A SURVEY OF UNFOLDING METHODS FOR PARTICLE PHYSICS

G. Cowan
Physics Department, Royal Holloway, University of London, UK

Abstract

A pedagogical introduction to the problem of unfolding probability distributions in particle physics is given. Several of the most commonly used methods are reviewed and compared.

1 INTRODUCTION

In this paper the problem of estimating probability distributions is reviewed in cases where no parametric form is available, and where the data are subject to additional random fluctuations due to limited resolution. In High Energy Physics this estimation procedure is usually called *unfolding*, although the same mathematics can be found under the general heading of *inverse problems*, and is also called *deconvolution* or *unsmearing*. The presentation here is based mainly on Ref. [1]. Related methods are treated in greater detail by V. Blobel elsewhere in these proceedings [2]. Further discussions of unfolding in High Energy Physics can be found in [3, 4, 5, 6].

In Section 2 the mathematical problem is formulated and notation defined. Alternatives to unfolding are considered in Section 3. Unfolding by inversion of the response matrix is discussed in Section 4, and a simple method based on correction factors is shown in Section 5. The main topic of this paper, regularized unfolding, is described in Section 6. This includes a survey of several regularization functions, and methods for estimating the variance and bias of the solution. An iterative method by D'Agostini is considered in Section 7, and further examples are shown in Section 8.

2 THE MATHEMATICAL PROBLEM

Suppose we wish to determine the probability density function (pdf) $f(y)$ of a random variable y using a sample of data $y_1, \ldots, y_n$. If a parametric form for the pdf is known, i.e., $f(y; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of parameters, then standard techniques such as the method of maximum likelihood can be used to obtain estimators $\hat{\boldsymbol{\theta}}$.

If no parametrization is available, we can simply construct a histogram of y with M bins. The value μ_i represents the expectation value of the number of entries in bin i of the histogram. This is often referred to loosely as the 'true histogram'. The sum is denoted by $\mu_{\rm tot} = \sum_i \mu_i$, and the probability for y to be found in bin i is thus $p_i = \mu_i / \mu_{\rm tot}$.

The goal of unfolding as formulated here is to construct estimators for the M parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$, or alternatively for the probabilities $\mathbf{p}$. The non-trivial aspect of this problem arises when the measured values of y are subject to further random fluctuations because of measurement errors. That is, each observation is characterized by a true (and unknown) value y and by a measured value x. The pdfs for x and y are related by a convolution,

$$f_{\rm meas}(x) = \int R(x|y)\, f_{\rm true}(y)\, dy \,, \tag{1}$$

where $R(x|y)$ is the response function. Here we will assume that this depends only on the measuring apparatus and is known. When we consider the usual case where the pdfs of x and y are both represented as histograms, equation (1) becomes

$$\nu_i = \sum_{j=1}^{M} R_{ij} \mu_j \,, \qquad i = 1, \ldots, N \,, \tag{2}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_{\mathbf{1}}, \ldots, \boldsymbol{\mu}_{\mathbf{M}})$ gives the expectation values for the histogram of y and $\boldsymbol{\nu} = (\boldsymbol{\nu}_{\mathbf{1}}, \ldots, \boldsymbol{\nu}_{\mathbf{N}})$ gives the expected number of events in bins of the observed variable x. The actual data are given as a vector of numbers $\mathbf{n} = (n_1, \ldots, n_N)$. These represent the actual (integer) number of entries observed in the histogram, which of course differ in general from the (non-integer) expectation values $\boldsymbol{\nu}$.

The *response matrix* R has the simple interpretation as a conditional probability:

$$R_{ij} = P(\text{observed in bin } i \,|\, \text{true value in bin } j) \ . \tag{3}$$

By summing R_{ij} over all possible bins of the observed value i, one obtains

$$\sum_{i=1}^{N} R_{ij} = P(\text{observed anywhere} \,|\, \text{true value in bin } j) = \varepsilon_j \ . \tag{4}$$

This gives the *efficiency*, ε_j, which depends in general on the bin j of the true histogram. A summary of the ingredients so far is illustrated in Figs. 1.

