Nuclear Instruments and Methods in Physics Research A 362 (1995) 487–498

ELSEVIER

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH
Section A

A multidimensional unfolding method based on Bayes' theorem

G. D'Agostini *

Università "La Sapienza" and INFN, Roma, Italy

Received 5 August 1994; revised form received 2 March 1995

Abstract

Bayes' theorem offers a natural way to unfold experimental distributions in order to get the best estimates of the true ones. The weak point of the Bayes approach, namely the need of the knowledge of the initial distribution, can be overcome by an iterative procedure. Since the method proposed here does not make use of continuous variables, but simply of cells in the spaces of the true and of the measured quantities, it can be applied in multidimensional problems.

1. Introduction

In any experiment, the distribution of the measured observables differs from that of the corresponding *true* physical quantities, due to physics and detector effects. For example, one may be interested in measuring in deep inelastic scattering (DIS) events the variables x and Q^2. In such a case one is able to build statistical estimators which have in principle a physical meaning similar to the true quantities, but which have a non-vanishing variance and are also distorted, due to QED and QCD radiative corrections, parton fragmentation, particle decay and limited detector performances. The aim of the experimentalist is to unfold the observed distribution from all these distortions so as to extract the true distribution. This requires a satisfactory knowledge of the overall effect of the distortions on the true physical quantity.

When dealing with only one physical variable the method mostly used to solve this problem is the so called *bin-to-bin* correction: one evaluates with a Monte Carlo simulation a *generalized efficiency* (it may even be larger than unity), calculating the ratio between the number of events falling in a certain bin of the reconstructed variable and the number of events in the *same* bin of the true variable. This efficiency is then used to estimate the number of true events from the number of events observed in that bin. Clearly this method requires the same subdivision in bins of the true and the experimental variable and hence cannot take into account large migrations of events from a bin to the others. Moreover it neglects the unavoidable correlations between adjacent bins. This approximation is valid only if the amount of migration is negligible and if the standard deviation of the smearing is smaller than the bin size.

Still dealing with the one dimensional case, an attempt to solve the problem of the migrations is sometimes done building a matrix which connects the number of events generated in one bin with the number of events observed in the other bins. This matrix is then inverted and applied to the measured distribution. This immediately yields inversion problems if the matrix is singular, since there is no reason, from a probabilistic point of view, why the inverse matrix should exist, as can be easily seen by taking as an example two bins of the true quantities both of which have the same probability to be observed in each of the bins of the measured quantity. This suggests that this way of treating probability distributions like vectors in space is clearly not correct, even in principle. Moreover, even if the matrix can be inverted (having for example a very large number of events to estimate its elements and choosing the binning in such a way as to make the matrix not singular) the method is not able to handle large statistical fluctuations. The easiest way to see this is to think of the unavoidable negative terms of the inverse of the matrix which, in some extreme cases, may yield negative numbers of unfolded events. Beside these theoretical considerations, the experience of the users of this method is rather discouraging, the results being strongly unstable.

A method which has been proposed to overcome the troubles encountered with the matrix inversion method is the "regularized unfolding" [1]. This produces satisfactory results, but it has never been widely used, probably because of certain technical complications. Unfortunately, since the true distribution is decomposed into orthogonal polynomials whose coefficients are estimated, this method only works in solving one dimensional problems.

* Corresponding author. E-mail 39942::dagostini or dagostini@vaxrom.roma1.infn.it.

0168-9002/95/$09.50 © 1995 Elsevier Science B.V. All rights reserved
SSDI 0168-9002(95)00274-X

This paper presents a different approach, based on Bayes' theorem, recognized by statisticians as the most powerful tool for making statistical inferences. The main advantages with respect to other unfolding methods are:

— it is theoretically well grounded;

— it can be applied to multidimensional problems;

— it can use cells of different sizes for the distribution of the true and the experimental values;

— the domain of definition of the experimental values may differ from that of the true values;

— it can take into account any kind of smearing and migration from the true values to the observed ones;

— it gives the best results (in terms of its ability to reproduce the true distribution) if one makes a realistic guess about the distribution that the true values follow, but, in case of total ignorance, satisfactory results are obtained even starting from a uniform distribution;

— it can take different sources of background into account;

— it does not require matrix inversion;

— it provides the correlation matrix of the results;

— it can be implemented in a short, simple and fast program, which deals directly with distributions and not with individual events.

2. Bayes' theorem

To stay close to the application of interest, let us state Bayes' theorem in terms of several independent *causes* (C_i, $i = 1,2,\cdots,n_C$) which can produce one *effect* (E). Let us assume we know the *initial probability* of the causes $P(C_i)$ and the conditional probability of the ith cause to produce the effect $P(E|C_i)$. The Bayes formula is then

$$P(C_i|E) = \frac{P(E|C_i)P(C_i)}{\sum_{l=1}^{n_C} P(E|C_l)P(C_l)}. \tag{1}$$

This can be read as follows: if we observe a single event (effect), the probability that it has been due to the ith cause is proportional to the probability of the cause times the probability of the cause to produce the effect.

For example, if we consider DIS events, the effect E can be the observation of an event in a cell of the measured quantities $\{\Delta Q^2_{\rm meas}, \Delta x_{\rm meas}\}$. The causes C_i are then all the possible cells of the true values $\{\Delta Q^2_{\rm true}, \Delta x_{\rm true}\}_i$.

One immediately sees that the $P(C_i|E)$ depends on the initial probability of the causes. This gives a first impression that this formula is sterile. In reality the Bayes formula has the power to increase the knowledge of $P(C_i)$ with the increasing number of observations. If one has no a priori prejudice on $P(C_i)$ the process of inference can be started from a uniform distribution.

The final distribution depends also on $P(E|C_i)$. These probabilities must be calculated or estimated with Monte Carlo methods. One has to notice that, in contrast to $P(C_i)$, these probabilities are not updated by the observations. So if there are ambiguities concerning the choice of $P(E|C_i)$ one has to try them all in order to evaluate their *systematic effects* on the results.

3. Unfolding an experimental distribution

If one observes $n(E)$ events with effect E, the expected number of events assignable to each of the causes is

$$\hat{n}(C_i) = n(E)P(C_i|E). \tag{2}$$

As the outcome of a measurement one has several possible effects E_j ($j = 1, 2,\cdots, n_E$) for a given cause C_i. For each of them the Bayes formula (1) holds, and $P(C_i|E_j)$ can be evaluated. For simplicity we will refer to conditional probabilities $P(C_i|E_j)$ as *smearing matrix* **S**, even if they describe cell-to-cell migration. Let us write Eq. (1) again in the case of n_E possible effects [1], indicating the initial probability of the causes with $P_0(C_i)$:

$$P(C_i|E_j) = \frac{P(E_j|C_i)P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j|C_l)P_0(C_l)}. \tag{3}$$

One has to note that:

— $\sum_{i=1}^{n_C} P_0(C_i) = 1$, as usual. Notice that if the probability of a cause is initially set to zero it can never change, i.e. if a cause does not exist it cannot be invented;

— $\sum_{i=1}^{n_C} P(C_i|E_j) = 1$: this normalization condition, mathematically trivial since it comes directly from Eq. (3), tells that each effect must come from one or more of the causes under examination. This means that if the observables contain also a non-negligible amount of background, this needs to be included among the causes;

— $0 \le \epsilon_i \equiv \sum_{j=1}^{n_E} P(E_j|C_i) \le 1$: there is no need for each cause to produce at least one of the effects taken under consideration. ϵ_i gives the *efficiency* of detecting the cause C_i in any of the possible effects.

