

Lawrence Berkeley National Laboratory

Recent Work

Title

Fits of Monte Carlo Distributions to Data

Permalink

<https://escholarship.org/uc/item/1vm984pv>

Journal

Nuclear Instruments and Methods in Physics Research A, 326

Authors

Eberhard, P.
Lynch, G.R.
Lambert, D.

Publication Date

1992-07-30



Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

Physics Division

Submitted to Nuclear Instruments and Methods in Physics Research A

Fits of Monte Carlo Distribution to Data

P. Eberhard, G. Lynch, and D. Lambert

July 1992



LOAN COPY
Circulates
for 4 weeks
Bldg. 50 Library.

LBL-32681

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-32681

Fits of Monte Carlo Distributions to Data¹

**Philippe Eberhard,
Gerald Lynch,
and
David Lambert**

**Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720**

July 31, 1992

¹This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics, of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

Fits of Monte Carlo Distributions to Data

Philippe Eberhard,
Gerald Lynch,
and
David Lambert

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

July 31, 1992

ABSTRACT

This paper concerns likelihood fits to binned data in which the contents of the bins are simple counts of independent events. It addresses the case where the distribution to be fitted is the sum of distributions of events due to different physical processes, and the prediction of each one of these processes is given by a distribution of events generated by a Monte Carlo program, binned like the real events. This paper suggests a method to analyze the properties of these individual processes.

1 The likelihood fit.

An ordinary likelihood fit to binned data is one in which one can calculate exactly the probability p_j , as a function of some parameters, that an event falls

in bin j . The content of the bin j is n_j . The log of the likelihood is

$$L_{00} = \sum_j n_j \ln p_j \quad (1)$$

and, of course, $\sum p_j = 1$ for any values of the parameters. One estimates the parameters involved in the expression of the p_j by maximizing L_{00} with respect to variations in these parameters [1].

The extended likelihood is one where the content of each bin is considered to be the outcome of a Poisson process. In some cases, when the total sample size is fixed, the distribution is really a multinomial distribution to be fitted with a fixed normalization. However, if another variable parameter is added to the others in the expression of the expected number of events and if this additional parameter allows free normalization, the extended likelihood method will give that additional parameter a value that normalizes the fitted distribution to the total number of events and the other parameters will acquire the same values as in the ordinary likelihood method. For the extended likelihood, the probability to get n_j events in bin j is

$$p_j = e^{-\lambda_j} \lambda_j^{n_j} / n_j! , \quad (2)$$

where λ_j is the expectation value of the content of the bin [1]:

$$\langle n_j \rangle = \lambda_j . \quad (3)$$

Then the log likelihood takes the form

$$L_0 = \sum_j (n_j \ln \lambda_j - \lambda_j) , \quad (4)$$

where the $n_j!$ term has been dropped because it generates only a constant in the fit of the parameters in L_0 of Eq. (4). In Eq. (4), $n_j \ln \lambda_j$ has to be taken to be zero if both n_j and λ_j are zero. In the fit the λ_j are fit independently and are not constrained to sum up to the total number of entries. However, in many cases, the fit will adjust the normalization to be equal to the total number of real events.

The maximum value that L_0 could possibly have (for a fixed number of events) is when all $\lambda_j = n_j$. If we subtract this maximum value from the log-likelihood we get the form

$$L_{overall} = \sum (n_j \ln \lambda_j - \lambda_j - n_j \ln n_j + n_j) . \quad (5)$$

This is a convenient form because if one defines $\omega = -2L_{overall}$, then this ω becomes an overall χ^2 in the limit of large statistics [1]. In Eq. (5), when n_j is zero, one must evaluate $n_j \ln n_j$ as zero.

2 Predictions given by Monte Carlo distributions.

In the case addressed by this paper, the distribution of the real events is the sum of contributions of N different processes, and the predictions for each of these processes are given in terms of a distribution of Monte Carlo events [2]. Let m_{ij} be the number of Monte Carlo entries generated for process i in bin j . There is an expectation value f_{ij} for m_{ij} . It is the average of m_{ij} , if the same Monte Carlo generation were repeated an infinite number of times:

$$\langle m_{ij} \rangle = f_{ij} . \quad (6)$$

We assume that there is no analytic form for the f_{ij} .

