Mon. Not. R. astr. Soc. (1983) **202,** 615–627

Two-dimensional goodness-of-fit testing in astronomy

J. A. Peacock *Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, Scotland*

Received 1982 May 14; in original form 1982 April 26

Summary. This paper deals with the techniques available to test for consistency between the empirical distribution of data points on a plane and a hypothetical density law. Two new statistical tests are developed. The first is a two-dimensional version of the Kolmogorov–Smirnov test, for which the distribution of the test statistic is investigated using a Monte Carlo method. This test is found in practice to be very nearly distribution-free, and empirical formulae for the confidence levels are given. Secondly, the method of power-spectrum analysis is extended to deal with cases in which the null hypothesis is not a uniform distribution. These methods are illustrated by application to the distribution of quasar candidates found on an objective-prism plate of the Virgo Cluster.

1 Introduction

A common problem in astronomy is the comparison of the empirical distribution of a given observed quantity (e.g. redshift, magnitude, colour, etc.) with a theoretical curve. Two related questions are involved: (i) given a set of data values, what can we conclude about the underlying distribution from which they are drawn? (ii) Given a guess for the underlying distribution, what is the probability that the observed data could have arisen from such a function? For the case of a single observed quantity, these questions have standard answers, but this is not so when several quantities are involved. Consider, for example, measuring the magnitude and redshift for all members of a magnitude-limited quasar sample: the distribution of points on the m–z plane is obviously related to the luminosity function. There is, however, no standard statistical test which can compare this distribution with a theoretical one, without losing efficiency through binning. The main purpose of this paper is to develop such a test.

Section 2 considers the ways in which a multi-dimensional goodness-of-fit statistic might be defined. Section 3 deals with a specific example – a two-dimensional analogue of the Kolmogorov–Smirnov statistic; the significance levels for this test have been derived using Monte Carlo techniques. The limitations of techniques of this sort are discussed in Section 4, by comparison with the more general method of power-spectrum analysis. These results are illustrated in Section 5 by application to the distribution on the sky of a sample of optically-selected quasars.

2 Data in several dimensions

The simplest way of assessing the fit between a model and multi-dimensional data is the χ^2 test: if the number of data points is large, then the data space may be divided into bins and the comparison is made in terms of the observed and predicted numbers in each bin. However, astronomical samples are often limited in size due to practical difficulties in observing large numbers of objects – in which case it is crucial to make optimum use of the data. For small numbers, it is well known that binning is inefficient; we really require a test which will make use of each data point individually.

One possible statistic of this type is the likelihood. For a multi-dimensional probability density function $f(x, y, \ldots)$, the likelihood, L, is defined by

$$L = \prod_i f(x_i, y_i, \ldots)$$

where the product is taken over all data points. This function represents, loosely, the probability of the observed data set arising from the hypothetical f. L is therefore in some sense a measure of goodness-of-fit between f and the data. Formally, using Bayes' theorem, we say that the *a posteriori* probability of f being the true density law from which the data were drawn is proportional to L. Hence, L is used as a measure of the *relative* probabilities of various alternative density laws; the criterion of maximum likelihood is thus commonly used in finding the best-fitting member of a family of parameterized model density laws, although some other test is normally used to find whether the fit between the final model and the data is acceptable in an *absolute* sense.

We could attempt to give the likelihood an absolute meaning: for n data points selected at random from a given density law $f(x, y, \ldots)$, the mean and variance of $ln\,(L)$ are

$$\mu\,[ln\,(L)] = n\,[\int f ln(f)\,dx\,dy\,\ldots]$$

$$\sigma^2\,[ln\,(L)] = n\,\{\int f\,[ln(f)]^2\,dx\,dy\,\ldots - [\int f ln\,(f)\,dx\,dy\,\ldots]^2\}\,.$$

For large n, the central limit theorem applies and the distribution of $ln\,(L)$ will be approximately normal. While a test based on L would be of some value, in that extremely low values of L would indicate a misfit, the likelihood cannot be used as a general test. To see this, consider the probability distribution of a single variable, $f(x)$; if f is constant in a certain range of x, we should expect a uniform distribution of data values in that range. However, L will be the same whether the values are uniformly distributed or clumped about one point in a way inconsistent with f. In short, because L takes no account of the relation between data points, any goodness-of-fit test based on L cannot be consistent against all alternatives.

