Mon. Not. R. astr. Soc. (1987) **225,** 155–170

A multidimensional version of the Kolmogorov–Smirnov test

G. Fasano *Osservatorio Astronomico, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*

A. Franceschini *Istituto di Astronomia, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy*

Accepted 1986 September 15. Received 1986 September 1; in original form 1985 September 6

Summary. We discuss a generalization of the classical Kolmogorov–Smirnov test, which is suitable to analyse random samples defined in two or three dimensions. This test provides some improvements with respect to an earlier version proposed by Peacock. In particular: (i) it is faster, by a factor equal to the sample size, n, and then usable to analyse quite sizeable samples; (ii) it fully takes into account the dependence of the test statistics on the degree of correlation of data points and on the sample size; (iii) it allows for a generalization to the three-dimensional case which is still viable as regards computing time. Supported by a large number of Monte Carlo simulations, we are ensured that this test is sufficiently distribution-free for any practical purposes. We also give a simple analytic expression to make easier the calculation of the critical values of the test probability distribution.

To illustrate how the test works, we use it to analyse models of the cosmological evolution of X-ray selected active galactic nuclei and we show that it is a much more sensitive goodness-of-fit test than the χ^2.

1 Introduction

It is rather common in astronomy to deal with samples of data points defined in two or more dimensions and limited by poor statistics, in spite of the big observational efforts involved. Binning these data and analysing them with a χ^2 technique or deriving the marginal distributions are known to be quite inefficient procedures. Recently, Peacock (1983) has proposed a two-dimensional version of the classical Kolmogorov–Smirnov (KS) test which, though not of general validity from a formal point of view, may be suitable in two-dimensional cases. The high efficiency of this test follows from the fact that full use is made of the information from each individual data point. However, a potentially serious limitation of this test arises from the over-long computation times required, that often makes its application cumbersome.

Here we propose a generalization of the Kolmogorov–Smirnov test based on a slightly different procedure which is faster (by a factor equal to the sample size, n) than the previous one,

yet maintains the same power, or even sometimes improves on it. This approach fully accounts for the dependence of the test statistics on the degree of correlation of data points and on the sample size n. Finally, we are also allowed to obtain a generalization to the three-dimensional case whose application implies still viable computing times.

Sections 2 and 3 are in particular devoted to the description of the new test. In Section 4 we present a simple generalization to the three-dimensional case. In Section 5 we analyse the two sample test for both two-dimensional and three-dimensional samples. In Section 6 we discuss some properties of the test, in particular its power and consistency. Section 7 contains the conclusions.

2 The two-dimensional KS test for a uniform uncorrelated distribution

2.1 THE PEACOCK'S TEST

As is well known, the classical one-dimensional Kolmogorov–Smirnov test makes use of the probability distribution of the quantity D_{KS}, defined as the largest absolute difference between the cumulative frequency distributions of the parent population and that of an n-point sample extracted from it. Since D_{KS} turns out to be approximately proportional to $1/\sqrt{n}$, one usually refers to the probability distribution of the quantity $Z_n \equiv D_{KS}\sqrt{n}$. For low n, the values of Z_n corresponding to a given significance level SL (let's call them $Z_{n,SL}$) slightly increase with n. For large n, the integral probability distribution $P(>Z_n)$ has the asymptotic expression (see Kendall & Stuart 1979)

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

which is satisfactory for $n \geq 80$.

The one-dimensional nature of the test implies that it does not depend in any way on the shape of the parent distribution. In particular, the Z_n distribution does not depend on which one of the two ways to cumulate the data on the axis is chosen.

Unfortunately, in the case of distributions in more than one dimension, the procedure to cumulate the information on the plane is not unequivocal. For data in two dimensions, the procedure devised by Peacock (1983) makes use of the maximum absolute difference D_{BKS} between the observed and predicted normalized cumulative distributions, when all four possible ways to cumulate data following the directions of the coordinate axes are considered. In practice, the Peacock's test requires that the cumulative distributions of both the observational data and the model function should be calculated in all $4n^2$ quadrants of the plane defined by:

$$(x<X_i, y<Y_j),\quad (x<X_i, y>Y_j),\quad (x>X_i, y<Y_j),\quad (x>X_i, y>Y_j)\qquad (i,j=1,\ldots,n) \tag{1}$$

for all possible combinations of the indices i and j. It is obvious that, unless n is small or the model distribution particularly simple to integrate, the application of the test is rather expensive in terms of computer time. For example, the analysis of a sample of $n=100$ data points on the plane would require $3n^2=30\,000$ double integrations of the model distribution. Of course, the use of two-dimensional interpolating routines for the cumulative distributions would alleviate the problem.

2.2 A NEW VERSION OF THE TWO-DIMENSIONAL KS TEST

In the following we propose a somewhat different generalization of the one-dimensional Kolmogorov–Smirnov test that has the advantage of a simpler and faster application than the

previous one. To this end, we define a statistic which, instead of considering all n^2 points (X_i, Y_j, $i, j=1, n$) of the plane as suitable places to cumulate data points and the model distribution, involves only those n points where a detection is found. We will see in Section 6 that this simplified procedure does not compromise the power and consistency of the test. Operationally, for any data point of coordinates (X_i, Y_i) we sum up observed data and predictions only in the four quadrants of the plane defined by

$$(x<X_i, y<Y_i), \quad (x<X_i, y>Y_i), \quad (x>X_i, y<Y_i), \quad (x>X_i, y>Y_i) \qquad (i=1, \ldots, n). \tag{2}$$

Then we calculate the largest absolute difference of these two quantities (both normalized to 1) within all four quadrants and define D_{BKS} as the largest of these differences when all data points are considered. As in the one-dimensional case, we define $Z_n \equiv D_{\mathrm{BKS}}\sqrt{n}$.

We have studied the statistics of Z_n by means of a Monte Carlo procedure using a uniform distribution within a square as the parent population (see model 1 of Peacock) and extending our analysis to comprise cases with $n=5000$. For any given n, we have produced a large number of simulations, recording for each of them the corresponding Z_n, in order to construct the integral probability distribution $P(>Z_n)$ with sufficient accuracy. Then we have calculated the critical values $Z_{n,\mathrm{SL}}$ corresponding to each relevant confidence level, SL.

The number of simulations produced for each given n varies from a maximum of 100 000 (for $n \leqslant 50$) to a minimum of 500 (for $n=5000$). Fig. 1 (see also Table A1) shows the critical values $Z_{n,\mathrm{SL}}$ resulting from these simulations, as functions of n, together with the polynomial fit that will be described in Appendix A. We notice that for all the values of SL the curves increase with increasing n and show clear signs of convergence only for $n>2000$ (that is approximately the square of the value of n at which the one-dimensional case converges).

Fig. 2 shows the dependence of the critical values $Z_{n,\mathrm{SL}}$ on the significance level SL, for some values of the sample size n.

