Lecture whichever +1.

"True" distribution y_j $j = 1, \cdots, m$

"Reconstructed" distribution x_i, $i = 1, \cdots, n$

Response matrix accounts for
- inefficiency
- resolution

$$x_i = \sum_{j=1}^{m} A_{ij} y_j$$

measurements

Unfolding: given m_i and an estimate for A_{ij}, determine y_j.

Matrix inversion:

$$\chi^2 = (\vec{m} - A\vec{y})^T V^{-1} (\vec{m} - A\vec{y})$$

$$V^{-1} = \begin{pmatrix} 1/m_1 & & 0 \\ & 1/m_2 & \\ 0 & & \ddots \end{pmatrix}$$

Because $\sigma_{x_i} = \sqrt{m_i}$

$$\vec{y} = (A^T V^{-1} A)^{-1} A^T V^{-1} \vec{m}$$

But this is wildly unstable. One reason is that adjacent measurements are highly anticorrelated with $\vec{y}$.

The minimization procedure will find the best y_j that describe m_i.

Tikhonov regularization

Penalize the minimization when y_j gets too large even though $y_j + y_{j\pm1}$ is small.

$$L_1 = (\vec{m} - A\vec{y})^T V^{-1} (\vec{m} - A\vec{y})$$

$$L_2 = \tau^2 (\vec{y} - f_b \vec{y}_0)^T L^T L (\vec{y} - f_b \vec{y}_0)$$

This is very general. Consider simple cases.

1. $f_b = 0$, $L = I$.

Then $L_2 = \tau^2 |\vec{y}|^2$ and the solution is penalized for large positive (or negative) values.

2. $f_b = 1$, $L = I$.

Then $L_2 = \tau^2 |\vec{y} - \vec{y}_0|^2$

Then the solution is penalized when it deviates from a prior estimate.

3. $f_b = 0$, $L = \begin{pmatrix} 1 & -1 & 0 & \cdots \\ 0 & 1 & -1 & 0 \cdots \\ 0 & 0 & 1 & -1 \\ \vdots & & & \ddots \end{pmatrix}$

Then $L_2 = \tau^2 \sum_{j=1}^{m-1} (y_j - y_{j+1})^2$

The solution is penalized when it changes rapidly (large first derivative).

Arbitrary choices:

- Binning (require $n > m$)
- Choice of f_0, $\vec{y}_0$
- Choice of regularization function (L)
- Choice of τ^2.

General considerations:

Usually require $n \gtrsim 2m$ to provide sufficient constraints.

Test the method on Monte Carlo (simulated data).

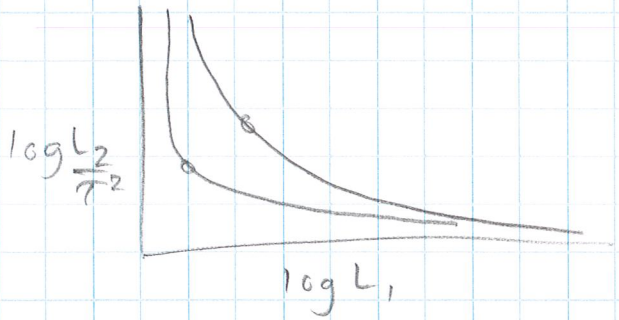
Try to demonstrate that the result is probably not biased ...

Optimization of τ^2 has been suggested via L-curve scans:

For a range of τ^2, minimize $L = L_1 + L_2$ and plot

$$\log \frac{L_2}{\tau^2} \text{ vs } \log L_1 .$$

When one or the other dominates, the curve will be straight.



Pick the point that is most sensitive to τ^2.

How else can we incorporate constraints?

$y_j < 0$ is unphysical. Instead, construct the Poisson likelihood.

$$L = \prod_i \frac{e^{-x_i} x_i^{m_i}}{m_i!}$$

$$x_i = \sum_j A_{ij} y_j .$$

$$\log L = \text{const} + m_i \log x_i - x_i$$

$$= \text{const} + \sum_i m_i \log \sum_j A_{ij} y_j - \sum_j A_{ij} y_j .$$

$$\frac{\partial \log L}{\partial y_k} = \frac{\sum_i m_i A_{ik}}{\sum_j A_{ij} y_j} - \sum_i A_{ik} = 0$$

This can be solved iteratively:

$$y_j^{(k+1)} = \frac{y_j^{(k)}}{\sum_{i=1}^{n} A_{ij}} \sum_{i=1}^{n} \frac{A_{ij} m_i}{\sum_{\ell=1}^{m} A_{i\ell} y_\ell^{(k)}}$$

and it actually converges quite rapidly.