For the one-dimensional case, the only efficient general goodness-of-fit tests are those based on differences in cumulative probability distributions. The most widely used of these is the Kolmogorov–Smirnov test, which considers simply the largest absolute difference between the two cumulative distributions as a measure of misfit. This test has a number of attractive features: it is highly efficient, as a general purpose test sensitive to any difference in the two distributions, since it makes use of each individual data point. Also, it is one of the class of tests called robust or distribution-free: the expected distribution of the test statistic is independent of the particular theoretical density law being considered. This latter property comes from the fact that any one-dimensional density law may always be mapped on to any other by a transformation which preserves the ordering of the variable, thus leaving the Kolmogorov–Smirnov statistic unaltered. Further, the result of the test is independent of the direction of ordering of the data [i.e. whether we consider the cumulative probabilities $P(>x)$ or $P(<x)$].

If we wish to test for differences in a specific property of the observed and hypothetical distributions (e.g. different mean), then there may be more efficient tests available, but even in these cases the gain in efficiency is usually only moderate. Generally, then, the most attractive statistic for goodness-of-fit testing in the one-dimensional case is the Kolmogorov–Smirnov test, and it is natural to ask if this type of test can be applied to data in several dimensions. As opposed to the likelihood, such a generalization of the Kolmogorov–Smirnov statistic is not straightforward.

We could define a cumulative probability distribution $P(X, Y, \ldots)$ as the probability that $x < X, y < Y, \ldots$; a statistic of the Kolmogorov–Smirnov type could certainly be defined as the difference between empirical and theoretical distributions of this kind. The definition of P is ambiguous however, since the *directions* in which we choose to order $x, y, \ldots$ are arbitrary. In one dimension this makes no difference, since $P(>x) = 1 - P(<x)$, but in N dimensions there are $2^N - 1$ independent ways of defining a cumulative probability distribution. The simplest way to deal with this problem so that our Kolmogorov–Smirnov-type statistic is independent of the particular orderings chosen is to specify that the statistic should be the largest difference found between empirical and theoretical cumulative distributions, when *all* possible ordering combinations are considered. Such a test would then retain one of the attractive features of the one-dimensional test. In two dimensions, for example, this corresponds to recognizing that all four quadrants of the plane defined by $(x < X, y < Y)$, $(x < X, y > Y)$, $(x > X, y < Y)$ and $(x > X, y > Y)$ are equally valid areas for the definition of the cumulative probability distribution. The procedure adopted here is to consider each in turn, and adopt the largest of the four differences in empirical and theoretical cumulative distributions as the final statistic.

A more serious difficulty, however, is that any test defined in this way need not be distribution-free. This is easily seen: any transformation of the coordinates $x, y, \ldots$ which preserves the ordering of data points (and hence the test statistic) cannot affect those features of the probability distribution which describe correlations between the coordinates. The proof of the distribution-free nature of the one-dimensional test, on the other hand, depends on being able to map any two density laws on to one another without affecting the test statistic. It is therefore clear that this proof does not apply in more than one dimension. Thus, while a statistic of this type has some distribution-free aspects (the statistic has the same distribution for all density laws which are factorizable – i.e. independent variables, for example), in general the result obtained can depend on the particular density law being considered. From a formal point of view, the test is therefore quite unappealing, since no universal significance levels can be established. Nevertheless, given that no suitable alternative exists, it is worth investigating tests of this sort further. The next section accordingly presents the results of a Monte Carlo investigation of the distribution of such a two-dimensional Kolmogorov–Smirnov statistic; this turns out to be sufficiently nearly distribution-free to be useful in practice.

3 The two-dimensional Kolmogorov–Smirnov test

3.1 PRELIMINARIES

As a guide to the properties we might expect a two-dimensional Kolmogorov–Smirnov test to have, consider the one-dimensional case. The statistic D_n is defined as the largest absolute difference in the empirical and theoretical cumulative probability distributions for a sample of size n. We expect this difference to be $\sim 1/\sqrt{n}$, so that it is more convenient to work with the quantity $Z_n \equiv \sqrt{n} D_n$. For large n, the integral probability distribution of Z is indepen-

dent of n and has the form (e.g. Kendall & Stuart 1946)

$$P(>Z) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2Z^2).$$

The distribution of Z_n for finite n was investigated by Birnbaum (1952), who derived analytic expressions for the cases $n = 2$ and 3 and tabulated values for higher n. The asymptotic and $n = 3$ distributions are shown in Fig. 1. This figure suggests that, over the range of interest ($0.01 \lesssim P(>Z) \lesssim 0.2$), the small-sample distribution may be represented very closely by the asymptotic distribution with Z decreased by a given factor. To investigate this further, we may use Birnbaum's tabulated values; consider Z'_n, the critical value of Z_n for a given significance level, and find the fractional difference from the asymptotic value $\delta \equiv 1 - Z'_n/Z'_\infty$. Fig. 2 shows a plot of $\log \delta$ against $\log n$ for the 5 and 1 per cent significance levels. For $n \gtrsim 5$ both sets of data are well represented by $\delta = 0.2n^{-0.6}$, confirming that the difference between the small-sample and asymptotic cases is very nearly a simple scaling in Z.

