Unfolding

Two parts:

1. Critical review of 3 unfolding methods:
   regularization and error assignment
2. Binning-free unfolding

Some initial remarks:

The purpose of a measurment is

- to establish or reject a theory
- or to provide input parameters for secondary applications.

It should be possible to combine measurements.

Measurements without error limits are useless.

Results should be published in such a way that they do not require extensive additional explications in the text, or even worse, in Refs.

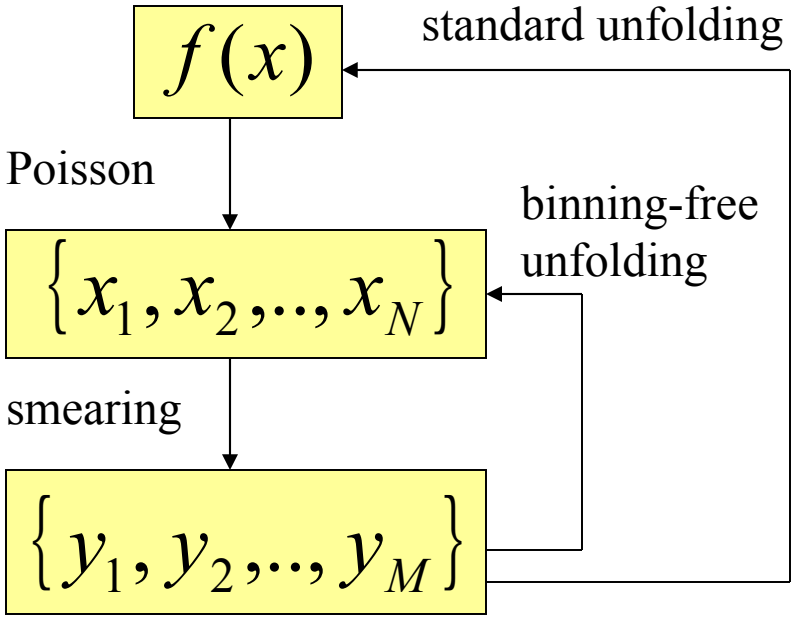
These requirement also apply to unfolding results

The problem

unknown true distribution

true data sample

observed data sample



We got y, we want to know f(x)

Situation A: f(x) is known up to unknown parameter values $f(x|\theta)$ (vast majority of cases: mass spectra, lifetime measurements, etc.)

solution: fit λ, standard techniques: maximum likelihood, χ^2, re-weight MC
(see G.B.+G.Z. http:/www-library.desy.de/elbook.html, page 157)

Situation B: f(x) is completely unknown

solution: parametrize f(x): histogram (or higher splines), parameters are the content of the true histogram bins
→ back to situation A
→ Poisson maximum likelihood or χ^2 fit

what is special? many parameters (order 100), strongly correlated, some smoothing inherent due to binning.

The parameters and the error matrix provide the necessary information for comparisons with theoetical predictions or other experiments and for combining measurements and should be published.

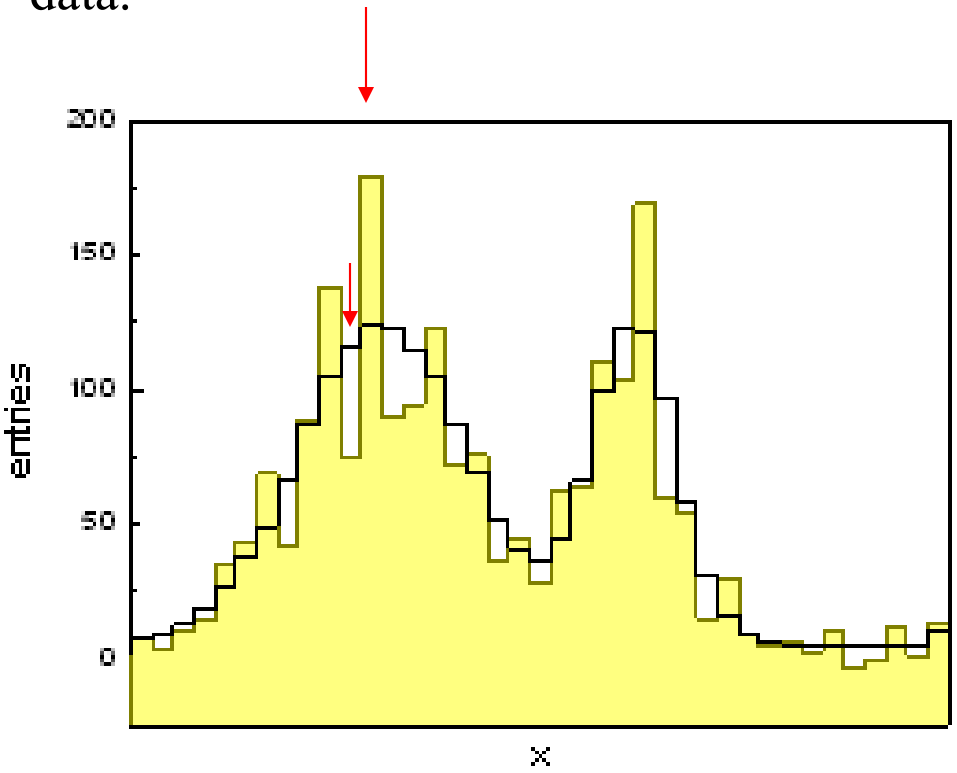
There are no major technical problems to compute these items.

Why are most physicists not satisfied with this solution?

Due to the correlations, it is not suited for a graphical representation! There are strong fluctuations from one bin to the next one.

Adjacent bins of the unfolded histogram are strongly correlated.

We could make them more equal without changing the likelihood or χ^2 by much. → Regularization
The corrected histogram is smooth and well compatible with the data.



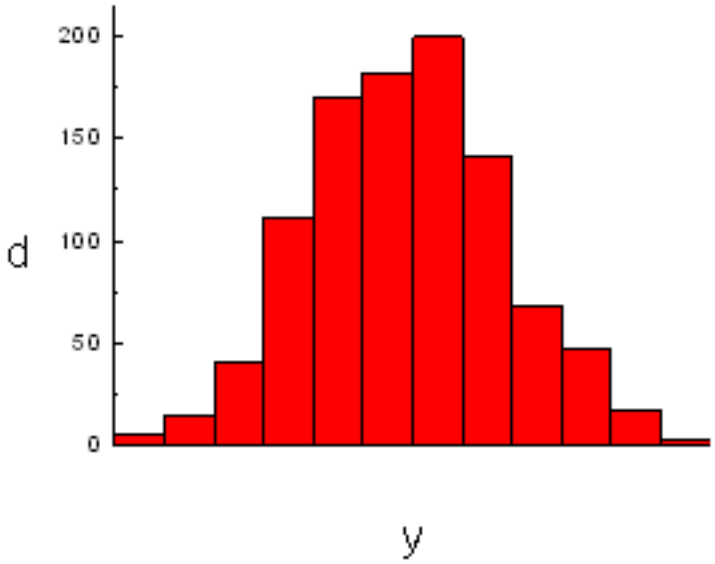
black: original distribution

yellow: unfolding result without regularization

The regularization scheme may be based on / include :
- the tranfer matrix
- the data error matrix
- the shape of the unfolded distribution

Notation

observed histogram = data vector



$$d = \begin{pmatrix} 6 \\ 15 \\ 41 \\ 111 \\ 169 \\ 182 \\ 199 \\ 141 \\ 68 \\ 48 \\ 17 \\ 3 \end{pmatrix}$$

d: (observed) data vector

ϑ: true distribution vector

T: transfer matrix (from MC)

$$d \cong T\vartheta$$

Regularization methods:

Method 1: Least square matrix inversion:

orthogonalize parameters, exclude non-significant parameters

- The regularization depends on measurement errors and transfer matrix
- The truncation (the regularization strength) is in most implementations somewhat arbitrary or not consistent with the standard error assignment.

We simplify, to show the principle:
same number of bins in true and observed distribution (T quadratic)

$$d = T\vartheta \quad \longrightarrow \quad \vartheta = T^{-1}d$$

Simple matrix inversion → oscillations. To understand this look at eigenvectors u_i of T and eigenvalue λ_i

Get eigenvectors of T: u_i

$$\lambda_i u_i = T u_i$$

A true distribution u_i produces an observed distribution v_i of the same shape but multiplied by the eigenvalue λ_i