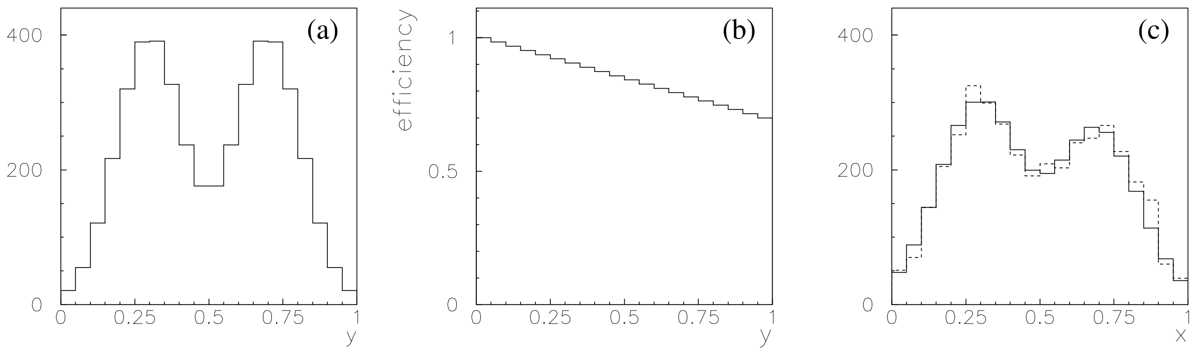

Fig. 1: Illustration of ingredients for unfolding: (a) a 'true histogram' $\boldsymbol{\mu}$, (b) a possible set of efficiencies $\boldsymbol{\varepsilon}$, and (c) the observed histogram $\mathbf{n}$ (dashed) and the corresponding expectation values $\boldsymbol{\nu}$ (solid).

In general, equation (2) must be modified to include the expected number of background events in bin i, β_i. The data $\mathbf{n}$, the corresponding expectation values $\boldsymbol{\nu}$, the response matrix R, the expectation values for the true histogram $\boldsymbol{\mu}$, and the expected number of background events $\boldsymbol{\beta}$ are finally related by

$$E[\mathbf{n}] = \boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta} \ . \tag{5}$$

3 WHY UNFOLD?

Before proceeding it should be emphasized that in many problems it is not necessary to unfold the measured distribution, in particular if the goal is to compare the result with the prediction of an existing theory. In that case one can simply modify the prediction $\boldsymbol{\mu}$ to include the distortions of the detector, and this can be directly compared with the measurement. That is, one finds $\boldsymbol{\nu}$ and compares this to the data $\mathbf{n}$. This procedure is considerably simpler than unfolding the measurement and comparing it with the original (unmodified) theory.

Without unfolding, however, the measurement cannot be compared with the results of other experiments, for which the response matrix will in general be different. It can also happen that a new theory is developed many years after a measurement has been carried out, and the response matrix may no longer be available. If a particularly important measured distribution is to retain its value, then both the data and the response matrix should be preserved. Unfortunately, this is often impractical.

By unfolding, i.e., constructing estimators $\hat{\boldsymbol{\mu}}$, one obtains a result which can directly be compared with those of other experiments as well as with theoretical predictions. Other reasons for unfolding exist in applications such as image reconstruction, where certain features may not be recognizable in the uncorrected distribution. In this paper we will assume that these arguments have been considered and that the decision has been made to unfold.

4 INVERTING THE RESPONSE MATRIX

Suppose the response matrix can be inverted (generally possible in practice) so we can write

$$\boldsymbol{\mu} = R^{-1}(\boldsymbol{\nu} - \boldsymbol{\beta}) \ , \tag{6}$$

and suppose further that the components of the data vector $\mathbf{n}$ are independent and Poisson distributed,

$$P(n_i; \nu_i) = \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i} \ . \tag{7}$$

It is then simple to show that the maximum likelihood (ML) estimators for $\boldsymbol{\nu}$ are $\mathbf{n}$, and thus we can take as estimators for $\boldsymbol{\mu}$,

$$\hat{\boldsymbol{\mu}} = R^{-1}(\mathbf{n} - \boldsymbol{\beta}) \ . \tag{8}$$

If the off-diagonal elements of the response matrix are too large, i.e., if the bin size is too small compared to the measurement resolution, then the ML estimators $\hat{\boldsymbol{\mu}}$ from (8) have catastrophically large variances and strong negative correlations between neighbouring bins. This is illustrated in the Fig. 2 (in this example the efficiencies are taken to be unity).

