Statistics for HEP

Lecture 3: Further topics

http://indico.cern.ch/conferenceDisplay.py?confId=202569

69th SUSSP
LHC Physics
St. Andrews
20-23 August, 2012

Glen Cowan
Physics Department
Royal Holloway, University of London
g.cowan@rhul.ac.uk
www.pp.rhul.ac.uk/~cowan

Outline

Lecture 1: Introduction and basic formalism
Probability, statistical tests, parameter estimation.

Lecture 2: Discovery and Limits
Quantifying discovery significance and sensitivity
Frequentist and Bayesian intervals/limits

→ Lecture 3: Further topics

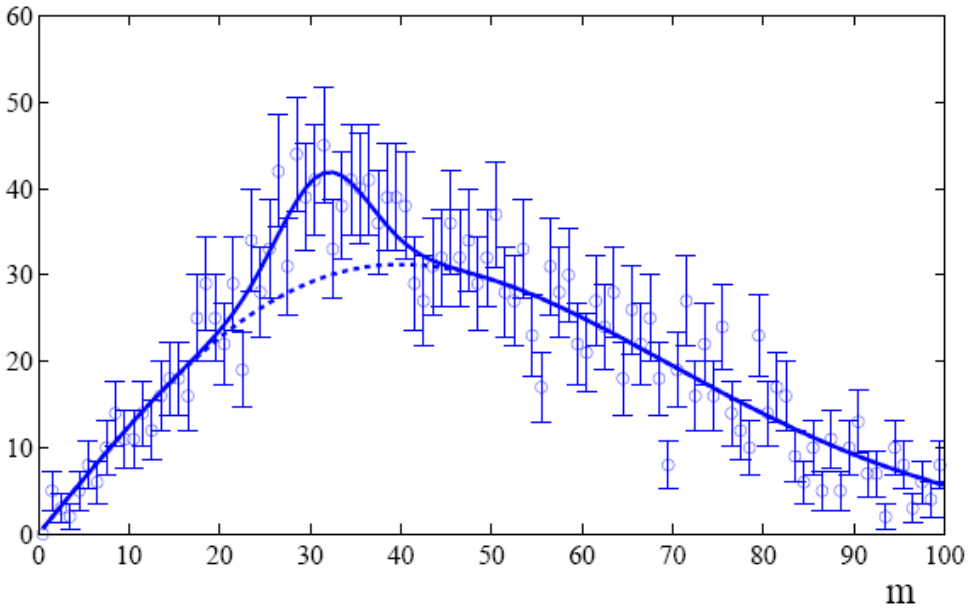
The Look-Elsewhere Effect
Unfolding (deconvolution)

Gross and Vitells, EPJC 70:525-530,2010, arXiv:1005.1891

The Look-Elsewhere Effect

Suppose a model for a mass distribution allows for a peak at a mass m with amplitude μ.

The data show a bump at a mass m_0.



How consistent is this with the no-bump ($\mu = 0$) hypothesis?

p-value for fixed mass

First, suppose the mass m_0 of the peak was specified a priori.

Test consistency of bump with the no-signal ($\mu = 0$) hypothesis with e.g. likelihood ratio

$$t_{\text{fix}} = -2 \ln \frac{L(0, m_0)}{L(\hat{\mu}, m_0)}$$

where "fix" indicates that the mass of the peak is fixed to m_0.

The resulting *p*-value

$$p_{\text{fix}} = \int_{t_{\text{fix,obs}}}^{\infty} f(t_{\text{fix}}|0) dt_{\text{fix}}$$

gives the probability to find a value of t_{fix} at least as great as observed at the specific mass m_0.

p-value for floating mass

But suppose we did not know where in the distribution to expect a peak.

What we want is the probability to find a peak at least as significant as the one observed anywhere in the distribution.

Include the mass as an adjustable parameter in the fit, test significance of peak using

$$t_{\text{float}} = -2 \ln \frac{L(0)}{L(\hat{\mu}, \hat{m})}$$

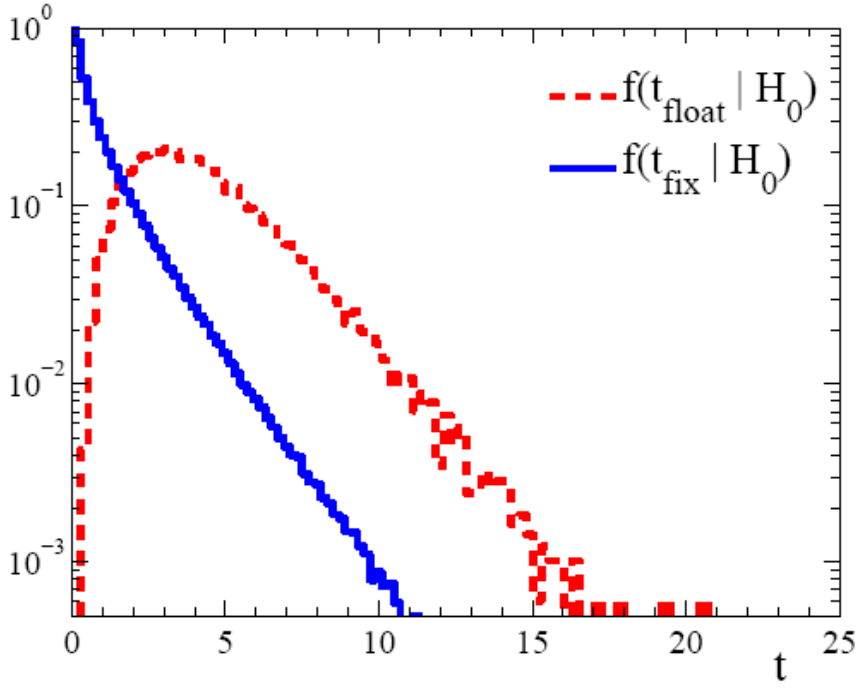
(Note m does not appear in the $\mu = 0$ model.)

$$p_{\text{float}} = \int_{t_{\text{float,obs}}}^{\infty} f(t_{\text{float}}|0) dt_{\text{float}}$$

Distributions of t_{fix}, t_{float}

For a sufficiently large data sample, t_{fix} ~chi-square for 1 degree of freedom (Wilks' theorem).

For t_{float} there are two adjustable parameters, μ and m, and naively Wilks theorem says t_{float} ~ chi-square for 2 d.o.f.



In fact Wilks' theorem does not hold in the floating mass case because on of the parameters (m) is not-defined in the $\mu = 0$ model.

So getting t_{float} distribution is more difficult.

Approximate correction for LEE

We would like to be able to relate the *p*-values for the fixed and floating mass analyses (at least approximately).

Gross and Vitells show the *p*-values are approximately related by

$$p_{\text{float}} \approx p_{\text{fix}} + \langle N(c) \rangle$$

where $\langle N(c) \rangle$ is the mean number "upcrossings" of $-2\ln L$ in the fit range based on a threshold

$$c = t_{\text{fix}} = Z_{\text{fix}}^2$$

and where Z_{fix} is the significance for the fixed mass case.

So we can either carry out the full floating-mass analysis (e.g. use MC to get *p*-value), or do fixed mass analysis and apply a correction factor (much faster than MC).

