# Channel likelihood: An extension of maximum likelihood for multibody final states*

Paul E. Condon† and Paul L. Cowell†
*University of California at Irvine, Irvine, California 92664*
(Received 31 December 1973)

We present here an extension of the maximum-likelihood method. We demonstrate that the estimation of certain parameters, which are important in the analysis of multibody final states in nuclear interactions, is equivalent in power to that of maximum likelihood. We indicate how this new method gives valuable guidance in improving the nature of the hypothesis under test, and how the method can be interpreted as a technique for separating a body of experimental data into a number of resonant channels on an event-by-event basis.

## INTRODUCTION

In this paper we present an extension of maximum-likelihood analysis which we believe is particularly well suited to the analysis of multibody final states. We present first a heuristic development of the technique, second a derivation of the important equations of the technique from the postulates of maximum-likelihood analysis, a very brief discussion of some problems not covered in the derivation, and finally a description of a computer program which applies the ideas presented here. We emphasize that the second part of the paper is a derivation from the accepted principles of statistical inference, so the other discussion in the paper is an illustration of, rather than experimental evidence for, the validity of a mathematical theorem. In subsequent papers we will present the results of some applications of this technique.

## HEURISTIC INTRODUCTION OF CHANNEL LIKELIHOOD

In the analysis of multibody final states an important problem is the determination of which resonant channels are present and the relative contribution of each channel to the total body of data. (In addition to resonant channels, one would include a nonresonant phase-space channel.) It would be desirable, if possible, to identify for each event the channel which was responsible for the production of that event. We feel that the technique described here comes very close to achieving this goal. This is possible because different channels populate different regions of allowed phase space differently. In the extreme, unrealistic case where each channel populates a different (nonoverlapping) region of phase space, there is no problem in seeing that an appropriate partition of phase space will separate events by channel. Unfortunately it is not that simple in reality. There are regions of phase space which are populated by more than one channel, and we must consider how to handle the events which occur in these regions. We will work in a full set of phase-space variables, so that the regions overlap must be smaller than in any possible mass projection. In this way we retain the fullest possible information about each event. In a region of overlap, we try to apportion an event between the overlapping channels, giving some percentage of its full weight to each of the channels. Since, in general, all channels overlap to a small extent everywhere, this apportioning is really done for every event in the data and it is therefore crucial that it be done rationally. To do this we introduce some notation.

There are $N$ events of data. Each is described by a full set of variables, which we think of as a vector, or point in phase space. $\vec{R}_i$ is the phase-space point of the $i$th event in the data.

There are $M$ channels in our hypothesis (including phase space). For each channel $j$ we have a phase-space density function denoted by $\rho_j(\vec{R})$. These functions usually are Breit-Wigner resonant shapes, but they can include factors accounting for angular dependence. The normalization of the $\rho_j(\vec{R})$ is

$$I_j = \int_{\text{LIPS}} \rho_j(\vec{R}) d\vec{R}, \tag{1}$$

where LIPS indicates an integration over Lorentz-invariant phase space.

The number of events which are produced via the $j$th channel is $N_j$. There is a constraint that

$$\sum_{j=1} N_j = N. \tag{2}$$

We attempt to find the "probability that the $i$th event came from the $j$th channel." This depends on the magnitude of phase-space density $\rho_j$ evaluated at the point $\vec{R}_i$, which is to say $\rho_j(\vec{R}_i)$. It also depends on the relative sizes of the $N_j$. Suppose an event were found with an $\vec{R}_i$ such that $\rho_j(\vec{R}_i) = \rho_k(\vec{R}_i)$ for $k \neq j$ (different channels, but

same event) and suppose we already knew from other information that there are nine times as many events in channel $j$ as in channel $k$. We would say event $i$ is 90% channel $j$ and only 10% channel $k$. We define $f_{ji}=\rho_j(\vec{R}_i)$ for conciseness of notation. We have then that

$$\frac{N_j f_{ji}}{I_j}$$

is the relative probability of event $i$ having come from channel $j$. We normalize this for each event,