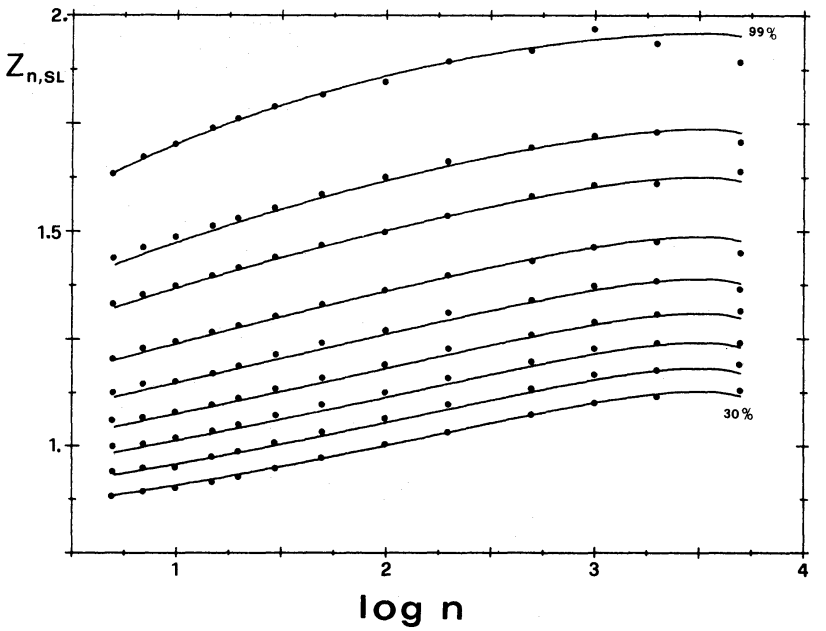

Figure 1. Critical values of the statistic Z_n as function of the sample size, n, and for values of the confidence level SL varying from 30 to 99 per cent (these data are also reported in Table A1). Recall that $SL \equiv 1-P(>Z_n)$. The model used is a uniform uncorrelated distribution on the plane. Curves are a third-order polynomial fit to all our $Z_{n,\mathrm{SL}}$ data, as described in Appendix A. Data for $n=5000$ are subjected to a considerable noise due to the limited number of simulations (see Table A1).

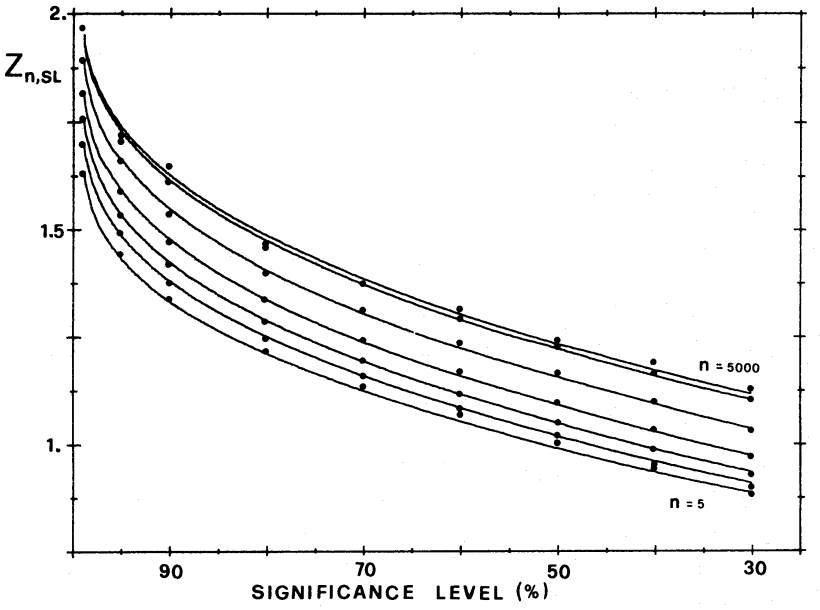

Figure 2. Critical values of Z_n as function of the significance level SL for $n=5, 10, 20, 50, 200, 1000, 5000$. The model is a uniform uncorrelated distribution (see also caption to Fig. 1).

3 Making the two-dimensional KS test distribution-free

The procedure to calculate Z_n described in Section 2.2 is such that, as the correlation of data points in the plane increases, the quadrants which predominantly contribute to the computation of Z_n tend to reduce to 2. For a perfect linear correlation this is rigorously verified and the distribution $P(>Z_n)$ has to reduce to that of the one-dimensional case.

Following this line of reasoning, we have extended our analysis to include parent distributions with correlation coefficients varying from 0 to 1. We have analysed, in particular, three schematic cases of uniform distributions defined within domains of the plane such that, varying some parameter Δ, the correlation coefficient (CC) spans continuously the range 0–1 (see Fig. 3). In

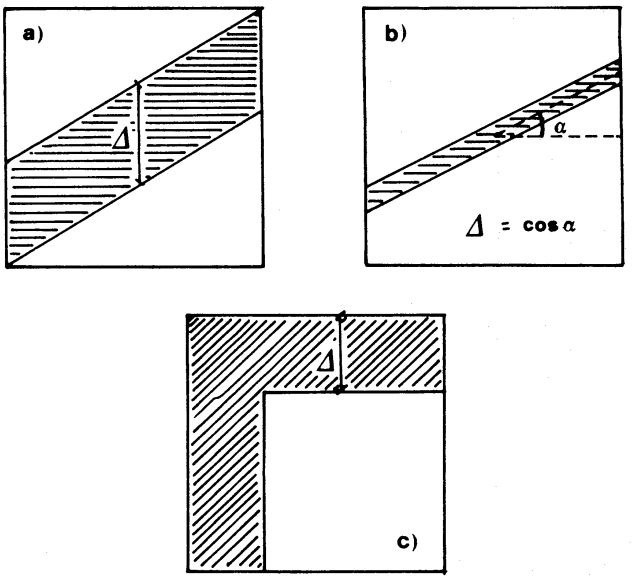

Figure 3. Models of uniform distributions with varying correlation coefficient. The side of the square was normalized to one. (a) The limits of $\Delta \to 0$ and $\Delta \to 1$ correspond to $CC=1$ and $CC=0$, respectively. (b) We have varied both the angle α and the width of the strip. (c) $\Delta \to 0$ and $\Delta \to 1$ give $CC=0$ and $CC=0{,}5$, respectively.

