THE BINNING ANALYSIS: TOWARDS A BETTER SIGNIFICANCE TEST

E. GOSSET* and B. LOUIS

Institut d'Astrophysique, Université de Liège, Cointe-Ougrée, Belgium

Abstract. We demonstrate that the problems encountered with the binning analysis essentially lie in the significance test which is used. We review and analyse most of the existing significance tests and show that the use of those based on randomization processes considerably improves the situation. For the one-dimensional binning analysis, we apply the '2 within 4' randomization test, whereas for the two- and the three-dimensional binning analyses we introduce two innovations: the '4 within 16' randomization test and the '8 within 64' randomization test, respectively.

1. Introduction

1.1. ASTRONOMICAL BACKGROUND

For a long time, in their quest towards understanding 'their Universe', many astronomers have been interested in the analysis of the spatial distribution of various categories of celestial objects. The repartitions of galaxies, extragalactic radio-sources and quasars are important since they act as many tracers of the intrinsic properties of the Universe (homogeneity, isotropy, ...). We do not have to recall that one of the bases of theoretical cosmology has been called the cosmological principle: this principle is still at the stage of being a working hypothesis rather than observational evidence. The large-scale structure of the Universe has yet to be investigated (Peebles, 1978); the distribution of quasars and radio-sources permit us to perform this task for very early epochs and for very large scales. On the other hand, a knowledge of the mode of distribution of these objects is also essential to provide some clues to the physical processes that govern their formation (isothermal or isentropic models; see for example Rees, 1982).

If a qualitative study is of interest at the beginning of an investigation, it cannot be considered an achievement. A rigorous approach implies the use of statistical tests elaborated for this particular case.

The problem can be defined as follows: we have an ensemble $\mathscr{E}$ of individuals X_i distributed in an n-dimensional space (usually $n = 1, 2, 3$ and the space is physical). Some questions then arise. Are these individuals randomly distributed? Do they have a tendency to cluster, to group together, to exhibit a contagious distribution or, on the contrary, do they exhibit a regular distribution (e.g., as atoms in a crystal lattice)? If they show a particular pattern, can the latter be described (nature, geometrical size, population size, ...)?

* Chercheur au 'Fonds de la Recherche Fondamentale Collective'.

The canonical statistical approach is to define a null hypothesis H_0 as simply as possible. For example, "The individuals are uniformly, independently, and randomly distributed in the field; the probability of finding one individual in an infinitesimal cell is proportional to the measure of this cell, is independent of the position of the individual and is independent of the position of any other individual."

One has then to build up a test or, more rigorously, a rule, i.e., an algorithm that will materialize itself into a test statistic $T = t(\mathbf{X})$ which is a reduction of the experimental data to a single value. Its distribution under H_0 is hopefully known and the observed value $T_{obs} = t(\mathbf{X})$ with $\mathbf{X} \in \mathscr{E}$ can be compared with it. A value of interest is usually one that may cast some doubt on H_0: this is to say that T_{obs} is in the tails of the distribution. For this reason, it is convenient to compute the probability under H_0 of obtaining at least such a value. For the sake of simplicity, one may prefer not to predefine a well-marked border between accepting and/or rejecting the null hypothesis. This leads us to call the above-mentioned probability the 'significance level'. Although of great practical interest, this definition is not common in statistical theories; we nevertheless adopt it.

Until now, many methods of analysis have been conceived: Binning Analysis, Statistical Reduction of Population, Power Spectrum Analysis, Extended Kolmogorov--Smirnov Test, Nearest Neighbours Analysis, and Correlation Function Analysis. All these have their own particular approach to the problem: we believe that an exhaustive study implies the use of several, if not all, of them together. This paper is concerned with an attempt to improve the Binning Analysis and to actually put it back in its right place.

1.2. The binning analysis (BA)

The BA is the oldest test: its use in astronomy dates back to Bok (1934). It has been one of the most extensively employed methods of analysis: Katz and Mulders (1942), Zwicky (1952, 1953, 1957), Neyman *et al.* (1954), de Vaucouleurs (1971), and more recently Osmer (1981). It is interesting to note that ecologists have used this method for many years and it was originally developed mainly by the Scandinavians, to whom we must be grateful for significant developments in the field. However, this test lost some of its importance, mainly because of the poor performances obtained when compared with more recent tests. We feel, however, that the nature of the BA is not the main reason. The majority of the problem stems from the level of the significance test: this paper intends to elaborate on a better system.

The BA consists of putting a cell at random on the field under investigation and counting the number of individuals in the cell.* The experience is repeated several times and one then observes the distribution of counts on which a significance test can be applied. If the null hypothesis is true, the individuals are found to be uniformly distributed, so the counts in the cells will be Poissonian as long as the expectation is

---

* A two-dimensional cell is called a quadrat.

small.* For high expectation, this is clearly no longer true and the counts will obey more closely a binomial distribution law. This method has the disadvantage of sampling the population of individuals. In addition, non-randomness is explored at only one scale: that of the cell.

An interesting improvement is due to Greig-Smith (1952). Instead of throwing cells at random, he uses a lattice of contiguous cells. He partitions the whole field, and all the individuals are taken into account. In this configuration it is possible to group the n-dimensional cells 2^n by 2^n in order to obtain the lattice of the next order and to analyse in such a way the randomness at another scale. This is called Multiple Binning Analysis. Again, observed counts can be compared to theoretical ones by performing a significance test. Those most commonly used are the χ^2 test and the variance/mean ratio test. Unfortunately, these methods have great deficiencies; using computer simulations, we shall outline, discuss and try to bypass them.

1.3. Synopsis of this work

In Section 2 we critically review the different existing significance tests linked to the BA and show that a better test can be obtained on the basis of randomization processes. In Section 3 we consider the one-dimensional binning analysis and associate with it the '2 within 4' randomization test: its power is then investigated. In Section 4 we consider the two-dimensional binning analysis and introduce the '4 within 16' randomization test. We also derive an interesting mathematical method for the randomization process. The power of such a method is also investigated. Section 5 deals with the three-dimensional case. In Section 6, some practical applications of the tests are presented. Conclusions are given in the last section.

2. Some Existing Significance Tests

We consider a field of measure A which contains n individuals: the density is $\rho = n/A$. We overlay this space with a lattice of N equal cells each of measure $a = A/N$. If, in the ith cell, we have r_i individuals, r_i is said to be the count of the ith cell.

2.1. The χ^2 test

Blackman (1935) used this obvious and very simple method. One first builds up the frequency distribution of the r_i. Let x_j be the number of cells having $r_i = j$. One can perform a conventional χ^2 test by comparing the observed frequencies x_j with the theoretical ones derived from the Poisson distribution

$$\chi^2 = \sum_{j=0}^{N_{\text{class}}} \frac{(x_j - NP_\mu(r = j))^2}{NP_\mu(r = j)} , \tag{2.1}$$

---

* This is an approximation; a detailed discussion can be found in Louis (1984) who shows, in essence, that if the counts in the cells are Poissonian, the distribution of individuals will be uniform but the reciprocal is not true in all cases. Such subtleties, however, have very little impact. It is essential to keep in mind that there may exist equifinality of fundamentally different generating processes.

where P_μ denotes the Poisson probability law for the mean $\mu \simeq \rho a = n/N$ and where the summation is performed over all the classes; as for the χ^2 goodness-of-fit test, if the expected number of members in one class is less than five, it has to be merged with another class (Cochran, 1952).

The first shortcoming of the χ^2 test is of a theoretical nature: for a Poisson distribution of random variables, the test is only asymptotically valid (Lindgren, 1976). Secondly, if the expectation is great in one cell, the use in Equation (2.1) of a multinomial distribution law instead of P_μ would be more exact. Nevertheless, the greatest problem is a practical one: if the density is small, then only classes $j = 0$ or 1 (and seldom 2 or 3) will be populated. This means that one has to perform the χ^2 test over an exceedingly small number of classes. This is almost impossible, especially as one of the degrees of freedom is absorbed by the computation of the mean of the distribution (in order to predefine the function P_μ) and also as the classes $j = 2$ and $j = 3$ are merged. The remaining number of degrees of freedom is consequently too small. It is also to be pointed out that the χ^2 test is unable to distinguish between the two partial alternative hypotheses H_{AC} (contagious distribution) and H_{AR} (regular distribution).

2.2. THE VARIANCE OVER MEAN RATIO TEST

Clapham (1936) used the variance/mean ratio v/m as a measure of the departure of the counts from the Poisson expectation, where v is the estimator of the variance and m, the mean, is

$$m = \rho a = \frac{n}{N}.$$

For a Poisson distribution, v/m is distributed around 1. A contagious distribution will induce an excess of cells with null counts and an excess of cells with high counts, i.e., an overdispersion (of the counts) and thus a value of v/m greater than 1. On the other hand, a regular distribution gives an underdispersion (of the counts) and thus a value of v/m less than 1. Under H_0, the ratio v/m is distributed with a mean of 1 and a variance

$$s^2 = \frac{2}{N}. \tag{2.2}$$

A demonstration and a discussion is given in Appendix A.

One of the failures of the test is that a markedly non-uniform distribution can produce a value v/m equal to 1 (see also Evans, 1952). So a significant deviation from 1 certainly indicates some kind of non-random arrangement, but the absence of any deviation does not necessarily imply true randomness.

In order to study in detail the behaviour of the test, we performed several simulations. In Figure 1 we show some of the runs of v/m against the characteristic scale investigated. The simulated populations are two-dimensional and their respective natures are:
    (a) 1000 individuals uniformly distributed;
    (b) 500 individuals uniformly distributed plus 100 individuals belonging to multiplets

containing on average two members separated by a distance of $1/100$ of the side of the square field;

(c) idem with 500, 200, $1/100$, respectively;

(d) idem with 500, 500, $1/100$, respectively;

(e) 1000 individuals belonging to multiplets containing on average four members separated by a distance of $1/1000$ of the side of the square field;

(f) idem but with a distance of $1/200$;

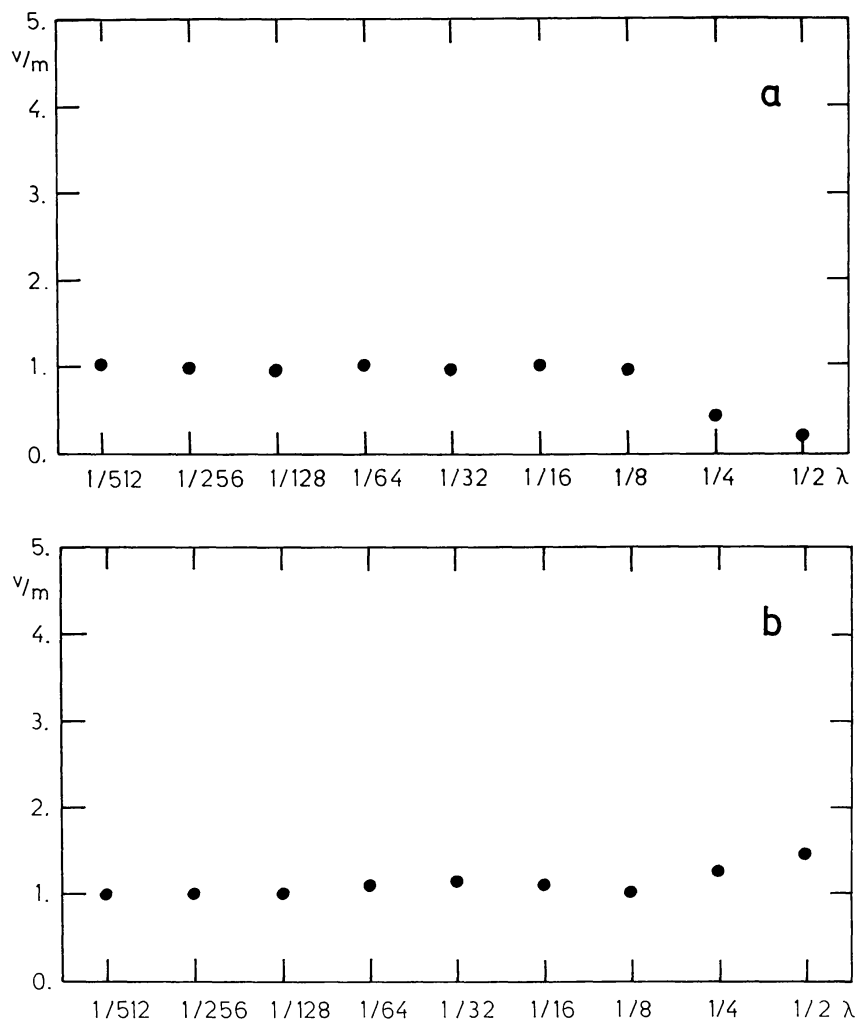(g) idem but with a distance of $1/100$;

(h) idem but with a distance of $1/50$.

Fig. 1(a)–(h).   Runs of the variance/mean ratio statistic against the side size of a cell expressed as a fraction of the field side length. The simulated populations are two-dimensional. Their respective characteristics are: (a) 1000 individuals uniformly distributed; (b) 500 individuals uniformly distributed plus 100 individuals belonging to multiplets containing, on average, two members separated by a distance of $1/100$ of the side of the square field; (c) idem but with 500, 200, $1/100$, respectively; (d) idem but with 500, 500, $1/100$, respectively; (e) 1000 individuals belonging to multiplets containing, on average, four members separated by a distance of $1/1000$ of the side of the square field; (f) idem but with a distance of $1/200$; (g) idem but with a distance of $1/100$; (h) idem but with a distance of $1/50$.
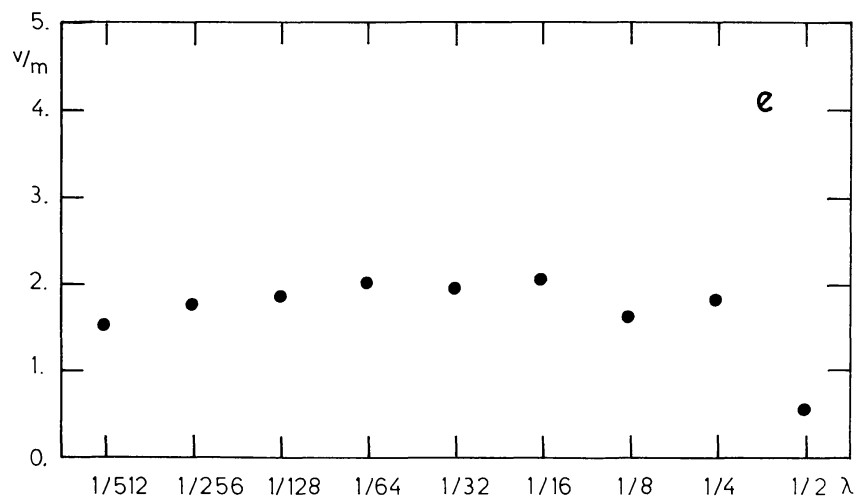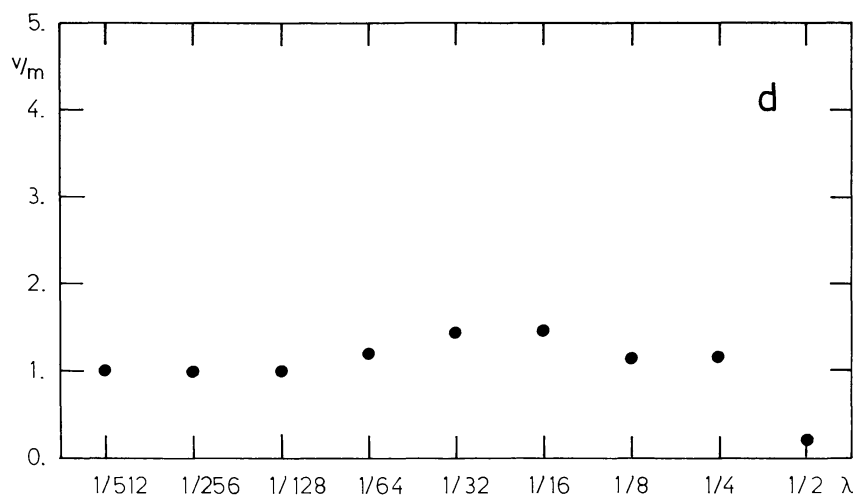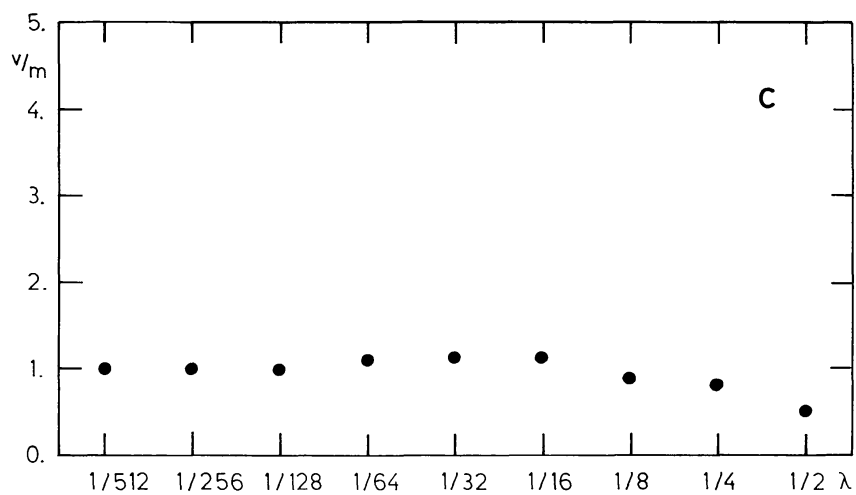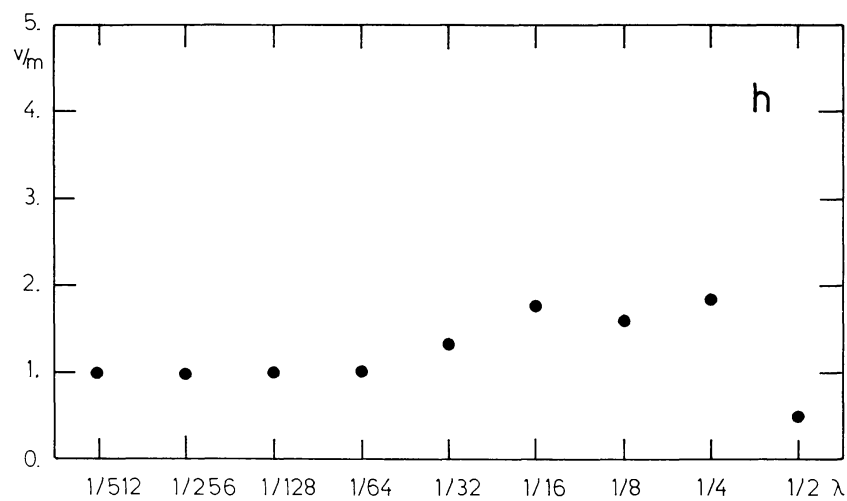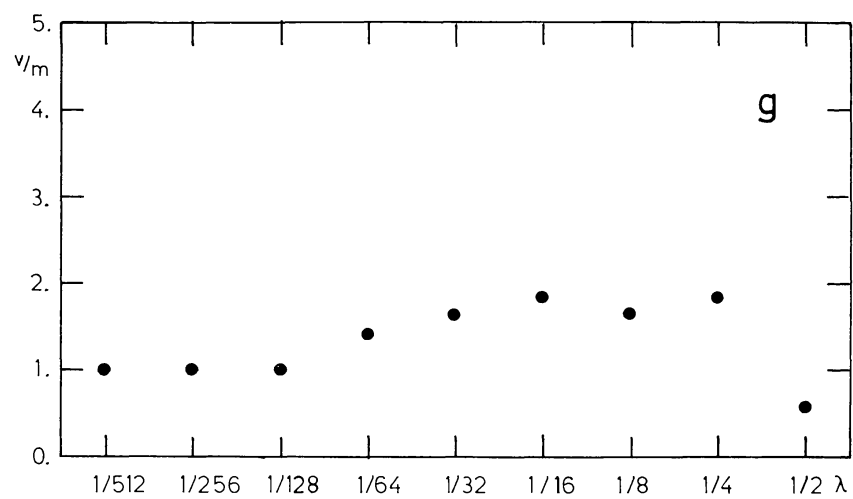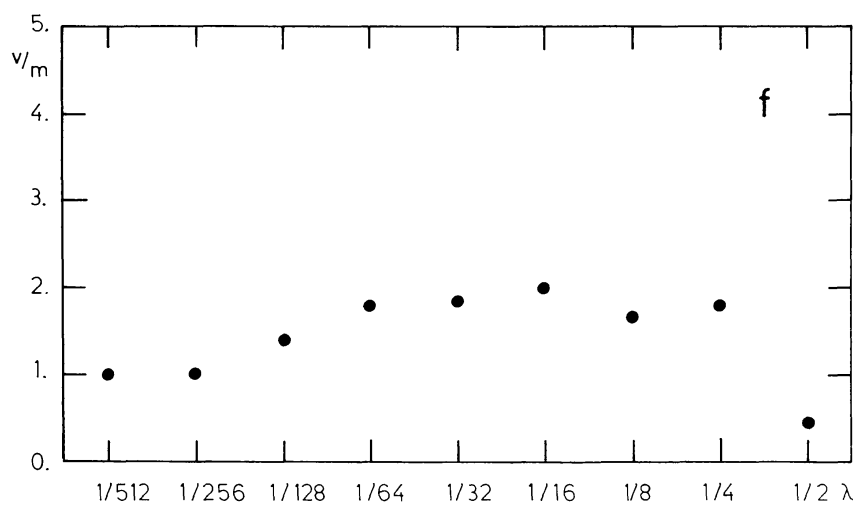
From part (a) of Figure 1, we can see that at scales $\frac{1}{4}$ and $\frac{1}{2}$ the test shows a strange behaviour. This is due to departure from the Poisson hypothesis. We can partly overcome this problem by using formulae (A.1) and (A.2) given in Appendix A. In the subsequent parts of Figure 1, the main shortcomings of the test are clearly outstanding. The v/m ratio markedly increases at the relevant scale of the association, the effect being proportional to the density of the latter. Although the behaviour at higher scales is very unusual, this can be easily explained by the presence of a pattern at some scale preventing the counts to be Poissonian at a greater scale. The effect, always towards overdispersion, is well marked even if the population shows no real pattern at higher scales. So the v/m ratio test permits a slight non-randomness to be detected, but as soon as one is detected, the test at higher scales is no longer reliable. In an elementary situation, such as that in Figure 1, the effect is not too severe, but in practice the test is quite intricate.