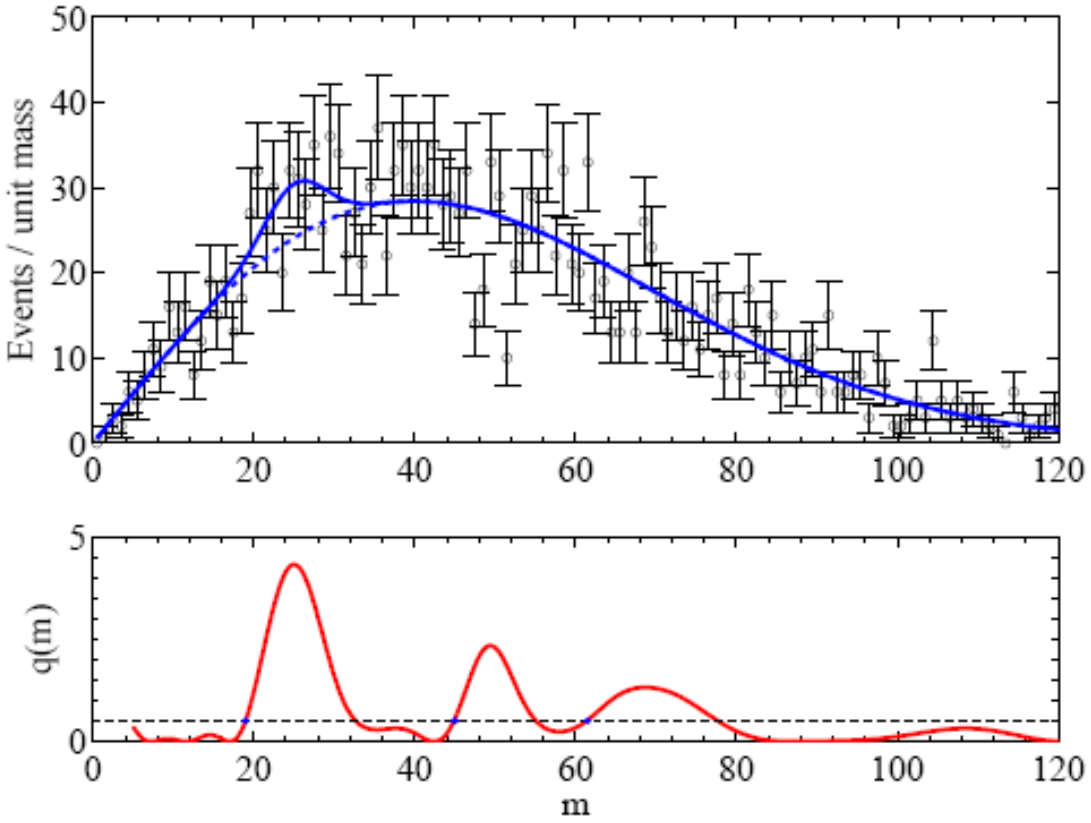
Upcrossings of $-2\ln L$

The Gross-Vitells formula for the trials factor requires $\langle N(c)\rangle$, the mean number "upcrossings" of $-2\ln L$ in the fit range based on a threshold $c = t_{\text{fix}} = Z_{\text{fix}}{}^2$.

$\langle N(c)\rangle$ can be estimated from MC (or the real data) using a much lower threshold c_0:

$$\langle N(c)\rangle \approx \langle N(c_0)\rangle e^{-(c-c_0)/2}$$

In this way $\langle N(c)\rangle$ can be estimated without need of large MC samples, even if the the threshold c is quite high.



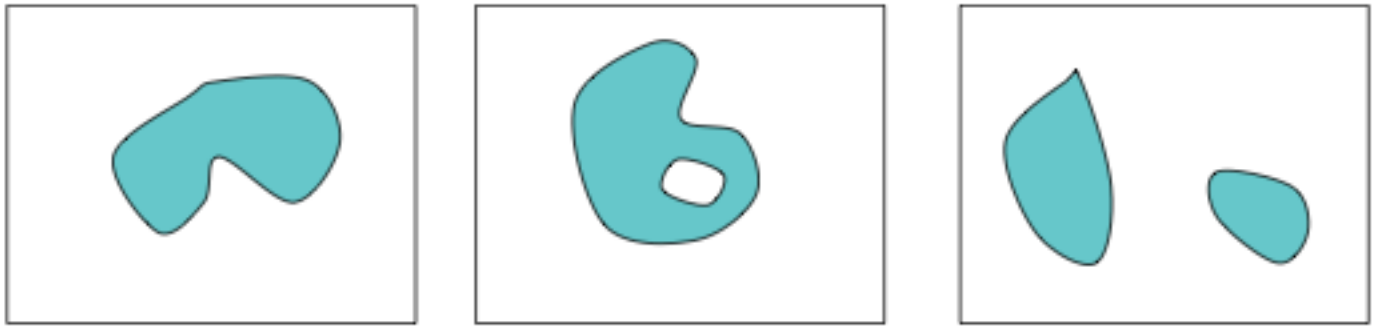
Vitells and Gross, Astropart. Phys. 35 (2011) 230-234; arXiv:1105.4355

Multidimensional look-elsewhere effect

Generalization to multiple dimensions: number of upcrossings replaced by expectation of Euler characteristic:

$$\mathrm{E}[\varphi(A_u)] = \sum_{d=0}^{n} \mathcal{N}_d \rho_d(u)$$

- Number of disconnected components minus number of `holes'



Applications: astrophysics (coordinates on sky), search for resonance of unknown mass and width, ...

Summary on Look-Elsewhere Effect

Remember the Look-Elsewhere Effect is when we test a single model (e.g., SM) with multiple observations, i..e, in mulitple places.

Note there is no look-elsewhere effect when considering exclusion limits. There we test specific signal models (typically once) and say whether each is excluded.

With exclusion there is, however, the analogous issue of testing many signal models (or parameter values) and thus excluding some even in the absence of signal ("spurious exclusion")

Approximate correction for LEE should be sufficient, and one should also report the uncorrected significance.

> "There's no sense in being precise when you don't even know what you're talking about." — John von Neumann

Why 5 sigma?

Common practice in HEP has been to claim a discovery if the p-value of the no-signal hypothesis is below 2.9×10^{-7}, corresponding to a significance $Z = \Phi^{-1}(1 - p) = 5$ (a 5σ effect).

There a number of reasons why one may want to require such a high threshold for discovery:

The "cost" of announcing a false discovery is high.

Unsure about systematics.

Unsure about look-elsewhere effect.

The implied signal may be a priori highly improbable (e.g., violation of Lorentz invariance).

Why 5 sigma (cont.)?

But the primary role of the *p*-value is to quantify the probability that the background-only model gives a statistical fluctuation as big as the one seen or bigger.

It is not intended as a means to protect against hidden systematics or the high standard required for a claim of an important discovery.

In the processes of establishing a discovery there comes a point where it is clear that the observation is not simply a fluctuation, but an "effect", and the focus shifts to whether this is new physics or a systematic.