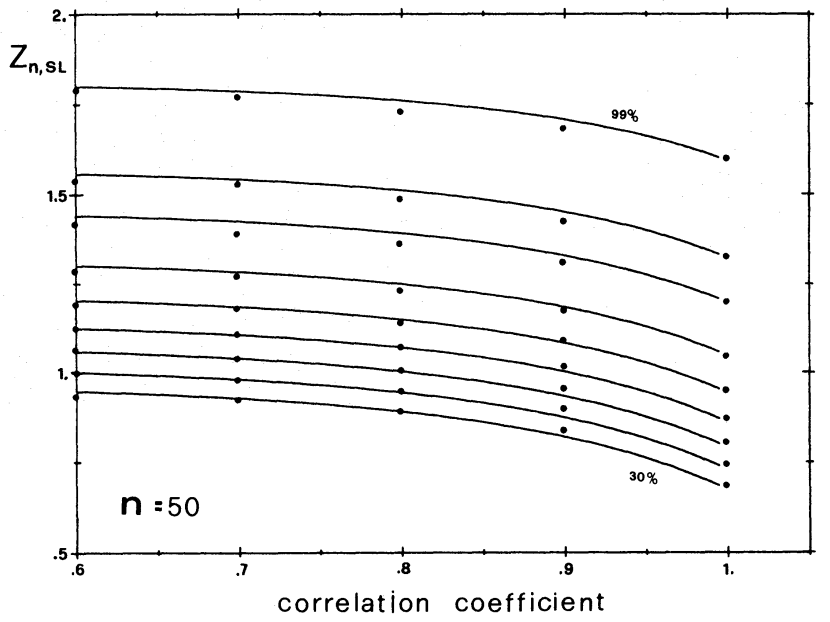

Figure 4. Critical values of Z_n as a function of the correlation coefficient CC for $n=50$. Simulations were based on model (a) of Fig. 3. However, we have found no difference wth the results of any other model for a given value of CC.

addition, to cover a wider range of possibilities, we have also considered more pathological distributions such as those used by Peacock (1983, see his fig. 3) to study the distribution-free nature of his test.

For any given CC value, we have found no differences, within the statistical uncertainties, between the $P(>Z_n)$ distributions corresponding to the above different cases. On the contrary, a marked dependence on the correlation coefficient is evident, at least for values of CC between 0.5 and 1 (see Fig. 4). In particular, the models of cases (a) and (b) in our Fig. 3 for $CC \rightarrow 1$ give values of $Z_{n,SL}$ equal to those of the one-dimensional KS test.

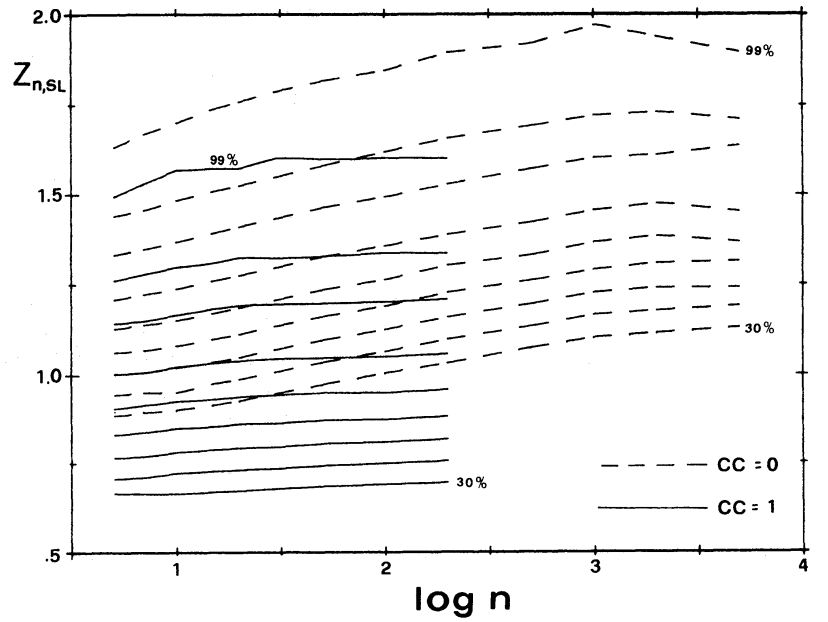

Figure 5. Comparison between the critical values of Z_n obtained by sampling a fully correlated (continuous lines) and an uncorrelated (dashed lines) distribution. $Z_{n,SL}$ values of the perfectly correlated case are not distinguishable from those of the classical one-dimensional test.

In Fig. 5 the perfectly correlated case is compared to the perfectly uncorrelated one, while Fig. 6 shows the results of the simulations for intermediate values of CC (0.6, 0.7, 0.8, 0.9). Further details are given in Appendix A.

We have also considered some cases of non-uniform distributions. In particular, we have analysed bivariate normal distributions having correlation coefficients varying from 0 upwards, and distributions for which there is a gradient of the density along arbitrary directions of the plane. Again, in all such cases the statistics of Z_n, turned out to be dependent on the parameter CC only. In conclusion, once the dependence of $P(>Z_n)$ on the CC is conveniently described, we may be confident that our test is sufficiently distribution-free for any practical applications.

Thus, if we want to test a statistical hypothesis concerning a given sample of data points on the plane, we must simply calculate the quantity Z_n, determine the correlation coefficient of the