$$S_i=\sum_{j=1}^{M}\frac{N_j f_{ji}}{I_j} \tag{3}$$

$$w_{ji}=\frac{N_j f_{ji}}{S_i I_j}\,, \tag{4}$$

so that the $w_{ji}$ have the property $\sum_{j=1}^{M} w_{ji}=1$ for every event $i$. This $w_{ji}$ is the "weight" of event $i$ in channel $j$. Notice that if there were no overlap, for each event $i$ all of $w_{ji}$ would be zero except for one channel $j$ such that $w_{ji}=1$. A new estimate of the number of events in channel $j$ is a sum of $w_{ji}$,

$$N_j'=\sum_{i=1}^{N} w_{ji}\,, \tag{5}$$

and always has the property that $\sum_j N_j'=N$. By computing the $w_{ji}$ and by weighting each event by $w_{ji}$ we expect to get enriched samples of events for channel $j$. We sum this enriched sample to get a better estimate of how many events there are in the channel. We iterate the procedure by using $N_j'$ in the place of $N_j$ and repeating the computation of the $w_{ji}$. The iteration converges to a set of $N_j$ which are the solutions of a set of equations,

$$N_j=\sum_{i=1}^{N}\frac{N_j f_{ji}}{S_i I_j}\,,\quad j=1,\ldots,M \tag{6}$$

or

$$1=\sum_{i=1}^{N}\frac{f_{ij}}{S_i I_j}\,. \tag{6'}$$

After this solution has been found, we use the $w_{ji}$ to weight events when making invariant-mass and angular distribution plots. These plots of weighted data should have distributions somewhat like the density function used in selecting them but need not agree exactly. If a slightly wrong mass or width is used in selecting a resonant channel, the events from that channel will still be selected and when plotted will display the actual mass and width of the resonance. (At least, they will be closer to the true value than the initial assumption was.)

If a resonant channel has been left out of the hypothesis, the events due to that channel will be picked up by some other channel (probably pure phase space) and will distort the distributions in that channel in a visible way, giving a strong indication that another channel should be included in the hypothesis. If a channel is included in the hypothesis which is not present in the data, the $N_j$ for that channel will be driven to zero in the iteration.

Since the plots of weighted or selected events approximate data samples from pure resonant channels, one is able to see directly in the histograms any features of the data which were not included in the original hypothesis. Thus, unlike the standard maximum-likelihood technique, the channel-likelihood approach gives useful indication as to how the structure of the hypothesis might be improved.

## DERIVATION FROM POSTULATE OF MAXIMUM LIKELIHOOD

The solution of Eqs. (5) is an estimation of parameters, $N_j$, in a hypothesis. We will show that these equations are a subset of the full set of equations which must be solved to find a maximum-likelihood fit of a hypothesis to the data. The likelihood function $\mathcal{L}$ is defined as a product over the data:

$$\mathcal{L}=\prod_{i=1}^{N}\left(\sum_{l=1}^{M} N_l\frac{f_{li}(\sigma_l)}{I_l(\sigma_l)}\right).$$

In this expression the adjustable parameters are the $N_l$ and the $\sigma_l$. The $\sigma_l$ are "shape parameters" within each channel. We display them explicitly so as to distinguish them clearly form the "size parameters," $N_j$.

The maximum of $\mathcal{L}$ must be found subject to the constraint given in Eq. (2). We use a Lagrange multiplier, $\lambda$, to introduce the constraint into an equation for the maximum of $\ln\mathcal{L}$:

$$\begin{aligned} w&=\ln\mathcal{L}-\lambda\left(\sum_{j=1}^{M} N_j-N\right)\\ &=\sum_i \ln S_i-\lambda\sum_{j=1}^{M} N_j+\lambda N, \end{aligned} \tag{7}$$

$$\begin{aligned} 0&=\frac{\delta w}{\delta N_j}\\ &=\sum_{i=1}^{N}\frac{1}{S_i}\frac{\delta S_i}{\delta N_j}-\lambda\\ &=\sum_{i=1}^{N}\frac{f_{ji}}{S_i I_j}-\lambda,\quad j=1,\ldots,M\,. \end{aligned} \tag{8}$$

Multiply by $N_j$ and sum these equations to determine $\lambda$:

$$0=\sum_i\frac{1}{S_i}\sum_j\frac{N_j f_{ji}}{I_j}-\lambda\sum N_j\,. \tag{9}$$

However,

$$\sum N_j = N$$

and

$$\sum N_j \frac{f_{ji}}{I_j} = S_i ,$$

so

$$\lambda = 1 .$$

The set of equations (8) reduces to the set (6′), thus completing the proof.