"D'Agostini iteration".

But if you iterate too many times you become sensitive to statistical fluctuations.

Other considerations:

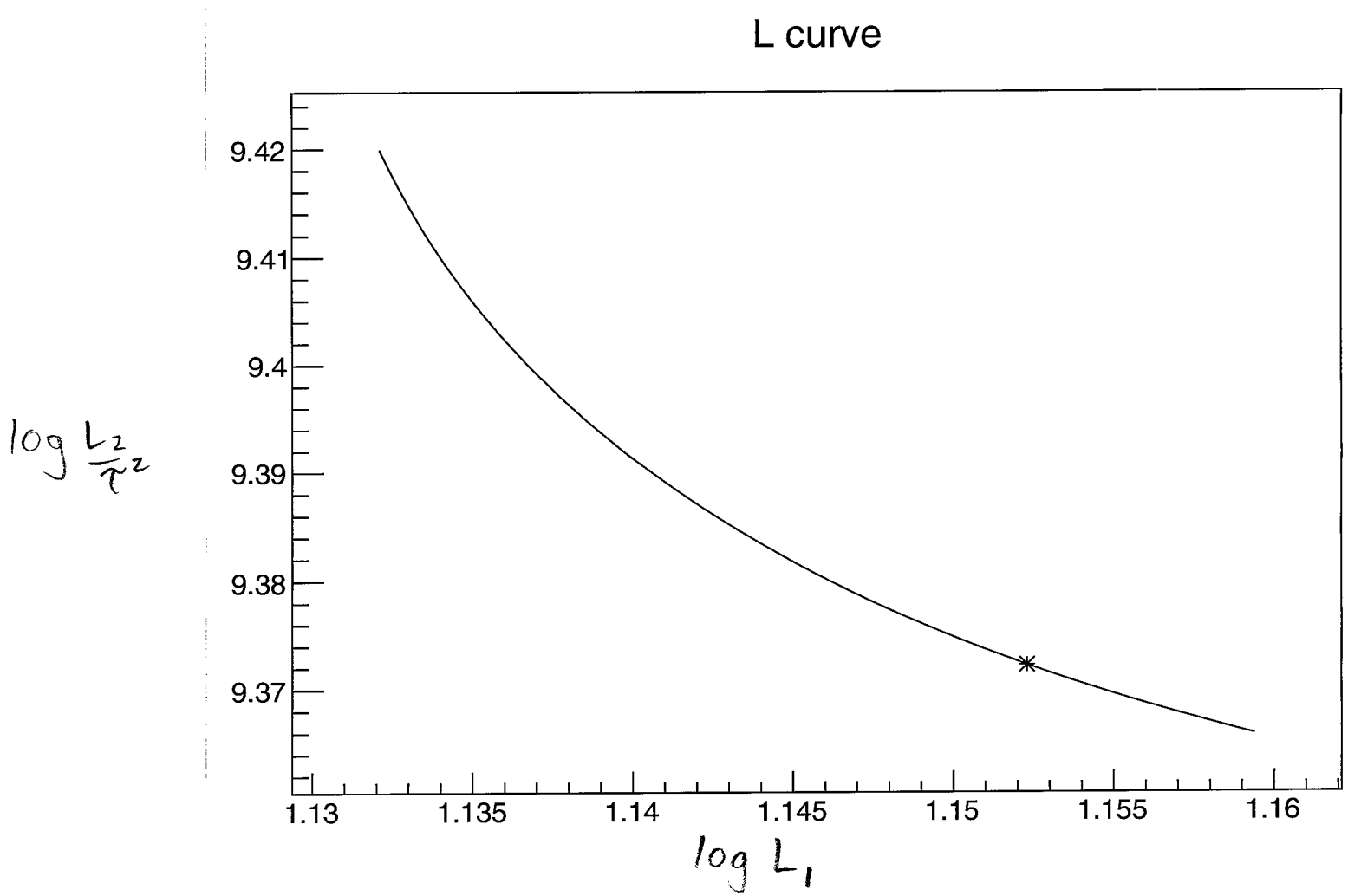
The response matrix depends on the shape of the "true" distribution used to generate it. If this doesn't match the actual (but unknown) distribution then biases can be introduced.