2.3. MOORE'S TEST

Trying to relieve ecologists from counting high-density quadrats, Moore (1953) developed a test based only on cells that have a low count ($r = 0, 1, 2$). The test statistic is

$$\phi = \frac{2x_0 x_2}{x_1^2} , \tag{2.3}$$

where the x_j are the observed frequencies, introduced in Section 2.1.

The first and second moment of the distribution of ϕ have been derived and a significant points table has been published.

In astronomy, and with the aid of modern computers, there is no reason to neglect high-density cells. Furthermore, the test exhibits the same problems as those encountered for the v/m test.

2.4. MORISITA'S TEST

Simpson (1949) has proposed a measurement of diversity to apply to classification schemes. In his paper, he defined λ (a measure of the degree of the concentration of the classification) and found an estimator l for this statistic. Taking advantage of this formalism and identifying the groups of the classification theory with the cells of the BA, Morisita (1959) proposed a test of which the statistic is

$$I = \frac{N \sum\limits_{i=1}^{N} r_i(r_i - 1)}{n(n - 1)} . \tag{2.4}$$

Crudely, this index has a value of about 1 under H_0, especially when n is large. Under H_{AC} it is located between 1 and N, whereas under H_{AR} it is smaller than 1. A detailed explanation of the I statistic and of its behaviour, as well as of the relevant significance test indirectly based on the F-distribution, can be found in the paper of Morisita (1959). I is characterized by a rather strong lack of sensitivity to the square-size as long as the latter is inferior to the clump size (in Morisita's terminology). This means that the

presence of a clustering at a given scale will make I large for most, if not all, of the smaller scales investigated. This inter-scale correlation makes the test slightly inefficient in deriving the size of the features existing in the investigated distribution. It is also necessary to consider several scales in order to gain information on only one.

One of the conditions for a safe application of Morisita's test is that n, the number of individuals on the field, must be greater than N, the number of quadrats. It is clear that this restriction is more easily satisfied in ecological applications than in astronomical ones. Despite all the above-mentioned remarks, the significance test is probably the best of those hitherto proposed.

2.5. THE LIKELIHOOD RATIO TEST

Mead (1974) has introduced a new test that is quite interesting in regards to the originality of the approach. One takes adjacent cells two by two; the counts in the first (respectively, second) are r_{2i-1} (respectively, r_{2i}), and the total is $n_i = r_{2i-1} + r_{2i}$. Each individual belonging to a pair has a probability p_i of being in the first cell. If the null hypothesis is true, the pairing will be random and p_i will be equal to $\frac{1}{2}$ for all i's. We can construct a likelihood ratio test of the hypothesis $p_i = \frac{1}{2}$ against the general alternative $p_i \neq \frac{1}{2}$. The relevant likelihoods are

$$L_{\max} = \prod_i \frac{n_i!}{(n_i - r_{2i-1})! \, r_{2i-1}!} \left[\frac{r_{2i-1}}{n_i} \right]^{r_{2i-1}} \left[\frac{n_i - r_{2i-1}}{n_i} \right]^{n_i - r_{2i-1}}$$

and

$$L_{1/2} = \prod_i \frac{n_i!}{(n_i - r_{2i-1})! \, r_{2i-1}!} \left[\frac{1}{2} \right]^{r_{2i-1}} \left[\frac{1}{2} \right]^{n_i - r_{2i-1}} ;$$

and the ratio

$$\lambda = \frac{L_{1/2}}{L_{\max}} .$$

One can easily derive that

$$-2 \log \lambda = 2 (\log 2) \sum_i n_i + 2 \sum_i [r_{2i-1} \log r_{2i-1} +$$

$$+ (n_i - r_{2i-1}) \log(n_i - r_{2i-1}) - n_i \log n_i] , \qquad (2.5)$$

where the summation is extended over all the N_0 pairs for which we have $n_i \geq 2$. From the properties of the likelihood ratio statistic (Lindgren, 1976), we can state that $-2 \log \lambda$ is asymptotically distributed as a χ^2 with N_0 degrees of freedom.

The method permits us to be free from the Poissonian hypothesis. Unfortunately, our simulations show that the test is not as perfect as one would hope: here, at small scales, the small number of pairs prevents us from reaching the asymptotic distribution. On the other hand, the expression of the likelihoods is still based on a theoretical law which

could be too restrictive. At large scales, slight deviations from this law are no longer blurred through the summation: this is due to the small number of pairs. Furthermore, this method suffers to some extent from the same problems as the previous tests, i.e., an interaction between different scales. The basic idea remains nevertheless quite interesting.

2.6. GENERAL COMMENT

All the methods described above, except that of the likelihood ratio, consider counts in cells independently of the location of the other cells. We feel that some attention should be given to the possible correlation existing between adjacent cells. On this point, the basic idea of the likelihood ratio test is of great help. However, such an approach is necessarily highly dependent on the nature of the investigated space: this is the reason we decided to treat one-, two-, and three-dimensional spaces separately in the remainder of the present paper.

It should also be noted that most of the tests described above have a disturbing property: the detection of a non-randomness at a given scale induces an erratic behaviour when other scales are investigated. Usually, the tests become inefficient at scales larger than that for which a deviation from randomness is present: this is always due to the implicit or explicit assumption that the counts obey an *a priori* theoretical law (usually the Poisson, sometimes the multinomial law). In the case of a significant departure from such a law at a given scale, the theoretical counts are, of course, no longer correct at larger scales: the tests thus lead to significant values of the statistic even when no pattern is actually present in the data! From this, it is clear that the solution consists of taking into account not a theoretical distribution but one which is drawn from the data themselves. An interesting procedure, then, is that of randomization: such a concept was probably first introduced by Fisher (1947). It is actually one of the modern features caused by the interaction between statistics and experimental designs. The basic philosophy is, while repeating an experiment, to shuffle all the experimental apparatus in order to eliminate accidental correlations arising from the algorithm that was used. More details can be found in the reference mentioned above.

Our approach appears to be slightly different: we wish to permute some features involved in our basic data and, each time, to compute the chosen statistic. We shall thus obtain, from the data, a distribution of the statistic against which the observed value can be tested. We are going to show that the basic philosophy is not very different. Let us first remark that, if we place two cells in a field containing randomly distributed individuals, the counts in the two cells will be uncorrelated as long as the cells remain disjoined. This is no longer true if the individuals are distributed in some other way. For example, let us consider a field containing clusters: if a cell coincides at least partially with a cluster, adjacent cells will have a greater chance to contain other parts of the cluster and a correlation between adjacent cells appears. The effect is the same for cells in voids between clusters. Further, the effect still exists in a field containing regularly distributed individuals. Shuffling or permuting cells eliminates the possible correlations: we then have to deal with a randomization process.

In the next sections, we shall take advantage of the features underlined above and more precisely define the statistic we favour. Before that, however, we must introduce a new and more adequate definition for the null hypothesis H_0 that is less restrictive:

H_0: "at a given scale, there is no pattern in the distribution of individuals".

As mentioned previously, to look for a pattern is less restrictive: to detect a pattern at a scale does not imply that H_0 is not correct at some other scale.

3. The One-Dimensional Binning Analysis

Let us consider a one-dimensional field of length l partitioned into four N bins. Again, the count in the ith bin is r_i. Let us pair adjacent bins such that the count in the jth pair is

$$n_j = \sum_{k=2j-1}^{k=2j} r_k .$$
(3.1)

Under H_0, there is no pattern at this scale; so pairings should be random. This can be tested from the observed population through a randomization process. We define a statistic which is a function of the n_j's and consider its distribution over random pairings: this distribution, called 'the randomization distribution', is obtained by performing all possible permutations of the counts r_i. The simplest statistic is the variance of the counts of the pairs, which is within some factor

$$\theta = \sum_{i=1}^{2N} n_i^2 - \frac{1}{2N} S_1^2 .$$
(3.2)

where

$$S_1 = \sum_{j=1}^{4N} r_j .$$

The expectation $E(\theta)$ and variance var(θ) over the randomization process can be easily obtained from simple, although tedious, calculations. The randomization distribution is usually too large to be tabulated, but we can use an approximate normal test using the statistic

$$Z = \frac{\theta - E(\theta)}{\sqrt{\text{var}(\theta)}} .$$
(3.3)

However, as pointed out by Mead (1974), the existence of a pattern at a scale larger than the one being investigated implies that the counts may consist of two or more populations whose characteristics are quite different. So, the randomization test is invalid. The way to overcome this difficulty is to reduce the application zone of the randomization process in order to ascertain that this zone contains only one type of population and then to add, through a normalization process, the intermediate statistics relevant to each zone.

3.1. The 2 within 4 randomization test

The most drastic reduction occurs to pairs of pairs, i.e., to four bins. Let us take as a criterion the absolute difference between the counts of the pairs. Explicitly, for the first pair of pairs, we have the counts r_1, r_2, r_3, r_4.

The criterion is

$$k_0 = |(r_1 + r_2) - (r_3 + r_4)| .$$
(3.4)

The randomization process gives also the values

$$k_1 = |(r_1 + r_3) - (r_2 + r_4)| ,$$

$$k_2 - |(r_1 + r_4) - (r_2 + r_3)| .$$

Thus the randomization distribution consists of three values and one can test the position of k_0 with respect to those of k_0, k_1, and k_2. This three-point randomization distribution is not quite large enough to perform an adequate test; however, we have N such pairs of pairs and the overall distribution then consists of 3^N values. This test, which is constructed to look at the randomness of the repartition of counts in four bins into counts in two pairs, is known (Mead, 1974) as the '2 within 4' randomization test. It is clear that if one of the r_i is large, it will have an important effect on the randomization distribution compared to the others. This leads directly to an attempt to find a method of normalization that is possible if we remark that the k_i are usually taking one of the three forms

$$(m, m, n) , \quad (m, n, s) , \quad \text{or} \quad (m, n, n) ,$$

where $m < n < s$ and the three numbers are all even or all odd.* A distribution-free version is then available if we replace the three forms above, respectively, by

$$(0, 0, 2) , \quad (0, 1, 2) , \quad \text{and} \quad (0, 2, 2) .$$

Their means and estimated variances are, respectively,

$$2/3 \text{ and } 8/9 ; \quad 3/3 \text{ and } 2/3 ; \quad 4/3 \text{ and } 8/9 .$$

One can use a normal test merging those three normalized distributions. If we have n_{10} observations corresponding to 0 and n_{12} observations corresponding to 2 from n_1 distributions $(0, 0, 2)$, and n_{20} corresponding to 0, n_{21} to 1 and n_{22} to 2 from n_2 distributions $(0, 1, 2)$ and, finally, n_{30} corresponding to 0 and n_{32} corresponding to 2 from n_3 distributions $(0, 2, 2)$ then

$$Z = \frac{n_{21} + 2(n_{12} + n_{22} + n_{32}) - (2n_1 + 3n_2 + 4n_3)/3}{\sqrt{2(4(n_1 + n_3) + 3n_2)/9}} ,$$
(3.5)

is approximately normally distributed with a mean of zero and a variance of 1 under

---

* It is to be noted that this is not the case if and only if three of the four counts are equal; in particular, a normal situation is when the counts are equal to zero.

the null hypothesis that there is no pattern in the repartition of four-bin counts into two-pair counts. Positive values of Z correspond to a tendency to cluster, and negative values to a tendency to show a regular distribution. The investigated scale is of the order of two bins, i.e., $l/2N$.

3.2. Study of the '2 within 4' randomization test

In this section we investigate the behaviour of the '2 within 4' randomization test. As the large number of simulations required cannot be given in detail in the present paper, we shall limit ourselves to the main results and conclusions.

3.2.1. *The Distribution of Z Under H_0*

Using Monte-Carlo techniques, we compute the distribution of Z for a uniform distribution of the individuals. Although sometimes irregular, the distribution is rather near normal for all scales. The mean is always almost equal to 0 and the standard deviation near 1; this confirms the assertion we gave in Section 3.1 and the conclusions of Mead (1974).

The main deviation comes when the number $(n_1 + n_2 + n_3)$ becomes too small. The discrete nature of the Z statistic then shows up. Such an effect happens in two cases: first, when we investigate a large scale and the number of bins is, therefore, small ($N = 1, 2$); and second, when the bin expectation is very small. This means that we have a great number of pairs of pairs where only one bin has a non-zero count. Such pairs of pairs, where three counts are the same, are not taken into account so that $(n_1 + n_2 + n_3)$ is reduced. In any case, it should be pointed out that the discrete nature of the statistic does not necessarily weaken the results of the test as long as the envelope of the peaks obeys the relevant distribution law; although not rigorously true, this is generally the case.

3.2.2. *Under the Alternative Hypothesis H_A*

Contrary to H_0, which is a simple hypothesis, the alternative hypothesis is composite. It can be considered as a union of hypotheses: symbolically we may write

$$H_A = H_{AC} + H_{AR} + H_{AU},$$

where H_{AC} and H_{AR} are already known and where H_{AU} includes all the other models of non-randomness.

