1988ASPC....2..281G

Statistical tests applied to the spatial distribution of quasars in several fields

E. GOSSET, J. SURDEJ, J.P. SWINGS
Institut d'Astrophysique, Université de Liège, Avenue de Cointe, 5
B-4200 Cointe-Ougrée (Belgium).

Abstract: We discuss the application of different statistical tests to the study of the spatial distribution of quasars. Emphasis is given to the necessity of simultaneously using fundamentally different methods in order to reach a good level in the understanding of the true nature of the distribution. Applications to data sets of optically selected quasars lead to the detection of a clustering at a typical scale of 10 arcminutes. Furthermore, the spatial distribution of quasars in a field around NGC 450 shows a deviation from randomness, towards clustering, at a scale of 10 h^{-1} Mpc.

INTRODUCTION

Over the last decade, and since the pioneering work of Osmer (1981), many astronomers have tried to analyze the distribution of quasars: the initial results suggested a random distribution. However, Shaver (1984) has pointed out a tendency towards clustering, tendency later confirmed by himself (Shaver, 1987) or by other authors such as Kruszewski (1987), Anderson et al. (1987), etc. Nevertheless, these results were based on a statistical method applied to a catalog and we feel that, if some observational biases can be taken into account, the existence of less well understood or even unknown effects cannot be ruled out. It seems to us complementary to study more homogeneous (although limited) samples of quasars in order to confirm the phenomenon. Both approaches are necessary, and we decided to concentrate on the latter. We have already reported preliminary results from the analysis of our surveys (Gosset et al., 1986); an up-to-date report is given in Swings et al. (1988). The present paper is dedicated to comments on the various statistical tests used for the analysis of the different samples.

THE STATISTICAL TESTS

The question is: *are quasars uniformly distributed at random ?* There is no doubt that, to give an answer to such a question, it is necessary to make use of statistical tests. Whereas a visual inspection may be very powerful in detecting anomalies or deviations from randomness, only a quantitative approach will help to know whether such effects are attributable to the underlying process or are simply due to statistical

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System

1988ASPC....2..281G

fluctuations of the sampling. Fortunately, it is now common practice to utilize a statistical method when analyzing data. However, our experience in this matter suggests that this is still not sufficient. Each method has its own advantages, sensitivity, property but also has its defects and weaknesses. Therefore, an exhaustive knowledge or understanding of the data is *not* accessible when one uses just one method. The combined results of several independent methods are, without any doubt, essential to reach a high degree of confidence in the conclusions of the analysis.

We have searched the literature in order to gather several good methods (details are given in Gosset, 1987a). We have retained a combination of five of them for their fundamentally different nature.

The first selected method is the MBA (Multiple Binning Analysis). This is the most ancient test and it has been widely used. Nevertheless, it lost some of its importance with respect to other more recent tests. Essentially, such a weakness is not due to the nature of the method of analysis but to the choice of the statistic. Gosset and Louis (1986) have put the MBA back to its right place among the most useful tests by introducing a statistic based on a randomization process. They have introduced the 4 within 16 randomization and the 8 within 64 randomization tests for the two- and the three-dimensional MBA's, respectively.

The second method is the CFA (Correlation Function Analysis). It consists in making counts in concentric rings sequentially built-up around each object of the sample. This method is extremely sensitive to edge effects and a correction has therefore to be applied. The only way to properly perform this correction is to calculate the exact measure of the domain actually explored when making counts. Such an approach can induce systematic effects, and the best way to estimate these consists nowadays in computing the mean cross-correlation function between simulated populations of uniformly distributed (usually) individuals and the data (as suggested by Sharp, 1979). One of the relevant estimators is unbiased, and its dispersion over the simulations gives a good approximation of the error associated with the auto-correlation function.