Here we have shown that, given a set of hypothetical resonances, the number of events attributed by this technique to each resonance is precisely the same number as determined by maximum likelihood. Concerning the shape parameters and histograms made using the weights $w_{ji}$ we can make the following comments:

If, for example, one makes a histogram of the invariant mass of a pair of particles that are supposed to form a resonance and in the histogram weights the events by the $w_{ji}$ corresponding to the resonance, one sees a resonant shape in the plot. Keep in mind that what is plotted is real events (but weighted).

One might ask whether the apparent shape is a result of an actual resonance or of the method of weighting. We have already shown that the $N_j$ is a best-fit value, so if the resonance is not actually there we expect that the corresponding $N_j$ would be quite small, and even if the histogram shape is only a reflection of the hypothesis the small $N_j$ value allows one to reject the hypothesis without being confused by the histogram. If, on the other hand, the $N_j$ value is large, indicating the resonance is actually there in a best fit, then, for every event in the histogram, there is one less event in the histogram of some other channel. Thus we can see that if the resonant shape in the histogram were purely an artifact of the weighting technique, there would be a dip in the histogram of some other channel.

If the resonance is actually present in the data, but the mass value used in selecting it is slightly wrong, the histogram of data will be biased away from the true value toward the value used in making the selection. The maximum weight will be given to events with mass at the peak of the selection function, and will decline for mass values beyond this peak. Thus the apparent peak can at most be pulled to conform to the hypothesis and is actually never pulled this far. Instead the histogram is always a plot of the selected events, but is biased somewhat toward the possibly erroneous mass hypothesis. Similar comments apply for the resonant width and for angular factors.

## THE COMPUTER PROGRAM

The elements of the simple computer program necessary to employ the channel-likelihood technique are briefly described here.

The computer program used to analyze bubble-chamber data reads a tape of preselected, fitted events, solves the coupled equations (6) for the number of events in each channel $N_j$, and creates histograms of the real data for any desired quantity. The input hypothesis for each channel includes a mass distribution for a particular group of particles and possibly production and decay angular distributions. The first step generates Monte Carlo events and uses them to integrate the matrix elements for each channel according to Eq. (1). Next, every event is read from the tape, the matrix elements $\rho_j(\vec{R}_i)$ for every channel are computed at the coordinates of the event, and the table of $h_{ji} = f_{ji}/I_j$ is saved. This array can be rather large, $N \times M$, for $N$ events and $M$ channels. Starting from an initial guess for the values of $N_j$, values of $w_{ji}$ from Eq. (4) are computed and new values for $N_j$ are obtained from Eq. (5). When Eqs. (6) are solved, by either the iterative method or a more conventional minimization technique, the weight $w_{ji}$ of each event in each channel is computed from the final values of $N_j$ using Eqs. (4).

The channel-likelihood method has been tested on several samples of data from a $\bar{p}$-$p$ exposure in the 30-in. BNL bubble chamber. The iterative procedure for solving the coupled equations (6) was used to determine the $N_j$. It is an extremely simple method and converges quickly (less than 20 iterations) when the number of channels is small (less than six). It is as efficient in time and number of iterations as the more general minimization procedures when used on a large number of channels.

The simplest task for this procedure to accomplish is separating the events by channel under the assumption that the matrix elements are completely known. This is illustrated using a sample of two-prong events with a vee which fit the hypothesis $K^{\pm}\pi^{\mp}K_S$. Five possible channels were considered. The events with $K^+\pi^-$ were not distinguished from those having $K^-\pi^+$ and were considered together as part of a $K^{*0}$ channel. Similarly, the $K_S\pi^+$ and $K_S\pi^-$ events were considered together in a channel defined for $K^{*+-}$ production. Both of these were analyzed for $K^*(890)$ and $K^*(1420)$. The final channel was simple phase space. No evidence for $K^*(1420)$ was found.

A channel is defined by the matrix element computed at the event coordinates. In this case, the matrix elements used were the simplest form for

the Breit-Wigner shape with values from Ref. 2:

$$b(M,M_0,\Gamma)=\left[1+\left(\frac{M-M_0}{\frac{1}{2}\Gamma}\right)^2\right]^{-1},$$

with $M_0=0.8917$ GeV and $\Gamma=0.0501$ GeV. The matrix element for phase space is just 1. The angular distribution included was flat.

