MULTIDIMENSIONAL ANALYSIS OF DATA OBTAINED IN EXPERIMENTS WITH X-RAY EMULSION CHAMBERS AND EXTENSIVE AIR SHOWERS

A.A.CHILINGARYAN, S.KH.GALFAYAN, M.Z.ZAZYAN
Yerevan Physics Institute, Markarian st.2
375036, Yerevan, Armenia, USSR

A.M.DUNAEVSKII
Lebedev Physical Institute
Leninsky prospect 53, Moscow, USSR

1. Introduction

The traditional approach to the analysis of data available from experiments with X-ray emulsion chambers consists in considering one-dimensional distributions or the dependences of one experimentally observed value on the other.

Recently the analysis of two- and three-dimensional distributions /1/ as well as the presentation of averages of two variables together with their errors led to the possibility of drawing a conclusion on the scaling violation in the secondary particle fragmentation region at the energies $\sim 10^{16}$ eV and estimating the degree of its violation /2/.

Thus, the increase of simultaneously analyzed features seems to be attractive, but it is apparent that the analysis of three and more features is connected with the necessity to have the quantitative measure of distinction of multidimensional distributions presented by limited samples.

The other main problem of cosmic ray physics may be formulated as the problem of determining the portion of experimental events belonging to one of described types. That is the problem of determining the portion of photon-hadron ($\gamma - h$) families generated by various primary nuclei or the determination of the primary component chemical composition by the EAS data.

In the present paper we shall show that the solution of the above problems may be realized in the multidimensional space by the nonparametric statistic methods developed in /3,5/. Note that the use of these methods for processing the experimental data from the cosmic ray physics installations has demonstrated their advantage over the traditionally applied techiques /4,6,7/.

This is a methodical work, i.e. the experimental events are replaced by the model ones. Thus we have an opportunity to determine the limits of applicability of the methods suggested and to estimate the expected accuracies of determining the desired physical values.

Among the models used, there is a pure-scaling one - M6, the models with increasing cross section and scaling violation in the pionization region - M4, Femin, Femax (the detailed description of the models may be found in /8/). Some models are obviously nonrealistic (e.g. M6), but for the methodical purposes the use of the family banks corresponding to these models is admissible.

2. The Distinction of Strong Interaction Models

The selection of the feature set optimal for the discrimination purposes is performed with account of the averages differences and the correlation information. The feature pairs with the statistically significant difference in correlations are included in the set. Finally each set is characterized by the so-called Bayesian risk R - the probability to misclassify the models (or the model and experimental data) in classification procedure performed with the optimal Bayesian decision rule (for details of the method and used features see /9/).

The R^B use in one-dimensional analysis leads to the same conclusions as the standard statistical methods of two samples averages difference significance calculation (T-test and Wilcoxon test). The more is the difference between R and 0.5 (corresponding to the classes total overlapping) the stronger is the difference between the distributions.

The Bayesian risk calculation method /10/ allows to obtain unbiased effective estimates and to judge of the model describing the experiment in the best way by the successive comparison of the experimental and alternative model training samples.

The estimation accuracy depends on the sizes of the used training samples. Besides there is an interrelation between the sample size and maximum dimensionality of the space where one may realize the effective local reconstruction of the probability multidimensional density. As we see from Table 1, the samples limitation (100÷400) leads to that the addition of low-informative features may even deteriorate the discrimination due to the scarcity of points in N-dimensional feature space.

Table 1. The comparison of the M4 and M6 models by means of γ-family various characteristics.

Combination	Space dimensionality	R^B
$\bar{R}_\gamma$	1	0.38±0.3
$\bar{R}_\gamma, \overline{ER}_\gamma$	2	0.37±0.3
$\bar{R}_\gamma\ \Sigma' E_\gamma, \alpha'_\gamma$	3	0.35±0.3
$\bar{R}_\gamma, \Sigma' E_\gamma, \alpha'_\gamma\ \beta'_\gamma$	4	0.34±0.3
$\Sigma E_\gamma, n_\gamma, \bar{R}_\gamma, \overline{ER}_\gamma, \alpha_\gamma, \Sigma' E_\gamma, \ldots$	10	0.38±0.3
$\Sigma' E_\gamma, n'_\gamma, \bar{R}'_\gamma, ER'_\gamma, \alpha'_\gamma, \beta_\gamma, \ldots$	15	0.40±0.3

The comparison of the M4 and M6 models by means of various feature combinations has shown that the addition of information from the hadronic block or the shower installation to the γ-family information does not reduce the classification errors. That is, in the problems related to the study of strong interaction cross sections and scaling violation in the pionization region it is enough to analyze only the photon family characteristics.