After $N_{\rm obs}$ experimental observations one obtains a distribution of frequencies $\boldsymbol{n}(E) \equiv \{n(E_1), n(E_2),\cdots, n(E_{n_E})\}$. The expected number of events to be assigned to each of the causes and only due to the ob-

[1] The broadening of the distribution due to the smearing suggests a choice of n_E larger then n_C. We would like to remark that there is no need to reject events where a measured quantity has a value outside the range allowed for the physical quantity. For example, for the case of DIS events, also cells with $x_{\rm meas} > 1$ or $Q^2_{\rm meas} < 0$ give information about the true distribution.

served events can be calculated applying Eq. (2) to each effect:

$$\hat{n}(\mathrm{C}_i)\,|_{\mathrm{obs}} = \sum_{j=1}^{n_\mathrm{E}} n(\mathrm{E}_j)\,P(\mathrm{C}_i\,|\,\mathrm{E}_j).$$

Taking into account the inefficiency [2], the best estimate of the true number of events is then

$$\hat{n}(\mathrm{C}_i) = \frac{1}{\epsilon_i}\sum_{j=1}^{n_\mathrm{E}} n(\mathrm{E}_j)\,P(\mathrm{C}_i\,|\,\mathrm{E}_j) \qquad \epsilon_i \neq 0. \tag{4}$$

From these unfolded events we can estimate the true total number of events, the final probabilities of the causes and the overall efficiency:

$$\hat{N}_{\mathrm{true}} = \sum_{i=1}^{n_\mathrm{C}} \hat{n}(\mathrm{C}_i),$$

$$\hat{P}(\mathrm{C}_i) \equiv P(\mathrm{C}_i\,|\,\boldsymbol{n}(\mathrm{E})) = \frac{\hat{n}(\mathrm{C}_i)}{\hat{N}_{\mathrm{true}}},$$

$$\hat{\epsilon} = \frac{N_{\mathrm{obs}}}{\hat{N}_{\mathrm{true}}}.$$

It is important to remark that $\hat{\epsilon}$ may differ from the a priori overall efficiency ϵ_0 calculated from the *rec*onstructed and *gen*erated Monte Carlo events

$$\epsilon_0 = \frac{N_{\mathrm{rec}}}{N_{\mathrm{gen}}} = \frac{\sum_{i=1}^{n_\mathrm{C}} \epsilon_i P_0(\mathrm{C}_i)}{\sum_{i=1}^{n_\mathrm{C}} P_0(\mathrm{C}_i)}.$$

If the initial distribution $\boldsymbol{P}_0(\mathrm{C})$ is not consistent with the data, it will not agree with the final distribution $\hat{\boldsymbol{P}}(\mathrm{C})$. The closer the initial distribution is to the true distribution, the better the agreement is. One can easily verify for simulated data (see e.g. the non-trivial cases shown in Section 6) that the distribution $\hat{\boldsymbol{P}}(\mathrm{C})$ lies between $\boldsymbol{P}_0(\mathrm{C})$ and the true one. This suggests to proceed iteratively. So the unfolding can be performed through the following steps:

1) choose the initial distribution of $\boldsymbol{P}_0(\mathrm{C})$ from the best knowledge of the process under study, and hence the initial expected number of events $n_0(\mathrm{C}_i) = P_0(\mathrm{C}_i)N_{\mathrm{obs}}$; in case of complete ignorance, $\boldsymbol{P}_0(\mathrm{C})$ will be just a uniform distribution: $P_0(\mathrm{C}_i) = 1/n_\mathrm{C}$;

2) calculate $\hat{\boldsymbol{n}}(\mathrm{C})$ and $\hat{\boldsymbol{P}}(\mathrm{C})$;

3) make a χ^2 comparison between $\hat{\boldsymbol{n}}(\mathrm{C})$ and $\boldsymbol{n}_0(\mathrm{C})$;

4) replace $\boldsymbol{P}_0(\mathrm{C})$ by $\hat{\boldsymbol{P}}(\mathrm{C})$, and $\boldsymbol{n}_0(\mathrm{C})$ by $\hat{\boldsymbol{n}}(\mathrm{C})$, and start again; if, after the second iteration the value of χ^2 is "small enough", stop the iteration; otherwise go to step 2. Some criteria about the optimum number of iterations will be discussed later.

4. Estimation of the uncertainties

After the iterative procedure described above has converged, one obtains the unfolded distribution $\hat{\boldsymbol{n}}(\mathrm{C})$. As far as the evaluation of the uncertainties is concerned, one cannot simply take the square root of these numbers. Even if the number of cells is large and the uncertainty on $P(\mathrm{E}_j\,|\,\mathrm{C}_i)$ is negligible, the quantities which are distributed according to a Poisson distribution are the observed numbers $\boldsymbol{n}(\mathrm{E})$ and not $\hat{\boldsymbol{n}}(\mathrm{C})$, since the latter get contributions from several $n(\mathrm{E}_j)$. Moreover, it is clear that the uncertainties on $n(\mathrm{C}_i)$ have some degree of correlation, since the observed number of events $n(\mathrm{E}_j)$ is shared between all the causes from which the events can be originated.

To see all the sources of uncertainties and of correlations on $\boldsymbol{n}(\mathrm{C})$ in detail, let us rewrite Eq. (4), making use of Eq. (3), as

$$\hat{n}(\mathrm{C}_i) = \sum_{j=1}^{n_\mathrm{E}} M_{ij}\,n(\mathrm{E}_j),$$

where

$$M_{ij} = \frac{P(\mathrm{E}_j\,|\,\mathrm{C}_i)P_0(\mathrm{C}_i)}{\left[\sum_{l=1}^{n_\mathrm{E}} P(\mathrm{E}_l\,|\,\mathrm{C}_i)\right]\left[\sum_{l=1}^{n_\mathrm{C}} P(\mathrm{E}_j\,|\,\mathrm{C}_l)P_0(\mathrm{C}_l)\right]}.$$

M_{ji} can be seen as the terms of the *unfolding matrix* **M**, which is clearly *not* the mathematical inverse of the smearing matrix **S**. Let us examine the various contributions to the covariance matrix of $\hat{n}(\mathrm{C}_i)$, denoted by **V**:

— $P_0(\mathrm{C}_i)$: we consider the initial probabilities without statistical error since they affect the results in a *systematic* way, to be evaluated by studying how stable the results are for a variation of the starting hypothesis. The values of $P_0(\mathrm{C}_i)$ used in the calculations of the uncertainties will be those obtained in the last but one iteration.

— $n(\mathrm{E}_j)$: the data sample is a realization of a multinomial distribution of which the parameter n has to be identified with the true number of events, estimated by $\hat{N}_{\mathrm{true}}$. The contribution to **V** due to $\boldsymbol{n}(\mathrm{E})$ is then

$$V_{kl}(\boldsymbol{n}(\mathrm{E})) = \sum_{j=1}^{n_\mathrm{E}} M_{kj}M_{lj}\,n(\mathrm{E}_j)\left(1 - \frac{n(\mathrm{E}_j)}{\hat{N}_{\mathrm{true}}}\right) - \sum_{\substack{i,j=1\\ i\neq j}}^{n_\mathrm{E}} M_{ki}M_{lj}\frac{n(\mathrm{E}_i)\,n(\mathrm{E}_j)}{\hat{N}_{\mathrm{true}}}.$$

If the relative frequency in each cell is sufficiently small, due to the large number of bins or to low efficiency, the numbers in each of the cells can be approximated by an independent Poisson distribution.