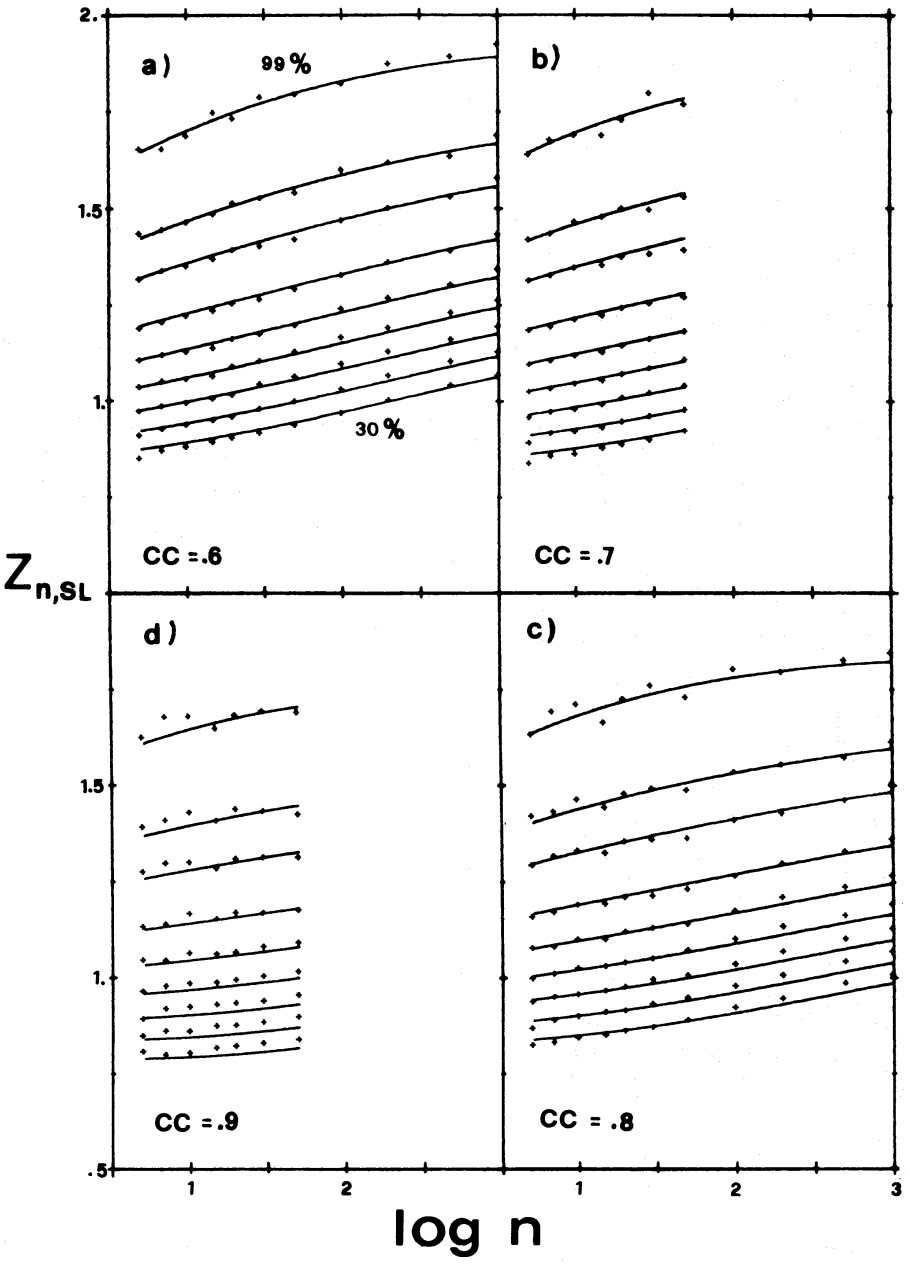

Figure 6. Critical values of Z_n for models with different values of the correlation coefficient CC. (a) CC=0.6. (b) CC=0.7. (c) CC=0.8. (d) CC=0.9.

model distribution, and derive the corresponding significance level at which the hypothesis can be rejected using the tables or figures that refer to the value of CC.

To make the comparison between the measured Z_n and the test critical values $Z_{n,\,\mathrm{SL}}$ easier, we have attempted to represent these latter values as a function of the sample size n, the correlation coefficient CC and the significance level SL, by means of a polynomial expansion of the form

$$Z_{n,\,\mathrm{SL}}=\sum_i\sum_j\sum_k a_{ijk}u^i v^j w^k, \tag{3}$$

where u, v, w are simple functions of n, CC and SL. In Appendix A we show that a third-order polynomial is a good enough representation of the data, and it can be reliably used to derive $Z_{n,\,\mathrm{SL}}$ for any value of n, CC and SL.

Figs 1, 2, 4 and 6 compare the results of this fitting procedure with some sets of data on $Z_{n,\,\mathrm{SL}}$. For high values of the significance level, SL, the fit turned out to be so good that we were tempted to extrapolate equation (3) to $SL>99$ per cent, where in general our data are not reliable due to the finite number of simulations (values of $Z_{n,\,\mathrm{SL}}$ with $SL>99$ per cent are omitted in the Appendix). For a few values of n ($n=5$, 30, 50 and 100) we have verified that this extrapolation gives acceptable fits to $Z_{n,\,\mathrm{SL}}$ up to $SL\simeq 99.7$ per cent, that corresponds to the classical 3σ level.

4 The KS test for three-dimensional samples

Since the computation times needed for the application of the two-dimensional KS test described in the previous sections are fairly moderate, it is now tempting to extend the analysis to the three-dimensional case. Likewise the two-dimensional case, we define a new $Z_{n,\,3\mathrm{D}}$ statistic as the absolute maximum difference (multiplied by $\sqrt{n}$) between the observed and predicted normalized integral distributions cumulated within the eight volumes of the three-dimensional space defined for each data point (X_i, Y_i, Z_i) by

$$(x<X_i, y<Y_i, z<Z_i),\quad \ldots,\quad (x>X_i, y>Y_i, z>Z_i)\qquad (i=1,\ldots,n). \tag{4}$$

To study the probability distribution of $Z_{n,\,3\mathrm{D}}$ we have carried out a set of Monte Carlo simulations using in particular as parent populations a three-dimensional uniform distribution and a trivariate Gaussian with $\sigma_{xx}=\sigma_{yy}=\sigma_{zz}=1$. A complication here comes from the fact that three parameters ($\varrho_{xy}=\sigma_{xy}/\sqrt{\sigma_{xx}\sigma_{yy}}$, ϱ_{xz}, ϱ_{yz}) are required to completely specify the correlation structure of the model distribution. However, we have found that unless two or more variables are very highly correlated just one parameter is enough to specify, as in the two-dimensional case, the probability distribution of $Z_{n,\,3\mathrm{D}}$. We found that the average $\bar{\varrho}$ of the three correlation coefficients ϱ_{xy}, ϱ_{xz} and ϱ_{yz} is suitable to this purpose.

We were able to reconstruct with sufficient accuracy the $Z_{n,\,3\mathrm{D}}$ statistics up to the value of the sample size $n=500$, for all the relevant values of $\bar{\varrho}$ (recall that the computation of Z_n in the three-dimensional case requires only $7n$ integrations of the model distribution). Fig. 7 shows the critical values (for $SL=80$ per cent) of Z_n as a function of the sample size n for some values of $\bar{\varrho}$, compared with the curves corresponding to the degenerate two-dimensional ($\varrho_{xy}=1$) and one-dimensional ($\varrho_{xy}=\varrho_{xz}=\varrho_{yz}=1$) cases.

As a check of consistency, we have found that, for the limit that two or all the three variables are perfectly correlated, the probability distribution of $Z_{n,\,3\mathrm{D}}$ coincides with the probability distributions of the two-dimensional or one-dimensional cases, respectively.

To verify to what extent the three-dimensional test is distribution-free, we have simulated samplings from a number of quite different parent distributions. A convenient way to do that was to place at arbitrary positions in the three-dimensional space a fairly large number of cubes ($\simeq$10 to 30) of different sizes and constant density. The cube was chosen to facilitate the computation of