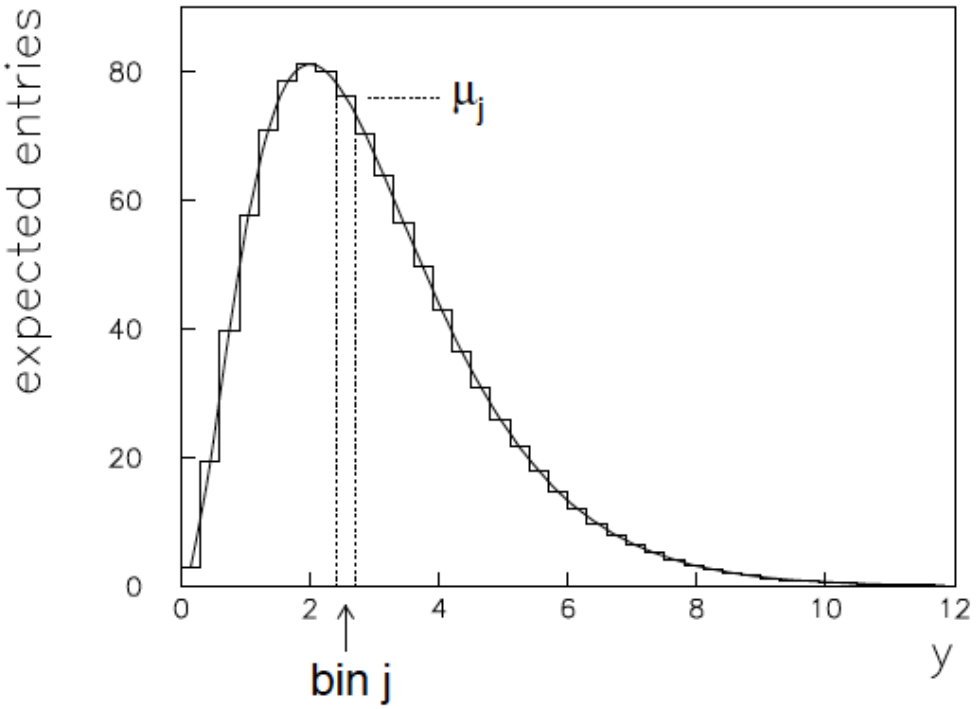
Providing LEE is dealt with, that threshold is probably closer to 3σ than 5σ.

Formulation of the unfolding problem

Consider a random variable y, goal is to determine pdf $f(y)$.

If parameterization $f(y;\theta)$ known, find e.g. ML estimators $\hat{\theta}$.

If no parameterization available, construct histogram:



$$p_j = \int_{\text{bin } j} f(y)\, dy$$

$$\mu_j = \mu_{\text{tot}} p_j$$

"true" histogram

New goal: construct estimators for the μ_j (or p_j).

Migration

Effect of measurement errors: y = true value, x = observed value,

migration of entries between bins,

$f(y)$ is 'smeared out', peaks broadened.

$$f_{\text{meas}}(x) = \int R(x|y) f_{\text{true}}(y)\, dy$$

discretize: data are $\mathbf{n} = (n_1, \ldots, n_N)$

$$\nu_i = E[n_i] = \sum_{j=1}^{M} R_{ij}\mu_j \,, \quad i = 1, \ldots, N$$

response matrix

$$R_{ij} = P(\text{observed in bin } i \,|\, \text{true in bin } j)$$

Note $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ are constants; $\boldsymbol{n}$ subject to statistical fluctuations.

Efficiency, background

Sometimes an event goes undetected:

$$\sum_{i=1}^{N} R_{ij} = \sum_{i=1}^{N} P(\text{observed in bin } i \,|\, \text{true value in bin } j)$$

$$= P(\text{observed anywhere} \,|\, \text{true value in bin } j)$$

$$= \varepsilon_j \qquad \longleftarrow \text{ efficiency}$$

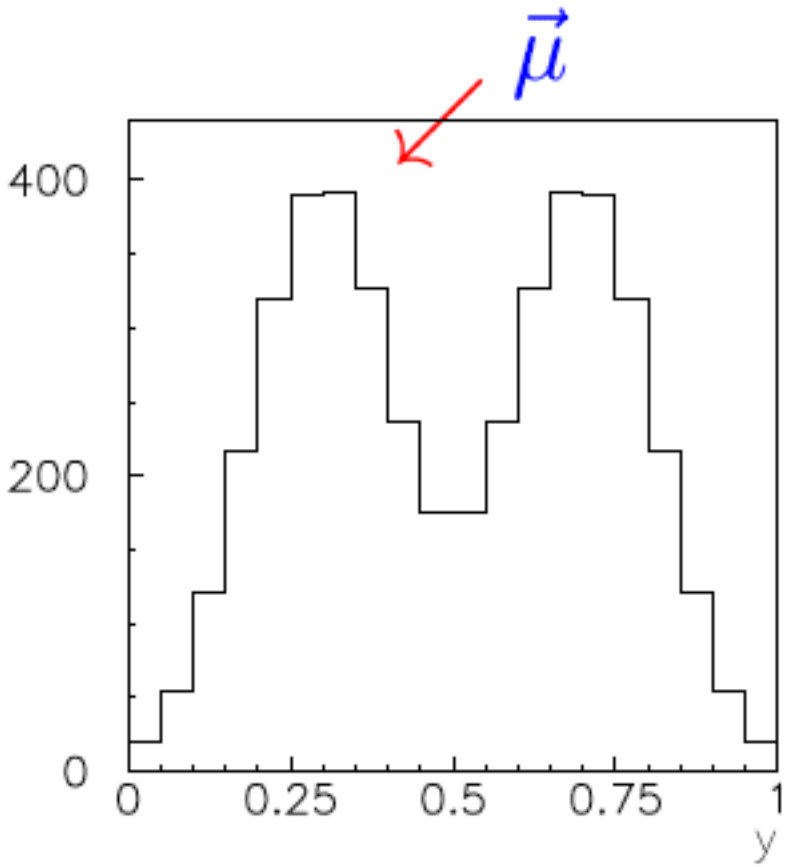
Sometimes an observed event is due to a background process:

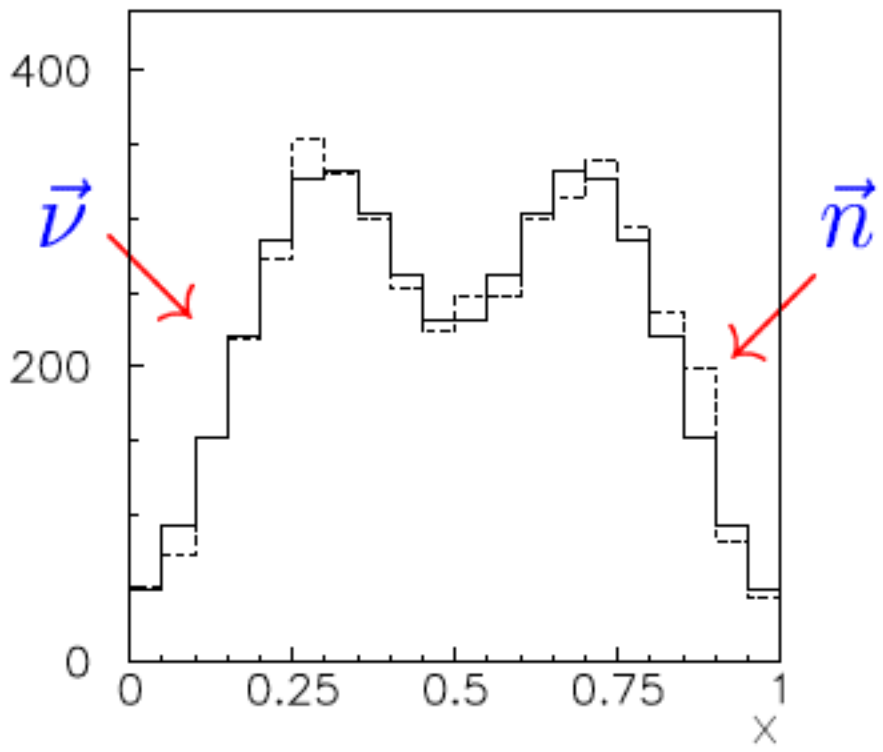
$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i$$

β_i = expected number of background events in *observed* histogram.

