Astron. Astrophys. 151, 264–269 (1985)

ASTRONOMY AND ASTROPHYSICS

The statistical significance of a large quasar inhomogeneity in the sky

J.V. Narlikar and K. Subramanian

Theoretical Astrophysics Group, Tata Institute of Fundamental Research, Bombay 400005, India

Received November 23, 1984; accepted May 9, 1985

Summary. Statistical techniques are devised to study the significance of apparent inhomogeneities in the distribution of points on a spherical surface, especially in the form of jets and chains. The methods are applied to complete samples of radio quasars in specific windows of magnitudes and redshifts. It is found that while most of the distributions studied show very little evidence of non-randomness, the inhomogeneity reported by Arp (1984a, b) is highly significant. There is also some evidence, although a statistically weaker one, for a chainlike structure in the distribution of low redshift quasars in another area of the sky.

Key words: quasars – statistical tests

1. Introduction

In the early days of the discovery of quassars, Strittmatter et al. (1966) had drawn attention to an apparent lack of isotropy in their surface distribution. In particular, they found that high redshift quasars ($z > 1.5$) were concentrated towards two antipodal directions. At that time the effect looked significant but one could have argued that it might not survive once complete and systematic surveys of quasars became available. Indeed the effect in its original form does not appear to be present now.

However, more recently Arp (1983, 1984a, b) has found that an effect in a somewhat different form indicating a large scale inhomogeneity in the distribution of quasars is present in complete samples of radio quasars from the Parkes and the 3CR surveys. In particular, quasars from these samples with redshifts in the range $1.4 < z < 2.7$ and apparent magnitudes in the range $17.5 < V < 19$ are not distributed isotropically but show a concentration in an elongated area running from R.A. $\approx 1^h30^m$, $\delta \approx 30°$ to R.A. $\approx 23^h$, $\delta = 0°$. Arp points out that the companion galaxy M33 in the Local Group lies at one end of this line. In an independent investigation Shastri and Gopal Krishna (1983) also report inhomogeneous distribution of quasars with $z > 2$.

To test the apparent non-randomness of distribution, Arp adds 13^h to the R.A. of all points in the above set to get another region in the Northern Galactic Hemisphere. The paucity of high redshift quasars in this region then becomes noticeable: the corresponding quasar density is found to be significantly higher than actually found in the NGH.

Such a large scale inhomogeneity, if real, would have profound implications for the cosmological interpretation of quasar redshifts, implying that the quasars are considerably nearer than indicated by Hubble's law. It is therefore essential to devise methods of statistical analysis that are rigorous as well as free from the criticism of being *a posteriori*.

In this paper we devise new methods of testing apparent non-randomness of points in the sky, methods that are *a priori* in nature and could be applied, besides quasars, to other groups or distributions of objects in the sky. The aim of these methods is to look for extended structures and to investigate the probability of their occurring by chance in a purely random process. We will consider two types of structures, straight ones designated as "jets" and curved winding ones designated as "chains". Although the human eye can make out jets and chains by looking at a distribution of points, it is more than desireable to have an objective assessment of the effect.

For example, Sofue et al. (1977) had applied a statistical technique to conclude that a linear configuration of seven peculiar galaxies emerging to the southwest of NGC 4889 is significant. The methods proposed here are different and will be outlined in general terms in the following section. In Sect. 3 we will describe specific applications to quasars in which the sample highlighted by Arp will be compared with other samples by way of "control". In Sect. 4 we will discuss the implications of these tests and the scope for improving them further.

2. Statistical techniques: a general discussion

Suppose we have a set of points distributed on the spherical surface. We will assume that the points are specified by spherical polar coordinates (θ, ϕ). Let us consider a large region Σ specified by the rectangle $\theta_1 < \theta < \theta_2$ and $\phi_1 < \phi < \phi_2$. Divide Σ into $N = nm$ smaller rectangles by lines given by $\theta = \text{const.}$ and $\phi = \text{const.}$ To ensure equal area rectangles, we choose the θ values in m equal steps of $\cos\theta$ between θ_1 and θ_2, while the ϕ-range is divided in n equal intervals of ϕ. We now outline four tests in increasing order of sophistication. The first two are well known but

---

Send offprint requests to: J.V. Narlikar we give them here for completeness and also to emphasize their inadequacy for the present purpose. In what follows each of the N rectangles will be referred to as an "area".