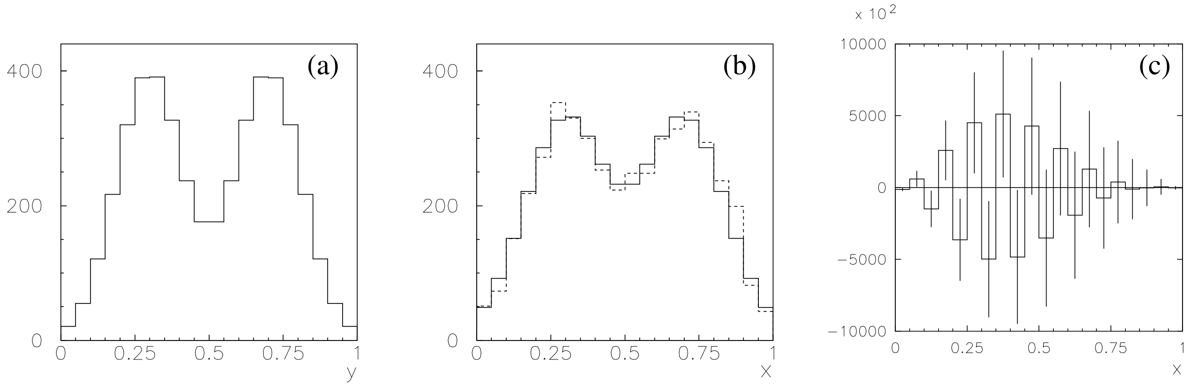

Fig. 2: Attempt to unfold using matrix inversion: (a) the 'true histogram', (b) the observed histogram $\mathbf{n}$ (dashed) and corresponding expectation values $\boldsymbol{\nu}$ (solid), (c) the estimators $\hat{\boldsymbol{\mu}}$ based on equation (8).

The reason for the large fluctuations can be understood qualitatively by considering what would happen if the true distribution $\boldsymbol{\mu}$ really did have a high degree of fine structure, as in Fig. 3(a). Applying the response matrix to obtain the expectation values for the observed histogram, $\boldsymbol{\nu}$, may wash out most of the structure although some remnants of the peaks remain, as shown in Fig. 3(b).

When we apply R^{-1} to the expectation values $\boldsymbol{\nu}$ we get back to $\boldsymbol{\mu}$. But of course we do not have $\boldsymbol{\nu}$ when we make a measurement, we only have the data $\mathbf{n}$. The actual data histogram $\mathbf{n}$ does indeed have small bumps from statistical fluctuations which, as far as applying R^{-1} is concerned, produce the same effect as if they were the residual effect of some initial fine structure. That is, the estimators $\hat{\boldsymbol{\mu}}$ from (8) wind up with large oscillations between neighbouring bins.

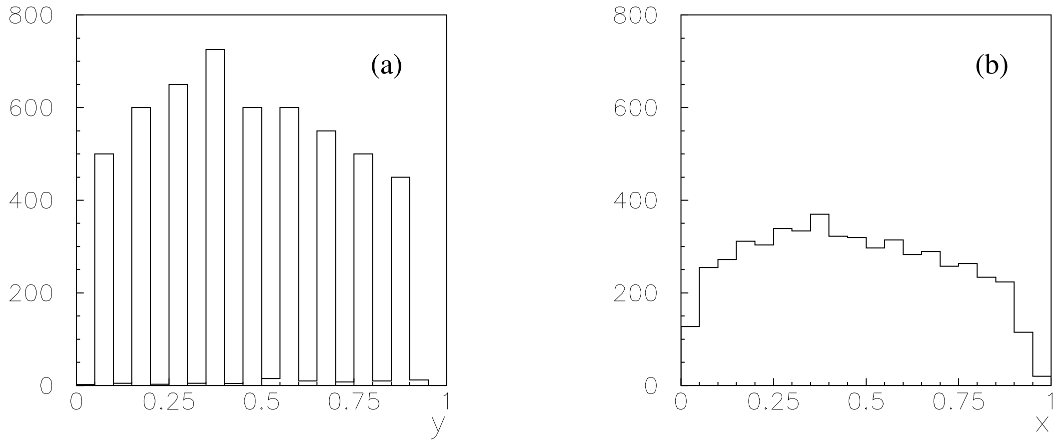

Fig. 3: (a) A true histogram $\boldsymbol{\mu}$ with a high degree of fine structure. (b) The corresponding expectation values for the observed histogram, $\boldsymbol{\nu} = R\boldsymbol{\mu}$, with most of the structure washed out.

Before discarding the ML estimators (8) as useless it is interesting to note that they are in fact unbiased, since

$$E[\hat{\boldsymbol{\mu}}] = R^{-1}(E[\mathbf{n}] - \boldsymbol{\beta}) = \boldsymbol{\mu} \,. \tag{9}$$

Furthermore their covariance is

$$U_{ij} = \mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = \sum_{k,l=1}^{N} (R^{-1})_{ik}(R^{-1})_{jl}\mathrm{cov}[n_k, n_l] = \sum_{k=1}^{N}(R^{-1})_{ik}(R^{-1})_{jk}\nu_k \,, \tag{10}$$

since $\mathrm{cov}[n_k, n_l] = \delta_{kl}\nu_k$ for the case where the n_i are independent Poisson variables. The minimum variance bound for the present case of zero bias,

$$(U^{-1})_{kl} = -E\left[\frac{\partial^2 \ln L}{\partial \mu_k \mu_l}\right] = \sum_{i=1}^{N} \frac{R_{ik}R_{il}}{\nu_i} \,, \tag{11}$$

gives, however, exactly the same covariance as equation (10). That is, the estimators obtained from inverting the response matrix have the smallest possible variance among all unbiased estimators, even though this variance was seen to be huge. So any estimators we construct that in some way reduce the variance will necessarily have a bias. The strategy we will follow is to accept a small bias (systematic error) in exchange for a large reduction in variance (statistical error).