Due to the complex nature of H_A, it is impossible to simulate it: we can only choose some simple cases of practical interest.

The first cases we are going to investigate are part of H_{AC}. We define in the field several points randomly distributed, each becoming the centre of a cluster of individuals randomly distributed with equal variances. Figure 2 shows the run of $\langle Z \rangle$ against the number of bins (which is related to the investigated scale). Figure 2(a) refers to 25 clusters of, on average, four members distributed with a maximal dispersion of $l/512$. Figures (2b) and (2c) refer to the same pattern with dispersions of $l/340$ and $l/256$,
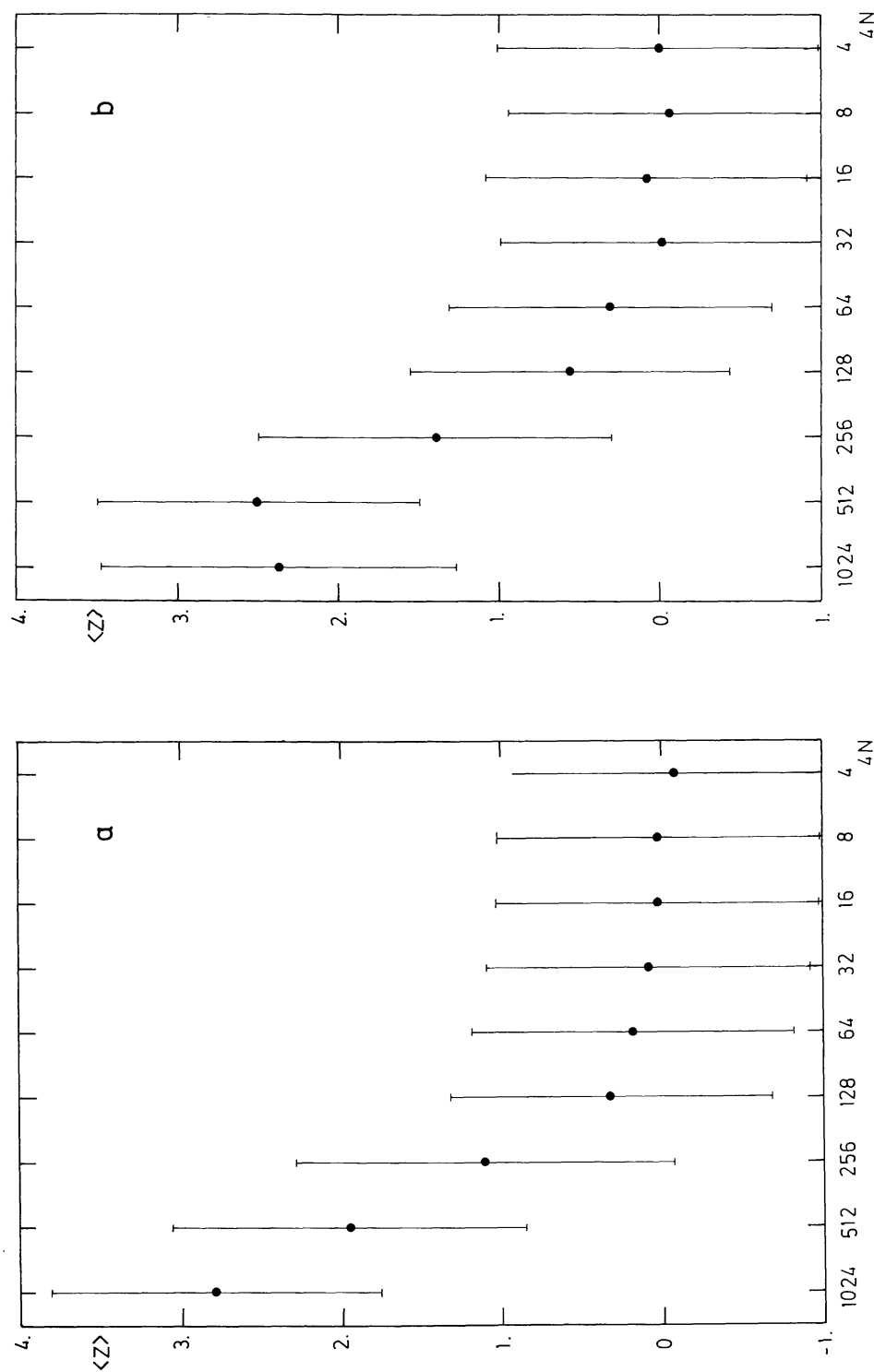
E. GOSSET AND B. LOUIS

Fig. 2(a)–(d). Runs of the mean $\langle Z \rangle$ over the simulated distribution of the Z statistic of the '2 within 4' randomization test against the number of bins of the investigating grid. Error bars are ± 1 standard deviation of Z over the simulated distribution. The characteristics of the distributions of individuals over the field are: (a) 25 clusters of, on average, four members distributed with a maximal dispersion of $l/512$; (b) idem but a dispersion of $l/340$; (c) idem but a dispersion of $l/256$; (d) 5 clusters of, on average, twenty members distributed with a maximal dispersion of $l/512$.
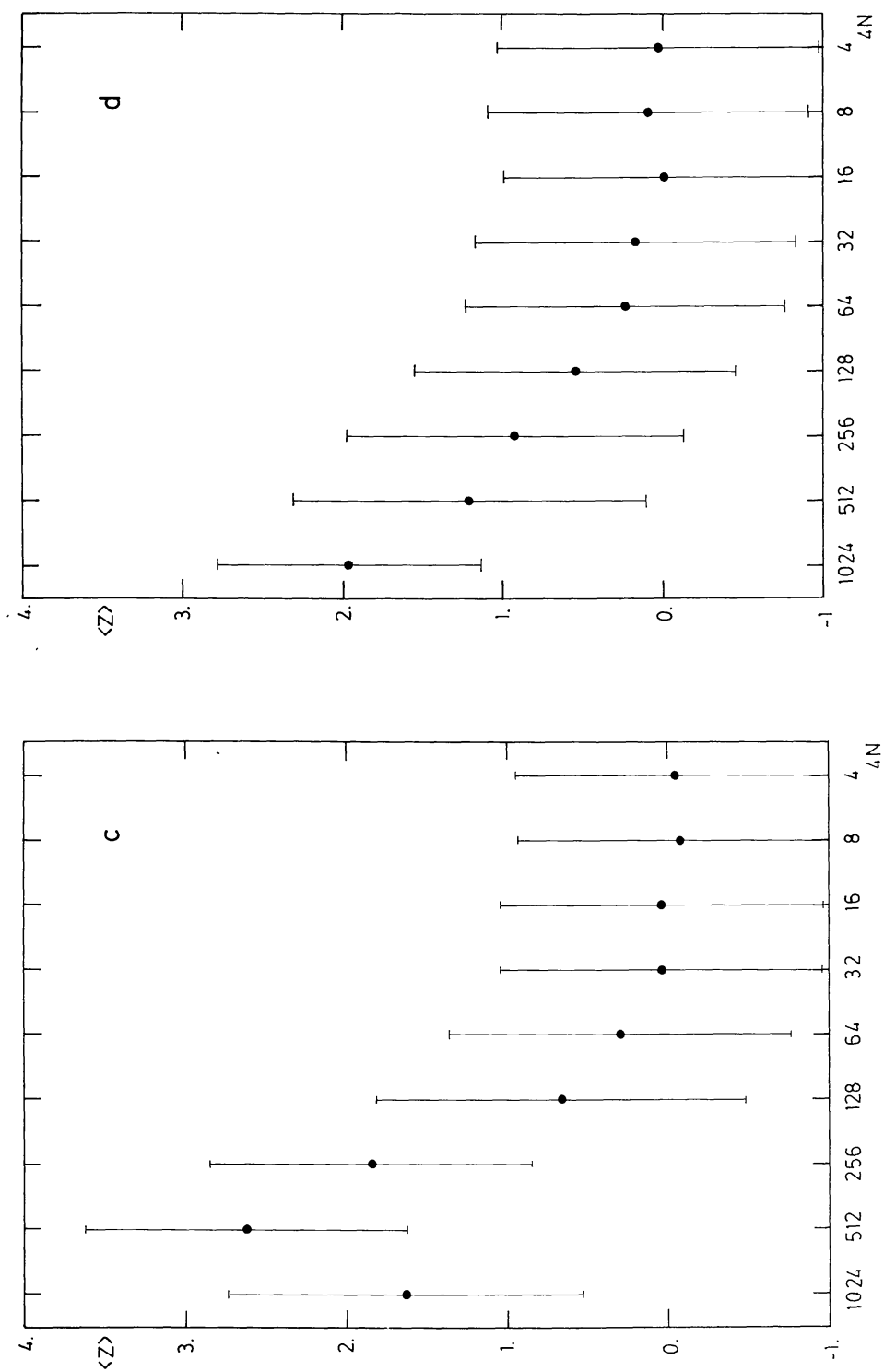
respectively. Figure 2(d) is for 5 clusters of, on average, 20 members distributed with a maximal dispersion of $l/512$. From these graphs, we can conclude that:

(a) the test reacts at the right place of $l/2N$, i.e., respectively at $4N = 1024$, ~ 680, 512, 1024;

(b) at $4N \leq 128$, it is particularly robust* in regard to the different kinds of clustering and the Z statistic, therefore, remains at totally insignificant values;

(c) at values adjacent to a significant one, there is a small but marked tendency to exhibit higher values; it is a remanence of the two populations effect;

(d) contrary to the PSA, the test is more sensitive to the number of clusters than to the number of individuals in each cluster (compare part (a) and part (d)); on this point, the two methods are quite complementary.

In summary, the test behaves well; however, in order to refine our approach, it is interesting to investigate its power. Let us define

$$\alpha^+ = \text{size of type I error} = P_{H_0} \, (\text{reject } H_0) \,,$$

$$\beta^+ = \text{size of type II error} = P_{H_{AC}} \, (\text{accept } H_0) \,.$$

P_{H_0} (which denotes the probability under H_0) is easy to compute as H_0 is a simple hypothesis; but this is not the case with $P_{H_{AC}}$ as H_{AC} is still composite. However, one may compute P for different simple hypotheses included in H_{AC}. Some results are presented in Figure 3. We have plotted $P_{H_0} \, (Z > K)$ as a function of K (≥ 0) and α^+ can be readily read off this graph as the ordinate of the curve. We have also plotted $P_{H_{AC}}$ $(Z < K)$ for different simple hypotheses (e.g., 25 clusters of on average four members distributed with a dispersion of $l/512$, the same with a dispersion of $l/256$ and, finally, 5 clusters of, on average, 20 members with a dispersion of $l/512$). β^+ is again easily read off the graph as the ordinate of those latter curves. Figure 3 will be of great help for those wishing to adopt a well-defined border between rejecting or accepting H_0.

If the power of the test is good, it is clearly not perfect. However, we think it is powerful enough to suggest its use more specifically since the general behaviour of the test is of great practical interest.

The second case we are going to investigate is part of H_{AR}. We define in the field a lattice of equidistant nodes separated by a distance of $l/128$. One of these nodes is chosen at random, and an individual is randomly positioned around this node with a maximal dispersion $\delta = l/512$. Again, another node is chosen at random, and an individual is randomly positioned around this point with the same dispersion. Figure 4(a) shows the run of $\langle Z \rangle$ against the number of bins. Figure 4(b) is the same for $\delta = l/256$, and Figure 4(c) is for $\delta = l/205$.

The following information can be seen from the graphs.

(a) The Z statistic shows a very small value at $4N = 256$ corresponding to an investigated scale of $l/128$. So, the regular distribution is well detected (see Figures 4(a) and (b)) in a significant way as long as δ is less than half the step of the regularity;

---

* By robust we mean 'almost insensitive to the type of the random variable distribution'. A more rigorous definition can be found in Kendall and Stuart (1967).
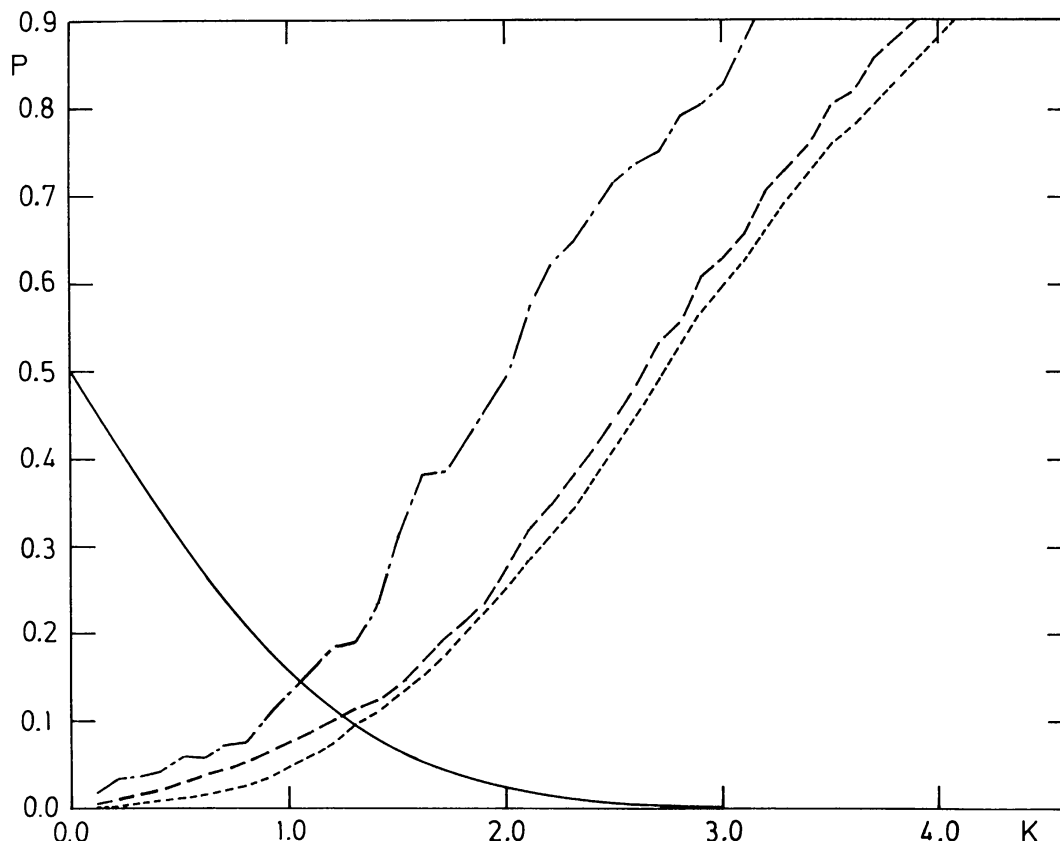
Fig. 3.   Probability $P_{H_0}(Z > K)$ as a function of the parameter $K \geq 0$ for the Z statistic of the '2 within 4' randomization test (continuous line). Three estimates of $P_{H_{AC}}(Z < K)$ are also represented. The three relevant simple hypotheses are: (1) 25 clusters of, on average, four members distributed with a dispersion of $l/512$ (dotted line); (2) idem but a dispersion of $l/256$ (dashed line); (3) 5 clusters of, on average, twenty members with a dispersion of $l/512$ (dashed-dotted line).

Figure 4(c) also exhibits a detection, but this is insignificant due to the large value of $\delta = l/205$.

(b) The Z statistic has a remarkable behaviour for $4N = 1024, 64, 32, 16, 8, 4$. The non-randomness detected at $4N = 256$ clearly does not affect the ability of the test to recognize randomness at these other scales. It has a marked robustness.

(c) The Z statistic at $4N = 128$ is slightly affected by the deviation from randomness detected at $4N = 256$. This is the only remanence of the above-mentioned two-population effect. However, the discrepancy ($Z \simeq - 0.38$) does not correspond to a significant deviation.

(d) At $4N = 512$, if $\delta = l/256$ or $\delta = l/205$, and at $4N = 1024$, if $\delta = l/512$, the test shows a pseudo-clustering which is no more than an artefact of the regular pattern.