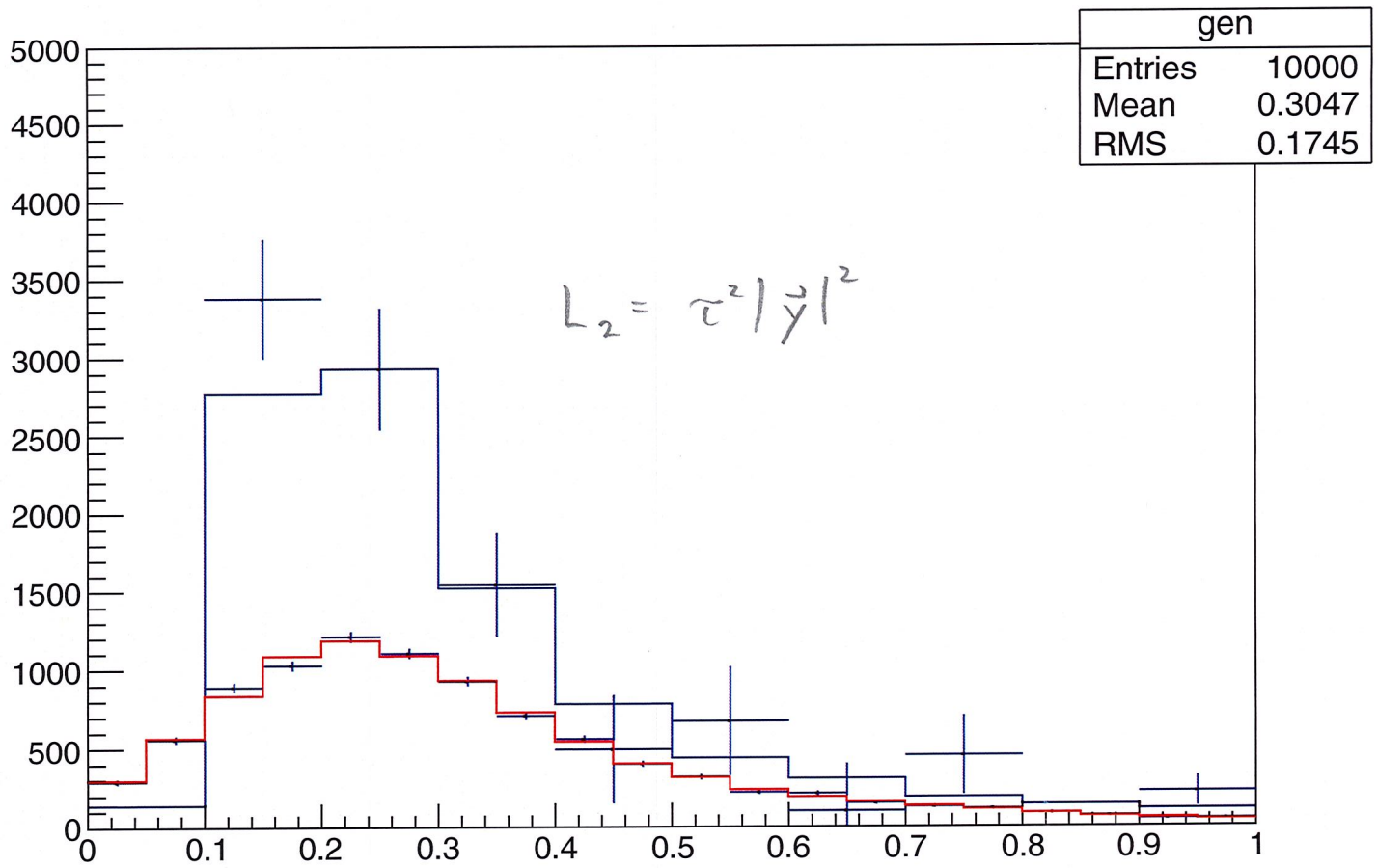
Example:

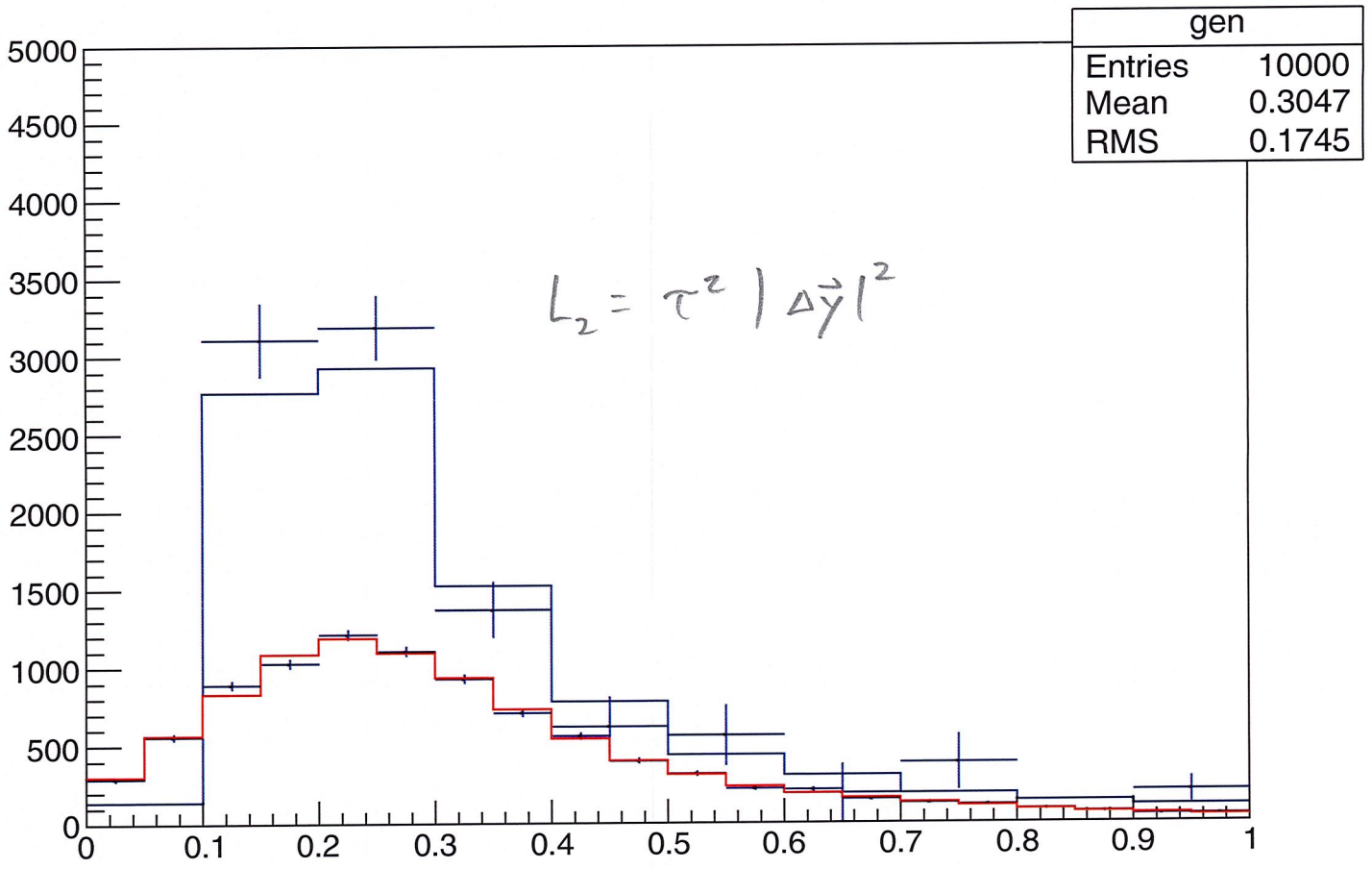
This motivates using more bins for the unfolded distribution but that reduces the number of degrees of freedom.

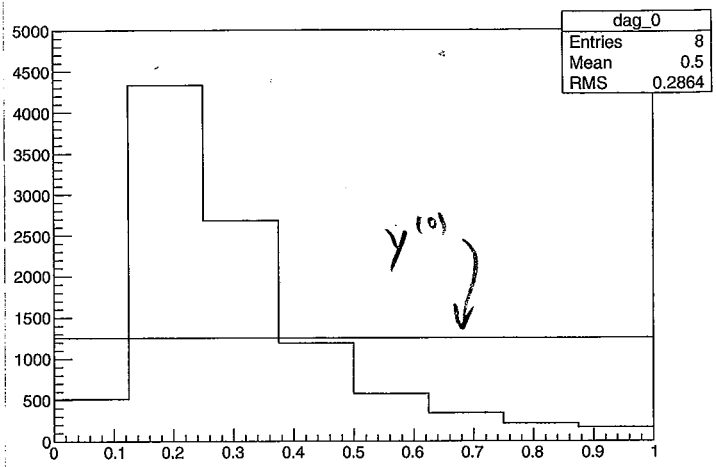
One can use the prior hypothesis for generation and iterate with subsequent estimates, but this requires making assumptions about smoothness.

