ISTITUTO NAZIONALE DI FISICA NUCLEARE

Sezione di Bari

INFN/TC-84/26
29 Ottobre 1984

G. Iaselli :

A FAST EVENT GENERATOR FOR PARTICLE PHYSICS

Servizio Documentazione
dei Laboratori Nazionali di Frascati

Istituto Nazionale di Fisica Nucleare
Sezione di Bari

INFN/TC-84/26
29 Ottobre 1984

A FAST EVENT GENERATOR FOR PARTICLE PHYSICS

G. Iaselli
Dipartimento di Fisica dell'Università di Bari, and INFN - Sezione di Bari

ABSTRACT

A fast event generation is achieved with a statistical approach. Through the "principal component" analysis few factors are determined which enable to reproduce in simple way the original data sample.

Monte Carlo generation is with no doubt a major tool in particle physics.

The production mechanism of a Monte Carlo generated event is usually fixed imputing analytical parametrizzations determined either in a theoretical framework or from fits to a data sample. Therefore one has to relay on the goodness of the theory and on the accuracy of the fits.

Moreover, if the final state is determined through many successive production steps, a large computing time may be needed to generate a statistically significant sample of events.

We propose a statistical approach which enables to generate events in a fast and simple way. A data sample is treated with the principal component analysis to extract information on the average common behaviour.

Each event is considered as a statistical case (Zi) of Z1...Zn variables, being Zi a measured physical quantity computed from the final state.

This approach enables us to reveal the underlying pattern of relations between the variables, and to reduce the data sample to a smaller set of 'factors' or 'components' to be used as generator of an event configuration (Zi) where the interrelations observed in the data are fully accounted.

The basic hypothesis of the method is that each zi can be expressed as:

$$z_i = a_{i1} F_1 + a_{i2} F_2 + \ldots + a_{im} F_m \qquad m \ll n \qquad (1)$$

where Fi is a hypothetical weighting factor and aij is a standarized correlation coefficient between the variable zi and the factor Fj (factor loading matrix).

Once the F1...Fm distributions and the factor loading matrix are known, formula (1) can be used as an event generator.

The method has been applied to a Monte Carlo data sample of charged current neutrino-neon interactions[1].

All the significant final state variables of the events have been used, such as charged and neutral hadronic momentum components, muon momentum components, charged and neutral multiplicity, total energy, muon and hadron transverse momentum and others.

Each event has been therefore rappresented as a set (Zi) of N=20 variables. A total of 5000 events have been processed through the SPSS statistical package[2] to extract the factor loading matrix and the number of principal components (factors) with their distributions.

A number of 10 factors is found to explain all the variance in the data sample. These factors are built as linear combination of the z1...zn variables, and are arranged in order of importance: the first factor is the most important component, and accounts for a relevant part of the variance; the second is the second most important and so on.

TABLE I - Percentage of variance loaded on each factor.

Factor	% Variance
1	26.6
2	20.3
3	19.8
4	6.9
5	5.4
6	4.9
7	4.3
8	4.0
9	2.8
10	1.8

Table I shows the percentage of variance loaded on each factor. A rotation in the 10-dimensional space is then applied to the factors in order to produce an orthogonal set. During this rotation the maximum variance is loaded on the minimum number of factors. This is per formed using the SPSS 'Varimax' rotation. After the rotation each element a^2ij gives the fraction of zi variance accounted by Fj. The total amount of variance of zi accounted by the 10 factors is then (communality):

$$h^2i = a^2i1 + a^2i2 + \dots + a^2i10 \quad (2)$$

In Table II h^2i is given for each zi.

TABLE II - Communality for each variable zi.

Variable:	z1	z2	z3	z4	z5	z6	z7	z8	z9	z10
Communality:	0.95	0.94	0.95	0.97	0.95	0.95	0.99	0.80	0.99	0.96
Variable:	z11	z12	z13	z14	z15	z16	z17	z18	z19	z20
Communality:	0.96	0.99	0.97	0.97	0.99	0.98	0.98	0.98	0.95	0.99

An estimate of each Fi for all the individual data cases can be calculated through the rotated factor loading matrix (aij) and the initial array z1...zn(2).

Figures 1 show the standarized distributions of the first 5 factors.