A crucial question concerns the influence of the relative phasing between the bin lattice and the node lattice. Figures 4(a), (b), (c) were constructed for nodes belonging to the first bins of the pairs of pairs. Further simulations show that the test reacts almost always at the right scale. However, for very particular relative phasings, the reaction takes place at half the expected scale.
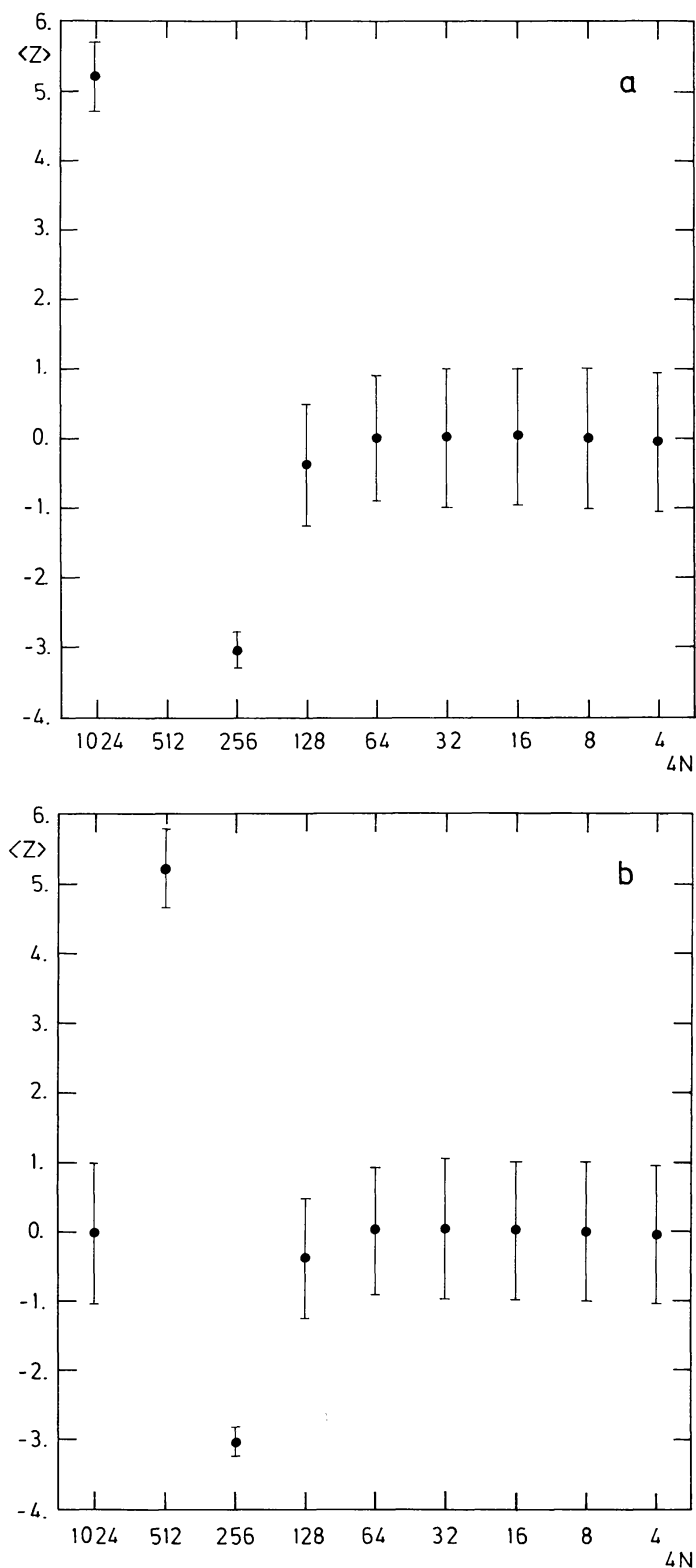
Fig. 4(a)–(c).   Runs of the mean $\langle Z \rangle$ over the simulated distribution of the Z statistic of the '2 within 4' randomization test against the number of bins of the investigating grid. Error bars are ± 1 standard deviation of Z over the simulated distribution. The distributions of individuals over the field are regular with a periodicity of $l/128$ and the dispersions of individuals around the nodes are: (a) $l/512$; (b) $l/256$; (c) $l/205$.
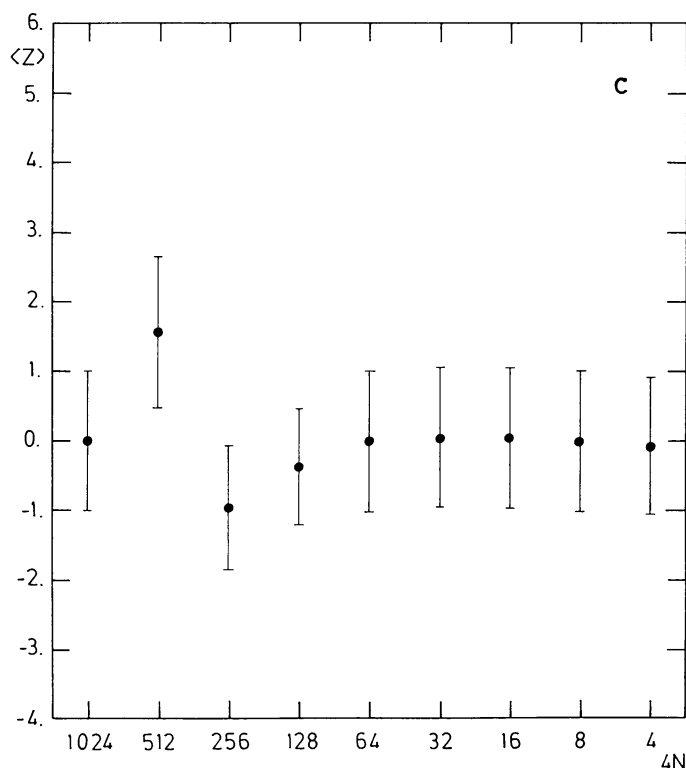
Although a little annoying, such a characteristic is not disastrous. Looking for a regular pattern, we shall probably have to try several phases and several scales. One can take advantage of the inter-scale robustness of the test to move the lower and the upper bounds of the investigated field in order to have a different phasing and a different set of scales $(l/2^n \Rightarrow (l + \Delta l)/2^n)$; the field beyond the true bounds remains, of course, empty.* Such an approach allows the deduction of the true periodicity.

We have also studied the power of the test with respect to the regular alternative hypotheses. One can easily derive the α^- and β^- (approximate size of type I and type II error, respectively) from Figure 5. We have plotted $P_{H_0}(X < K)$ and $P_{H_{AR}}(X > K)$ against $K\ (< 0)$. Again H_{AR} is not exhaustively defined; we give the curve in two simple cases: fixed phase and random phase, each with $\delta <$ half the periodicity. All practical situations of interest will be within the two curves. The test seems to be more powerful against H_{AR} than against H_{AC}.

4. The Two-Dimensional Binning Analysis

We now deal with a two-dimensional field: for the sake of simplicity, we consider a square of dimension $l \times l$ partitioned into $4N \times 4N = 16N^2$ quadrats. This assumption is not restrictive. It is clear that in order to test the distribution of individuals, one may

* It must be clear that such generated additional Z-statistic values are not independent of the basic values: although several trials are permitted in order to deduce the correct scale, only one significance level can be taken into account.
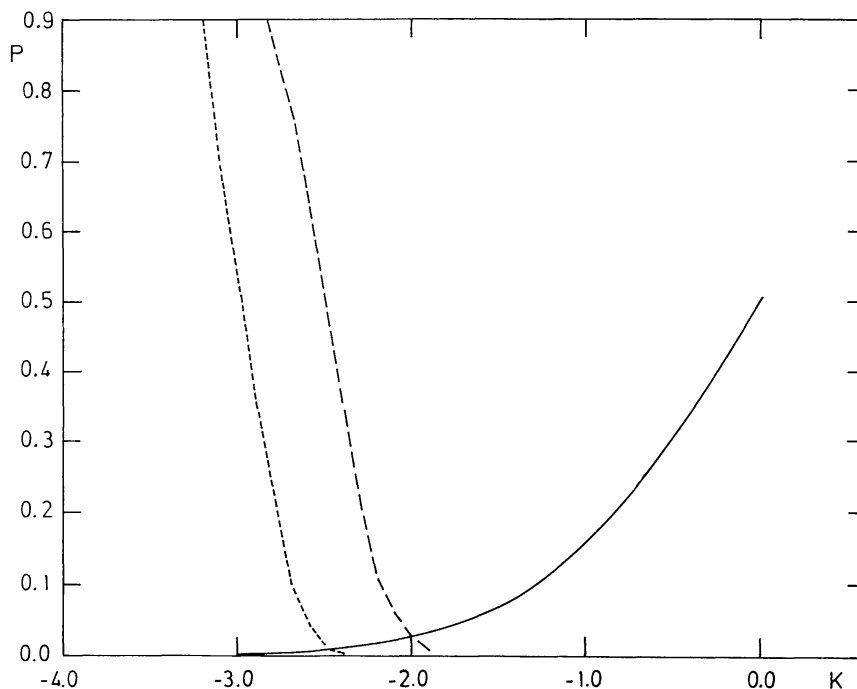
Fig. 5.   Probability $P_{H_0}(Z < K)$ as a function of the parameter $K \leq 0$ for the Z statistic of the '2 within 4' randomization test (continuous line). Two extreme estimates of $P_{H_{AR}}(Z > K)$ are also represented. The relevant simple cases are: (1) fixed phase (dotted line); (2) random phase (dashed line). See text for more details.

use the above-mentioned '2 within 4' randomization test: one has only to continuously rearrange the matrix of quadrat counts row by row, or column by column, and to put them into a vector mimicking a one-dimensional field. This procedure, which may appear arbitrary, is in fact the historical and, until this work, the only way the test has been used (Mead, 1974).

Two remarks are to be made. First, the test is anisotropic and, for the sake of completeness for astronomical applications, must be performed twice: row by row ('horizontally') and column by column ('vertically'). This leads to two sets of results which can be quite discordant and a subjective choice must be made. In any case, its anisotropic nature prevents the test from indicating only the clustering, and some directional properties of the distribution of the individuals are also enshrined in the results. Second, the elementary group of quadrats investigating the field has dimensions of 1 by 4 quadrats. So, an undesirable mixing of the scales is induced, which may be a problem. Despite this, performances of the '2 within 4' randomization test for the two-dimensional case are encouraging. An example is given in Figure 6: the field contains 14 randomly distributed individuals and 12 individuals in two by two clusters, separated by an average distance $l/10$. The run of Z against the quadrat size is given for the two possible arrangements of the quadrats four by four. The test reacts for a scale of $l/16$; i.e. the nearest to the correct $l/20$. A more suitable method will nevertheless be derived in the next section.
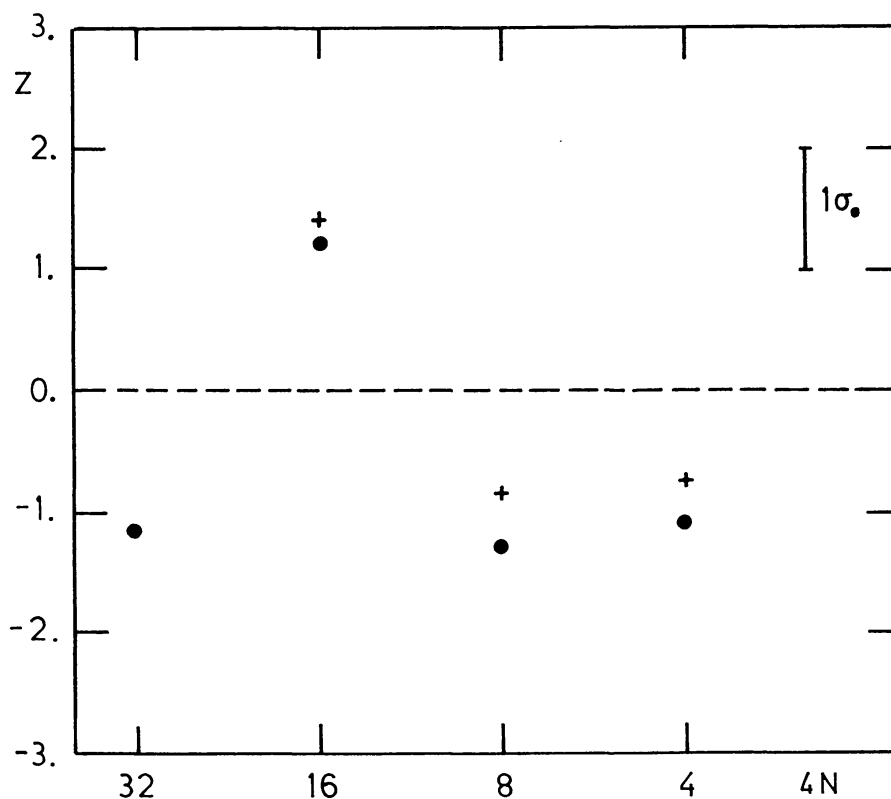
Fig. 6. Run of the Z statistic of the two-dimensional '2 within 4' randomization test against the number of quadrats on one line of the investigating grid. The analysed population consists of fourteen randomly distributed individuals and twelve individuals in cluster two by two and separated by, on average, a distance of $l/10$. Dots refer to a test performed with elongation of the basic quartet along the x-axis whereas crosses refer to an elongation along the y-axis. The error bar represents one standard deviation under the null hypothesis.

4.1. THE '4 WITHIN 16' RANDOMIZATION TEST

Let us first adopt a systematic numbering of quadrats such as that illustrated in Figure 7. The count in the ith quadrat is r_i. Let us group adjacent quadrats four by four (isotropically) in such a way that the count in the jth quartet of squares is

$$n_j = \sum_{k=4j-3}^{k=4j} r_k . \tag{4.1}$$

Under the null hypothesis H_0, there is no pattern at this scale: divisions in four should thus be random and this can be tested from the observed population through the randomization process. Shanks (1979) tried to directly transpose the method of Mead to this case. Instead of pairs of pairs, he considered quartets of quartets. The relevant criterion equivalent to 3.4 is for the first quartet,

$$k_0 = |n_1 - n_2| + |n_1 - n_3| + |n_1 - n_4| + |n_2 - n_3| +$$

$$+ |n_2 - n_4| + |n_3 - n_4| ; \tag{4.2}$$

1	2	5	6	17	18	21	22	65	66
3	4	7	8	19	20	23	24	67	68
9	10	13	14	25	26	29	30	73	74
11	12	15	16	27	28	31	32	75	76
33	34	37	38	49	50	53	54	97	98
35	36	39	40	51	52	55	56	99	100
41	42	45	46	57	58	61	62	105	106
43	44	47	48	59	60	63	64	107	108
129	130	133	134	145	146	149	150	193	194
131	132	135							

Fig. 7.   Illustration of the adopted systematic numbering of the quadrats.

and the randomization is obtained by Shanks (1979) through a Monte-Carlo technique. Such a method deserves two criticisms. First, no normalization such as that leading to Equation (3.5) is applied; a high value can, therefore, dominate the distribution and induce false conclusions. However, we must admit that a normalization on such a complex criterion as k_0 is quite difficult to conceive. Second, the use of Monte-Carlo techniques makes the method slow and tedious. In addition, we cast some doubt on the reliability of the method since the suggested number of simulations for the randomization is less than 1 to 20 000 of all the possibilities. Such a number is definitely too small to lead to an efficient approximation of the parent distribution.

In what follows, we shall derive an improved test. The statistic we suggest is the variance of the counts of the quartets, which is within some factor

$$\theta = \sum_{i=1}^{4N^2} n_i^2 - \frac{1}{4N^2} S_1^2 , \tag{4.3}$$

where

$$S_1 = \sum_{i=1}^{16N^2} r_i .$$

The expectation and variance over the randomization process are

$$E(\theta) = \frac{(4N^2 - 1)}{4N^2(16N^2 - 1)} \, [S_0S_2 - S_1^2] \, , \tag{4.4}$$

$$\mathrm{var}(\theta) = \frac{12(4N^2 - 1)}{(16N^2 - 1)^2 \, (16N^2 - 3)\,(8N^2 - 1)} \, [S_0S_2 - S_1^2]^2 \, +$$

$$+ \frac{12(4N^2 - 1)}{(16N^2 - 1)\,(16N^2 - 3)\,(8N^2 - 1)} \, [4S_3S_1 - 3S_2^2 - S_0S_4] \, , \tag{4.5}$$

where

$$S_j = \sum_{i=1}^{16N^2} r_i^j \, .$$

The demonstration of this result is given in Appendix B. As previously, the randomization distribution is always too large to be tabulated but we can also use an approximate normal test using a statistic similar to that given in Equation (3.3). For the same reason as for the one-dimensional case, such a randomization test is not valid due to the multipopulation effect. Again, the way to overcome the difficulty is to reduce the application zone. The necessary reduction is to quartets of quartets, i.e., to sixteen quadrats, from which the name of the test is derived: the '4 within 16' randomization test. It would be interesting to find a normalization process analogous to the one leading to the robust version of the '2 within 4' randomization test. However, the randomization is no longer on 3 values but on

$$\frac{16!}{(4!)^5} \sim 2.6 \times 10^6 \, \text{values} \, .$$

This is, of course, too large a number not only to be tabulated but even to make possible the derivation of a criterion such as the one leading to Equation (3.5).

Nevertheless, we can take advantage of this great number of values of the randomization distribution, since it ensures that the test on sixteen quadrats is completely realistic on its own. So, for the ith quartet of quartets, one assumes that the statistic

$$Z_i = \frac{\theta_i - E(\theta_i)}{\sqrt{\mathrm{var}(\theta_i)}} \, , \tag{4.6}$$

is almost normally distributed with a mean of 0 and a variance of 1 ($E(\theta_i)$ and $\mathrm{var}(\theta_i)$ are easily computed using Equations (4.4) and (4.5) with $N = 1$).

This is clearly not the case as the intrinsic nature of the statistic is that of the square of a random variable. Nevertheless, we believe that the normalization procedure of Equation (4.6) is sufficient to lead to a good approximation. It is still possible, of course, to use a more sophisticated procedure.

The Z statistic is then obtained simply by adding together the intermediate statistics Z_i. Thus, we have

$$Z = \frac{1}{\sqrt{\mathfrak{A}}} \sum_{i=1}^{\mathfrak{A}} Z_i ,$$

(4.7)

where the summation is extended to all the non-empty quartets of quartets for which $\text{var}(\theta_i)$ is defined. Obviously, we have $\mathfrak{A} \leq N^2$. Again, Z is distributed with a mean of 0 and a variance of 1, in particular when $\mathfrak{A}$ is large.

4.2. STUDY OF THE '4 WITHIN 16' RANDOMIZATION TEST

It is of great interest to analyse the behaviour of this new test. A very large number of trials and simulations was performed but cannot be presented here in detail: we shall restrict ourselves to the main results.