2.1. *Rectangular distribution*

If altogether there are Q points in Σ distributed uniformly over N equal areas then the expected number in each area should be Q/N. If the observed number in the i^{th} area is $q_i (i=1,\dots,N)$ then we define the *chi-square* by

$$\chi^2 \equiv \sum_{i=1}^{N} \frac{N}{Q}\left(q_i - \frac{Q}{N}\right)^2 \equiv \left(\sum_{i=1}^{N} \frac{q_i^2 N}{Q}\right) - Q. \tag{1}$$

The *reduced chi-square* is defined by

$$\chi_R^2 \equiv \frac{\chi^2}{N-1}. \tag{2}$$

The reduced chi-square can be compared with the tabulated values for $N-1$ degrees of freedom at various confidence limits.

2.2. *Poisson distribution*

The above method becomes unreliable if small numbers $Q/N < 5$ are expected in each area. In that case a better approach is to count the numbers of areas which have $q_i = 0$, $q_i = 1, \dots$ etc. Let f_r be the frequency of such areas containing r points, the mean number of points per area being

$$\frac{1}{N} \sum_r r f_r = \frac{Q}{N} \equiv \langle r \rangle. \tag{3}$$

Then by the Poisson distribution the expected number of areas with r points is

$$e_r = N e^{-\langle r \rangle} \frac{\langle r \rangle^r}{r!}. \tag{4}$$

The chi-square for this distribution

$$\chi^2 = \sum_r \frac{(f_r - e_r)^2}{e_r} \tag{5}$$

then gives indication of non-randomness.

Both these tests, however, do not tell us anything about extended structures like jets and chains. For example, jets or chains may not be apparent even if either of the above tests shows significant non-randomness. This can happen, for example, if only a few unconnected areas are highly populated and the rest are empty. Conversely, jets and chains could exist even if both these tests show randomness. This is because areas with $r > \langle r \rangle$ may happen to be suitably juxtaposed. New tests are therefore needed to decide how likely is the emergence of structures through a chance juxtaposition of areas with more than average density of population.

2.3. *A search for chains*

Regardless of whether tests 2.1 or 2.2 display non-randomness in the distribution we may search for chains in the following way. First we define a chain C_l^r as a sequence of l contiguous areas with each area containing at least r points. Before proceeding further we comment on the above definition.

We expect the chain to be noticeable, i.e., it should stand out as a dense and longish distribution of points (not necessarily in a straight line) against a relatively sparse background. Thus each area in the chain should contain a significantly higher number of points than the average $\langle r \rangle$. In practice we expect $\langle r \rangle$ to be less than unity and so $\sigma = \langle r \rangle^{1/2}$ represents the standard deviation of the Poisson distribution in (4). Hence if $r > \langle r \rangle + 2\sigma$, we have a significantly dense area at the 2σ level. For example, for $\langle r \rangle = 0.5$, we have $\sigma = 0.707$. For $r = 2$, $(2-0.5)/0.707 > 2$ and hence an area is significantly dense for any r not less than 2.

Barring "edge effects", each area will have 8 neighbours contiguous to it. So a chain, from a given area could start in 8 possible directions. For it to develop further at least one contiguous area must be significantly dense in the above sense. If there is none there is no chain from that point.

With these comments we now proceed to estimate the number of C_l^r chains that are likely to arise in a distribution purely randomly. If the observed number of chains turns out to be significantly larger than the expected number, we have grounds to suspect some physical effect in operation. For a given r we define p_r to be the probability of finding at least r points in a given area; so that

$$p_r = \sum_{s \geq r} \frac{e^{-\langle r \rangle} \cdot \langle r \rangle^s}{s!}. \tag{6}$$

Now start with any of the N areas, ignoring the edge effect due to the boundary, which we shall consider later. Suppose we have a C_l^r chain from here. The chance of this area containing at least r points is p_r. The chain can proceed from here in 8 possible ways provided the next area has at least r points. The expected number of such areas is therefore $8p_r$.