$$v_i = \lambda_i u_i = T u_i$$

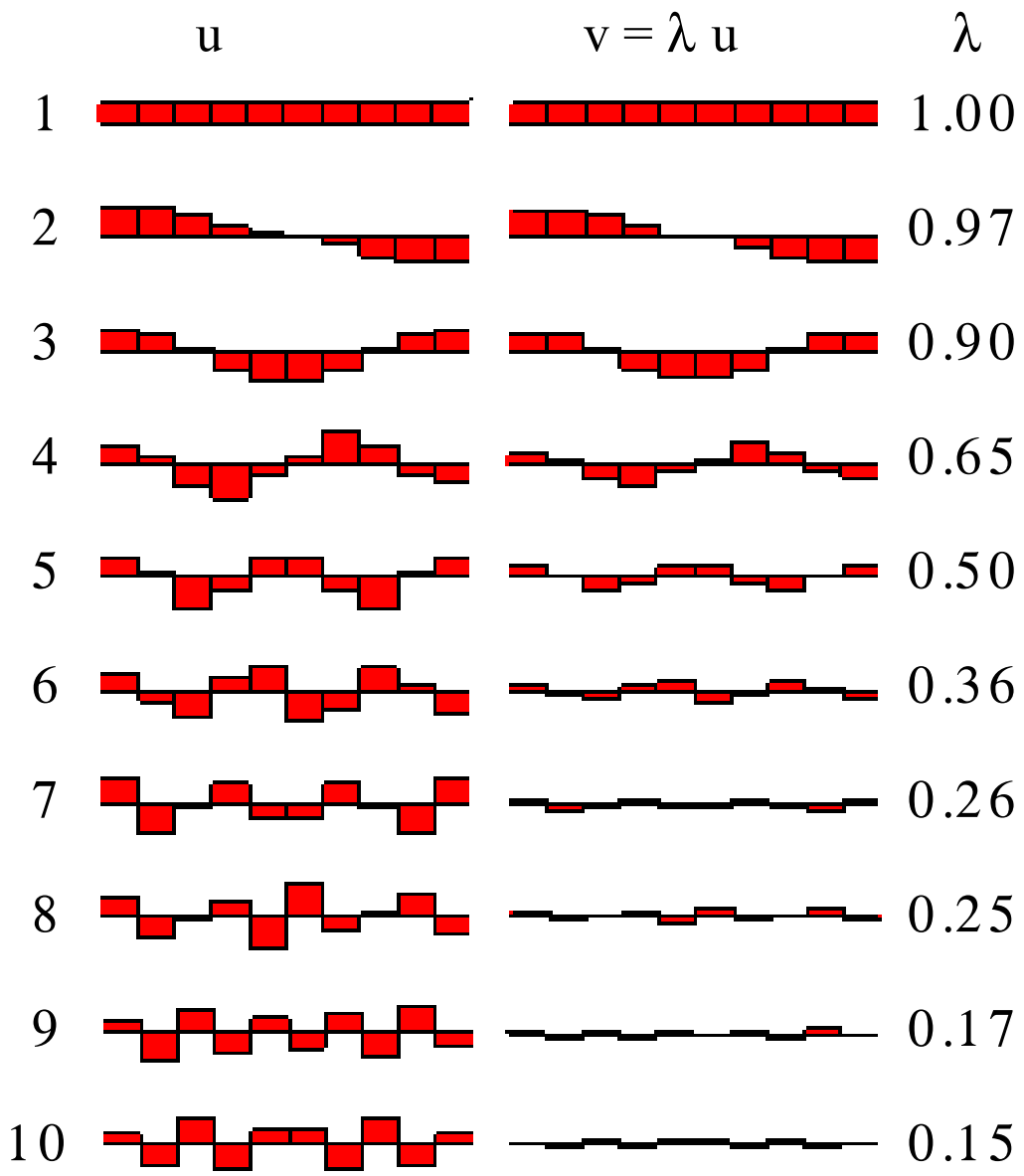
$$u_i = T^{-1} v_i = \frac{1}{\lambda_i} v_i$$

Adding v_i to the observed data → adding u_i to the unfolded distribution.
Small values of λ lead to large contributions!

σ
μ
α
λ

Example: 10 x 10 transfer matrix

.6	.3	.1							
.3	.5	.2							
.1	.2	.5	.2				0		
		.2	.5	.2	.1				
			.2	.5	.2	.1			
			.1	.2	.5	.2			
				.1	.2	.5	.2		
		0				.2	.5	.2	.1
							.2	.5	.3
							.1	.3	.6



The true distribution is a superposition of the eigenvectors u.

The eigenvectors are orthogonal.

A contribution v to the data
→ contribution u to the result.

Eliminate contributions 5 to 10 with small eigenvalues. (v compatible with zero) The unfolded distribution is then a superposition of u_1 to u_5

(or reduce their contribution)

One should use less bins for the true histogram than for the observed data.
→T is rectangular.

The least square fit has to include the error matrix E of the data vector d.
(weight matrix $V=E^{-1}$) . We transform to a quadratic matrix C:

$$T \Rightarrow C = (T^T V)T \qquad d \Rightarrow b = (T^T V)d$$

$$C\vartheta = b \qquad \vartheta = C^{-1}b$$

Eliminate small eigenvalues of C,

(see V. Blobel's talk at the Desy workshop on unfolding)

Likelihood (or χ^2) fit with curvature regularization

- The result depends on measurement errors, transfer matrix and data vector (smoothing even with diagonal transfer matrix)
- The regularization strength can be controlled by chi-squared.

$$\ln L = -\sum_k \left[d_k \sum_j T_{kj} \vartheta_j - \sum_j T_{kj} \vartheta_j \right] - R_{regu}(\vartheta)$$

$$\chi^2 = \sum_k \frac{\left(d_k - \sum_j T_{kj} \vartheta_j \right)^2}{\delta_k^2} + R_{regu}(\vartheta)$$

d_k: observed histogram, δ_k error

ϑ : true histogram to be fitted

T: transfer matrix from Monte Carlo, R: penalty for curvature

Standard: curvature regularization → prefers a linear distribution.

$$R_{regu} = r(2\vartheta_i - \vartheta_{i-1} - \vartheta_{i+1})^2$$

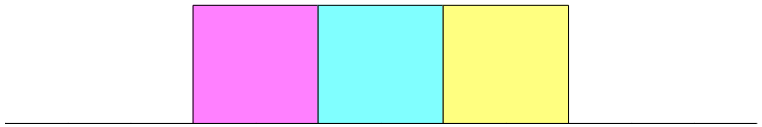
r: regularisation strength

You can use a different penalty term and apply for instance the regularization to the deviation of the fitted from the expected shape of the distribution (iterate).

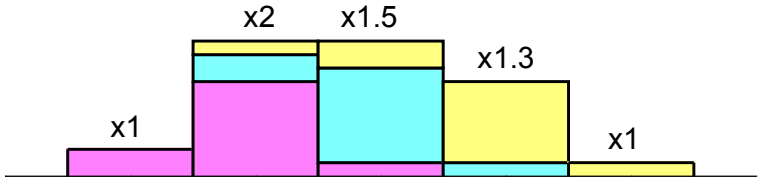
Iterative adjustment (Últai et al. (1986) , D‘Agostini (1994))

- The regularization depends on transfer matrix and presumed true distribution, independent from measurement errors
- A chi-squared control of regularization strength ispossible

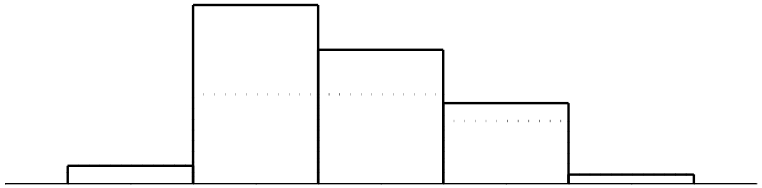
a) starting true distribution



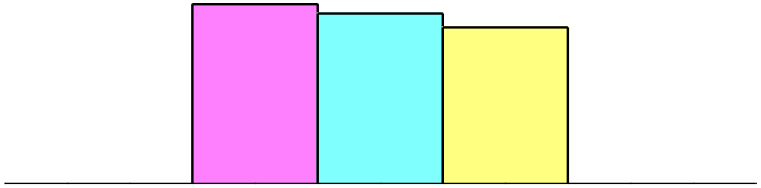
b) corresponding folded distribution.



c) observed data distribution

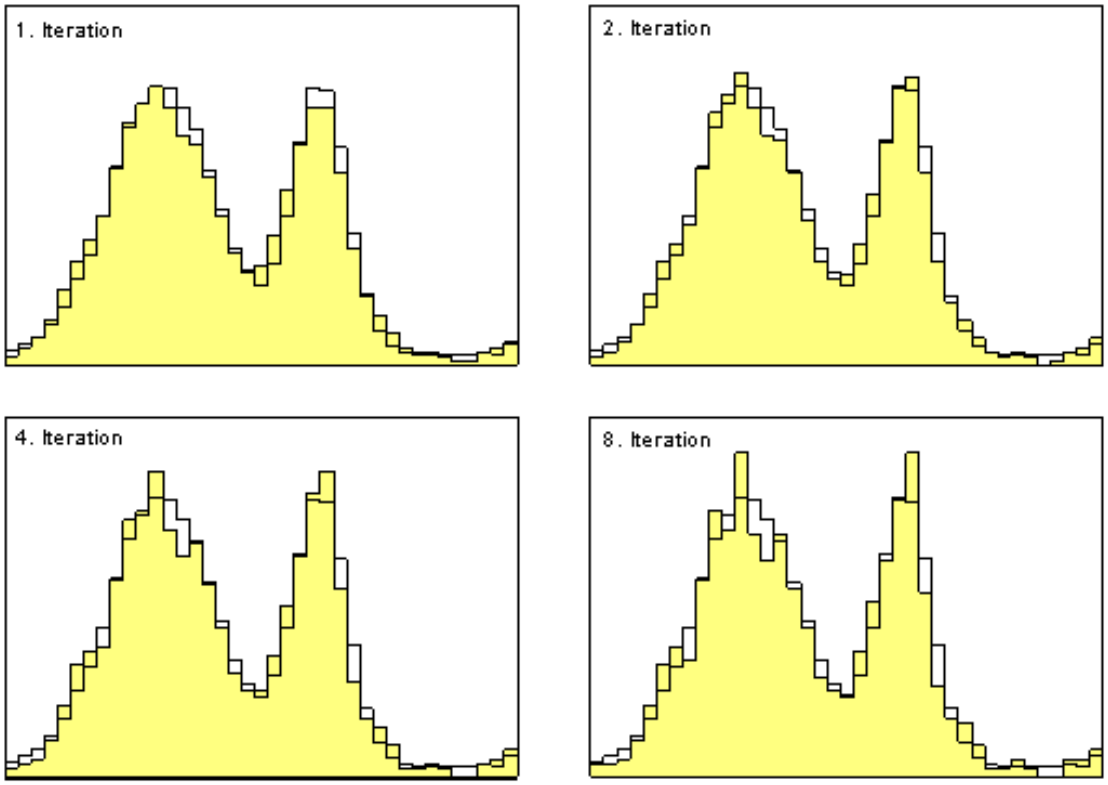


d) 1. Iteration



Multiply each contribution by a factor to get aggreement with the observed distribution and put it back to the true distribution

