Characterizations, Sub and Resampling, and Goodness of Fit
Author(s): L. Brown, Anirban DasGupta, John Marden, Dimitris Politis
Source: *Lecture Notes-Monograph Series*, Vol. 45, A Festschrift for Herman Rubin (2004), pp. 180-206
Published by: Institute of Mathematical Statistics
Stable URL: http://www.jstor.org/stable/4356309
Accessed: 25/03/2010 16:24

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ims.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *Lecture Notes-Monograph Series.*

http://www.jstor.org

A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 180–206
© Institute of Mathematical Statistics, 2004

Characterizations, Sub and resampling, and goodness of fit

L. Brown[1], Anirban DasGupta[2], John Marden[3] and Dimitris Politis[4]

University of Pennsylvania, Purdue University, University of Illinois at Urbana-Champaign, University of California, San Diego

Abstract: We present a general proposal for testing for goodness of fit, based on resampling and subsampling methods, and illustrate it with graphical and analytical tests for the problems of testing for univariate or multivariate normality. The proposal shows promising, and in some cases dramatic, success in detecting nonnormality. Compared to common competitors, such as a Q-Q plot or a likelihood ratio test against a specified alternative, our proposal seems to be the most useful when the sample size is small, such as 10 or 12, or even very small, such as 6! We also show how our proposal provides tangible information about the nature of the true cdf from which one is sampling. Thus, our proposal also has data analytic value. Although only the normality problem is addressed here, the scope of application of the general proposal should be much broader.

1. Introduction

The purpose of this article is to present a general proposal, based on re or subsampling, for goodness of fit tests and apply it to the problem of testing for univariate or multivariate normality of iid data. Based on the evidence we have accumulated, the proposal seems to have unexpected success. It comes out especially well, relative to its common competitors, when the sample size is small, or even very small. The common tests, graphical or analytical, do not have much credibility for very small sample sizes. For example, a Q-Q plot with a sample of size 6 would be hardly credible; neither would be an analytical test, such as the Shapiro-Wilk, the Anderson-Darling or the Kolmogorov-Smirnov test with estimated parameters (Shapiro and Wilk (1965), Anderson and Darling (1952,1954), Stephens (1976), Babu and Rao (2004)). But, somewhat mysteriously, the tests based on our proposal seem to have impressive detection power even with such small sample sizes. Furthermore, the proposal is general, and so its scope of application is broader than just the normality problem. However, in this article, we choose to investigate only the normality problem in detail, it being the obvious first application one would want to try. Although we have not conducted a complete technical analysis, we still hope that we have presented here a useful set of ideas with broad applicability.

The basic idea is to use a suitably chosen characterization result for the null hypothesis and combine it with the bootstrap or subsampling to produce a goodness

[1]Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Hunstman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340, USA. e-mail: `lbrown@wharton.upenn.edu`

[2]Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2068, USA. e-mail: `dasgupta@stat.purdue.edu`

[3]Department of Statistics, University of Illinois at Urbana-Champaign, 116B Illini Hall, 725 S. Wright St., Champaign, IL 61820, USA. e-mail: `jimarden@uiuc.edu`

[4]Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA. e-mail: `dpolitis@ucsd.edu`

Keywords and phrases: bootstrap, characterization, consistency, goodness of fit, normal, multivariate normal, power, Q-Q plot, scatterplot, subsampling.

AMS 2000 subject classifications: 62G09, 62E10.

of fit test. The idea has been mentioned previously. But it has not been investigated in the way or at length, as we do it here (see McDonald and Katti (1974), Mudholkar, McDermott and Srivastava (1992), Mudholkar, Marchetti and Lin (2002) and D'Agostino and Stephens (1986)). To illustrate the basic idea, it is well known that if $X_1, X_2, \ldots, X_n$ are iid samples from some cdf F on the real line with a finite variance, then F is a normal distribution if and only if the sample mean $\overline{X}$ and the sample variance s^2 are independent, and distributed respectively, as a normal and a (scaled) chisquare. Therefore, using standard notation, with G_m denoting the cdf of a chisquare distribution with m degrees of freedom, the random variables $U_n = \Phi(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma})$ and $V_n = G_{n-1}(\frac{(n-1)s^2}{\sigma^2})$ would be independent $U[0,1]$ random variables. Proxies of U_n, V_n can be computed, in the usual way, by using either a resample (such as the ordinary bootstrap), or a subsample, with some subsample size b. These proxies, namely the pairs, $w_i^* = (U_i^*, V_i^*)$ can then be plotted in the unit square to visually assess evidence of any structured or patterned deviation from a random uniform like scattering. They can also be used to construct formal tests, in addition to graphical tests. The use of the univariate normality problem, and of $\overline{X}$ and s^2 are both artifacts. Other statistics can be used, and in fact we do so (interquartile range/s and s, for instance). We also investigate the multivariate normality problem, which remains to date, a notoriously difficult problem, especially for small sample sizes, the case we most emphasize in this article.

We begin with a quantification of the statistical folklore that Q-Q plots tend to look linear in the central part of the plot for many types of nonnormal data. We present these results on the Q-Q plot for two main reasons. The precise quantifications we give would be surprising to many people; in addition, these results provide a background for why complementary graphical tests, such as the ones we offer, can be useful.

The resampling based graphical tests are presented and analyzed next. A charming property of our resampling based test is that it does not stop at simply detecting nonnormality. It gives substantially more information about the nature of the true cdf from which one is sampling, if it is not a normal cdf. We show how a skillful analysis of the graphical test would produce such useful information by looking at key features of the plots, for instance, empty corners, or a pronounced trend. In this sense, our proposal also has the flavor of being a useful data analytic tool.

Subsampling based tests are presented at the end. But we do not analyze them with as much detail as the resampling based tests. The main reason is limitation of space. But comparison of the resampling based tests and the test based on subsampling reveals quite interesting phenomena. For example, when a structured deviation from a uniform like scattering is seen, the structures are *different* for the re and subsampling based tests. Thus, we seem to have the situation that we do not need to necessarily choose one or the other. The resampling and subsampling based tests complement each other. They can both be used, as alternatives or complements, to common tests, and especially when the sample sizes are small, or even very small.

To summarize, the principal contributions and the salient features of this article are the following:

1. We suggest a flexible general proposal for testing goodness of fit to parametric families based on characterizations of the family;

2. We illustrate the method for the problems of testing univariate and multivariate normality;

3. The method is based on re or subsampling, and tests based on the two methods nicely complement each other;

4. Graphical tests form the core of our proposal, and they are especially useful for small sample sizes due to lack of credible graphical tests when the sample size is small;

5. We give companion formal tests to our graphical tests with some power studies; but the graphical test is more effective in our assessment;

6. We provide a theoretical background for why new graphical tests should be welcome in the area by providing some precise quantifications for just how misleading Q-Q plots can be. The exact results should be surprising to many.

7. We indicate scope of additional applications by discussing three interesting problems.

2. Why Q-Q plots can mislead

The principal contribution of our article is a proposal for new resampling based graphical tests for goodness of fit. Since Q-Q plots are of wide and universal use for that purpose, it would be helpful to explain why we think that alternative graphical tests would be useful, and perhaps even needed. Towards this end, we first provide a few technical results and some numerics to illustrate how Q-Q plots can be misleading. It has been part of the general knowledge and folklore that Q-Q plots can be misleading; but the results below give some precise explanation for and quantification of such misleading behavior of Q-Q plots.

Q-Q plots can mislead because of two reasons. They look approximately linear in the central part for many types of nonnormal data, and because of the common standard we apply to ourselves (and teach students) that we should not overreact to wiggles in the Q-Q plot and what counts is an overall visual impression of linearity. The following results explain why that standard is a dangerous one. First some notation is introduced.