To look for the next contiguous area we have 7 possibilities since we want to avoid the direction we came from. So the expected number of ways of the chain continuing further is $7p_r$. The same argument applies at each subsequent stage. Thus the expected number of C_l^r chains is given by

$$E_l^r < N \times p_r \times 8p_r \times (7p_r)^{l-2}/l! = \frac{8N}{49} \frac{(7p_r)^l}{l!}. \tag{7}$$

The factor $l!$ appears in the denominator to correct for overcounting the same sequence of areas.

The inequality expresses the fact that the right hand side of the above expression overestimates the probability of random chains. One reason for the overestimate is the fact that we have ignored the edge effects which would give less numbers for areas on the borders and also the fact that we have ignored the possibility of a long chain crossing itself.

It could be argued that the criterion for chains as given here would not be able to distinguish between winding spiral chains from "amorphous blobs". A possible way of avoiding this problem and ensuring one-dimensional structure in the distribution of points would be to look for anticorrelation in directions orthogonal to the link in the chain. It is hard to quantify the number of such chains expected by chance but in any case it would necessarily be *less* than the number given by (7). Thus if an observed distribution is found significant with the help of (7) it would have to be significant under a more refined procedure like that outlined above.

Now suppose that the observed number of C_l^r chains is O_l^r. Define the Poisson probability function by

$$P(m,r) = e^{-m} \frac{m^r}{r!}. \tag{8}$$

The probability of seeing at least O_l^r chains by chance is less than

$$p = \sum_{q \geqq O_l^r} P(E_l^r, q) . \quad (9)$$

This probability may be compared with typical significance levels like 5% or 1%.

2.4. A search for jets

A chain can meander in many ways as mentioned above. A more clear cut appearance in a non-random distribution is that of a linearly aligned sequence of densely populated areas. We define a jet A_l^r as a linear sequence of l contiguous areas, each containing at least r points. The calculation of the expected number of jets of type A_l^r proceeds as follows.

Start with the first area anywhere. For it to contain at least r points the probability is p_r. The next contiguous area can be chosen in 8 ways with the expected number $8p_r$. Subsequent areas have to be in the straight line direction determined by the first two areas. The expected number of such A_l^r is then given by

$$\xi_l^r < N \times 8 \times p_r^l/2! = 4Np_r^l . \quad (10)$$

(The inequality recognizes that edge effects have been ignored.) If the observed number of jets of type A_l^r in a given sample is L_l^r then the probability of chance occurrence is less than

$$p = \sum_{q \geqq L_l^r} P(\xi_l^r, q) . \quad (11)$$

Again, this probability may be compared with a specified level of significance, e.g. 5%.

Because of the discreteness of the problem, the number of directions in which a jet can be identified in this way is eight. This number could be increased to twelve by choosing a grid of "equilateral" triangles to cover the specified area on the sphere. However, since we are not looking for very precisely aligned structures but for set of points which have a relatively denser distribution of points around straight lines, it is unlikely that the above method misses out jets in intermediate directions. In any case since the comparison with observations will be made with the same grid of areas any underestimate of ξ_l^r is compensated by a similar undercounting of L_l^r.

2.5. Criteria for binning

The above two tests require the binning of the point distribution into areas of discrete sizes. How large should these areas be? Clearly, for a meaningful search for structure it is necessary to set some numerical limits on the number of points in significant areas and their sizes. These limits are as follows.

For example, we have already indicated how a significantly dense area is identified: The number r of points in such an area should be at least about 2σ larger than the overall mean for the population. There is another criterion that is relevant here. Suppose that the number of areas with points not less than r is $\lambda(r)$. Then the average separation between two such areas in a random distribution is $[N/\lambda(r)]^{1/2}$. In order that such areas are not juxtaposed frequently by chance we should ensure that

$$\left[\frac{N}{\lambda(r)}\right]^{1/2} \gtrsim 2 . \quad (12)$$

If this condition is not satisfied, there will be an overcrowding of squares that would clump most bins in the form of chains by a purely random process. Further, to ensure stability of the result we should not choose $r=1$ even if the condition (12) is satisfied then. For, $r=1$ is the least number one can expect in a discrete distribution; for a real effect to be seen we should ask for more stringent conditions, e.g., $r=2$.