For now, assume the β_i are known.

The basic ingredients





"true"

"observed"

Summary of ingredients

'true' histogram: $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M), \quad \mu_{\text{tot}} = \sum_{j=1}^{M} \mu_j$

probabilities: $\mathbf{p} = (p_1, \ldots, p_M) = \boldsymbol{\mu}/\mu_{\text{tot}}$

expectation values for observed histogram: $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$

observed histogram: $\mathbf{n} = (n_1, \ldots, n_N)$

response matrix: $R_{ij} = P(\text{observed in bin } i \,|\, \text{true in bin } j)$

efficiencies: $\varepsilon_j = \sum_{i=1}^{N} R_{ij}$

expected background: $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$

These are related by:

$$E[\mathbf{n}] = \boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta}$$

Maximum likelihood (ML) estimator from inverting the response matrix

Assume $\boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta}$ can be inverted: $\boldsymbol{\mu} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta})$

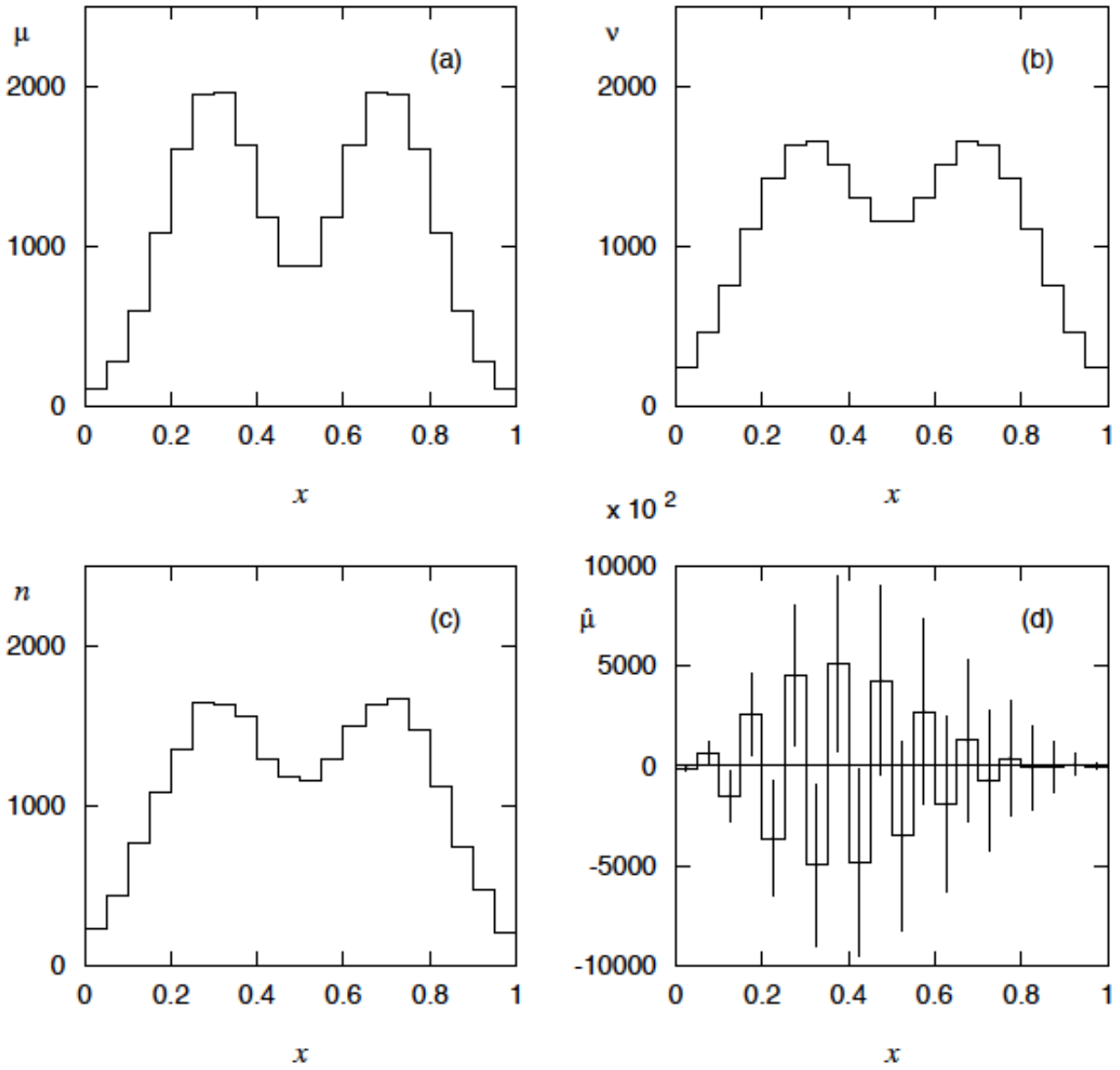
Suppose data are independent Poisson: $P(n_i; \nu_i) = \dfrac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$

So the log-likelihood is $\ln L(\boldsymbol{\mu}) = \sum_{i=1}^{N} (n_i \ln \nu_i - \nu_i)$

ML estimator is $\hat{\boldsymbol{\nu}} = \mathbf{n}$

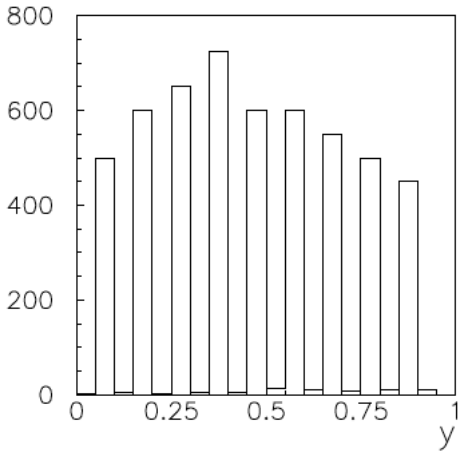
$\longrightarrow$ $\hat{\boldsymbol{\mu}} = R^{-1}(\mathbf{n} - \boldsymbol{\beta})$

Example with ML solution



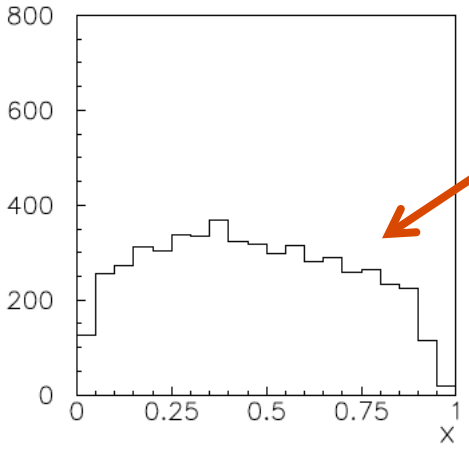
Catastrophic failure???

What went wrong?



Suppose $\boldsymbol{\mu}$ really had a lot of fine structure.

$\vec{\mu}$



$\vec{\nu} = R\vec{\mu}$

Applying R washes this out, but leaves a residual structure:

Applying R^{-1} to $\vec{\nu}$ puts the fine structure back: $\vec{\mu} = R^{-1}\vec{\nu}$.

But we don't have $\boldsymbol{\nu}$, only $\boldsymbol{n}$. R^{-1} "thinks" fluctuations in $\boldsymbol{n}$ are the residual of original fine structure, puts this back into $\hat{\mu}$.

