A Comment on "A multidimensional unfolding method based on Bayes' theorem"

Michael Betancourt

February 8, 2010

D'Agostini [1] offers a seemingly unique approach to the notoriously difficult problem of deconvolving a measured spectrum $\mathbf{n}\left(E\right)$ to recover the underlying true distribution $\mathbf{n}\left(C\right)$. Instead of focusing on the ill-posed inversion of the transfer matrix

$$n\left(E_i\right) \propto \sum_j P\left(E_i | C_j\right) n\left(C_j\right) \equiv \sum_j H_{ij} n\left(C_j\right)$$

the approach begins with the inverse problem

$$n\left(C_i\right) = \sum_j P\left(C_i | E_j\right) n\left(E_j\right) \equiv \sum_j M_{ij} n\left(E_j\right).$$

If the true distribution is known a priori, the M_{ij} are given by Bayes Theorem

$$M_{ij} = \frac{H_{ji} P\left(C_i\right)}{\sum_k H_{jk} P\left(C_k\right)}.$$

Were the true distribution $P\left(C_k\right)$ available, however, the problem would be trivial. D'Agostini offers an iterative approach. Beginning with a guess for the true distribution, the M_{ij} are computed and used to estimate the true distribution from the measured spectrum,

$$\hat{P}\left(C_i\right) = \frac{\hat{n}\left(C_i\right)}{\sum_j \hat{n}\left(C_j\right)} = \frac{\hat{n}\left(C_i\right)}{\hat{N}_{\mathrm{true}}}$$

where

$$\hat{n}\left(C_i\right) = \frac{\sum_j M_{ij} n\left(E_j\right)}{\sum_j H_{ji}} \equiv \frac{\sum_j M_{ij} n\left(E_j\right)}{\epsilon_i}$$

The $\hat{P}\left(C_i\right)$ become the new guess, and the calculation repeated.

No proof is given that the algorithm estimates $\hat{P}(C_i)$ will converge to the true distribution $P(C_i)$; D'Agostini relies entirely on simulation studies that converge to reasonable results within a few estimates. If the algorithm is allowed to continue past these initial iterations, however, the estimates begin to diverge.

This behavior can be illuminated by taking a deeper look at the update equations. First, define

$$\hat{n}(E_i) = \sum_j H_{ij} \hat{n}(C_j)$$

$$\hat{n}(E_i) = \hat{N}_{\text{true}} \sum_j H_{ij} \hat{P}(C_j).$$

D'Agostini's application of Bayes Theorem becomes

$$M_{ij} = \frac{H_{ji} P(C_i)}{\sum_k H_{jk} P(C_k)}$$

$$M_{ij} = \hat{N}_{\text{true}} \frac{H_{ji} P(C_i)}{\hat{n}(E_j)}$$

with the resulting estimates

$$\hat{P}(C_i) \propto \hat{n}(C_i)$$

$$\hat{P}(C_i) \propto \frac{1}{\epsilon_i} \sum_j M_{ij} n(E_j)$$

$$\hat{P}(C_i) \propto \frac{1}{\epsilon_i} \sum_j \hat{N}_{\text{true}} \frac{H_{ji} P(C_i)}{\hat{n}(E_j)} n(E_j)$$

$$\hat{P}(C_i) \propto \frac{1}{\epsilon_i} \left[\sum_j H_{ji} \frac{n(E_j)}{\hat{n}(E_j)} \right] P(C_i).$$

$$\hat{P}(C_i) \propto \frac{1}{\epsilon_i} \left[\sum_j H_{ji} d(E_j) \right] P(C_i).$$

where

$$d(E_i) \equiv \frac{n(E_i)}{\hat{n}(E_i)}.$$

Each $P(C_i)$ is scaled by the weighted sum of observational deviations $d(E_j)$, with the weights given by the probability of C_i contributing to E_j. In fact, if one considers the columns of the transfer matrix as probabilities then the update equation can be written as the average deviation caused by C_i,

$$\hat{P}(C_i) \propto \frac{1}{\epsilon_i} \left[\sum_j H_{ji} d(E_j) \right] P(C_i).$$

$$\hat{P}(C_i) \propto \frac{\sum_j H_{ji} d(E_j)}{\sum_j H_{ji}} P(C_i).$$

$$\hat{P}(C_i) \propto \langle d(E) \rangle_i P(C_i).$$

The corrections push the $\hat{P}(C_i)$ towards a solution consistent with the data (Fig 1). As the deviations $d(E_i)$ decrease, the corrections become smaller until the algorithm finally converges to a solution with

$$\frac{n(E_i)}{\hat{n}(E_i)} = d(E_i) = 1$$

and

$$\hat{P}(C_i) \propto \langle d(E) \rangle_i P(C_i).$$

$$\hat{P}(C_i) \propto \langle 1 \rangle_i P(C_i).$$

$$\hat{P}(C_i) \propto P(C_i).$$

If the algorithm converges to a consistent solution they why does apparently divergent behavior arise in the simulation studies presented by D'Agostini?