3. Separation of Families from the Light and Heavy Nuclei. Determination of the Portion of Families from Fe Nuclei.

Let us take as prototypes (training samples) the samples containing the events from the light nuclei and, respectively, the events from the banks Femin and Femax. Various combinations of both types from the events not included in the training samples will be taken as the "experimental" one. Such a choice imitates the case of precise knowledge of the strong interaction model. The portion of "iron" events in these samples is set to be $P_H = 0.05$, 0.07, 0.1.

The estimate of the portion of families initiated by the Fe nuclei will be obtained after the experimental data classification and calculation of the probabilities to misclassify the events of both types.

$$\hat{R}_H = \frac{R_H^* - R_{L\to H}}{1 - R_{H\to L} - R_{L-H}}$$

where P_H^* is the portion of families referred to the "iron" type, $P_{L\to H}$ and $R_{H\to L}$ are the probabilities of the classification possible errors.

Table 2 shows that the reconstructed P_H value is rather close to the true one. Besides, it may be shown that the classification allows one to enrich 5÷7 times the selected events with the families from the heavy nuclei, this possibly enabling one to study Fe - N^{14} interaction at the energies more than 10^{16} eV.

Table 2. Reconstruction of the portion of families from the Fe nuclei. The training and control samples are taken from the banks M6 and Femin.

Installation	Features	P_H	P_H^*	$\hat{P}_H$
γ-block	$\Sigma E_\gamma, \bar{R}_\gamma, \alpha_\gamma$	0.05	0.194	0.042±0.060
		0.07	0.202	0.055±0.058
		0.1	0.217	0.080±0.058
γ-block + shower part	$\Sigma E_\gamma, \bar{R}_\gamma, \alpha_\gamma, E_0$	0.05	0.09	0.048±0.025
		0.07	0.115	0.074±0.024
		0.1	0.138	0.098±0.023
γ-h -block	$\bar{R}_\gamma, n_h, \bar{R}_h$	0.05	0.125	0.044±0.028
		0.07	0.149	0.076±0.026
		0.1	0.171	0.099±0.025

However, if the strong interaction model is unknown (this case was simulated by the use of "experimental" data from the banks not coinciding with the prototypes), the reconstruction is carried out with great errors. Therefore, to treat the real experimental data one should use either more realistic models or combinations of features weakly dependent on the strong interaction model, but simultaneously highly sensitive to the primary nucleus type.

4. Conclusion

The use of nonparametric statistic methods allows one to carry out the quantitative comparison of the model and experimental data. The same methods enable one to select the events initiated by the heavy nuclei and to determine the portion of the corresponding events. For this purpose it is necessary to have the banks of artificial events describing the experiment sufficiently well. At present, the model with the small scaling violation in the fragmentation region /11/ is the closest to the experiments. Therefore, the treatment of γ -families obtained in "Pamir" experiment is being carried out at present with the application of these models.

REFERENCES

/1/ Danilova T.V., Erlykin A.D., VANT, ser.tekhn.fiz.eksp., issue 1(13), p.13-19, 1983.
/2/ Pamir Collaboration, 18th ICRC, v.5, p.425-428, 1983.
/3/ Chilingaryan A.A., VANT, ser.tekhn.fiz.eksp., issue 2(8), p.59-68, 1981.
/4/ Chilingaryan A.A., VANT, ser.tekhn.fiz.eksp.,issue 3(12), p.90-91, 1982.
/5/ Chilingaryan A.A., 18th ICRC, v.5, p.524-526, 1983.
/6/ Chilingaryan A.A., VANT, ser.tekhn.fiz.eksp., issue 1(13), p.81-84, 1983.
/7/ Chilingaryan A.A., VANT, ser.tekhn.fiz.eksp., issue 1(13), p.61-68, 1983.
/8/ Dunaevskii A.M., Uryson A.V., Preprint FIAN, No.150.
/9/ Chilingaryan A.A., Galfayan S.Kh., Zazyan M.Z., Dunaevskii A.M., 18th ICRC, v.5, p.487-490, 1983.
/10/ Galfayan S.Kh., Chilingaryan A.A., in: Statistical Problems of Control, issue 66, Vilnius, 1985.
/11/ Vernov S.N., Khristiansen G.B. et al., 15th ICRC, v.8, p.32-35, 1976.