DELPHI Collaboration

DELPHI 2002-0 XXXX 000
28 February, 2002

Statistically regularized deconvolution of unbinned spectra containing event-by-event resolution functions in a unified Bayesian-Frequentist Monte Carlo Framework.

Niels Jørgen Kjær
CERN, CH - 1211 Geneva 23, Switzerland
e-mail: Niels.Jorgen.Kjaer@cern.ch

Abstract

A recursive Bayesian algorithm is applied to update a generator function in the case having a set of events having individual Bayesian resolution functions. The recursive process is locally regularized by computing the statistical significance of the proposed update from the data themselves. The significance is evaluated in a Frequentist's framework using the Jackknife estimator. The algorithm is stopped when all proposed updates have significances which are below a regularization parameter. The choice of this parameter to be one standard deviation is shown to give a deconvoluted spectrum which is optimal and information conserving at the level of one standard deviation.

The output of the algorithm is formally a Monte Carlo generator function, with a full covariance matrix, which ensures full usage of the spectral information after deconvolution. Only localized information and features are retained by applying the algorithm, which explains why it is possible to obtain information conservation.

The algorithm is extendable to any number of dimensions, but relies on the concept of locality, which is only uniquely defined in 1-d. With an additional required definition of locality the algorithm is shown also to be information conserving in 2-d deconvolution problems at the level of one standard deviation.

This rapidly converging algorithm is unique since it correctly deals with event-by-event resolution functions and does not produce any artificial features which other unbinned algorithms are prone to do.

Contributed paper to "Advanced Statistical Techniques in Particle Physics",
Grey College, Durham, 18 - 22 March 2002

1 Introduction

The problem of rigorous deconvolution has occupied statisticians and mathematicians basically forever. Numerous prescriptions exist with varying degrees of reliability. This paper addresses the problem of deconvolution for few events with non-trivial event-by-event resolution functions. The resolution functions should be treated as log-likelihood functions with a uniform Bayesian prior. How the resolution functions are obtained is outside the scope of this paper. In [1] it was shown how to construct optimal log-likelihood function using the k2-filter technique. In this framework the resolution functions should be extracted using MC reweighted to the limit of negligible width of the generator function compared to the width of a typical resolution function. Such a procedure will automatically include features like backgrounds, phase space, etc.

The approach used in the work presented in this paper is a MC framework, where there exists an unknown generator function $g(x)$, which generates the spectrum of interest. Each event, i, is then stochastically smeared according to resolution functions $r_i(x)$.

The idea behind the method presented in this paper is to “train” the generator function to the observed spectrum in a similar manner as resolution functions are “trained” in MC analyses like the one described in [1]. From a conceptual point of view there should be no way to distinguish between deconvoluting a generator function or to construct optimal resolution functions. The training described in Section 2 is Bayesian in its foundation while the regularization described in Section 3 uses Frequentist’s methods to decide when there is no more significant information to be extracted and therefore that the algorithm has converged. Frequentist’s methods can finally also be applied to extract the statistical properties of deconvoluted spectra, which is fundamentally impossible in a Bayesian framework.

2 The master formula

The basic formula for deconvolution can be derived in many ways.

- From an Bayesian optimization of the combined convoluted likelihood.
- From asking that when looked at as a MC generator problem, the generator becomes self-consistent and when reweighted does not change the total probability.
- From heuristic arguments saying that the generator part should be uncorrelated with the resolution functions.

It basically all boils down to the master formula shown as Eq. 1:

$$\sum_{i=1}^{n} \frac{r_i(x)dx}{\int g(x)r_i(x)dx} = 1 \quad \forall x \tag{1}$$

This formula has very peculiar properties. On the one hand it is highly unstable. When solved directly for $g(x)$, $g(x)$ and $g(x+dx)$ become 100% anti-correlated. On the other hand Eq. 1 implies that

$$g(x) = \sum_{i=1}^{n} \frac{g(x)r_i(x)}{\int g(x)r_i(x)dx}, \tag{2}$$

which suggest an indirect iterative solution:

$$g_{n+1}(x) = \sum_{i=1}^{n} \frac{g_n(x) r_i(x)}{\int g(x) r_i(x) dx}. \tag{3}$$

In this case, $g(x)$ and $g(x+dx)$ become 100% positively correlated. This also means that the algorithm has recursive properties which show up as a growth of amplitudes which continues indefinitely. The algorithm itself is basically logarithmic in its behaviour. Recursive instabilities have exponential behaviour and the resulting behaviour is nearly linear in the number of recursions. Figure 1 shows the evolution of one typical deconvolution as a function of the number of recursions. If the algorithm is not regularized the final spectrum will end up being a sum of delta functions, with the length between delta spikes being related to effective widths of the resolution functions.