4.2.1. *The Distribution of Z Under H_0*

The distribution of Z under H_0 is illustrated in Figure 8. It refers to 2000 simulations of a field containing 200 randomly distributed individuals. Part (a) shows the distribution of Z corresponding to a quadrat size $l/128 \times l/128$; part (b) to a quadrat size $l/64 \times l/64$, where it is readily seen that the relevant distribution is strongly near normal. In fact, deviations occur in two cases. First, when $\mathfrak{A}$ is too small, the Central Limit Theorem can no longer be applied: the crude normalization procedure of Equation (4.6) is no longer blurred through the summation of Equation (4.7). This is illustrated in Figure 8(c) where the Z corresponds to a quadrat size of $l/8 \times l/8$, i.e., $\mathfrak{A} = 4$. One can
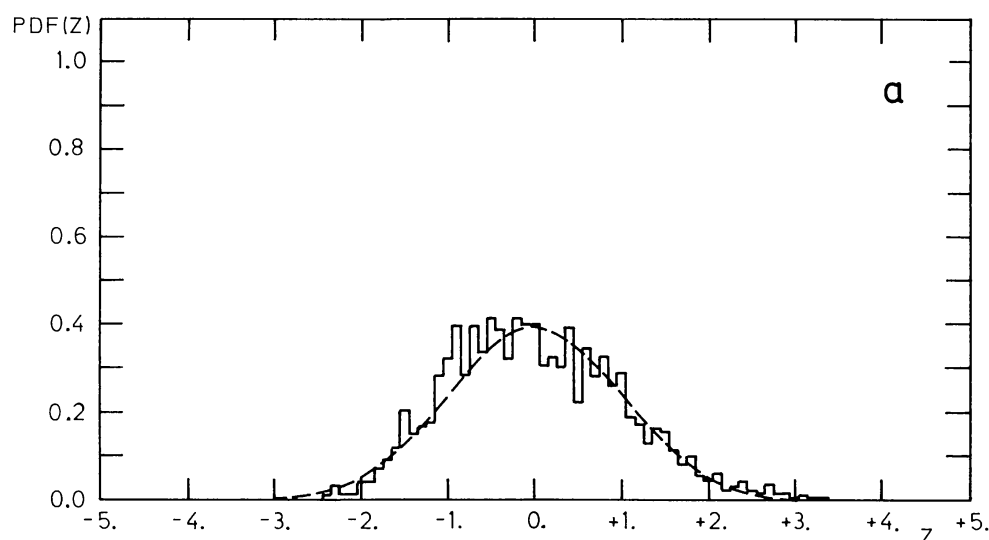

Fig. 8(a)–(c). Simulated probability density function of the Z statistic of the '4 within 16' randomization test under the null hypothesis for: (a) a quadrat size of $l/128 \times l/128$; (b) a quadrat size of $l/64 \times l/64$; (c) a quadrat size of $l/8 \times l/8$. The population consists in 200 individuals. The dotted curve is the theoretical normal probability density function (zero mean, standard deviation of 1).

PDF (Z)

b

PDF (Z)

c
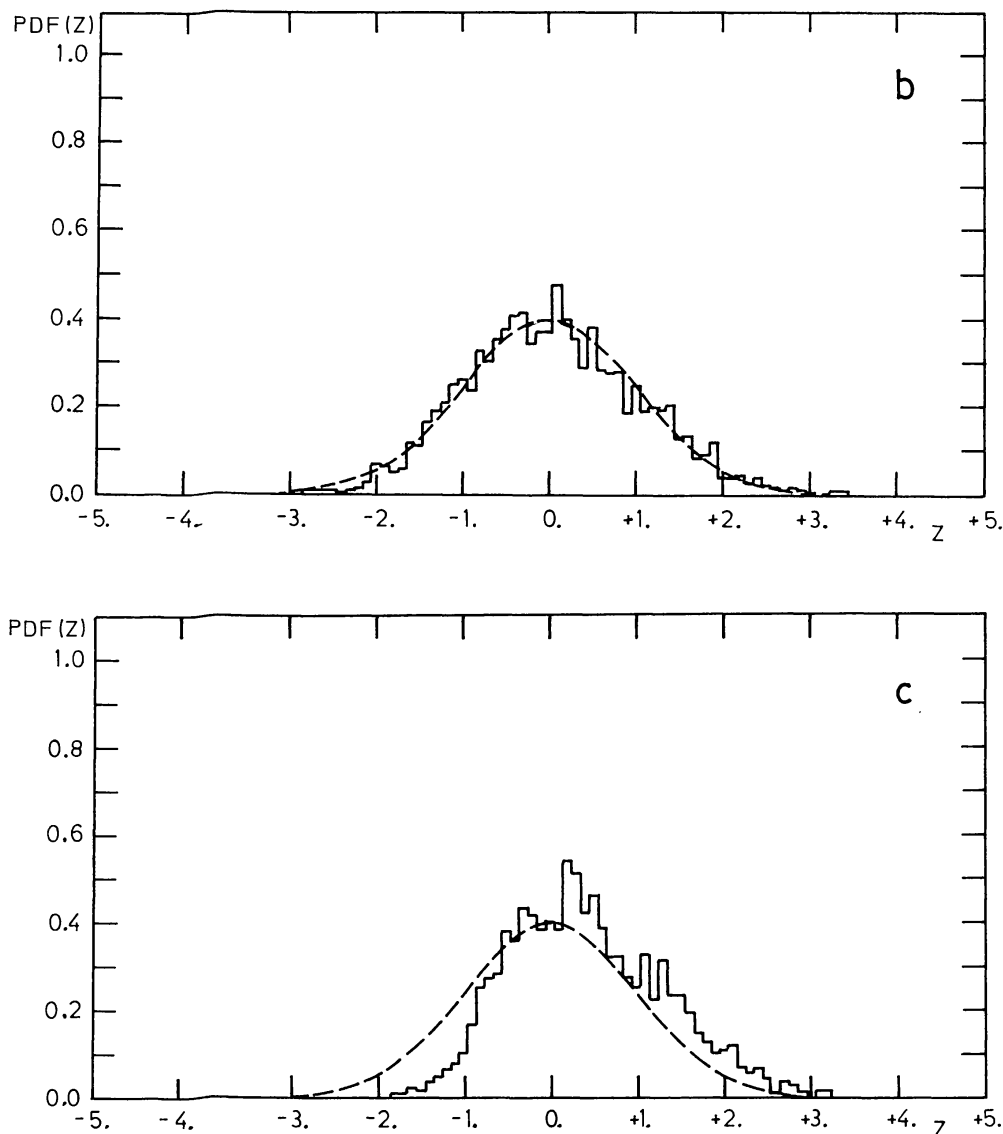
easily see the asymmetrical nature of the distribution around $Z = 0$. It is not a great problem and it is always possible to use an algorithm other than Equation (4.6). Second, deviations from a normal distribution occur when the expectation is too small in quadrats: then the θ statistic reveals its discrete nature. Although well marked, the effect is less dramatic than for the '2 within 4' randomization test.

4.2.2. *Under the Alternative Hypothesis* H_A

As in Section 3.2.2, let us first investigate the behaviour under the partial alternative hypothesis H_{AC}. The clusters are built up in the same way as previously except that they are two-dimensional. Figure 9 shows the run of $\langle Z \rangle$ against the number of quadrats on one line; the number of individuals is approximately 200. All the clusters have a dispersion of $l/16$. Figure 9(a), (b), (c), (d) refer, respectively, to 100 clusters of, on
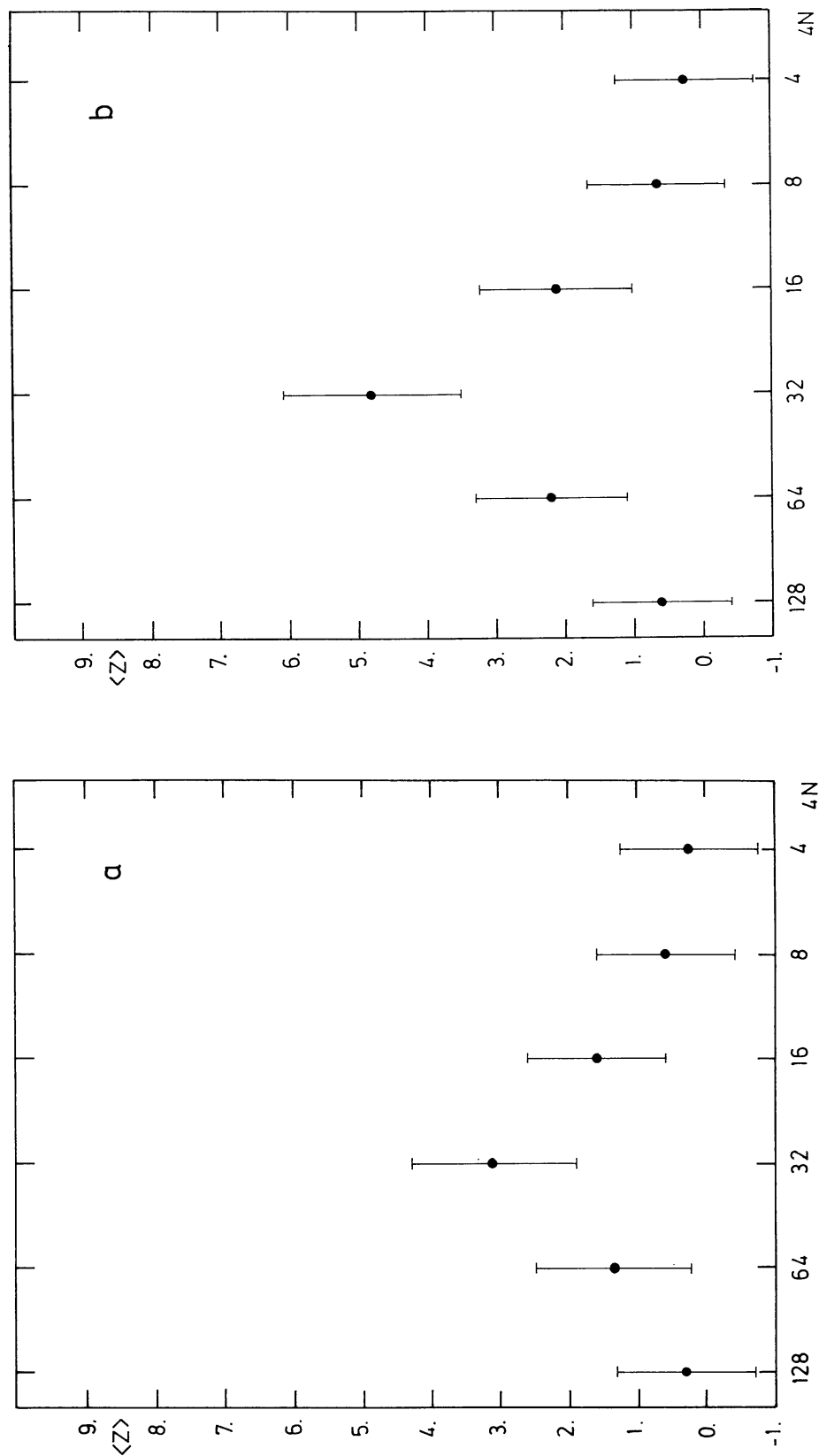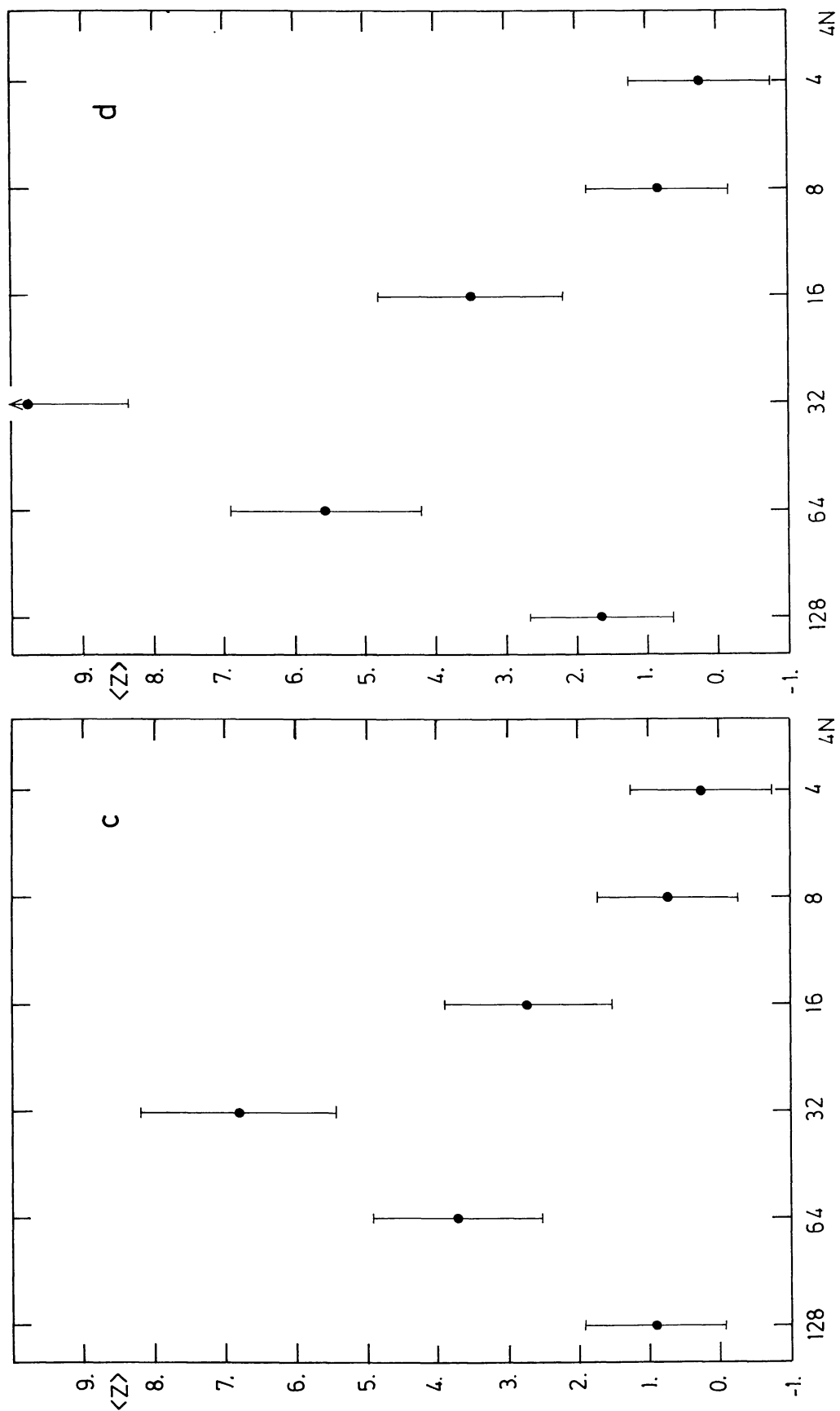
Fig. 9(a)–(d). Runs of the mean $\langle Z \rangle$ over the simulated distribution of the Z statistic of the '4 within 16' randomization test against the number of quadrats on one line of the investigating grid. Error bars are ± 1 standard deviation of Z over the simulated distribution. The characteristics of the distributions of the 200 individuals over the field are: (a) 100 clusters of, on average, two individuals; (b) 67 clusters of, on average, three individuals; (c) 40 clusters of, on average, five individuals; (d) 20 clusters of, on average, ten individuals. All the clusters have a maximal dispersion of $l/16$.
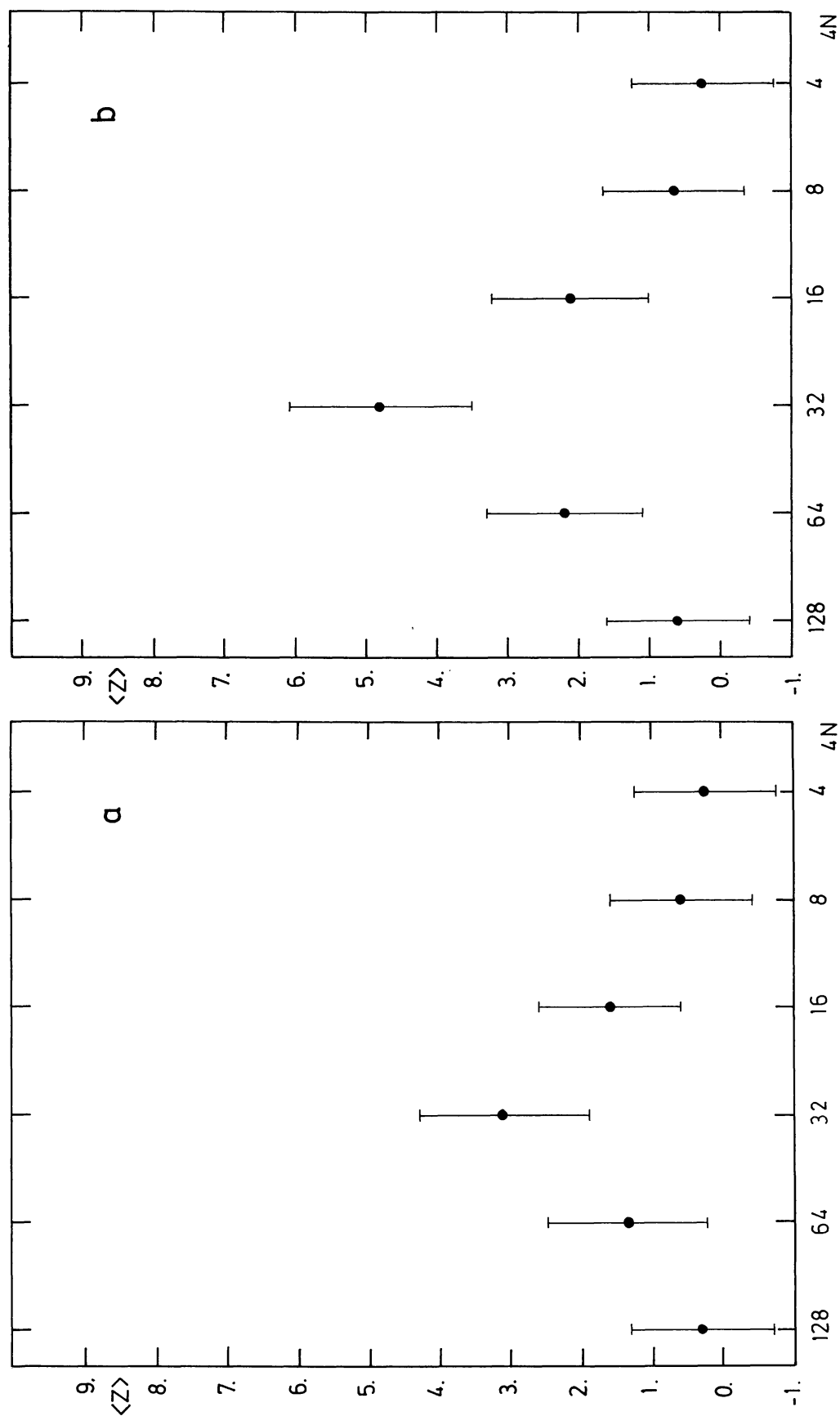
E. GOSSET AND B. LOUIS

Fig. 10(a)–(b). Runs of the mean $\langle Z \rangle$ over the simulated distribution of the Z statistic of the '4 within 16' randomization test against the number of quadrats on one line of the investigating grid. Error bars are ± 1 standard deviation of Z over the simulated distribution. The characteristics of the distributions of the 100 individuals over the field are: (a) 50 clusters of, on average, two individuals; (b) 20 clusters of, on average, five individuals. All the clusters have a maximal dispersion of $l/16$.

average, 2 individuals, to 67 clusters of, on average, 3 individuals, to 40 clusters of, on average, 5 individuals and to 20 clusters of, on average, 10 individuals. Figure 10 refers to a total number of individuals of about 100; in the case of part (a), they are arranged in 50 clusters of, on average, 2 individuals whereas for part (b) we have 20 clusters of, on average, 5 members. From those graphs, we can draw the following conclusions.