The expected number of real events in bin j , λ_j is a sum of N contributions of the processes i :

$$\lambda_j = \sum_i a_i f_{ij} . \quad (7)$$

The parameter a_i expresses the relative importance of process i in the distribution of the real events. One wants to obtain these parameters from the fit.

The m_{ij} are affected by Monte Carlo fluctuations but, within statistical error, they are equal to the f_{ij} 's. We can treat the f_{ij} as if they were parameters to be determined from their estimate m_{ij} in the same procedure that determines the a_i 's. For this purpose one can maximize

$$L = \sum_j \left[n_j \ln \lambda_j - \lambda_j + \sum_i (m_{ij} \ln f_{ij} - f_{ij}) \right] . \quad (8)$$

This problem now involves the determination of many parameters. Fortunately, these parameters do not need to be determined by the methods used by the standard maximizing programs. The problem can be broken down into several pieces. L of Eq. (8) can be maximized using a standard program while varying only the N values of a_i 's and, at each step of the procedure, computing the best values of f_{ij} for the current set of a_i . For this there is a straight-forward algorithm equivalent to a one parameter fit, which determines λ_j and then the f_{ij} in one bin after the other. These best f_{ij} can then be introduced in the evaluation of L .

In the expression of L of Eq. (8), any one of the f_{ij} is involved only in the term relative to the corresponding bin j . It follows that the derivative with respect to one f_{ij} does not contain any $f_{ij'}$ for any $j' \neq j$:

$$u_{ij} = \frac{\partial L}{\partial f_{ij}} = \left(\frac{n_j}{\lambda_j} - 1 \right) a_i + \frac{m_{ij}}{f_{ij}} - 1 . \quad (9)$$

At the maximum of L , $u_{ij} = 0$ for all i and j , except of course for those i and j with f_{ij} stuck against the boundary of the physical region. The physical region

is defined as

$$f_{ij} \geq 0. \quad (10)$$

Thus, at the maximum of L in the physical region, u_{ij} may not be zero if f_{ij} is.

2.1 The general case, all $m_{ij} \neq 0$.

If $m_{ij} \neq 0$, f_{ij} cannot be zero without making L of Eq. (8) equal to $-\infty$. Then, for $m_{ij} \neq 0$, u_{ij} of Eq. (9) is zero at the maximum of L . Setting $u_{ij} = 0$ and solving for f_{ij} ,

$$f_{ij} = \frac{m_{ij}}{1 - a_i \left(\frac{n_j}{\lambda_j} - 1 \right)}. \quad (11)$$

This is an implicit equation since λ_j is bound to the f_{ij} 's for the same j by Eq. (7). The only solutions that make sense are those for which

$$\frac{n_j}{\lambda_j} < \frac{1 + a_i}{a_i} \quad (12)$$

for all i 's with $m_{ij} \neq 0$. Otherwise f_{ij} and m_{ij} could not both be positive.

From Eq. (11) for all the f_{ij} for a given j where $m_{ij} \neq 0$, and from Eq. (7), we get

$$\lambda_j = \sum_i a_i f_{ij} = \sum_i \frac{a_i m_{ij}}{1 - a_i \left(\frac{n_j}{\lambda_j} - 1 \right)} = \lambda_j \sum_i \frac{c_i m_{ij}}{\lambda_j - c_i n_j}, \quad (13)$$

where

$$c_i = \frac{a_i}{1 + a_i}. \quad (14)$$

2.1.1 Solving for λ_j .

Let a_M be the largest of the a_i 's for which $m_{ij} \neq 0$. Multiplying both sides by $\frac{1}{\lambda_j}(\lambda_j - c_M n_j)$, we get

$$\lambda_j - c_M n_j = \sum_{i \neq M} c_i m_{ij} \frac{\lambda_j - c_M n_j}{\lambda_j - c_i n_j} + c_M m_{Mj} \quad (15)$$

$$\lambda_j = c_M n_j + F(\lambda_j), \quad (16)$$

where

$$F(\lambda) = c_M m_{Mj} + \sum_{i \neq M} c_i m_{ij} \frac{\lambda - c_M n_j}{\lambda - c_i n_j}. \quad (17)$$

For $\lambda = c_M n_j$,

$$F(\lambda) = c_M m_{Mj} > \lambda - c_M n_j. \quad (18)$$

For $\lambda \rightarrow \infty$,

$$F(\lambda) \rightarrow F(\infty) = \sum_i c_i m_{ij} < \lambda - c_M n_j . \quad (19)$$

Therefore, because $F(\lambda)$ is continuous, there is at least one value λ_j of λ , between $\lambda = c_M n_j$ and $\lambda = +\infty$, that satisfies Eq. (16).