The fact that the "least count" in our data set is 1 (there cannot be fractions of a quasar) can also be used to get an upper bound in the number of areas N. For if N is so large that $\langle r\rangle + \sigma < 1$ for the data set then even if the quasar distribution is non-random the tests outlined above would not be able to make this out. On the other hand, if N is too small, then the area size will exceed the scale of the non-random effect we are trying to detect.

2.6. The classification of chains

The above criteria can be translated into numbers only after the actual point distribution is given. We will further specify rules for identifying chains in order to ensure that the effect found is not spurious or trivial:

a) Choose only the largest chain. Rule out a single square chain ($l=1$) even if significant.

b) Test whether the chain structure survives even after changing N within the limits discussed above. Spurious effects will in general go away under such a variation. Also test whether the chains picked up for different N's all lie in the same part of the sky (i.e., they all should imply the same chain structure).

c) If more than one sample distributions on the sky are available, it would be useful to apply the same binning to all and compare. For example, if one sample is claimed to have structure and the others are random, then the latter should serve as "controls" on the former.

3. Application to radio quasars

The analysis of the preceding section lays out a general procedure and calculates a priori probabilities for non-random distributions, chains and jets. We can now apply this analysis to the distribution of quasars on the sky and see if any non-random signal emerges.

To this end we have, somewhat arbitrarily divided the total sample of Parkes and 3CR radio quasars into a) different redshift ranges b) different magnitude ranges and c) different areas of the sky. Table 1 contains the summary of our investigations. The very first sample listed there includes Arp's sample (1983), wherein the effect was first noticed. Our aim has been to see if a) the effect found by Arp is significant by any of the statistical criteria discussed above and b) the effect is also present in other samples.

It will be seen that with possibly one or two exceptions all the samples other than Arp's chosen sample show a random distribution on the sky. We will therefore discuss these cases in some detail, beginning with Arp's sample. The sample numbers are according to Table 1. The optical data are from the catalogue of Hewitt and Burbidge (1979).

Sample 1

This covers the region $-45° \leqq \delta \leqq 45°$, $20^h \leqq$ R.A. $\leqq 6^h$ and consists of quasars with redshifts in the range [1.4, 2.7] and magnitudes in the range [17.5, 19].

To begin with the region was divided into $10 \times 10 = 100$ areas. Such a division satisfies the binning criteria discussed earlier. As there are 50 quasars, the mean occupancy of an area is $\langle r\rangle = 0.5$. Thus $\sigma = 0.707$. The reduced χ_R^2 as computed according to (2) is

Table 1. Statistical analysis of large-scale distributions of radio quasars

Sample No.	Sky area	Redshift range	Magnitude range	Results of statistical tests[a] R: Random NR: Non Random			
				(i)[b]	(ii)	(iii)	(iv)
1	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[1.4, 2.7]	[17.5, 19.0]	NR	NR	NR	NR
2.	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[2.7, 3.8]	[17.5, 19.0]	R	R	R	R
3.	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[0, 1.4]	[17.5, 19.0]	R	R	R	R
4.	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[0.27–0.47]	[13.5, 16.8]	R	R	R	R
5.	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[1.4–2.7]	[13, 17.5]	NR	R	R	R
6.	$-45° \leqq \delta \leqq 45°$, $20^h \leqq RA \leqq 6^h$	[1.4, 2.7]	[19, 21]	R	R	R	R
7.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[0, 1.4]	[17.5, 19]	R	R	NR	R
8.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[1.4, 2.7]	[17.5, 19]	R	R	R	R
9.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[2.7, 3.8]	[17.5, 19]	R	R	R	R
10.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[0, 1.4]	[12, 17.5]	R	R	R	R
11.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[0, 1.4]	[19, 21]	NR	R	R	R
12.	$-45° \leqq \delta \leqq 45°$, $8^h \leqq RA \leqq 18^h$	[0.6, 1.4]	[18.5, 21]	R	R	R	R

