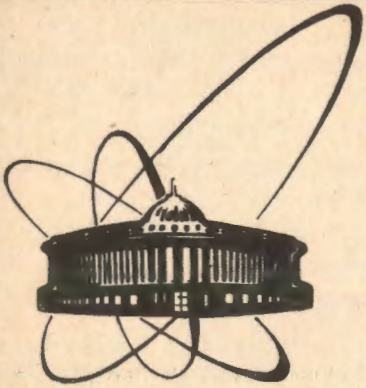

сообщения
объединенного
института
ядерных
исследований
дубна

E10-90-579

V.I.Ilyushchenko, N.I.Chernov

AN ANALYSIS OF SPECTRAL DATA WITH RARE EVENTS STATISTICS

1990

3. THE CASE OF SEVERAL NONSUMMABLE EXPERIMENTAL RUNS

Here we proposed an algorithm for testing a hypothesis about the presence of a persistent peak in different nonsummable experimental spectra.

Let us denote by $n_1,...,n_K$ the numbers of hits in an isolated histogram cell; and by $l_1,...,l_K$, the theoretically predicted numbers of hits, i.e., $l_i = n_i * p_i$, where p_i is the probability to occupy the i-th histogram cell according to an available theoretical distribution curve. In our case $n_i > l_i$ for the overwhelming majority of the numbers i in expected peak locations.

The values of n_i are known to be distributed according to the Poisson law with a mean value l_i. The Poisson distribution can be effectively approximated by a normal one, i.e. that with $n \sim n(l_i, l_i)$, where $n(a, p^2)$ stands for a normal distribution specified by a mean value a and a variance p^2.

Due to the independence of experimental runs the joint probability distribution density will be described as

$$f_0(n_1,...,n_K) = \frac{1}{(2\pi)^{K/2} \prod_{i=1}^{K} \sqrt{l_i}} \exp\left[-\sum_{i=1}^{n} \frac{(n_i - l_i)^2}{2l_i}\right]. \quad (1)$$

An alternative hypothesis states that the "projected" mean values n_i are somewhat higher and equal to l_i^*, where $l_i^* > l_i$; thereby exact values l_i^* are unknown. Then $n_i \sim n(l_i^*, l_i^*)$ and the joint probability density will be

$$f_1(n_1,...,n_K) = \frac{1}{(2\pi)^{K/2} \prod_{i=1}^{K} \sqrt{l_i}} \exp\left[-\sum_{i=1}^{n} \frac{(n_i - l_i^*)^2}{2l_i^*}\right]. \quad (2)$$

According to the optimum Neyman-Pearson test [5] the decision rule is defined by the following inequality:

$$\frac{f_1(n_1,...,n_K)}{f_0(n_1,...,n_K)} > C. \quad (3)$$

On satisfying this inequality we accept a hypothesis about the presence of persistent peaks in experimental spectra.

The value of the C is used to be chosen from the condition of the probability of the type 1 error not exceeding 1-5%.

In our case the inequality (3) results in

$$\left[\sum_{i=1}^{K} \frac{(n_i - l_i)^2}{2l_i} - \sum_{i=1}^{K} \frac{(n_i - l_i^*)^2}{2l_i^*}\right] > C' \quad (4)$$

that can be reduced to

$$\sum_{i=1}^{K} n_i^2\left(\frac{1}{2l} - \frac{1}{2l^*}\right) > C'', \tag{5}$$

where C' and C'' are constants depending on C in (3).

Here it is possible to take $l_i^* = (1 + \alpha) * l_i$, i.e., to assume hypothetical systematical effects to exceed the background by $(1 + \alpha)$ times in each experimental run. Then the decision rule will be defined as

$$\sum_{i=1}^{K} \frac{n_i^2}{l_i} > C. \tag{6}$$

4. THE CHOICE OF TEST CRITERION

Now let us choose the test criterion C.

4.1 From the standpoint of theoretical statistics it is possible to take a confidence level equal to, say, 99%. This implies the probability of type 1 error to be 1%. Then, the value of C can be computed from the condition that if our hypothesis about the presence of persistent peaks is false, it should be accepted with a probability of 1%. Thus, the results of observations, $n_1,...,n_K$ have a distribution density $f_0(n_1,...,n_K)$ and the probability of accepting our hypothesis is calculated as

$$\frac{1}{(2\pi)^{K/2} \prod_{i=1}^{K} \sqrt{l_i}} \int\cdots\int_{\sum_{i=1}^{K} n_i^2/l_i > C} \exp\left[-\sum_{i=1}^{K} \frac{(n_i - l_i)^2}{2l_i}\right] dn_1 \ldots dn_K = 0.01. \tag{7}$$

This integral provides a value of C and, thus, a solution of the problem under study.

