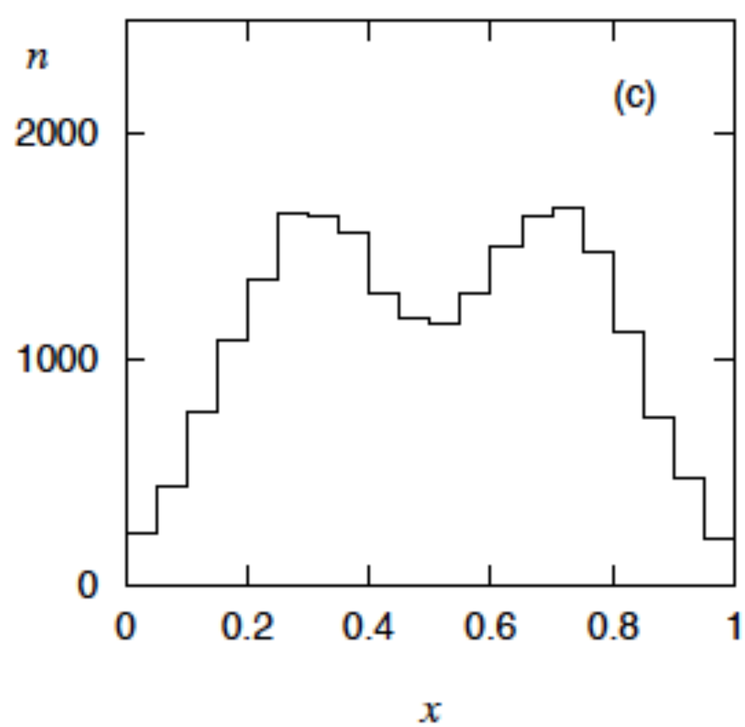

Consider data example

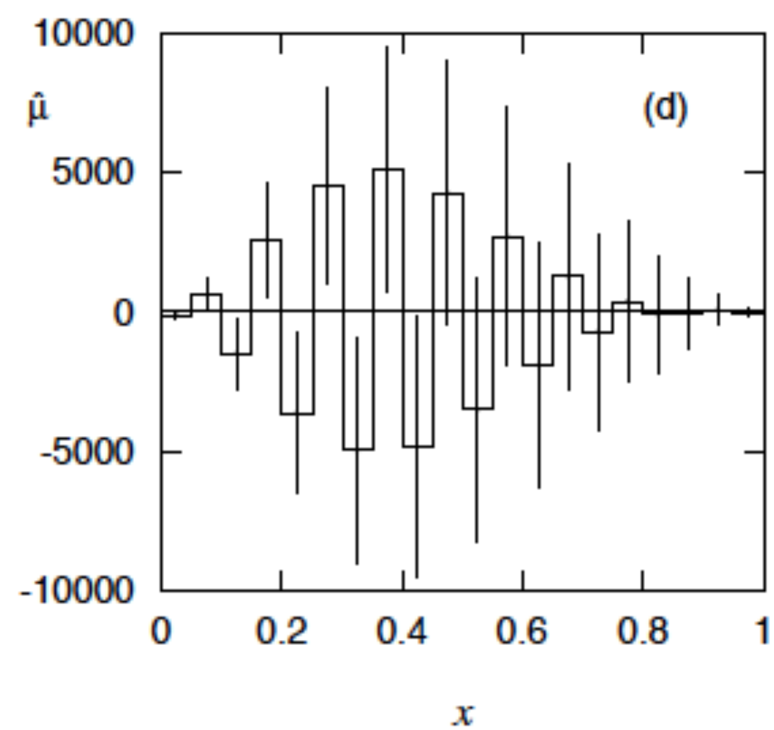
?

x 10 2

The result from

$\boldsymbol{\mu}_{ML} = R^{-1} (\boldsymbol{\nu} - \boldsymbol{\beta})$

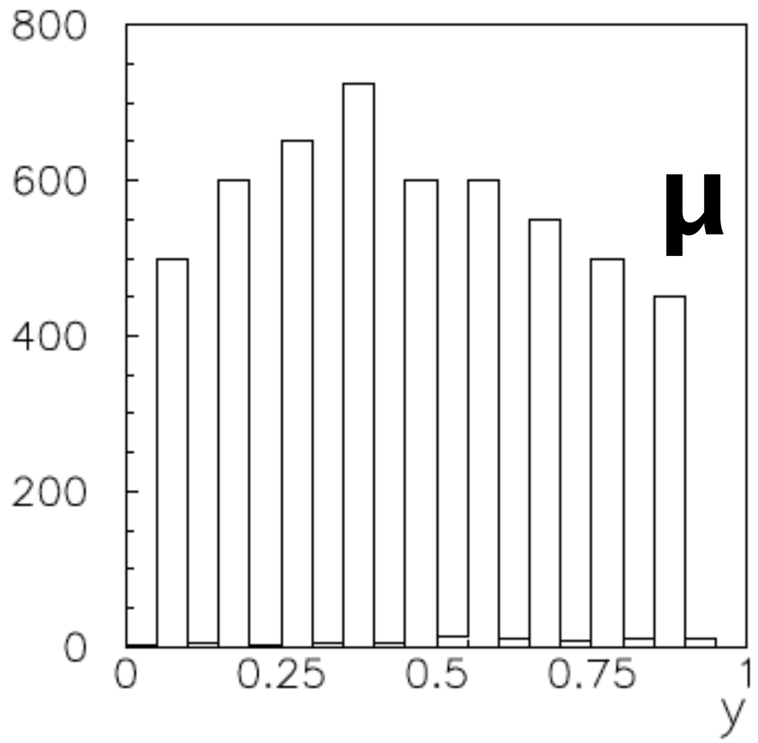


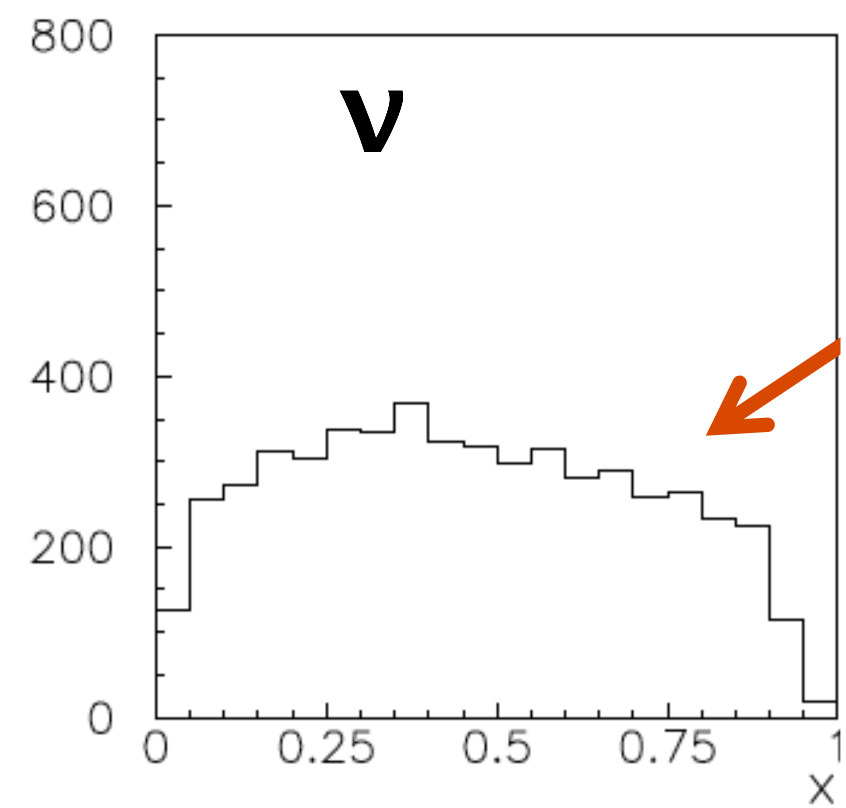

- Major failure?

# ML solution: What was wrong?

G Cowan, A survey of unfolding methods for Part Phys ,PHYSTAT2002




- Take the case where **μ** really had a lot of fine structure




- The response R dilutes the info (smoothen), but allows residual structure to be present

- The application of $R^{-1}$ restores the structure $\boldsymbol{\mu}_{ML} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta})$

  BUT we do not have **ν, we have n:** R "assumes" the fluctuations in **n** are the residual of the "real" original structure and puts the pattern back into **ν** to get **μ** ( i.e. "magnifies" flucts back)

# ML solution: before you leave (1)

- Bias: the ML solution is an unbiased estimator

$$E[\boldsymbol{\mu}_{ML}] = E\,[R^{-1}\,(\mathbf{n} - \boldsymbol{\beta})] = R^{-1}\,(E[n]-\boldsymbol{\beta}) = R^{-1}\,(\boldsymbol{\nu}-\boldsymbol{\beta})$$

- Its covariance is

$$U_{ij} = \mathrm{cov}[\mu_{ML,i}\ \mu_{ML,j}] = \sum_{k,l=1}^{N} (R^{-1})_{ik}\,(R^{-1})_{jl}\,\mathrm{cov}[n_k\,,n_l]$$
$$= \sum_{k,l=1}^{N} (R^{-1})_{ik}\,(R^{-1})_{jl}\,\delta_{k,l}\,\nu_k = \sum_{k=1}^{N} (R^{-1})_{ik}\,(R^{-1})_{jk}\,\nu_k$$

# ML solution: important properties (2)

- The Cramér-Rao inequality states that for unbiased estimators the co-variance has a minimum value (lower bound)

$$(U^{-1}_{bound})_{kl} = - E\left[\frac{\partial^2 \log L}{\partial \mu_k \, \partial \mu_l}\right] = \sum_{i=1}^{N} \frac{R_{ik} R_{jl}}{\nu_i}$$

- If we invert it we get

$$U_{ij,bound} = \sum_{k=1}^{N} (R^{-1})_{ik} (R^{-1})_{jk} \nu_k = U_{ij}$$

this IS the variance!