From the expression of $F(\lambda)$ in Eq. (17), it is clear that the first derivative of F is always positive and the second derivative is always negative. F increases continuously from $c_M m_{Mj}$ to $F(\infty)$ as shown on Fig. 1. On that figure, we also show the straight line $\lambda - c_M n_j$. The solution λ_j is at the intersection of the line and the curve. There is only one solution. And it satisfies inequality (12).

A procedure to reach the solution consists of taking any value of $\lambda > c_M n_j$ as a potential value for λ_j , compute a new value of λ_j using Eq. (16), and use this λ_j again in Eq. (16); and so on in an iterative procedure. In Fig. 1, one can see that this procedure is equivalent to taking a point on the curve, move horizontally till one reaches the straight line, then vertically till we reach the curve, and so on. The procedure obviously converges.

As a convergence criterion, one can use the following recipe. Let ϵ be the limit we tolerate for the difference between the true value of λ_j and the value we are going to obtain. When any step makes a difference $< \epsilon$ between two subsequent values of λ_j , take a step of size ϵ instead of the one computed. Then stop when the computed step changes sign.

After obtaining a value for λ_j , one can compute all values of f_{ij} for all i 's using Eq. (11).

2.2 Special cases: some $m_{ij} = 0$.

Special cases are cases where, for a bin j , there are some processes i for which $m_{ij} = 0$. Then, for that i ,

$$u_{ij} = \frac{\partial L}{\partial f_{ij}} = \left(\frac{n_j}{\lambda_j} - 1 \right) a_i - 1 . \quad (20)$$

Now f_{ij} may be zero at the maximum of L in the physical region defined by Eq. (10). The best value of L corresponds to derivatives u_{ij} equal to zero for the $f_{ij} \neq 0$ and negative for the $f_{ij} = 0$, i.e. at the edge of the physical region. This way all changes within the physical region result in L being decreased.

2.2.1 Some $m_{ij} = 0$ but not for the largest a_i .

In this case, we also call a_M the largest of the a_i 's. Then $m_{ij} = 0$ only for parameters $a_i < a_M$. The solution is to apply Eq. (11) to parameters a_i for which $m_{ij} > 0$ and take $f_{ij} = 0$ for the special a_i 's for which $m_{ij} = 0$. That procedure insures $u_{ij} = 0$ for $m_{ij} \neq 0$ and condition (12) for a_M as for the other

a_i 's. Thus for these special a_i where $m_{ij} = 0$, using Eq. 20,

$$u_{ij} = \left(\frac{n_j}{\lambda_j} - 1 \right) a_i - 1 \leq \left(\frac{n_j}{\lambda_j} - 1 \right) a_M - 1 < 1 + a_M - a_M - 1 = 0, \quad (21)$$

therefore u_{ij} is negative when $f_{ij} = 0$.

Then L is maximum in the physical region. Note that f_{ij} becomes 0 anyway for $m_{ij} = 0$ when using Eq. (11). Therefore Eq. (11) can be used for all the i 's as in the general case of Sect. 2.1. This special case does not require special treatment.

2.2.2 $m_{ij} = 0$ for the largest of the a_i 's.

In this case we call $a_{M'}$ the largest of the a_i 's and a_M the largest of the a_i 's for which $m_{ij} \neq 0$. That definition of a_M is consistent with the one used in the previous cases. One can use the standard procedure of Sect. 2.1 to solve Eqs. (16) and (17) with the present definition of a_M introduced in both of them. Three cases may arise.

- (a) The standard procedure with the present definition of a_M introduced in Eqs. (16) and (17), yields a value of λ_j such that condition (12) is satisfied for $i = M'$. Then it is also satisfied for all i , as can be shown writing inequalities like (21) but where $M \rightarrow M'$. It follows that, for $m_{ij} = 0$, all u_{ij} are negative for that value of λ_j and the best values satisfying inequality (10) are $f_{ij} = 0$. Then λ_j has to be taken at the value obtained by the iteration procedure described above and, as in the case where all $m_{ij} \neq 0$, all f_{ij} 's are finally given by Eq. (11).
- (b) The standard procedure with the present definition of a_M yields a value λ_j such that $a_{M'}$ violates condition (12). Then $u_{M'j}$ is positive for that value of λ_j and L can be increased by taking $f_{M'j}$ positive rather than zero. A solution with derivatives equal to zero for $f_{ij} \neq 0$ and negative for $f_{ij} = 0$ can be obtained by defining