[a] Nonrandomness is assumed for probabilities <5%, although in most NR cases they are much lower
[b] This is not reliable since $\langle r \rangle < 1$ in all cases

Table 2. Sample 1: Sky-distribution of quasars in the redshift range [1.4–2.7] and the magnitude range [17.5–19]

Declination in degrees	Right ascension in hours									
	6–5	5–4	4–3	3–2	2–1	1–0	24–23	23–22	22–21	21–20
45.00 – 34.45	0	0	0	0	0	0	0	0	0	0
34.45 – 25.10	0	0	0	1	2	0	0	0	0	
25.10 – 16.43	0	0	0	0	2	2	0	3	2	0
16.43 – 8.13	0	0	0	1	0	1	3	0	1	0
8.13 – 0.00	0	2	0	0	1	0	3	2	3	0
0.00 – (−8.13)	0	1	0	3	0	2	1	1	0	0
(−8.13)–(−16.43)	0	2	0	0	0	0	0	0	0	0
(−16.43)–(−25.10)	0	1	0	1	0	1	1	1	0	0
(−25.10)–(−34.45)	0	0	0	1	1	0	1	1	0	0
(−34.45)–(−45.00)	0	0	1	0	0	0	0	0	1	2

1.4343 which at 99 degrees of freedom is *significant* with probability of χ_R^2 exceeding this value being less than $5\ 10^{-3}$.

However, the low value of $\langle r \rangle$ makes this test unreliable. So we go to the next test described in §2.2. The value of χ^2 computed according to (5) is found to be *significant* with the probability less than $5\ 10^{-4}$.

Thus the distribution is definitely non-random. Do we actually see any chains or jets which are significant by tests in Sects. 2.3 and 2.4? The actual distribution is given in Table 2, for the 10×10 binning.

Let us put $r = 2$, i.e., consider those areas significant which contain at least 2 quasars. Table 2 shows that the longest chain is of length $l = 7$. For $l = 7$, $r = 2$ and $p_2 = 0.0902$, formula (7) gives

$$E_7^2 < \frac{800}{49} \frac{(7 \times 0.0902)^7}{7!} = 1.3\ 10^{-4}. \tag{13}$$

Since $O_7^2 = 1$, the probability of such a long chain developing by chance is less than the right hand side of (13).

Table 2 shows the longest jet at $r = 2$ to be of length $l = 4$. Using (10) we get

$$\xi_4^2 < 400 \times (0.0902)^4 = 2.65\ 10^{-2}. \tag{14}$$

Both the longest chain and the longest jet are therefore statistically significant and they both form part of the area singled out by Arp as having excess quasars. M33 lies almost towards the top of this region.

We could look for and find many chains of shorter length in this region and use their observed numbers against the expected ones to test for significance. The probabilities come out in the range of $10^{-4} - 10^{-2}$. However, if we are looking for the largest inhomogeneity on the sky then we should test against the largest value of l observed.

Sample 7

This covers the region $-45° \leqq \delta \leqq 45°$, $8^h \leqq R.A. \leqq 18^h$, lying "on the other side" of the sky compared to the region of Sample 1. The

Table 3. Sample 7: Sky-distribution of quasars in the redshift range [0, 1.4] and the magnitude range [17.5, 19]

Declination in degrees	Right ascension in hours									
	18–17	17–16	16–15	15–14	14–13	13–12	12–11	11–10	10–9	9–8
45.00– 34.45	0	0	0	0	0	1	1	0	2	0
34.45 – 25.10	0	2	0	0	1	0	1	1	0	0
25.10 – 16.43	0	0	0	1	0	2	0	4	1	2
16.43 – 8.13	0	0	1	1	0	0	2	0	1	1
8.13 – 0.00	1	1	2	0	3	1	1	2	2	1
0.00 – (−8.13)	0	0	0	3	1	4	0	0	2	1
(−8.13)–(−16.43)	0	0	0	1	1	1	1	0	0	0
(−16.43)–(−25.10)	0	0	0	2	0	0	1	0	1	0
(−25.10)–(−34.45)	0	0	0	1	0	1	0	0	0	0
(−34.45)–(−45.00)	0	0	0	1	1	0	1	0	0	0