- The ML solution provides the smallest variance amongst the unbiased estimators, albeit large

# Turning point: bias vs variance

- ML estimator for unfolding attain minimum variance amongst unbiased estimators

- Estimators providing a reduction in variance will necessarily introduce bias

- The balance between bias and variance is the name of the game in unfolding/deconvolving/smearing

- Important to understand from where the problem is coming from: understanding source means understanding the cure and its validity

# Take a step back: correction factors

- Use same binning for **μ ν and take** $\mu_{i,est} = C_i\,(n_i - \beta_i)$ where
  - $C_i = \dfrac{\mu_i^{MC}}{\nu_i^{MC}}$ — $\mu_i^{MC}$ and $\nu_i^{MC}$ result from simulation (no bkg included)
  - $U(i,j) = cov[\mu_{est,i}\ \mu_{est,j}] = C_i^2\ cov[n_i\ ,n_j]$
- $C_i$ is often of O(1) so stat errors are much smaller than ML case
- However bias is $b = \left( \dfrac{\mu_i^{MC}}{\nu_i^{MC}} - \dfrac{\mu_i}{\nu_i^{sig}} \right) \nu_i^{sig}$ $\qquad \nu_i^{sig} = \nu_i - \beta_i$
- No bias only if MC=Nature; bias pulls results to MC
- Note: assuming R to be diagonal, while it might well (and usually) not be so

# Take a step back: correction factor uncertainties (II)

(Example from R. Cousins)

- Assume the for some bin i one has

$$C_i = 0.1 \qquad \beta_i = 0 \qquad n_i = 100$$

- Then

$$\mu_{i,est} = C_i n_i = 10 \qquad \sigma_{\mu_{i,est}} = C_i \sqrt{n_i} = 1$$

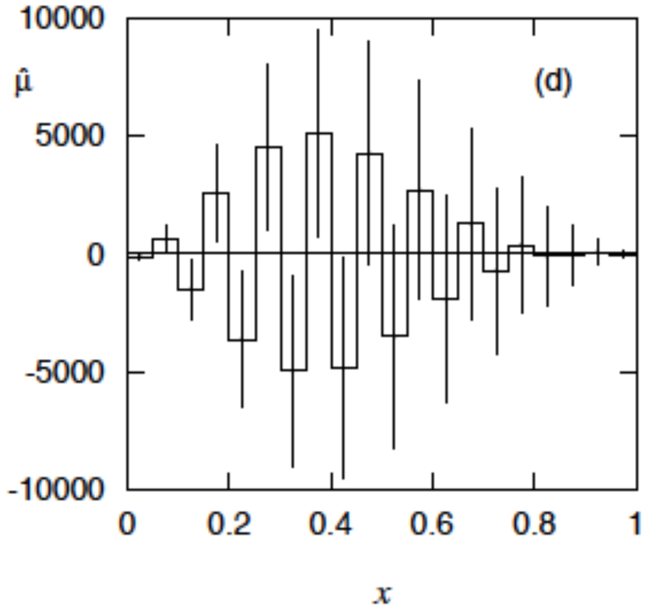
- However the estimate maintains that only 10 of 100 events observed in the bin really belong to it while the rest “migrated” in from outside the bin

- How can it be possible to have a 10% measurement if only 10 events are really carrying information about the bin content?

# Take a step back: correction factors (III)

- Features
  - C depends on the assumed distribution which one is trying to find
  - Bin-to-bin correlations are completely neglected
  - Sum of estimated #events in truth can be different from sum of observed data
- Reduction of stat uncertainty is obtained in exchange for bias (standard unfolding). Hard to quantify the bias (also hard in other cases)
- Bias is reduced if bin width is large compared to resolution i.e. if migrations are small (non diagonal elements in R are
- Useful for quick solution, if bias << other uncertainties
- Best to avoid using it

# Back to basics: where to from ML?




A problem is called improper when large and sometimes infinite changes in the solution could correspond to small changes in the input data.

I.P Nedelkov, *Improper problems in computation physics*, Com.Phys Comm 4 (1972) 157

- Propagation of uncertainties is a measure of stability

- When solving $\mathbf{n}-\boldsymbol{\beta}=R\boldsymbol{\mu}_{est}$ for $\boldsymbol{\mu}_{est}$ ,consider the maximum ratio of the relative inaccuracy on $\boldsymbol{\mu}_{est}$ to the one on $\mathbf{y}=\mathbf{n}-\boldsymbol{\beta}$

$$c(R) = \max_{y,\delta y} ||\delta\boldsymbol{\mu}_{est}||/||\boldsymbol{\mu}_{est}|| / ||\delta\mathbf{y}|/||\mathbf{y}||$$

- c(R) is called the *condition* of matrix R : upper bound on magnification factor for the input data uncertainties.
- Large c(R) implies instability under small fluctuations in the data i.e. sensitivity to noise

see for instance <u>S Leach *SVD A primer*</u> and ref. therein

# Back to basics:where to from ?ML(II) $\mathbf{n}-\boldsymbol{\beta}=E[\boldsymbol{\nu}]-\boldsymbol{\beta}=R\boldsymbol{\mu}$

- Given $V_{\mathbf{y}}$,cov matrix of **y**, "rotate" R and **y** such that $V(\mathbf{y}')=\mathbf{1}$ (identity matrix) → normalized variables according to uncertainty
  - i.e. $(R\boldsymbol{\mu}_{est} - \mathbf{y})V_{\mathbf{y}}^{-1}(R\boldsymbol{\mu}_{est} - \mathbf{y}) = (R'\boldsymbol{\mu}_{est} - \mathbf{y}')(R'\boldsymbol{\mu}_{est} - \mathbf{y}')$

- Consider Singular Value Decomposition of R matrix

  Every matrix R of dimension MxN can be decomposed as

  $$R = U\,\Sigma\,V^T$$

  such that $U=(\mathbf{u}_1,\ldots\mathbf{u}_N) \in R^{MxN}$ and $V=(\mathbf{v}_1,..,\mathbf{v}_N) \in R^{NxN}$ are unitary matrices ($U^TU=\mathbf{1}$) and $\Sigma = U^TRV \in R^{MxN}$ is diagonal = $\{\sigma_1,..,\sigma_n\}$

- If R' is inverted using SVD decomposition and $\sigma_j \neq 0 \;\; \forall j$

$$\boldsymbol{\mu}_{est}=R'^{-1}(\mathbf{n}-\boldsymbol{\beta})=\mathbf{R}'^{-1}\mathbf{y}' = (V\,\Sigma^{-1}\,U^T)\mathbf{y} = \sum_{j=1}^{N}\frac{1}{\sigma_j}(\mathbf{u}^T_j\,\mathbf{y})\,\mathbf{v}_j = \sum_{j=1}^{N}\frac{1}{\sigma_j}\mathbf{c}_j\,\mathbf{v}_j$$

  - if ordered in value, $c_j$ decreases with j : often steeply (exponentially for Gaussian response)
  - Contribution of $c_j$ is weighted with inverse of singular value: small singular values → large fluctuations

see also V Blobel, in Proc of PHYSTAT2011

# Back to basics: Where to from ML? (III)

- Is there a connection with magnification of uncertainties?

$$||\delta \boldsymbol{\mu}_{est}|| / ||\boldsymbol{\mu}_{est}|| \,\boldsymbol{/}\, ||\delta \mathbf{y}|| / ||\mathbf{y}|| \quad = \quad ||\mathbf{R}(\delta \mathbf{y})|| / ||\mathbf{R}\mathbf{y}|| \,\boldsymbol{/}\, ||\delta \mathbf{y}|| / ||\mathbf{y}||$$

- It can be shown that

$$c(R) = ||R|| \, ||R^{-1}|| = \sigma_{max} / \sigma_{min}$$

from SVD decomposition of R

- The condition of R matrix can be read off its SVD decomposition

- $|| \; ||$ *is the operator norm of a matrix induced by the euclidean norm*

*if* $R: R^N \rightarrow R^M$ *with euclidean norm* ($||x|| = \sum_i |x_i|^2$) *a norm on the matrix is induced as*

$$||R|| = \sqrt{(\text{max eigenval of } R^T R)}$$

# Where to from ML? The picture

- SVD decomposition gives insight into the unfolding problem: small effects can lead to large changes in ML estimator→ large sensitivity to small fluctuations, high frequency →large condition number

- Need to suppress “noise” info i.e. reduce impact of high frequency, noisy components while preserving as much “signal” as possible. Come to terms with condition number.