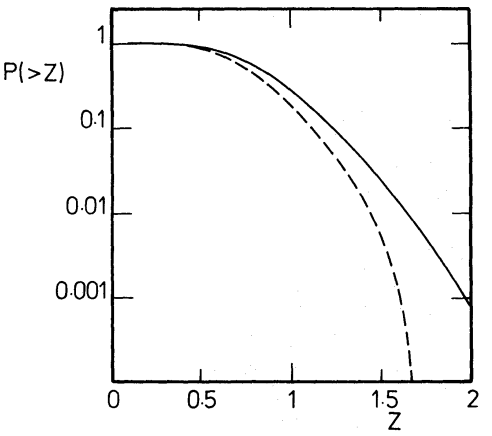

Figure 1. Theoretical $P(>Z)$ curves for the one-dimensional Kolmogorov–Smirnov test for $n = 3$ (– – –) and ∞ (——).

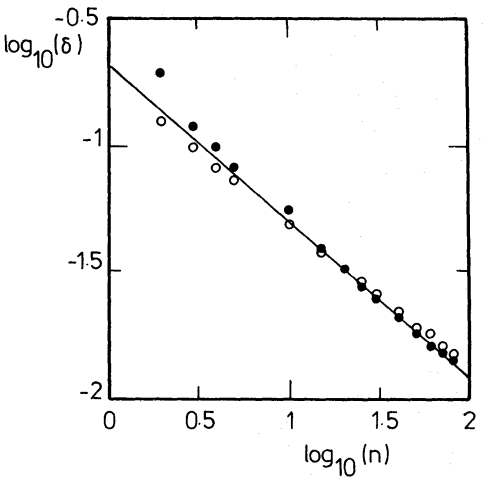

Figure 2. A log–log plot of $\delta (\equiv 1 - Z'_n/Z'_\infty)$ against n for the 5 per cent (indicated by ○) and 1 per cent (indicated by •) points of the one-dimensional Kolmogorov–Smirnov test. The line shown is $\delta = 0.20\, n^{-0.6}$.

3.2 MONTE CARLO RESULTS

For a given two-dimensional density law, the distribution of Z_n may be derived by selecting n points at random from that density law, evaluating Z_n and repeating the process. In practice, 5000 values of Z_n were generated in one simulation: this allows the important significance levels $[P(>Z) \cong 0.05]$ to be estimated with reasonable accuracy (with $P(>Z)$ in error due to Poisson noise by a factor of only ~ 1.05). Most initial simulations used $n = 30$, which yields results very close to the asymptotic values (see Section 3.3). Various density laws were investigated; these are shown in Fig. 3. A deliberate attempt was made to pick highly pathological examples, in the hope of seeing how much effect the particular density law being considered could have on the distribution of Z_n.

Some results are shown in Fig. 4: this gives $P(>Z)$ for the asymptotic one-dimensional case and the $n = 30$ simulated distributions for density laws 1 and 2 – i.e. the cases of (i) totally independent and (ii) perfectly correlated variables. The latter two curves differ, confirming that the test is distribution-dependent. However, the difference is not marked: for a given Z, the significance levels for the two cases differ by less than a factor of 1.5. We might expect that these extremes of correlation would yield extreme forms for $P(>Z)$, and this is borne out dramatically: the $P(>Z)$ curves for density laws 3–9 are *identical*, within the accuracy of the simulation, to that found for law 1 – the uniform distribution on a square. This makes it extremely plausible that all density laws, except those which are highly correlated, will have critical values of Z which are very similar. Of course, no proof of this can be given – but the fact that $P(>Z)$ shows so little variation even over the set of density laws in Fig. 3 makes the conjecture that this approximate invariance will apply in all practical cases seem entirely reasonable.