The exact definition of the Q-Q plot varies a little from source to source. For the numerical illustrations, we will define a Q-Q plot as a plot of the pairs $(z_{(i-1/2)/n}, X_{(i)})$, where $z_\alpha = \Phi^{-1}(1-\alpha)$ is the $(1-\alpha)$th quantile of the standard normal distribution and $X_{(i)}$ is the ith sample order statistic (at other places, $z_{(i-1/2)/n}$ is replaced by $z_{(i+1/2)/(n+1)}, z_{(i+1/2)/(n+3/4)}$, etc. Due to the asymptotic nature of our results, these distinctions do not affect the statements of the results). For notational simplicity, we will simply write z_i for $z_{(i-1/2)/n}$. The natural index for visual linearity of the Q-Q plot is the coefficient of correlation

$$r_n = \frac{\sum_{i=1}^n z_i(X_{(i)} - \overline{X})}{\sqrt{\sum_{i=1}^n z_i^2 \sum_{i=1}^n (X_{(i)} - \overline{X})^2}} = \frac{\sum_{i=1}^n z_i X_{(i)}}{\sqrt{\sum_{i=1}^n z_i^2 \sum_{i=1}^n (X_{(i)} - \overline{X})^2}}.$$

As we mentioned above, the central part of a Q-Q plot tends to look approximately linear for many types of nonnormal data. This necessitates another index for linearity of the central part in a Q-Q plot. Thus, for $0 < \alpha < 0.5$, we define the *trimmed correlation*

$$r_\alpha = r_{n,\alpha} = \frac{\sum_{i=k+1}^{n-k} z_i X_{(i)}}{\sqrt{\sum_{i=k+1}^{n-k} z_i^2 \sum_{i=k+1}^{n-k} (X_{(i)} - \overline{X_k})^2}},$$

where $k = [n\alpha]$, and $\overline{X_k}$ is the corresponding trimmed mean. In other words, r_α is the correlation in the Q-Q plot when $100\alpha\%$ of the points are deleted from each tail of the plot. r_α typically is larger in magnitude than r_n, as we shall see below.

We will assume that the true underlying CDF F is continuous, although a number of our results do not require that assumption.

2.1. Almost sure limits of r_n and r_α

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ be iid observations from a CDF F with finite variance σ^2. Then*

$$r_n \to \rho(F) = \frac{\int_0^1 F^{-1}(x)\Phi^{-1}(x)\,dx}{\sigma}$$

with probability 1.

Proof. Multiply the numerator as well as each term within the square-root sign in the denominator by n. The term $\frac{1}{n}\sum_{i=1}^n z_i^2$ converges to $\int_0^1 (\Phi^{-1}(x))^2\,dx$, being a Riemann sum for that integral. The second term $\frac{1}{n}\sum_{i=1}^n (X_{(i)} - \overline{X})^2$ converges a.s. to σ^2 by the usual strong law. Since $\int_0^1 (\Phi^{-1}(x))^2\,dx = 1$, on division by n, the denominator in r_n converges a.s. to σ.

The numerator needs a little work. Using the same notation as in Serfling (1980) (pp. 277–279), define the double sequence $t_{ni} = (i - 1/2)/n$ and $J(t) = \Phi^{-1}(t)$. Thus J is everywhere continuous and satisfies for every $r > 0$ and in particular for $r = 2$, the growth condition $|J(t)| \le M[t(1-t)]^{1/r-1+\delta}$ for some $\delta > 0$. Trivially, $\max_{1\le i\le n} |t_{ni} - i/n| \to 0$. Finally, there exists a positive constant a such that $a.\min_{1\le i\le n}\{i/n, 1 - i/n\} \le t_{ni} \le 1 - a.\min_{1\le i\le n}\{i/n, 1 - i/n\}$. Specifically, this holds with $a = 1/2$. It follows from Example A and Example A^* in pp. 277–279 in Serfling (1980) that on division by n, the numerator of r_n converges a.s. to $\int_0^1 F^{-1}(x)\Phi^{-1}(x)\,dx$, establishing the statement of Theorem 1.

The almost sure limit of the truncated correlation r_α is stated next; we omit its proof as it is very similar to the proof of Theorem 1. □

Theorem 2. *Let $X_1, X_2, \ldots, X_n$ be iid observations from a CDF F. Let $0 < \alpha < 0.5$, and*

$$\mu_\alpha = \frac{\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x\,dF(x)}{1 - 2\alpha}.$$

Then, with probability 1,

$$r_\alpha \to \rho_\alpha(F) = \frac{\int_\alpha^{1-\alpha} F^{-1}(x)\Phi^{-1}(x)\,dx}{\sqrt{\int_\alpha^{1-\alpha}(\Phi^{-1}(x))^2\,dx \cdot \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)}(x-\mu_\alpha)^2\,dF(x)}}.$$

Theorem 1 and 2 are used in the following Table to explain why Q-Q plots show an overall visual linearity for many types of nonnormal data, and especially so in the central part of the plot.

Discussion of Table 1

We see from Table 1 that for each distribution that we tried, the trimmed correlation is larger than the untrimmed one. We also see that as little as 5% trimming from each tail produces a correlation at least as large as .95, even for the extremely skewed Exponential case. For symmetric populations, 5% trimming produces a nearly perfectly linear Q-Q plot, asymptotically. Theorem 1, Theorem 2,

Table 1: Limiting correlation in Q-Q plots.

F	No Trimming	5% trimming
Uniform	.9772	.9949
Double Exp.	.9811	.9941
Logistic	.9663	.9995
t(3)	.9008	.9984
t(5)	.9832	.9991
Tukey distribution	.9706	.9997
		(defined as .9N(0,1) + .1N(0,9))
chisquare(5)	.9577	.9826
Exponential	.9032	.9536

and Table 1 vindicate our common empirical experience that the central part of a Q-Q plot is very likely to look linear for all types of data: light tailed, medium tailed, heavy tailed, symmetric, skewed. Information about nonnormality from a Q-Q plot can only come from the tails and the somewhat pervasive practice of concentrating on the overall linearity and ignoring the wiggles at the tails renders the Q-Q plot substantially useless in detecting nonnormality. Certainly we are not suggesting, and it is not true, that everyone uses the Q-Q plot by concentrating on the central part. Still, these results suggest that alternative or complementary graphical tests can be useful, especially for small sample sizes. A part of our efforts in the rest of this article address that.

3. Resampling based tests for univariate normality

3.1. Test based on $\overline{X}$ and s^2

Let $X_1, X_2, \ldots, X_n$ be iid observations from a $N(\mu, \sigma^2)$ distribution. A well known characterization of the family of normal distributions is that the sample mean $\overline{X}$ and the sample variance s^2 are independently distributed (see Kagan,Linnik and Rao (1973); a good generalization is Parthasarathy (1976). The generalizations due to him can be used for other resampling based tests of normality). If one can test their independence using the sample data, it would in principle provide a means of testing for the normality of the underlying population. But of course to test the independence, we will have to have some idea of the joint distribution of $\overline{X}$ and s^2, and this cannot be done using just one set of sample observations in the standard statistical paradigm. Here is where resampling can be useful.

Thus, for some $B > 1$, let $X_{i1}^*, X_{i2}^*, \ldots, X_{in}^*, i = 1, 2, \ldots, B$ be a sample from the empirical CDF of the original sample values $X_1, X_2, \ldots, X_n$. Define,

$$\overline{X}_i^* = \frac{1}{n}\sum_{j=1}^{n} X_{ij}^*, \qquad \text{and} \qquad s_i^{2*} = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij}^* - \overline{X}_i^*\right)^2.$$

Let Φ denote the standard normal CDF and G_m the CDF of the *chisquare* distribution with m degrees of freedom. Under the null hypothesis of normality, the statistics

$$U_n = \Phi\left(\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}\right) \qquad \text{and} \qquad V_n = G_{n-1}\left(\frac{(n-1)s^2}{\sigma^2}\right)$$

are independently distributed as $U[0, 1]$.

Motivated by this, define: for $i = 1, 2, \ldots, B$,

$$u_i^* = \Phi\left(\frac{\sqrt{n}(\overline{X}_i^* - \overline{X})}{s}\right) \qquad \text{and} \qquad v_i^* = G_{n-1}\left(\frac{(n-1)s_i^{2*}}{s^2}\right).$$

Let $w_i^* = (u_i^*, v_i^*), i = 1, 2, \ldots, B$. If the null hypothesis is true, the w_i^* should be roughly uniformly scattered in the unit square $[0,1] \times [0,1]$. This is the graphical test we propose in this section. A subsampling based test using the same idea will be described in a subsequent section. We will present evidence that this resampling based graphical test is quite effective, and relatively speaking, is more useful for *small* sample sizes. This is because for small n, it is hard to think of other procedures that will have much credibility. For example, if $n = 6$, a case that we present here, it is not very credible to draw a Q-Q plot. Our resampling based test would be more credible for such small sample sizes.

The following consistency theorem shows that our method will correctly identify the joint distribution of (U_n, V_n), asymptotically. Although we use the test in small samples, the consistency theorem still provides some necessary theoretical foundation for our method.

