Fermilab Library
0 1160 0033823 0

CRN-93-65

BOURG

CRN 93-65
December 1993

FERMILAB
MAY 02 1994

Correction of measured multiplicity distributions by a genetic-like algorithm

Mohamed Hafidouni

CENTRE DE RECHERCHES NUCLEAIRES
STRASBOURG

IN2P3
CNRS

UNIVERSITE
LOUIS PASTEUR

CRN 93-65
December 1993

Correction of measured multiplicity distributions by a genetic-like algorithm

Mohamed Hafidouni*

Centre de Recherches Nucléaires
B.P. 20, 67037 Strasbourg, France

Abstract. An algorithm is proposed for the correction of observed multiplicity distributions by a maximum entropy method. This algorithm gives results which are compatible with those from more time consuming simulated annealing method.

Introduction.

Observed multiplicity distributions have to be corrected for detector efficiency and limited geometrical acceptance. A method [1] to correct observed multiplicity distributions consists in minimizing

$$F = -S + \lambda\, \chi^2 \tag{1}$$

where

$$S = -\sum_{n=1}^{N} t_n \; ln \; t_n, \qquad t_n = \frac{T_n}{\sum_{i=1}^{N} T_i} \tag{2}$$

T_n is the fraction of events with true multiplicity in bin n and N the total number of bins.

$$\chi^2 = \sum_{m=1}^{M} \left(\frac{O_m - \sum_{n=1}^{N} P_{mn} T_n}{\sigma_m} \right)^2 . \tag{3}$$

O_m being the fraction of detected events with observed multiplicity in the bin m, P_{mn} is the probability that an event with a true multiplicity in the bin n be observed as an event with observed multiplicity

* CRNVAX::HAFIDOUNI or HAFIDOUNI@CRNVAX.IN2P3.FR

in the bin m and σ_m is the statistical error of bin m. λ is a Lagrange multiplier that determines the relative strength of S and χ^2.

In the following an algorithm to minimize (the cost function) F is described.

The algorithm.

One starts with the configuration $(T_1^0, T_2^0, ..., T_N^0)$, which we call C_1, where the T_i^0's are given by

$$T_1^0 = T_2^0 = ... = T_N^0 = \frac{N_{ev}}{N} \tag{4}$$

N_{ev} being the total number of events and N the maximum number of bins. Since the T_i's are still undetermined, N is chosen equal to the number of bins of the multiplicity distribution from event generator used to compute the P_{mn}'s. The first step consists in lowering the number of events in a bin, say i, by $\Delta T_i = 1$ and raising the number of events of another bin, say j, by $\Delta T_j = \Delta T_i = 1$. This is called a move and the new configuration is called C_2. In a second step one computes

$$\Delta F = F(C_2) - F(C_1) \tag{5}$$

where $F(C_2)$ and $F(C_1)$ correspond, respectively, to the cost function of configurations C_1 and C_2. If $\Delta F < 0$ the move is accepted, otherwise it is rejected. Steps 1 and 2 are repeated for the moves between bin i and the other bins. When i spans bins 1 through N, a sweep is completed. For each sweep one counts the number of accepted moves and the algorithm is stopped when no further move is accepted.

The above algorithm may be used to minimize any function. Also, in a move one may lower or raise the number of events in a bin i by an arbitrary ΔT_i; one may also use different ΔT_i for different sweeps. For example, one may use a large ΔT_i in the first sweeps to accelerate the algorithm and a small one when approaching the minimum.

This algorithm may be regarded as a genetic algorithm (see ref. [2] for an introduction). Within the framework of genetic algorithms the T_i's correspond to genetic populations, a move to a mutation between two different populations and F to the fitness function.

Application to the multiplicity distribution.

The algorithm described above is applied here to the correction of observed multiplicity distribution of negative particles produced in S-Pb interactions at 200 GeV/c per nucleon (see ref. [3] for details). The observed negative multiplicity distribution to be corrected in the following is shown in figure 1.