After the fit had been made to find the best values of $N_j$, the number of events in the $j$th channel, the weight or probability that the $i$th event is in that channel is found from Eq. (4). To find the distribution of events in the $j$th channel the kinematic quantities are plotted for all of the given events, but weighted by $w_{ji}$.

The mass distributions in Figs. 1 and 2 obtained from events at two different energies show the characteristic shape of the Breit-Wigner amplitude. The plots contain exactly as many events as occurred in the channel. No cuts were necessary to obtain these estimates of pure and complete samples of $K^*$ events.

No attempt to adjust the shape parameters was made in this example. From the plots shown, the agreement between the hypothesis and the data can be evaluated. The plots made using weighted

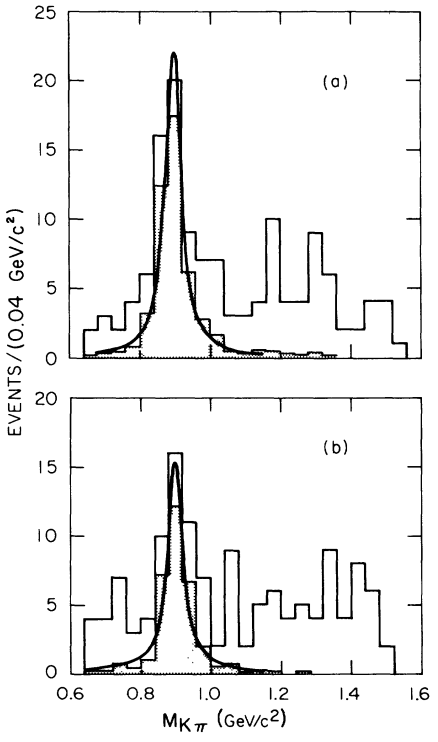


FIG. 1. Mass of $K\pi$ in 132 $K^{\pm}K_S\pi^{\mp}$ events at 0.0862 GeV/$c$ The mass distribution of all events is shown, and underneath it the mass distribution of the same events weighted (see text) for (a) the $K^{*0}$ channel and (b) the $K^{*\pm}$ channel. The curves represent the matrix element hypothesis for each channel.

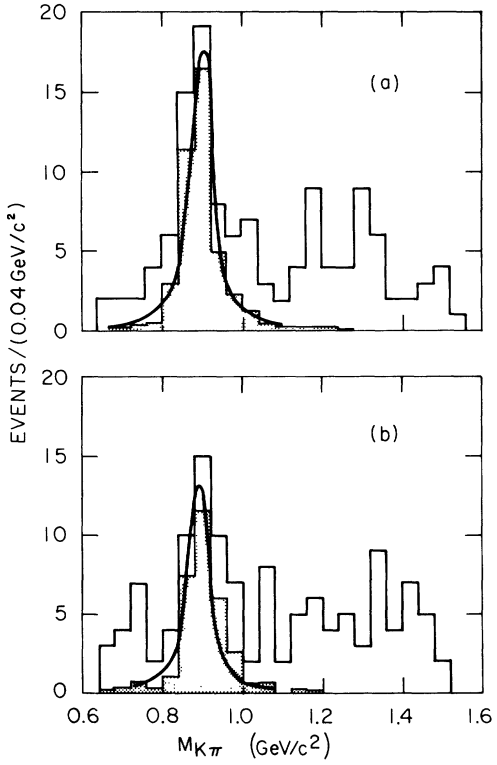


FIG. 2. Mass of $K\pi$ in 124 $K^{\pm}K_S\pi^{\mp}$ events at 0.926 GeV/$c$ (see Fig. 1).

data are not necessarily identical to the initial hypothesis. Note that the widths of the matrix elements in Figs. 1 and 2 are somewhat too small to describe the data accurately. Either the resonances are somewhat wider than the published values or a measurement error should be taken into account.