Theorem 3. *Using standard notation,*

$$\sup_{0 \le u \le 1, 0 \le v \le 1} \left| P_*\big(U^* \le u, V^* \le v\big) - P_F\big(U_n \le u, V_n \le v\big) \right| \to 0$$

in probability, provided F has four moments, where F denotes the true CDF from which $X_1, X_2, \ldots, X_n$ are iid observations.

Proof. We observe that the ordinary bootstrap is consistent for the joint distribution of $(\overline{X}, s^2)$ if F has four moments. Theorem 3 follows from this and the uniform *delta* theorem for the bootstrap (see van der Vaart (1998)). □

Under the null hypothesis, (U_n, V_n) are uniformly distributed in the unit square for each n, and hence also asymptotically. We next describe the joint asymptotic distribution of (U_n, V_n) under a general F with four moments. It will follow that our test is not consistent against a specific alternative F if and only if F has the same first four moments as some $N(\mu, \sigma^2)$ distribution. From the point of view of common statistical practice, this is not a major drawback. To have a test consistent against all alternatives, we will have to use more than $\overline{X}$ and s^2.

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ be iid observations from a CDF F with four finite moments. Let μ_3, μ_4 denote the third and the fourth central moment of F, and $\kappa = \frac{\mu_4}{\sigma^4}$. Then,*

$(U_n, V_n) \Rightarrow H$, where H has the density

$$\begin{aligned} h(u,v) \;=\; & \sqrt{\frac{2}{\kappa - 1}} \frac{1}{\sqrt{1 - \frac{\mu_3^2}{(\kappa-1)\sigma^6}}} \exp\Bigg\{ -\frac{1}{2(\mu_3^2 - (\kappa-1)\sigma^6)} \\ & \times \Big[2\sqrt{2}\mu_3 \sigma^3 \Phi^{-1}(u)\Phi^{-1}(v) + (\kappa - 3)\sigma^6 \big(\Phi^{-1}(v)\big)^2 \\ & - \mu_3^2 \big(\big(\Phi^{-1}(u)\big)^2 + \big(\Phi^{-1}(v)\big)^2 \big) \Big] \Bigg\}. \end{aligned} \tag{1}$$

Proof. Let

$$Z_{1n} = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}, \qquad Z_{2n} = \frac{\sqrt{n}(s^2 - \sigma^2)}{\sqrt{\mu_4 - \sigma^4}}.$$

Then, it is well known that $(Z_{1n}, Z_{2n}) \Rightarrow (Z_1, Z_2) \sim N(0, 0, \Sigma)$, where $\Sigma = ((\sigma_{ij}))$, with $\sigma_{11} = 1, \sigma_{12} = \frac{\mu_3}{\sigma^3\sqrt{\kappa-1}}$, and $\sigma_{22} = 1$.

Hence, from the definitions of U_n, V_n, it follows that we only need the joint asymptotic distribution of $(\Phi(Z_{1n}), \Phi(\sqrt{\frac{\kappa-1}{2}} Z_{2n}))$.By the continuity theorem for weak convergence, therefore, $(U_n, V_n) \Rightarrow (\Phi(Z_1), \Phi(\sqrt{\frac{\kappa-1}{2}} Z_2))$. Thus, we need to derive the joint density of $(\Phi(Z_1), \Phi(\sqrt{\frac{\kappa-1}{2}} Z_2))$, which will be our $h(u,v)$.

Let $f(x,y)$ denote the bivariate normal density of (Z_1, Z_2), i.e., let

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)}.$$

Then,

$$\begin{aligned} H(u,v) &= P\left(\Phi(Z_1) \le u, \Phi\left(\sqrt{\frac{\kappa-1}{2}} Z_2\right) \le v\right) \\ &= P\left(Z_1 \le \Phi^{-1}(u), Z_2 \le \sqrt{\frac{2}{\kappa-1}} \Phi^{-1}(v)\right) \\ &= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\sqrt{\frac{2}{\kappa-1}}\Phi^{-1}(v)} f(x,y)\, dy\, dx. \end{aligned}$$

The joint density $h(u,v)$ is obtained by obtaining the mixed partial derivative $\frac{\partial^2}{\partial v \partial u} H(u,v)$. Direct differentiation using the chain rule gives

$$h(u,v) = \sqrt{\frac{2}{\kappa-1}} \frac{1}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} f\left(\Phi^{-1}(u), \sqrt{\frac{2}{\kappa-1}} \Phi^{-1}(v)\right),$$

on some algebra.

From here, the stated formula for $h(u,v)$ follows on some further algebra, which we omit. □

3.2. Learning from the plots

It is clear from the expression for $h(u,v)$ that if the third central moment μ_3 is zero, then U, V are independent; moreover, U is marginally uniform. Thus, intuitively, we may expect that our proposal would have less success for distinguishing normal data from other symmetric data, and more success in detecting nonnormality when the population is skewed. This is in fact true, as we shall later see in our simulations of the test. It would be useful to see the plots of the density $h(u,v)$ for some trial nonnormal distributions, and try to synchronize them with actual simulations of the bootstrapped pairs w_i^*. Such a synchronization would help us learn something about the nature of the true population as opposed to just concluding nonnormality. In this, we have had reasonable success, as we shall again see in our simulations. We remark that this is one reason that knowing the formula in Theorem 4 for the asymptotic density $h(u,v)$ is useful; other uses of knowing the asymptotic density are discussed below.

It is informative to look at a few other summary quantities of the asymptotic density $h(u,v)$ that we can try to synchronize with our plots of the w_i^*. We have in mind summaries that would indicate if we are likely to see an upward or downward trend in the plot under a given specific F, and if we might expect noticeable departures from a uniform scattering such as empty corners. The next two results shed some light on those questions.

Theorem 5. *Let $(U,V) \sim h(u,v)$. Then, $\rho := Corr(U,V)$ has the following values for the corresponding choices of F:*

$\rho \approx .69$ *if* $F =$ Exponential;
$\rho \approx .56$ *if* $F =$ Chisquare(5);
$\rho \approx .44$ *if* $F =$ Beta(2, 6);
$\rho \approx .50$ *if* $F =$ Beta(2, 10);
$\rho \approx .53$ *if* $F =$ Poisson(1);
$\rho \approx .28$ *if* $F =$ Poisson(5).

The values of ρ stated above follow by using the formula for $h(u, v)$ and doing the requisite expectation calculations by a two dimensional numerical integration.

A discussion of the utility of knowing the asymptotic correlations will follow the next theorem.

Theorem 6. *Let* $p_{11} = P(U \leq .2, V \leq .2)$, $p_{12} = P(U \leq .2, V \geq .8)$, $p_{13} = P(U \geq .8, V \leq .2)$ *and* $p_{14} = P(U \geq .8, V \geq .8)$.

Then, $p_{11} = p_{12} = p_{13} = p_{14} = .04$ *if* $F =$ Normal;
$p_{11} = .024$, $p_{12} = .064$, $p_{13} = .0255$, $p_{14} = .068$ *if* $F =$ Double Exponential;
$p_{11} = .023$, $p_{12} = .067$, $p_{13} = .024$, $p_{14} = .071$ *if* $F =$ t(5);
$p_{11} = .01$, $p_{12} = .02$, $p_{13} = .01$, $p_{14} = .02$ *if* $F =$ Uniform;
$p_{11} = .04$, $p_{12} = .008$, $p_{13} = .004$, $p_{14} = .148$ *if* $F =$ Exponential;
$p_{11} = .04$, $p_{12} = .012$, $p_{13} = .006$, $p_{14} = .097$ *if* $F =$ Beta(2, 6);
$p_{11} = .045$, $p_{12} = .01$, $p_{13} = .005$, $p_{14} = .117$ *if* $F =$ Beta(2, 10).