$$\lambda'_j = n_j c_{M'} = n_j \frac{a_{M'}}{1 + a_{M'}}, \quad (22)$$

writing

$$\lambda_j = \lambda'_j, \quad (23)$$

and computing all f_{ij} for $i \neq M'$ by Eq. (11). Then

$$f_{M'j} = \frac{\lambda'_j - \sum_{i \neq M'} a_i f_{ij}}{a_{M'}}. \quad (24)$$

It is easy to check that $\lambda_j = \lambda'_j$ makes $u_{M'j} = 0$ in Eq. (20). For all other i 's with $m_{ij} = 0$, f_{ij} is zero and u_{ij} given by Eq. (20) is negative, because they correspond to $a_i < a_{M'}$. Eq. (11) also insures that all derivatives u_{ij} are zero for $m_{ij} \neq 0$. Therefore we have a solution if $f_{M'} > 0$. To verify that $f_{M'} > 0$, we use Eqs. (17), (11), and (24), to deduce

$$F(\lambda'_j) = (\lambda'_j - c_M n_j) \sum_i \frac{c_i m_{ij}}{\lambda'_j - c_i n_j} \quad (25)$$

$$= \frac{\lambda'_j - c_M n_j}{\lambda'_j} \sum_{i \neq M'} c_i (1 + a_i) f_{ij} \quad (26)$$

$$= \frac{\lambda'_j - c_M n_j}{\lambda'_j} (\lambda'_j - a_{M'} f_{M'j}) . \quad (27)$$

Since the solution of Eq. (16) violates condition (12), the solution of Eq. (16) is $< \lambda'_j$. Therefore, Fig. 1 shows that

$$\lambda'_j - c_M n_j > F(\lambda'_j) = (\lambda'_j - c_M n_j) \left(1 - \frac{a_{M'} f_{M'j}}{\lambda'_j} \right) . \quad (28)$$

Therefore we have indeed

$$f_{M'j} \text{ of Eq. (24)} > 0 . \quad (29)$$

- (c) The procedure spelled out above cannot be applied because all m_{ij} 's are zero. The solution is all f_{ij} 's = 0, except

$$f_{M'j} = \frac{n_j}{1 + a_{M'}} . \quad (30)$$

Of course, if all m_{ij} and n_j are zero, then λ_j and all f_{ij} are zero.

2.3 Normalization.

Whatever the procedure to get λ_j is, the condition of $u_{ij} = 0$ except if f_{ij} is zero, and Eq. (9) imply

$$f_{ij} u_{ij} = \left(\frac{n_j}{\lambda_j} - 1 \right) a_i f_{ij} + m_{ij} - f_{ij} = 0 . \quad (31)$$

Therefore using Eq. (8),

$$\sum_j f_{ij} = \sum_j m_{ij} + a_i \frac{\partial L}{\partial a_i} . \quad (32)$$

When L is maximum, $\frac{\partial L}{\partial a_i}$ is zero for all a_i except for those a_i 's at the edge of the physical boundary, i.e. those a_i 's equal to zero. Therefore

$$a_i \frac{\partial L}{\partial a_i} = 0 . \quad (33)$$

Therefore, at the maximum for f_{ij} 's and a_i 's,

$$\sum_j f_{ij} = \sum_j m_{ij} . \quad (34)$$

the normalization of f_{ij} is the same as the ones of the Monte Carlo events. If Eq. (31) is summed over i , using Eq. (3)

$$(n_j - \lambda_j) + \sum_i (m_{ij} - f_{ij}) = 0 . \quad (35)$$

Therefore, using Eq. (34), the sum of the fitted values λ_j over all bins equal the number of real events,

$$\sum_j \lambda_j = \sum_j n_j . \quad (36)$$

3 Applications.

Finding out the number of events due to each process is often not the only goal. Sometimes one has to find out also what kind of properties the events produced by each individual process have.