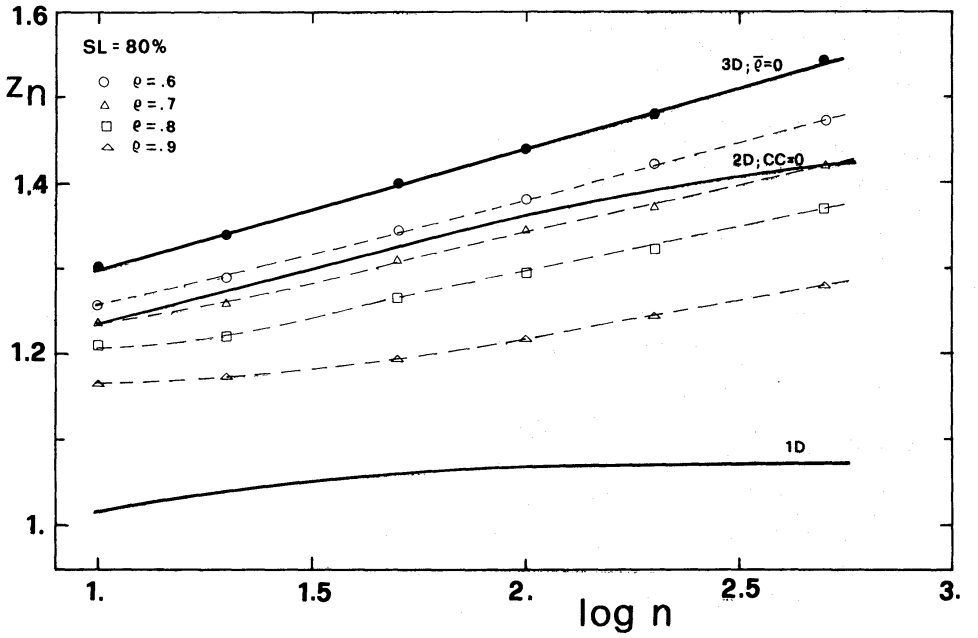

Figure 7. Critical values of the three-dimensional KS test statistics for different values of the average correlation coefficient $\bar{\varrho}$, compared with those of the two- and one-dimensional cases. The heavy lines correspond to the three-, two- and one-dimensional uncorrelated cases. The light dashed lines refer to the three-dimensional statistics with $\bar{\varrho}=0.6$, 0.7, 0.8 and 0.9. The two- and three-dimensional cases are the limiting cases of the three-dimensional statistics at increasing degree of correlation of the samples.

the predicted cumulative distributions. By changing the position and size of each cube, one can easily simulate a number of different distributions including many pathological cases also. It turned out that indeed the key parameter of the statistics is the average correlation coefficient $\bar{\varrho}$. We did not find departures larger than 5 per cent for the critical values $Z_{n,3D}$, unless the correlation coefficient connecting two of the three variables is higher than 0.95: this is the only case in which the above three-dimensional procedure cannot be applied. Note, however, that in this case the two-dimensional (or one-dimensional) statistics can be used.

In Appendix B we report the relevant information needed for the application of the three-dimensional KS test.

5 The two-sample test

The generalization of the two-sample one-dimensional KS test to our multidimensional case is not obvious. The construction of a self-consistent two-sample test for two-dimensional or three-dimensional samples has to satisfy the following three conditions (see also Peacock 1983): (i) the test has to be symmetric with respect to samples 1 and 2 (of sizes n_1 and n_2); (ii) it must reduce to the one-sample case for $n_1 \to \infty$ or $n_2 \to \infty$; (iii) its probability distribution should be easily recovered from that of the one-sample tests. Relying again on a set of Monte Carlo simulations we have found that, if we calculate the maximum absolute difference as the average $\bar{D}_{BKS}$ of the two values of D_{BKS} obtained by cumulating data according to both samples 1 and 2, then the probability distribution $P(>Z)$ of the statistic,

$$Z=\bar{D}_{BKS}\sqrt{n_1 n_2/(n_1+n_2)}, \tag{5}$$

is indistinguishable from the $P(>Z_n)$ distribution of the one-sample case, where $n=n_1 n_2/(n_1+n_2)$.

Therefore, if we need to compare two independent samples defined in two-dimensional or three-dimensional spaces, we can calculate Z as in equation (5) and then derive the significance levels by using as critical values of Z those already obtained for the one-sample cases [with $n=n_1n_2/(n_1+n_2)$] (see Appendices A and B). The only assumption here is that the correlation coefficients CC of the two samples must not be substantially different. However, using the probability distribution of the CC values of a random sample, and using the particular dependence of $Z_{n,\,SL}$ on CC (see Fig. 4), we can show that this uncertainty implies errors on the final determination of SL not larger than the statistical uncertainties ($\approx$5 per cent) due to the limited number of our simulations.

6 Discussion

6.1 COMPARISON BETWEEN OUR TWO-DIMENSIONAL KS TEST AND PEACOCK'S (1983) VERSION

While for the application of our two-dimensional KS test, as discussed in Section 2, it is sufficient to compute the four observed and predicted cumulative distributions at any detected points of coordinates $(X_i, Y_i, i=1, n)$, Peacock's (1983) version requires that these should be computed for each of the n^2 points of coordinates $(X_i, Y_j, i, j=1, n)$. To check whether our simplified procedure is causing any loss of information, we have carried out a set of simulations applying both our test and Peacock's test to several different samples of data points. We have still used a bivariate Gaussian model with varying standard deviation and correlation coefficients. The results show that, within the statistical uncertainties, for uncorrelated distributions there is no difference between the power of the two tests. On the contrary, when the correlation coefficient of the model distribution approaches unity ($CC>0.9$), the power of our two-dimensional test tends to be higher. This is not unexpected, however, if we consider that Peacock's test neglects a residual (weak) dependence on the correlation of the model distribution.

From a more general point of view, since our statistic makes use of the differences between the cumulative distributions calculated *only* at the observed points, one might question the ability of our test to detect, even in the limit $n\to\infty$, a difference between any two arbitrary distributions, that is to say its consistency.* We measure the consistency of the test as follows. In the case where the domains of the model distributions and of the sampled points do not coincide, we must always find at least one data point, for example at the boundary between the two domains, for which the difference between the cumulative distributions (for at least one of the four rankings) is different from zero. As $n\to\infty$, $D_{BKS}\neq 0$ implies $Z_n\to\infty$ and then $SL\to 1$. If instead these two domains coincide it is also obvious that, as $n\to\infty$, the sampled points tend to cover all the domain and any difference between model and parent distributions will be reflected in a difference between the cumulative distributions relative to at least one of the data points and at least one ranking. Further evidence of consistency will be discussed in Section 6.2.

6.2 PROPERTIES OF THE TWO-DIMENSIONAL KS TEST

Kendall & Stuart (1967) report some general results concerning the power of the one-dimensional KS test compared with that of the χ^2 statistics. They conclude that KS is a very much more powerful test than χ^2, even for sample sizes as large as $n=200$. We have no general procedure by which to compare the power of the two-dimensional KS test with that of other goodness-of-fit tests, nor have we done thorough runs of simulations to study it. However, some impression can be gained by working out specific examples.

*Test consistency may also be defined as the ability to reject, as $n\to\infty$, a wrong hypothesis with probability $\to 1$.

The problem under study concerns the cosmological evolution of X-ray selected active galactic nuclei. Two samples exist with complete flux (S_x) and redshift (z) information, with a total of 78 X-ray sources which fall into two nearby domains of the (S_x, z) plane. We have used these data to test a couple of models of evolution, with two different values of the cosmological deceleration parameter (q_0=0.1 and q_0=0.5). Both models allow only sources with X-ray luminosity larger than $10^{44}\,\mathrm{erg\,s^{-1}}$ (2–10 keV band) to undergo luminosity evolution (details can be found in Danese *et al.* 1985). A χ^2 technique is applied to the binned marginal distributions on the S_x and z axes, as opposed to our two-dimensional KS technique that makes use of the information from each of 78 data points in the (S_x, z) plane.