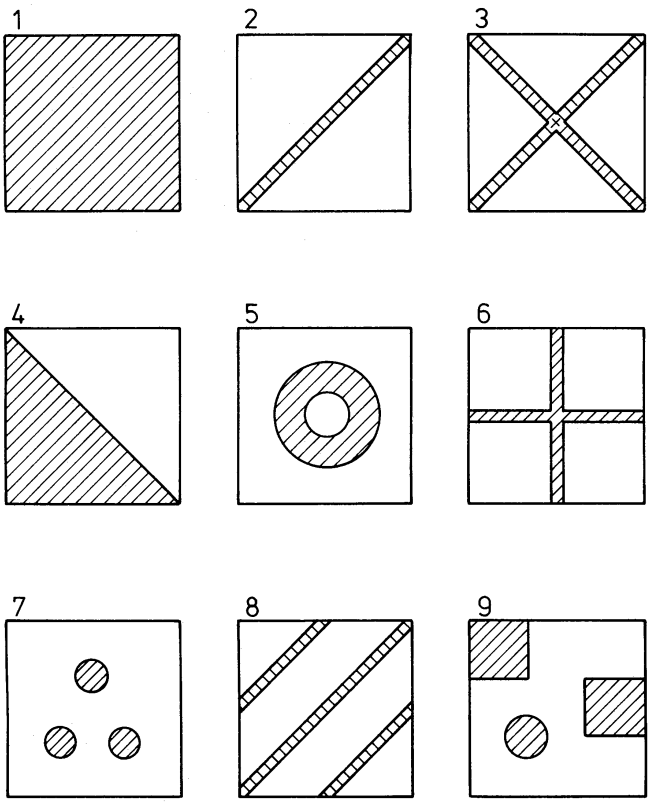

Figure 3. The density laws for which simulations were carried out. Hatched areas are of constant density, blank areas zero density.

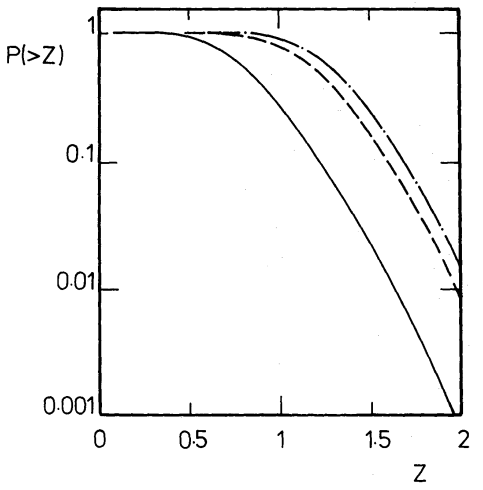

Figure 4. Empirical $P\,(>Z)$ curves for density laws 1 and 2 compared with the asymptotic one-dimensional distribution. Note the similarity in shape. —·—·— law 1, – – – – law 2.

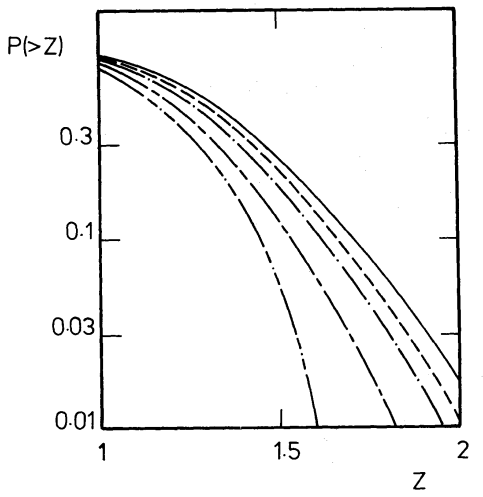

Figure 5. Empirical $P\,(>Z)$ curves for the uniform distribution on a square and $n = 3, 5, 10, 20, 50$. ——— $n = 50$, – – – – $n = 20$, —·—·— $n = 10$, – – – - $n = 5$, —·····— $n = 3$.

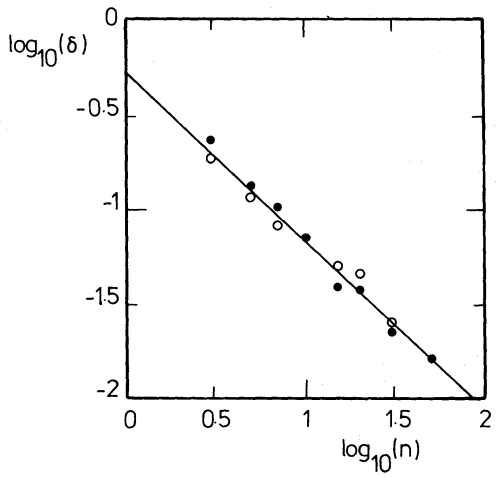

Figure 6. A log–log plot of $\delta\,(\equiv 1 - Z'_n/Z'_\infty)$ against n for the 5 per cent (indicated by ○) and 1 per cent (indicated by ●) points of the two-dimensional test. The line shown is $\delta = 0.53\,n^{-0.9}$.

The conclusion to be drawn from this is the following: if the Z statistic is evaluated for a given density law and the $P(>Z)$ curve for the uniform distribution on a square adopted, then $P(>Z)$ will be greater than the true value by at most ~ 50 per cent. Moreover, for any 'reasonable' (i.e. not strongly correlated) density law, the error in $P(>Z)$ will be much smaller. Obviously, from the point of view of accepting or rejecting a hypothesis, uncertainties of this order are unimportant. The test is therefore one which may be used with confidence in practice, despite its lack of formal elegance.