- “Regularize” the problem, by accepting some bias in exchange for reduced variance

# Regularized unfolding- General view (I)

- LKL (or sum of squares~-2logL ) quantifies the distance between data n and expectation **ν**.

- Take a step back and consider region of **μ around ML solution**

$$\log L(\boldsymbol{\mu}) \geq \log L_{max} - \Delta \ln L$$

- Out of these estimators choose the “smoothest” (one with less fluctuations) according to some measure i.e. maximize

$$\phi(\boldsymbol{\mu}) = \alpha \log L(\boldsymbol{\mu}) + S(\boldsymbol{\mu}) \quad \text{or} \quad \phi(\boldsymbol{\mu}) = \log L(\boldsymbol{\mu}) + \tau S(\boldsymbol{\mu})$$

$S(\boldsymbol{\mu})$: regularization function (to measure smoothness)

$\alpha$ or $\tau$: regularization parameter (to give the desired $\Delta \ln L$)

# Regularized unfolding- General view (II)

- Possibly add $n_{tot} = \sum_{i=1}^{N} \nu_i$ if one wants solutin to provide an unbiased estimate of total entries and consequently maximize

$$\phi(\boldsymbol{\mu}, \lambda) = \alpha \log L(\boldsymbol{\mu}) + S(\boldsymbol{\mu}) + \lambda(n_{tot} - \sum_{i=1}^{N} \nu_i)$$

over $\boldsymbol{\mu}$ and $\lambda$

*function of* $\boldsymbol{\mu_i}$ *as* $\nu_i = \sum_{i=1}^{N} \boldsymbol{\mu_i} + \beta_i$

- $\alpha = 0$ smoothest solution (data are ignored! impose S shape)
- $\alpha \to \infty$ recover ML solution (S carries no weight in the maximization)

Ingredients

$S(\boldsymbol{\mu})$

prescription for $\alpha$

# Regularization: Tikhonov scheme

- Consider the mean square of the k[th] derivative = measure of smoothness

$$S[f(y)] = \int \left(\frac{d^k f(x)}{dy^k}\right)^2 dy$$

with k=1,2,3...

- Using k=2 one has

$$S(\boldsymbol{\mu}) = \sum_{j=1}^{M-2} \left[ (\boldsymbol{\mu}_{i+1} - \boldsymbol{\mu}_i) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1}) \right]^2$$

*numerical 2nd derivative*

and summing it with $\log L = -1/2\ \chi^2$, $\chi^2 = (R\boldsymbol{\mu} - \mathbf{y})^T V_{\mathbf{y}}^{-1}(R\boldsymbol{\mu} - \mathbf{ny})$ with $\mathbf{y}=\mathbf{n}-\boldsymbol{\beta}$, the result is

$$\phi(\boldsymbol{\mu}, \lambda) = -\alpha/2\ \chi^2(\boldsymbol{\mu}) + S(\boldsymbol{\mu})$$

quadratic in $\boldsymbol{\mu}$

↓

- First derivatives of $\phi(\boldsymbol{\mu}, \lambda)$ w.r.t. $\boldsymbol{\mu}, \lambda$ return linear equations

# Tikhonov with k=2 + SVD

Hoecker, V Kartvelishvili Nucl Instr Meth A 372 1996 469

see also V Blobel, in Proc of PHYSTAT2011A

- Minimize

$$\phi(\boldsymbol{\mu}, \lambda) = -1/2\, \chi^2 (\boldsymbol{\mu}) + \tau \sum_{j=1}^{M-2} [ (\boldsymbol{\mu}_{i+1}-\boldsymbol{\mu}_i) - (\boldsymbol{\mu}_i-\boldsymbol{\mu}_{i-1})]^2$$

- **"Rotate"** R and **y** = **n-β** so that $V_{\mathbf{y}} \sim \mathbf{Id}$ ($\rightarrow$ in $\chi^2$ $\mathbf{Id}^{-1} = \mathbf{Id}$)

- In matrix format, minimize

$$(R'\boldsymbol{\mu} - \mathbf{y}')^T (R'\boldsymbol{\mu} - \mathbf{y}') + \tau\, (C\, \boldsymbol{\mu})^T (C\, \boldsymbol{\mu}) \longrightarrow \begin{pmatrix} R'\boldsymbol{\mu} \\ \sqrt{\tau} C \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \rightarrow \begin{pmatrix} R'C^{-1} \\ \sqrt{\tau}\, \mathbf{Id} \end{pmatrix} C\boldsymbol{\mu} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix}$$

*C encodes 2nd derivaties*

- Expand $R'C^{-1}$ with SVD and express solution as a function of τ and of the solution for τ=0

Equivalent to

$$\boldsymbol{\mu}_{\mathbf{est}} = \sum_{j=1}^{N} \frac{1}{\sigma_j} \boldsymbol{\phi}_{\mathbf{j}}\, \mathbf{c}_{\mathbf{j}}\, \mathbf{v}_j \text{ with } \boldsymbol{\phi}_{\mathbf{j}} = \frac{\sigma_j^2}{\sigma_j^2 + \tau}$$

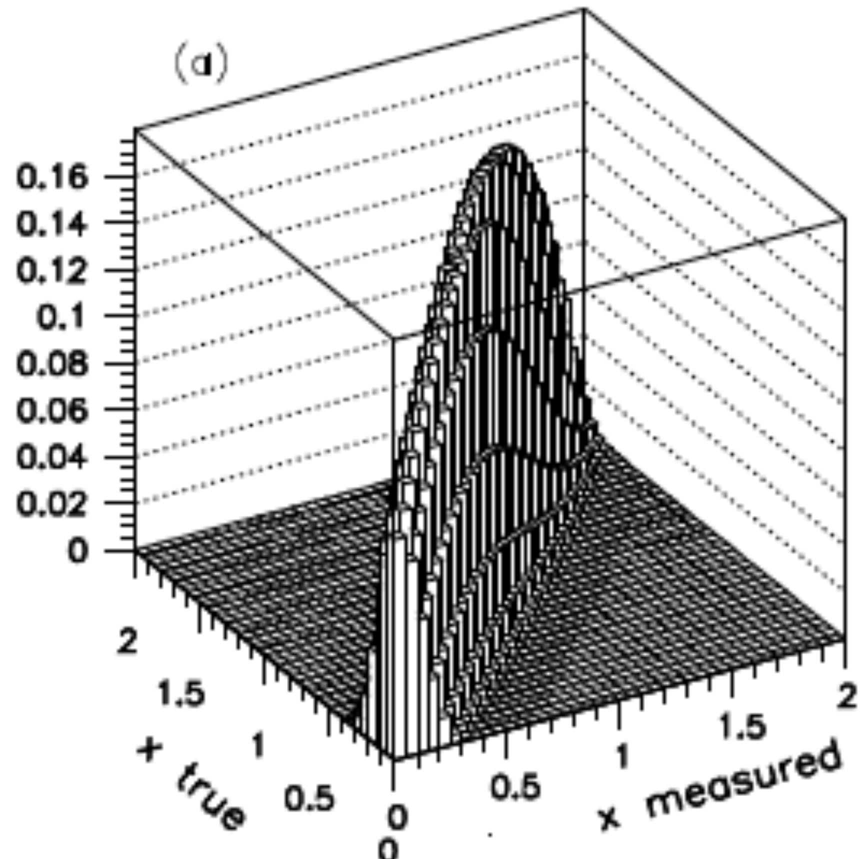
low pass filter with a preference for small curvature