— $P(\mathrm{E}_j\,|\,\mathrm{C}_i)$: these terms are usually estimated by Monte Carlo. They are affected by statistical and systematic errors. The latter come from the assumptions made in the simulation and have to be treated with appropriate

[2] If $\epsilon_i = 0$ then $\hat{n}(\mathrm{C}_i)$ will be set to zero, since the experiment is not sensitive to the cause C_i.

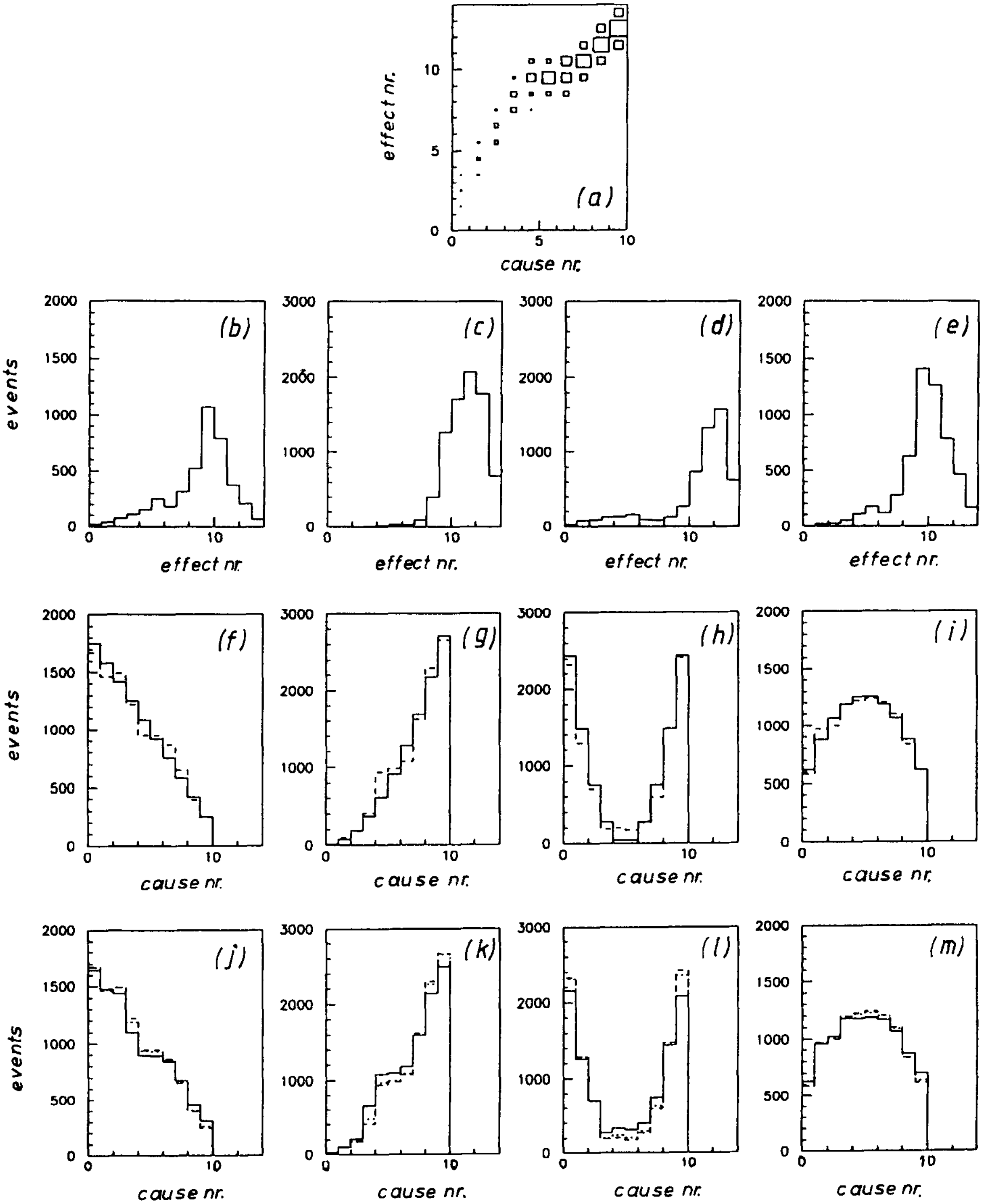


Fig. 1. (a) Visualization of the smearing matrix $\mathbf{S}_1$ (see text): in the abscissa and the ordinate are the true and the measured quantities, respectively; the box area is proportional to the migration probability. (b–e) Smeared distributions of f_{1-4} (see text) obtained from 10000 generated events. (f–i) True distributions of f_{1-4} (solid lines) compared with the results of a 3-step unfolding (dashed lines). (j–m) Unfolded distributions after the first, second and third step (solid, dotted and dashed lines) of f_{1-4}.

methods. The statistical errors come instead from the limited number of simulated events. Like the uncertainties on $n(E_j)$, they also induce correlations between the results. In fact, the total number of events generated for each of the cause-cells C_i is shared between the effect-cells E_j and their distribution is multinomial. If the migration effect is not very strong, i.e. a cell C_i is observed only in a small number of cells E_j, one cannot neglect the covariance