Example



Empty histogram: original distribution

yellow: unfolding result

At the 4-th iteration already oscillations appear → stop after the first or second iteration

How to fix the increase in χ^2 caused by the regularization

- compute χ^2 without regularization (ML fit) → χ^2_0

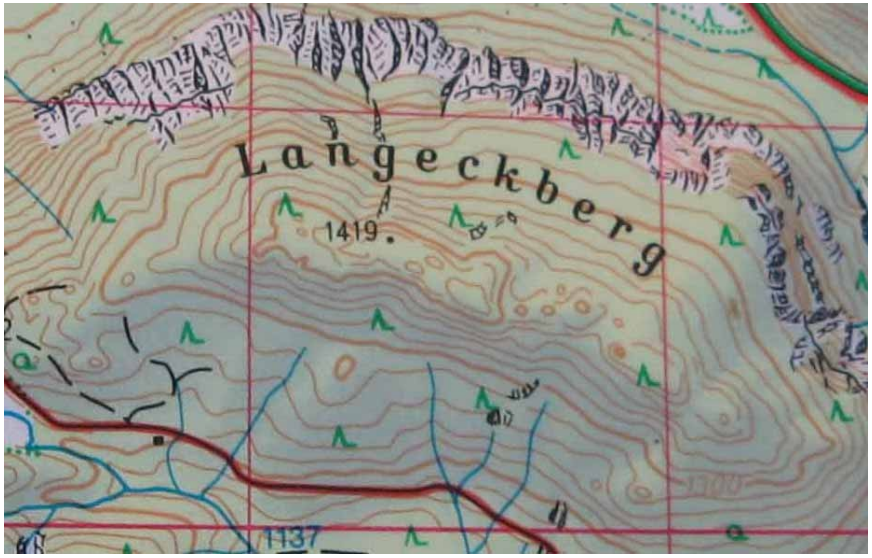
- compute χ^2 with regularization → χ^2_r

- p-value: $p(\Delta\ \chi^2) = 90\ \%$, → $\chi^2_r = \chi^2_0 + \Delta\chi^2$,

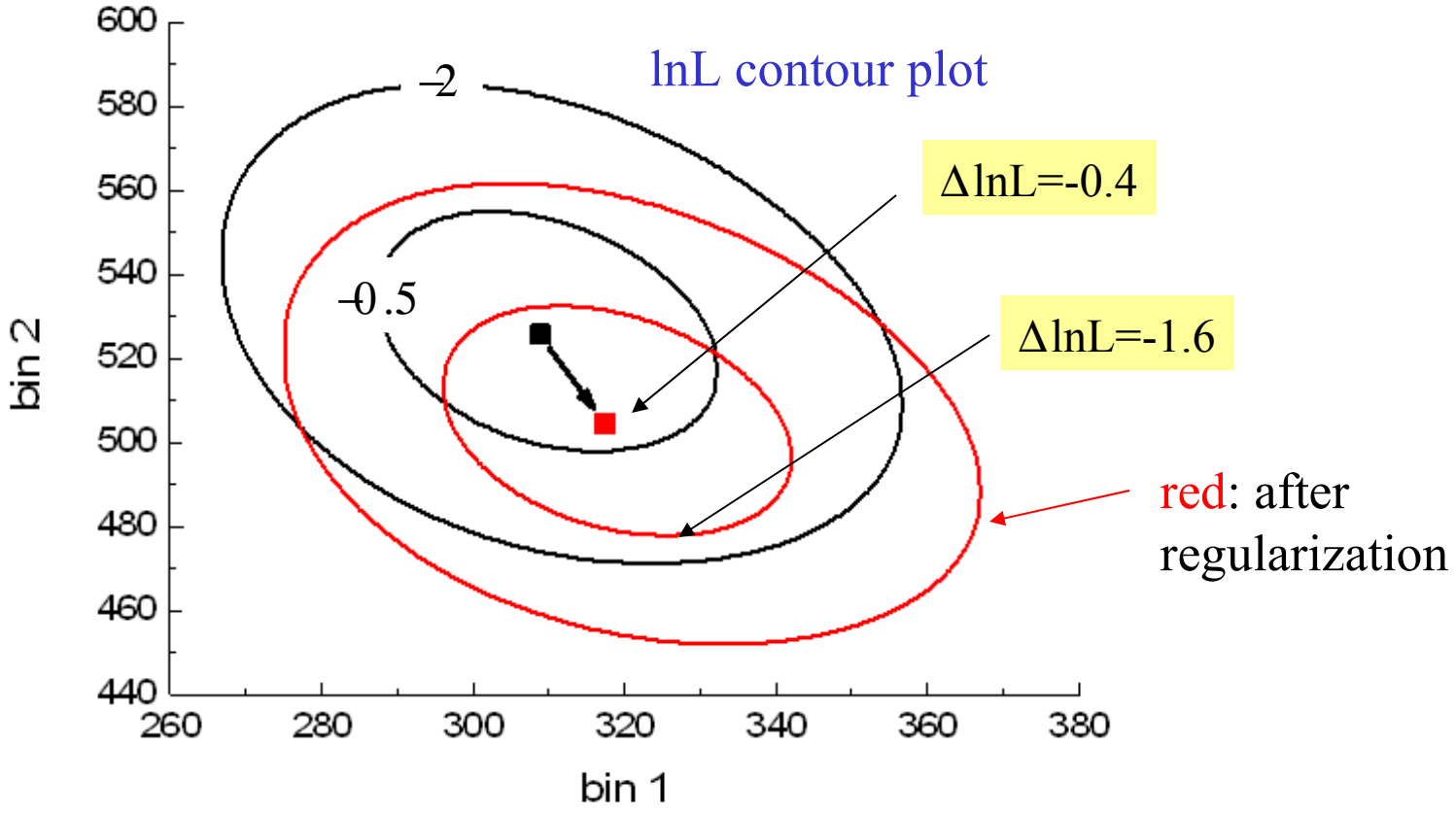
  (degrees of freedom = number of data bins)

Regularization and error

Assume somebody measures the altitude profile of a soft mountain region with a GPS device. He determines the location of the peak and of contour lines. He will not be able to fix location of the highes point with good precision, but the position of the contour lines where the slope is steep will be rather well determined.



Error contour of two adjacent bin contents with and without soft regularization. The shifted maximum (to make the distribution smooth) is well compatible with the measurement, but the new error contours are not! Regularization moves the whole mountain!



Standard implementations apply regularization and justify this by the compatibility of the modified distribution with the observed data.

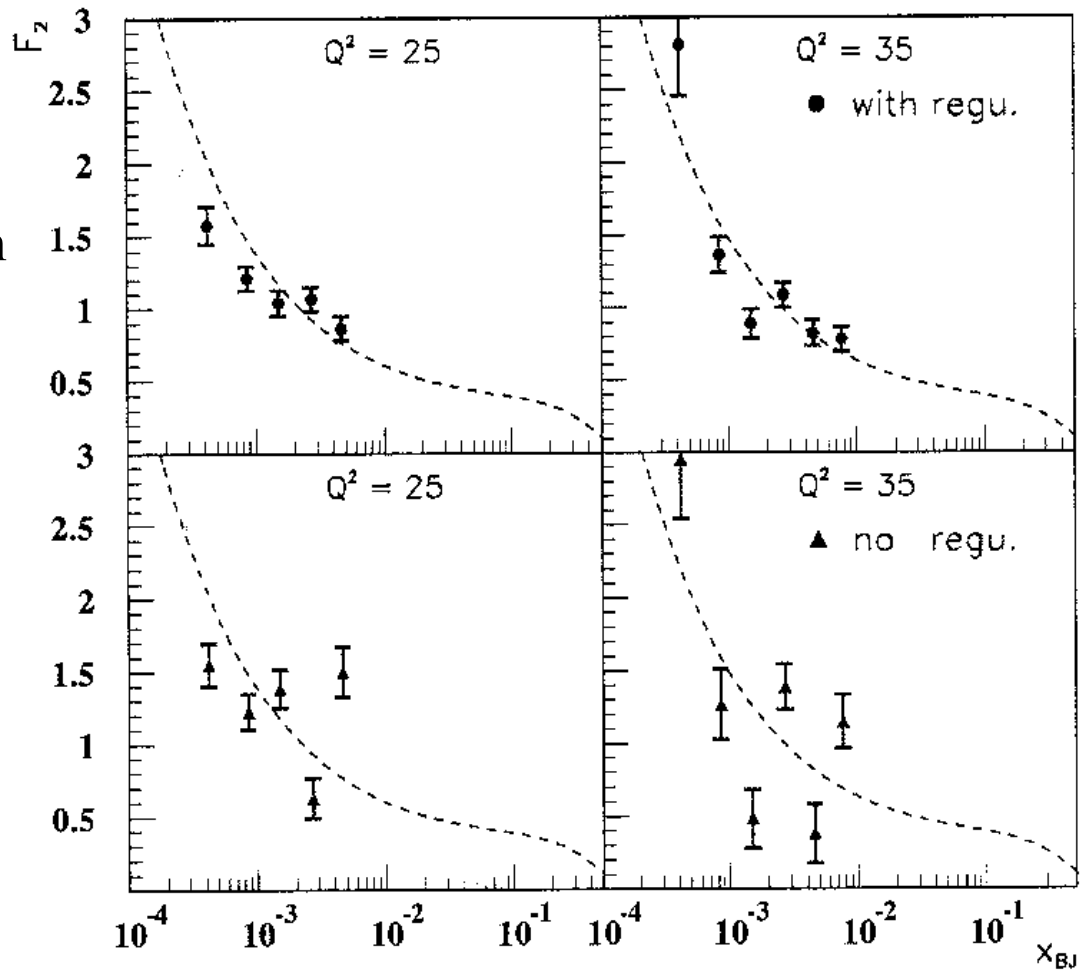
However they silently apply error propagation and produce errors depending on the regularization strength (reduced errors and shifted contours) which are not justified by the observations.