To minimize F, 20 sweeps were sufficient. Figures 2 and 3 show, respectively, the number of accepted moves and χ^2 as functions of the number of sweeps. As F approaches the minimum less and less moves

are accepted; at the minimum no more moves are accepted. At this point $\chi^2 \approx M$ as it is expected; M being the number of bins of the observed distribution.

Figure 4 shows the corrected multiplicity distribution; error bars are purely statistical. This distribution is less smooth than the one from the simulated annealing method [4]. One may explain this fact noting that the simulated annealing method gives a better estimate of the minimum of the cost function while genetic algorithms give a rough estimate of the minimum [5]. However, it takes 30 minutes on IBM 3090 to minimize F with the present algorithm while with the simulated annealing method it takes 500 minutes.

The mean multiplicity, dispersion $D_- = \sqrt{< n_-^2 > - < n_- >^2}$ and their ratio are

$$< n_- >= 57.40 \pm 0.40 \tag{6}$$

$$D_- = 44.27 \pm 0.97 \tag{7}$$

$$\frac{< n_- >}{D_-} = 1.297 \pm 0.030 \tag{8}$$

Conclusion.

An algorithm for the correction of observed multiplicity distributions is presented. The corrected distribution is compatible with the one obtained by the simulated annealing method. The algorithm is less demanding in terms of computing time.

Acknowledgements.

Thanks to Walter Geist for reading the manuscript.

References.

[1] M. Schmelling, Ph.D. thesis, Dortmund University, Germany (1987); CERN preprint CERN-CN/93-03 (1993).

[2] S. Forrest, Science 261 (1993) 872.

[3] M. Hafidouni, Doctorat de l'Université Louis Pasteur de Strasbourg, CRN/HE 92-35 (1992).

[4] M. Hafidouni, Strasbourg preprint CRN 93-32 (1993), *to appear in Nuclear Instruments and Methods in Physics Research A.*

[5] D.A. Diver, J. Phys. A 26 (1993) 3503.