Table 4a. A list of NR chains with their probabilites for various binning (x indicates no NR chains)

Sample No.	No. of QSO's	7×7	8×8	10×10	11×11	12×12	15×15
1	50	10^{-5}	$4\ 10^{-3}$	$1.3\ 10^{-4}$	$3.2\ 10^{-3}$	x	x
3	54	$1.3\ 10^{-4}$(A)	$8\ 10^{-3}$(E)	x	x	x	x
7	63	$2\ 10^{-6}$(A)	$3\ 10^{-2}$	$3.8\ 10^{-3}$	$2.6\ 10^{-3}$	$6.3\ 10^{-3}$	x
10	59	$3\ 10^{-2}$(E)	x	x	x	x	x
12	18	$8\ 10^{-3}$	$2\ 10^{-2}$	x	x	x	x

(A) means chain spans the window. (E) means chain goes to edge of window

Table 4b. A list of NR jets with their probabilities for various binning (x indicates no NR chains)

Sample No.	No. of QSO's	7×7	8×8	10×10	11×11	12×12	15×15
1	50	$2.4\ 10^{-4}$	$2\ 10^{-2}$	$2.7\ 10^{-2}$	$8.7\ 10^{-3}$	x	x
3	54	10^{-4}	x	x	x	x	x
7	63	x	x	x	x	$2\ 10^{-2}$	x
10	69	x	x	x	x	x	$4\ 10^{-2}$
12	18	$8\ 10^{-3}$	$3\ 10^{-3}$	x	x	x	x

redshift range is [0, 1.4] and the magnitude range is [17.5, 19]. The division of this region into 10×10 areas and the distribution of quasars therein are shown in Table 3.

There are 63 quasars with mean $\langle r\rangle = 0.63$, and $\sigma = 0.794$. The small value of $\langle r\rangle$ again tells us to drop the first test and go to the second. The value of χ^2 is *not* significant for the Poisson distribution. However, a glance at Table 3 reveals two chains of length 5 and a jet of length 4 at $r = 2$. Let us therefore investigate the tests of Sects. 2.3 and 2.4 for this distribution.

We have $p_2 = 0.1319$ and (7) gives

$$E_5^2 < \frac{800}{49} \times \frac{(7 \times 0.1558)^5}{120} \simeq 0.0912. \tag{15}$$

Since two chains of length 5 are observed, the Poisson probability (9) works out to be $\simeq 4\ 10^{-3}$. Also, for $l = 4$, formula (10) gives

$$\xi_4^2 < 400 \times (0.1319)^4 \simeq 0.121. \tag{16}$$

The probabilities here are not as low as in Sample 1 and the jet could certainly have arisen by chance. The choice of $r = 2$ implies a population density only 1.7σ in excess of the mean, which is why p_2 is not very low. We therefore have here a considerably weaker, perhaps a marginal case for a chain, even though to the eye the chain in Table 3 looks impressive.

We next considered the variation of N, by increasing or decreasing m in the $m \times m$ grid, from the value $m = 10$ used above. Tables 4a and 4b list the outcome of such studies for chains and jets. The entries in columns 3–8 are probabilities for finding the chains by chance, wherever the probabiliteis are low enough to be significant. The blanks imply that no non-random structures are found. The tables go as far as $m = 15$ although the criteria discussed in Sect. 2.6 restrict m to less than 12. Thus the blanks for $m = 12$, 15 are not important for our discussion.

The chains and jets listed in Tables 4a and 4b were further tested using the criterion (c) of Sect. 2.6. The chains in samples 10 and 3 do not survive when N is increased (with N still below the upper limit) and as far as the chain in sample 7 is concerned different N's pick up structures in different parts of the sky. Only the chain and the jet in sample 1 survives as a significantly non-random structure under the variation of N; and only in this case do the different binnings pick up the chain and jet in the same area of the sky. (This chain is marginally random for $m=12$, but for this value of m the condition $\langle r \rangle + \sigma > 1$ is not satisfied.)