The phase-space plot in Fig. 3 contains both $\pi$-$K$

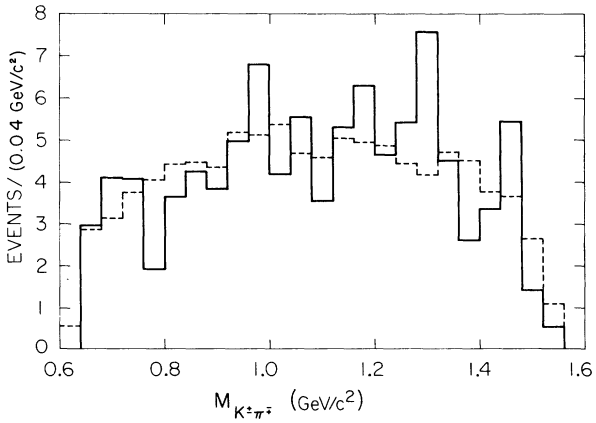


FIG. 3. Mass of $K^{\pm}\pi^{\mp}$ in 132 $K^{\pm}K_S\pi^{\mp}$ events at 0.862 GeV/$c$ weighted for phase space. For comparison, the phase-space mass distribution is shown (dashed line) generated from a Monte Carlo program.

mass combinations of the real events weighted by their weight for pure phase space and, for comparison, the same quantity plotted for Monte Carlo events. Notice that the resonance has been removed by this weighting.

A second example used a sample of four charged $\pi$ events. By extending the method further, parameters in the matrix element can be determined from the data with corresponding distributions created from Monte Carlo events. An extremely simple method for improving the mass and width parameters of the matrix element was found to work well. A parabola was fitted to both the data and the Monte Carlo mass distribution near the resonant mass. The difference in the values of the parameters was used to estimate a better mass and width for the matrix element.

It is not guaranteed that the parameters calculated each way are identical to each other, and in general they will not be. The data "pull" the distributions determined by the input matrix element toward more correct values, that is, away from the parameters put into the matrix element. After a few tries, using the fits to parabolas to quantify the disagreement and make a reasonable correction, the distributions did converge to a stable limit, which presumably corresponds to the best values. The general agreement of the actual distributions with the input matrix element can be compared if it is felt that some confirmation of the form of the matrix element is needed. For example, the various forms for a mass- or momentum-dependent width produce markedly different agreement. Angular distributions can also be determined by starting with a flat angular distribution and fitting the resulting angular distribution to Legendre polynomials, for instance. Putting this hypothesis into the next iteration will cause the angular distribution to change further until it converges to a stable limit after several steps.

## DISCUSSION

We set out to find a way of identifying which channel was responsible for each event in a body of data. In our examples we did find that usually for each event, $i$, there was one channel, $j$, for which the weight $w_{ji}$ was much larger than the weight for any other channel. We identify event $i$ as having been produced by channel $j$.

Our claim that this most probable channel is the channel causing the event is in our opinion no different in principle from the claim that the values of parameters found by maximum likelihood are the values which should be used in describing nature.

Unlike the usual way of presenting mass plots, this technique does not cut the data and throw away what fails the cut. All data must be compared to some part of the hypothesis, and one is not finished until all comparisons show consistency. Our feeling is that the technique offers valuable guidance to the experimentalist in improving the form of the total hypothesis under test. In addition we have proved that, given an hypothetical form, the fit is the best (i.e., maximum likelihood).

The present work was inspired by the paper of Brau *et al.*[1] on "prism plots." However, we do not use prism plot variables, nor, in fact, do we use any special set of variables. When doing channel likelihood analysis, we use a variety of variables, mostly invariants masses, choosing at each stage the variables most convenient for that part of the computation. Our method of computing a channel likelihood for each event seems to be heuristically similar to the event tagging mentioned in their paper. There are four significant differences. We are able to connect our method with the established method of maximum likelihood. We use Monte Carlo techniques solely for the purpose of integrating over phase space; there is no underlying hypothesis about the nature of the interactions taking place (such as the supposition that they are highly peripheral). It is now seen that the $N_j$ are a solution of a set of equations for maximum likelihood, and can be solved by any method, such as common minimization procedures, and not just by the illustrative iterative procedure developed heuristically.

---

*Work supported by the U. S. Atomic Energy Commission.

†On leave; present address: Stanford Linear Accelerator Center, Stanford, California 94305.

[1]J. E. Brau, F. T. Dao, M. F. Hodous, I. A. Pless, and R. A. Singer, Phys. Rev. Lett. 27, 1481 (1971).

[2]Particle Data Group, Rev. Mod. Phys. 45, S1 (1973).