KERNFORSCHUNGSZENTRUM

KARLSRUHE

Januar 1969 KFK 965

Institut für Experimentelle Kernphysik

Detection of Inhomogeneity in Scanning Efficiencies

J.S. Ausländer

E.M. Friedländer (CERN)

GESELLSCHAFT FÜR KERNFORSCHUNG M.B.H.

KARLSRUHE

NUCLEAR INSTRUMENTS AND METHODS 67 (1969) 71-73; © NORTH-HOLLAND PUBLISHING CO.

DETECTION OF INHOMOGENEITY IN SCANNING EFFICIENCIES

J. S. AUSLÄNDER

Laboratoire de Physique Corpusculaire de l'Université de Strassbourg and Institut für Experimentelle Kernphysik, Universität und Kernforschungszentrum, Karlsruhe

and

E. M. FRIEDLÄNDER*

CERN, Geneva, Switzerland

Received 18 July 1968

A simple, statistical criterion is presented, establishing whether in visual scanning all events are detectable with the same probability, or not. In this way misleading conclusions drawn from the usual method of two-fold scanning can be avoided.

1. In scanning of events recorded by visual detectors such as track chambers, nuclear emulsions, etc., a certain fraction of events is inevitably lost. It is usually assumed that this loss may be corrected for if the same events are inspected two (or more) times; the true number of events is taken as[1]):

$$Q = R_1 R_2 / D, \tag{1}$$

where R_i are the numbers of events detected in scan no. i, and D is the number of events detected in scan no. 1 as well as in scan no. 2.

It has been repeatedly pointed out [2-4]), that eq. (1) is valid only if within a given scanning run the probability to detect an event is the same for all the events; whenever this condition is satisfied we shall speak of homogeneous detection efficiencies.

As long as the events can be labelled with respect to the parameter(s) causing the inhomogeneity (e.g. stars with different numbers of prongs), an inhomogeneous problem can be reduced to the homogeneous case in an obvious way.

Quite often, however, it may happen that the physical quantity(ies) controlling inhomogeneity does not appear in such a manifest manner.

In such a case some guess must be made, but then the reliability of the double-scan procedure is open to considerable doubt.

To our knowledge, at present there exists no general method for detecting inhomogeneities which would be free from the risks of a such a more or less arbitrary guess)†.

Actually it seems surprising that a very simple solution of this problem has passed unnoticed, hitherto. In fact, if the *same* events are inspected several times, the number of events, R, observed per scanning run, is distributed according to different laws in the homogeneous and in the inhomogeneous case. These differences can be used to detect an existing inhomogeneity without any need for a guess about its causes.

2. Consider a number m of independent scanning runs, performed on a fixed number Q of events. For the sake of simplicity we shall assume that the scanning efficiencies do not vary from run to run.

If all the Q events are of the same kind, i.e. they have the same detection probability p, then the results R_i ($i = 1, 2, \ldots, m$) of the different scanning runs are individual values of a random variable R, obeying the Bernoulli law

$$B(R|p,Q) = [Q!/\{R!(Q-R)!\}]\, p^R (1-p)^{Q-R}. \tag{2}$$

If, however, the Q events are not all of the same kind, i.e. if the sample is inhomogeneous, this will no longer be true.

We shall now discuss the special case of two kinds a and b of events+, present in numbers Q_a and Q_b, and detected with probabilities p_a and p_b respectively. Then the total number R of events detected in a given run is the sum of the two unknown quantities R_a and R_b. Since we have assumed that the subsamples a and b are homogeneous, both R_a and R_b are binomially distributed, too. Their sum R, however is distributed binomially if, and only if

$$p_a = p_b, \tag{3}$$

(for proof, see appendix).

Hence the problem of checking a given sample for homogeneity reduces to that of checking whether R

* On leave from Institute of Atomic Physics, Bucharest.

† As will be shown in a forthcoming paper, the criterion given[5,6]) by Tolstov may often be inconclusive.

\+ The generalization to the case of more than two different kinds of events is obvious.