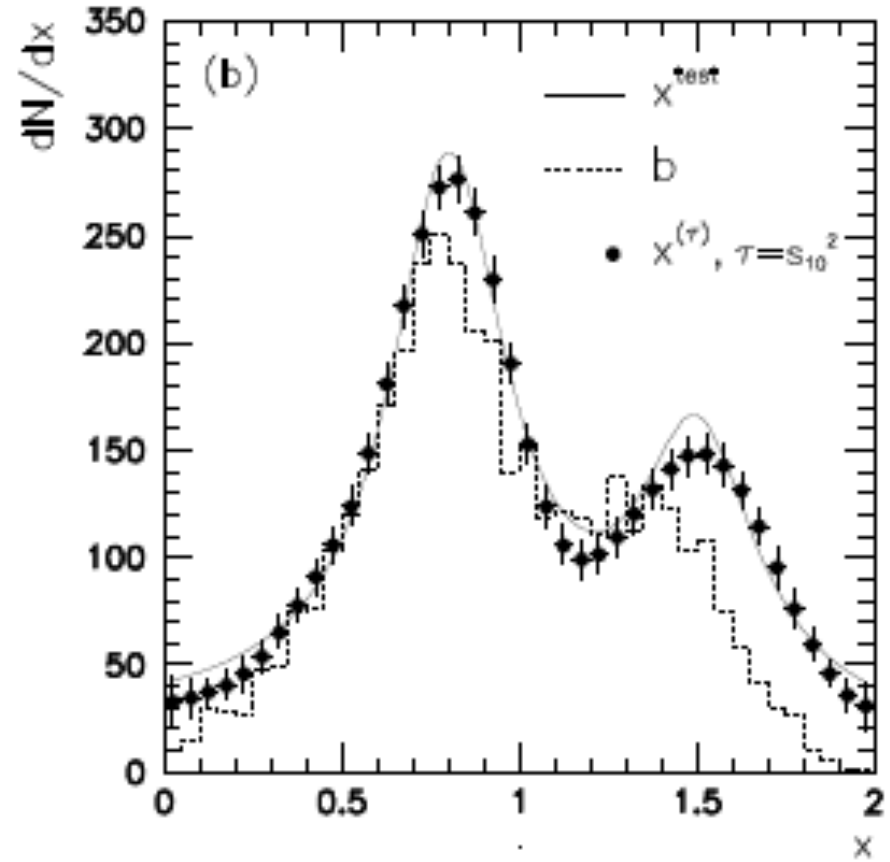
- Choice of τ
  - number of transformed values significantly different from zero
  - unfold folded test distribution vs τ: choose τ with best $\chi^2$(unfolded, true)

# Tikhonov with n=2 + SVD

A Hoecker, V Kartvelishvili Nucl Instr Meth A 372 1996 469

Response matrix




solid line: true dist

dashed hist : measured

dots: unfolded

$d_i$ = normalized, SVD-transformed, observed bin content

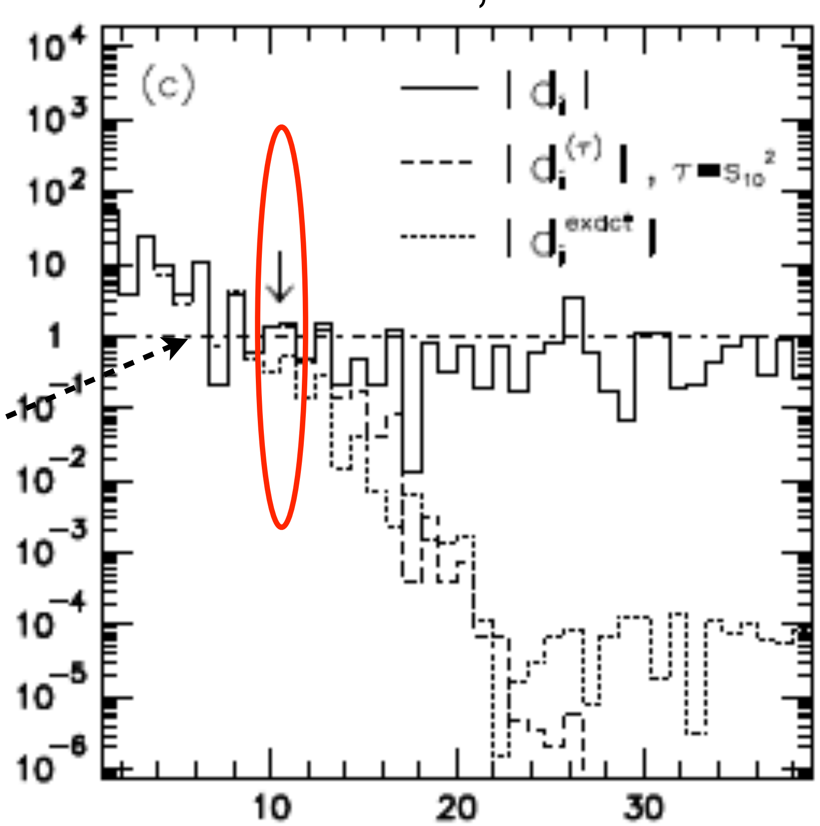
solid: true

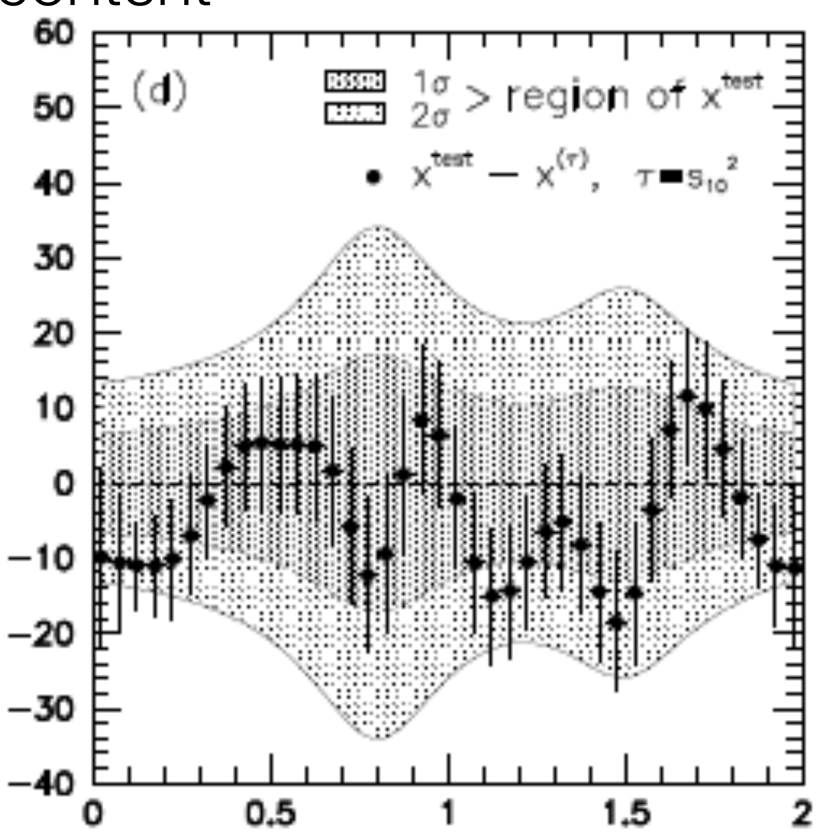
dashed: regularized

dotted: no stat fuctuations

stat errors in $d_i$

significant/non-significant boundary




deviation of unfolded from true

# Iterative unfolding: the idea

L.B.Lucy, *An iterative technique for the rectification of observed distributions* Astronomical Journal 79 (6) (1974) 745

- Consider Bayes' theorem with "true" and "reco" labels

known wanted guess

(I) $$P(\text{true}|\text{obs}) = \frac{P(\text{obs}\,|\,\text{true})\, f(\text{true})}{\int f(\text{true}) P(\text{obs}|\,\text{true})\, d\text{true}} = \frac{P(\text{obs}\,|\,\text{true})\, f(\text{true})}{g(\text{obs})}$$

(II) $$f(\text{true}) = \int g(\text{obs}) P(\text{true}|\text{obs})\, d\text{obs}$$

Note from (I) P(true|obs) is function of f (true), the real inverse kernel for f needs to be function of P(obs | true) only

Iterate following steps

- Guess f(true) and use (I) to estimate P(true|obs) using known P(obs | true)

$$g^r(\text{obs}) = \int f^r(\text{true})\, P(\text{obs},\text{true})\ \text{true}$$

$$P^r(\text{true}|\text{obs}) = \frac{f^r(\text{true})\, P(\text{obs},\text{true})}{g^r(\text{obs})}$$

- Use **the ansatz** of (II) and integrate P(true|obs) over ĝ (obs) **estimated from data**

$$f^{r+1}(\text{true}) = \int \hat{g}(\text{obs})\, P^r(\text{true}|\text{obs})\ d\text{obs}$$

$$f^{r+1}(\text{true}) = f^r(\text{true}) \int \frac{\hat{g}(\text{obs})}{g^r(\text{obs})} P(\text{obs},\text{true}) d\text{true}$$

- Check quality measure

# Iterative unfolding (II): implementation

see G.Zech in PHYSTAT 2011 proceedings

- Idea: if R is positive definite, invert relation $\mathbf{x} = \mathbf{n} - \boldsymbol{\beta} = E[\boldsymbol{\nu}] - \boldsymbol{\beta} = R\boldsymbol{\mu}$ iteratively
- Start with a guess of $\boldsymbol{\mu}^{(0)}$ **calculate** $\mathbf{x}^{(0)} = \mathbf{R}\,\boldsymbol{\mu}^{(0)}$

Iterate

known wanted

"Fold"

- Predict $\mathbf{x}_i^{(k)}$ from k-th estimate of $\boldsymbol{\mu}_j$

- $\mathbf{x}_i^{(k)} = \sum_j R_{ij}\, \boldsymbol{\mu}_j^{(k)}$

"Unfold"