ML solution revisited

For Poisson data the ML estimators are unbiased:

$$E[\hat{\boldsymbol{\mu}}] = R^{-1}(E[\mathbf{n}] - \boldsymbol{\beta}) = \boldsymbol{\mu}$$

Their covariance is:

$$U_{ij} = \mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = \sum_{k,l=1}^{N} (R^{-1})_{ik}(R^{-1})_{jl}\,\mathrm{cov}[n_k, n_l]$$

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

(Recall these statistical errors were huge for the example shown.)

ML solution revisited (2)

The information inequality gives for unbiased estimators the minimum (co)variance bound:

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

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

This is the same as the actual variance! I.e. ML solution gives smallest variance among all unbiased estimators, even though this variance was huge.

In unfolding one must accept some bias in exchange for a (hopefully large) reduction in variance.

Correction factor method

Use equal binning for $\vec{\mu}$, $\vec{\nu}$ and take $\hat{\mu}_i = C_i(n_i - \beta_i)$, where

$$C_i = \frac{\mu_i^{\rm MC}}{\nu_i^{\rm MC}}$$

$\nu_i^{\rm MC}$ and $\mu_i^{\rm MC}$ from Monte Carlo simulation (no background).

$$U_{ij} = {\rm cov}[\hat{\mu}_i, \hat{\mu}_j] = C_i^2\, {\rm cov}[n_i, n_j]$$

Often $C_i \sim O(1)$ so statistical errors far smaller than for ML.

But the bias $b_i = E[\hat{\mu}_i] - \mu_i$ is $\quad b_i = \left(\frac{\mu_i^{\rm MC}}{\nu_i^{\rm MC}} - \frac{\mu_i}{\nu_i^{\rm sig}} \right)$

Nonzero bias unless MC = Nature.

$$\nu_i^{\rm sig} = \nu_i - \beta_i$$

Reality check on the statistical errors

Suppose for some bin *i* we have:

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

$$\rightarrow \quad \hat{\mu}_i = C_i n_i = 10 \qquad \sigma_{\hat{\mu}_i} = C_i \sqrt{n_i} = 1.0$$

(10% stat. error)

But according to the estimate, only 10 of the 100 events found in the bin belong there; the rest spilled in from outside.

How can we have a 10% measurement if it is based on only 10 events that really carry information about the desired parameter?

Discussion of correction factor method

As with all unfolding methods, we get a reduction in statistical error in exchange for a bias; here the bias is difficult to quantify (difficult also for many other unfolding methods).

The bias should be small if the bin width is substantially larger than the resolution, so that there is not much bin migration.

So if other uncertainties dominate in an analysis, correction factors may provide a quick and simple solution (a "first-look").

Still the method has important flaws and it would be best to avoid it.

Regularized unfolding

Consider 'reasonable' estimators such that for some $\Delta \log L$,

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

Out of these estimators, choose the 'smoothest', by maximizing

$$\Phi(\vec{\mu}) = \alpha \log L(\vec{\mu}) + S(\vec{\mu}),$$

$S(\vec{\mu}) =$ regularization function (measure of smoothness),

$\alpha =$ regularization parameter (choose to give desired $\Delta \log L$)

Regularized unfolding (2)

In addition require $\sum_{i=1}^{N} \nu_i = \sum_{i,j} R_{ij}\mu_j = n_{\text{tot}}$, i.e. maximize

$$\varphi(\vec{\mu}, \lambda) = \alpha \log L(\vec{\mu}) + S(\vec{\mu}) + \lambda \left[n_{\text{tot}} - \sum_{i=1}^{N} \nu_i \right]$$

where λ is a Lagrange multiplier, $\partial\varphi/\partial\lambda = 0 \rightarrow \sum_{i=1}^{N} \nu_i = n_{\text{tot}}$.

$\alpha = 0$ gives smoothest solution (ignores data!),

$\alpha \rightarrow \infty$ gives ML solution (variance too large).

We need: regularization function $S(\vec{\mu})$,

a prescription for setting α.

Tikhonov regularization

Take measure of smoothness = mean square of kth derivative,

$$S[f_{\text{true}}(y)] = -\int \left(\frac{d^k f_{\text{true}}(y)}{dy^k} \right)^2 dy \ , \text{ where } k = 1, 2, \ldots$$

If we use Tikhonov ($k = 2$) with $\log L = -\frac{1}{2}\chi^2$,

$$S(\boldsymbol{\mu}) = -\sum_{i=1}^{M-2} (-\mu_i + 2\mu_{i+1} - \mu_{i+2})^2$$

$$\varphi(\vec{\mu}, \lambda) = -\frac{\alpha}{2}\chi^2(\vec{\mu}) + S(\vec{\mu}) \quad \text{quadratic in } \mu_i,$$

$\rightarrow$ setting derivatives of φ equal to zero gives linear equations.

Solution using Singular Value Decomposition (SVD).

SVD implementation of Tikhonov unfolding

A. Hoecker, V. Kartvelishvili, NIM A372 (1996) 469;
(TSVDUnfold in ROOT).

Minimizes $$\chi^2(\boldsymbol{\mu}) + \tau \sum_i \left[(\mu_{i+1} - \mu_i) - (\mu_i - \mu_{i-1})^2 \right]$$

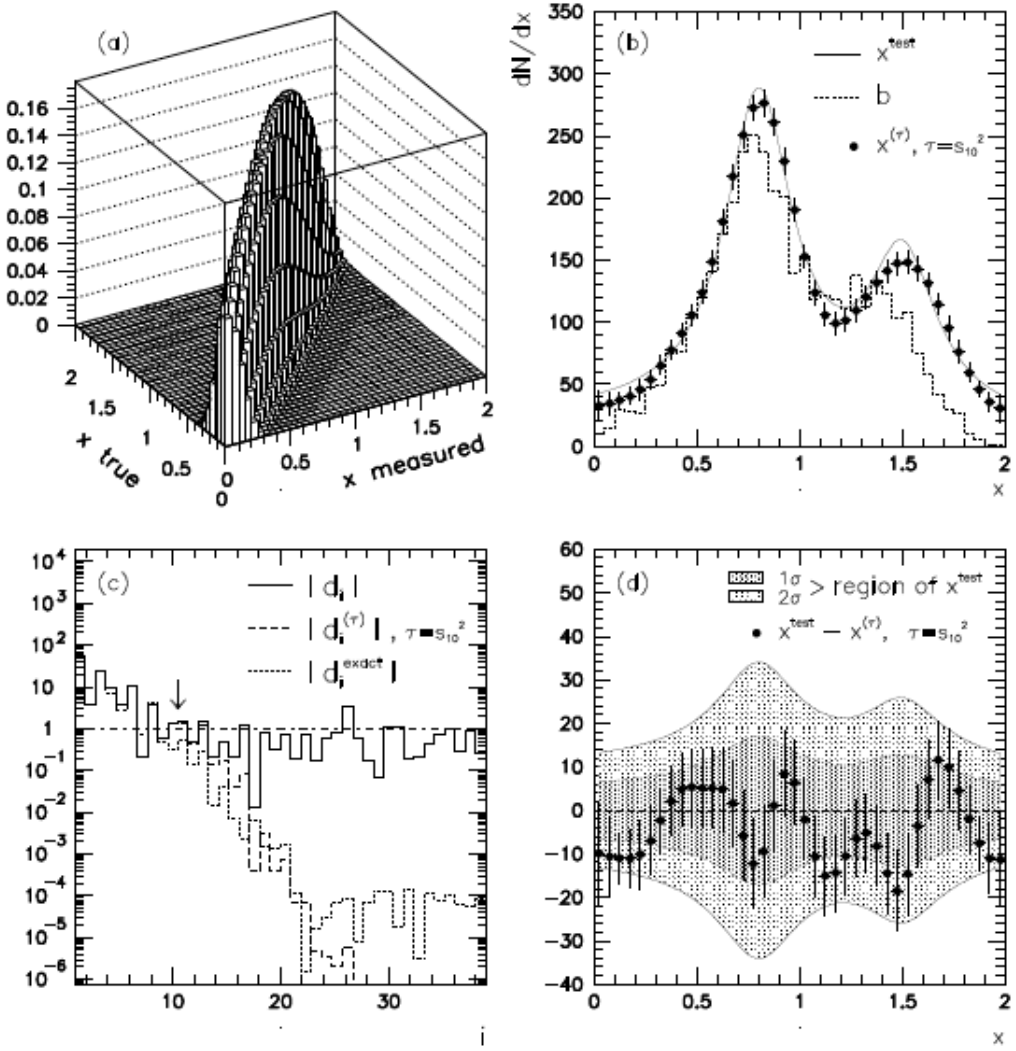
Numerical implementation using Singular Value Decomposition.