Proof. Again, the values stated in the Theorem are obtained by using the formula for $h(u, v)$ and doing the required numerical integrations. □

3.3. Synchronization of theorems and plots

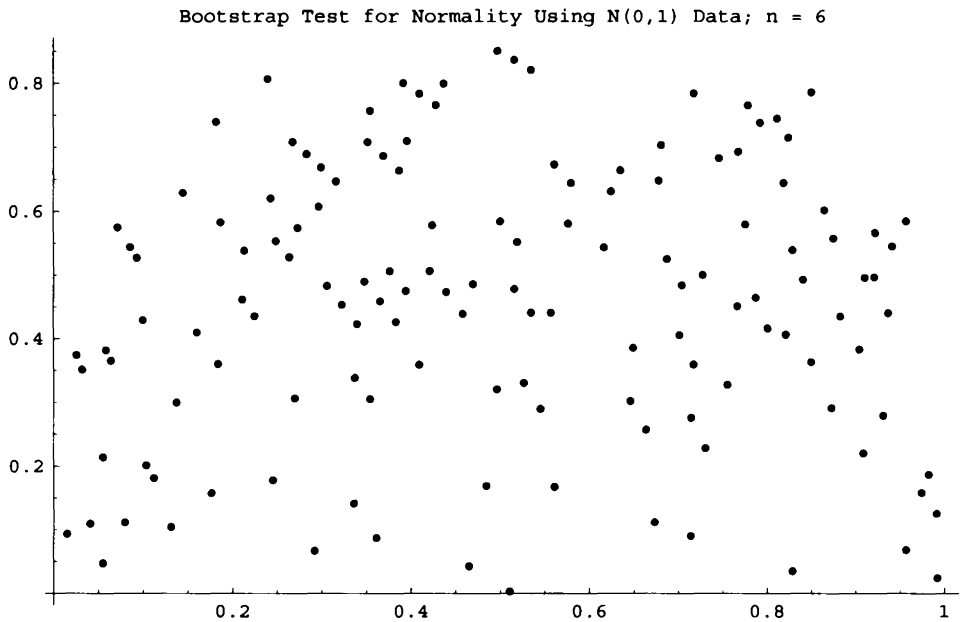
Together, Theorem 5 and Theorem 6 have the potential of giving useful information about the nature of the true CDF F from which one is sampling, by inspecting the cloud of the w_i^* and comparing certain features of the cloud with the general pattern of the numbers quoted in Theorems 5 and 6. Here are some main points.

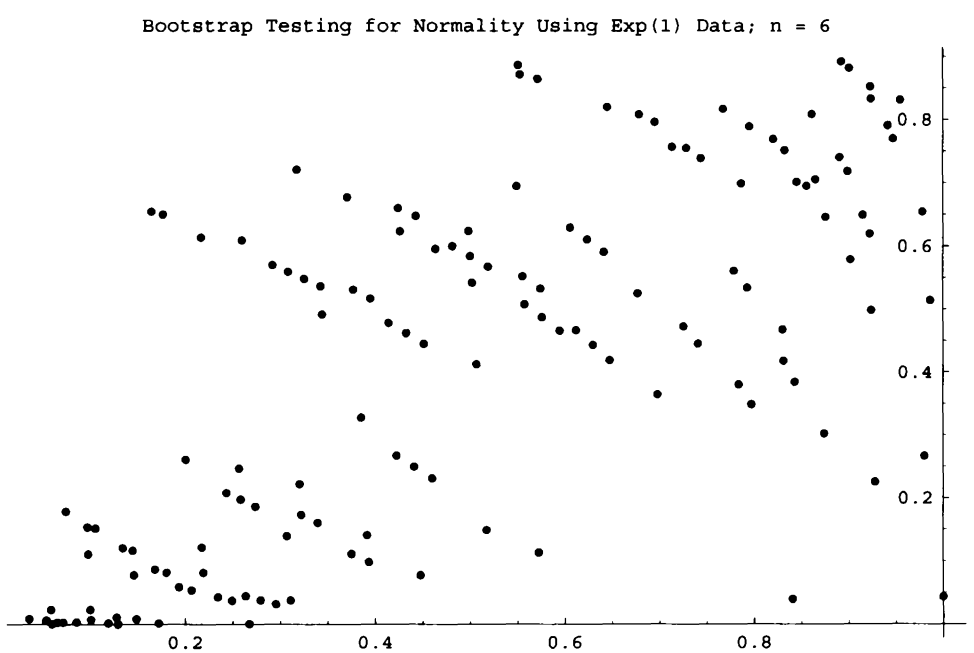
1. A pronounced upward trend in the w_i^* cloud would indicate a right skewed population (such as Exponential or a small degree of freedom chisquare or a right skewed Beta, etc.), while a mild upward trend may be indicative of a population slightly right skewed, such as a Poisson with a moderately large mean.

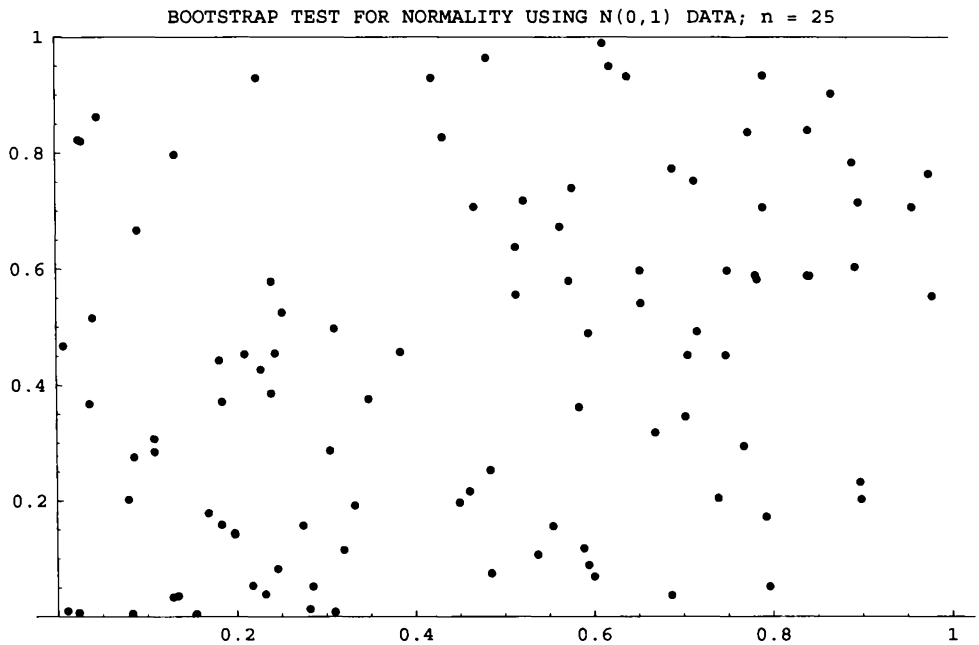
2. To make a finer distinction, Theorem 6 can be useful. $p_{11}, p_{12}, p_{13}, p_{14}$ respectively measure the density of the points in the lower left, upper left, lower right, and the upper right corner of the w_i^* cloud. From Theorem 6 we learn that for right skewed populations, the upper left and the lower right corners should be rather empty, while the upper right corner should be relatively much more crowded. This is rather interesting, and consistent with the correlation information provided by Theorem 5 too.

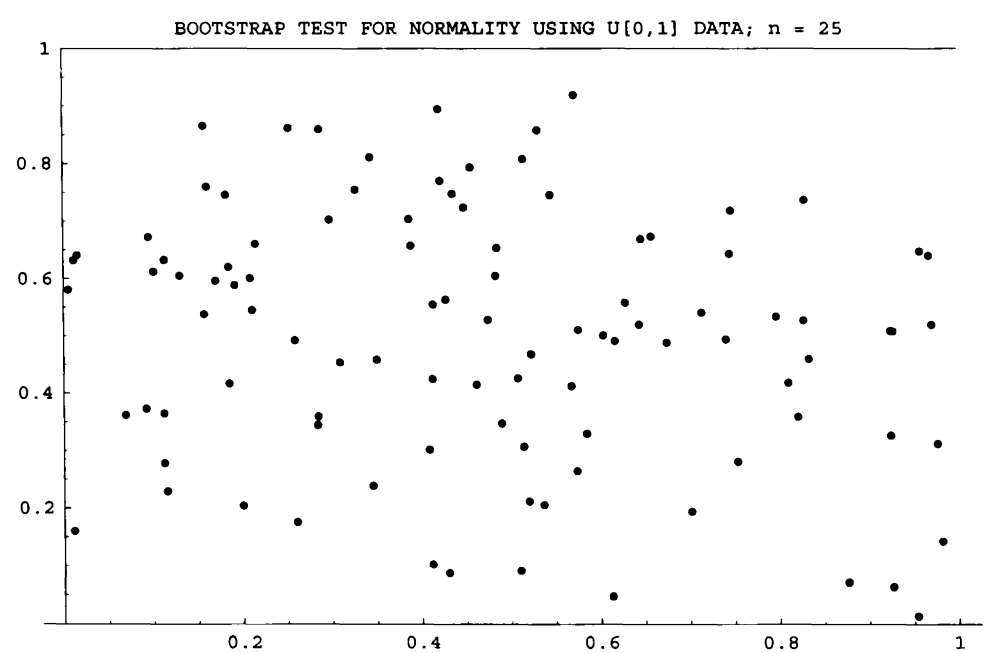
3. In contrast, for symmetric heavy tailed populations, the two upper corners should be relatively more crowded compared to the two lower corners, as we can see from the numbers obtained in Theorem 6 for Double Exponential and t(5) distributions. For uniform data, all four corners should be about equally dense, with a general sparsity of points in all four corners. In our opinion, these conclusions that one can draw from Theorems 5 and 6 together about the nature of the true CDF are potentially quite useful.