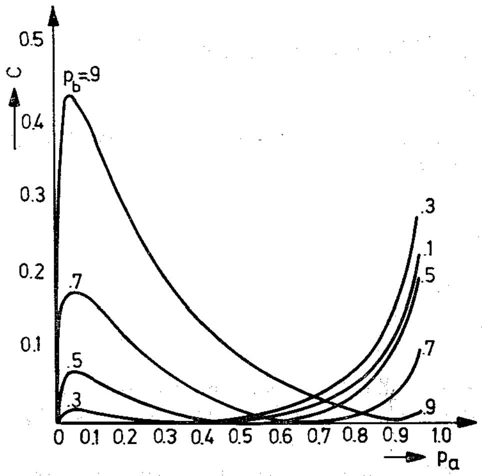

Fig. 1. Plots of the test quantity C vs efficiency p_a, for different values of p_b; $Q_a = 5$, $Q_b = 5$.

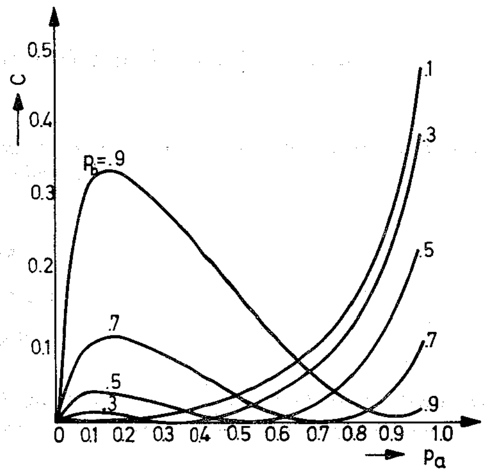

Fig. 2. Same as fig. 1, for $Q_a = 5$, $Q_b = 1$.

is binomially distributed or not*. A convenient criterion for such a check is described in the following section.

3. Denote by ρ, σ^2 and α respectively, the expectation value, variance and skewness of the distribution of R.

In the homogeneous, i.e. binomial, case:

$$\rho = Qp, \tag{4}$$

$$\sigma^2 = Qp(1-p), \tag{5}$$

$$\alpha = (1-2p)\cdot\{Qp(1-p)\}^{-\frac{1}{2}}. \tag{6}$$

Eliminating Q and p from these equations we obtain a relationship between ρ, σ and α, which in terms of the quantity

$$C \equiv \tfrac{1}{2}(1+\alpha\sigma)-\sigma^2/\rho, \tag{7}$$

can be written:

$$C = 0. \tag{8}$$

In the inhomogeneous case

$$C \neq 0 \tag{9}$$

is to be expected. Then, ρ, σ and α, have to be expressed in terms of Q_a, Q_b, p_a and p_b, as follows:

$$\rho = Q_a p_a + Q_b p_b, \tag{10}$$

$$\sigma^2 = Q_a p_a(1-p_a)+Q_b p_b(1-p_b), \tag{11}$$

$$\alpha\sigma = \{Q_a p_a(1-p_a)(1-2p_a)+Q_b p_b(1-p_b)(1-2p_b)\}\cdot \cdot\{Q_a p_a(1-p_a)+Q_b p_b(1-p_b)\}^{-1}. \tag{12}$$

Figs. 1 and 2 show plots of C computed by means of eqs. (10), (11), (12), against p_a, for fixed p_b and two different combinations of Q_a and Q_b.

It is important to stress here that C is independent of Q in the homogeneous as well as in the inhomogeneous case, when it depends only on the ratio Q_a/Q_b.

The problem of the statistical significance of non-nil estimates $\hat{C}$ for C will be dealt with in the following section.

4. For practical purposes it is essential to know the spread of the distribution of $\hat{C}$. The problem is complicated by the fact that the estimates $\hat{\rho}$, $\hat{\sigma}$ and $\hat{\alpha}$ are not statistically independent, the correlation being especially strong for low values of m, the number of scanning runs. On the other hand, in practice it is desirable to keep m as low as possible.

TABLE 1

Estimates C^*, eq. (13) and rms deviations of $\hat{C}$ values, obtained by Monte Carlo simulation.

Input data: $Q = Q_a + Q_b = 6$; $m = 4$; $K = 2500$.
I: $p_a = 0.6$; $p_b = 0.3$.
II: $p_a = 0.8$; $p_b = 0.2$.