Recommendations for setting regularization parameter τ:

Transform variables so errors ~ Gauss(0,1);
number of transformed values significantly different from zero gives prescription for τ;
or base choice of τ on unfolding of test distributions.

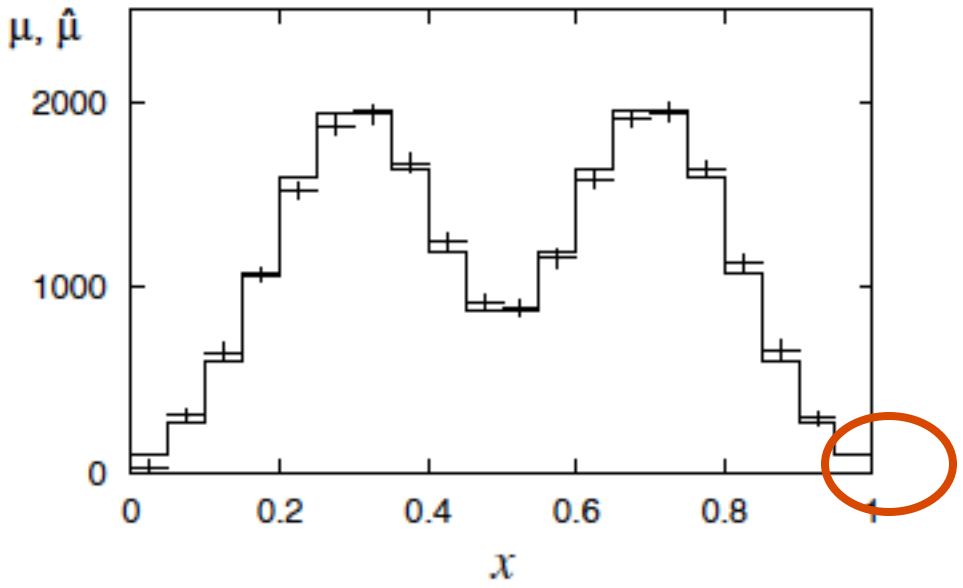
SVD example

A. Höcker, V. Kartvelishvili, NIM A**372** (1996) 469.



Edge effects

Regularized unfolding can lead to "edge effects". E.g. in Tikhonov regularization with Gaussian data, solution can go negative:



Solution pushed negative.

Important e.g. if New Physics would appear as a longer tail of a distribution.

Regularization function based on entropy

Shannon entropy of a set of probabilities is

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

All p_i equal $\rightarrow$ maximum entropy (maximum smoothness)

One $p_i = 1$, all others $= 0$ $\rightarrow$ minimum entropy

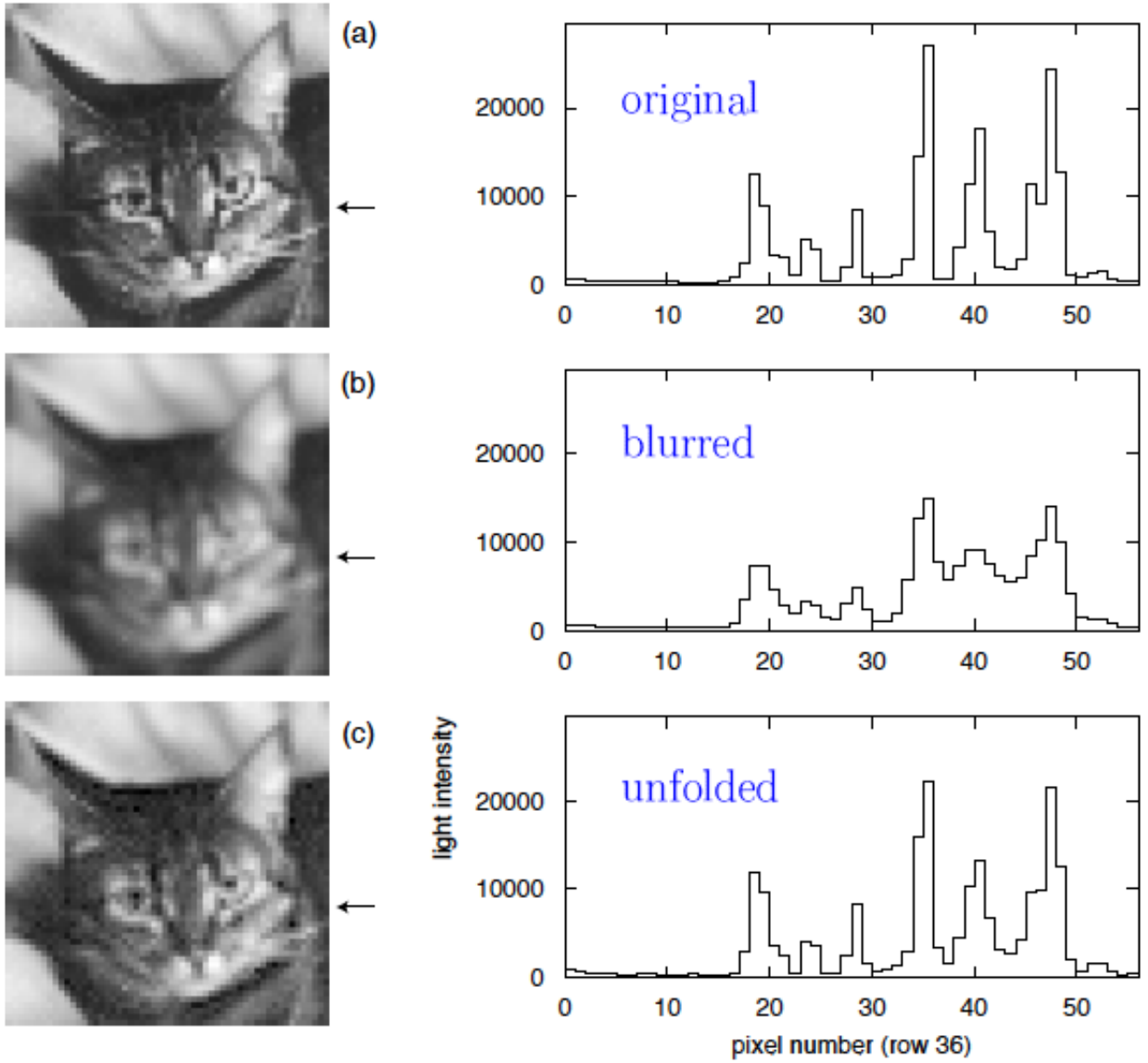
Use entropy as regularization function,

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

$$\propto \log(\text{number of ways to arrange } \mu_{\rm tot} \text{ entries in } M \text{ bins})$$

Can have Bayesian motivation: $S(\vec{\mu}) \rightarrow$ prior pdf for $\vec{\mu}$

Example of entropy-based unfolding



Estimating bias and variance

In general, the equations determining $\hat{\vec{\mu}}(\vec{n})$ are nonlinear.