4. Conclusion

The statistical techniques developed here do imply that the inhomogeneity found by Arp by visual inspection is very significant. This conclusion gains strength when contrasted with the performance of other samples as the test parameters are varied.

One might wonder at this stage whether the sample or the coordinate, redshift and magnitude ranges have been chosen so as to optimize the effect. Two comments are appropriate in this context. First, Arp's survey included two complete samples and so it is unlikely that there are any selection effects in the choice of the data. Secondly, Arp argues (Arp, 1984b) from an analysis of his various surveys that the quasars with high redshifts and bright apparent magnitude (which constitute sample 1) are likely to be the nearest quasars. Thus his sample 1 has been chosen "A priori" to test whether one can find any evidence for the association of these quasars with nearby galaxies. And it is indeed such an association in the form of a jet from M 33 that our analysis has shown to be statistically significant. The conventional (Hubble law) alternative to Arp's interpretation on the other hand translates the above angular inhomogeneity into a spatial inhomogeneity extending to ~ 1000 Mpc. Our aim here, however, is not to discuss the merits of theoretical alternatives but to draw attention to an effect that deserves further study.

A question that we have not tackled here is that of multiple branching where a chain can extend in two or more directions. The Sample 7 shows such a situation. Multiple branching leads to overlapping chains of varying lengths, and the theoretical estimate of the probability of their occurring by chance is harder to compute. This problem is currently under investigation.

Finally, we should like to mention an alternative approach to the problem, an approach which does not involve binning of quasar positions. This is to associate with each quasar a circular lamina of radius r, (which is chosen so that if two quasars are within an angular distance r they can be considered to be associated) and then ask whether the clumping of these laminae is random. Such questions have already been tackled in literature (see for example Roach, 1968) in connection with counts and distributions of dust particles and bacteria. A drawback of this approach, however, is the arbitrariness in choosing the value of r. We are exploring ways of overcoming this difficulty.

The methods outlined here could also be applied to the study of other inhomogeneities reported by Arp (1984) as well as to the suspected sequential structures in superclusters.

Acknowledgements. We are grateful to Dr. Arp for comments while this work was in progress. One of us (JVN) thanks the Kitt Peak National Observatory for hospitality.

Postscript

We have now carried out computer simulations of randomly distributed quasars as an independent check on the analytical estimates of the chain and jet probabilities given in § 2. In each simulation we put 50 points (corresponding to the number of quasars in Sample I) randomly on a 10×10 grid and counted the number of chains and jets of various lengths according to the criteria given in § 2. In order to get reasonably good statistics 500 such simulations were run. We find very good agreement between the observed and predicted numbers of "1-" to "4-chains", about twice as many "5-" and "6-chains" as that estimated in § 2 and no 7-chains. We also find that the number of long ($l \geqq 3$) jets is less than that estimated analytically. In the 500 simulations we find only one "4-jet" of the kind present in sample I, which gives a probability that the "4-jet" in sample I arising by chance $\sim 2\ 10^{-3}$. The computer simulations therefore do bear out in the main, our conclusion that the inhomogeneity found by Arp (Sample I) is significant. The details of the results from the computer simulations will be presented elsewhere.

References

Arp, M.: "Groups, Concentrations and Associations of Quasars", Liège Conference, 1983 June, Institut d'Astrophysique, Universite Liège
Arp, H.: 1984a, *Astrophys. J. Letters* **277**, L27
Arp, H.: 1984b, *J. Astrophys. Astron.* **5**, 31
Hewitt, A., Burbidge, G.R.: 1984, *Astrophys. J. Suppl.* **43**, 57
Roach, S.A.: 1968, *The Theory of Random Clumping* (Methuen), Chap. 4
Shastri, P., Gopal Krishna: 1983, *J. Astrophys. Astron.* **4**, 109
Sofue, Y., Fujimoto, M., Jugaku: 1977, *Publ. Astron. Soc. Japan* **29**, 27
Strittmatter, P., Faulkner, J., Walmsley, M.: 1966, *Nature* **212**, 1441
Veron-Cetty, M.P., Veron, P.: 1984, ESO Scientific Report No. 1