3.3 BEHAVIOUR FOR FINITE SAMPLES

Restricting attention now to the case of the uniform distribution on a square, we investigate the distribution of the statistic for finite n. Fig. 5 shows $P(>Z)$ for $n = 3, 5, 10, 20, 50$. The behaviour is similar to that in Fig. 1: the distributions are related approximately by a simple scaling in Z. Fig. 6 shows the analogue of Fig. 2: $\log \delta$ against $\log n$ ($\delta \equiv 1 - Z'_n/Z'_\infty$) for the 5 and 1 per cent points. Both sets of points are well fitted by the relation

$$1 - Z'_n/Z'_\infty = 0.53\, n^{-0.9}.$$

This formula may be used to convert a value of Z_n into Z_∞, for estimating $P(>Z)$.

3.4 SIGNIFICANCE LEVELS

In case where the distribution of a test statistic is derived numerically, it is usual to tabulate critical values. Table 1 gives the values of Z for which $(P>Z) = 0.2, 0.1, 0.05, 0.01$, for various values of n, for the case of the uniform distribution on a square. Since this test is not exact, however, it may be more useful to have an analytic approximation. We have seen in Section 3.3 how to convert Z_n to Z_∞; the form for the distribution of Z_∞ is suggested by Fig. 4. From this, it appears that $P(>Z)$ for the two-dimensional case is very similar to the one-dimensional distribution with a constant offset in Z. A detailed comparison reveals difference in the distributions at low Z [$P(>Z) \cong 0.8$], but at higher Z the offset appears

Table 1. Significance levels.

	P(>Z)			
n	0.20	0.10	0.05	0.01
3	1.37	1.44	1.52	1.61
5	1.43	1.53	1.64	1.82
7	1.45	1.59	1.71	1.88
10	1.48	1.62	1.73	1.95
15	1.50	1.65	1.76	2.02
20	1.53	1.67	1.78	2.03
30	1.54	1.69	1.81	2.05
50	1.57	1.70	1.83	2.06

constant. For significance levels of interest [$P(>Z) \lesssim 0.2$], the asymptotic distribution of Z in the two-dimensional case is well-represented by

$$P(>Z_\infty) = 2 \exp\left[-2(Z_\infty - 0.5)^2\right]$$

(*cf.* $P(>Z_\infty) \cong 2 \exp(-2Z_\infty^2)$ for large Z_∞ in the one-dimensional case).

3.5 TWO-SAMPLE TESTS

A final extension of these results is to the case of comparison of two independent samples, rather than testing one sample against a theoretical curve. The one-dimensional case is quite straightforward: we now define D as the maximum absolute difference between the two empirical cumulative probability distributions. The distribution of this statistic differs slightly from the single-sample case, since the statistic now has only a finite number of discrete values possible (see e.g. Conover 1971). But, for cases where the sizes of both samples exceed 10, the confidence levels for the one-sample test are very close to those for the two-sample test (typical error in $P(>D)$ for $n = 10$ is only a factor of 1.1–1.2), provided we now define the Z statistic by

$$Z = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D$$

where n_1 and n_2 are the sizes of the two samples. In other words, provided n_1 and n_2 are both $\gtrsim 10$, the two-sample test may be analysed as a one-sample test with effective sample size $n_1 n_2/(n_1 + n_2)$.

The previously found similarities between the one- and two-dimensional tests strongly suggest that the above result will also apply for a two-dimensional, two-sample test. To verify this conjecture, the distribution of the two-sample, two-dimensional test statistic was investigated by Monte Carlo simulation, in the same way as for the one-sample test.

The results confirm our expectation: even for sizes of the two samples as small as 10, the values of $P(>Z)$ found for the uniform distribution on a square were within a factor of ~ 1.2 of those obtained using the results of Section 3.4, with $n = n_1 n_2/(n_1 + n_2)$. The same degree of approximate distribution-independence was found, with $P(>Z)$ for perfectly correlated variables again approximately 50 per cent smaller than $P(>Z)$ for the uniform distribution on a square. We therefore conclude that, as with the one-sample case, the two-dimensional Kolmogorov–Smirnov test may be used in practice for comparison of two independent samples, without making any important error in the derived significance level.