The point estimate is ok, but not the interval estimate!

Top and bottom are the same data. The constraint due to regularization reduces the errors. The error assignment makes no sense.

with regularization

no regularization



Summary and conclusions

1. Regularization moves the raw error contours and may exclude admissible distributions.
2. → Publish results without regularization, include the full error matrix.
3. Use chi-square probability (90 %, soft regularization) to fix the regularization strength. Avoid arbitrary choices.
4. Unfolded bins should have large containment (> 70 %).
5. There is no objective way to reject a certain regularization scheme. For instance, starting with a certain distribution in the iterative unfolding is perfectly ok.
6. D'Agostini's way to stop the iteration is arbitrary.

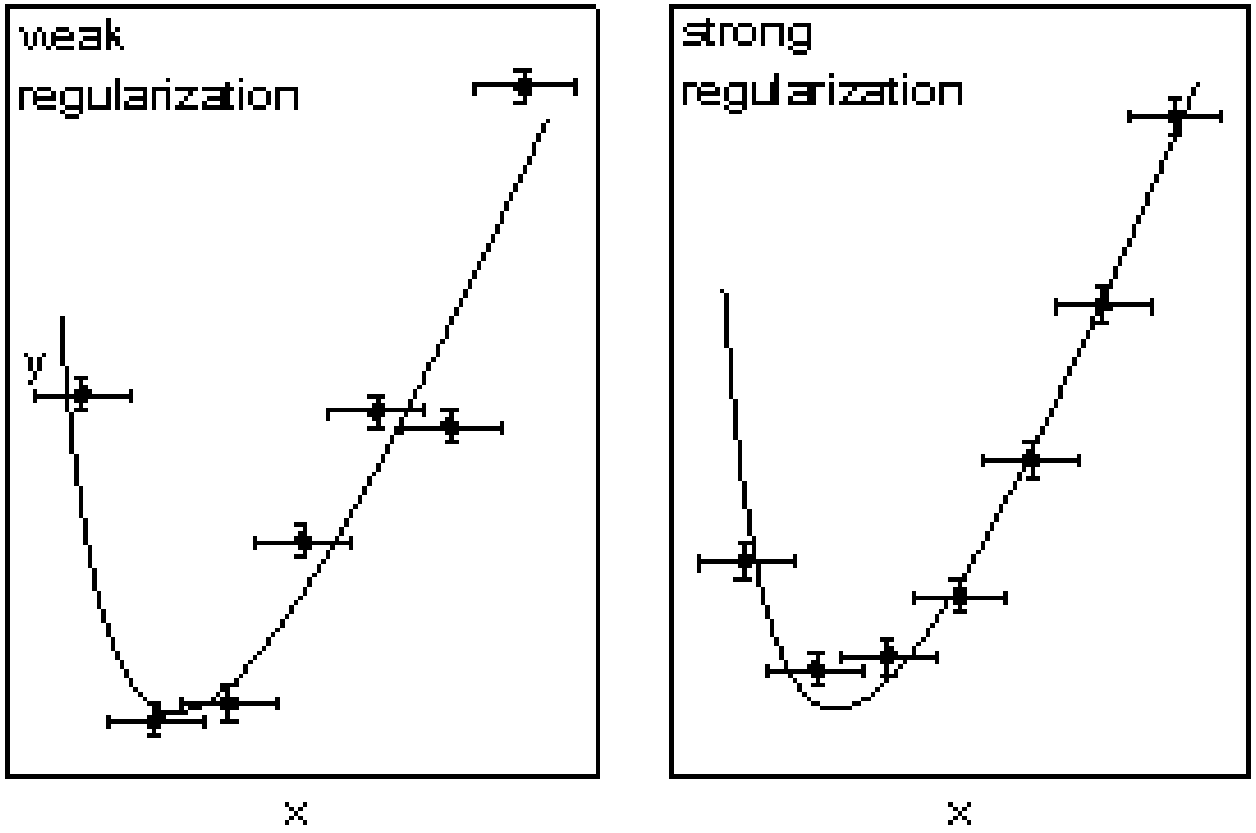
7. The 3 approaches differ in what enters into the regularization: All 3 depend of course on the transfer matrix
   + error matrix of data vector (matrix least square method)
   + first guess of true distribution (iterative unfolding)
   + error matrix of data vector + shape of fitted distribution (likelihood with curvature regularization)
8. My prefered method is the likelihood fit with curvature regularization.

Some further remarks

- Check effect from uncertainty in the resolution (How good is the Monte Carlo simulation?)
- Check effect of MC input distribution, used to get T. If necessary, iterate (re-weighting).
- Try to avoid a contribution to the statistical error from MC (→ difficult error estimate)
- Do not use programs off-the-shelf unless you have understood the method and share the underlying philosophy.
- It is not sufficient to show that the results are reasonable for a certain selection of examples.
- Unbiasedness is not an essential criterion. The quotient $\theta/(\delta\theta)^2$ has to be unbiased.

A posssible way to present the unfoded data

Proposal: vertical bar → statistical from # of events,
horizontal bar → resolution



Binning-free Unfolding

Günter Zech, Desy May 2010

Observed event → true location

- motivation
- basic idea: likelihood, analytic resolution function
- how to find the minimum
- regularization
- results
- include Monte Carlo resolution
- further steps

Motivation

There are situations where binned unfolding suffers from serious difficulties:

- low statistics (for example 40 events)
- events located on unknown curves or points (astronomy)
- multi-dimensional distributions (structure functions)
  (imagine 1000 events, 3 dimensions, 5 bins each
  → 125 bins and in average only 8 events per bin

Advantages:

- apply cuts after unfolding
- define histogram parameters after unfolding
- define histogram variables after unfolding
  (unfold p_x, p_y, plot E)
- consistent histograms of projections

Basic idea

As in parameter fitting, apply single event likelihood

$$\ln L(x_1,..,x_N) = \sum_{i=1}^{N} \ln \sum_{j=1}^{N} f(x_i' | x_j)$$

Notation:

- analytic resolution function $f(x', x_i)$
- True location of point i: x_i (free parameters in the fit. 10000 events, 2 dimensions → 20000 parameters
- Observed location x_i'

(For simplicity written in 1 dimension, but all variables could be vectors)

Minimum search

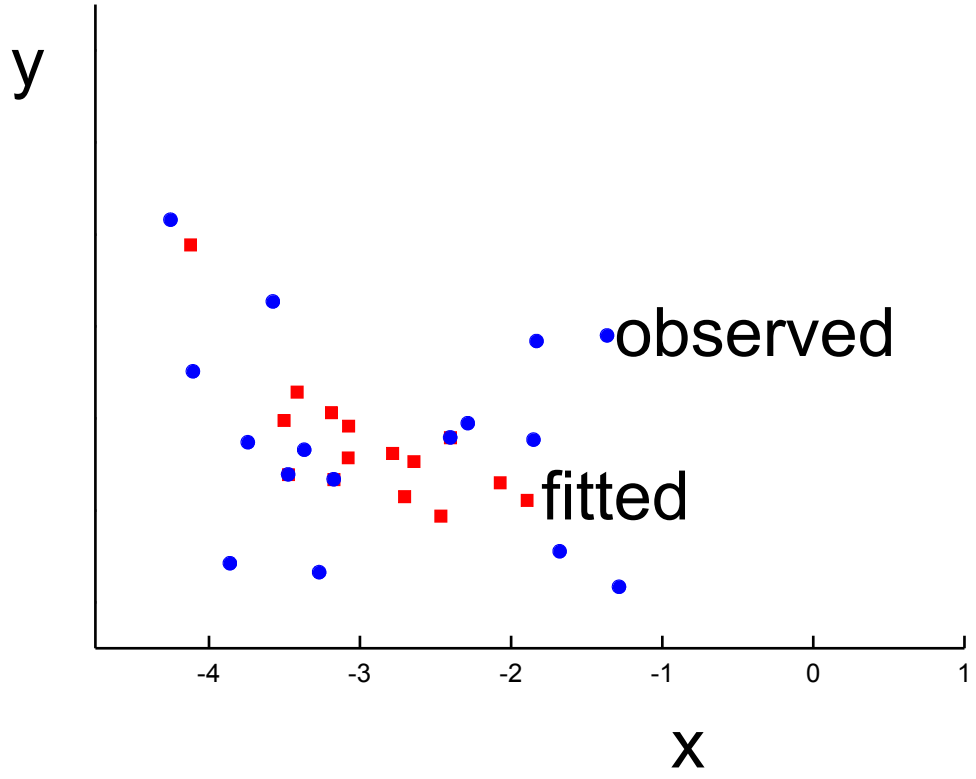
We have a huge number of parameters, but:

- it is easy to select good starting values
- the minima of –lnL are rather shallow
- no dangerous local minima

Minimum search by random migration:

- select randomly a true point move by random step according to a uniform distribution
- accept if the likelihood increases
- repeat until result is satisfying

At the beginning, the true points (red) were sitting on top of the observed data points (blue). They move in such a way that the likelihood increases.