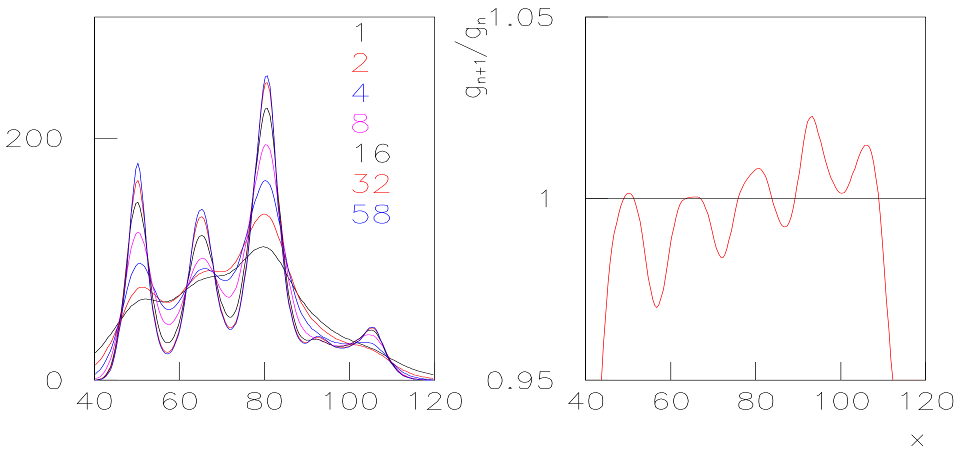

Figure 1: Plots showing the evolution of the deconvolution including regularization. The left hand depicts a typical deconvoluted spectrum after 1, 2, 4, 8, 16, 32 and 58 iterations. The right hand plot shows the "proposed" update of $g_{58(x)}/g_{57}(x)$.

3 Regularization

Figure 1 shows the final "proposed" updates of $g_{58}(x)/g_{57}(x)$ for a deconvolution including regularization. In average the update must be unity and due to the positive correlations a pattern arises with a scale related to the widths of the resolution functions. Now, the question can be posed whether a proposed update $g_{n+1}(x)/g_n(x)$ is statistically significant or not. If the question is posed only at one given value of x the answer is normally no! However, the question can also be posed semi-locally, e.g., just between two points, x_k and x_{k+1}, where all proposed updates have the same sign.

The proposed update can be computed as the total integral locally with

$$g_{n+1}(x \in [x_k; x_{k+1}])/g_n(x \in [x_k; x_{k+1}]) = \int_{x_k}^{x_{k+1}} \sum_{i=1}^{n} \frac{r_i(x) dx}{\int g(x) r_i(x) dx}. \tag{4}$$

Since the proposed update is a simple sum over all events the significance can be computed using the Jackknife method. In a first version of the algorithm, only the updates which had a local significance larger than a certain number of standard deviations were allowed. Notice that the "pivot" points where the proposed updates are zero pose no numerical problem. If only one side of the pivot is updated, the updated curve remains smooth at the pivot point. Later investigations showed that this algorithm does not take into account the correlations over the pivot points. A detailed study shows that the algorithm becomes more stable and performant when the decision of whether to update one segment also contains the information from the neighbor segments. When integrating over the neighbor segments the proposed update changes sign which has to be taken into account changing the sign of the integrator appropriately. The choice of including all the information from the neighbor segments is purely phenomenologically motivated, and more studies are needed in order to find out why this is the case.

The choice of the number of standard deviations has also been studied as a phenomenologically problem. This number remains a free regularization parameter, but the studies showed no reasons why this number should not normally be unity. The studies were performed on generated Breit-Wigner and delta spectra, with various resolution functions. After the deconvolution the spectra were used directly to measure the generator parameters, and the best performances and smallest biases were consistently found when the regularization parameter was set to one standard deviation. The results also showed that reliable measurements can be performed using deconvoluted spectra, but compared to a convolution measurement where the full event-by-event information is used, the performance was in general one standard deviations worse. Also the fluctuations in general and unwanted features are kept at one standard deviation. When the regularization parameter is set lower the precision can be enhanced, but at the cost of generating unwanted fluctuations, which cannot be controlled.