The configuration of an event are quantities such as number of tracks, particle identities, angles and momenta. These quantities we will refer to by the symbol x . The bin j in which the event falls depends on the configuration x . Event number k has a configuration x_k and falls into bin j_k . The properties of the processes are characterized by parameters b_l 's and by the probability distribution $P_{ij}(b_l, x)$ of configuration x if the event is due to process i and if it falls into bin j . If events due to different processes were contained in different samples, one could estimate the b_l 's by varying them and maximizing

$$L_i = \sum_k \ln P_{ij}(b_l, x_k) , \quad (37)$$

with the sum extended to the events due to each process i separately. However, in general, samples contain events from several processes. In such a case, one can first determine the a_i 's and the f_{ij} 's in a preliminary fit using the method of Sect. 2, then estimate the parameters b_l 's by varying them and maximizing

$$L_{00} = \sum_k \ln \left[\sum_i a_i f_{ij_k} P_{ij}(b_l, x_k) \right] , \quad (38)$$

where j_k is the bin that contains event k . In this second fit the a_i 's and the f_{ij} 's are fixed at those values obtained in the preliminary fit. After that second fit, one may want still to improve the determination of the b_l 's and the a_i 's by letting both a_i 's and b_l 's vary and maximizing

$$L_0 = L_{00} - \sum_{ij} a_i f_{ij} = \sum_k \ln \left[\sum_i a_i f_{ij} P_{ij}(b_l, x_k) \right] - \sum_{ij} a_i f_{ij} . \quad (39)$$

Of course the best accuracy would be obtained if one could then let all a_i 's, f_{ij} 's and b_l 's free and maximize

$$L = L_0 + \sum_{ij} (m_{ij} \ln f_{ij} - f_{ij}) . \quad (40)$$

However, the number of free parameters would usually be too large to be introduced in a maximizing routine. If, as in Sect. 2, one tries to write equations to maximize L as a function of the f_{ij} for each bin separately, one ends up with equations with more than one unknown for each bin. This was not the case in Sect. 2 because, there, the only unknown was λ_j to be determined by Eq. (16). The existence of several unknowns may raise convergence problems much harder to control. Therefore, we recommend making the fit maximizing L_0 varying only a_i and b_l at this point. The result will still be a consistent estimate of the parameters a_i and b_l , though not quite as accurate in principle as if one would vary all parameters including the f_{ij} together and maximize L of Eq. (40).

3.1 Error matrix.

If only the a_i 's are estimated and if the procedure of Sect. 2 is used, the estimation is made by maximizing the likelihood function L of Eq. (8) in the space of the a_i 's and of the f_{ij} 's. Suppose the second derivative matrix in the space of only the a_i 's is evaluated numerically with the f_{ij} 's adjusted according to the rules of Sect. 2, i.e. having the f_{ij} 's constantly optimized for each set of value of a_i . Then the Monte Carlo uncertainties are taken into account by that readjustment of the f_{ij} 's for each set of a_i . The error matrix is the inverse of the second derivative of L of Eq. (8) in the space of the a_i 's and it includes the Monte Carlo error.

If other parameters b_l 's are estimated along with the a_i 's and if the procedure of Sect. 3 is used, the final fit involves maximizing L_0 of Eq. (40) varying the a_i 's and the b_l 's but keeping the f_{ij} 's constant. The inverse of the second derivative matrix of that L_0 in the space of the a_i 's and b_l 's is an estimate of the error matrix on the a_i 's and b_l 's due to the statistical uncertainties of the real events, i.e. if there were no error on the Monte Carlo. Another error matrix due to Monte Carlo errors has to be added to it. That other error matrix can be estimated in the way that follows.

Suppose large numbers N_i^{MC} of Monte Carlo events have been generated for each process i and used in the fitting procedure. They can be divided in, let us say, 10 samples of $\frac{N_i^{MC}}{10}$ events for each process. We number them by the index ν running from 1 to 10. Then 10 fits of the type described as “final” fits in Sect. 3 can be made with the same real events but with the 10 different Monte Carlo samples. The fit with sample # ν will give results having values $\alpha_{i,\nu}$ and $\beta_{\ell,\nu}$ for a_i and b_ℓ respectively. We can compute the matrix elements

$$E_{i,i'}^{MC} = \frac{1}{100} \sum_{\nu} \alpha_{i,\nu} \alpha_{i',\nu} - \frac{1}{1000} \sum_{\nu} \alpha_{i,\nu} \sum_{\nu} \alpha_{i',\nu} \quad (41)$$

$$E_{i,\ell}^{MC} = \frac{1}{100} \sum_{\nu} \alpha_{i,\nu} \beta_{\ell,\nu} - \frac{1}{1000} \sum_{\nu} \alpha_{i,\nu} \sum_{\nu} \beta_{\ell,\nu} \quad (42)$$

$$E_{\ell,\ell'}^{MC} = \frac{1}{100} \sum_{\nu} \beta_{\ell,\nu} \beta_{\ell',\nu} - \frac{1}{1000} \sum_{\nu} \beta_{\ell,\nu} \sum_{\nu} \beta_{\ell',\nu} \quad (43)$$