- kth estimate of true $\boldsymbol{\mu}_j$ formed by "integrating" $R_{ij}$ over updating function $\mathbf{x}_i/\mathbf{x}_i^{(k)}$

- $\boldsymbol{\mu}_j^{(k+1)} = 1/\varepsilon_j \sum_i R_{ij}\, \boldsymbol{\mu}_j^{(k)}\, (\mathbf{x}_i/\mathbf{x}_i^{(k)})$

if $\mathbf{x}_i/\mathbf{x}_i^{(k)} \sim 1$, $\sum_i R_{ij} = 1$ so $\boldsymbol{\mu}_j^{(k+1)} = \boldsymbol{\mu}_j^{(k)}$

- Dividing by efficiency $\varepsilon_j$ corrects for acceptance losses

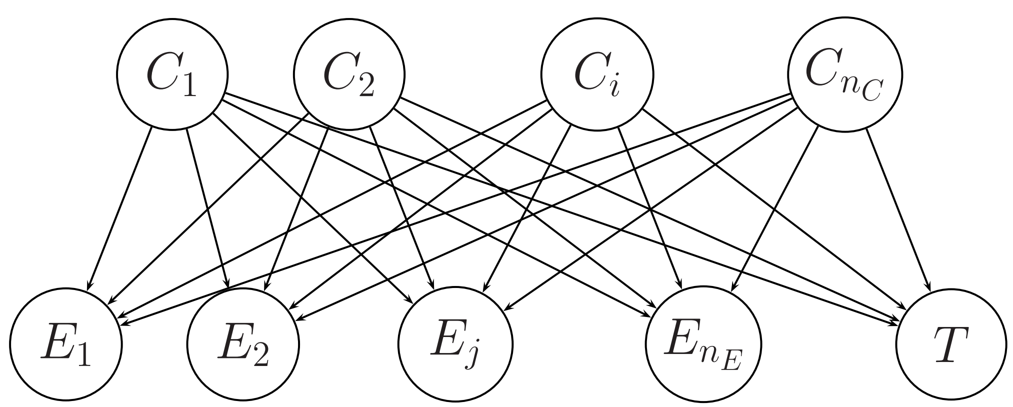
converge to ML solution for Poisson uncertainties (empirical)

# Iterative unfolding and regularization

G. D'Agostini Nucl. Instr. Meth A 362 1995 (487)
arxiv:1010.0632
and ref therein

- Standard basic iterative steps:
  - initial guess is $p_i = 1/M$, so $\mu_i = n_{tot}\, p_0$
  - estimate kth observed with kth guess of "true" dist ("fold")
  - get k+1st estimate of "true" dist by integrating data-scaled kth estimate over observed ("unfold")

*No final estimator in terms of prior updating rule inspired to Bayes*




$$P(C_i|E_j) = \frac{P(E_j|C_i)\cdot P_o(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l)\cdot P_o(C_l)}$$

$$\hat{n}(C_i) = \frac{1}{\epsilon_i}\sum_{j=1}^{n_E} n(E_j)\cdot P(C_i|E_j)$$

$\mu_i^{(1)}$ → $\hat{n}(C_i)$; $x_i$ → $n(E_j)$

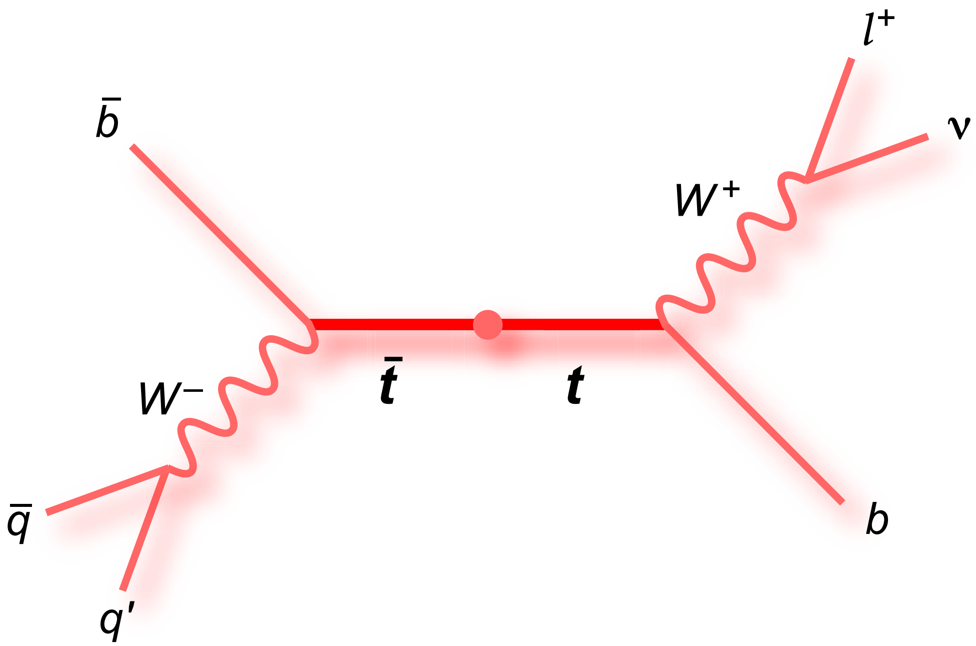
$$\hat{n}(C_i) = \sum_{j=1}^{n_E} M_{ij}\cdot n(E_j)$$

$R_{ij}$ → $P(E_j|C_i)$; $\mu_i^{(0)}$ → $P_o(C_i)$; $x_i^1$ → denominator

$$M_{ij} = \frac{P(E_j|C_i)\cdot P_o(C_i)}{\left[\sum_{l=1}^{n_E} P(E_l|C_i)\right]\cdot\left[\sum_{l=1}^{n_C} P(E_j|C_l)\cdot P_o(C_l)\right]}$$

- **Distinctive**:
  - Smoothen estimated distribution in each iteration step before "unfold" step (not last) : by polynomial fit (user can change it)
  - Continue until solution is stable ($\chi^2$ test with previous iteration )

# Example Iterative unfolding: tt charge asymmetry with ATLAS @ LHC

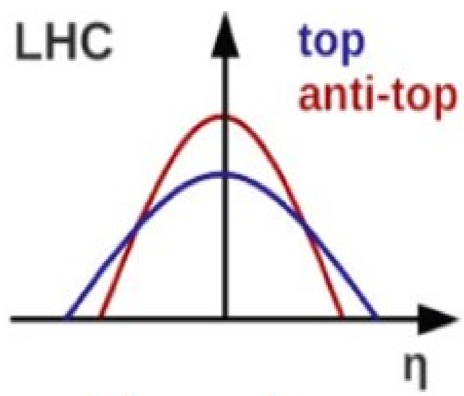



$\int Ldt$ = **1 fb$^{-1}$** *(2011)*

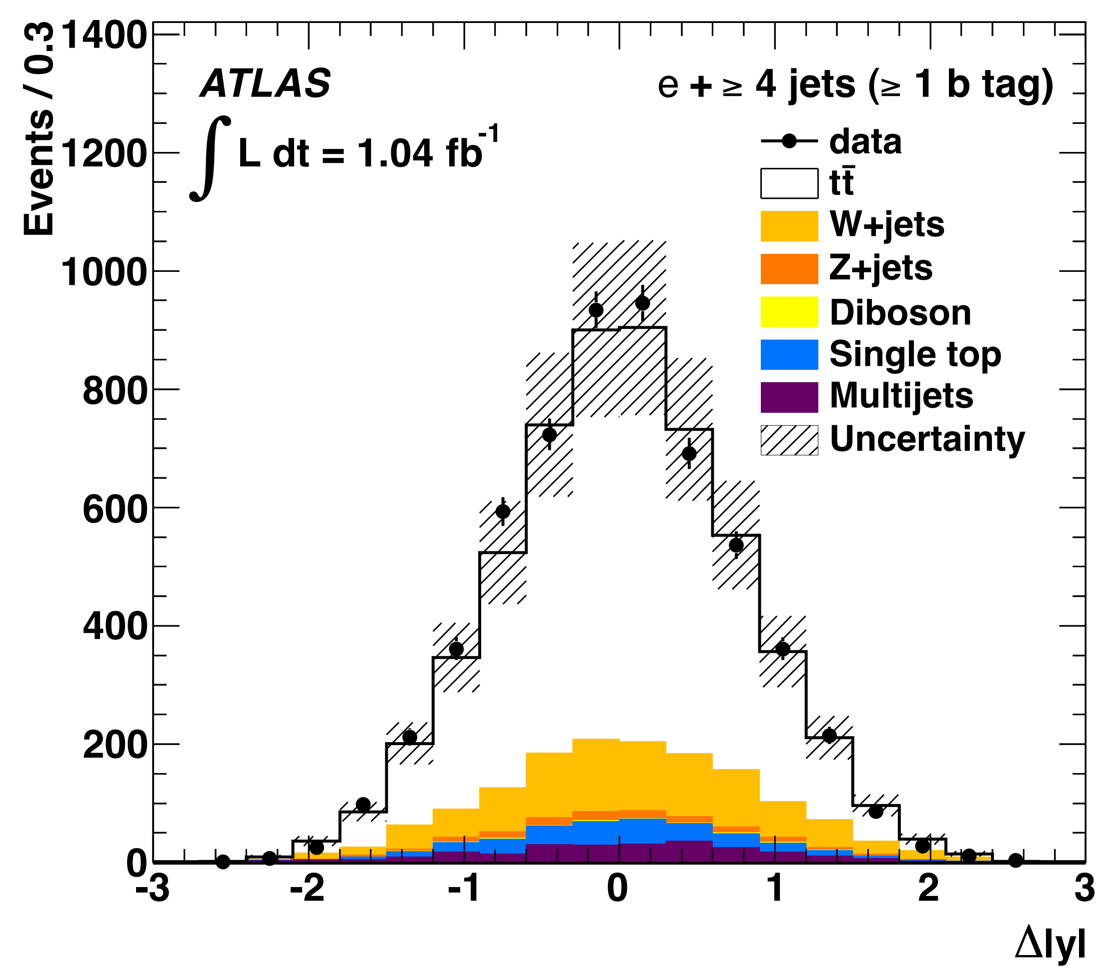
accepted by Eur.Phys.J
30th May2012

arxiv:1203.4211[hep-ex]