The third method is the NNA (Nearest Neighbours Analysis) which is based on the mean distances to neighbours. It is sensitive to edge effects too; no rigorous correction can be applied, although a good approximation can be obtained by using simulated populations mimicking the data. This test is somewhat less powerful but gives some additional information such as, for example, the number of individuals per cluster.

The fourth method, the PSA (Power Spectrum Analysis; Webster, 1976), as well as its generalized version (GPSA; Peacock, 1983), has a good reputation of flexibility and of great sensitivity. However, we found that this characteristic is overrated and we would just like to show a simple example. We have generated a 2-D synthetic population heavily clustered at a scale of 0.05; the clusters were built in such a way that they have a tendency to repel each other, and the characteristic scale of the inhibition was taken to be 0.12. So, the clusters are not randomly distributed but, rather, exhibit a deviation towards regularity. Figure 1 illustrates the variation of the statistic Q' of the PSA as a function of the spatial frequency explored. A peak at $1/\lambda = 7.5$ is clearly visible; it corresponds to the inter-cluster regularity ($7.5 \sim 1/0.12$). It is nevertheless worth noticing that, besides this, there is no visible trace of clustering ($20 \sim 1/0.05$). To be sure that the clustering is present, we have applied the MBA in the configuration of the 4 within 16 randomization test. The run of the relevant normal statistic Z is shown in Figure 2 as a function of the investigated characteristic scale. We obtain a deviation greater than 4σ towards clustering, in good