Q_a	C^*		$\sigma(\hat{C})$	
	I	II	I	II
1	0.014±0.007	0.126±0.007	0.351±0.025	0.332±0.023
2	0.015±0.009	0.155±0.008	0.441±0.031	0.394±0.028
3	0.011±0.009	0.182±0.009	0.449±0.032	0.423±0.030
4	0.003±0.010	0.243±0.009	0.525±0.037	0.451±0.032
5	0.005±0.011	0.262±0.008	0.527±0.038	0.406±0.026

* It is easily understood that the same is true for D in the case of m independent pairs of scans (i.e. m double-scans), etc.

Table 1 shows the results of a Monte Carlo simulation of multiple scans ($m = 4$) for a fixed $Q = 6$, and different values of Q_a, p_a and p_b. From table 1 and figs. 1 and 2 it can be seen, that for the same changes of the inhomogeneity parameters Q_a, Q_b, p_a, and p_b:

a. The quantity C (and obviously also its estimates varies considerably, while

b. The rms deviation of $\hat{C}$ varies only slightly.

Furthermore, if C is in excess of, say 0.1, the inhomogeneity can be revealed, given reasonable statistics, even just by a fourfold scan. Such values of C appear, as can be seen from figs. 1 and 2, as soon as one of the efficiencies, say p_a, is large enough ($\gtrsim 0.8$).

Practically an experiment can be performed as follows.

A large number K of cells, frames, etc., each containing a relatively small population of events, is scanned m times. Since three quantities (viz. ρ, σ and α) must be estimated, one must choose $m \geqslant 3$. Each cell will then yield an estimate $\hat{C}_k^{(m)}$ $(k = 1, 2, \ldots, K)$ for C which will have a rather broad distribution for low m. These estimates must now be averaged:

$$C^* \equiv (1/K) \sum_{k=1}^{K} \hat{C}_k^{(m)}; \tag{13}$$

obviously C^* will be $K^{-\frac{1}{2}}$ times narrower distributed than $\hat{C}^{(m)}$. Roughly one may say that the standard error of C^*, which for a given non-nil value of C is decisive from the practical point of view, is $\approx 0.7\,(mK)^{\frac{1}{2}}$.

For illustration we quote the following numerical example: $Q = 6$; $Q_a = 1$; $p_a = 0.8$; $p_b = 0.2$; $m = 4$. Then,

$$C = 0.126, \qquad \sigma(\hat{C}) = 0.332;$$

if three standard deviation confidence is required, this yields $K = 62$. This implies that effectively ≈ 375 events have to be inspected 4 times, which seems quite reasonalbe, as far as scanning effort is concerned.

Concluding one may say that detection of an existing inhomogeneity may often be possible without any guess as to its causes and without a prohibitive volume of scanning work.

One of us (E.M.F.) expresses his gratitude to Mrs. M. Busi for her advice in the use of the CERN CDC computers.

Appendix

Consider two independent binomial random variables R_a and R_b, with parameters Q_a, Q_b and p_a, p_b, and their sum

$$R = R_a + R_b.$$

The moment generating function (mgf) $G_R(\lambda)$ is then given by the product of the partial mgf's:

$$G_R(\lambda) = G_{Ra}(\lambda) \cdot G_{Rb}(\lambda)$$
$$= (1 + p_a\, e^{\lambda} - p_a)^{Q_a} \cdot (1 + p_b\, e^{\lambda} - p_b)^{Q_b}. \tag{A.1}$$

Only if

$$p_a = p_b,$$

the product in the rhs of eq. (A.1) can be brought to the form of a binomial mgf.

Note added in proof : Actually the Monte Carlo simulations of table 1 were performed with $m = 100$ and the results recalculated for $m = 4$. Since then we have performed more extensive Monte Carlo calculations which show that for low values of m the estimate $\hat{C}$ is biassed (displaced). This gives rise to some complications in the practical use of the method, which will be discussed in a more detailed account, to be published later.

References

1) H. Geiger and A. Werner, Z. Physik **21** (1924) 187.
2) J. Barnothy and M. Forro, Rev. Sci. Instr. **22** (1951) 415.
3) C. I. Waddington, Suppl. Nuovo Cimento **19** (1961) 37.
4) D. A. Evans and W. H Barkas, Nucl. Instr. and Meth. **56** (1967) 289.
5) K. D. Tolstov, JINR Report P-864 (Dubna, 1962).
6) E. Fett, Kinematical identification of bubble chamber events, École Int. de la Physique des particules élémentaires (Herceg-Novi, 1965).