Consider the original problem of inverting the transfer matrix. The matrix is singular, and the inversion fails, because there are infinitely many input spectra that will produce the measured spectrum. Most of these spectra are far from the smooth distribution expected from theory.

The D'Agostini algorithm isn't diverging - it's just converging to one of the infinite irregularly shaped distributions. In the first few iterations the estimates $\hat{P}(C_i)$ feature residual smoothness from the initial guess that make is appear well behaved. As the iterations proceed, however, the memory of this smoothness vanishes in favor of an estimate better fitting the statistical fluctuations in the data.

From this perspective, the D'Agostini algorithm is basically an iterative matrix inversion algorithm (or pseudo inversion if the transfer matrix isn't square), vulnerable to all of the problems it claimed to have solved.

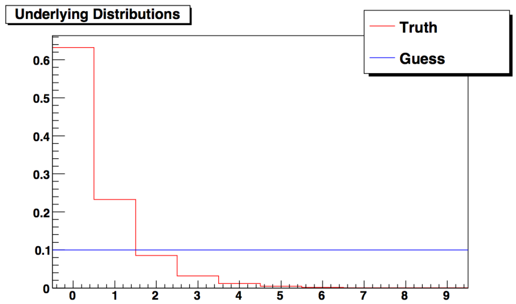
Modern Bayesian approaches [2] [3], place priors on the smoothness of the underlying distribution, regularizing the problem and providing well behaved solutions consistent with the underlying true distribution.

References

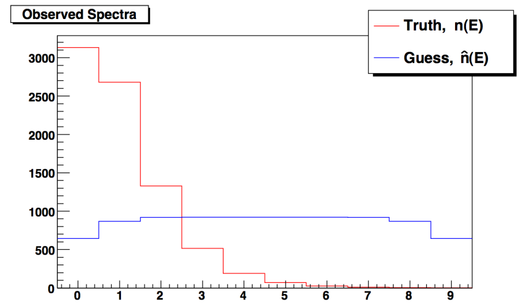
[1] D'Agostini, G. *Nucl. Instr. and Methods A* 362:487 (1995)

[2] Sivia, D. S. with Skilling, J. (2006) *Data Analysis.* Oxford, New York

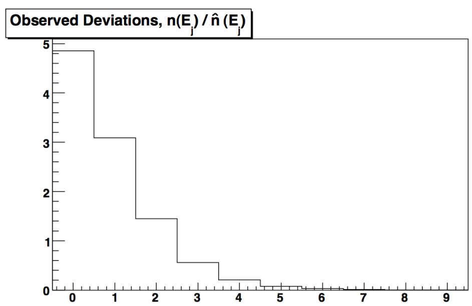
[3] Skilling, J. (1989). Classic maximum entropy. In *Maximum Entropy and Bayesian methods* (ed J. Skilling). Kluwer, Dordrecht.



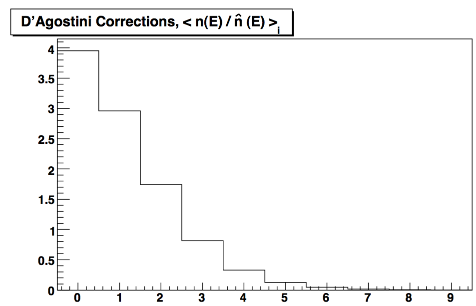
(a)



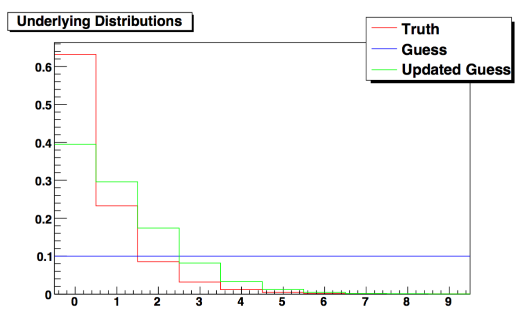
(b)



(c)



(d)



(e)

Figure 1: Example of the D'Agostini algorithm in practice. (a) One starts with an initial guess that may have no resemblance to the true distribution. (b) The observed spectrum consistent with the guess is compared to the measured spectrum and (c) the deviations $d(E)$ computed. (d) The deviations are used to compute the corrections, which (e) give the updated guess.