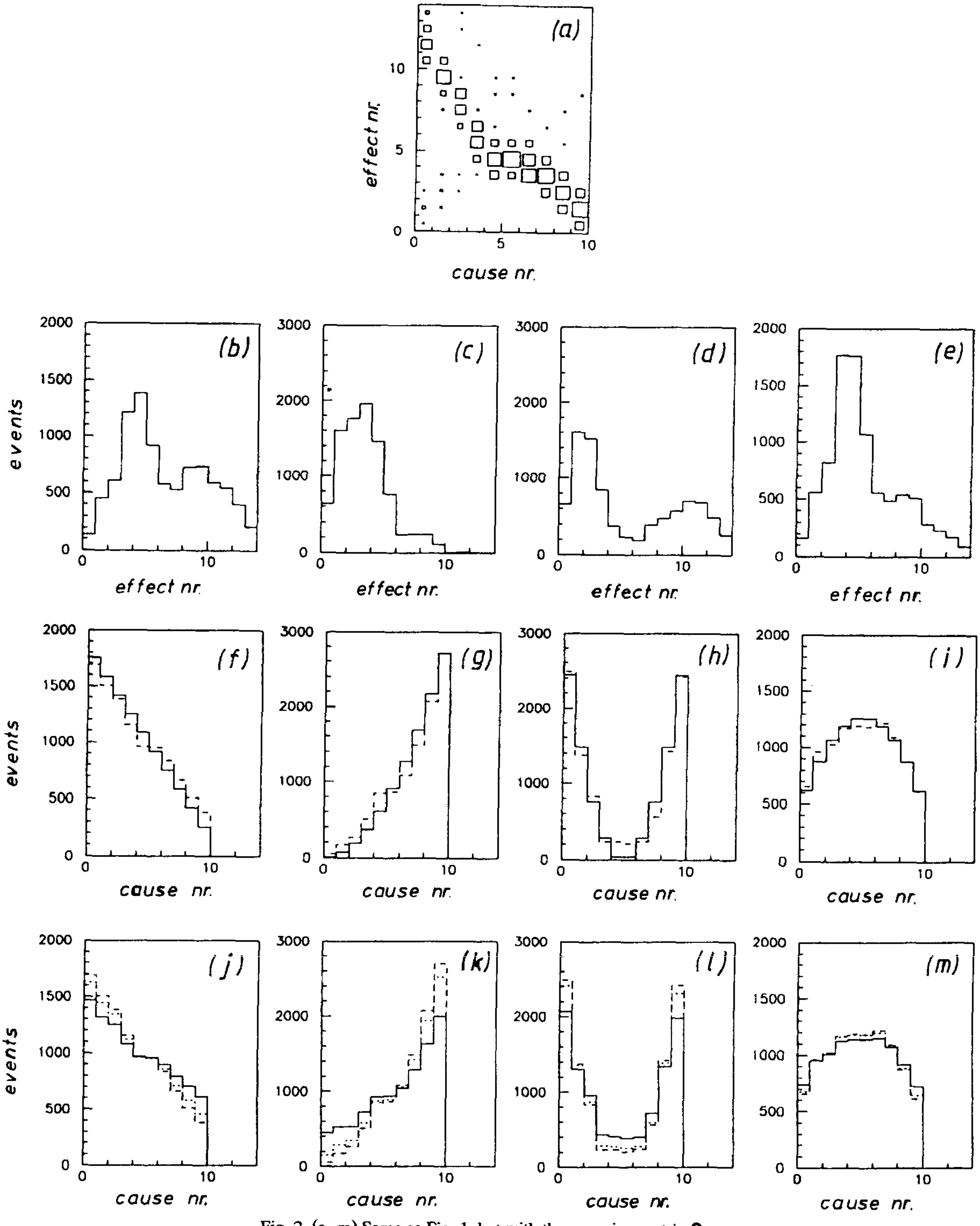


Fig. 2. (a–m) Same as Fig. 1, but with the smearing matrix $\mathbf{S}_2$.

between the terms $P(\mathrm{E}_{j1}|\mathrm{C}_i)$ and $P(\mathrm{E}_{j2}|\mathrm{C}_i)$. We can instead reasonably neglect the covariance between two terms related to different cause-cells.

Under these hypotheses the contribution to **V** from the uncertainty of **M** is

$$V_{kl}(\mathbf{M}) = \sum_{i,j=1}^{n_{\mathrm{E}}} n(\mathrm{E}_i)n(\mathrm{E}_j)\mathrm{Cov}(M_{ki}, M_{lj}),$$

where:

$$\mathrm{Cov}(M_{ki}, M_{lj}) = \sum_{\{ru\},\{su\}} \frac{\partial M_{ki}}{\partial P(\mathrm{E}_r|\mathrm{C}_u)} \frac{\partial M_{lj}}{\partial P(\mathrm{E}_s|\mathrm{C}_u)} \times \mathrm{Cov}[P(\mathrm{E}_r|\mathrm{C}_u), P(\mathrm{E}_s|\mathrm{C}_u)];$$

$$\frac{\partial M_{ki}}{\partial P(\mathrm{E}_r|\mathrm{C}_u)} = M_{ki}\left[\frac{\delta_{ku}\delta_{ri}}{P(\mathrm{E}_r|\mathrm{C}_u)} - \frac{\delta_{ku}}{\epsilon_u} - \frac{\delta_{ri}M_{ui}\epsilon_u}{P(\mathrm{E}_i|\mathrm{C}_u)}\right];$$

$$\mathrm{Cov}[P(\mathrm{E}_r|\mathrm{C}_u), P(\mathrm{E}_s|\mathrm{C}_u)] = \begin{cases} \dfrac{1}{n_u}P(\mathrm{E}_r|\mathrm{C}_u)[1-P(\mathrm{E}_r|\mathrm{C}_u)] & (r=s) \\ -\dfrac{1}{n_u}P(\mathrm{E}_r|\mathrm{C}_u)P(\mathrm{E}_s|\mathrm{C}_u) & (r\neq s). \end{cases}$$

In the last expression n_u represents the number of events generated in the cell C_u in order to evaluate the smearing function.

The sum of the two contributions gives the elements of the covariance matrix of the unfolded numbers:

$$V_{kl} = V_{kl}(\boldsymbol{n}(\mathrm{E})) + V_{kl}(\mathbf{M}).$$

5. Treatment of background

The unfolding based on Bayes' theorem can take into account in a natural way the presence of background, simply adding it to the possible causes responsible for the observables. Even several sources of background can be treated. For example, in case of a single contribution, one adds to the physical cells an extra $\mathrm{C}_{n_\mathrm{C}+1}$, with initial probability $P(\mathrm{C}_{n_\mathrm{C}+1})$. The conditional probabilities $P(\mathrm{E}_j|\mathrm{C}_{n_\mathrm{C}+1})$ will be just the unnormalized shape (in the sense that $\epsilon_{n_\mathrm{C}+1}$ may be smaller than unity) of the background distributions. The result of the unfolding will then provide the number of events to be assigned to the background.

In principle this method could also be used to disentangle the true distributions of several physics processes contributing to the same distribution of the observable. The problem will not be further discussed in the rest of the paper, but this method is likely to work only in very simple cases, and other more sophisticated methods — like neural network algorithms — should yield better results.

6. Results

6.1. The program

The above method has been implemented in a short self-contained Fortran code available on request from the author together with examples.

The user can provide either the smearing matrix or, more directly, the number of MC events produced in a cell of the true quantities together with the number of events which fall in each cell of the measured quantities. Only in the latter case it is possible to take into account the uncertainties due to the limited MC statistics. It is interesting to remark that there is no need to generate the MC events according to a realistic physical distribution of the variable under study, and in fact it can be more efficient to have several runs in different regions and to merge the results at the end, or to use a uniform distribution in order to populate well all the kinematical regions.

The covariance matrix of $\hat{\boldsymbol{n}}(\mathrm{C})$ calculated by the program allows the user to redefine the cell sizes (which may be not all equal) a posteriori in order to reduce the correlation of the results. On the other end, the presence of common sources of uncertainty make it impossible to redefine the cells so as to have uncorrelated results. The full covariance matrix should be used to exploit at best the results in further analysis.