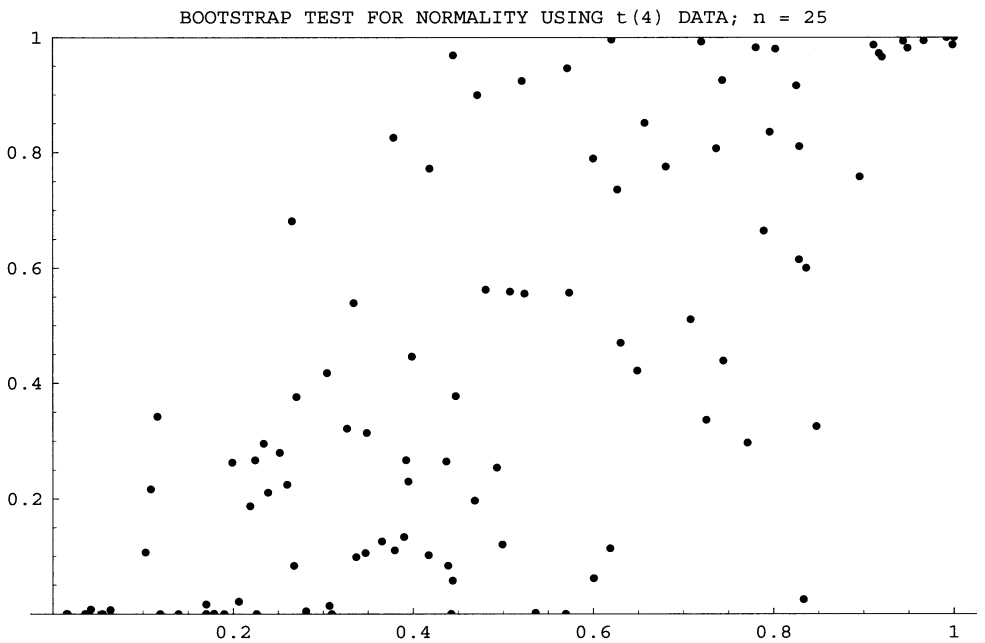
We next present a selection of scatterplots corresponding to our test above. Due to reasons of space, we are unable to present all the plots we have. The plots we present characterize what we saw in our plots typically; the resample size B varies

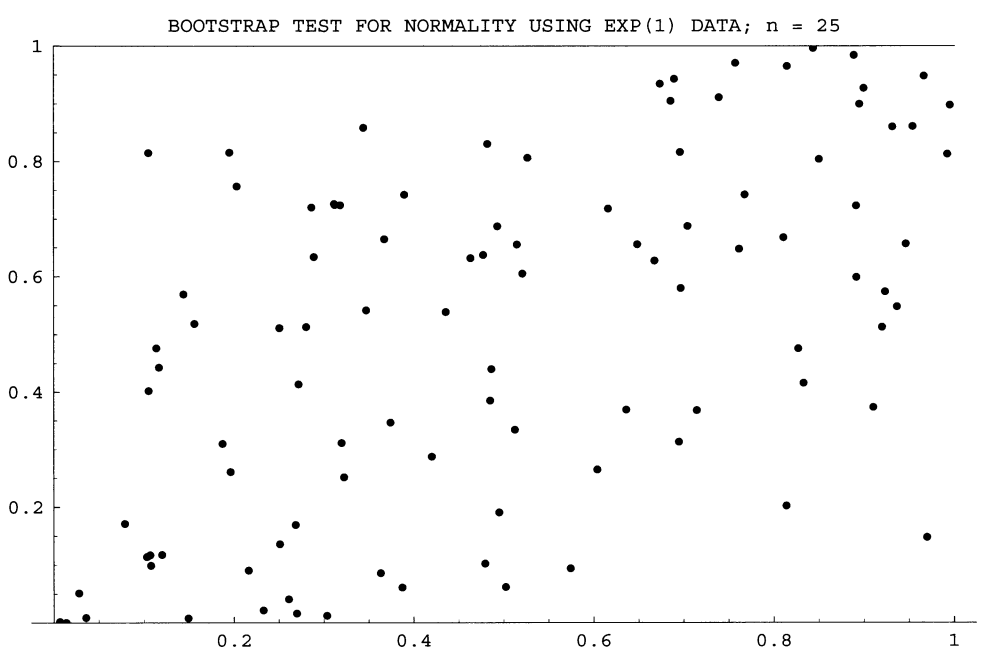

Bootstrap Test for Normality Using N(0,1) Data; n = 6
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


Bootstrap Testing for Normality Using Exp(1) Data; n = 6
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8


BOOTSTRAP TEST FOR NORMALITY USING N(0,1) DATA; n = 25
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


BOOTSTRAP TEST FOR NORMALITY USING U[0,1] DATA; n = 25
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


BOOTSTRAP TEST FOR NORMALITY USING t(4) DATA; n = 25
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


BOOTSTRAP TEST FOR NORMALITY USING EXP(1) DATA; n = 25
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1

between 100 and 200 in the plots. The main conclusions we draw from our plots are summarized in the following discussion.

The most dramatic aspect of these plots is the transparent structure in the plots for the right skewed Exponential case for the extremely small sample size of $n = 6$. We also see satisfactory agreement as regards the density of points at the corners with the statements in Theorem 6. Note the relatively empty upper left and lower right corners in the Exponential plot, as Theorem 6 predicts, and the general sparsity of points in all the corners in the uniform case, also as Theorem 6 predicts. The plot for the t case shows mixed success; the very empty upper left corner is not predicted by Theorem 6. However, the plot itself looks very nonuniform in the unit square, and in that sense the t(4) plot can be regarded as a success. To summarize, certain predictions of Theorems 5 and 6 manifest reasonably in these plots, which is reassuring.

The three dimensional plots of the asymptotic density function $h(u, v)$ are also presented next for the uniform, t(5), and the Exponential case, for completeness and better understanding.

3.4. Comparative power and a formal test

While graphical tests have a simple appeal and are preferred by some, a formal test is more objective. We will offer some in this subsection; however, for the kinds of small sample sizes we are emphasizing, the chi-square approximation is not good. The correct percentiles needed for an accurate application of the formal test would require numerical evaluation. In the power table reported below, that was done.

The formal test

The test is a standard chisquare test. Partition the unit square into subrectangles $[a_i, b_j]$, where $a_i = b_i = .2i$, and let in a collection of B points, O_{ij} be the observed number of pairs w^* in the subrectangle $[a_i, b_j]$. The expected number of points in each subrectangle is $.04B$. Thus, the test is as follows:

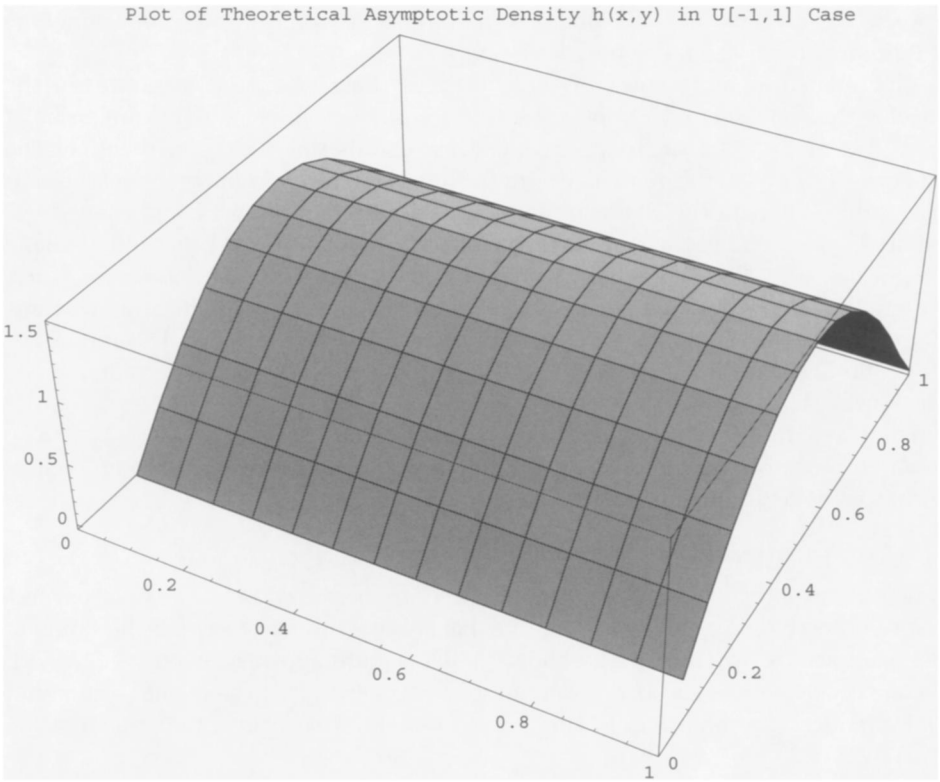
Calculate $\chi^2 = \sum \frac{(O_{ij} - .04B)^2}{.04B}$ and find the P-value $P(\chi^2(24) > \chi^2)$.

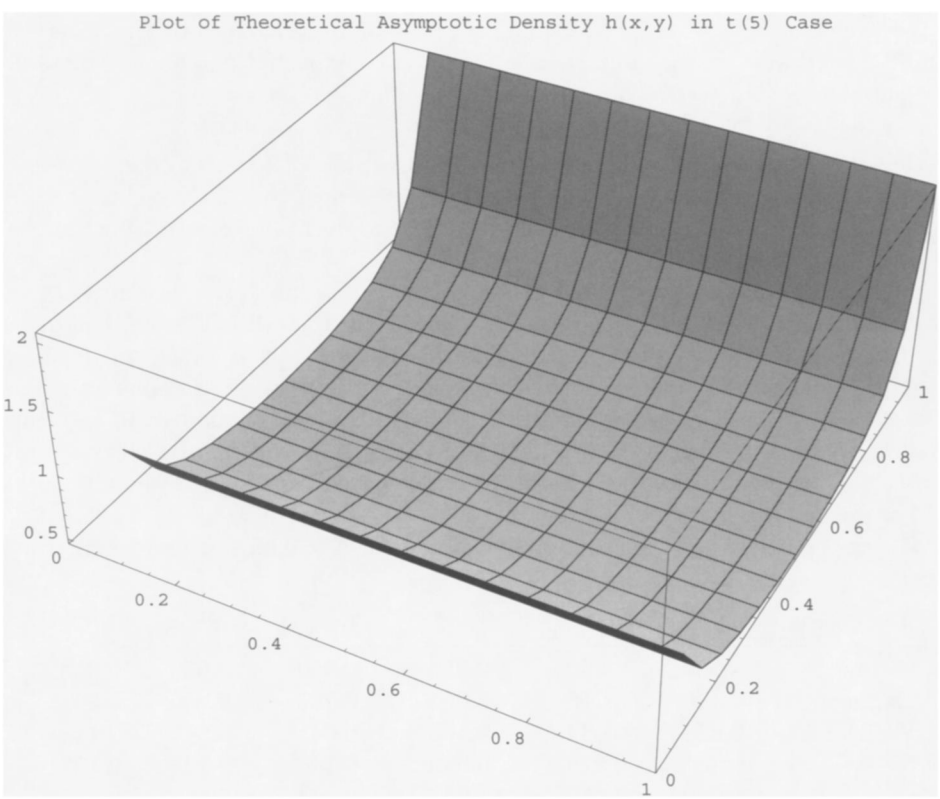
How does the test perform? One way to address the issue is to see whether a test statistic based on the plot has reasonable power. It is clear that the plot-based tests cannot be more powerful than the best test (for a given alternative), but maybe they can be competitive.

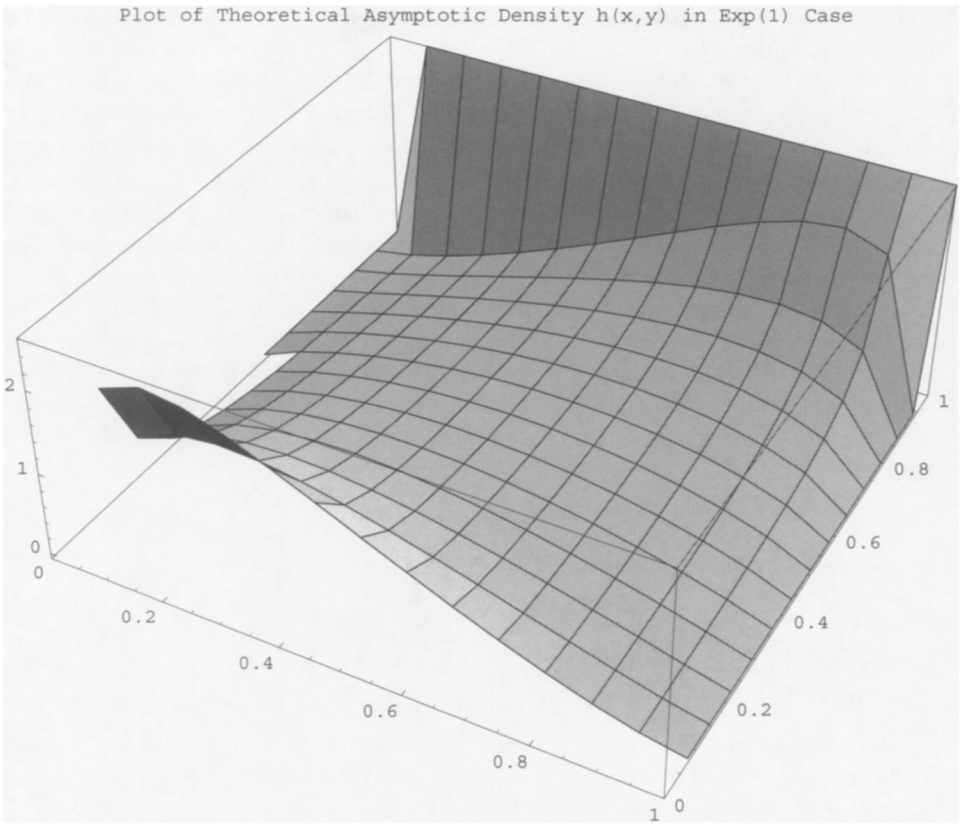
We take the best test to be the likelihood ratio test for testing the alternative versus the normal, using the location-scale family for each distribution. The plot-based tests include the χ^2 test in the paper, two based on the $MAD(v_i^*)$ (median absolute deviation of the v_i^*'s), one which rejects for large values and one for small values, and two based on $Correlation(u_i^*, v_i^*)$. Note the likelihood ratio test can *only* be used when there is a specified alternative, but the plot-based tests are *omnibus*. Thus, what counts is whether the plot-based tests show some all round good performance.

The tables below have the estimated powers (for $\alpha = 0.05$) for various alternatives, for $n = 6$ and 25.

$n = 6$	χ^2	MAD(>)	MAD(<)	Corr(>)	Corr(<)	LRT
Normal	0.050	0.050	0.050	0.050	0.050	0.050
Exponential	0.176	0.075	0.064	0.293	0.006	0.344
Uniform	0.048	0.033	0.105	0.041	0.044	0.118
t_2	0.185	0.079	0.036	0.146	0.138	0.197
t_5	0.070	0.059	0.043	0.064	0.067	0.089


Plot of Theoretical Asymptotic Density h(x,y) in U[-1,1] Case
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1


Plot of Theoretical Asymptotic Density h(x,y) in t(5) Case
2
1.5
1
0.5
0
0.2
0.4
0.6
0.8
1
0
0.2
0.4
0.6
0.8
1



$n = 25$	χ^2	MAD(>)	MAD(<)	Corr(>)	Corr(<)	LRT
Normal	0.050	0.050	0.050	0.050	0.050	0.050
Exponential	0.821	0.469	0.022	0.930	0.000	0.989
Uniform	0.164	0.000	0.506	0.045	0.038	0.690
t_2	0.553	0.635	0.003	0.261	0.264	0.721
t_5	0.179	0.208	0.011	0.104	0.121	0.289

The powers for $n = 6$ are naturally fairly low, but we can see that for each distribution, there is a plot-based test that comes reasonably close to the LRT. For the Exponential, the correlation (>) test does very well. For the uniform, the best test rejects for small values of MAD. For the t's, rejecting for large values of MAD works reasonably well, and the χ^2 and two correlation tests do fine. These results are consistent with the plots in the paper, i.e., for skewed distributions there is a positive correlation between the u_i^*'s and v_i^*'s, and for symmetric distributions, the differences are revealed in the spread of the v_i^*'s . On balance, the Corr(>) test for suspected right skewed cases and the χ^2 test for heavy-tailed symmetric cases seem to be good plot-based formal tests. However, further numerical power studies will be necessary to confirm these recommendations.