Regularization

Two possibilities, either

1. Stop migration process, or
2. Curvature regularization by probability density estimation using side bands

$$R = r\,\frac{(2n_C - n_L - n_R)^2}{n_C + n_L + n_R} \qquad \ln L = \ln L_{stat} - R$$

Correspondingly in higher dimensions

r: regularization constant

n_c: number of events in central region

n_L, n_R: number of events in left and right hand side bands

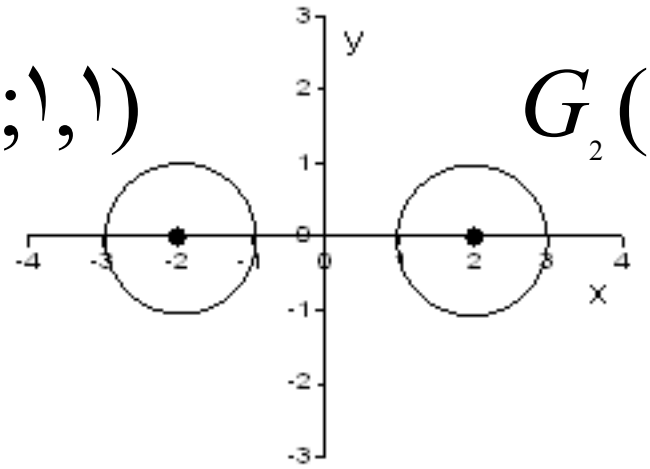
Some Results

Generate 2 Gaussians at x = -2, y = 0 and at x = +2, y = 0 and widths $s_x = 1$, $s_y = 1$ for both

fraction left: 0.6, fraction right: 0.4
1000 events

Gaussian smearing with width s = 1

$$G_1(x, y \mid -2, 0; 1, 1)$$



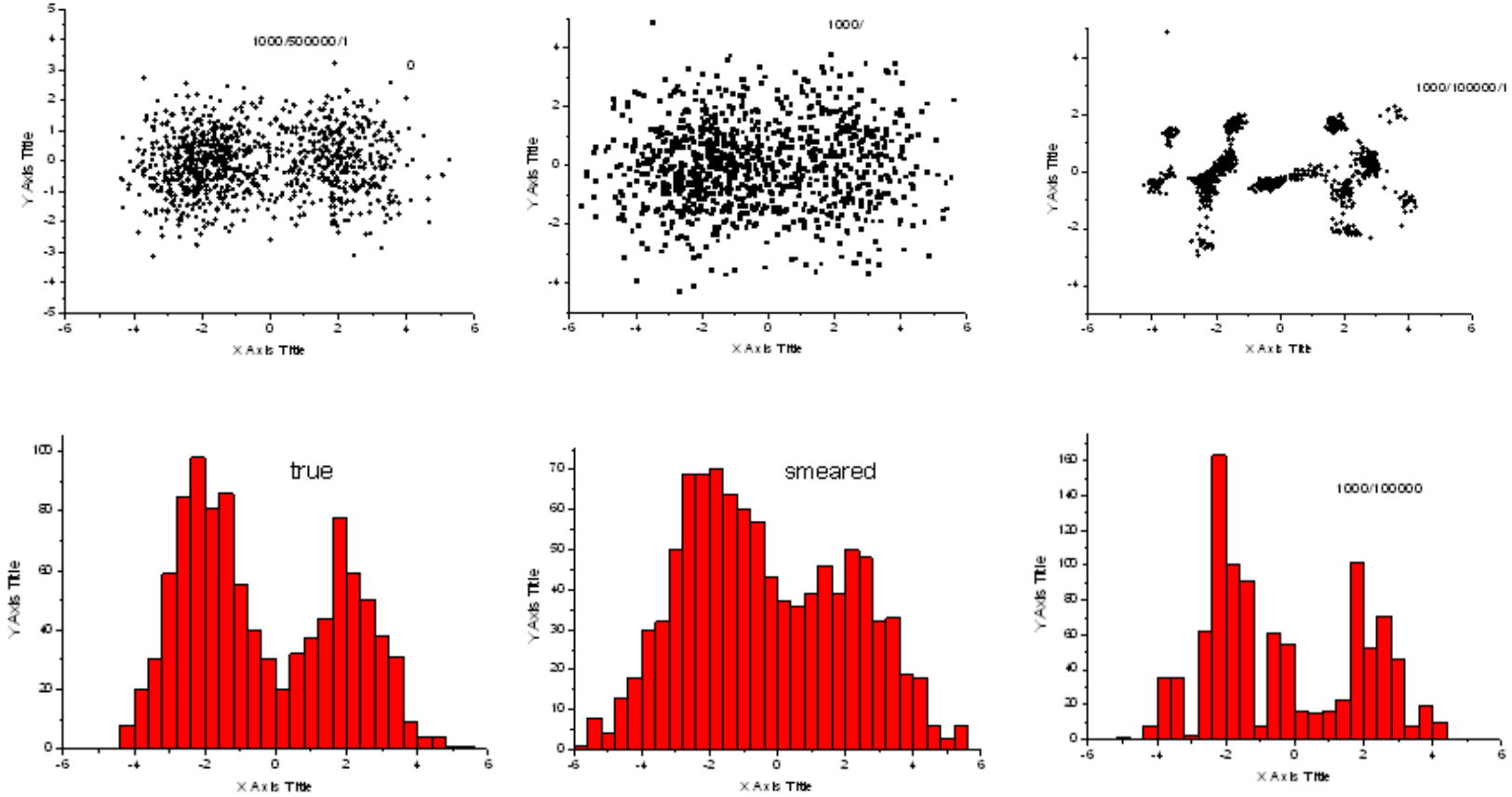
$$G_2(x, y \mid +2, 0; 1, 1)$$

$$f(x', y' \mid x, y) = G(x', y' \mid x, y; 1, 1)$$

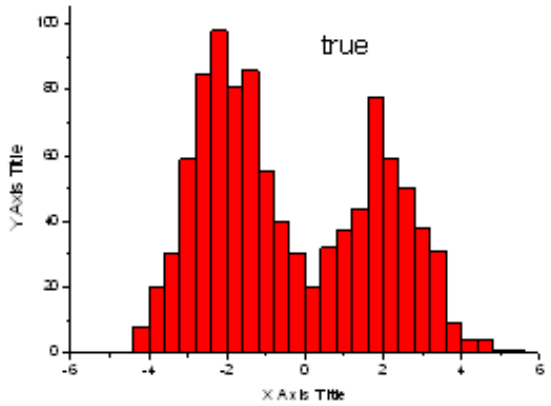
true

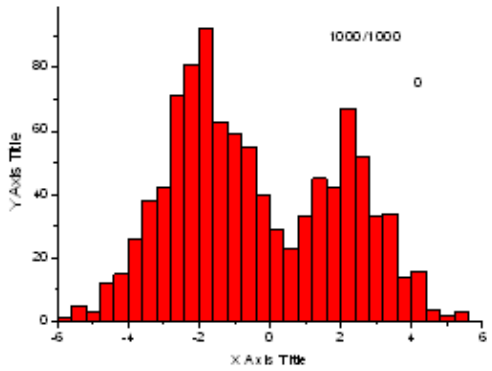
smeared

unfolded
(no regularization)