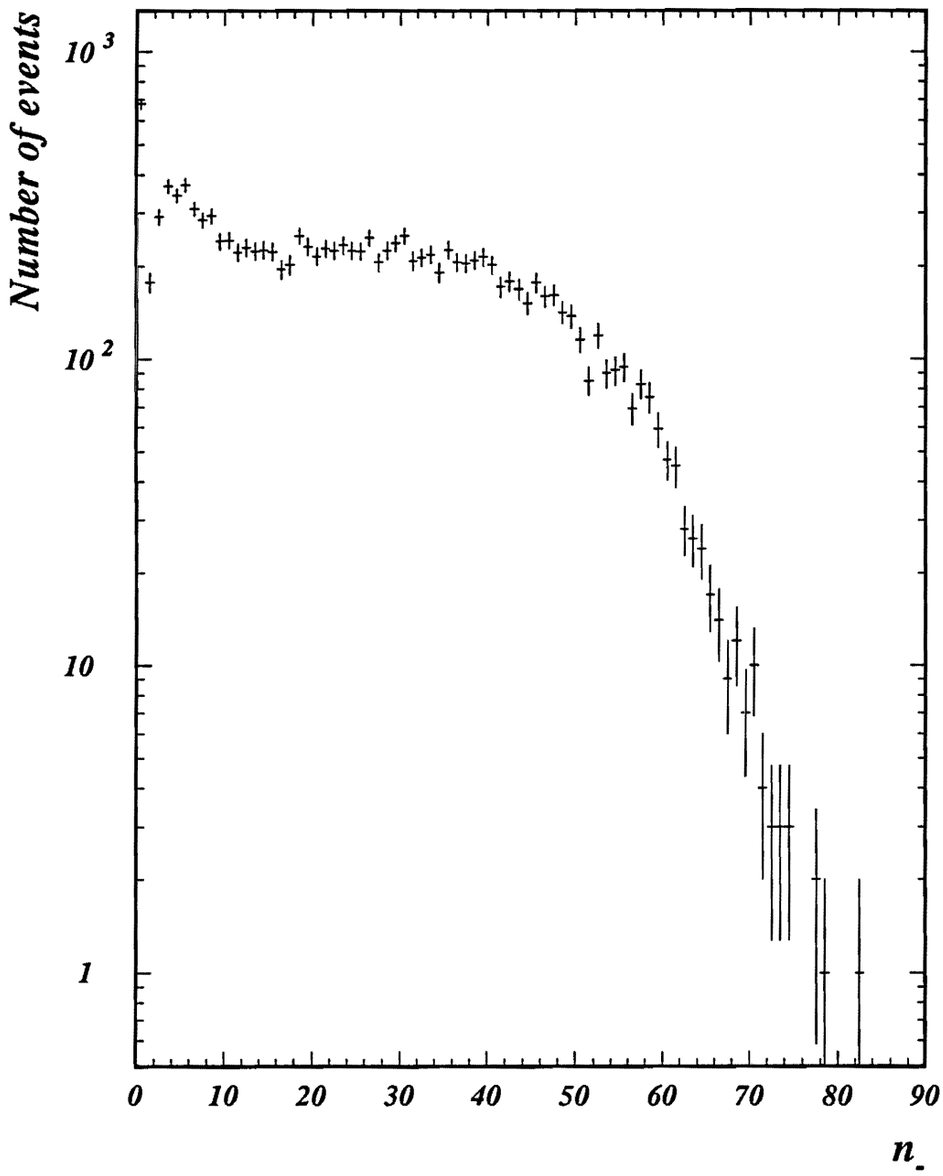

Fig. 1. Observed negative multiplicity distribution.