5 CORRECTION FACTORS

A relatively simple method to construct estimators with much smaller variances is based on multiplicative correction factors derived from Monte Carlo simulations. The estimator for bin i of the unfolded distribution is taken to be

$$\hat{\mu_i} = C_i(n_i - \beta_i) \,, \tag{12}$$

where the correction factor C_i is determined by the ratio

$$C_i = \frac{\mu_i^{\mathrm{MC}}}{\nu_i^{\mathrm{MC}}} \,. \tag{13}$$

Here $\nu_i^{\rm MC}$ and $\mu_i^{\rm MC}$ are obtained from Monte Carlo models for the signal process only, since background is subtracted separately. The covariance matrix is

$$U_{ij} = \mathrm{cov}[\hat{\mu}_i, \hat{\mu}_j] = C_i^2 \mathrm{cov}[n_i, n_j] \,. \tag{14}$$

Thus as long as the correction factors themselves are of order unity, the variances of the estimators do not become much larger than what is dictated by the Poisson statistics of the data.

The bias of the estimators, $b_i = E[\hat{\mu}_i] - \mu_i$, is found to be

$$b_i = \left(\frac{\mu_i^{\rm MC}}{\nu_i^{\rm MC}} - \frac{\mu_i}{\nu_i^{\rm sig}} \right) \nu_i^{\rm sig} \,, \tag{15}$$

where $\nu_i^{\rm sig} = \nu_i - \beta_i$. Thus there is a nonzero bias unless the model is correct, which of course is not known to be true before making the measurement. In fact the bias often tends to pull the estimators towards the predictions of the Monte Carlo model used for the correction factors, which makes it particularly difficult to test the model. Such tests are better carried out by a direct comparison of the data $\mathbf{n}$ and the model's predicted $\boldsymbol{\nu}$. If correction factors are used, an estimate of the bias must be included in the systematic uncertainty of the measurement, and this should be taken into account in any model tests.

6 REGULARIZED UNFOLDING

We have seen that the estimators based on inversion of the response matrix can have extremely large variances. These estimators are what one obtains from the methods of maximum likelihood or least squares. The log-likelihood quantifies in some sense the 'distance' between the data $\mathbf{n}$ and the predicted expectation values $\boldsymbol{\nu}$. (If using the method of least squares one typically has $\chi^2 = -2 \ln L$.) We can then consider the region of $\boldsymbol{\mu}$-space around the ML solution where the log-likelihood is within some prespecified $\Delta \ln L$ of its maximum, i.e.,

$$\ln L(\boldsymbol{\mu}) \geq \ln L_{\max} - \Delta \ln L \,. \tag{16}$$

Estimators for $\boldsymbol{\mu}$ can then be constructed by choosing the smoothest solution out of this region, according to some appropriate measure of smoothness.

One can easily show (e.g., with a Lagrange multiplier) that this approach is equivalent to maximizing not the log-likelihood but rather

$$\Phi(\boldsymbol{\mu}) = \alpha \ln L(\boldsymbol{\mu}) + S(\boldsymbol{\mu}) \,. \tag{17}$$

Here $S(\boldsymbol{\mu})$ is the *regularization function*, which represents the smoothness of the distribution, and α is the *regularization parameter*, which can be chosen to correspond to a given $\Delta \ln L$. The regularization parameter can equivalently be taken to multiply S rather than $\ln L$, as is done in some other references.

One may wish to insist that the unfolded solution provide an unbiased estimate for the total number of entries, i.e., one wants $\sum_i \nu_i = \sum_i n_i = n_{\rm tot}$. This can be achieved by maximizing

$$\varphi(\boldsymbol{\mu}, \lambda) = \alpha \ln L(\boldsymbol{\mu}) + S(\boldsymbol{\mu}) + \lambda \left[n_{\rm tot} - \sum_{i=1}^{N} \nu_i \right] \,, \tag{18}$$

where λ is a Lagrange multiplier. Note that the last term is effectively a function of $\boldsymbol{\mu}$ since $\boldsymbol{\nu} = R\boldsymbol{\mu} + \boldsymbol{\beta}$ always holds.