The model with q_0=0.1 gives a total χ^2=20.32 with 22 data bins and five free parameters (thus we must compare this quantity with the critical values of a χ^2 probability distribution with 17 degrees of freedom). This implies that the model can be rejected only at the 75 per cent confidence level. The same model has Z_{78}=1.385 and a correlation coefficient CC=0.73: by using the interpolating expression of equation (A1) we can reject the model at the ≈90 per cent confidence level.

The model with q_0=0.5 gives χ^2=20.35 with the same number of degrees of freedom and can again be rejected only with 75 per cent confidence. In this case Z_{78}=1.64 and CC=0.74, so that this model can be rejected at better than 98 per cent confidence by the two-dimensional KS test.

We have also investigated the properties of the two-dimensional KS as a parameter point estimator and compared it with the widely used maximum likelihood method. We have simulated a repeated sampling from a bivariate normal distribution: for each simulation we have considered as known parameters the means $\bar{x}$=0 and $\bar{y}$=0 and estimated the standard deviation σ by minimizing Z_n and maximizing the likelihood. In addition, we have also kept as a known parameter σ and estimated $\bar{x}=\bar{y}$. Then we have derived the corresponding distributions of the estimates. This analysis has shown that the variance of the two-dimensional KS estimator is usually larger than that of the ML, in spite of a slightly lower bias. In particular, in the case of the estimate of the parameter σ, the ratio of the variance of the two-dimensional KS estimator to that of the ML was found to be larger for low values of the sample size n and to approach unity only for high n: for n=20 we find $\sigma_{KS}/\sigma_{ML}\simeq 2.24$, for n=50 $\sigma_{KS}/\sigma_{ML}\simeq 2.1$ and for n=500 $\sigma_{KS}/\sigma_{ML}\simeq 1.2$. If we add that usually the two-dimensional KS test requires a larger computation time than the ML [$3n$ integrations of the density function $f(x, y)$, instead of n simple determinations of $f(x, y)$], we must conclude that the Maximum Likelihood is to be preferred to the two-dimensional KS as a point estimator.

From the above discussion, we may also get some further indications concerning the consistency of the test. According to Kendall & Stuart (1979), if a test of goodness-of-fit concerning the value of a parameter θ 'is based on a statistic which is a consistent estimator of θ, it is immediately obvious that the test will be consistent too'. On the other hand, a consistent estimator will be characterized by a variance converging to zero as $n\to\infty$. We have seen that, in all the cases explored, the variance of the two-dimensional KS is converging at least as quickly as that of the ML estimator. Since the latter is known to be a consistent estimator, we may infer from that the consistency of the two-dimensional KS at least in all the cases we have directly studied.

We finally mention that the property of the test of being distribution-free allows us to set confidence limits to a distribution as a whole (see Kendall & Stuart 1979, for a similar proposal concerning the one-dimensional test). If $d_{SL}=Z_{n,SL}/\sqrt{n}$ is the critical value of D_{BKS} at the significance level SL and for a sample of size n, we can set up a band of width $\pm d_{SL}$ around the sample distribution, such that it must contain with probability SL the true parent distribution. If we use as reference value of CC that of the sample distribution (but more conservative values can also be used), again, following a reasoning similar to that of Section 5, we can show that the uncertainty on CC cannot invalidate the result.

6.3 THE THREE-DIMENSIONAL KS VERSUS THE TWO-DIMENSIONAL KS TEST

We have also briefly investigated the power of the three-dimensional KS test, compared to that of the two-dimensional KS. Several random samples with different sizes, n, were extracted from a trivariate Gaussian with $\varrho_{xy}=\varrho_{xz}=\varrho_{yz}=0.5$ and analysed on the basis of an uncorrelated ($\varrho_{xy}=\varrho_{xz}=\varrho_{yz}=0$) model. For each value of n we have simulated 30 random samples recording the significance levels at which that model is rejected by the three-dimensional KS test and by the two-dimensional KS test applied to the projections on the three coordinate planes. For samples of $n=50$ data points, we have found an average probability of rejection based on the three-dimensional test of 94 per cent, while the two-dimensional test applied on the (x, y), (x, z) and (y, z) planes gave probabilities of 81, 87 and 85 per cent, respectively. For samples of $n=100$ data points, we found that the three-dimensional test rejects the hypothesis at better than 99 per cent probability, while the two-dimensional test gives probabilities of 95, 97 and 98 per cent. We conclude that the three-dimensional KS test is a safe procedure to testing statistical hypotheses on three-dimensional samples.

7 Conclusions

This paper deals with a generalization of the classical Kolmogorov–Smirnov test to analyse data points in two or three dimensions. This version of the test turns out to be faster (by a factor equal to the size of the samples under study) with respect to an earlier version by Peacock (1983). This allows us to thoroughly explore the test statistics, and, in particular, to fully account for its dependence on the degree of correlation of data points and on the sample size. Then we are also allowed to construct a more general test suitable for analysing data points in three dimensions. As a counter-check, the test statistics are found to converge to the one- and two-dimensional cases for highly correlated data.

We have derived for the two-dimensional case an analytic expression yielding the critical values $Z_{n,\,SL}$ for any value of sample size, significance level and correlation coefficient, within the ranges explored by our simulations. We suggest that this expression extrapolated to values of SL slightly larger than those directly explored does still work.

We have used our test to analyse a particular model of cosmological evolution for the X-ray selected active galactic nuclei and found that it sets much tighter constraints on the models than the χ^2 technique. On the other hand, the test turns out to be usually less efficient than the maximum likelihood method in the parameter estimation.

We still do not have any formal proof of the general validity of the test, but our extensive Monte Carlo simulations make us confident in its applicability to many astrophysical problems.

Acknowledgments

We thank J. Peacock for expanding on his 1983 paper and for helpful criticisms. We are also indebted to L. Danese and G. De Zotti for stimulating discussions and to A. Bianchini for a careful reading of the manuscript. This work would not have been possible without the computing facilities provided by the Faculty of Sciences, University of Padova, through the Centro di Calcolo. The work was supported by the CNR and MPI.

References

Danese, L., De Zotti, G., Fasano, G. & Franceschini, A., 1985. *Astr. Astrophys.*, **161,** 1.
Kendall, M. G. & Stuart, A., 1979. *The Advanced Theory of Statistics,* Griffin, London.
Peacock, J. A., 1983. *Mon. Not. R. astr. Soc.*, **202,** 615.