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System

1988ASPC....2..281G

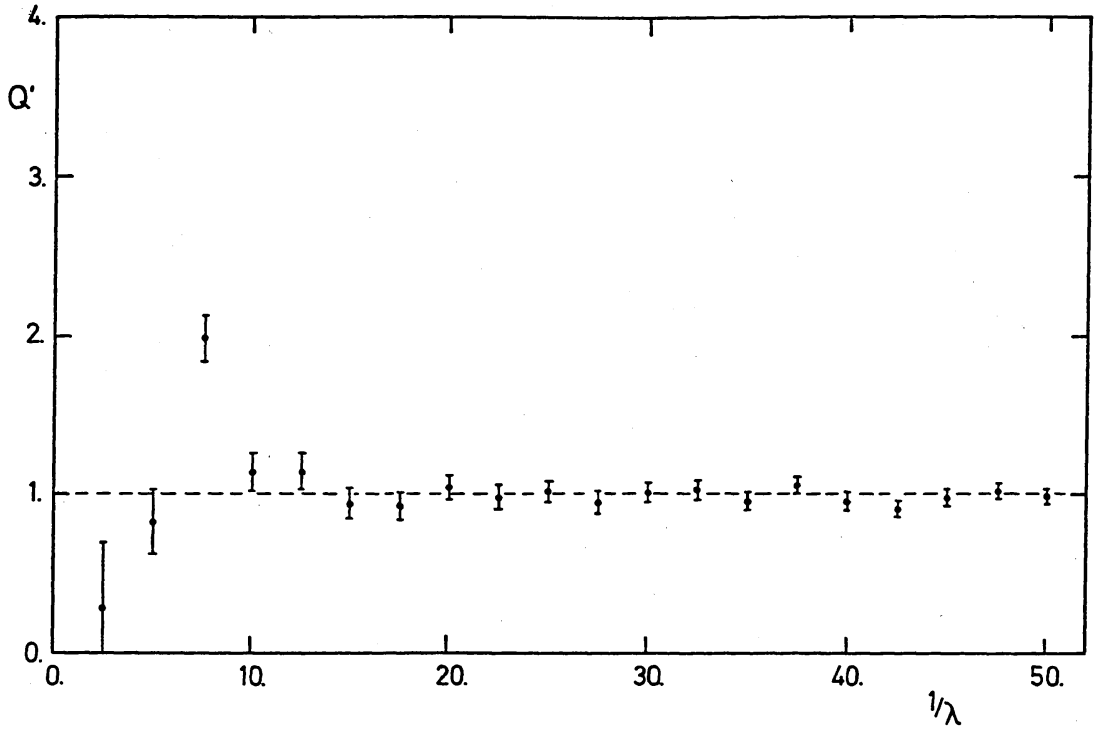

Figure 1

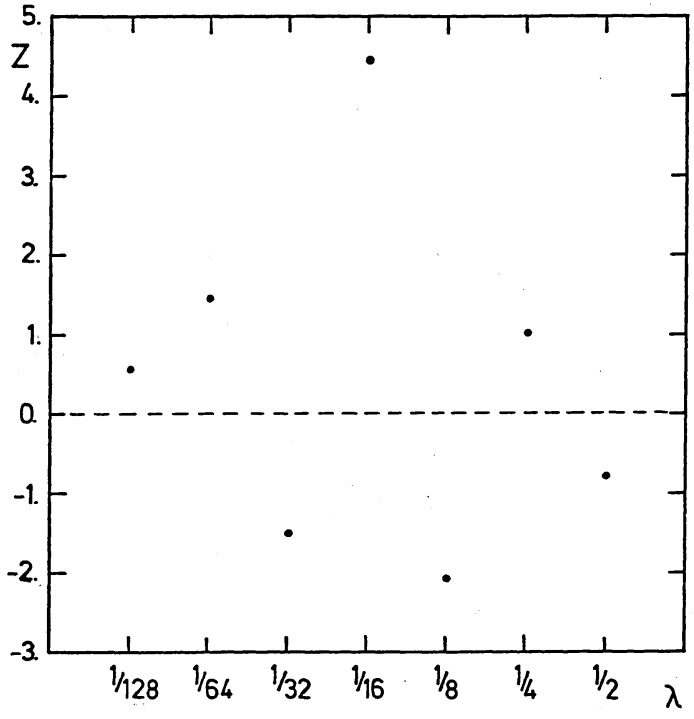

Figure 2

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System

1988ASPC....2..281G

agreement with the known characteristic of the synthetic population. The MBA also reveals a deviation towards regularity at a scale of 1/8, in good agreement with the PSA. Three things are to be retained as conclusions at this particular point: (1) the PSA is sometimes difficult or impossible to interpret and can lead to false conclusions; (2) this is another proof of the interest of the randomization tests; (3) again, the use of several methods is most useful.

Finally, a fifth method has been taken into account. Effectively, the first four are not well suited for analyzing data at a typical scale of the order of that of the whole field. For the one-dimensional space, such a test, only sensitive to large scale deviations, exists: it is the well-known Kolmogorov-Smirnov test. Peacock (1983) has proposed an Extended Kolmogorov-Smirnov test for the two-dimensional case while, recently, Gosset (1987b) has introduced the three-dimensional EKS test. The EKS tests are also helpful in adjusting a probability density function to the distribution of the individuals. They can be used jointly with the NNA in order to correct for edge effects or as a preliminary step to the application of the GPSA.

We conceived a set of computer codes in order to test those five selected methods and we will use them in the analysis of the data.

THE DATA ANALYSIS

Data from different fields have been analyzed. A brief description of the three main surveys is given in Swings et al. (1988). In a preceding report (Gosset et al., 1986), we concluded that the quasars in the NGC 450 field were clustered on the celestial sphere with a typical scale of 10 arcminutes. We have subsequently analyzed the distribution of quasars in the NGC 520 field and in the ESO field #300. Both of these samples show the same deviation towards clustering at a similar scale of about 10 arcminutes.