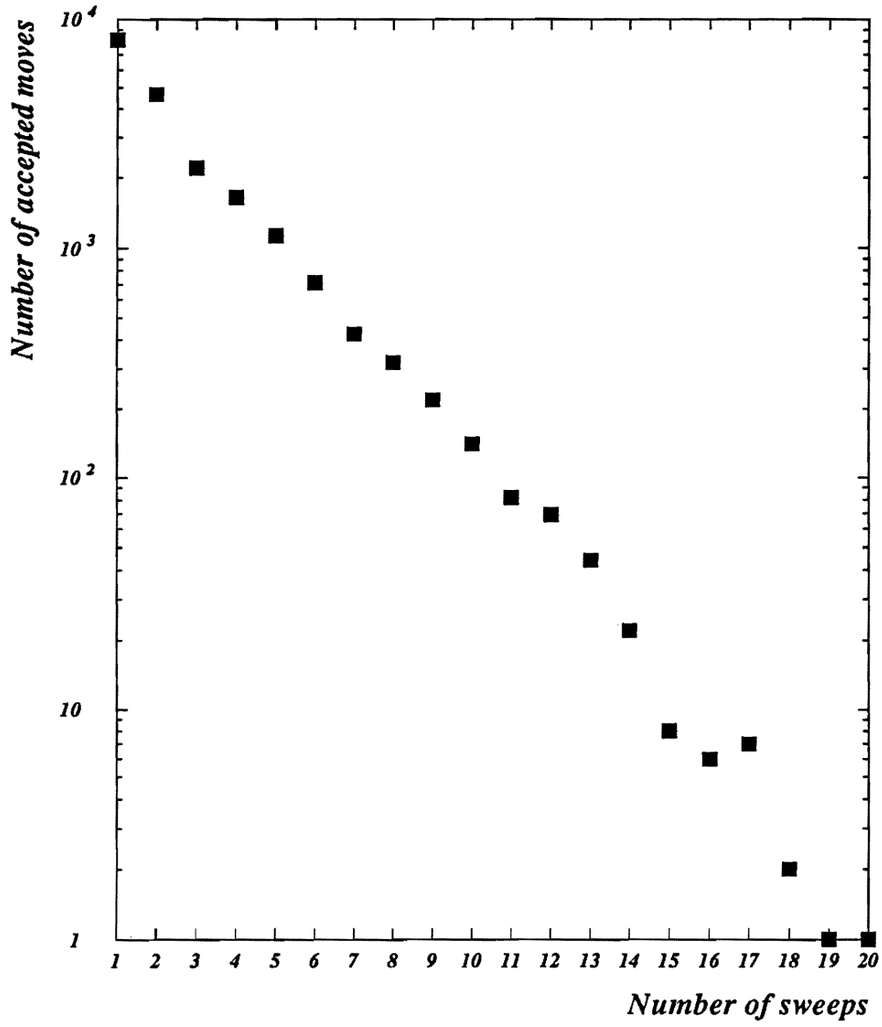

Fig. 2. Number of accepted moves *vs* number of sweeps.

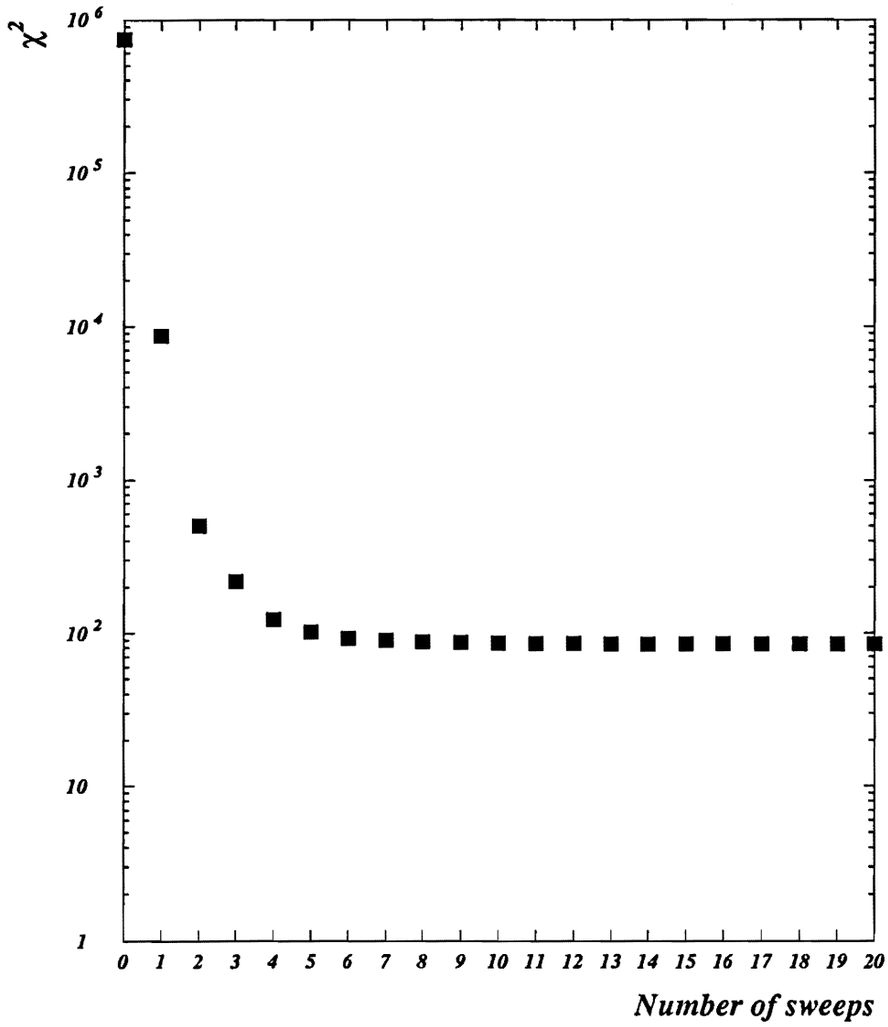

Fig. 3. χ^2 *vs* number of sweeps.