4 Power-spectrum analysis

The results of Section 3 have provided us with a method whereby the difference between empirical and hypothetical density laws in two dimensions may be quantified, allowing poor hypotheses to be rejected. However, if the value of $P(>Z)$ is high, this does not necessarily mean that the observed data points could reasonably have been selected at random from the hypothetical density law. Formally, the Kolmogorov–Smirnov test compares two hypotheses: H_0 – the data were drawn at random from a given density law; H_1 – the data were drawn at random from some other density law. Thus, even when H_0 is not rejected in favour of H_1, H_0 need not be true. There are two possibilities: either the number of data points is too small to reveal the difference between the true and hypothetical density laws, or the data may not have been selected at random. The first case is straightforward – we have no evidence against H_0 and therefore accept it as a working hypothesis. The second case is more interesting – there are likely to be differences between random and non-random distribu-

tions of data points which obey the same overall density law (consider, e.g. a uniform grid of pairs of points; this is clearly non-random, but would be consistent with a uniform density according to the Kolmogorov–Smirnov test). For many applications, a high value of $P(>Z)$ will be sufficient, since it verifies that the hypothetical law represents well the overall distribution of the data. In cases such as the study of clustering, on the other hand, we are concerned specifically with departures from randomness in the data – a more general analysis than the above is required to test for these.

A suitable technique was described by Webster (1976) – the method of power-spectrum analysis (PSA). This considers the coefficients r_{uv}, which are the amplitudes of terms in the Fourier expansion of the empirical distribution of points on the plane:

$$r_{uv}^2 = |\sum_k \exp 2\pi i (ux_k + vy_k)|^2$$

where $u, v = 0, 1, 2, \ldots$ and the coordinates x and y are normalized to lie in the range 0–1. If the null hypothesis is of a uniform distribution, then the values of r_{uv} are generated by a random walk on the Argand diagram, for which $P(r > R) = \exp(-R^2/n)$ (assuming n, the number of data points, to be large). Thus, deviations from randomness on any scale may be detected by extreme values of r_{uv}. Generally, the power of the test is increased by adding together values of r_{uv}^2 for a range of wavelengths. The individual r_{uv}^2 values are not quite independent, but this was shown by Webster to be unimportant provided the number of terms added together in any one bin was $\ll n$. While this method is more complex than analyses based on a single statistic, it is considerably more general. But the test as presented by Webster deals only with the null hypothesis of a uniform density over a plane; we now consider how to extend PSA so that it may be applied for any density law.

For a given order of Fourier component (u, v), consider the phase factor

$$\phi_k = 2\pi(ux_k + vy_k)$$

where x and y are in the range 0.1. For a uniform plane, ϕ is uniformly distributed between 0 and 2π (values in excess of 2π being expressed modulo 2π), so that the value of the PSA statistic

$$r_{uv}^2 = |\sum_k \exp i\phi_k|^2$$

is determined by a random walk in two dimensions, while if the null hypothesis is of a non-uniform distribution, we have a weighted random walk, with different values of ϕ having probability $f(\phi)$. This process may be analysed by defining variables

$$c = \sum_{k=1}^{n} \cos\phi_k$$

$$s = \sum_{k=1}^{n} \sin\phi_k.$$

Now, knowing $f(\phi)$, we may find the mean, variance and covariance of these quantities:

$$\langle c \rangle \equiv \bar{c} = n \int \cos\phi \, f \, d\phi$$

$$\langle s \rangle \equiv \bar{s} = n \int \sin\phi \, f \, d\phi$$

$$\langle (c - \bar{c})^2 \rangle \equiv \sigma_c^2 = n \int \cos^2\phi \, f \, d\phi - n(\int \cos\phi \, f \, d\phi)^2$$

$$\langle (s - \bar{s})^2 \rangle \equiv \sigma_s^2 = n \int \sin^2\phi \, f \, d\phi - n(\int \sin\phi \, f \, d\phi)^2$$

$$\langle (c - \bar{c})(s - \bar{s}) \rangle \equiv \sigma_{cs}^2 = n \int \sin\phi \cos\phi \, f \, d\phi - n \int \cos\phi \, f \, d\phi \int \sin\phi \, f \, d\phi.$$

For large n, the central limit theorem ensures that the distribution of c and s is jointly normal. We may therefore define a generalized version of I_{uv}, the PSA statistic, for which

$$P(>I_{uv}) = \exp(-I_{uv}/2)$$

$$I_{uv} = \frac{(c-\bar{c})^2\sigma_s^2 + (s-\bar{s})^2\sigma_c^2 - 2(c-\bar{c})(s-\bar{s})\sigma_{cs}^2}{(\sigma_c^2\sigma_s^2 - \sigma_{cs}^2\sigma_{cs}^2)}.$$

Thus, for a given hypothetical density law, we find as above the mean, variance and covariance of the real and imaginary parts (c and s) of the (u, v) component of the Fourier expansion of the data. From these, the generalized PSA statistic I_{uv} is formed, which is distributed like χ^2 for two degrees of freedom.