3.5. Another pair of statistics

One of the strengths of our approach is that the pair of statistics that can be used to define U_n, V_n is flexible, and therefore different tests can be used to test for normality. We now describe an alternative test based on another pair of statistics.

It too shows impressive power in our simulations in detecting right skewed data for quite small sample sizes.

Let $X_1, X_2, \ldots, X_n$ be the sample values and let Q, s denote respectively the interquartile range and the standard deviation of the data. From Basu's theorem (Basu (1955)), $\frac{Q}{s}$ and s are independent if $X_1, X_2, \ldots, X_n$ are samples from any normal distribution. The exact distribution of $\frac{Q}{s}$ in finite samples is cumbersome. So in forming the quantile transformations, we use the asymptotic distribution of $\frac{Q}{s}$. This is, admittedly, a compromise. But at the end, the test we propose still works very well at least for right skewed alternatives. So the compromise is not a serious drawback at least in some applications, and one has no good alternative to using the asymptotic distribution of $\frac{Q}{s}$. The asymptotic distribution of $\frac{Q}{s}$ for any population F with four moments is explicitly worked out in DasGupta and Haff (2003). In particular, they give the following results for the normal, Exponential and the Beta(2,10) case, the three cases we present here as illustration of the power of this test.

(a) $\sqrt{n}(\frac{IQR}{s} - 1.349) \Rightarrow N(0, 1.566)$ if F = normal;

(b) $\sqrt{n}(\frac{IQR}{s} - 1.099) \Rightarrow N(0, 3.060)$ if F = Exponential.

(c) $\sqrt{n}(\frac{IQR}{s} - 1.345) \Rightarrow N(0, 1.933)$ if F = Beta(2,10).

Hence, as in Subsection 3.1, define:

$u_i^* = \Phi(\frac{\sqrt{n}}{\tau}(\frac{Q_i^*}{s_i^*} - \frac{Q}{s}))$, and $v_i^* = G_{n-1}((n-1)\frac{s_i^{2*}}{s^2})$ and $w_i^* = (u_i^*, v_i^*)$; note that τ^2 is the appropriate variance of the limiting normal distribution of $\frac{IQR}{s}$ as we indicate above. As in Subsection 3.1, we then plot the pairs w_i^* and check for an approximately uniform scattering, *particularly lack of any striking structure.*

The plots below are for the normal, Exponential and Beta(2,10) case; the last two were chosen because we are particularly interested in establishing the efficacy of our procedures for picking up skewed alternatives. It is clear from the plots that for the skewed cases, even at a small sample size $n = 12$, they show striking visual structure, far removed from an approximately uniform scattering. In contrast, the plot for the normal data look much more uniform.

Exactly as in Subsection 3.1, there are analogs of Theorem 3 and Theorem 4 for this case too; however, we will not present them.

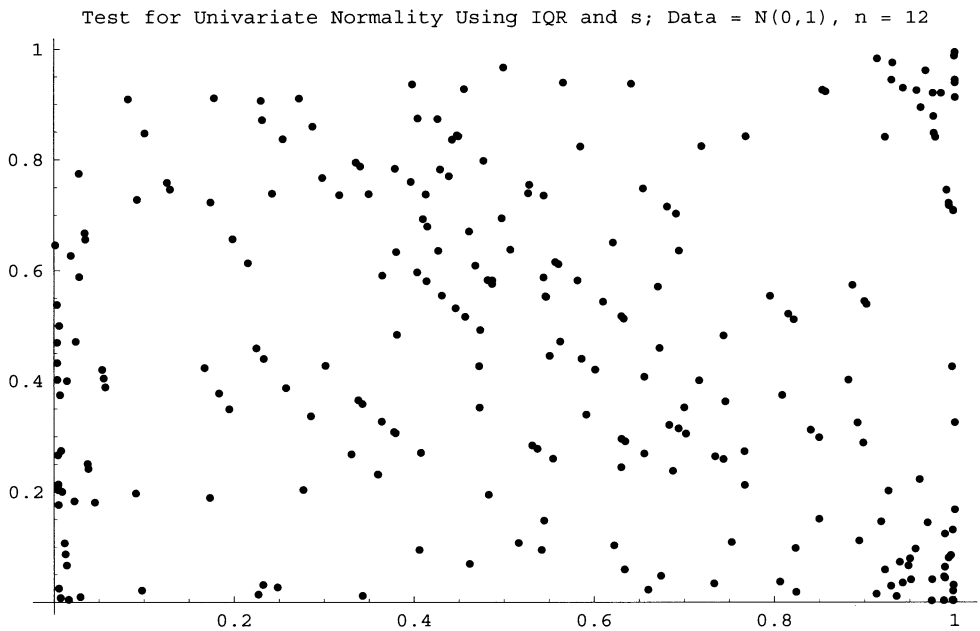
We now address the multivariate case briefly.

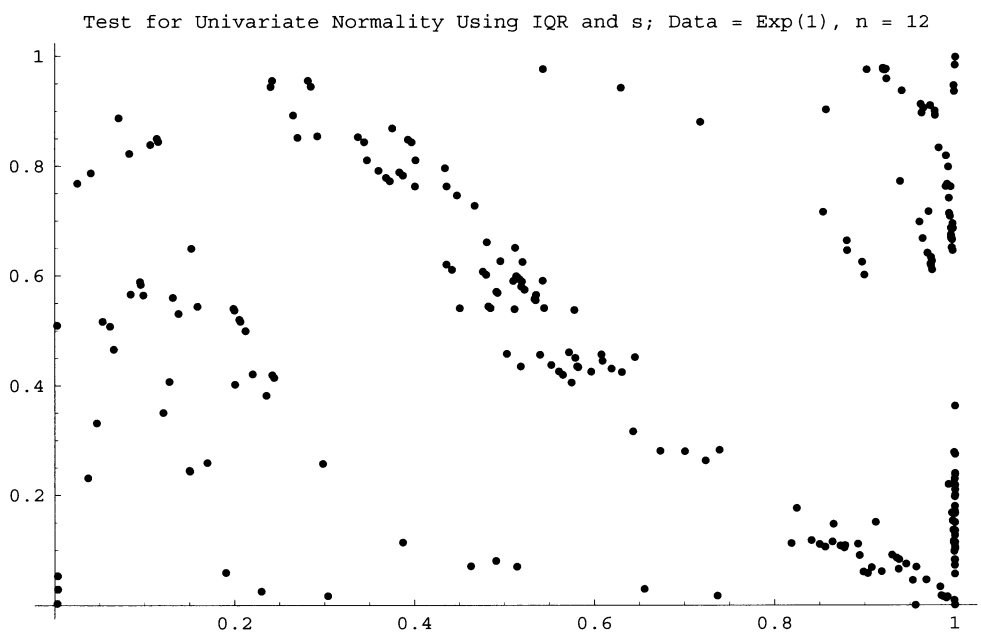
4. Resampling based tests for multivariate normality