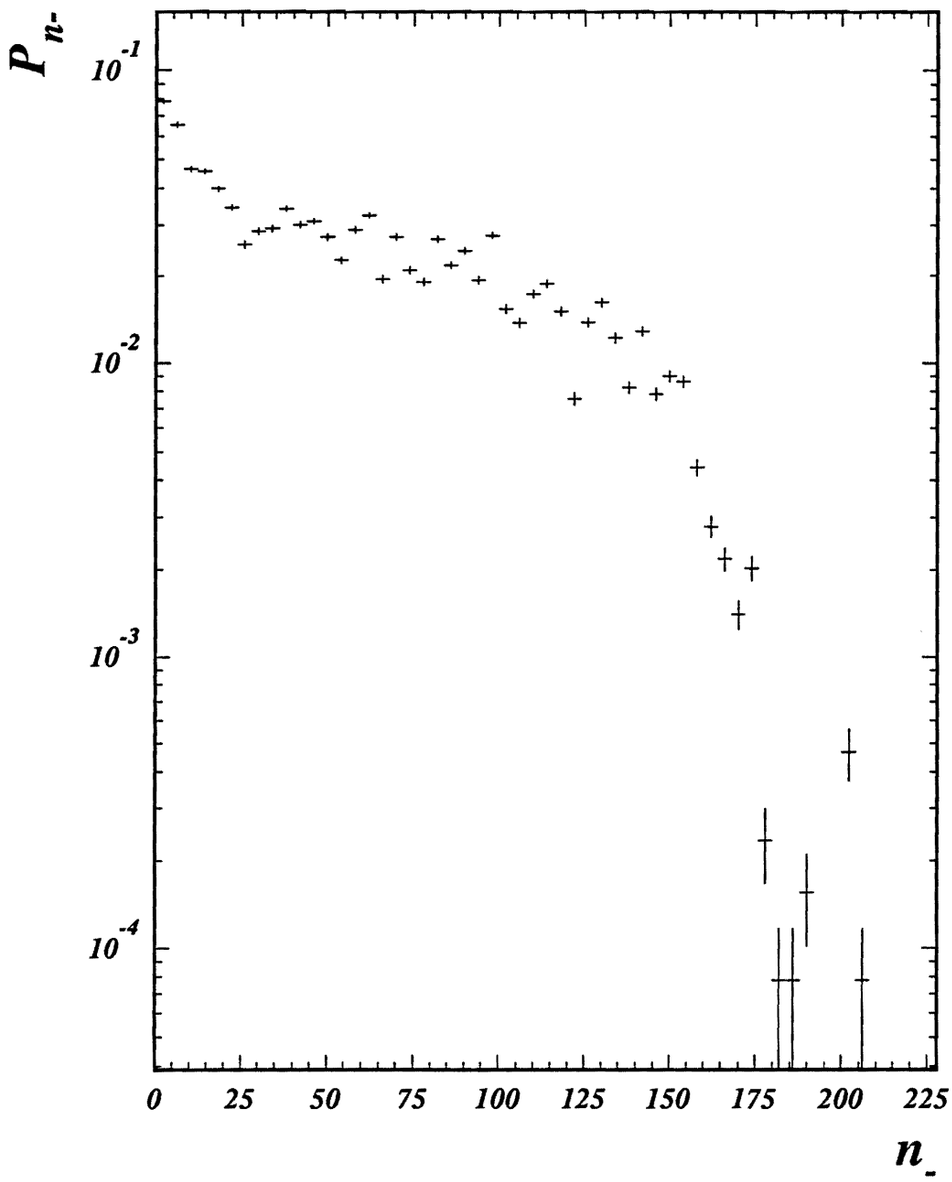

Fig. 4. Corrected negative multiplicity distribution.