The elements $E_{i,i'}^{MC}$, $E_{i,\ell}^{MC}$, and $E_{\ell,\ell'}^{MC}$ are the matrix elements of the Monte Carlo error matrix E^{MC} . They should be added to the inverse of the second derivative matrix of L_0 of Eq. (39) to obtain the total error matrix, i.e. one that includes Monte Carlo errors too.

4 An example of these methods.

These methods are planned to be used in an analysis of jets produced by the creation of quark pairs of various flavors in e^+e^- annihilation. The analysis has two parts:

1. measure what fraction of the events produced at any given energy are light-, charm-, and bottom-quark events; and
2. measure how many pions, kaons, and protons are produced, as a function of momentum, for these three types of events.

There is no analytic expression for the distribution of events in the space of measurable parameters. However, for each one of the three quark types, there are Monte Carlo algorithms to generate events having distributions expected from events due to that process. Therefore, one can generate specific distributions of Monte Carlo events corresponding to each process, bin them in variables expected to show a large difference of behavior between the three processes, and apply the methods of this paper.

Three samples of Monte Carlo events, each one composed entirely of events due to only one process, are generated by the program Jetset 6.3 [3] and used in both parts of the analysis. Our choice of variables to distinguish the three processes involves a neural network. From the kinematical data of a real or

a Monte Carlo event, a neural network computes quantities called “outputs” which it considers giving the best separation among the different distributions for different processes.¹ The first part of the analysis consists of fitting the binned distribution of neural network outputs for data to a linear combination of the binned distributions of outputs for the Monte Carlo samples of light-, charm-, and bottom-quark jets, as described in Sect. 2. Following the notation of Sect. 2, m_{ij} is the number of Monte Carlo jets in bin j of quark type i , and n_j is the number of jets from data in bin j . To optimize the log-likelihood L of Eq. (8) with respect to the three a_i , the program MINUIT [4] is used. During each call to the user-supplied routine FCN, which computes the function that MINUIT minimizes, the f_{ij} for the current values of the a_i are calculated, bin by bin, using the algorithm described in Sects. 2.1 and 2.2.

The method has been tried with another set of Monte Carlo events replacing the experimental data. These other events were generated by the three processes at the same time, each in a known amount. In this trial, the fits of the a_i easily converge to values consistent with the known amounts. The a_i have reasonably parabolic one-sigma errors, and are highly correlated in this analysis, since the three Monte Carlo distributions (especially the light- and charm-quark ones) look quite similar.

For the second part of the analysis, events are binned according to the characteristics of one jet (the tagging jet) and the number of tracks of all kinds is studied in the other (tagged) jet. The characteristics of the tagged jet are assumed to depend only on the quark flavor, not on any other quantity correlated with the variables of the tagging jet used to do the binning. Different bins are treated as samples of events with different proportions of each quark flavor. From the characteristics of particles in the tagged jet, i.e. from their momenta and dE/dx values, we determine the average charged hadron multiplicities and hadron spectra for the three types of events using the method of Sect. 3.

Again, the neural network is used to determine the variables of the tagging jet used for the binning labeled j in this paper. The charged hadron multiplicity of each particle type r in each jet flavor i is assumed to be Poisson distributed; the corresponding expectation values b_{ir} are the parameters to be fitted. Thus $P_{ij}(b_L, x_k)$ in Eq. (37) is

$$\exp\left(-\sum_r b_{ir}\right) \prod_{t_k} \left[\sum_r b_{ir} G_r(R_{t_k}) \right], \quad (44)$$

¹The neural network is used as follows: using only two-jet events, the track information in a jet is used to compute a set of inputs to a feed-forward neural network that was trained to distinguish, as far as possible, light-, charm-, and bottom-quark jets. For each set of inputs, the neural network produces a set of two independent outputs that indicate the likelihood that the jet used to compute the inputs originated from a light-, a charm-, or a bottom-quark. We use a neural network only because it is a particularly effective method of achieving flavor separation. One can apply the methods described in this paper regardless of the variables used to differentiate the different processes.

where r is the particle type (e , π , K , or p), t_k is the index of tracks in jet k , R is dE/dx , and $G_r(R)$ is a Gaussian-distributed probability density that a track of type r has dE/dx R . Note that this probability is independent of the bin number j , since correlations between tagging jet and track information of the opposite jet, except for flavor tagging of course, are neglected.