Expand $\hat{\vec{\mu}}(\vec{n})$ about $\vec{n}_{\rm obs}$ (observed data set),

Use error propagation to get covariance $U_{ij} = {\rm cov}[\hat{\mu}_i, \hat{\mu}_j]$,

and estimators for the bias, $b_i = E[\hat{\mu}_i] - \mu_i$,

$$\hat{b}_i = \sum_{j=1}^{N} \frac{\partial \hat{\mu}_i}{\partial n_j} (\hat{\nu}_j - n_j),$$

where $\hat{\vec{\nu}} = R\hat{\vec{\mu}} + \vec{\beta}$. (N.B. $\hat{\vec{\nu}} \neq \vec{n}$.)

Choosing the regularization parameter

$\alpha = 0 \rightarrow \hat{\vec{\mu}}$ maximally smooth (ignores data).

$\alpha \rightarrow \infty \rightarrow$ ML solution (no bias, very large variance).

Possible criteria for best trade-off between bias and variance:

Minimize mean squared error,

$$\text{MSE} = \frac{1}{M} \sum_{i=1}^{M} \left(U_{ii} + \hat{b}_i^2 \right), \text{ or}$$

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

Choosing the regularization parameter (2)

Or look at changes in χ^2 from unregularized (ML) solution,

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

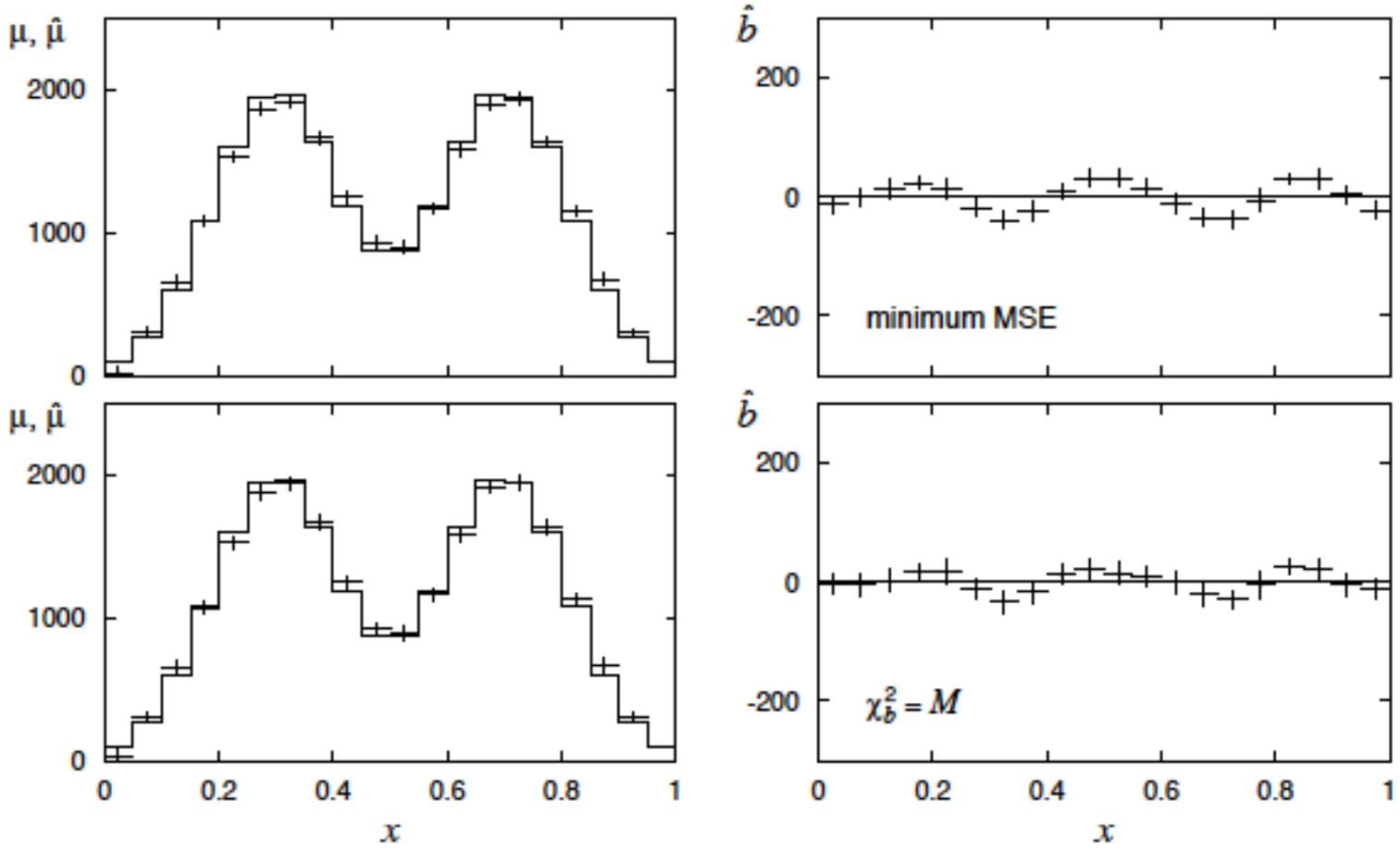
Or require that bias be consistent with zero to within its own error,

$$\chi^2_b = \sum_{i=1}^{M} \frac{\hat{b}_i^2}{W_{ii}} = M \quad \text{where } W_{ij} = \text{cov}[\hat{b}_i, \hat{b}_j].$$