- Expect



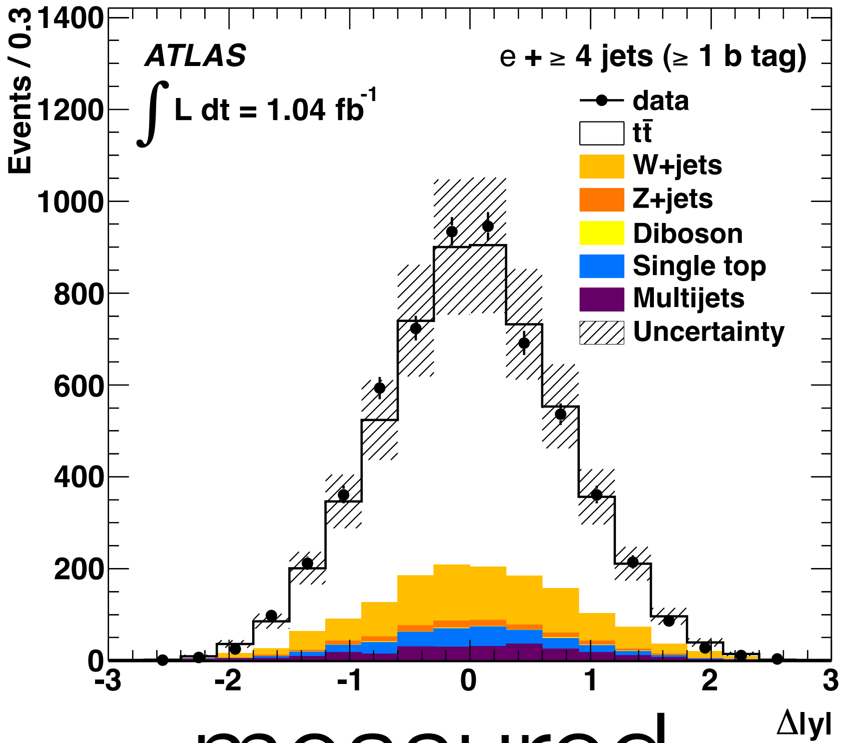

**MC@NLO@ 7TeV LHC predicts $A_c$ = 0.006 +/- 0.002**



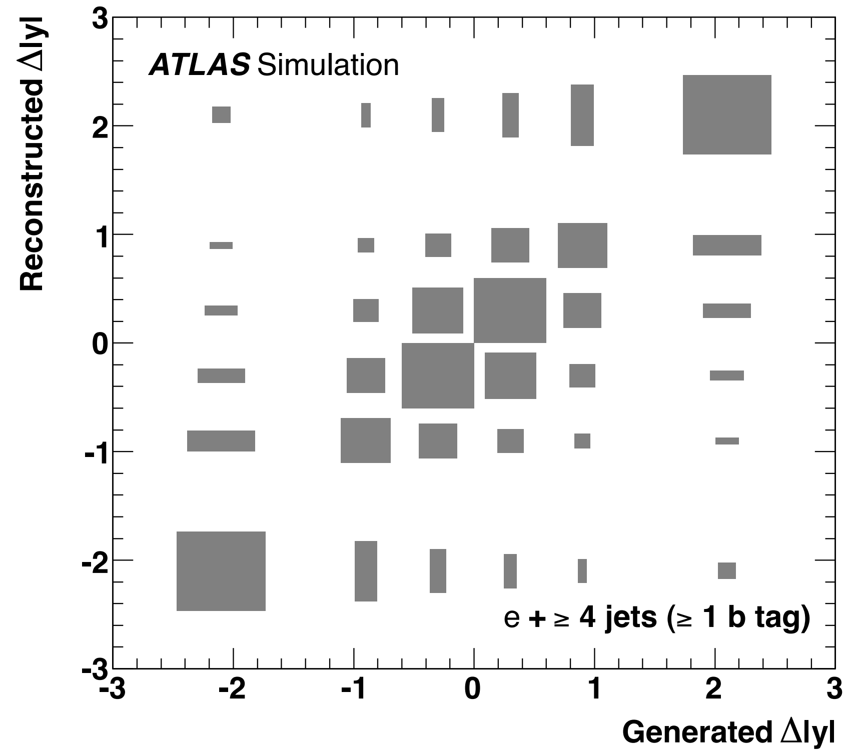

- Reconstruct $t\bar{t}$ and study

$$A_C = \frac{N(\Delta|Y| > 0) - N(\Delta|Y| < 0)}{N(\Delta|Y| > 0) + N(\Delta|Y| < 0)}$$

# Example of iterative unfolding: $t\bar{t}$ charge asymmetry with ATLAS @ LHC




measured

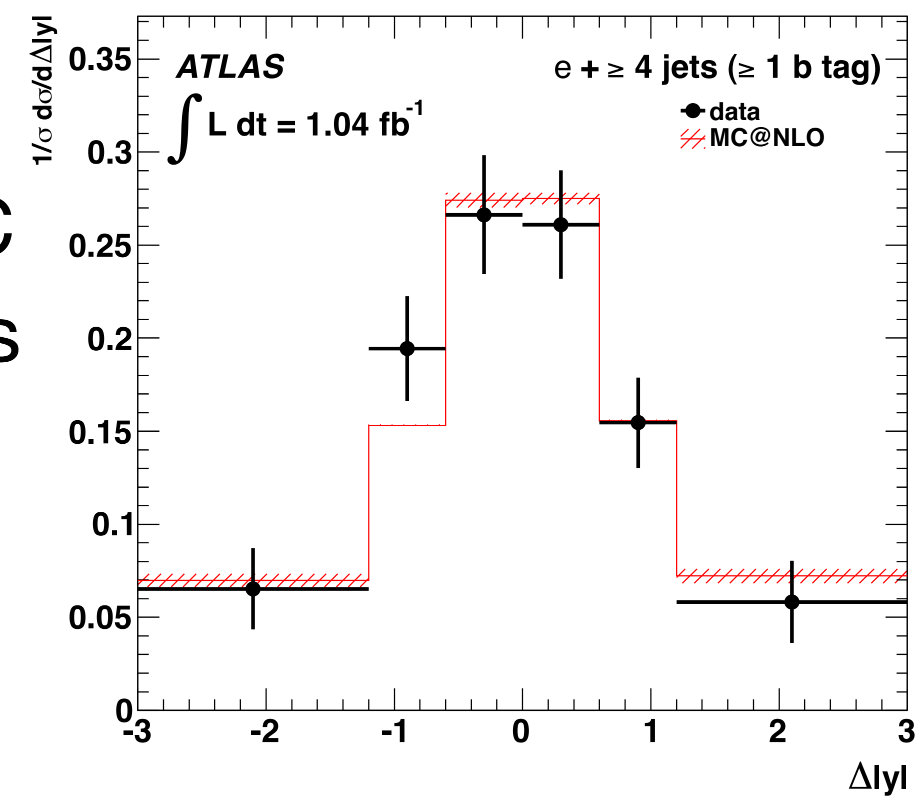



$\int L dt$ = **1 fb$^{-1}$** *(2011)*

accepted by Eur.Phys.J 30th May2012

arxiv:1203.4211[hep-ex]

unfolded




- Stop when $A_C$ changes by less than 0.1% on MC
- Stat uncertainty checked with pseudoexpriments
- Syst uncertainty propagated to response matrix and bkg
- Re-weight $t\bar{t}$ events to vary $A_C$ and check unfolding linearity.