For the charged hadron spectrum, the same approach is taken. We assume the number of particles of each type in each jet in different momentum bins s is Poisson distributed. The expected number of particles of type r in bin s produced in jets of flavor i are the parameters b_{irs} to be fitted. Thus, the function $P_{ij}(b_\ell, x_k)$ in Eq. (37) is

$$\exp\left(-\sum_{rs} b_{irs}\right) \prod_{t_k} \left[\sum_{rs} b_{irs} G_r(R_{t_k}) \delta_s(t_k) \right], \quad (45)$$

where $\delta_s(t)$ is 1 if track t falls in momentum bin s and 0 otherwise.

5 Summary.

To take all cases into account we recommend the following procedure:

5.1 Preliminary fit.

Maximize L of Eq. (8) changing the a_i 's only. For each set of a_i 's, the value of all f_{ij} 's in all bin j 's, should be determined, bin by bin, using the following procedure:

1. If $m_{M'j} \neq 0$, ($a_{M'}$ is the largest of the a_i 's), solve for λ_j using Eq. (16) iteratively, starting with any value of $\lambda_j \geq c_{M'} n_j$. Note that, in this case, our definitions imply $M = M'$. Then compute all f_{ij} using Eq. (11).
2. If $m_{M'j} = 0$, start the procedure of solving for λ_j as above starting at the value $\lambda_j = \lambda'_j$ given by Eq. (22) (Note that in this case, $\lambda'_j > c_{M'} n_j$). The first step of iteration and the possibility to compute it will indicate if we deal with case (a), (b) or (c) of Sect. 2.2.2. Indeed, as can be seen on Fig. 1, the first step is in the same direction as the solution λ_j of Eq. (16) thus, at $\lambda = \lambda'_j$ it tells us if condition (12) is satisfied for $a_{M'}$. Therefore:
 - (a) If the first step of iteration leads to a value $< \lambda'_j$, continue iteration and proceed as in item 1 where $m_{M'j}$ was not zero.
 - (b) If the first step gives a value $< \lambda'_j$, stop iterating. Make $\lambda_j = \lambda'_j$ and compute f_{ij} for $i \neq M'$ using Eq. (11) and $f_{M'j}$ using Eq. (24).
 - (c) If all the $m_{ij} = 0$, all f_{ij} should be set to zero except $f_{M'j}$, which should have the value given by Eq. (30).

5.2 Fit of the b_l 's.

Maximize L of Eq. (40) changing only the b_l 's and using the values of a_i and f_{ij} obtained in the preliminary fit of Sect. 5.1.

5.3 Fit of the a_i 's and of the b_l 's.

Maximize L of Eq. (40) changing both the a_i 's and the b_l 's. The f_{ij} 's have still the values determined by the preliminary fit of Sect. 5.1.

5.4 The error matrix.

Divide the sample of Monte Carlo events in 10 (for instance) samples of equal numbers of events, so that, for each process, the number of events be 10 times less than what it is in the original sample. Repeat the fit described just above in Sect. 5.3 with each one of these 10 samples. For the fit obtained with sample $\# \nu$, the results are $a_i = \alpha_{i,\nu}$ and $b_l = \beta_{l,\nu}$. Elements of a matrix E^{MC} are obtained using Eqs. (41), (42), and (43). That matrix represents the contribution of the Monte Carlo error to the uncertainty of the results. The contribution of the statistical error on the real events is the inverse of the second derivative matrix of L_0 of Eq. (39), maximized in the procedure of Sect. 5.3. The two matrices should be added to get the total error matrix.

Acknowledgements

The authors are indebted to Dr. M.T. Ronan for discussions and comments about this paper.