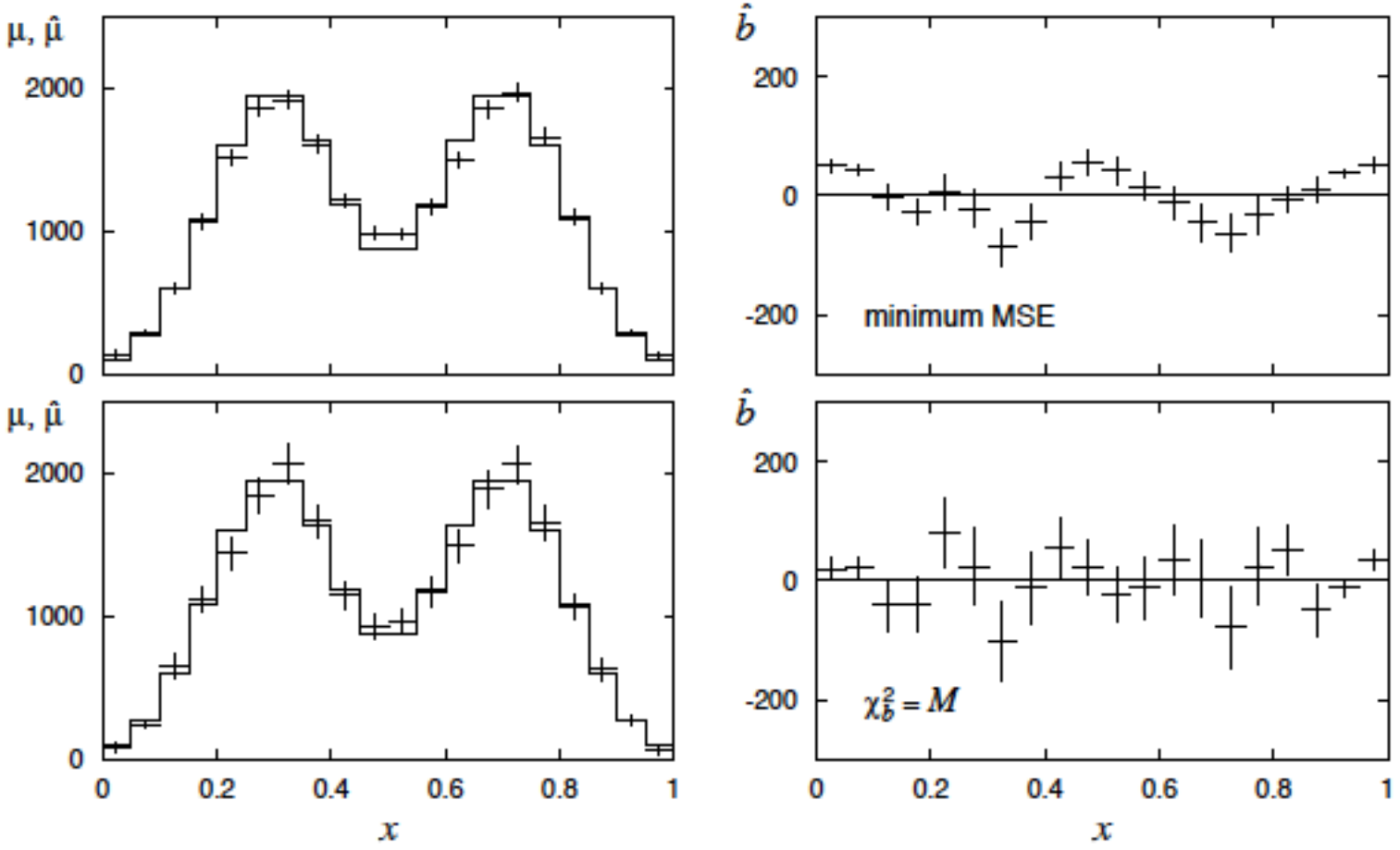
i.e. if bias significantly different from zero, we would subtract it;

$\rightarrow$ equivalent to going to smaller $\Delta \log L$ or larger α (less bias).

Some examples with Tikhonov regularization



Some examples with entropy regularization



Iterative unfolding ("Bayesian")

G. D'Agostini, NIM A **362** (1995) 487; see also arXiv:1010.0632.

Goal is to estimate probabilities: $\mathbf{p} = (p_1, \ldots, p_M)$

For initial guess take e.g. $p_i = 1/M$

Initial estimators for $\boldsymbol{\mu}$ are $\hat{\boldsymbol{\mu}}_0 = n_{\text{tot}}\,\mathbf{p}_0\,,$

Update according to the rule

$$\hat{\mu}_i = \frac{1}{\varepsilon_i}\sum_{j=1}^{N} P(\text{true value in bin } i \,|\, \text{found in bin } j)\, n_j$$

uses Bayes' theorem here

$$= \frac{1}{\varepsilon_i}\sum_{j=1}^{N}\left(\frac{R_{ij}p_i}{\sum_k R_{jk}p_k}\right) n_j$$

Continue until solution stable using e.g. χ^2 test with previous iteration.

Estimating systematic uncertainty

We know that unfolding introduces a bias, but quantifying this (including correlations) can be difficult.

Suppose a model predicts a spectrum

$$f(y;\theta) \sim 1/y^{\theta} \;\rightarrow\; \boldsymbol{\mu}(\theta)$$

A priori suppose e.g. $\theta \approx 8 \pm 2$. More precisely, assign prior $\pi(\theta)$. Propagate this into a systematic covariance for the unfolded spectrum:

$$U_{ij}^{(\theta)} = \int (\hat{\mu}_i - \mu_i(\theta))(\hat{\mu}_j - \mu_j(\theta))\, \pi(\theta)\, d\theta$$

(Typically large positive correlations.)

Often in practice, one doesn't have $\pi(\theta)$ but rather a few models that differ in spectrum. Not obvious how to convert this into a meaningful covariance for the unfolded distribution.

Stat. and sys. errors of unfolded solution

In general the statistical covariance matrix of the unfolded estimators is not diagonal; need to report full

$$U_{ij} = \mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j]$$

But unfolding necessarily introduces biases as well, corresponding to a systematic uncertainty (also correlated between bins).

This is more difficult to estimate. Suppose, nevertheless, we manage to report both U_{stat} and U_{sys}.

To test a new theory depending on parameters $\boldsymbol{\theta}$, use e.g.

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

Mixes frequentist and Bayesian elements; interpretation of result can be problematic, especially if U_{sys} itself has large uncertainty.

Folding

Suppose a theory predicts $f(y) \rightarrow \boldsymbol{\mu}$ (may depend on parameters $\boldsymbol{\theta}$).

Given the response matrix R and expected background $\boldsymbol{\beta}$, this predicts the expected numbers of observed events:

$$\boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta}$$

From this we can get the likelihood, e.g., for Poisson data,

$$L(\mathbf{n}|\boldsymbol{\nu}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

And using this we can fit parameters and/or test, e.g., using the likelihood ratio statistic

$$q = -2 \ln \frac{L(\mathbf{n}|\boldsymbol{\nu})}{L(\mathbf{n}|\hat{\boldsymbol{\nu}})} \sim \chi^2_N$$

Versus unfolding

If we have an unfolded spectrum and full statistical and systematic covariance matrices, to compare this to a model $\boldsymbol{\mu}$ compute likelihood

$$L(\hat{\boldsymbol{\mu}}|\boldsymbol{\mu}) \sim e^{-\chi^2/2}$$

where

$$\chi^2 = (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^T (U_{\text{stat}} + U_{\text{sys}})^{-1} (\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})$$

Complications because one needs estimate of systematic bias U_{sys}.

If we find a gain in sensitivity from the test using the unfolded distribution, e.g., through a decrease in statistical errors, then we are exploiting information inserted via the regularization (e.g., imposed smoothness).

ML solution again

From the standpoint of testing a theory or estimating its parameters, the ML solution, despite catastrophically large errors, is equivalent to using the uncorrected data (same information content).

There is no bias (at least from unfolding), so use

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

The estimators of $\boldsymbol{\theta}$ should have close to optimal properties: zero bias, minimum variance.

The corresponding estimators from any unfolded solution cannot in general match this.

Crucial point is to use full covariance, not just diagonal errors.

Summary/discussion

Unfolding can be a minefield and is not necessary if goal is to compare measured distribution with a model prediction.

Even comparison of uncorrected distribution with *future* theories not a problem, as long as it is reported together with the expected background and response matrix.

> In practice complications because these ingredients have uncertainties, and they must be reported as well.

Unfolding useful for getting an actual estimate of the distribution we think we've measured; can e.g. compare ATLAS/CMS.

Model test using unfolded distribution should take account of the (correlated) bias introduced by the unfolding procedure.

Summary of Lecture 3

Bayesian treatment of limits is conceptually easy (integrate posterior pdf); appropriate choice of prior not obvious.

Look-Elsewhere Effect

> Need to give probability to see a signal as big as the one you saw (or bigger) anywhere you looked. Hard to define precisely; approximate correction should be adequate.

Why 5 sigma? If LEE taken in to account, one is usually convinced the effect is not a fluctuation much earlier (at 3 sigma?)

Unfolding

Extra slides