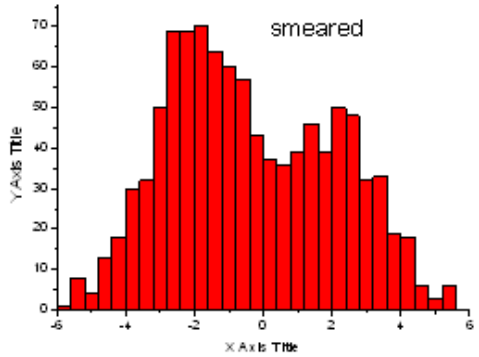


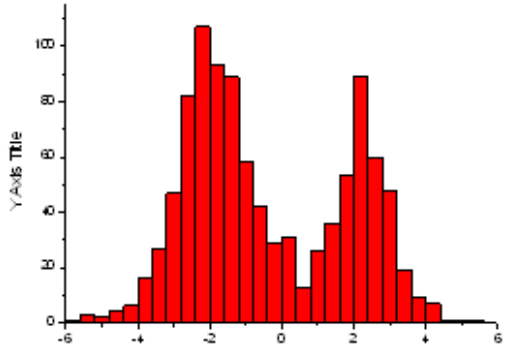
unfolded





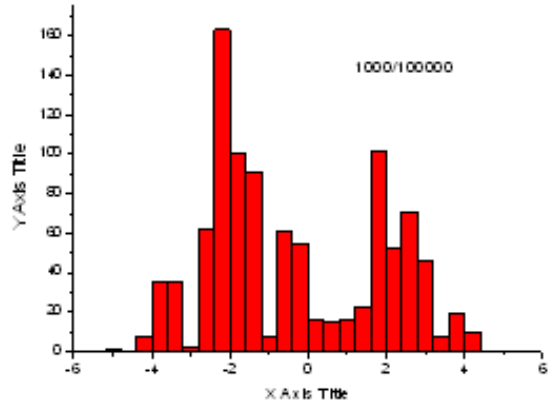
1000 moves





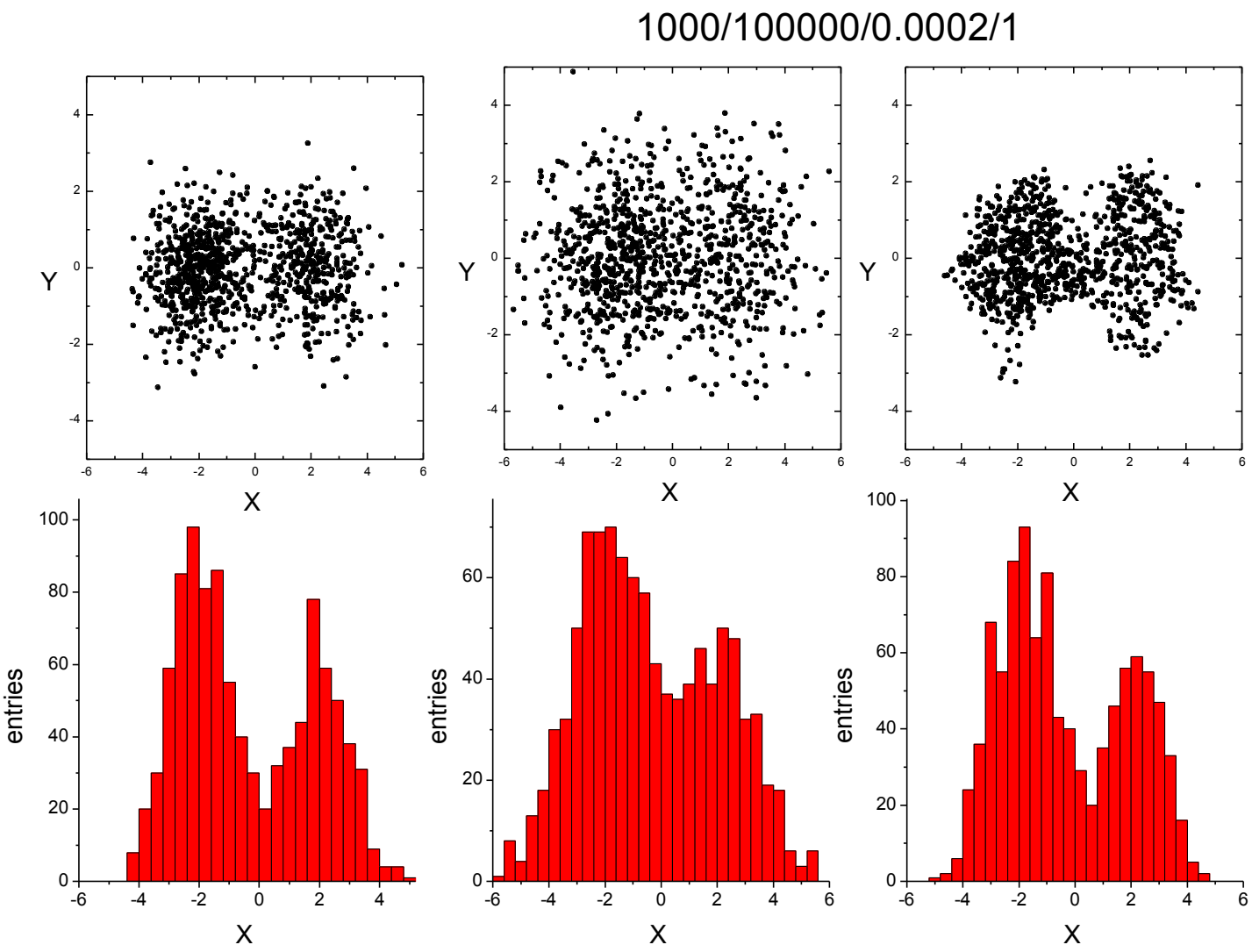
2000 moves

Regularization by limiting the number of moves

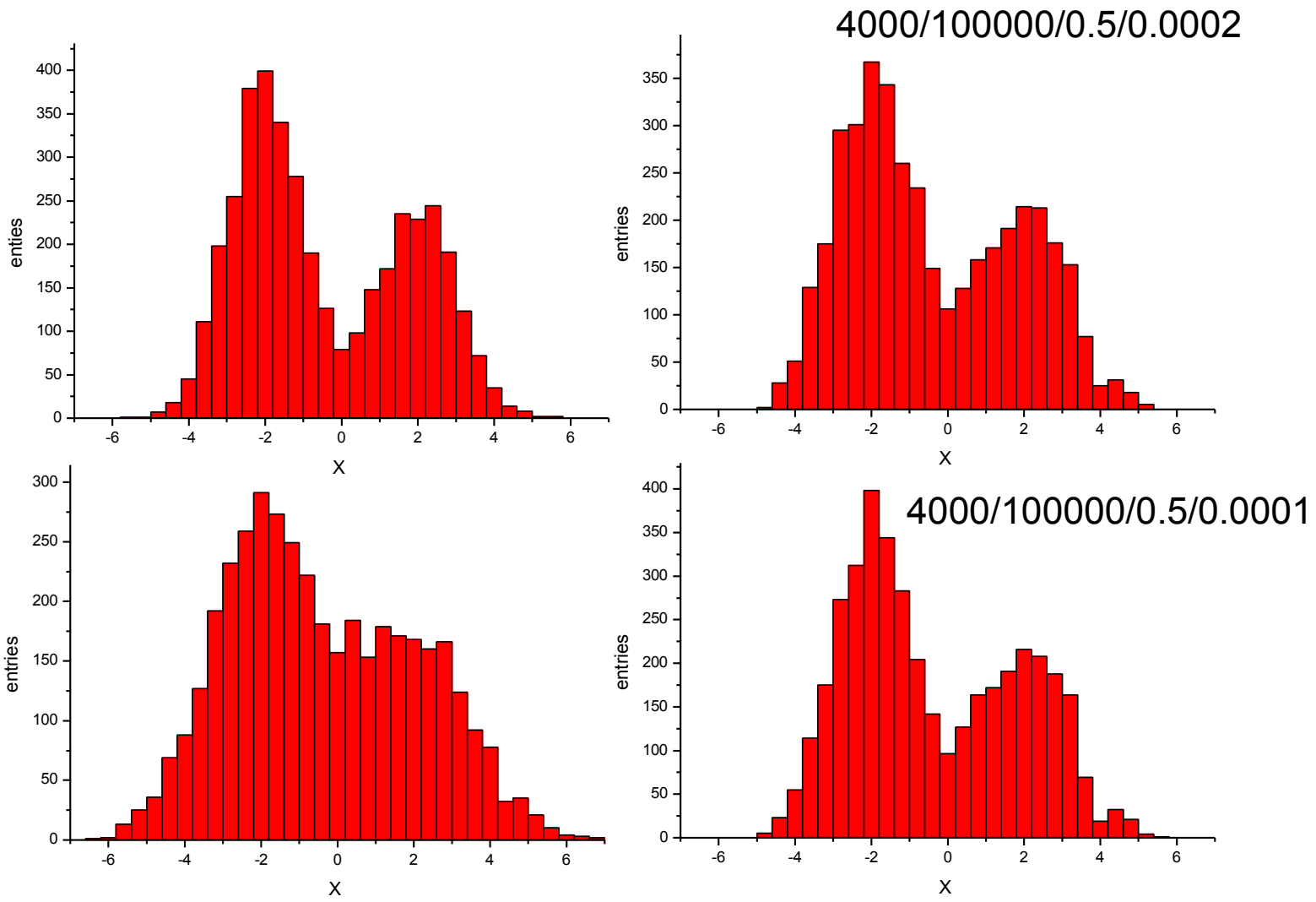


100000 moves

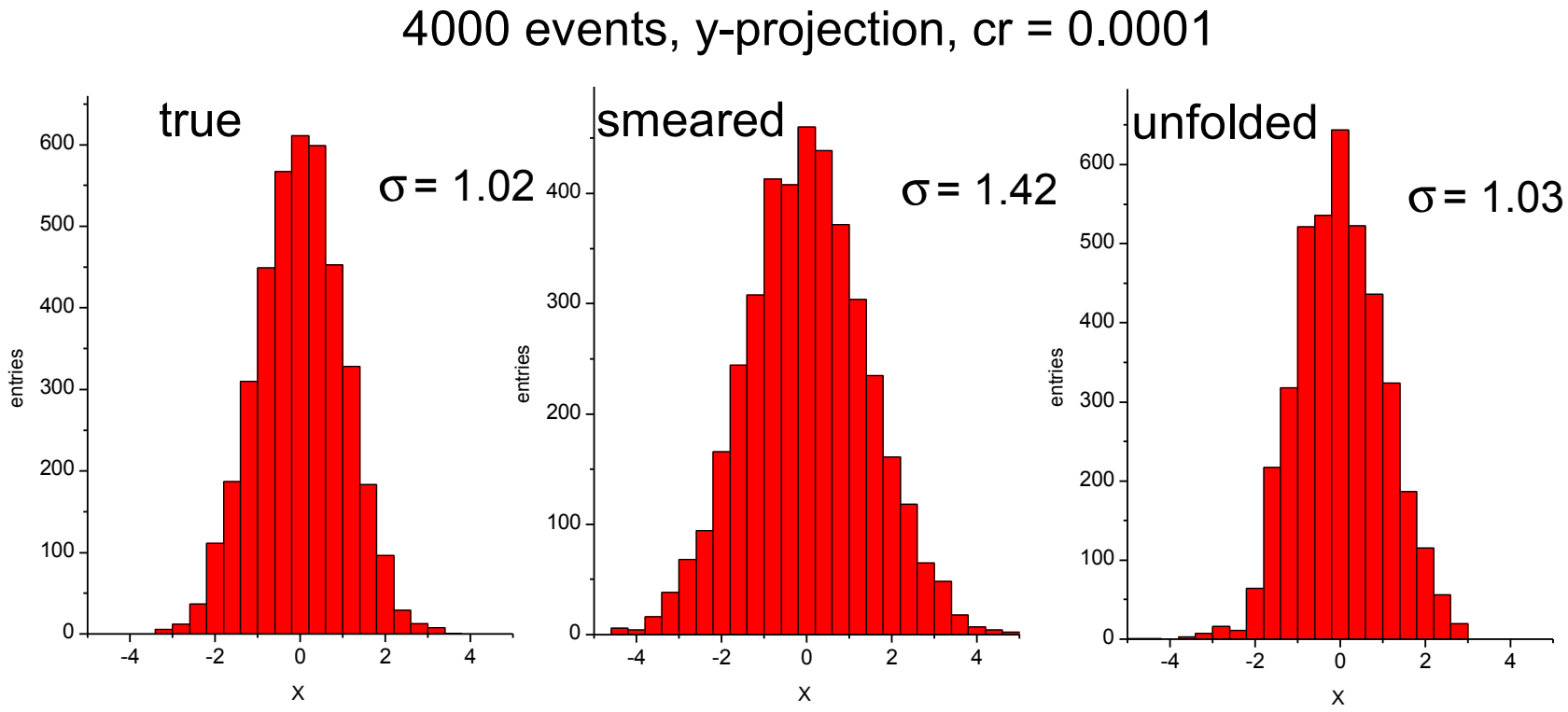
Side band regularization



4000 events, side band regularuzation, x projection

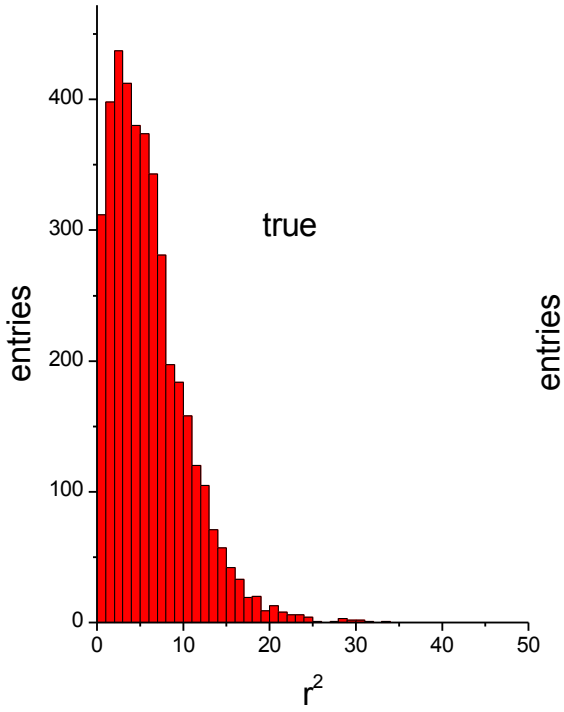


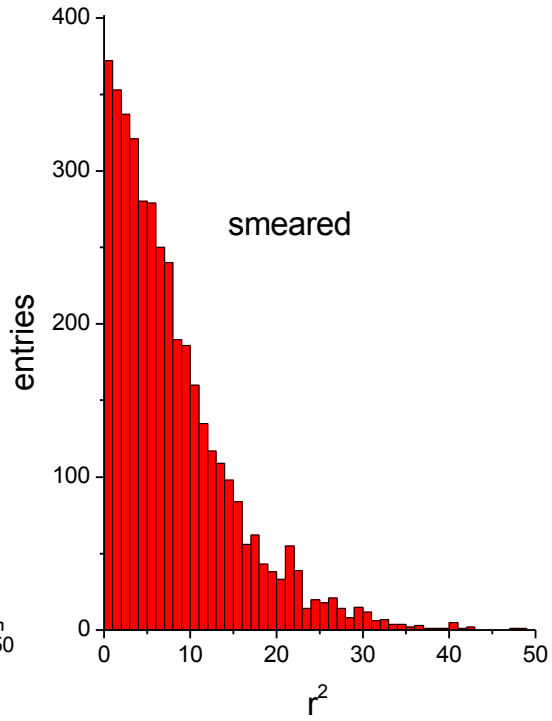
y projection

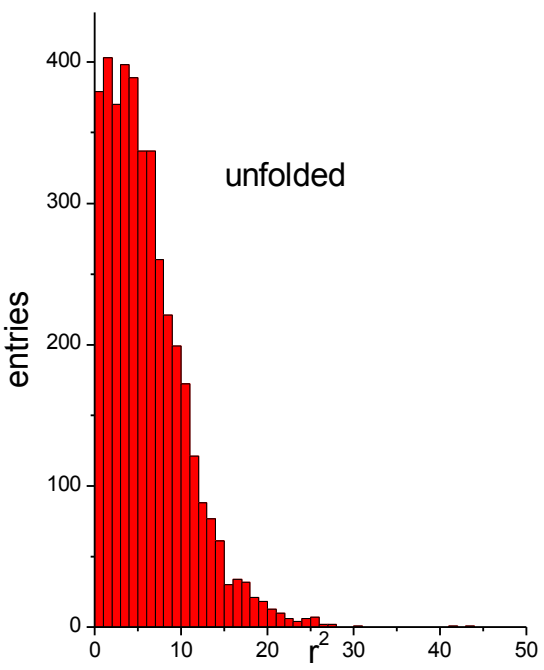


Distribution of $r^2 = x^2 + y^2$

(not possible with binning)







$\sigma = 4.5$ $\sigma = 7.0$ $\sigma = 4.7$

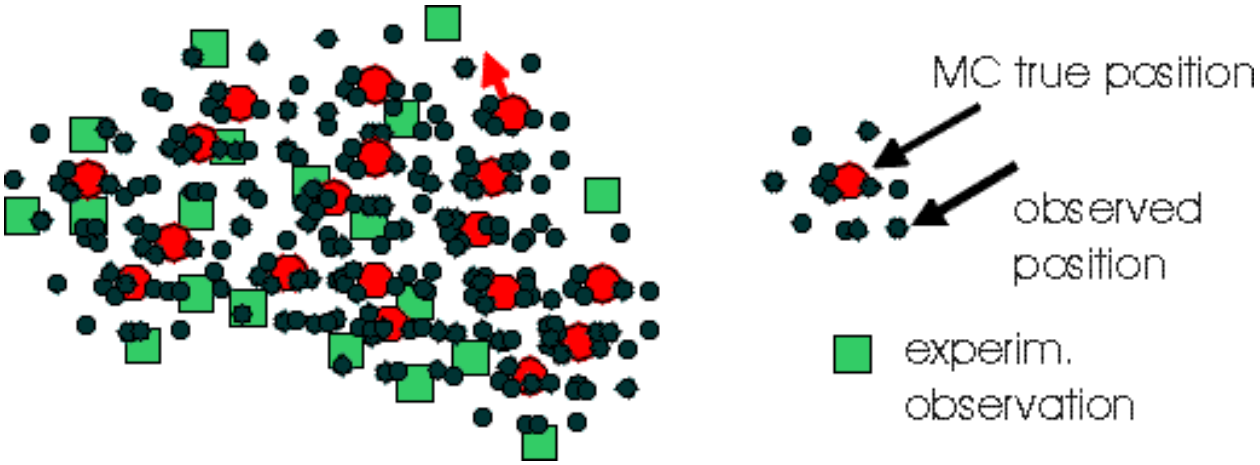
Acceptance and resolution from Monte Carlo