***consistent with SM,*** *main syst: parton shower, top mass , ISR/FSR, jet scale*

$$A_C = -0.018 \pm 0.028\ (\text{stat.}) \pm 0.023\ (\text{syst.})$$

# Regularization with Entropy: the idea

See for instance
M Schmelling ,Nucl. Instr. Meth. A 340 (1994) 400-412

- Shannon's entropy is

$$\mathbf{H} = - \sum_{i=1}^{M} p_i \log p_i$$

  - $p_i$ are equal → maximal smoothness
  - one $p_i$ =1 m all others = 0 → minimum entropy

- Use H as regularization function

$$\mathbf{S(\mu)= H(}\mu\mathbf{)=} \sum_{i=1}^{M} \frac{\mu_i}{\mu_{tot}} \log \frac{\mu_i}{\mu_{tot}}$$

~log (number of ways to arrange $\mu_{tot}$ in M bins)

- Bayesian justification: S is a prior pdf for **μ**

# Regularization with entropy - ARU

Automatic Regularized Unfolding

http://aru.hepforge.org/

H. P. Dembinski, M.Roth in Proc PHYSTATProc2011

- 1 dimensional non parametric unfolding
- Parametrize unfolded distribution as sum of B-spline functions
- Fold it with detector Kernel (calibration, efficiency , resolution..)

$$f(y) = \int K(y,x)b(x)\mathrm{d}x = \sum_j c_j \int K(y,x)b_j(x)\mathrm{d}x = \sum_j c_j f_j(y),$$

- Fit to data with extended ML method: minimize -log L

$$L(\boldsymbol{c}) = L_1(\boldsymbol{c}) + wL_2(\boldsymbol{c})$$

- where L1 is the negative log of the likelihood...

$$L_1(\boldsymbol{c}) = \sum_j c_j F_j - \sum_i \ln f(y_i),$$

- ... and L2 is the regularization term

$$L_2(\boldsymbol{c}) = \int b(x) \ln \frac{b(x)}{g(x)} \mathrm{d}x - \sum_j c_j B_j$$

Choose uniform g

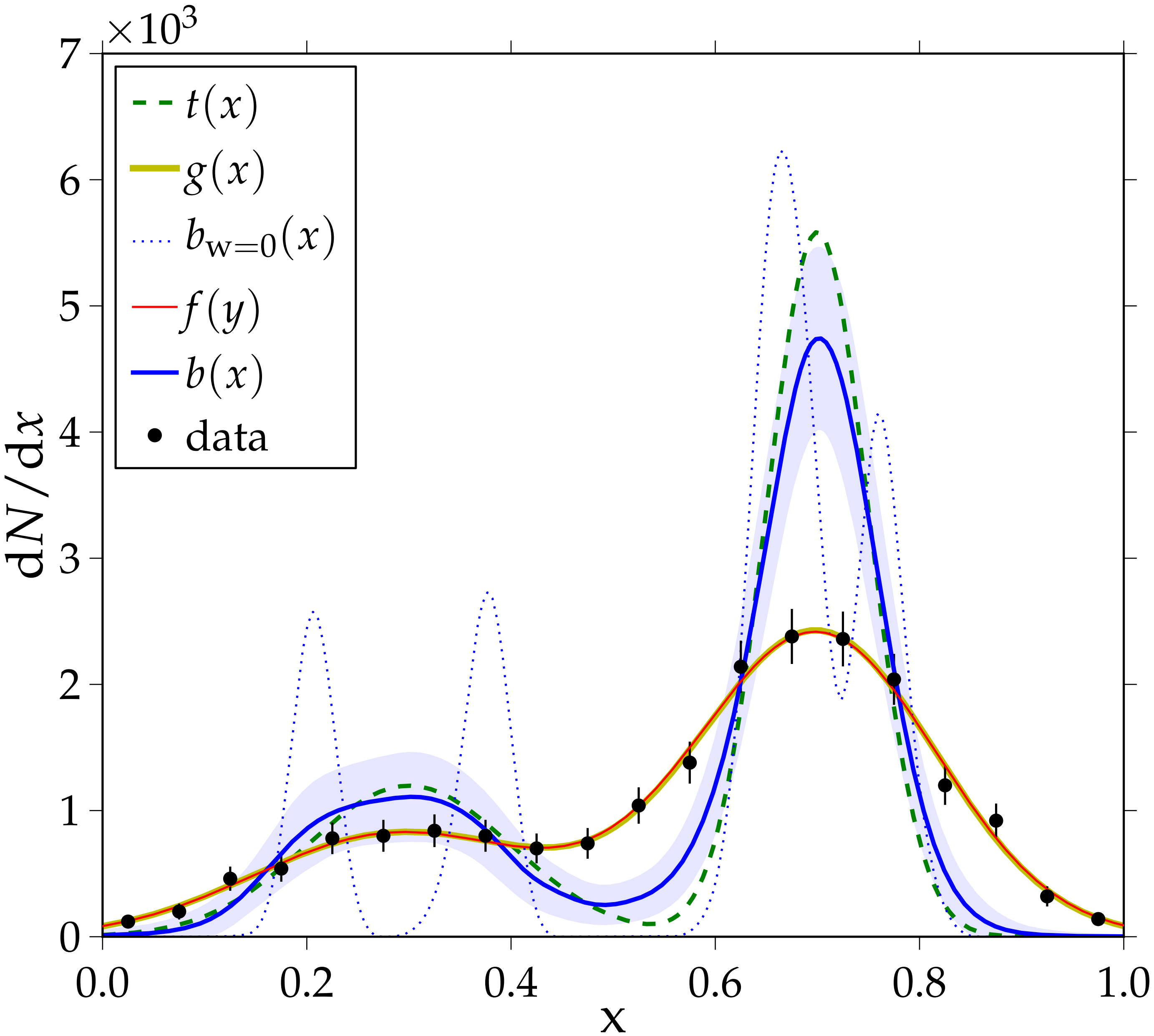
Choose w by minimizing

$$\mathrm{MISE}\big(f(y)\big) = \int \mathrm{d}y\, E\Big[\big(f(y) - f_{\text{true}}(y)\big)^2\Big]$$

$$\int \mathrm{d}y\, \Big\{V[f(y)] + \big(f(y) - f_{\text{true}}(y)\big)^2\Big\}.$$

# Regularization with entropy - ARU

http://aru.hepforge.org/
H. P. Dembinski, M.Roth in
Proc PHYSTATProc2011




- Two Gaussians smeared out with Gaussian kernel
- Perform 2000 pseudo-exp: uncertainty is consistent with stand dev. from ARU

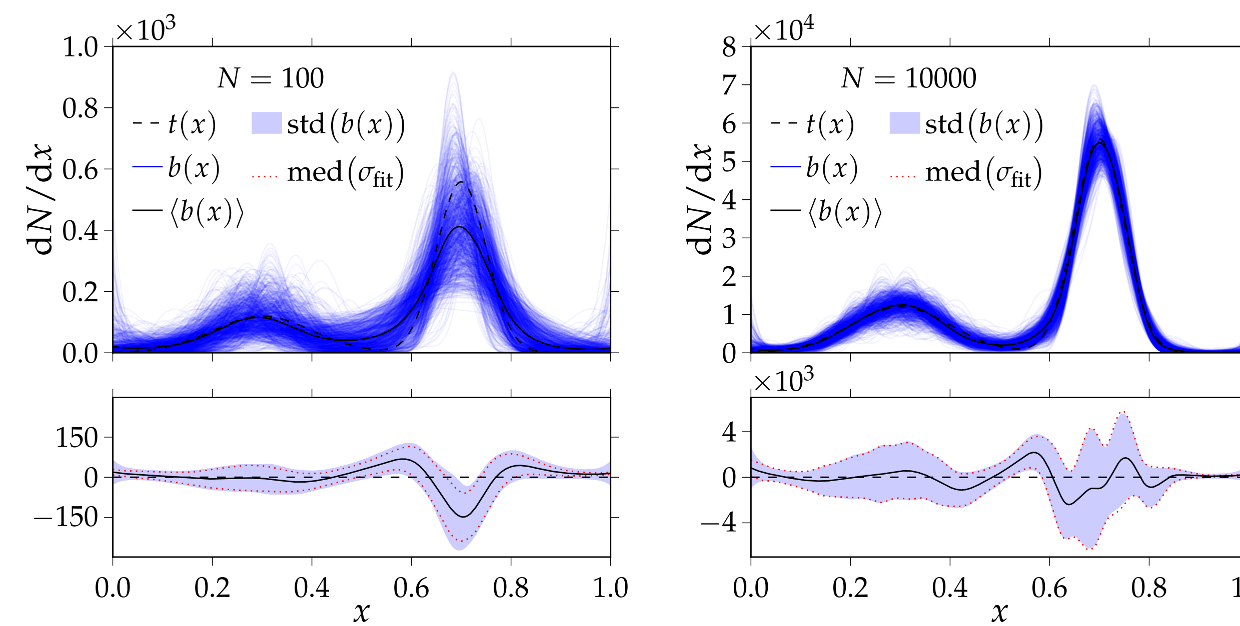



Figure 1: Unfolding of a toy data set of 1000 events. $t(x)$ is the true distribution, the points show a histogram of the smeared data. In this case, the folded solution $f(y)$ is on top of the reference distribution $g(x)$ used for regularization. The regularized solution $b(x)$ shows no undesired oscillations, in contrast to the solution $b_{w=0}(x)$, which is obtained if no regularization is applied.