Appendix A: Relevant data on the two-dimensional KS test

We report here the relevant information required to recover quickly the critical values $Z_{n,\,SL}$ needed for the application of the two-dimensional KS test. As detailed in Section 3, we have performed Monte Carlo simulations based on parent distributions characterized by correlation coefficients CC covering all the range from 0 to 1. Here we report in Tables A1–A5 the results

Table A1. Critical values of Z_n for a uniform uncorrelated distribution.

n[b]	$SL(\%)$[a] # of simul.	30	40	50	60	70	80	90	95	99
5	100000	.883	.943	1.00	1.06	1.13	1.21	1.33	1.44	1.63
7	"	.894	.950	1.01	1.07	1.14	1.22	1.35	1.46	1.67
10	"	.901	.950	1.02	1.08	1.15	1.24	1.37	1.48	1.70
15	"	.915	.975	1.03	1.10	1.17	1.26	1.39	1.51	1.74
20	"	.927	.987	1.05	1.11	1.19	1.27	1.41	1.52	1.76
30	"	.948	1.01	1.07	1.13	1.21	1.30	1.43	1.55	1.79
50	"	.972	1.03	1.09	1.16	1.23	1.33	1.46	1.58	1.82
100	10000	1.00	1.06	1.12	1.19	1.26	1.36	1.49	1.62	1.85
200	5000	1.03	1.10	1.16	1.23	1.30	1.39	1.53	1.66	1.89
500	"	1.07	1.13	1.20	1.26	1.33	1.42	1.57	1.69	1.92
1000	"	1.10	1.17	1.23	1.29	1.37	1.45	1.60	1.72	1.97
2000	1000	1.11	1.18	1.24	1.31	1.38	1.47	1.61	1.73	1.94
5000	500	1.13	1.19	1.24	1.31	1.36	1.45	1.64	1.71	1.89

[a] Significance level ($SL \equiv 1 - P(> Z_n)$); [b] Size of the sample

Table A2. Critical values of Z_n for a uniform distribution with CC=0.6.

$SL(\%)$ / n	30	40	50	60	70	80	90	95	99
5	.851	.911	.974	1.04	1.11	1.19	1.32	1.44	1.65
7	.868	.926	.986	1.05	1.12	1.20	1.34	1.44	1.65
10	.878	.937	.994	1.06	1.13	1.22	1.35	1.46	1.68
15	.891	.949	1.00	1.06	1.14	1.23	1.37	1.48	1.74
20	.902	.958	1.01	1.09	1.16	1.25	1.39	1.51	1.73
30	.914	.978	1.04	1.10	1.17	1.26	1.40	1.52	1.78
50	.935	.997	1.06	1.13	1.19	1.29	1.42	1.54	1.79
100	.967	1.03	1.09	1.16	1.24	1.33	1.47	1.60	1.82
200	1.00	1.06	1.13	1.19	1.27	1.36	1.50	1.62	1.87
500	1.04	1.10	1.16	1.23	1.30	1.39	1.53	1.63	1.89
1000	1.07	1.13	1.19	1.26	1.34	1.43	1.58	1.69	1.92

Table A3. Critical values of Z_n for a uniform distribution with CC=0.7.

$SL(\%)$ / n	30	40	50	60	70	80	90	95	99
5	.839	.892	.959	1.02	1.10	1.18	1.31	1.42	1.64
7	.856	.915	.970	1.03	1.10	1.19	1.32	1.43	1.67
10	.861	.920	.977	1.04	1.12	1.21	1.35	1.46	1.69
15	.876	.930	.990	1.05	1.12	1.22	1.35	1.48	1.69
20	.886	.945	1.01	1.07	1.14	1.24	1.37	1.50	1.73
30	.898	.960	1.02	1.08	1.16	1.25	1.38	1.49	1.80
50	.921	.976	1.04	1.11	1.18	1.27	1.39	1.53	1.77

Table A4. Critical values of Z_n for a uniform distribution with CC=0.8.

SL(%)	30	40	50	60	70	80	90	95	99
n									
5	.825	.869	.939	1.00	1.08	1.16	1.29	1.42	1.63
7	.830	.888	.949	1.01	1.08	1.17	1.32	1.43	1.68
10	.842	.898	.956	1.02	1.10	1.19	1.33	1.46	1.70
15	.849	.909	.966	1.03	1.10	1.19	1.32	1.44	1.66
20	.860	.913	.976	1.04	1.12	1.21	1.35	1.48	1.72
30	.870	.931	.996	1.05	1.13	1.21	1.36	1.49	1.76
50	.889	.947	1.01	1.07	1.14	1.23	1.36	1.49	1.73
100	.922	.979	1.04	1.10	1.17	1.27	1.41	1.53	1.80
200	.946	1.01	1.07	1.13	1.21	1.30	1.43	1.55	1.79
500	.987	1.04	1.10	1.16	1.24	1.33	1.46	1.57	1.83
1000	1.01	1.07	1.13	1.19	1.27	1.36	1.50	1.61	1.84

Table A5. Critical values of Z_n for a uniform distribution with CC=0.9.

SL(%)	30	40	50	60	70	80	90	95	99
n									
5	.808	.848	.893	.965	1.05	1.13	1.28	1.39	1.61
7	.798	.859	.918	.977	1.04	1.14	1.30	1.41	1.67
10	.803	.859	.922	.985	1.06	1.17	1.30	1.43	1.67
15	.816	.873	.929	.987	1.06	1.15	1.28	1.41	1.64
20	.821	.875	.933	.994	1.07	1.17	1.31	1.44	1.68
30	.829	.883	.938	1.00	1.08	1.17	1.31	1.43	1.69
50	.837	.896	.953	1.01	1.09	1.18	1.31	1.42	1.68

Table A6. Coefficients a_{ijk} of the polynomial fit to all data on $Z_{n,\,\mathrm{SL}}$.

	order of term		
i	j	k	a_{ijk}
0	0	0	0.7107
0	0	1	0.9853E-2
0	0	2	0.8561E-2
0	0	3	0.2800E-2
0	1	0	-0.5126
0	1	1	-0.2539
0	1	2	0.3745E-1
0	2	0	-0.3298
0	2	1	-0.1408E-1
0	3	0	-0.7693E-1
1	0	0	-0.6124E-1
1	0	1	-0.1918E-1
1	0	2	0.1064E-1
1	1	0	-0.5042E-1
1	1	1	0.3670E-1
1	2	0	0.7750E-3
2	0	0	0.7775
2	0	1	-0.2322
2	1	0	-0.1273
3	0	0	-0.9915

corresponding to $CC=0, 0.6, 0.7, 0.8, 0.9$. In the limit $CC\rightarrow 1$ the probability distribution of the one-dimensional KS applies. We have omitted to give details for intermediate values of CC, as they do not differ markedly from the case with $CC=0$. From Tables A2–A5 the number of simulations was 5000 for all values of n.