Since the "proposed" updates contain the sum over the events the Jackknife can be used to estimate the variance of the final result. However since the algorithm is recursive, not all correlations are taken into account. To obtain a full correlation matrix, it would be needed to do a full bootstrap on the event samples. Since the deconvolution algorithm is not incredibly fast, such a bootstrap should be avoided if possible.

The algorithm works on binned data, where the binning can be as fine as required. For each recursive step it requires one convolution of all the resolution functions. The convergence rate depends mainly of the width of the resolution functions w.r.t. the final structures, but are typically 5-10^2 for normal problems. The final cpu-use is proportional to the number of events and the number of bins squared.

3.1 Boundary conditions

The boundary conditions are important for generator functions which have a base outside the widow of interest. In the studies for this paper the boundary condition was to fix $g(x)$ outside the boundaries and include an integration outside the boundary limit to obtain the best possible estimators for the proposed updates.

3.2 Multi-dimensional deconvolution

The method described above can in principle easily be generalized to any number of dimensions. There is however one caveat which makes the practical implementation cumbersome. In 1-d it was easy to identify the pivot points used to determine whether to perform the proposed update or not. In more dimensions there might be saddle points where both the proposed update and its derivative are zero. If this happens a more detailed decision on how to determine whether to update is necessary. It will probably be easier to return the method where the neighbor segments are not used in the decision making.

4 Examples

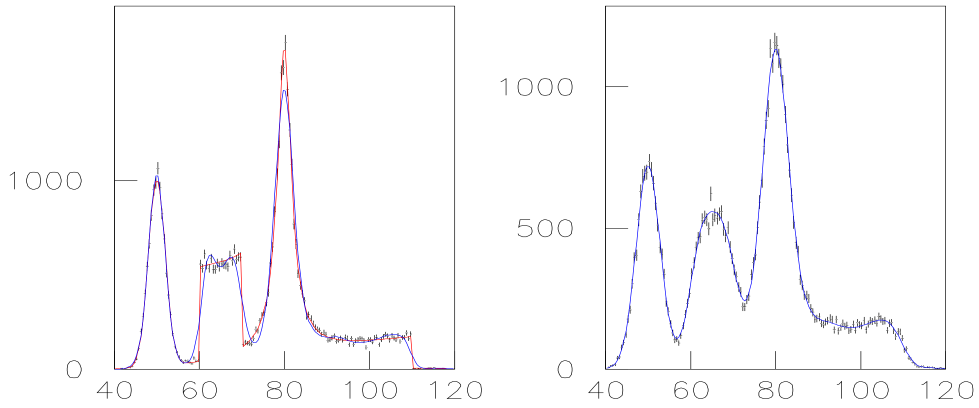

Figure 2: Generated, measured and deconvoluted spectra for 50000 events generated with the red spectrum in the left hand plot. The events for which the spectrum is shown in black have been convoluted with Gaussian resolution functions with width of 2. The blue curves correspond to the deconvoluted spectrum. In the right hand plot this is reconvoluted with the resolution functions to obtain a comparison with the measured spectrum in black.

This section depicts the behaviour of the proposed deconvolution algorithm for several examples. A complicated generator function is built in order to check all possible features. The events are then convoluted with a resolution function, which here for simplicity is chosen as a Gaussian[1]. In some cases the widths of the Gaussian is fixed, while in other it is allowed to vary on an event-by-event basis.

Figure 2 shows an example with large statistics and small widths of the resolution functions. The Gaussian part generated at 50 is always very well deconvoluted, while the more abrupt step function behaviour evidently cannot be perfectly determined. The Breit-Wigner peak will always be underestimated, which is due to its peculiar statistical properties. Notice finally that step functions will tend to give small wiggles close to the edges. The quality of the deconvolution can be checked by the right hand plot in

[1]More complicated and realistic resolution functions have also been investigated with similar conclusions.

Figure 2. There is a perfect agreement, except at the critical places (57,73,80) mentioned above. There the algorithm will underestimate the abruptness typically at the 1 standard deviation level.