To carry out this procedure we therefore need a regularization function, $S(\boldsymbol{\mu})$, and a prescription for determining the regularization parameter α. These will determine the bias and variance of the unfolded distribution, which can be estimated by expanding the estimators about the estimates obtained from the actual data (see [1]). This gives estimates for the bias, $b_i = E[\hat{\mu}_i] - \mu_i$, of the form

$$\hat{b}_i = \sum_{j=1}^{N} \frac{\partial \hat{\mu}_i}{\partial n_j} (\hat{\nu}_j - n_j) \,, \tag{19}$$

where $\hat{\boldsymbol{\nu}} = R\hat{\boldsymbol{\mu}} + \boldsymbol{\beta}$. Recall that in general we do not have $\hat{\boldsymbol{\nu}} = \mathbf{n}$.

6.1 Choosing the regularization parameter

The regularization parameter α determines the relative weight placed on the data (through $\ln L$) compared to the degree of smoothness (through S). Taking $\alpha = 0$ gives maximally smooth estimators, which do not depend at all on the data. They therefore have zero variance (!) but a clear bias. Taking α very large gets us back to the highly oscillating ML solution with zero bias.

Many possible criteria can be used to define the trade-off between bias and variance. For example, one could choose α to minimize the mean squared error averaged over the bins of the histogram,

$$\mathrm{MSE} = \frac{1}{M} \sum_{i=1}^{M} (U_{ii} \,+\, \hat{b}_i^2) \,. \tag{20}$$

Alternatively one could argue that the contribution to the mean squared error should be different for different bins depending on how accurately they are measured. Since the variance of a Poisson variable with mean value μ_i is equal to μ_i, one can define a weighted MSE,

$$\mathrm{MSE}' = \frac{1}{M} \sum_{i=1}^{M} \frac{U_{ii} \,+\, \hat{b}_i^2}{\hat{\mu}_i}, \tag{21}$$

One could, for example, require that the estimated bias squared, $\hat{b}_i^2$, be not larger than its own estimated variance, $\widehat{W}_{ii}$, where, $W_{ii} = \mathrm{cov}[\hat{b}_i, \hat{b}_i]$. This can be achieved by finding the value of α for which

$$\chi_b^2 = \sum_{i=1}^{M} \hat{b}_i^2 / \widehat{W}_{ii} \;=\; M \,. \tag{22}$$

The rationale behind this prescription is that if one had a statistically significant bias, i.e., with $\hat{b}_i^2 \gg \widehat{W}_{ii}$, then this could be subtracted. But to obtain estimators with smaller bias it is equivalent to go to a larger value of α.

Other criteria based on the eigenvalues of the response matrix are discussed in [2, 7]. Further prescriptions for determining the regularization parameter are discussed in [1, 8].

6.2 The regularization function

The user must define what constitutes a 'smooth' distribution by specifying a regularization function $S(\boldsymbol{\mu})$. One possibility is to take the mean square of the second derivative, which is clearly related to the amount of curvature. Using finite differences for the derivatives, one has

$$S(\boldsymbol{\mu}) = -\sum_{i=1}^{M-2} (-\mu_i + 2\mu_{i+1} - \mu_{i+2})^2 \,. \tag{23}$$

Functions based on other derivatives of the distribution can also be used. This approach is often called *Tikhonov* regularization [9, 10]. It has been implemented in programs widely used in High Energy Physics such as RUN by Blobel [2, 3, 11] and GURU by Höcker and Kartvelishvili [7].

An alternative measure of smoothness is based on the entropy, which for a set of probabilities $\mathbf{p} = (p_1, \ldots, p_M)$ is

$$H = -\sum_{i=1}^{M} p_i \ln p_i \ . \tag{24}$$

One can easily show that this is a maximum for all p_i equal (i.e., maximal smoothness) and it is a minimum for one $p_i = 1$ and all the rest zero. The entropy can be used as the regularization function $S(\boldsymbol{\mu})$ by substituting $p_i = \mu_i/\mu_{\rm tot}$ in (24).

This approach is often called the method of maximum entropy or MaxEnt, although of course it is a combination of entropy and log-likelihood that is maximized. It can be motivated through Bayesian statistics by relating the entropy to the prior probability for $\boldsymbol{\mu}$. Some drawbacks of this point of view are discussed in [1]. MaxEnt estimation is often used in astronomical image processing [12].