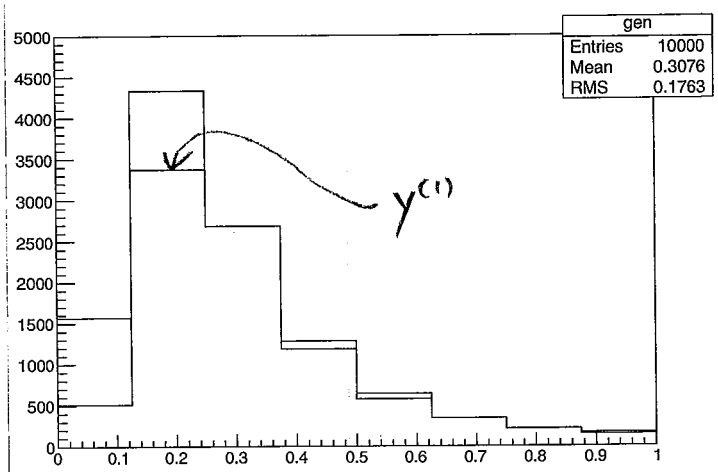
In general, unfolded distributions can have large correlations between adjacent bins. The full covariance matrix may need to be provided to fully quantify the uncertainties in the result.

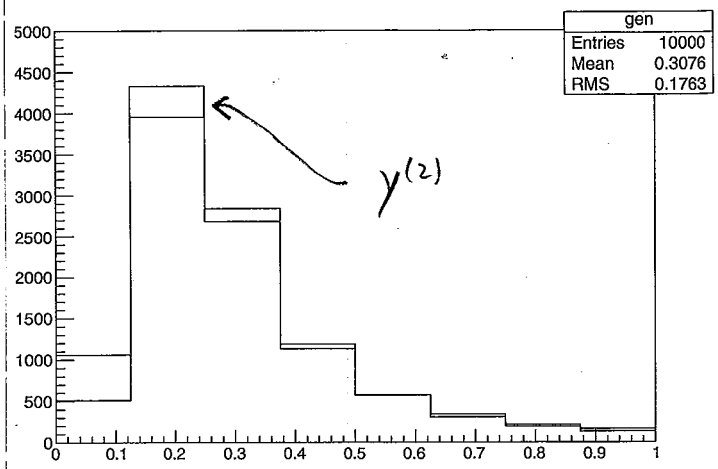

L curve
9.42
9.41
9.4
9.39
9.38
9.37
1.13
1.135
1.14
1.145
1.15
1.155
1.16
log L1
log L2/τ2

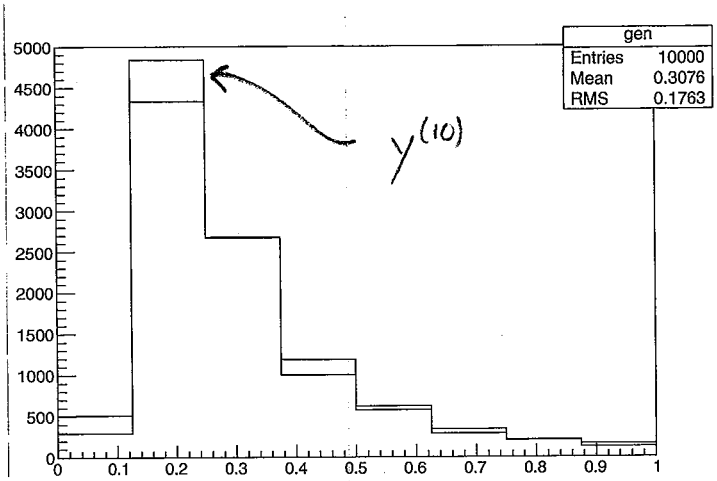

gen
Entries 10000
Mean 0.3047
RMS 0.1745
$L_2 = \tilde{\tau}^2 / \tilde{y}^2$
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


gen
Entries 10000
Mean 0.3047
RMS 0.1745
$L_2 = \tau^2 |\Delta\vec{y}|^2$
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


dag_0
Entries 8
Mean 0.5
RMS 0.2864
y(0)


gen
Entries 10000
Mean 0.3076
RMS 0.1763
y(1)


gen
Entries 10000
Mean 0.3076
RMS 0.1763
y(2)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1


gen
Entries 10000
Mean 0.3076
RMS 0.1763
y(10)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1