(a) The test reacts at the right place in each cases, i.e., at a square size of $l/32 \times l/32$ ($4N \times 4N = 32 \times 32$): the response is actually often quite large.

(b) At $4N \times 4N = 128 \times 128$, 8×8, and 4×4, the test is relatively robust; the Z statistic remains at almost insignificant values.

(c) At adjacent values of a significant one, there is a marked tendency to exhibit higher Z. It is again a remanence of the unavoidable two-population effect. We would like to point out that Shanks (1979) erroneously attributed this behaviour to 'an effect of testing for circular clustering with a necessarily square grid'.

(d) Contrary to the '2 within 4' randomization test, the '4 within 16' test is strongly sensitive to the number of members of the clusters (compare Figures 2, 9, and 10).
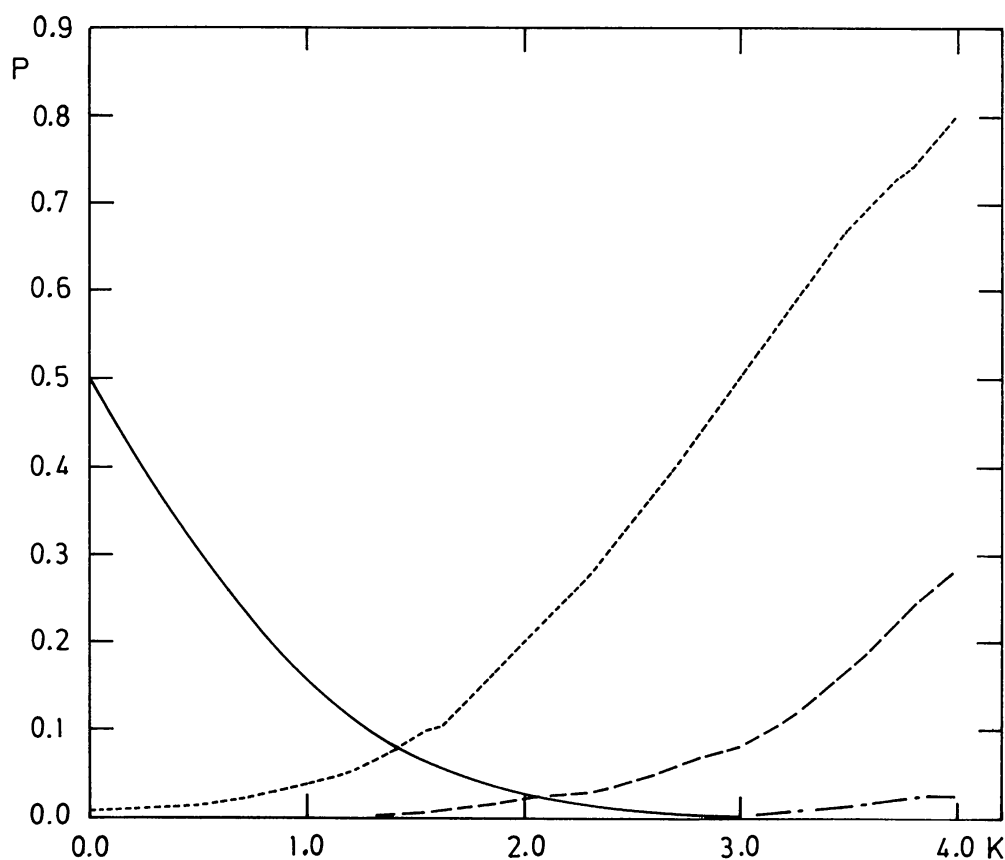

Fig. 11.   Probability $P_{H_0}(Z > K)$ as a function of the parameter $K \geq 0$ for the Z statistic of the '4 within 16' randomization test (continuous line). Three estimates of $P_{H_{AC}}(Z < K)$ are also represented. The three relevant simple hypotheses are: (1) 100 clusters of, on average, two members with a dispersion of $l/16$ (dotted line); (2) 67 clusters of, on average, three members with a dispersion of $l/16$ (dashed line); (3) 40 clusters of, on average, five members with a dispersion of $l/16$ (dashed-dotted line).

In summary, the test behaves well and seems to be more sensitive than the previous one. A concrete investigation confirms this statement. In Figure 11 are plotted $P_{H_0}(Z > K)$ as a function of $(K \geq 0)$ and $P_{H_{AC}}(Z < K)$ for different simple hypotheses (these are: 100 clusters of, on average, 2 members distributed with a dispersion $l/16$; 67 clusters of, on average, 3 members distributed with a dispersion $l/16$; 40 clusters of, on average, 5 members distributed with a dispersion $l/16$).

In Figure 12, we have plotted the same functions but for two different simple alternative hypotheses (these are: 50 clusters of, on average, 2 members and 20 clusters of, on average, 5 members).
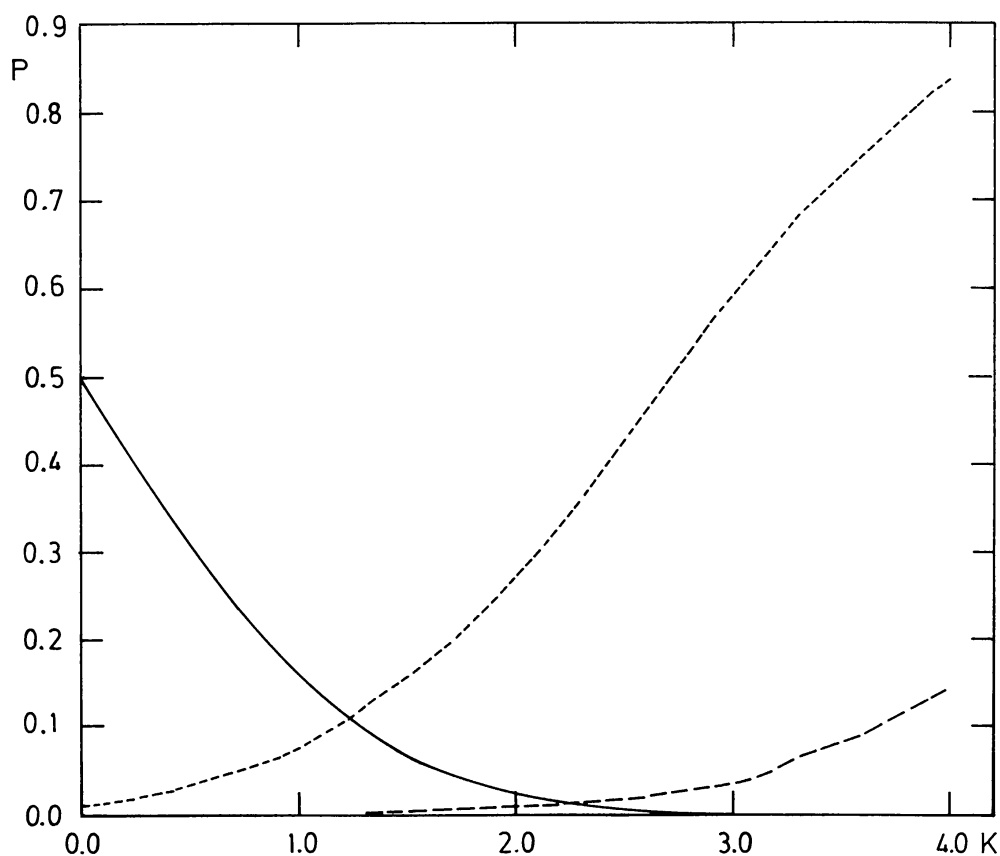

Fig. 12.   Same as Figure 11 but where two different estimates of $P_{H_{AC}}(Z < K)$ are considered. The two relevant simple hypotheses are: (a) 50 clusters of, on average, two members (dotted line); (b) 20 clusters of, on average, five members (dashed line).

Again α^+ and β^+ can be easily read off those curves. The power of the test is clearly quite good. As long as we have more than two members by cluster, the α^+ and β^+ can even both be made quite small. This result is very encouraging; nevertheless, it must be kept in mind that these situations are probably not very realistic.

We investigate next the behaviour under the partial alternative hypothesis H_{AR}. The regular distributions are built up in the same way as previously but are two-dimensional.
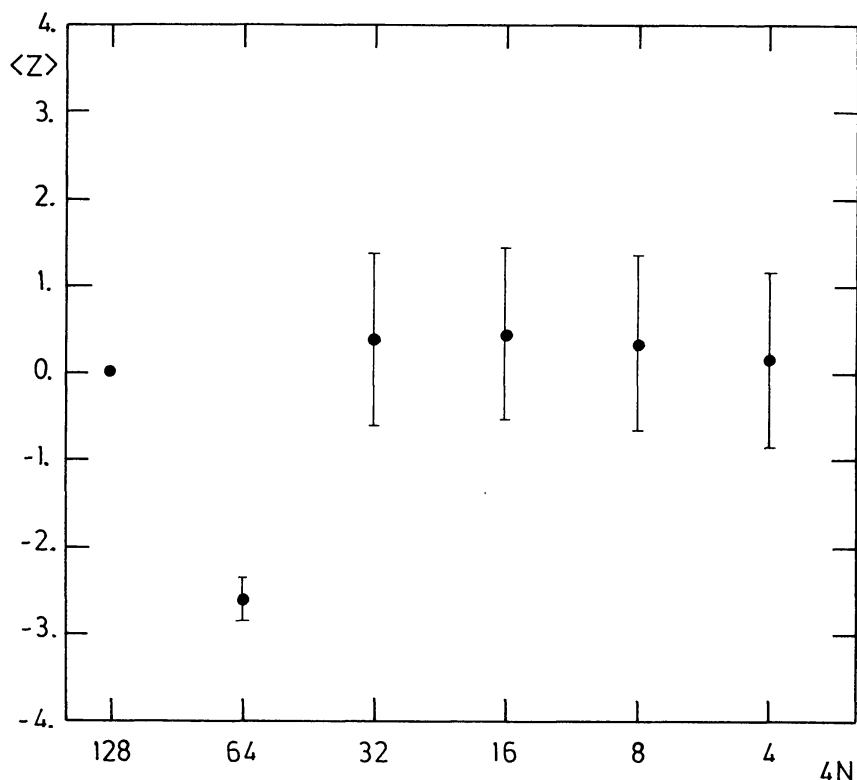
Fig. 13. Run of the mean $\langle Z \rangle$ over the simulated distribution of the Z statistic of the '4 within 16' randomization test against the number of quadrats on one line of the investigating grid. Error bars are ± 1 standard deviation of Z over the simulated distribution. The distributions of 150 individuals over the field are regular with a periodicity of $l/32$ and a random phasing.

An illustration is given in Figure 13 where one can see the run of $\langle Z \rangle$ against the number of quadrats on one line for a regular distribution of 150 individuals with a spatial periodicity of $l/32$. As long as the maximal dispersion of the positioning of an individual around a node is less than about half the periodicity, the '4 within 16' randomization test turns out to be sufficiently powerful. It is, however, sensitive to the mean number of individuals on any one node and to the relative kind of regularity over the two coordinates. Nevertheless, we tried also to better refine its power and to estimate the β^-. Figure 14 shows the obvious α^- and the two most extreme curves representing the β^-. Any realistic regular distribution will hopefully give a β^- within these two functions. One can see again that, without being perfect, it is at least possible to derive interesting results with this newly developed test as it presently stands.

5. The Three-Dimensional Binning Analysis

Encouraged by the behaviour of the two-dimensional test, we thought it would be interesting to consider the three-dimensional case. It is easy to derive the corresponding procedure; it is no more than a direct extension of the '4 within 16' randomization test. The name will, of course, be the '8 within 64' randomization test. If we consider a
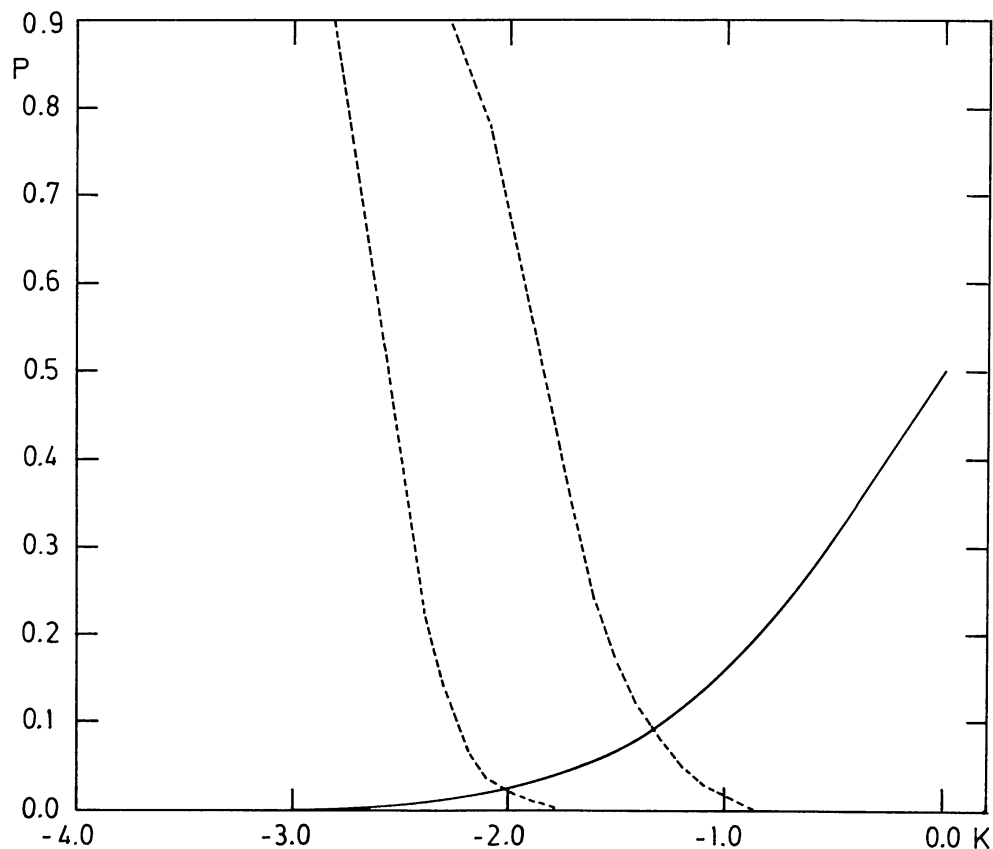
Fig. 14.   Probability $P_{H_0}(Z < K)$ as a function of the parameter $K \le 0$ for the Z statistic of the '4 within 16' randomization test (continuous line). Two extreme estimates of $P_{H_{AR}}(Z > K)$ are also represented (dotted line). See text for more details.

three-dimensional field partitioned in $64N^3$ cells, the relevant statistic is

$$\theta = \sum_{i=1}^{8N^3} n_i^2 - \frac{1}{8N^3} S_1^2 , \qquad (5.1)$$

where

$$n_i = \sum_{k=8i-7}^{k=8i} r_k$$

and

$$S_j = \sum_{i=1}^{64N^3} r_i^j .$$

Formulae equivalent to Equation (4.4) and Equation (4.5) are also needed. The expectation over the randomization process is given by

$$E(\theta) = \frac{(8N^3 - 1)}{8N^3(64N^3 - 1)} [S_0 S_2 - S_1^2] ; \qquad (5.2)$$

and the variance is given by

$$\text{var}(\theta) = \frac{56(8N^3 - 1)}{(64N^3 - 1)^2 (64N^3 - 3)(32N^3 - 1)} [S_0 S_2 - S_1^2]^2 +$$

$$+ \frac{56(8N^3 - 1)}{(64N^3 - 1)(64N^3 - 3)(32N^3 - 1)} [4S_3 S_1 - 3S_2^2 - S_0 S_4].$$

$$(5.3)$$

The remaining part of the procedure is the same as in Section 4. The relevance and the results of the present test may be expected to be as good as those for the '4 within 16' randomization test, and actually even better since the randomization distribution is larger still (about 4.5×10^{47}).

6. Applications

At this stage, it is interesting to apply both the '2 within 4' and the '4 within 16' randomization tests to concrete cases in order to demonstrate their utility. Our sole aim is to illustrate the subject.