So far we had assumed an analytic resolution function. Normally we know it from a Monte Carlo simulation.

We replace the analytic function by Monte Carlo satellites: Each MC true point is surrounded by *k* observed points (stallites) which are simulated measurements.

We move the true point together with its satellites until the observed points are compatible with the experimental data.

To do so, we need a binning-free goodness-of-fit statistic to measure the aggreement of the simulation with the data: energy test statistic or k nearest neighbor statistic. (see Refs.)



The MC points move until the distribution of the black dots agrees with the distribution of the green boxes

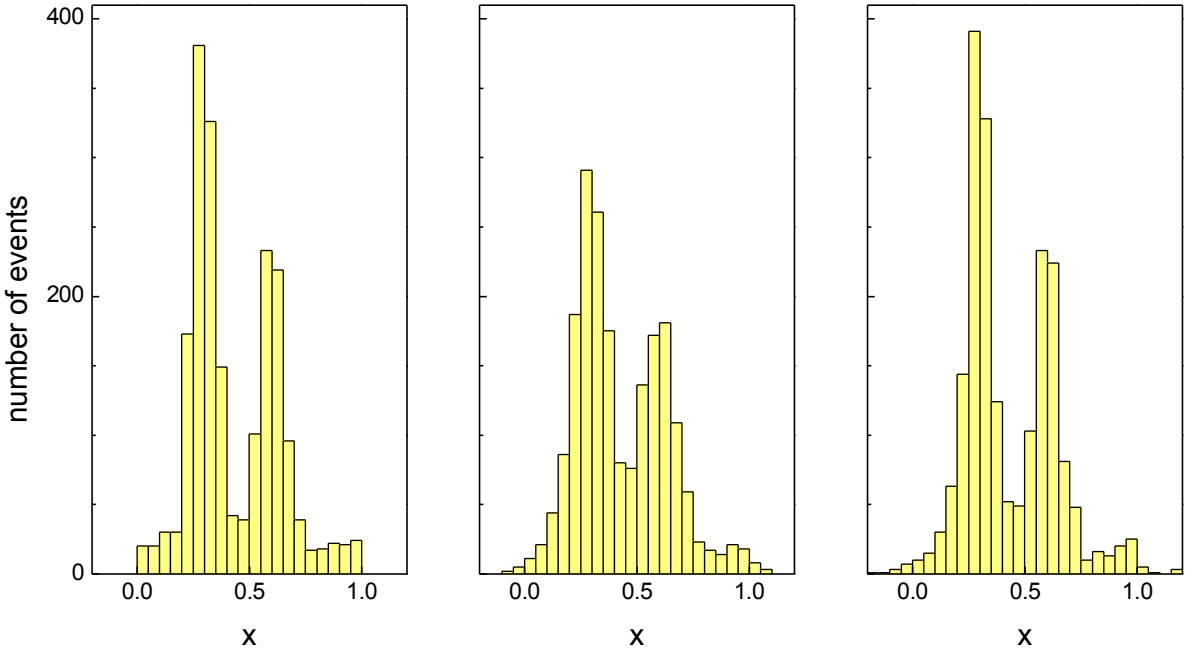
Remarks:

- The smearing is reduced by factor $\sqrt{k}$
- Result is independent of the distance function.
- Result is independent of migration step width.
- Regularization strength depends on k
- Regularization can be steered by stopping the process

Examples

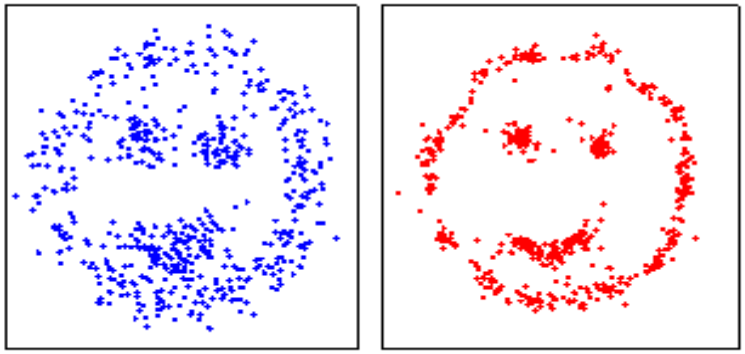
One-dimensional distribution

(unfolded binning-free, presented as histogram)



original smeared unfolded

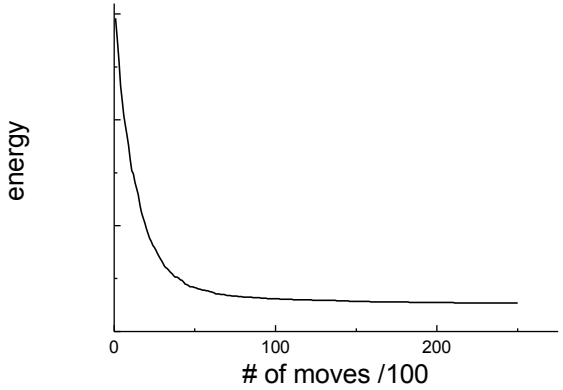
Two-dimensional drawing (not feasable with binning)



600 experimental observations,

$k = 25$ observation per MC true point

20 000 random moves



Some complications

Acceptance losses

Solution: weighting

During generation of observations remember $w_j = k$ / # of trials
→ MC observations are weighted.
After unfolding, weights are included in the error calculation.

Variation of resolution and acceptance with position

(similar problem as in binned case)

Solution: iteration, repeat the simulation periodically

What about speed?

With analytic function, 2 dimensions, N=1000 events + side band regularization, 100000 moves: 100 s
N=4000 → 15 min. $t \sim N^2$
(on a 5 years old slow labtop)

With MC satellites time increases proportional to the number k^2 of satellites

Speed can be increased:

- faster computer,
- migration in two steps. step 1: use approximate analytic function
  step 2: simulate satellites and iprove precision.
- consider only points in neighborhood → $t \sim N$
- increase # of satellites during process

Future improvements

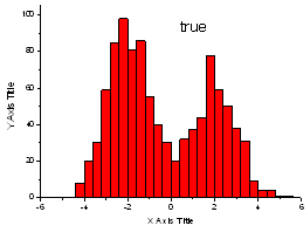
- include side band regularization into MC scheme
- combine analytic and MC approaches.
  - step 1: use approximate analytic function
  - step 2: simulate satellites and improve precision.
- increase speed by storing addresses of neighboring points
- automatic parameter setting based on data

More details can be found in:

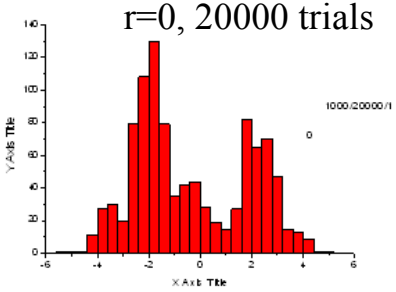
1. G. Bohm, G. Zech, Einführung in Statistik und Messwertanalyse für Physiker , E-book, Desy Library
2. G. Bohm, G. Zech, Introduction to Statistics and Data Analysis for Physicists , E-book, Desy Library (considerably extended w.r. to German version) (soon available)
3. B. Aslan and G. Zech, Statistical energy as a tool for binning- free goodness-of-fit tests, two sample comparison and unfolding. NIM A 537 (2005) 626
4. B. Aslan and G. Zech, \emph{New Test for the Multivariate Two-Sample Problem based on the concept of Minimum Energy}, J. Statist.Comput. Simul. 75, 2 (2004), 109

Side-band regularization in x

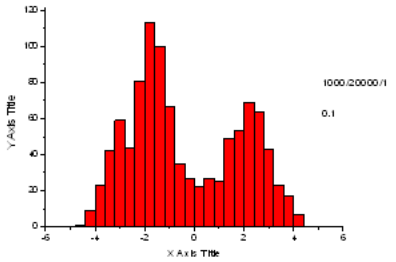
true



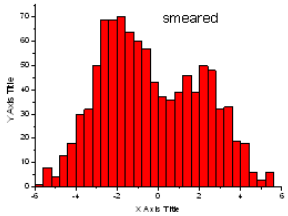
r=0, 20000 trials



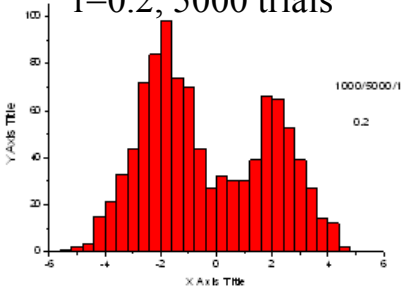
r=0.1, 20000 trials



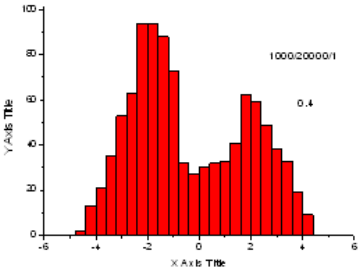
smeared



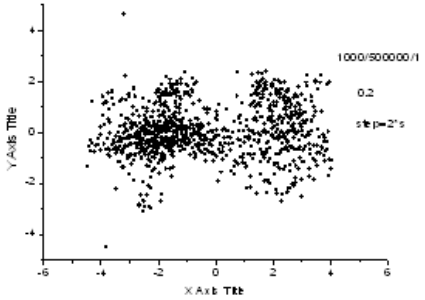
r=0.2, 5000 trials



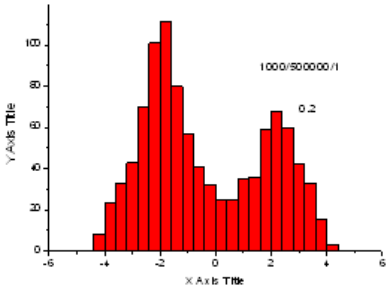
r=0.4, 20000 trials



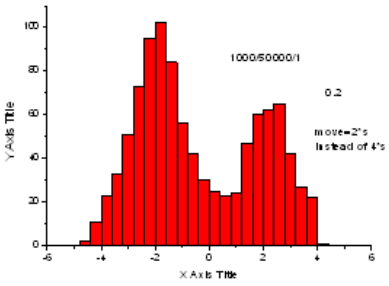
r=0.2, 500000 trials



r=0.2, 500000 trials



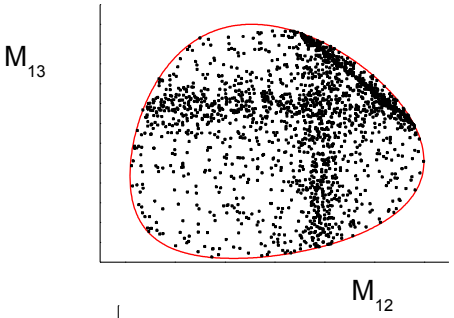
r=0.2, 50000 trials
double step width



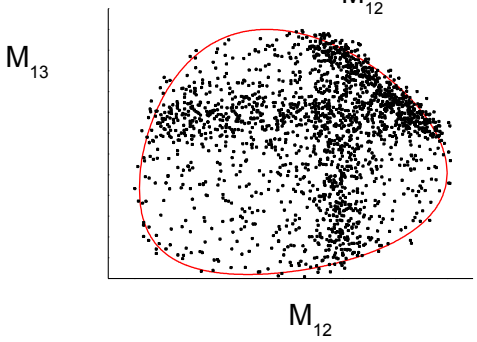
Dalitz plot with 25 satellites

2000 events, K*, ϕ

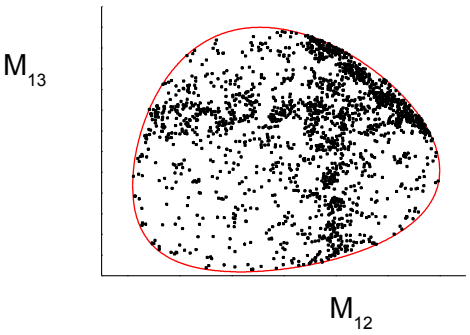
k =25



original data



smeared data



unfolded data