The NGC 450 sample has been analyzed in a three dimensional space (α, δ, z). Because the number of identified quasars ($\sim$ 60) is quite small, it results a very low volume density. For this reason, some difficulties may arise. The GPSA is difficult to apply and no convincing result in any direction can be drawn. Provided we authorize some mixing of the scales, it is possible to apply the MBA. The 8 within 64 randomization test detects a deviation towards clustering with a significance level of 0.04 and at a typical scale somewhere between 6 h^{-1} Mpc and 70 h^{-1} Mpc. The CFA detects a deviation, greater than 4σ, towards clustering and with a characteristic scale of about 12 h^{-1} Mpc. In order to take into account selection effects on the redshift, we also computed the standard deviation on the basis of non-uniform simulated populations obtained by randomizing the redshifts. This attitude leads to a highly conservative test. The relevant significance level for the 12 h^{-1} Mpc clustering is 0.04 and is to be considered as an upper limit. The agreement with the results from the MBA is interesting. The NNA test indicates that we have to deal with a tendency to form pairs. Finally, we would like to note that the deviation in this particular field is mainly caused by the presence of the quasars Q0107-025A and Q0107-025B, whose spectra are described in Surdej et al. (1986).

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System

1988ASPC....2..281G

CONCLUDING REMARKS

We believe that a good and thorough analysis of the distribution of a sample of quasars for the detection of clustering, contagion or regularity, requires the simultaneous use of several methods (at least MBA, CFA, NNA, PSA and EKS). We have suggested what appear to presently be the best configurations for these tests. The analysis of our samples leads to the general detection of a clustering on the celestial sphere at a typical scale of 10 arcminutes. A deviation towards spatial clustering with a highly conservative significance level of 0.04 and a scale of approximately 10 h^{-1} Mpc is also reported for the field of optically selected quasars around NGC 450.

REFERENCES

Anderson, N., Kunth, D., Sargent, W. L. W.: 1987, *Astron. J.*, in press.
Gosset, E.: 1987a, *Analyse de nuages de points. Applications astronomiques et étude de la distribution des quasars*, Ph.D. Dissertation, Université de Liège.
Gosset, E.: 1987b, *Astron. Astrophys.*, **188**, 258.
Gosset, E., Louis, B.: 1986, *Astrophys. Space Sci.*, **120**, 263.
Gosset, E., Surdej, J., Swings, J. P.: 1986, in *Quasars*, IAU Symposium no.119, eds.: G. Swarup, V. K. Kapahi, Reidel, Dordrecht, 45.
Kruszewski, A.: 1987, *Preprint.*
Osmer, P. S.: 1981, *Astrophys. J.*, **247**, 762.
Peacock, J. A.: 1983, *Monthly Notices Roy. Astron. Soc.*, **202**, 615.
Sharp, N. A.: 1979, *Astron. Astrophys.*, **74**, 308.
Shaver, P. A.: 1984, *Astron. Astrophys.*, **136**, L9.
Shaver, P. A.: 1987, *ESO Preprint*, no.**534**.
Surdej, J., Arp, H., Gosset, E., Kruszewski, A., Robertson, J. G., Shaver, P. A., Swings, J. P.: 1986, *Astron. Astrophys.*, **161**, 209.
Swings, J. P., Surdej, J., Gosset, E.: 1988, in *Optical surveys for quasars*, these proceedings.
Webster, A. S.: 1976, *Monthly Notices Roy. Astron. Soc.*, **175**, 61.

CAPTION FOR FIGURES

Figure 1: Run of the statistic Q' of the bi-dimensional PSA as a function of the spatial frequency $1/\lambda$ (see text for details).

Figure 2: Run of the normal statistic Z of the MBA in the configuration of the 4 within 16 randomization test as a function of the investigated characteristic scale (see text for details).