6.2. Unfolding a distribution not affected by statistical fluctuations

The ideal performances of the method have been studied with samples equivalent to having infinite statistics. This was done calculating the expected number of events from the true distribution and the smearing matrix, namely

$$n(\mathrm{E}_j) = \sum_i P(\mathrm{E}_j|\mathrm{C}_i)P_{\mathrm{true}}(\mathrm{C}_i)N_{\mathrm{true}},$$

or in a more compact way

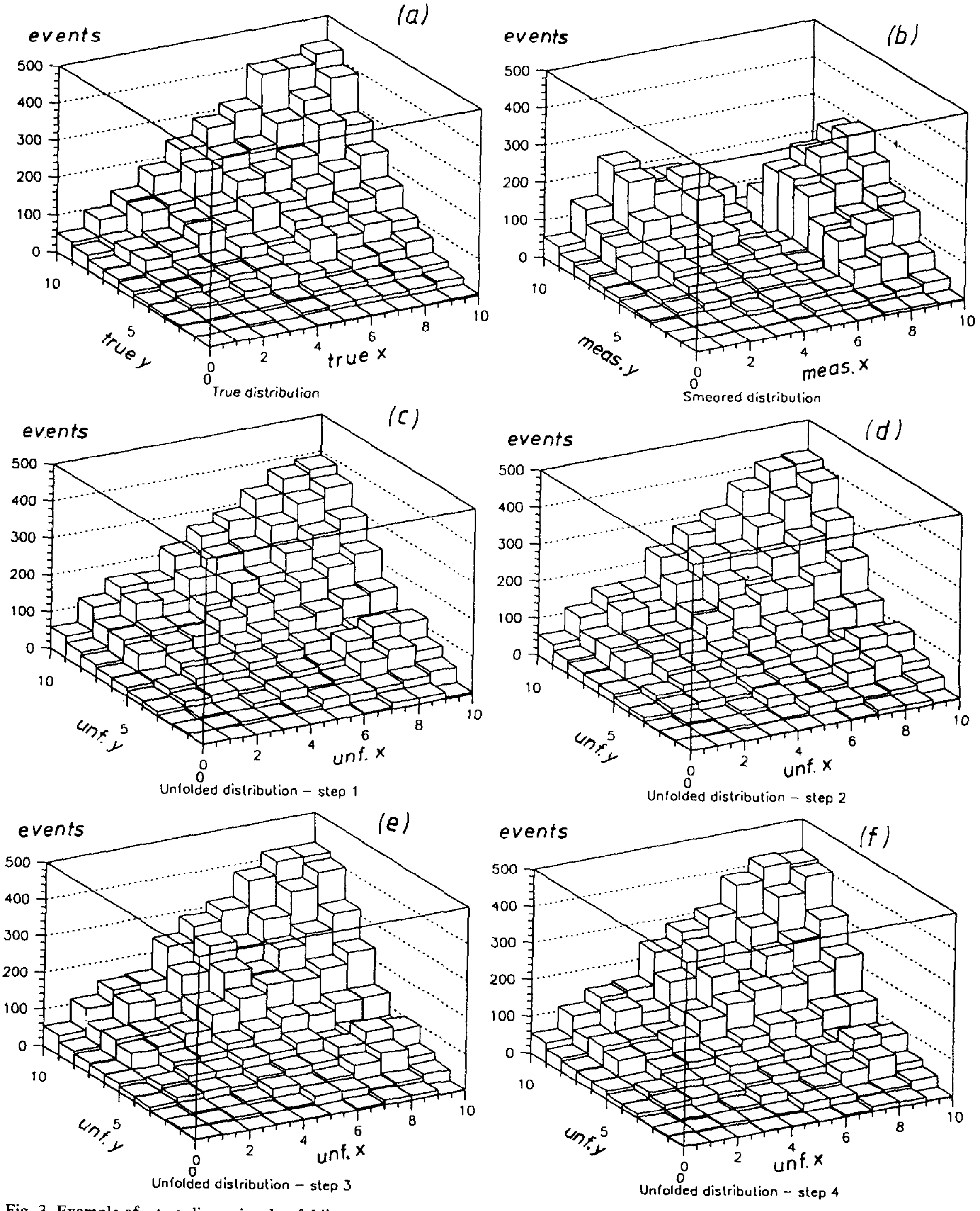
$$\boldsymbol{n}(E) = \mathbf{S}\boldsymbol{P}_{\mathrm{true}}N_{\mathrm{true}}.$$

Figs. 1a and 2a show the smearing matrices, called hereafter $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively, used for the tests. The abscissa and the ordinate represent the true and measured quantities, respectively, and the box area is proportional to the probability $P(\mathrm{E}_j|\mathrm{C}_i)$. In order to test the unfolding performances, nontrivial smearing matrices have been chosen: in $\mathbf{S}_1$ the causes have all different efficiencies; $\mathbf{S}_2$ shows an unusual anticorrelation and also long range migrations; both are almost flat in some points; the domains of the true and measured value are different. The number of effect cells has been chosen larger than the number of causes, as the smearing makes the distributions usually broader. The true value distributions are shown with solid lines in Figs. 1f–1i, repeated also in Fig. 2. We will refer to these test distributions with f_1, f_2, f_3 and f_4.

Monte Carlo samples of smeared distributions obtained from 10000 generated events (those without random fluctuations do not look much different) are shown in Figs. 1b–1e and 2b–2e for the two smearing matrices.

In all cases the initial distribution of the true values has been assumed to be uniform.

Already after the first iteration the unfolded distribution is close to the true one (as shown in the case of limited



True distribution

Smeared distribution

Unfolded distribution – step 1

Unfolded distribution – step 2

Unfolded distribution – step 3

Unfolded distribution – step 4

Fig. 3. Example of a two dimensional unfolding: true distribution (a), smeared distribution (b) and results after the first 4 steps (c–f).

statistics with solid line of Figs. 1j–1m and Figs. 2j–2m). The agreement increases constantly with the number of iterations and eventually, for all the test distribution and smearing matrices, the true distribution is recovered. This means that **MS** → **1**. Obviously, this cannot be a rigorous result, since both matrices have all the elements defined not negative, and hence there must be a correct combination of the zeros in the row of **M** with the zeros of the column of **S** in order to give the null off-diagonal elements of the product matrix. An easy case where this limit cannot hold is when two causes have the same probabilities to produce the effects, i.e. $P(E_j|C_i) = P(E_j|C_k)\ \forall j$. In this case the result will be that $\hat{n}(C_i) = \hat{n}(C_k)$ independently of the true probabilities of C_i and C_k. This is in fact the best result that the hypothesis allows. An extreme case is when the elements of the smearing matrix are all equal. One finds then that the final probabilities are equal to the initial ones, since under this condition the observations do not increase the knowledge at all.

6.3. Simulation with limited statistics

To make a more realistic evaluation of the performances of the method, the role of observed distributions has been played by Monte Carlo events simulated according to the true distribution and the smearing matrix. For each of the configurations 10000 events have been generated. The observed distributions are shown in Figs. 1b–1e and 2b–2e. As the smearing matrices are rather severe there is no similarity at all with the true distributions.

After the experience gained with the sample having no statistical fluctuation, the first results are a bit surprising. After a few iterations the unfolded distribution becomes very close to the true one, but if one performs a very large number of iterations, it converges toward a distribution which shows strong fluctuations around the true one. The reason is simple. As stated before, an infinite iteration loop yields an unfolding matrix which is — in some sense — the inverse of the smearing matrix. So one gets exactly the same problems as discussed in the introduction about the matrix inversion method. The reason is that each of the bins in the true value distribution acts as a independent degree of freedom and after an infinite number of iterations one reaches a very fluctuating solution — a kind of amplification of the statistical fluctuations — similar to the result of a fit of a large number n of points with a polynomial of order $n - 1$.

Figs. 1m–1p and 2m–2p show the result after the 1st (solid line), 2nd (dotted line) and 3rd step (dashed line). The latter is compared to the true distribution (solid line) in Figs. 1i–1e and 2i–2e. The agreement is qualitatively good also in the case of difficult situations, like f_3 with $\mathbf{S}_2$. An example of 2-D unfolding is shown in Fig. 3, where the true distribution, the smeared one, and the results of the first 4 steps are shown.

In order to quantify the goodness of the results, Figs.