So far, this is a straightforward extension of the PSA technique, but when we come to add together values of I_{uv} for a given wavelength range, there are additional complications. Webster (1976) showed that the sum of m values of I_{uv} was distributed very nearly as χ^2 with $2m$ degrees of freedom, despite the fact that the I_{uv} are not independent, because the I_{uv} are nevertheless uncorrelated. In the present case, however, it is clear that the degree of correlation of the I_{uv} depends on the hypothetical density law being considered; if this is non-uniform then the I_{uv} *will* be correlated. In practice, this will probably not be very important; for smooth density laws, the function $f(\phi)$ will not show a strong preference for any particular value of ϕ, and any two I_{uv} would thus be approximately uncorrelated. Nevertheless, it is preferable to have a method of analysis which applies in the general case.

Consider calculating m Fourier components in a given range of wavelength for a sample of n data points; For a given value of (u, v) and a given point on the x–y plane, we can find the phase ϕ, enabling the $2m$ values of $\cos\phi$ and $\sin\phi$ for the point (x, y) to be evaluated. We may represent these values as a single point in a $2m$-dimensional space, whose coordinates will be called $w_i\,(i = 1, \ldots 2m)$. The probability density in this space may be found from the hypothetical x–y distribution, since each cell on the x–y plane maps on to an element in w-space. Now, when n independent points in w-space are added to produce the $2m$ values of

$$\sum_k \cos\phi_k$$

and

$$\sum_k \sin\phi_k,$$

the central limit theorem guarantees that the final distribution in w-space will be a multivariate normal distribution for large n:

$$dp \propto \exp\left[-\left(\frac{1}{2}\right)\sum_{i,j} A_{ij}^{-1}(w_i - \bar{w}_i)(w_j - \bar{w}_j)\right]\prod_{i=1}^{2m} dw_i$$

where $\bar{w}_i$ is the mean value of w_i, and A_{ij} is the covariance matrix

$$A_{ij} = \langle(w_i - \bar{w}_i)(w_j - \bar{w}_j)\rangle.$$

The quantities $\bar{w}_i$ and A_{ij} may be found by numerical integration over the hypothetical x–y probability distribution (in the same way as for a single Fourier component; see above). Then, the quantity

$$I = \sum_{i,j} A_{ij}^{-1}(w_i - \bar{w}_i)(w_j - \bar{w}_j)$$

is distributed like χ^2 for $2m$ degrees of freedom. Alternatively, for $m \gtrsim 5$, the variable

$$Q' = I/2m$$

has a distribution which is approximately normal with mean 1 and standard deviation $1/\sqrt{m}$.

This provides us with the analogue to standard PSA for non-uniform distributions which we were seeking. The run of Q' with λ (or $1/\lambda$, which is usually chosen in practice) will show any non-random behaviour of the data points. If the points are in clusters each of q points, with the clusters randomly scattered, then $< Q' > = q$ for all scales greater than the cluster size. Problems concerning the interdependence of individual $I_{u\,v}$ terms now do not arise; the only uncertainty in this analysis is how large n must be for the central limit approximation to be adequate. This of course depends on the probability distribution being considered and also on $2m$, the number of dimensions, but the approximation should be good if $n \gg m$. Since the shortest wavelength of interest is often ~ the interpoint spacing, the maximum number of (u, v) components to be considered is ~ n, and the normal approximation will therefore apply unless the number of bins in $1/\lambda$ is small. Although the above method of analysis is rather more complex than the standard PSA of Webster (1976), it does allow the application of PSA to non-uniform data without the introduction of any additional statistical uncertainties. The required numerical techniques of integration and matrix inversion are sufficiently standard that their introduction into the PSA method should not cause any unreasonable difficulties.

5 Application to quasars in Virgo

As an example of a specific use of the above results in astronomical model-fitting, consider Fig. 7. This shows the distribution of 71 quasar candidates found in the central 5° square of an objective-prism plate of the Virgo cluster (He *et al.* in preparation). The question is, how many of the apparent features of the quasar distribution are likely to be real? Let us first consider the hypothesis that the objects are uniformly scattered on the plate. The value of Z_n for this case is 1.95, which converts to $Z_\infty = 1.97$, for which $P(> Z)$ is 2.7 per cent. We can therefore state that the quasars are non-uniform, with 97.3 per cent confidence. The next most likely hypothesis might be that there is some radial variation of quasar surface density, perhaps corresponding to variations in plate sensitivity. To test this, the variation was assumed to be of the form $\rho \propto \exp(\alpha R)$ where R is the radius from the plate centre, expressed in units of 5°. The maximum-likelihood value of α was -2.6, for which Z_n became 1.61, with $P(> Z) = 0.16$. Hence, the hypothesis that $\rho \propto \exp(-2.6R)$ is not strongly ruled out by the data. Lower values of Z are found by allowing ρ to have a more general two-

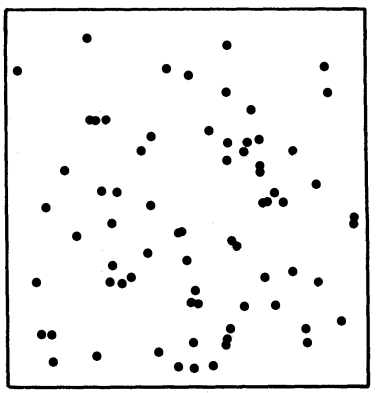

Figure 7. The distribution of quasar candidates over a 5° square covering the Virgo cluster.