7 AN ITERATIVE METHOD

A different approach to unfolding is taken in the iterative method proposed by D'Agostini [13], which contains elements of Bayesian statistics. Here one starts with a set of initial probabilities $\mathbf{p} = (p_1, \ldots, p_M)$ for an event to be found in each bin. In the absence of further information one can take $p_i = 1/M$ for bins of equal size. Initial estimators for $\boldsymbol{\mu}$ are

$$\hat{\boldsymbol{\mu}}_0 = n_{\rm tot}\mathbf{p}_0 \ , \tag{25}$$

where $n_{\rm tot} = \sum_i n_i$ is the total observed number of entries. These estimators are updated using the rule

$$\hat{\mu}_i = \frac{1}{\varepsilon_i} \sum_{j=1}^{N} P(\text{true value in bin } i | \text{ found in bin } j)\, n_j = \frac{1}{\varepsilon_i} \sum_{j=1}^{N} \left(\frac{R_{ij} p_i}{\sum_k R_{jk} p_k} \right) n_j \ . \tag{26}$$

Here Bayes' theorem has been used to write the conditional probability that the event originated in bin i given that it was observed in bin j in terms of the response matrix R and the prior probabilities $\mathbf{p}$.

The updated estimators can then be compared to those of the previous iteration by, for example, using a χ^2 test. If the χ^2 is too large, the procedure can be iterated with the new prior probabilities taken as the solution at the previous step, i.e., $\mathbf{p} = \hat{\boldsymbol{\mu}}/\mu_{\rm tot}$. In practice this is found to converge to a reasonable solution in several iterations.

One must not of course simply iterate until the solution looks right and then stop. The number of iterations should be decided upon before looking at the actual data, for example by using Monte Carlo test data. Continuing to iterate brings increasingly large variances and the estimators eventually approach the oscillating solution from matrix inversion.

As the procedure uses Bayes' theorem in an intermediate step it has been called a 'Bayesian method'. This should imply, however, that the estimators summarize in some way a joint posterior pdf for $\boldsymbol{\mu}$, and one would need to specify a joint prior for $\boldsymbol{\mu}$ as well. That is, in Bayesian statistics one relates the joint posterior pdf for $\boldsymbol{\mu}$ given the data $\mathbf{n}$, $p(\boldsymbol{\mu}|\mathbf{n})$, to the likelihood (probability of $\mathbf{n}$ given the hypothesis $\boldsymbol{\mu}$) and the prior probability $\pi(\boldsymbol{\mu})$:

$$p(\boldsymbol{\mu}|\mathbf{n}) \propto L(\mathbf{n}|\boldsymbol{\mu})\pi(\boldsymbol{\mu}) \ . \tag{27}$$

Bayesian estimators could be defined as, for example, the mode of $p(\boldsymbol{\mu}|\mathbf{n})$. This is not the case in the iterative method. The $\mathbf{p}$ above are initial probabilities for an event to be in each bin, i.e., they are initial guesses for $\boldsymbol{\mu}/\mu_{\rm tot}$; they do not represent a probability density in $\boldsymbol{\mu}$-space. As the Bayesian aspect of the procedure is confined to particular step, the term 'iterative method' is more appropriate.

8 SOME EXAMPLES

Figure 4 shows the result of unfolding the test distribution considered previously in Section 4. The left-hand plots show the true distribution as a solid histogram and the estimators as points with error bars. The right-hand plots show the estimated biases from equation (19); note these are not simply the residuals. In (a) the regularization parameter was determined by finding the minimum mean squared error, and in (b) the criterion $\chi_b^2 = M$ was used. Note that in (a) some of the biases are significantly different from zero, whereas in (b) they are on average within one standard deviation of zero by construction.