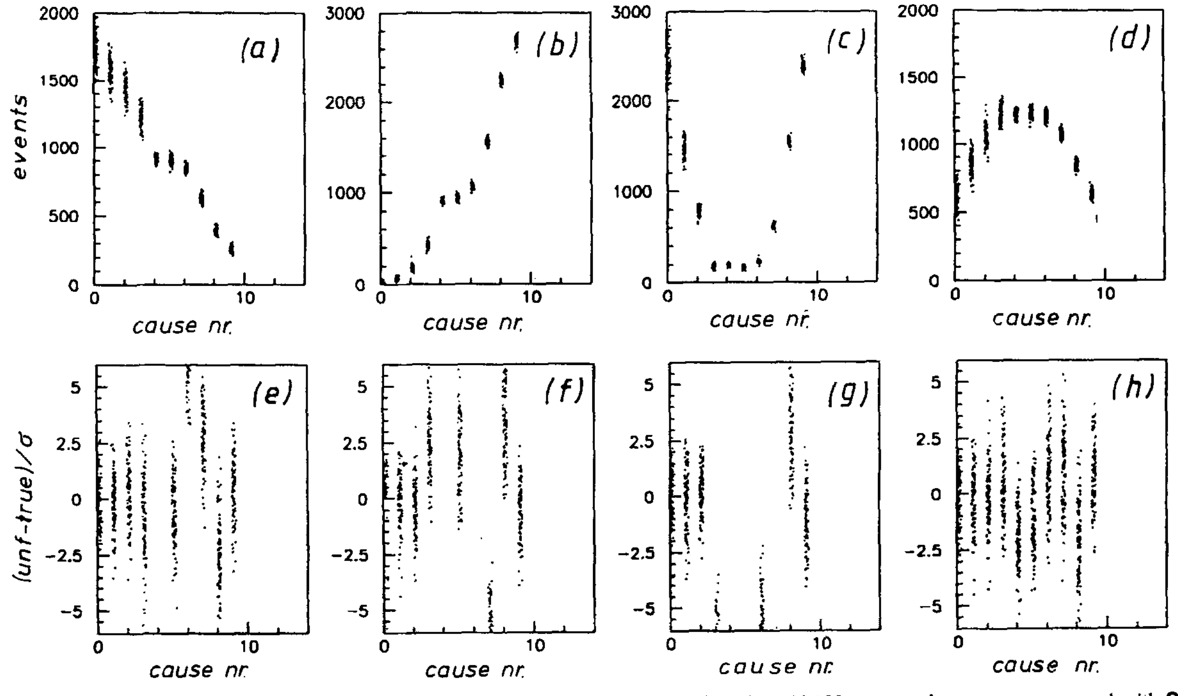


Fig. 4. Results obtained by a 3-step unfolding of 100 independent data sets, each based on 10000 generated true events smeared with $\mathbf{S}_1$: (a–b) distribution on the unfolded numbers for true distributions f_{1-4}; (e–h) distribution of the difference between the unfolded and the true numbers, divided by the extimated standard deviation, for true distributions f_{1-4}.

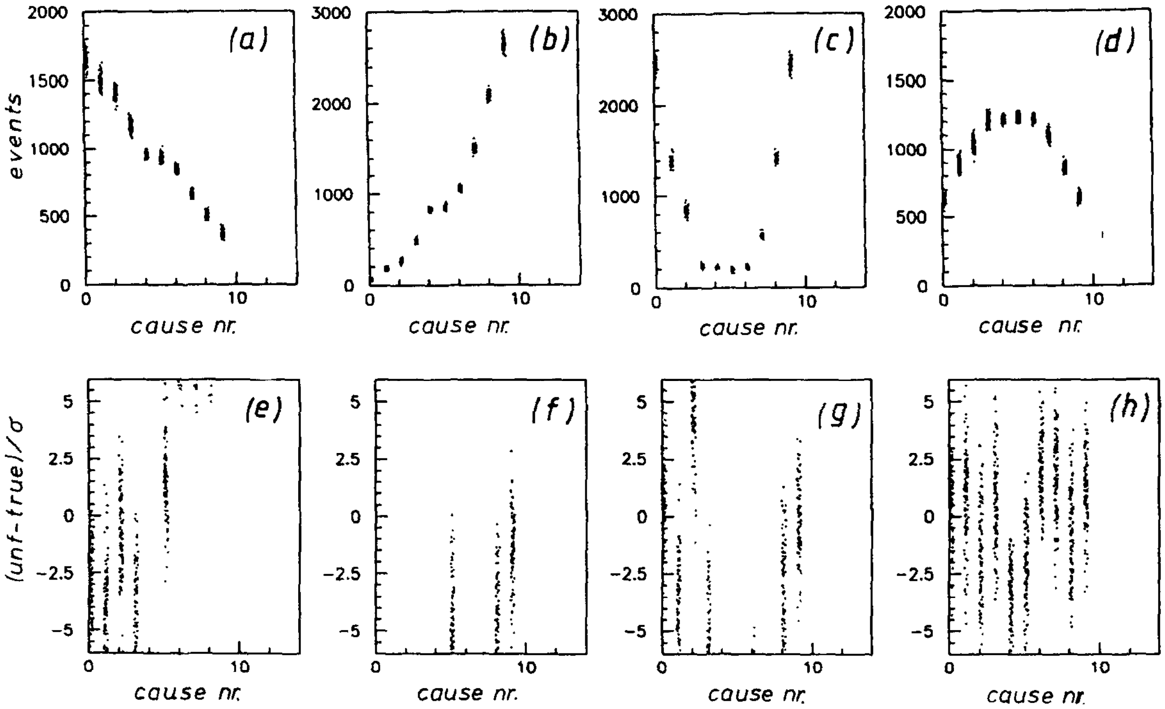


Fig. 5. Same as Fig. 4 in case of the smearing matrix $\mathbf{S}_2$.

4a–4d and Figs. 5a–5d show, respectively for $\mathbf{S}_1$ and $\mathbf{S}_2$, the results obtained unfolding 100 independent MC data sets, each of 10000 generated events, and where a 3-step procedure has been used. Figs. 4e–4h and Figs. 5e–5h give for each of the true bins the difference between unfolded and true numbers of events, divided by the standard deviation calculated in each unfolding. This gives an idea of the bias of the method and of the goodness of