6.1. APPLICATION 1

The first data set is one-dimensional and consists of the redshifts of 58 confirmed quasars detected in a UV-excess survey performed in the field of NGC 520 by Swings, Surdej, and Gosset (to be published). Due to the method of selection of the candidates, the redshifts range from $z = 0$ to $z = 2.3$ and the repartition is clearly non-uniform. The distribution of redshifts has been investigated using the '2 within 4' randomization test on the field of length $l = 2.3$ and with random phasing. The result is shown in Figure 15.
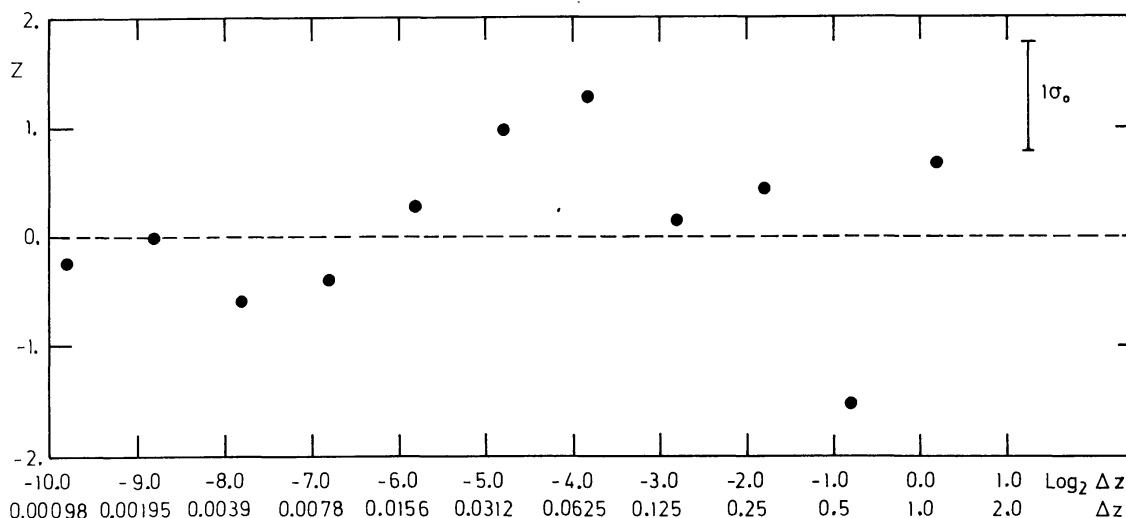

Fig. 15. Analysis of the distribution of quasar redshifts in the field of NGC 520: run of the Z statistic of the one-dimensional '2 within 4' randomization test against the logarithm in base 2 of the investigated redshift scale; the relevant scale is also given.

There are three important deviations. The first is a clustering at a scale of $l/2$, i.e., $\Delta z \simeq 1.15$. The second denotes a regularity at a scale of $l/4$, i.e., $\Delta z \simeq 0.6$. Both are due to the presence of peaks caused by the emission lines of different quasars coming into the photometric bands. Generated peaks lie at $z \simeq 0.3$ (Mg II), $z \simeq 1.4$ (Si IV) and $z \approx 2.0$ (Lα). The effect is the same as that detected by Box and Roeder (1984) at a periodicity of $\Delta z = 0.8625$ (see, for example, the peak at $N = 4$ in their Figure 5) for the quasars in the catalogue of Hewitt and Burbidge. The third deviation is at a scale of $l/32$ (with a satellite at $l/64$), i.e., $\Delta z \simeq 0.07$. The positive value indicates a clustering. The relevant significance level is 0.10; this value is not conclusive but is sufficiently large to be of some interest. It is to be associated mainly with a concentration of quasars around $z \simeq 0.95$. This peculiarity has already been reported for the nearby field around NGC 450 (Swings *et al.*, 1983, 1985).

A complete and detailed statistical analysis will be published elsewhere. It is quite interesting to note that the large-scale anomaly (which is an observational one) does not prevent us from detecting the smallest one, and this without any preliminary assumption.

6.2. APPLICATION 2

More than a decade ago, Gouguenheim (1969) published a list of galaxy redshifts $v_{\rm HI}$ measured from H I 21 cm lines. Most of the observed galaxies had optical redshifts $v_{\rm opt}$ known from the literature. In Figure 2 of her paper, Gouguenheim (1969) shows an histogram of $v_{\rm HI} - v_{\rm opt}$ which, according to her, suggests that the differences must be distributed almost normally with a standard deviation of 60 km s^{-1}, in agreement with the two estimated errors of 35 and 50 km s^{-1}, respectively, on the radio and optical redshifts. We apply here the '2 within 4' randomization test to the distribution of these differences. The investigated field is -220 km s^{-1} to $+220$ km s^{-1}. The run of Z against the characteristic scale is shown in Figure 16. Z is near to 0 for $\Delta(v_{\rm HI} - v_{\rm opt}) = 220$ km s^{-1}, confirming the symmetry of the distribution. The next point ($\Delta(v_{\rm HI} - v_{\rm opt}) = 110$ km s^{-1}) denotes a marked clustering; this is the gaussian feature. All other values are not significant, except one: for $\Delta(v_{\rm HI} - v_{\rm opt}) = 55$ km s^{-1}, Z equals -1.92, denoting a regularity at a significance level of 0.0275. At first sight, this is quite surprising. Some *a posteriori* refinements indicate that the period of the regularity is probably closer to a higher value, perhaps reaching 70 km s^{-1}. This value is only approximate but it is to be pointed out that such a regularity is readily visible in Figure 2 of Gouguenheim (1969). Several bins are overpopulated with respect to the theoretical expectations, mainly near -200, -150, -60, 0, $+50$, $+150$, $+200$ km s^{-1}. Of course, the spectral resolution element of the radio telescope used in deriving the redshifts is 59 km s^{-1} wide, which could explain the regularity. It is tempting to compare this result with the well-known conclusions of Tifft (1982a, b) who claims the detection of a quantization of the redshifts. However, such an effect, if it exists, should be blurred when making the difference between optical and radio redshifts as long as the errors on the two are of the same order; this is clearly the case for the data set of Gouguenheim (1969). This may be due to an instrumental effect although a definite answer is beyond the scope of this paper. A detailed statistical analysis of the data of Arp and Sulentic (1985) as well as of Tifft (1982a, b) will be published subsequently.
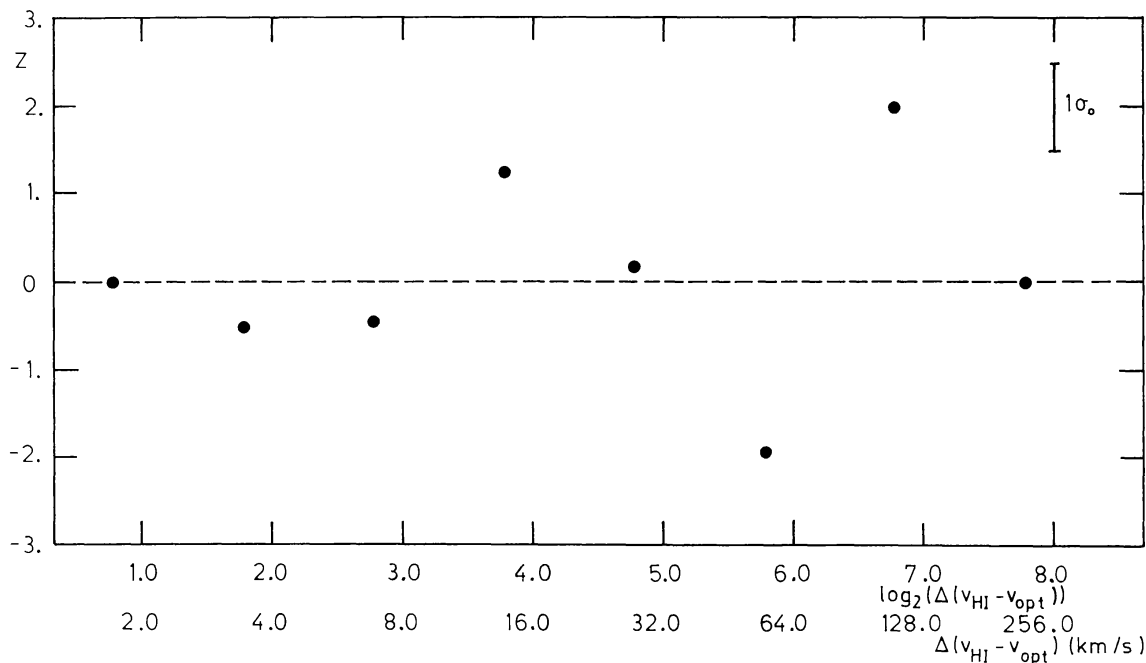
Fig. 16. Analysis of the distribution of the differences between optical and radio redshifts of the sample of galaxies of Gouguenheim (1969): run of the Z statistic of the one-dimensional '2 within 4' randomization test against the logarithm in base 2 of the investigated velocity scale (expressed in km s^{-1}), the relevant scale is also given.

6.3. APPLICATION 3

Let us now investigate a two-dimensional field. We shall take the 71 detected quasars as reported in the first version of the He *et al.* (1984) survey. The $5° \times 5°$ field has already been analysed by Peacock (1983) using the two-dimensional Kolmogorov–Smirnov test and non-uniform (or generalized) power spectrum analysis. The run of the Z statistic of the '4 within 16' randomization test is shown in Figure 17. Let us analyse the two largest deviations from randomness. At $l/4$ – i.e., $\Delta\theta = 1°25$ and $\log_2 \Delta\theta = 0.32$ – the Z reaches a value of 2.25 at a significance level of 0.0122. We rediscover the non-uniformity previously reported by Peacock (1983) and tested through the two-dimensional Kolmogorov–Smirnov method. The relevant significance level was 0.027, which is the same order of magnitude as our value. The other deviation is again towards a clustering and occurs at $l/32$, i.e., $\Delta\theta = 0°16$ and $\log_2 \Delta\theta = -2.68$ or an angular scale of about 10 arc min. In this case, Z equals to 1.24, leading to a significance level of 0.11. Again, this is not conclusive but it is of some importance in the framework of this paper.

Such a feature is actually not surprising in a visual survey since, when one candidate is discovered, it is tempting to search for a second. We are nevertheless interested in this small-scale clustering more especially for its existence in the data than for its reality as a significant physical feature. So, we have also computed the two-point autocorrelation function of this field of quasars. The result is visible in Figure 18. The statistic
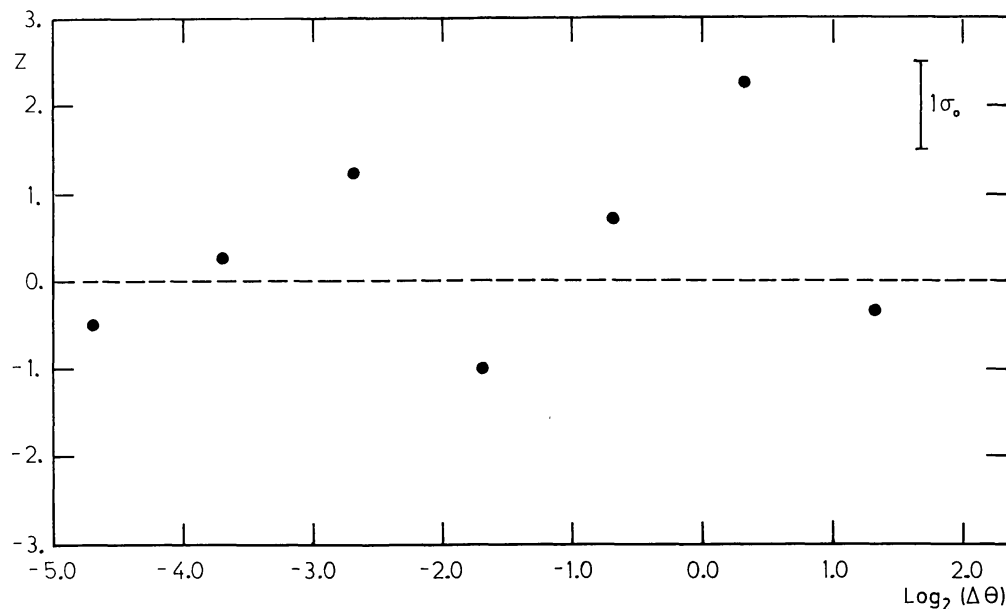
Fig. 17.   Analysis of the distribution of the two-dimensional positions of the quasars in a field near the Virgo cluster. The data are taken from the first version of the survey of He *et al.* (1984) as published by Peacock (1983). The figure represents the run of the Z statistic of the '4 within 16' randomization test against the logarithm in base 2 of the investigated angular scale $\Delta\theta$, where θ is expressed in degrees.
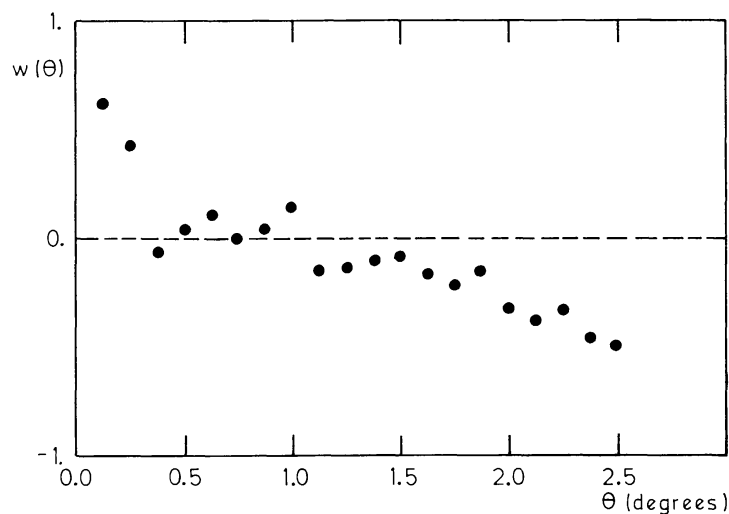

Fig. 18.   Investigation of the same field as for Figure 17 but with the two point autocorrelation function analysis: run of the $w(\theta)$ statistic against the angular separation θ expressed in degrees.

$w(\theta)$ (Fall, 1979) clearly exhibits discrepant values for θ of less than 15 arc min! All this confirms the power of the new test and its inter-scale robustness.

7. Conclusions

In this paper, we have described, discussed and analysed the best known significance tests usually associated with the binning analysis (BA). We have shown that better ones

can be obtained when using randomization processes. We have studied their behaviour. For one-dimensional BAs, we used the '2 within 4' randomization test almost in its original version as proposed by ecologists (Mead, 1974). For two- and three-dimensional BAs, we have suggested new randomization tests: respectively, the '4 within 16' and the '8 within 64'. The methods have been explained and their power tentatively investigated for some practical situations.

These latter randomization tests have certain advantages.

(a) They have an inter-scale robustness: the value of the statistic Z is quasi-independent of the features present at smaller or higher scales and this without the necessity of a model for these scales (which differs from the Generalized PSA where a probability density function is needed).

(b) They can discern between the two kinds of deviation from randomness: a positive value of Z indicates a clustering; a negative one a regularity. For example, the χ^2 test is unable to recognize such a type of tendency.

(c) They are able to detect clusterings at different scales even when the clusters are densely piled on top of each other. (The PSA also enables this.)

(d) They are easy and quite fast to apply with any kind of computer and a well-conceived algorithm.

(e) Due to the inter-scale robustness, it is possible to compute the Z statistic at scales other than $l/2^n$ ($n = 1, 2, 3, 4, 5, \ldots$): for example, $(l + \Delta l)/2^n$.

Their main disadvantage is the discrete nature of the investigating grid. We do not wish to present these tests as the universal panacea, but we nevertheless think they are sufficiently efficient to take their place beside the others (PSA, CFA, ...) and that they ought to be used jointly with them.

Acknowledgements

Appendix A

In the following, we consider the statistic

$$\theta = v/m \,,$$

which usually is assumed to be distributed under the null hypothesis of Section 1.1 with an expectation

$$E(\theta) = 1$$

and a variance

$$\mathrm{var}(\theta) = \frac{2N}{(N-1)^2} \,,$$

or

$$\mathrm{var}(\theta) = \frac{2}{N-1} \,,$$

depending on the authors (Greig-Smith, 1952). At first, it should be noted that the counts are Poissonian only by approximation. In fact, if one decides to be rigorous, the probability law to be used is at least the multinomial one. In this case, the statistic

$$\theta = v/m = \frac{1}{n} \sum_{i=1}^{N} \left(r_i - \frac{n}{N} \right)^2 ,$$

is distributed with an expectation

$$E(\theta) = \frac{N-1}{N} \tag{A.1}$$

and a variance

$$\mathrm{var}(\theta) = \frac{2(N-1)^2}{N^3} - \frac{2N-3}{nN^2} \,. \tag{A.2}$$

If N, the number of cells, is large with respect to 1, we shall have

$$E(\theta) \sim 1 \,, \qquad \mathrm{var}(\theta) \sim \frac{2}{N} \left(1 - \frac{1}{n} \right)$$

and in order to recover Equation (2.2), one still has to assume that n, the number of individuals, is also large. It is clear that both assumptions are not always true and we strongly suggest that use be made of the more rigorous formulae (A.1) and (A.2) given above.