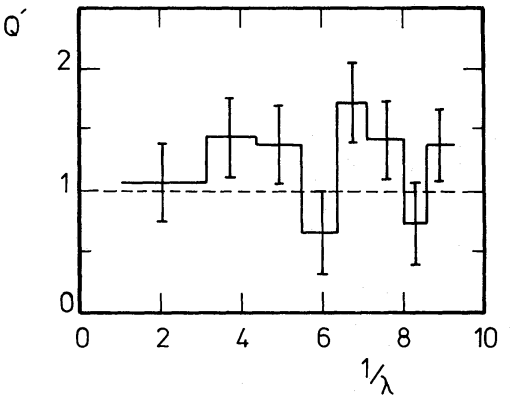

Figure 8. The run of the PSA statistic Q' with (wavelength)$^{-1}$ for the Virgo quasar candidates. Wavelength is in units of 5°.

dimensional variation, but as the hypothesis of a simple radial variation is not rejected, more complex forms for ρ will simply over-fit the data.

The application of the generalized PSA test with the radial form for ρ yields the run of Q' against $1/\lambda$ shown in Fig. 8 (again, λ is in units of 5°). Clearly, there is no compelling evidence for any clustering of the quasar candidates. In this case, the values of Q' are similar to those obtained from the standard analysis, with the exception of the longest-wavelength bin, since ρ contains no short-wavelength terms. However, in many cases we do not have data from a square or rectangular search area to deal with. Where the survey boundary is irregular, the standard PSA can be made to operate only on the largest rectangle contained within the surveyed region, so that data are rejected. The generalized PSA, on the other hand, uses *all* the data and can deal directly with cases such as a survey region containing 'holes' (due to bright stars for example), which previously had to be analysed via Monte Carlo simulation.

We conclude from the present data that, while there is evidence for a general variation of quasar density on the plate, other smaller features (such as the apparent central 'hole') could well have arisen by chance. The smooth form for ρ may now be used as a hypothetical law in other analyses. He *et al.* (in preparation) consider the question of association between the quasar candidates and galaxies on the plate. They find that there is no evidence for association, provided the variation of quasar density is taken into account. The two-dimensional tests were thus of value in allowing a consistent null hypothesis to be established.

6 Summary and conclusions

This paper has presented a two-dimensional analogue of the Kolmogorov–Smirnov test, which is applied as follows.

(1) Find D_n, the maximum absolute difference in the two cumulative probability distributions being compared, all four possible ranking combinations being considered.

(2) Set $Z_n = \sqrt{n} D_n$ and convert Z_n to Z_∞ by $1 - Z_n/Z_\infty = 0.53\, n^{-0.9}$.

(3) Calculate the significance from $P(> Z_\infty) = 2 \exp\left[-2\,(Z_\infty - 0.5)^2\right]$. This probability will be too large by at very most a factor of ~ 1.5 and generally within ~ 1.1.

(4) For a two-sample test, replace n by $n_1 n_2/(n_1 + n_2)$, provided n_1 and $n_2 \gtrsim 10$.

The test statistic is relatively straightforward to compute: $P(x < X, y < Y)$ need be evaluated only where X and Y are values found in the data (not necessarily both from the same data point), since otherwise D_n will not be a maximum. The distribution of the statistic

has been shown plausibly to be nearly distribution-free, so that no errors of practical importance will be made in $P(>Z)$ when the above procedure is applied.

Deviations from randomness in a two-dimensional distribution of data points may be tested for with the more general (but less straightforward) technique of power-spectrum analysis. This method has been extended to deal with distributions which are non-uniform and survey areas which are not rectangular.

These results provide powerful techniques for analysing the distribution of data in two dimensions, and should be useful in a wide range of astronomical problems.

References

Birnbaum, Z. W., 1952. *J. Am. Statist. Ass.*, **47**, 425.
Conover, W. J., 1971. *Practical nonparametric statistics.* Wiley, New York.
Kendall, M. G. & Stuart, A., 1946. *The Advanced Theory of Statistics,* Vol. 2, Griffin, London.
Webster, A. S., 1976. *Mon. Not. R. astr. Soc.*, **175**, 61.