G. Cowan, Statistical Data Analysis, OUP (1998) Ch. 11

# Choice of regularization par:example criteria

$\alpha = 0 \rightarrow \hat{\vec{\mu}}$ $\qquad \alpha \rightarrow \infty$

- Minimize

$$\mathrm{MSE} = \frac{1}{M} \sum^{M} (U_{ii} + \hat{b}_i^2)$$

$$\mathrm{MSE}' = \frac{1}{M} \sum_{i=1}^{M} \frac{U_{ii} + \hat{b}_i^2}{\hat{\mu}_i}$$

- Consider changes in Chi2 from unregularized solution

$$\Delta\chi^2 = 2\Delta \log L = N$$

- Bias consistent with zero withing its own uncertainty

$$\chi_b^2 = \sum_{i=1}^{M} \frac{\hat{b}_i^2}{W_{ii}} = M \;\; \text{where } W_{ij} = \mathrm{cov}[\hat{b}_i, \hat{b}_j]$$

if bias is non zero, one should correct for it

# Fully Bayesian Unfolding (FBU)

G. Choudalakis
arxiv:1201.4612

- Unfolding question: find Truth spectrum **T** given Data **D** and migration model *P*. Give Bayesian answer

$$p(\mathbf{T}|\mathbf{D} \wedge \mathcal{P}) \propto P(\mathbf{D}|\mathbf{T} \wedge \mathcal{P})\; p(\mathbf{T} \wedge \mathcal{P})$$

$p(\mathbf{T}|\mathbf{D} \wedge \mathcal{P})$ : The posterior p.d.f. of **T**.

$P(\mathbf{D}|\mathbf{T} \wedge \mathcal{P})$ : The likelihood of **D**, as a function of **T** and $\mathcal{P}$

$p(\mathbf{T} \wedge \mathcal{P})$ : The prior p.d.f. of **T** and $\mathcal{P}$.

**T**: The truth-level binned spectrum. $\mathbf{T} \in \mathbb{R}^{N_t}$.

**D**: The observed binned spectrum; $\mathbf{D} \in \mathbb{N}^{N_r}$, if Poisson.

$\mathcal{P}$: The conditional migrations matrix: $\mathcal{P}_{t,r} \equiv P(r|t) = P_{t\to r}$.

computed from the migrations matrix, $\mathcal{M}_{tr} \equiv P(t,r)$ its efficiency, $\epsilon_t \equiv \frac{\sum_r P(r|t)}{P(t)}$

**Result is posterior pdf** p(**T**|**D**, P) defined in space of possible spectra (**not estimator and variance**).

# FBU: general ides

G. Choudalakis
arxiv:1201.4612

- For Likelihood one can choose Poisson

$$P(\mathbf{D}|\mathbf{T}) \quad = \quad \prod_{r-1}^{N_r} \text{Poisson}(D_r|\mathbf{T}) \quad = \quad \prod_{r=1}^{N_r} \frac{R_r^{D_r}}{D_r!} e^{-R_r}$$

with

$$R_r = \sum_{t=1}^{N_t} T_t \, P_{t \to r} = \sum_{t=1}^{N_t} T_t \, \frac{\mathcal{M}_{t,r}}{\epsilon_t^{-1} \sum_{k=1}^{N_r} \mathcal{M}_{t,k}}$$

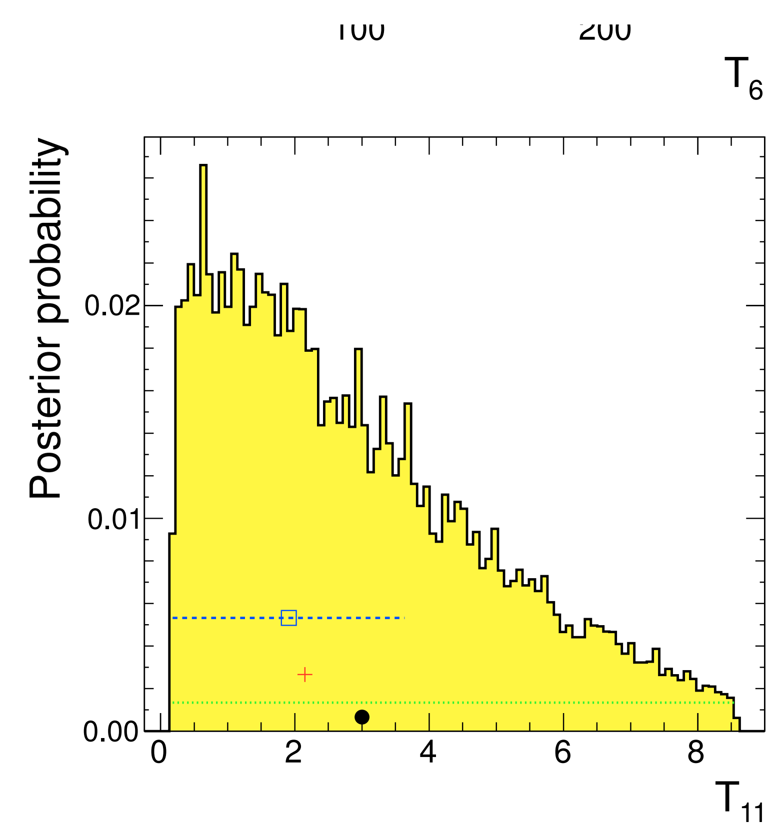
- Regularization in standard form

$$P(\mathbf{D}|\mathbf{T}) \cdot e^{-\alpha \, S(\mathbf{T})}$$

For FBU that role is played by the Prior

$$p(\mathbf{T}) = e^{-\alpha \, S(\mathbf{T})}$$

- Name of the game: integral calculation



# Examples of other unfolding schemes

- IDS: iterative dynamically stabilized, B. Malaescu, arxiv: 0907.3791 [phys.data-an]
  - used in ATLAS paper http://arxiv.org/abs/1112.6297

- Binning free Iterative Deconvolution, Lindemann, Zech, Nucl.Instr. Meth A 354 (1995) 516-521

- ***Satellite Method***, see G. Bohm and G. Zech, *Introduction to Statistics and Data Analysis for Physicists*, Verlag Deutsches Elektronen-Synchrotron (2010), available at http://www-library.desy.de/elbook.html

- ***SPlot***, M Pivk, F. Le Diberder, arXiv:physics/0402083v

by no means exhaustive (more in Nucl. Instr. Meth for instance)

# 2 cents on optimization/choice of technique

- Choices  strongly analysis-dependent

- Always consider/produce/report un-regularized solution
  - no bias form unfolding. Powerful to test a theory using **full covariance** matrix

$$\chi^2(\boldsymbol{\theta}) = (\boldsymbol{\mu}(\boldsymbol{\theta}) - \hat{\boldsymbol{\mu}}_{\mathrm{ML}})^T U_{\mathrm{stat}}^{-1} (\boldsymbol{\mu}(\boldsymbol{\theta}) - \hat{\boldsymbol{\mu}}_{\mathrm{ML}})$$

  - consider SVD decomposition diagnostic & condition number for response matrix, also in the light of syst uncertainties

- Carefully consider the possible impact of the regularization on your analysis
  - Can I afford to suppress bumps/large curvature?

- If  regularizing a discrete estimator, choose bins using full stat and systematic uncertainty andfully propagate in analysis on simulated data (your best prediction)

# 2 cents on systematic uncertainties

- Include syst in your analysis: vary all elements in LKL according to their dependence on syst
  - response matrix, bkg

- Possible inclusion of syst: use pseudoexpriments with given priors/hypothesis for distribution or resulting from ancillary measurement: take into account correlations induced by unfolding
  - for instance hybrid bayesian: marginalize max of lkl with pseudo exp

- Crucial to devise tests for stability and bias
  - stress unfolding response with distorted shapes / varying parameter of interest in simulated events: unfold folded test distributions to check for bias, compare with overall expected syst+stat uncertainty or use $\chi^2$ with model

# Additional references

- G Cowan, Lecture 4, CERN academic lectures, available at http://indico.cern.ch/conferenceDisplay.py?confId=173729

- V Blobel, in CSC84, CERN-85-09

- PHYSTAT 2011 proceedings available at
  - Agenda :http://indico.cern.ch/conferenceDisplay.py?confId=107747
  - File: http://cdsweb.cern.ch/record/1306523/files/CERN-2011-006.pdf

- G. Bohm and G. Zech, *Introduction to Statistics and Data Analysis for Physicists*, Verlag Deutsches Elektronen-Synchrotron (2010), available at http://www-library.desy.de/elbook.html.

## Tools and repositories

- The Unfolding Framework project at

  https://www.wiki.terascale.de/index.php/Unfolding_Framework_Project

- RooUnfold by T. Adye