This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics, of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

References

- [1] W.T. Eadie, D. Drijard, F.E. James, M. Roos, and B. Sadoulet, *Statistical Methods in Experimental Physics* (North Holland, London, 1971).
- [2] F. James, Monte Carlo Theory and Practice, *Rep. Prog. Phys.* 43 (1980) 1145; reprinted in: *Experimental Techniques in High Energy Physics*, T. Ferbel, ed. (Addison-Wesley, Menlo Park, CA, 1987).
- [3] T. Sjöstrand and M. Bengtsson, Lund Report LU-TP-86-22, Univ. of Lund, 1986.
- [4] F. James and M. Roos, *Comp. Phys. Comm.* 10 (1975), 343.

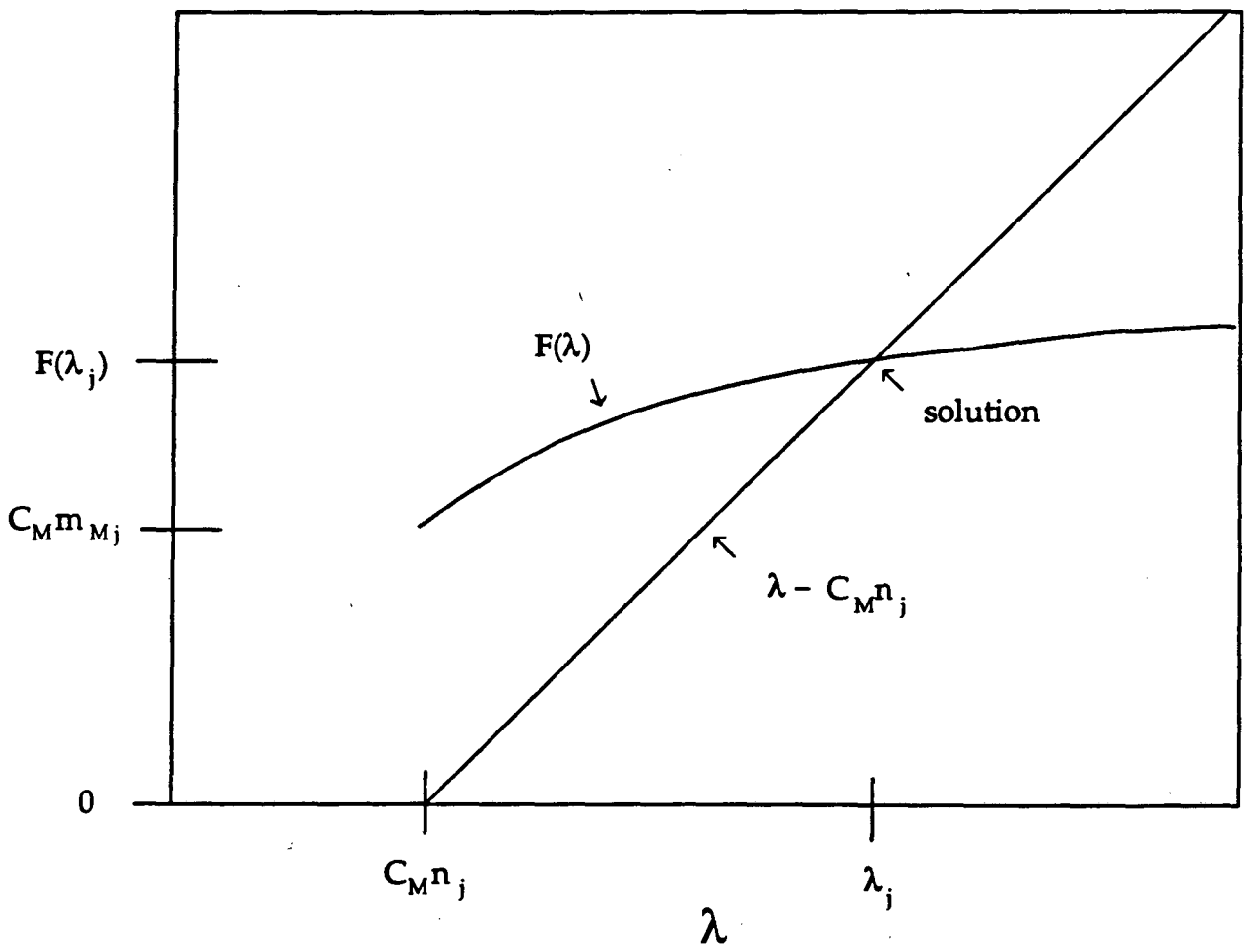


Figure 1

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720