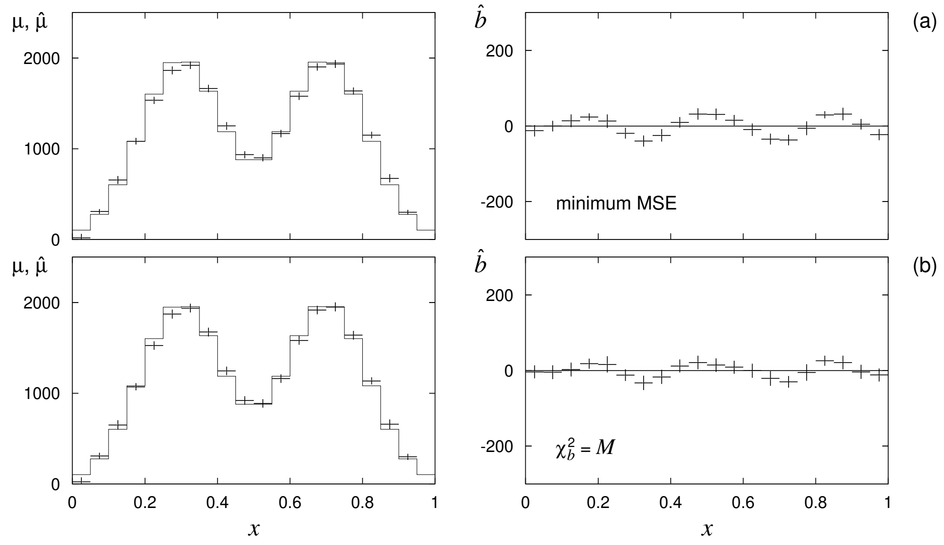

Fig. 4: Unfolded distributions using Tikhonov regularization shown as points with the true distribution shown as a histogram (left) and the estimated biases (right). The regularization parameter α was determined in (a) by the minimum mean squared error, and in (b) using $\chi_b^2 = M$ (see text).

A particular feature of Tikhonov regularization is that the solution is not forced to be positive. In fact in this example the right-most bin has a negative estimate. With the regularization function is based on entropy, however, the solution is forced to be positive. This can be seen in the corresponding set of plots in Fig. 5.

Unfolding has been widely used in high energy physics for structure functions, since there one usually wants to combine results from different experiments and later to use the structure functions to predict other cross sections. One often performs many checks with Monte Carlo test data in order to investigate how the estimators from different methods will differ and how they depend on the regularization parameter.

As an example, Figure 6(a) from the OPAL experiment [14] shows a test unfolding of the photon structure function using the programs `GURU` [7], `RUN` [11] and the iterative method (`'BAYES'`) [13]. Although the three methods agree reasonably well, the spread in the results is not negligible. Figure 6(b) shows the unfolded results from the program `GURU` corresponding to three different values of the regularization parameter (labelled by the effective parameter `NDF`). Higher values of `NDF` corresponds to a less regularized solution and therefore to larger variance, as can be seen from the size of the error bars.

Much work on unfolding has taken place in other fields, where it is usually called deconvolution. Examples of deconvolution for image reconstruction from astronomy [12, 15], medical imaging [16] and

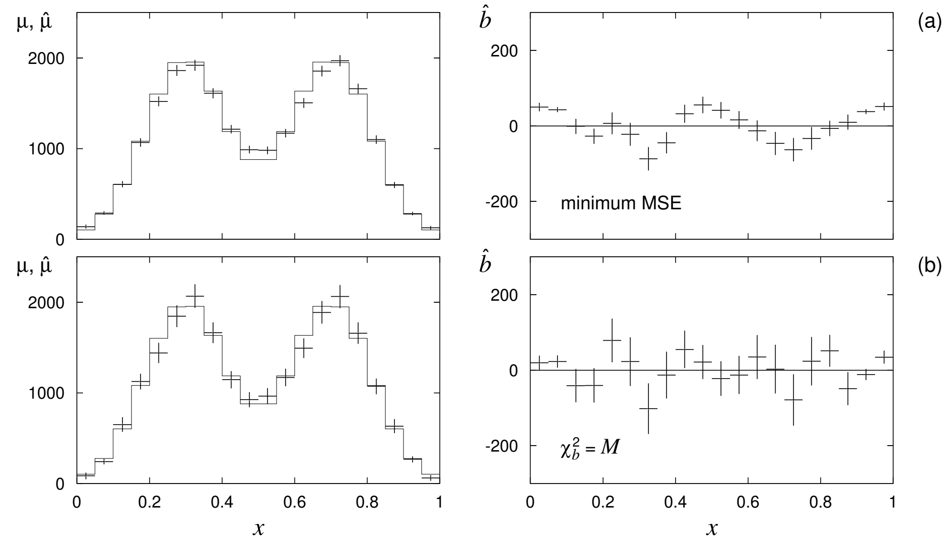

Fig. 5: Unfolded distributions using MaxEnt regularization shown as points with the true distribution shown as a histogram (left) and the estimated biases (right). The regularization parameter α was determined in (a) by the minimum mean squared error, and in (b) using $\chi_b^2 = M$ (see text).

even for digital archiving of works of art [17] are in some ways mathematically more challenging than the corresponding problems in particle physics, since the number of pixels in an image is in general much greater than the number of bins in a one-dimensional distribution.

9 CONCLUSIONS

Unfolding is a complicated business and one is well advised to ask in each problem if it can be avoided. There are nevertheless many problems where an unfolded distribution is needed. In choosing a technique one has to find a trade-off between bias and variance, although there is no clear winner for how the optimum should be achieved. Simplicity is also a consideration, and in many applications a technique such as multiplicative correction factors may be sufficient. Deconvolution techniques have been developed in many fields outside particle physics, particularly in image reconstruction. My glance at those efforts has only been sufficient to convince me that we can learn a great deal from them.