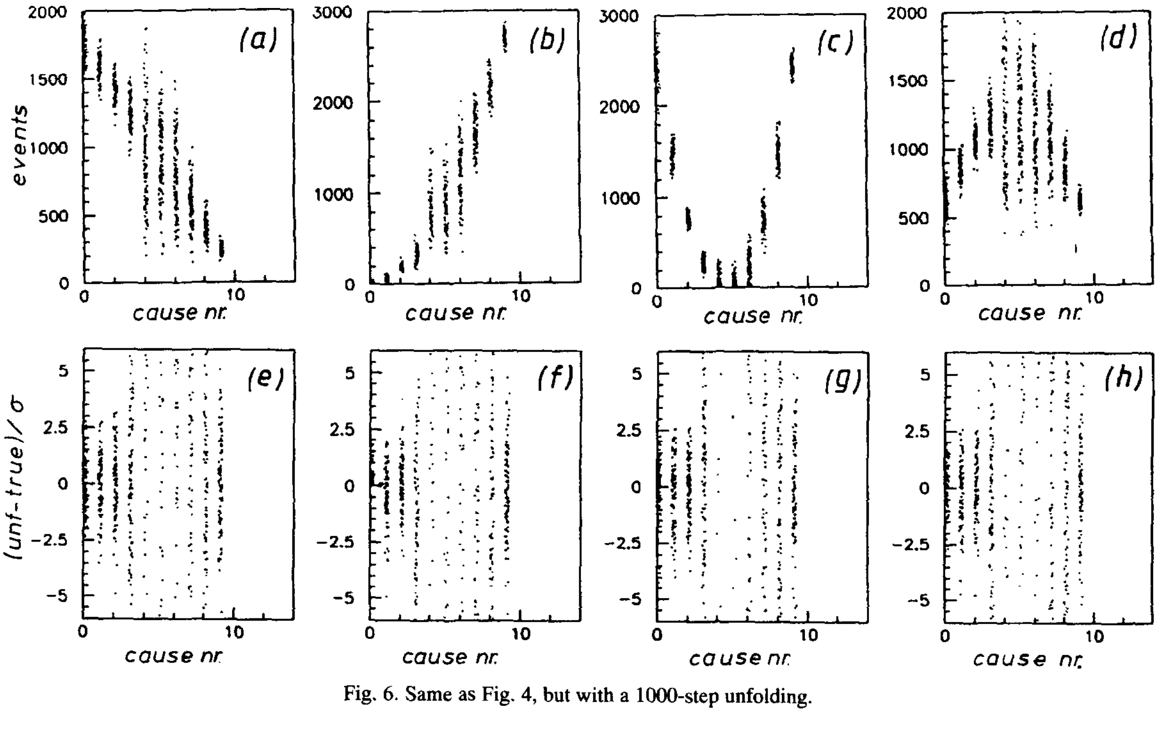


Fig. 6. Same as Fig. 4, but with a 1000-step unfolding.

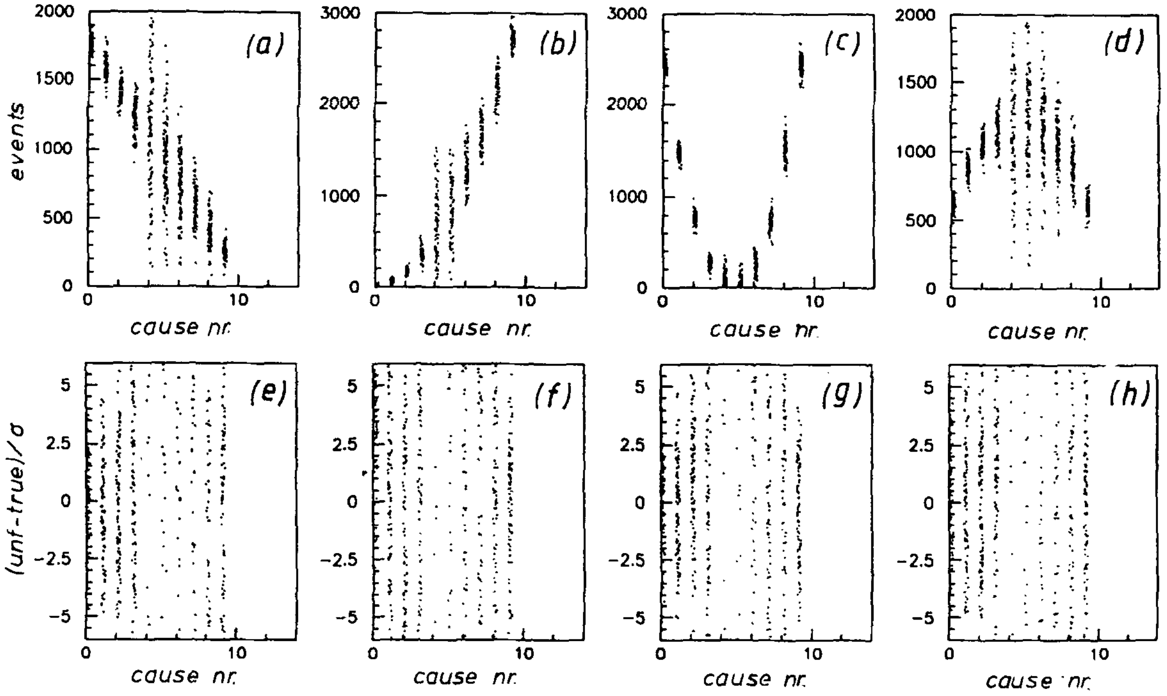


Fig. 7. Same as Fig. 5, but with a 1000-step unfolding.

the uncertainty estimation. Only some cases show quantitatively satisfactory results. Figs. 6 and 7 show what happens after 1000 steps. Very large fluctuations appear around the true value, as discussed above.

In order to avoid the problem of wild fluctuations with the increasing number of steps some possible ways out may be considered:

— reduce the number of degrees of freedom putting some constraints between the probabilities of the true values, for example imposing that they follow a particular function. This is exactly equivalent to making a maximum-likelihood fit to the data (it is in fact known that the maximum-likelihood principle can be derived from Bayes' theorem). It implies we know the expression of the function a priori, and it is the best way to proceed if one is just interested in finding the parameters of a particular model;

— find a criterion concerning the optimum number of iterations, which may depend on the kind of problem: playing with the distributions f_{1-4} one can realize that in most of the cases a good agreement is reached after a few iterations. Some particularly difficult cases for which this is not the case have to be attributed to the smearing matrices chosen;

— smooth the results of the unfolding before feeding them into the next step as "initial probability".

The third possibility has been chosen, as it turns out to produce stable results and moreover to be consistent with the spirit of the Bayesian inference. We remind the reader that in this frame knowledge is achieved by making use of Bayes' theorem, initial hypotheses and empirical observations. The hypothesis that most of the physical distributions of interest, in particular structure functions, are smooth is well proven by experience. For this reason using the result of the first iteration with all its fluctuations as initial probability for the second step is from the Bayesian point of view even wrong in principle, since one is "telling" the unfolding that the physical distribution can be of that kind, with all those wild fluctuations. It is preferable instead to feed into the program, as the initial distribution of the next step, a continuous and smooth function, whose shape is already influenced by the observations.

One has to notice that there is no reason to worry that the smoothing procedure produces biased results or hides strong peaks if they are significantly present in the data, since the smoothed distribution used as initial probability for the second iteration is nothing but an hypothesis more realistic that the first one and can only give better results than the uniform distribution.

For the test distributions a rough smoothing has been performed for all of them by a polynomial fit of 3rd degree.[3] Superior results, with respect to the previous

[3] The smoothing has not been put inside the unfolding program and must be done by the user, who knows the topology of the cells in the space of the physics quantities. It is recommanded, whenever it is possible, to choose the smoothing function which reflects at best the physics case. This can be of particular relevance in multidimensional problems, and moreover it can provide a very effective procedure for a simultaneous unfolding and fitting.