Appendix B

We consider $16N^2$ squares. The count in the ith quadrat is r_i and the expression of the statistic is

$$\theta = \frac{1}{4N^2} \left[4N^2 \sum_{i=1}^{4N^2} n_i^2 - \left(\sum_{i=1}^{16N^2} r_i \right)^2 \right] \tag{B.1}$$

with

$$n_i = \sum_{k=4i-3}^{k=4i} r_k \, .$$

The randomization process is performed by permutations of the r_i. There are $16N^2!$ possible permutations. We can symbolize one of them by the function $p(i)$ and write $r_{p(i)}$ for the permuted counts. The statistic is then

$$\theta_p = \frac{1}{4N^2} \left[4N^2 \sum_{i=1}^{4N^2} n_i^2 - \left(\sum_{i=1}^{16N^2} r_{p(i)} \right)^2 \right] \qquad (B.2)$$

with

$$n_i = \sum_{k=4i-3}^{k=4i} r_{p(k)} \, .$$

However, for the sake of simplicity, we shall drop this heavy notation but keep it implicitly. The expectation over the randomization process is given by

$$E(\theta) = \sum_p \frac{1}{16N^2!} \, \theta_p \, ,$$

where the summation is extended to all the $16N^2!$ possible permutations.

$$E(\theta) = \frac{1}{16N^2!} \frac{1}{4N^2} \left[4N^2 \sum_p \sum_{i=1}^{16N^2} r_i^2 + \right.$$

$$+ 4N^2 \times 2 \sum_p \sum_{i=1}^{4N^2} (r_{4i-3} r_{4i-2} + r_{4i-3} r_{4i-1} + r_{4i-3} r_{4i} +$$

$$\left. + r_{4i-2} r_{4i-1} + r_{4i-2} r_{4i} + r_{4i-1} r_{4i}) - \sum_p \left(\sum_{i=1}^{16N^2} r_i \right)^2 \right] \, .$$

It should be noted that, in the first and third terms, the internal sum is invariant with respect to the permutations. Let us consider the second term. We have six subterms of the type

$$\sum_p \sum_{i=1}^{4N^2} r_{f(i)} r_{g(i)} \, ,$$

where $f(i)$ and $g(i)$ have four permitted values, namely $4i - 3$, $4i - 2$, $4i - 1$, $4i$. The effect of the summation over all the possible permutations is to remove the restrictions imposed by $f(i)$ and $g(i)$. So, we have $16N^2! \times 4N^2 \times 6$ terms.

Of course, only

$$C_{16N^2}^2 = \frac{16N^2(16N^2 - 1)}{2}$$

of them are different. Consequently, the number of times that each appears is

$$\frac{16N^2! \times 4N^2 \times 6 \times 2}{16N^2(16N^2 - 1)} = \frac{16N^2! \times 3}{(16N^2 - 1)} \ .$$

We can recreate them from a double sum. Then, the expression of the second term becomes

$$\frac{16N^2! \times 3}{(16N^2 - 1)} \ 4N^2 \sum_{\substack{i = 1 \\ i \neq k}}^{16N^2} \sum_{k = 1}^{16N^2} r_i r_k \ .$$

A factor of 2 is included in the double sum. So, we have

$$E(\theta) = \frac{1}{4N^2} \left[4N^2 \sum_{i = 1}^{16N^2} r_i^2 + \frac{3 \times 4N^2}{(16N^2 - 1)} \sum_{\substack{i = 1 \\ i \neq k}}^{16N^2} \sum_{k = 1}^{16N^2} r_i r_k - \sum_{i = 1}^{16N^2} \sum_{k = 1}^{16N^2} r_i r_k \right] =$$

$$= \frac{1}{4N^2} \left[4N^2 \left(1 - \frac{3}{16N^2 - 1} \right) \sum_{i = 1}^{16N^2} r_i^2 - \left(1 - \frac{3 \times 4N^2}{16N^2 - 1} \right) \sum_{i = 1}^{16N^2} \sum_{k = 1}^{16N^2} r_i r_k \right] =$$

$$= \frac{(4N^2 - 1)}{4N^2(16N^2 - 1)} \left[16N^2 \sum_{i = 1}^{16N^2} r_i^2 - \sum_{i = 1}^{16N^2} \sum_{k = 1}^{16N^2} r_i r_k \right] =$$

$$= \frac{(4N^2 - 1)}{4N^2(16N^2 - 1)} \ [S_0 S_2 - S_1^2] \ . \tag{B.3}$$

The variance can be computed from

$$\mathrm{var}(\theta) = E(\theta^2) - (E(\theta))^2 \ . \tag{B.4}$$

We still have to compute $E(\theta^2)$. We find

$$E(\theta^2) = \frac{1}{16N^2!} \sum_p \theta_p^2 =$$

$$= \frac{1}{16N^2!} \sum_p \frac{1}{(4N^2)^2} \left[(4N^2)^2 \left(\sum_{i = 1}^{4N^2} n_i^2 \right)^2 - \right.$$

$$\left. - 2.4N^2 \sum_{i = 1}^{4N^2} n_i^2 \left(\sum_{i = 1}^{16N^2} r_i \right)^2 + \left(\sum_{i = 1}^{16N^2} r_i \right)^4 \right] \ . \tag{B.5}$$

Concerning the third term, we obviously have

$$E \left[\frac{1}{(4N^2)^2} \left(\sum_{i = 1}^{16N^2} r_i \right)^4 \right] = \frac{1}{16N^2!} \sum_p \frac{1}{(4N^2)^2} \left(\sum_{i = 1}^{16N^2} r_i \right)^4 = \frac{1}{(4N^2)^2} \ S_1^4 \ . \tag{B.6}$$

For the second term of B.5, using the same method as for the computation of $E(\theta)$, we have

$$E\left[-\frac{2}{4N^2}\left(\sum_{i=1}^{16N^2} r_i\right)^2 \sum_{i=1}^{4N^2} n_i^2\right] =$$

$$= -\frac{2}{4N^2} S_1^2 \left(S_2 + \frac{3}{16N^2 - 1} S_1^2 - \frac{3}{16N^2 - 1} S_2\right) =$$

$$= -\frac{2}{4N^2(16N^2 - 1)} (4(4N^2 - 1)S_1^2 S_2 + 3S_1^4). \qquad (B.7)$$

The first term of B.5 remains

$$E\left(\sum_{i=1}^{4N^2} \sum_{k=1}^{4N^2} n_i^2 n_k^2\right).$$

If we develop the argument of the expectation, we obtain $4N^2 \times 4N^2 \times 100$ terms. Out of them, $4N^2 \times 4N^2 \times 16$ are of the form

$$r_{f(i)}^2 r_{g(k)}^2$$

and through the summation over all the permutations, they give

$$\sum_{i=1}^{16N^2} \sum_{k=1}^{16N^2} r_i^2 r_k^2 = S_2^2. \qquad (B.8)$$

From the remaining terms, $4N^2 \times 4N^2 \times 48$ are of the form

$$2r_{f(i)}^2 r_{g(k)} r_{h(k)}.$$

Obviously, $r_{f(i)}^2$ will give an S_2 term whereas $r_{g(k)} r_{h(k)}$ will give $(S_1^2 - S_2)$. We obtain

$$\frac{2 \times 3}{16N^2 - 1} [S_2 S_1^2 - S_2^2]. \qquad (B.9)$$

There remain $4N^2 \times 4N^2 \times 36$ terms of the form

$$4r_{f(i)} r_{g(i)} r_{h(k)} r_{t(k)}.$$

If $i \neq k$, all are different and will be taken into account later. If $i = k$, we have $4N^2 \times 36$ terms, of which $4N^2 \times 6$ are of the form

$$4r_{f(i)}^2 r_{g(i)}^2$$

and they give

$$\frac{2 \times 3}{16N^2 - 1} [S_2^2 - S_4]. \qquad (B.10)$$

We also have $4N^2 \times 24$ terms of the form

$$4r_{f(i)}^2 r_{g(i)} r_{h(i)} \, .$$

With the permutations, we have $16N^2! \times 4N^2 \times 24$ terms and only

$$C_{16N^2}^3 \, \frac{3!}{2!}$$

of them are different. This means

$$16N^2(16N^2 - 1)(8N^2 - 1) \, .$$

Consequently, the number of times each term appears is

$$\frac{16N^2! \times 4N^2 \times 24}{16N^2(16N^2 - 1)(8N^2 - 1)} \, .$$

So, we have

$$\frac{4}{2} \, \frac{4N^2 \times 24}{16N^2(16N^2 - 1)(8N^2 - 1)} \sum_{\substack{i=1 \\ i \neq k \\ i \neq l}}^{16N^2} \sum_{\substack{k=1 \\ k \neq l}}^{16N^2} \sum_{l=1}^{16N^2} r_i^2 r_k r_l \, .$$

The 2 in the denominator comes from a correction for a redundancy in the triple sum. Finally,

$$\frac{12}{(16N^2 - 1)(8N^2 - 1)} \, [S_2 S_1^2 - S_2^2 - 2S_3 S_1 + 2S_4] \, . \tag{B.11}$$

There remain $4N^2 \times 6$ terms of the form

$$4r_{f(i)} r_{g(i)} r_{h(i)} r_{t(i)} \, .$$

It should be noted that, as only four different values exist for the functions f, g, h, t, all these terms are necessarily identical. We have $16N^2! \times 4N^2 \times 6$ terms of which only $C_{16N^2}^4$ are different. Each one appears

$$\frac{16N^2! \times 4N^2 \times 6 \times 4 \times 3 \times 2}{16N^2(16N^2 - 1)(16N^2 - 2)(16N^2 - 3)}$$

times and we obtain, temporarily,

$$\frac{18 \times 4}{(16N^2 - 1)(16N^2 - 3)(8N^2 - 1) \times 24} \sum_{\substack{i=1 \\ i \neq k \\ i \neq l \\ i \neq m}}^{16N^2} \sum_{\substack{k=1 \\ k \neq l \\ k \neq m}}^{16N^2} \sum_{\substack{l=1 \\ l \neq m}}^{16N^2} \sum_{m=1}^{16N^2} r_i r_k r_l r_m \, .$$

$$\tag{B.12}$$

The 24 in the denominator comes from a correction for a redundancy in the quadruple sum.

We still have to take into account the $4N^2 \times (4N^2 - 1) \times 36$ terms with $i \neq k$. In this case, we obtain $16N^2! \times 4N^2 \times (4N^2 - 1) \times 36$ terms, of which only $C^4_{16N^2}$ are different. Thus,

$$\frac{18(4N^2 - 1)}{(16N^2 - 1)(16N^2 - 3)(8N^2 - 1)} \sum_{\substack{i=1 \\ i \neq k \\ i \neq l \\ i \neq m}}^{16N^2} \sum_{\substack{k=1 \\ k \neq l \\ k \neq m}}^{16N^2} \sum_{\substack{l=1 \\ l \neq m}}^{16N^2} \sum_{m=1}^{16N^2} r_i r_k r_l r_m. \qquad (B.13)$$

The quadruple sum involved in Equations (B.12) and (B.13) can be expressed in terms of the S_j.

$$\sum_{\substack{i=1 \\ i \neq k \\ i \neq l \\ i \neq m}}^{16N^2} \sum_{\substack{k=1 \\ k \neq l \\ k \neq m}}^{16N^2} \sum_{\substack{l=1 \\ l \neq m}}^{16N^2} \sum_{m=1}^{16N^2} r_i r_k r_l r_m =$$

$$= \sum_{i=1}^{16N^2} \sum_{k=1}^{16N^2} \sum_{l=1}^{16N^2} \sum_{m=1}^{16N^2} r_i r_k r_l r_m - 6 \sum_{\substack{i=1 \\ i \neq k \\ i \neq l}}^{16N^2} \sum_{\substack{k=1 \\ k \neq l}}^{16N^2} \sum_{l=1}^{16N^2} r_i^2 r_k r_l -$$

$$- 3 \sum_{\substack{i=1 \\ i \neq k}}^{16N^2} \sum_{k=1}^{16N^2} r_i^2 r_k^2 - 4 \sum_{\substack{i=1 \\ i \neq k}}^{16N^2} \sum_{k=1}^{16N^2} r_i^3 r_k - \sum_{i=1}^{16N^2} r_i^4 =$$

$$= S_1^4 - 6S_2 S_1^2 + 6 \sum_{\substack{i=1 \\ i \neq k}}^{16N^2} \sum_{k=1}^{16N^2} r_i^2 r_k^2 +$$

$$+ 2.6 \sum_{\substack{i=1 \\ i \neq k}}^{16N^2} \sum_{k=1}^{16N^2} r_i^3 r_k + 6 \sum_{i=1}^{16N^2} r_i^4 - 3S_2^2 +$$

$$+ 3 \sum_{i=1}^{16N^2} r_i^4 - 4S_3 S_1 + 4 \sum_{i=1}^{16N^2} r_i^4 - S_4 =$$

$$= S_1^4 - 6S_2 S_1^2 + 6S_2^2 - 6S_4 + 12S_3 S_1 - 12S_4 + 6S_4 - 3S_2^2 +$$

$$+ 3S_4 - 4S_3 S_1 + 4S_4 - S_4 =$$

$$= S_1^4 + 3S_2^2 - 6S_2 S_1^2 + 8S_3 S_1 - 6S_4. \qquad (B.14)$$

Combining (B.12), (B.13), and (B.14), we obtain

$$\frac{18(4N^2 - 1) + 3}{(16N^2 - 1)(16N^2 - 3)(8N^2 - 1)} [S_1^4 + 3S_2^2 - 6S_2 S_1^2 + 8S_3 S_1 - 6S_4].$$

$$(B.15)$$

Now, we have to combine (B.6), (B.7), (B.8), (B.9), (B.10), (B.11), (B.15), and the expression of $-(E(\theta))^2$ as deduced from (B.3). Through a long and tedious calculation, we may group those terms in two large entities. The first consists of the terms in $S_2 S_1^2$, S_1^4 and partly S_2^2. The second will contain terms in $S_3 S_1$, S_4, and the remaining S_2^2. The

first group gives

$$\frac{12(4N^2 - 1)}{(16N^2 - 1)^2 (16N^2 - 3)(8N^2 - 1)} [S_0 S_2 - S_1^2]^2 \tag{B.16}$$

and the second gives

$$\frac{12(4N^2 - 1)}{(16N^2 - 1)(16N^2 - 3)(8N^2 - 1)} [4S_3 S_1 - 3S_2^2 - S_0 S_4] . \tag{B.17}$$

References

Arp, H. and Sulentic, J. W.: 1985, *Astrophys. J.* **291**, 88.

Blackman, G. E.: 1935, *Ann. Bot.* **49**, 749.

Bok, B. J.: 1934, *Harvard Obs. Bull.*, No. 895, p. 1.

Box, T. C. and Roeder, R. C.: 1984, *Astron. Astrophys.* **134**, 234.

Clapham, A. R.: 1936, *J. Ecol.* **24**, 232.

Cochran, W. G.: 1952, *Ann. Math. Statistics* **23**, 315.

De Vaucouleurs, G.: 1971, *Publ. Astron. Soc. Pacific* **83**, 113.

Evans, F. C.: 1952, *Contr. Lab. Vertebr. Biol. Univ. Mich.* **54**, 1.

Fall, S. M.: 1979, *Rev. Mod. Phys.* **51**, 21.

Fisher, R. A.: 1947, *The Design of Experiments*, Oliver and Boyd, Edinburgh.

Gouguenheim, L.: 1969, *Astron. Astrophys.* **3**, 281.

Greig-Smith, P.: 1952, *Ann. Bot. N.S.* **16**, 293.

He, X. T., Cannon, R. D., Peacock, J. A., Smith, M. G., and Oke, J. B.: 1984, *Monthly Notices Roy. Astron. Soc.* **211**, 443.

Katz, L. and Mulders, G. F. W.: 1942, *Astrophys. J.* **95**, 565.

Kendall, M. G. and Stuart, A.: 1967, *The Advanced Theory of Statistics*, Vol. 2, C. Griffin and Co., London.

Lindgren, B. W.: 1976, *Statistical Theory*, Collier Macmillan Publ. Co., New York.

Louis, B.: 1984, 'Etude stochastique d'un nuage de points', Thesis, Université de Liège.

Mead, R.: 1974, *Biometrics* **30**, 295.

Moore, P. G.: 1953, *Ann. Bot. N.S.* **17**, 57.

Morisita, M.: 1959, *Mem. Fac. Sci. Kyushu Univ. Ser. E* **2**, 215.

Neyman, J., Scott, E. L., and Shane, C. D.: 1954, *Astrophys. J. Suppl.* **8**, 1, 269.

Osmer, P. S.: 1981, *Astrophys. J.* **247**, 762.

Peacock, J. A.: 1983, *Monthly Notices Roy. Astron. Soc.* **202**, 615.

Peebles, P. J. E.: 1978, in M. S. Longair and J. Einasto (eds.), *The Large-Scale Structure of the Universe*, p. 217.

Rees, M. J.: 1982, in H. A. Brück, G. V. Coyne, and M. S. Longair (eds.), *Astrophysical Cosmology*, p. 3.

Shanks, T.: 1979, *Monthly Notices Roy. Astron. Soc.* **186**, 583.

Simpson, E. H.: 1949, *Nature* **163**, 688.

Swings, J. P., Arp, H., Surdej, J., Henry, A., and Gosset, E.: 1983, in 'Quasars and Gravitational Lenses', *24th Liège Astrophysical Colloq.*, p. 37.

Swings, J. P., Surdej, J., Henry, A., Gosset, E., and Arp, H.: 1985, *Rev. Mexicana Astron. Astrofis.* **10**, 91 (Proceedings of the Latin-American IAU Regional Meeting, No. 3, Buenos Aires, 1983).

Tifft, W. G.: 1982a, *Astrophys. J.* **257**, 442.

Tifft, W. G.: 1982b, *Astrophys. J.* **262**, 44.

Zwicky, F.: 1952, *Publ. Astron. Soc. Pacific* **64**, 247.

Zwicky, F.: 1953, *Helv. Phys. Acta* **26**, 241.

Zwicky, F.: 1957, *Morphological Astronomy*, Springer-Verlag, Berlin.