4.2 The above solution is rigorous and complete from the standpoint of theoretical statistics, but it is inconvenient to be used in practice due to the very complicated form of the integral (7). Really, the applied statistics prefers to begin with computing the value of

$$C = \sum_{i=1}^{K} \frac{n_i^2}{l_i} \tag{8}$$

with the subsequent integration in relation (7). If the value of (7) is close to zero, e.g., being at a level of 0.01, then the hypothesis about the presence of persistent peaks will be accepted.

4.3 The integration in (7) is an involved operation requiring some sophisticated solving means. Therefore it is desirable to have a simpler testing procedure.

Let us take

$$n_i = \sqrt{l_i} * x_i + l_i, \tag{9}$$

where $x_i \sim n(0, 1)$, if our hypothesis is false. Then from (8)

$$C = \sum_{i=1}^{K} \frac{(\sqrt{l_i} * x_i + l_i)^2}{l_i} = \sum_{i=1}^{K} (x_i^2 + 2\sqrt{l_i} * x_i + l_i) \tag{10}$$

so that for mathematical expectation $E(C)$ and variance $D(C)$ one obtains

$$E(C) = K + \sum_{i=1}^{K} l_i, \tag{11}$$

$$D(C) = \sum_{i=1}^{K} D(x_i^2 + 2\sqrt{l_i} * x_i) = 2K + 4 \sum_{i=1}^{K} l_i. \tag{12}$$

The hypothesis about the presence of persistent peaks can be accepted, e.g., for

$$C \geq E(C) + 3\sqrt{D(C)} \tag{13}$$

or

$$C \geq E(C) + 4\sqrt{D(C)}, \tag{14}$$

i.e., on the basis of $3\sigma-$ or 4σ-rules even though the C-distribution is not normal (in reality, it is closer to χ^2-distribution).

For example, there are realized K experimental runs with $l_1 = \ldots = l_K = l \sim 20$. Then $E(C) = K + Kl \sim Kl$ and $D(C) = 2K + 4Kl \sim 4Kl$, i.e., $\sqrt{D(C)} \sim 2\sqrt{Kl}$.

Let us recall that $E(n) = D(n) = l$ and let us assume that in each run $n = l + y\sqrt{l}$, i.e., for $y = 2$ the result differs from mean value by about 2σ, which is clearly insufficient to accept the hypothesis. In this case $C \sim Kl + 2Ky\sqrt{l} = E(C) + (y\sqrt{K}) * \sqrt{D(C)}$ and statistic C will deviate from mean value by $(y * \sqrt{K})\sigma$. For example, at $y = 2$ and $K = 4$ the final deviation will be 4σ and the hypothesis about the presence of persistent peaks must be accepted.

Thus, this method acts as if to increase by $\sqrt{K}$ times the deviation observed in each experimental run.

5. CONCLUSION

The statistical analysis performed for the case of spectral rare events experiments with nonsummable total statistical demonstrates the confidence level of the persistent peak identification hypothesis to be proportional to $\sqrt{K}$, where K is the number of individual experimental runs.

This testing criterion can be effective in application to the same spectral pattern produced from an elementary particle interacion with different reaction input channels or different experimental setup configurations.

The cooperation of Drs. Yu.A.Troyan and V.N.Pechenov is gratefully acknowledged.

REFERENCES

1. Troyan Yu.A. et al. - Preprint JINR D1-88-329, Dubna, 1988.

2. Tatischeff B. et al. - Phys.Rev., 1987, C36, p.1995.

3. Kozma P., Ilyushchenko V.I. - An Analysis of Dibaryon Resonance Spectra by GSAP, AUTSPAN and SAMPO-80 Codes, (accepted for publication in Czech. J. Phys., 1990).

4. Hnatowicz V., Ilyushchenko V.I., Kozma P. - CPC, 1990, 60, p.111.

5. Cramer H. - Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

Received by Publishing Department
on December 24, 1990.