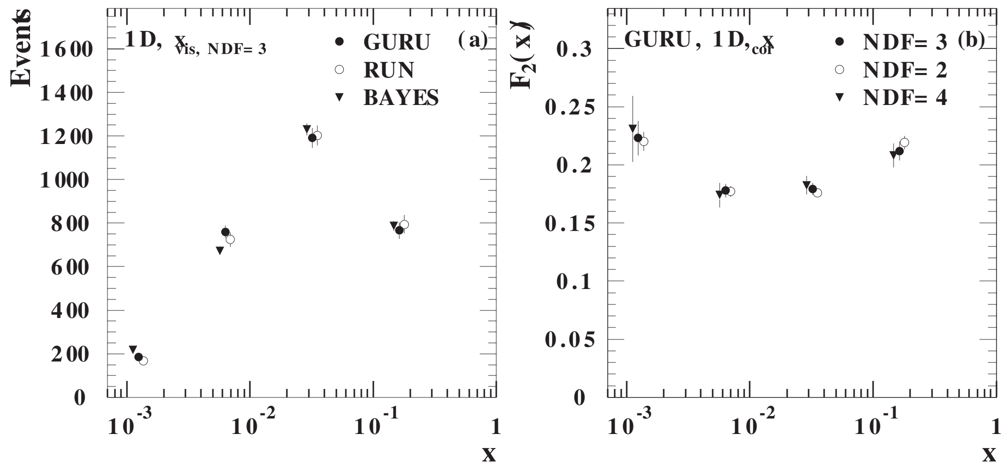

Fig. 6: Unfolding tests from OPAL's measurement of the photon structure function [14]. (a) The results of different unfolding programs. (b) The results using different values for the regularization parameter.

ACKNOWLEDGEMENTS

I would like to thank Louis Lyons, Fred James, James Stirling, Mike Whalley and Linda Wilkinson for a very interesting and enjoyable conference. I benefited greatly from discussions on unfolding with Michael Goldstein and Volker Blobel.

References

[1] G. Cowan, Statistical Data Analysis, Oxford University Press (1998).

[2] V. Blobel, An Unfolding Method for High Energy Physics Experiments, in these proceedings.

[3] V. Blobel, Unfolding Methods in High Energy Physics, DESY 84-118 (1984); also CERN 85-02 (1985).

[4] G. Zech, Comparing statistical data to Monte Carlo simulation – parameter fitting and unfolding, DESY 95-113 (1995).

[5] V.B. Anykeev, A.A. Spiridonov, V.P. Zhigunov, Comparative Investigation of Unfolding Methods, NIM A **303** (1991) 350.

[6] R. Barlow, SLUO Lectures on Numerical Methods in HEP (Lecture 9: Unfolding), available at `www.hep.man.ac.uk/people/roger`, (2000).

[7] A. Höcker and V. Kartvelishvili, NIM A **372** (1996) 469; code available at `www.hep.man.ac.uk/~vato`.

[8] Michael Schmelling, The method of reduced cross-entropy. A general approach to unfold probability distributions, Nucl. Instrum. Methods **A340** (1994) 400.

[9] A.N. Tikhonov, On the solution of improperly posed problems and the method of regularization, Sov. Math. **5** (1963) 1035.

[10] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM **9** (1962) 84.

[11] V. Blobel, The RUN manual: regularized unfolding for high-energy physics experiments, OPAL Technical Note TN361, March 8, 1996; code available from `www.desy.de/~blobel/wwwrunf.html`.

[12] R. Narayan and R. Nityananda, Maximum entropy image restoration in astronomy, Ann. Rev. Astron. Astrophys. **24** (1986) 127.

[13] G. D'Agostini, NIM A **362** (1995) 487; see also G. D'Agostini, Comments on the Bayesian unfolding, Zeus Note 95-166 (1995).

[14] G. Abbiendi *et al.* (OPAL), Eur. Phys. J. C **18** (2000) 15.

[15] K. Bouyoucef *et al.*, Astron. Astrophys. Suppl. Ser. **121** (1997) 575.

[16] A.K. Louis, Medical imaging: state of the art and future development, Inverse Probl. **8** (1992) 709.

[17] D. Biagi Maino *et al.*, Proceedings of the ENEA Workshop "Intelligenza Artificiale per i Beni Culturali", Bologna, 1999; see `studi131.casaccia.enea.it/enea/it/wsbc.html`.