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System

1988ASPC....2..281G

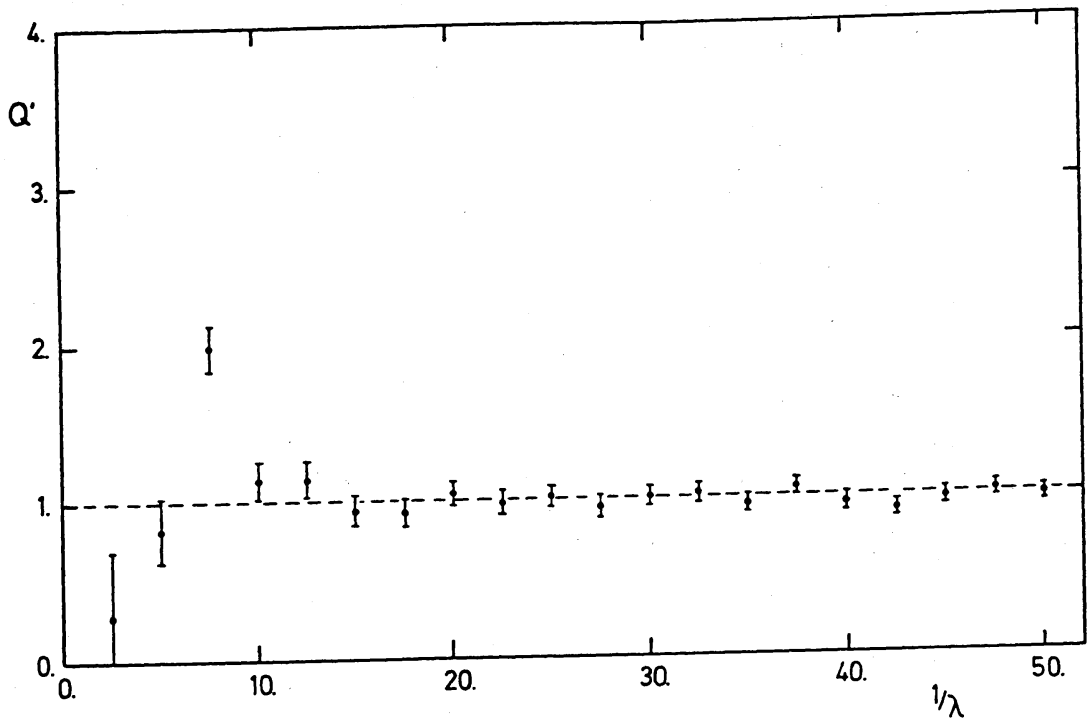

Figure 1

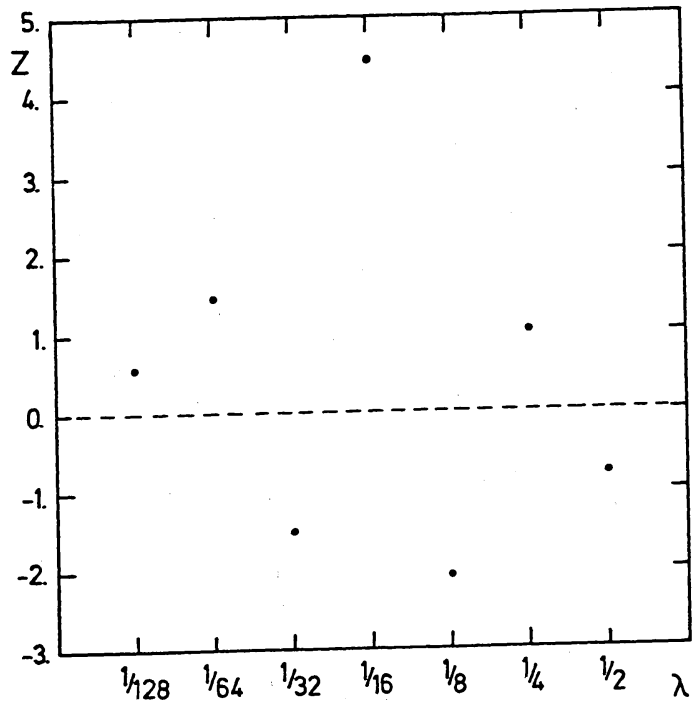

Figure 2

© Astronomical Society of the Pacific • Provided by the NASA Astrophysics Data System