As anticipated in Section 3, we have fitted all our values of $Z_{n,\,\mathrm{SL}}$ with the following polynomial expansion of degree m

$$Z_{n,\,\mathrm{SL}}=\sum_{i=0}^{m}\sum_{j=0}^{(m-i)}\sum_{k=0}^{(m-i-j)} a_{ijk}u^i v^j w^k \tag{A1}$$

with

$$u\equiv -\log[(1.268-CC)/1.41]$$

$$v\equiv \log(100-SL)-2$$

$$w\equiv s[4.989/(4.989-s)]^{-0.09418}$$

$$s\equiv \log(n)+1.074.$$

Here SL is the percentage significance level. We have fitted the relation A1 to the data by means of a χ^2 technique and found that confining ourselves to a third-order expansion ($m=3$) gives a good enough fit. Table A6 reports the corresponding 20 values of the matrix a_{ijk}. The above expressions for the quantities u, v, w and s were chosen so as to make easier the optimization of the coefficients a_{ijk}. The amplitude of the residuals of the best-fit are never larger than the uncertainties in $Z_{n,\,\mathrm{SL}}$ due to the limited number of simulations.

Appendix B: The probability distributions of the three-dimensional KS test

The application of the three-dimensional KS test requires that the statistic $Z_{n,\,3\mathrm{D}}$ should first be calculated according to the prescriptions of Section 4. The second step is to determine the partial correlation coefficients ϱ_{xy}, ϱ_{xz} and ϱ_{yz} of the model distribution. Finally, unless one or more of them is higher than 0.95, one should take the average $\bar{\varrho}$ of the three CC values and then use the corresponding critical values of $Z_{n,\,3\mathrm{D}}$. These critical values are reported in Tables B1–B6 for $\bar{\varrho}=0$, 0.5, 0.6, 0.7, 0.8 and 0.9, respectively. The number of simulations carried out for each value of n is reported in Table B1. We have omitted the intermediate values of $\bar{\varrho}$ as they do not differ markedly from the $\bar{\varrho}=0$ case.

Table B1. Critical values of $Z_{n,\,3\mathrm{D}}$ for a uniform uncorrelated distribution in three dimensions.

n[b]	$SL(\%)$[a] # of simul.	30	40	50	60	70	80	90	95	99
10	10000	.992	1.05	1.10	1.16	1.22	1.30	1.42	1.53	1.75
20	"	1.02	1.07	1.13	1.19	1.26	1.34	1.47	1.57	1.80
50	5000	1.06	1.12	1.18	1.24	1.31	1.39	1.52	1.63	1.84
100	"	1.12	1.17	1.23	1.29	1.36	1.44	1.56	1.68	1.88
200	2000	1.15	1.21	1.27	1.33	1.41	1.48	1.60	1.71	1.88
500	1000	1.22	1.28	1.33	1.39	1.46	1.54	1.66	1.76	1.95

[a] Significance level [$SL\equiv 1-P(>Z_n)$].
[b] Size of the sample.

Table B2. Critical values of $Z_{n,3D}$ for a Gaussian three-dimensional distribution with $\bar{\varrho}=0.5$.

$SL(\%)$	30	40	50	60	70	80	90	95	99
n									
10	.960	1.01	1.07	1.12	1.19	1.27	1.39	1.51	1.74
20	.980	1.03	1.09	1.15	1.22	1.30	1.43	1.55	1.78
50	1.02	1.08	1.14	1.20	1.27	1.37	1.50	1.61	1.86
100	1.07	1.12	1.18	1.24	1.32	1.41	1.54	1.67	1.90
200	1.11	1.17	1.22	1.29	1.36	1.45	1.58	1.70	1.93
500	1.17	1.24	1.29	1.36	1.42	1.50	1.62	1.72	1.94

Table B3. Critical values of $Z_{n,3D}$ for a Gaussian three-dimensional distribution with $\bar{\varrho}=0.6$.

$SL(\%)$	30	40	50	60	70	80	90	95	99
n									
10	.940	.990	1.05	1.11	1.17	1.26	1.38	1.49	1.72
20	.960	1.01	1.07	1.13	1.20	1.29	1.42	1.53	1.78
50	1.00	1.06	1.12	1.18	1.25	1.34	1.47	1.58	1.83
100	1.04	1.10	1.16	1.22	1.30	1.38	1.52	1.63	1.86
200	1.09	1.15	1.20	1.26	1.33	1.42	1.56	1.66	1.91
500	1.15	1.20	1.26	1.32	1.39	1.47	1.61	1.72	1.91

Table B4. Critical values of $Z_{n,3D}$ for a Gaussian three-dimensional distribution with $\bar{\varrho}=0.7$.

$SL(\%)$	30	40	50	60	70	80	90	95	99
n									
10	.914	.970	1.02	1.09	1.15	1.24	1.36	1.48	1.72
20	.927	.985	1.04	1.10	1.17	1.26	1.39	1.51	1.76
50	.970	1.02	1.08	1.14	1.22	1.31	1.44	1.55	1.80
100	1.01	1.06	1.12	1.19	1.25	1.34	1.48	1.59	1.83
200	1.05	1.10	1.16	1.22	1.29	1.37	1.52	1.62	1.86
500	1.11	1.16	1.21	1.27	1.33	1.42	1.54	1.64	1.90

Table B5. Critical values of $Z_{n,3D}$ for a Gaussian three-dimensional distribution with $\bar{\varrho}=0.8$.

$SL(\%)$	30	40	50	60	70	80	90	95	99
n									
10	.890	.942	1.00	1.06	1.12	1.21	1.33	1.45	1.66
20	.893	.946	1.00	1.06	1.13	1.22	1.36	1.48	1.74
50	.929	.982	1.04	1.10	1.17	1.26	1.40	1.52	1.79
100	.960	1.01	1.07	1.13	1.20	1.29	1.43	1.56	1.83
200	.990	1.05	1.16	1.23	1.32	1.32	1.46	1.59	1.83
500	1.06	1.11	1.16	1.22	1.29	1.37	1.48	1.60	1.85

Table B6. Critical values of $Z_{n,3D}$ for a Gaussian three-dimensional distribution with $\bar{\varrho}=0.9$.

$SL(\%)$	30	40	50	60	70	80	90	95	99
n									
10	.846	.895	.953	1.01	1.08	1.16	1.30	1.41	1.63
20	.844	.896	.947	1.01	1.08	1.17	1.31	1.44	1.70
50	.867	.917	.968	1.03	1.10	1.19	1.33	1.47	1.75
100	.890	.943	.997	1.06	1.13	1.22	1.37	1.51	1.77
200	.926	.974	1.03	1.09	1.15	1.24	1.40	1.52	1.83
500	.958	1.01	1.07	1.13	1.19	1.28	1.41	1.53	1.84

If, instead, one or more of the three partial correlation coefficients is $\gtrsim 0.95$, then Tables B1–B5 cannot be used to derive the critical values of $Z_{n,\,3D}$. In this case, however, one of the corresponding variables can be represented as a function of the others and the three-dimensional case degenerates to the two-dimensional (or one-dimensional) ones.