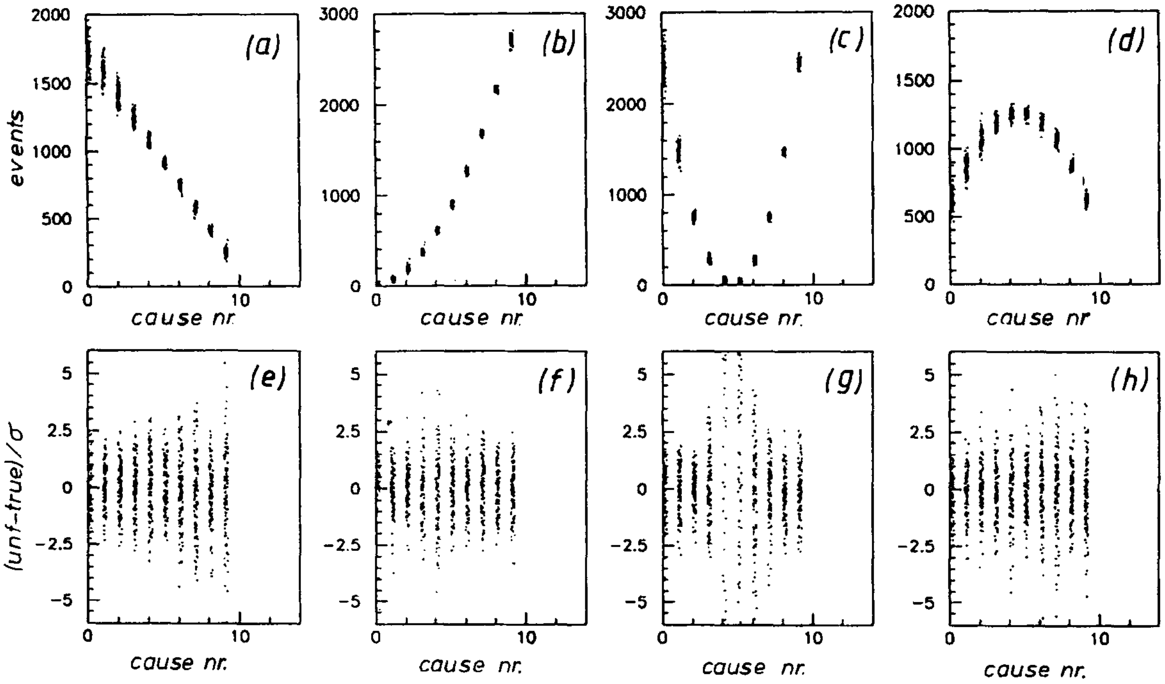


Fig. 8. Same as Fig. 4, but with a 20-step unfolding and smoothing the probability distribution between the steps.

case, are achieved after a few steps and the convergence is obtained between 3 steps (as in the case of f_1 and f_4 with $\mathbf{S}_1$) and 15 steps (f_2 with $\mathbf{S}_2$). Figs. 8 and 9 show the result of 100 data samples after a 20-step unfolding with intermediate smoothing. No oscillations are present and the results do not change with the increasing the number of

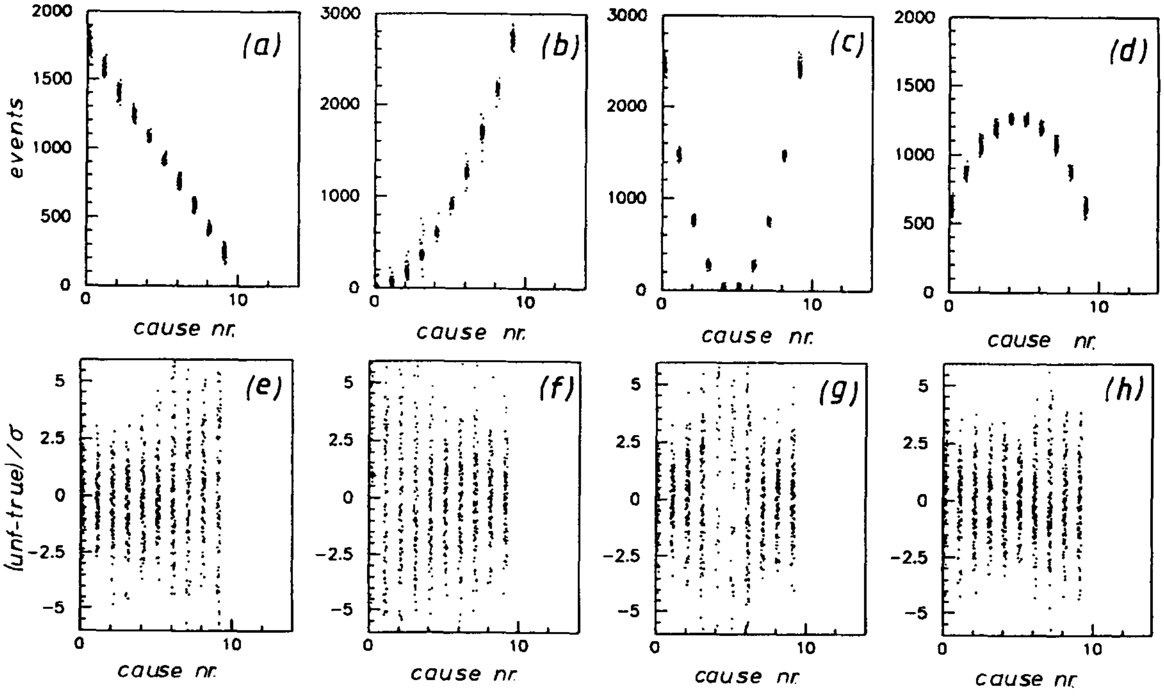


Fig. 9. Same as Fig. 5, but with a 20-step unfolding and smoothing the probability distribution between the steps.

steps, indicating that the procedure has converged. Moreover, the differences between the unfolded and the true numbers, normalized to the calculated standard deviations, show that the latter are well enough estimated.

7. Conclusions

The iterative use of Bayes' theorem provides a promising method to unfold multidimensional distributions. With the smoothing of the resulting distribution between the steps a fast convergence is reached, and, for normal applications, the results are stable with respect to variations of the initial probabilities and of the smoothing procedures.

The covariance matrix of the result, which can also take into account the uncertainties due to the limited Monte Carlo statistics used to evaluate the smearing matrix, is provided. A Monte Carlo study has shown that the estimated standard deviations turn out to be close to those calculated from the dispersion of the data around the mean values, and that the method does not bias the results.

Acknowledgements

I am grateful to several colleagues of the ZEUS Collaboration for the discussions we had on the subject. In particular, I would like to thank Lutz Feld for criticisms on the proposed method and for a careful check of the program.

Note added

After the appearance of the preprint of this paper, I have received two claims of similar methods used in the analysis of high energy physics data: Stefan Kluth pointed out the use of formula (4) — although stated differently, not justified by the Bayes' theorem and not taking into account the inefficiency — together with the iteration procedure in an OPAL publication [2]; François Le Diberder, in order to perform a 4-dimensional unfolding of ALEPH data [3], has used a method which seems identical to the one here proposed (unfortunately no detailed information is available in the paper) and which also makes use of the smoothing between the iterations. In both cases there was no attempt to calculate directly the uncertainties, and they were estimated from the dispersion of the results in subsamples of the data sets.

References

[1] V. Blobel, Proc. 1984 CERN School of Computing, Aiguablava, Catalonia, Spain, 9–12 September 1984 (CERN 1985) p. 88.

[2] OPAL Collaboration, P.D. Acton et al., Z. Phys. C 59 (1993) 1.

[3] ALEPH Collaboration, D. Baskulic et al., Phys. Lett. B 307 (1993) 209.