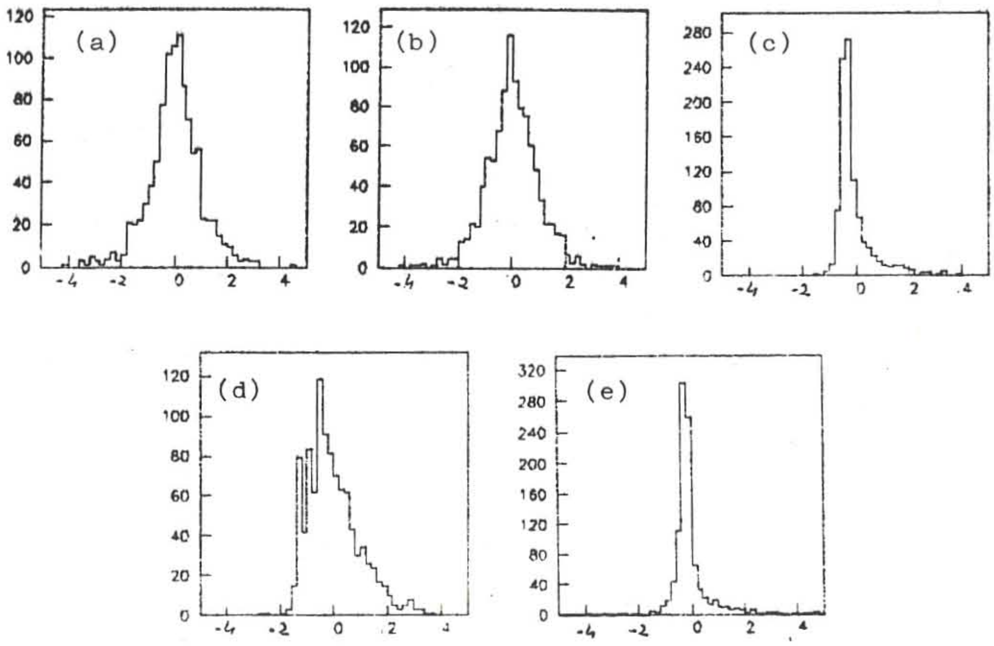

FIG. 1 - The factor distributions (a-e) obtained at the end of the SPSS procedure. They are all standardized distributions.

Since SPSS requires a standardized form for all the variables, at the end of the procedure no information on the mean and the standard values of zi's is stored in the factors and in the factor loading matrix. Therefore the user has to supply this information independently.

In order to generate events, a standardized configuration is initially obtained as:

$$zi' = \sum_{j=1}^{10} aij\ Fj \qquad (3)$$

where aij is the element of the know factor loading matrix and Fj are extracted randomly from the distributions in Figs. 1.

The final configuration is then evaluated as:

$$zi = zi'^{*}\sigma(i) + \mu(i) \qquad (4)$$

being $\sigma(i)$ and $\mu(i)$ the values of the standard deviation and the mean of variable Zi in the original data sample.

Results are shown in Figs. 2, respectively for muon longitudinal momentum, total hadronic longitudinal momentum, charged hadronic longitudinal momentum and neutral hadronic longitudinal momentum.

The generation is successful in reproducing the original distributions with good accuracy. The same accuracy is achieved for all the other variables.

Moreover, being the generation made via simple formula (1), the computing time required to obtain high statistics is negligible.

This approach is usefull when the generation of a sample with features similar to the data is required.

However it could also be used to help large conventional generation by factor analyzing a small sample of Monte Carlo events and employing it for high statistic production.

REFERENCES

(1) - P.Kasper, WA59 Collaboration, Internal comunication.

(2) - SPSS (McGraw-Hill Company, 1975), and references therein.

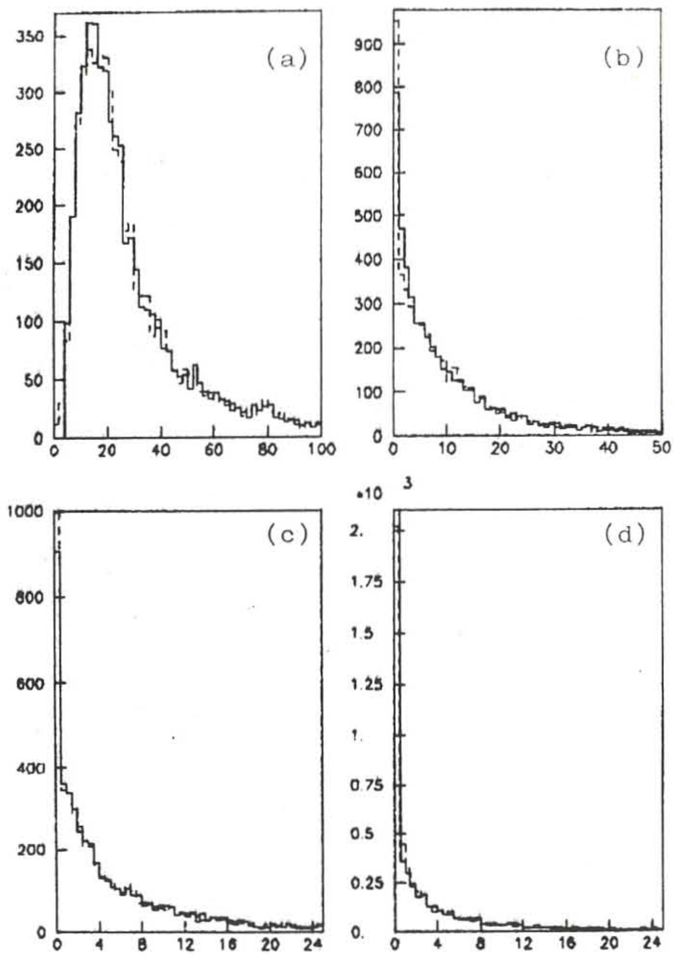

FIG. 2 - Distributions as obtained from the data sample (full line) and as generated with this method (dotted lines) for the muon longitudinal momentum (a), the total hadronic longitudinal momentum (b), the charged hadronic longitudinal momentum (c), and the neutral hadronic longitudinal momentum (d).