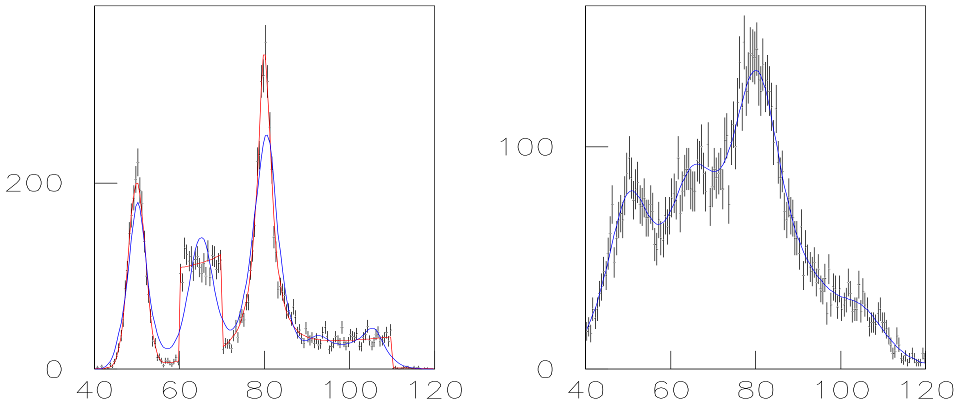

Figure 3: Same as Figure 2 with 10000 events generated with Gaussian resolution functions with widths varying from 2 to 10

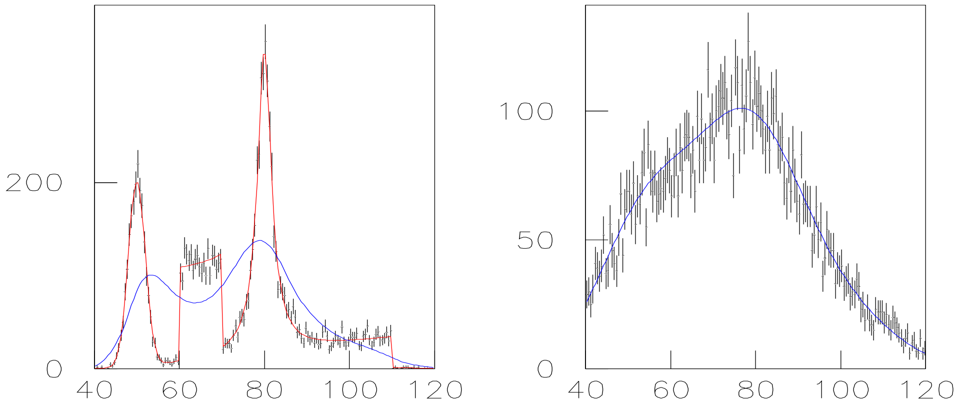

Figure 4: Same as Figure 2 with 10000 events generated with Gaussian resolution functions with widths fixed at 10

Figures 3 and 4 shows again spectra with large statistics. It can be seen that when the resolution becomes worse, features of the generator function gradually disappear.

Finally Figures 5-7 show examples of very low statistics spectra. Basically all the deconvolutions behave as the author would expect. The properties of the deconvolution in the limit of very few events is quite spectacular. If events do not have significant overlap with any other events the resulting local spectrum becomes $r_i(x)$, while when more than two events have exactly the same value[2] and no significant overlap with other events, the algorithm will continue to produce a delta function.

[2]Which is highly improbable

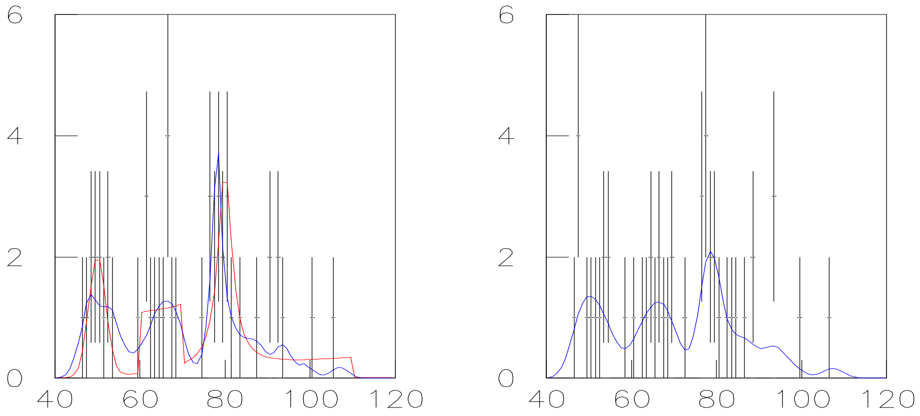

Figure 5: Same as Figure 2 with 50 events generated with Gaussian resolution functions with widths fixed to 2

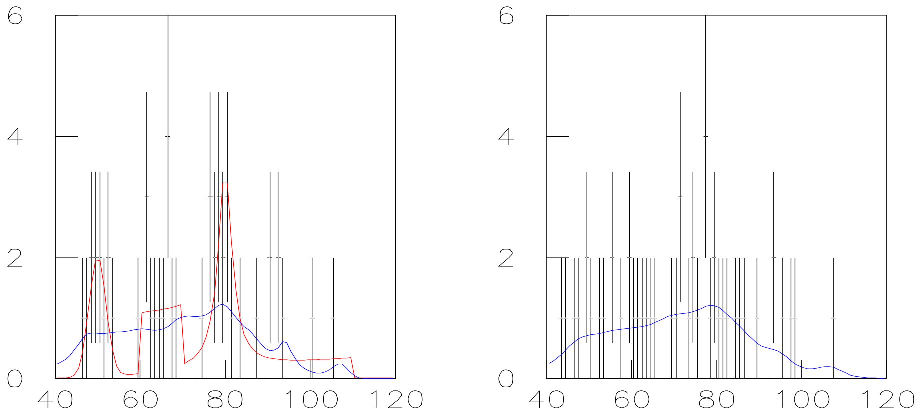

Figure 6: Same as Figure 2 with 50 events generated with Gaussian resolution functions with widths varying from 2 to 10

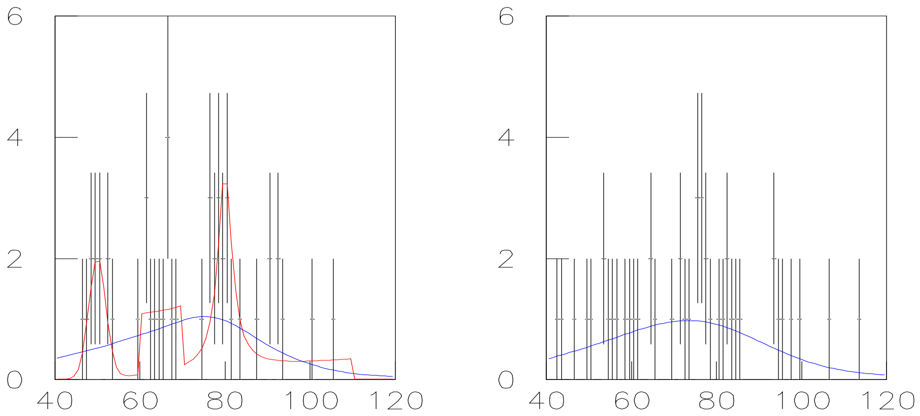

Figure 7: Same as Figure 2 with 10000 events generated with Gaussian resolution functions with widths fixed to 10

5 Conclusion

This paper has been devoted to a description of a deconvolution algorithm. The algorithm is Bayesian in nature and recursive. The deconvolution is regularized using a Frequentist's approach, where the significance of one recursion is computed locally and an update only performed when it is found to be statistically significant. The regularization is not 100% unique since it depends on the definition of local, which turns out to be a significant problem when applied to multi-dimensional deconvolution.

The algorithm is simple to implement in a few lines of computer code and converges reasonably rapid, depending on the resolution functions. The resulting generator function is found to be information conserving at the one standard deviation level, and no unwanted features of the deconvolution has been observed.

In principle a deconvolution algorithm can also be used for smoothing. The concepts of locality and MC generators suggests that this might be possible to develop a performant algorithm within the framework presented in this paper. The idea would be to use a kernel for for the resolution functions with widths greater than the found variations. The smoothing then becomes iterative with each step consisting of a convolution and a subsequent deconvolution. After each step a new local kernel is constructed as the product of the wide fixed kernel with the obtained smoothed function. Such an approach would automatically make the kernel broad in places where there is no information and narrow and non-symmetric in information containing regions, like close to phase space.

It is clear from the above discussion that this paper constitutes a tiny part of a framework for addressing optimal deconvolution and smoothing.

References

[1] N. J. Kjær, J. D'Hondt
Exact optimization of likelihood analyses in a unified Bayesian-Frequentist Monte Carlo Framework: The "k2-filter"
Contributed paper to "Advanced Statistical Techniques in Particle Physics",
Grey College, Durham, 18 - 22 March 2002