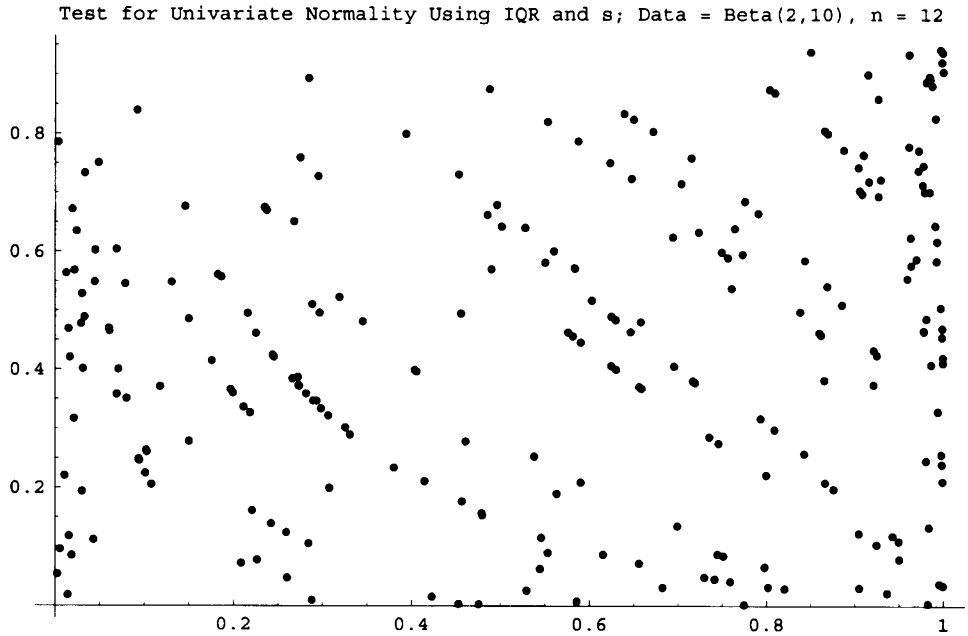
As in the univariate case, our proposed test uses the independence of the sample mean vector and the sample variance-covariance matrix. A difficult issue is the selection of two statistics, one a function of the mean vector and the other a function of the covariance matrix, that are to be used, as in the univariate case, for obtaining the w_i^* via use of the quantile transformation. We use the statistics $c'\overline{X}$, and either $\mathrm{tr}(\Sigma^{-1}S)$, or $\frac{|S|}{|\Sigma|}$. Our choice is exclusively guided by the fact that for these cases, the distributions of the statistics in finite samples are known. Other choices can (and should) be explored, but the technicalities would be substantially more complex.

Test 1. Suppose $X_1, X_2, \ldots, X_n$ are iid p-variate multivariate normal observations, distributed as $N_p(\mu, \Sigma)$. Then, for a given vector c, $c'\overline{X} \sim N_p(c'\mu\ ,\ \frac{1}{n}c'\Sigma c)$, and $\mathrm{tr}(\Sigma^{-1}S) \sim$ *chisquare*$(p(n-1))$. Thus, using the same notation as in Section 3.1,

$$U_n = \Phi\left(\frac{\sqrt{n}(c'\overline{X} - c'\mu)}{\sqrt{c'\Sigma c}}\right) \qquad \text{and} \qquad V_n = G_{p(n-1)}\big(\mathrm{tr}(\Sigma^{-1}S)\big)$$


Test for Univariate Normality Using IQR and s; Data = N(0,1), n = 12
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


Test for Univariate Normality Using IQR and s; Data = Exp(1), n = 12
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1



are independently $U[0,1]$ distributed. For $i = 1, 2, \ldots, B$, define

$$u_i^* = \Phi\left(\frac{\sqrt{n}(c'\overline{X_i}^* - c'\overline{X})}{\sqrt{c'Sc}}\right) \qquad \text{and} \qquad v_i^* = G_{p(n-1)}\left(\operatorname{tr}\left(S^{-1}S_i^*\right)\right),$$

where $\overline{X_i}^*, S_i^*$ are the mean vector and the covariance matrix of the ith bootstrap sample, and $\overline{X}, S$ are the mean vector and the covariance matrix for the original data. As before, we plot the pairs $w_i^* = (u_i^*, v_i^*), i = 1, 2, \ldots, B$ and check for an approximately uniform scattering.

Test 2. Instead of $\operatorname{tr}(\Sigma^{-1}S)$, consider the statistic $\frac{|S|}{|\Sigma|} \sim \prod_{i=1}^{p} \chi^2(n-i)$, where the *chisquare* variables are independently distributed.

For the special case $p = 2$, the distribution can be reduced to that of $\frac{\chi^4(2n-4)}{4}$ (see Anderson (1984)). Hence, U_n (as defined in Test 1 above), and

$$V_n = G_{2n-4}\left(2\frac{|S|^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}}\right)$$

are independently $U[0,1]$ distributed. Define now u_i^* as in Test 1 above, but

$$v_i^* = G_{2n-4}\left(2\frac{|S_i^*|^{\frac{1}{2}}}{|S|^{\frac{1}{2}}}\right),$$

and plot the pairs $w_i^* = (u_i^*, v_i^*)$ to check for an approximately uniform scattering.

The CDF of $\frac{|S|}{|\Sigma|}$ can be written in a reasonably amenable form also for the case $p = 3$ by using the Hypergeometric functions, but we will not describe the three dimensional case here.

As in the univariate case, we will see that Tests 1 and 2 can be quite effective and especially for small samples they are relatively more useful than alternative

tests used in the literature. For example, the common graphical test for bivariate normality that plots the *Mahalanobis* D^2 values against *chisquare* percentiles (see Johnson and Wichern (1992))would not have very much credibility at sample sizes such as $n = 10$ (a sample size we will try).

Corresponding to Theorem 3 , we have a similar consistency theorem.

Theorem 7. $\sup_{0\le u\le 1, 0\le v\le 1} |P_*(U^* \le u,\ V^* \le v) - P_F(U_n \le u, V_n \le v)| \to 0$ *in probability, provided the true CDF F has four moments (in the usual sense for a multivariate CDF).*

The nonull asymptotics (i.e., the analog of Theorem 4) are *much* harder to write down analytically. We have a notationally messy version for the bivariate case. However, we will not present it due to the notational complexity.

The plots of the pairs w_i^* corresponding to both Test 1 and Test 2 are important to examine from the point of view of applications. The plots corresponding to the first test are presented next. The plots corresponding to the second test look very similar and are omitted here.

The plots again show the impressive power of the tests to detect skewness, as is clear from the Bivariate Gamma plot (we adopt the definition of Bivariate Gamma as $(X, Y) = (U + W, V + W)$, where U, V, W are independent Gammas with the same scale parameter; see Li (2003) for certain recent applications of such representations.) The normal plot looks reasonably devoid of any structure or drastic nonuniformity. Considering that testing for bivariate normality continues to remain a very hard problem for such small sample sizes, our proposals appear to show good potential for being useful and definitely competitive. The ideas we present need to be examined in more detail, however.

5. Subsampling based tests

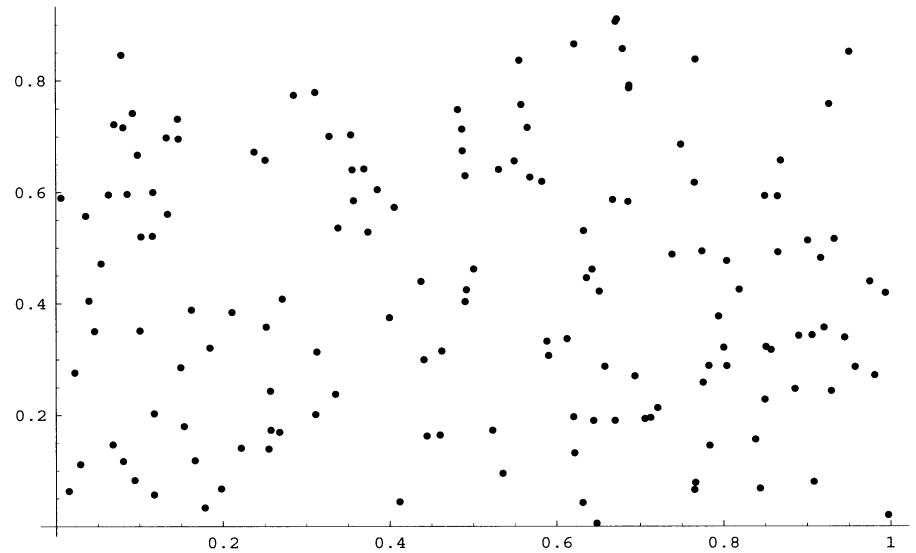
An alternative to the resampling based tests of the preceding sections is to use subsampling. From a purely theoretical point of view, there is no reason to prefer subsampling in this problem. Resampling and subsampling will both produce uniformly consistent distribution estimators, but neither will produce a test that is consistent against all alternatives. However, as a matter of practicality, it might be useful to use each method as a complement to the other. In fact, our subsampling based plots below show that there is probably some truth in that. In this section we will present a brief description of subsampling based tests. A more complete presentation of the ideas in this section will be presented elsewhere.

5.1. Consistency

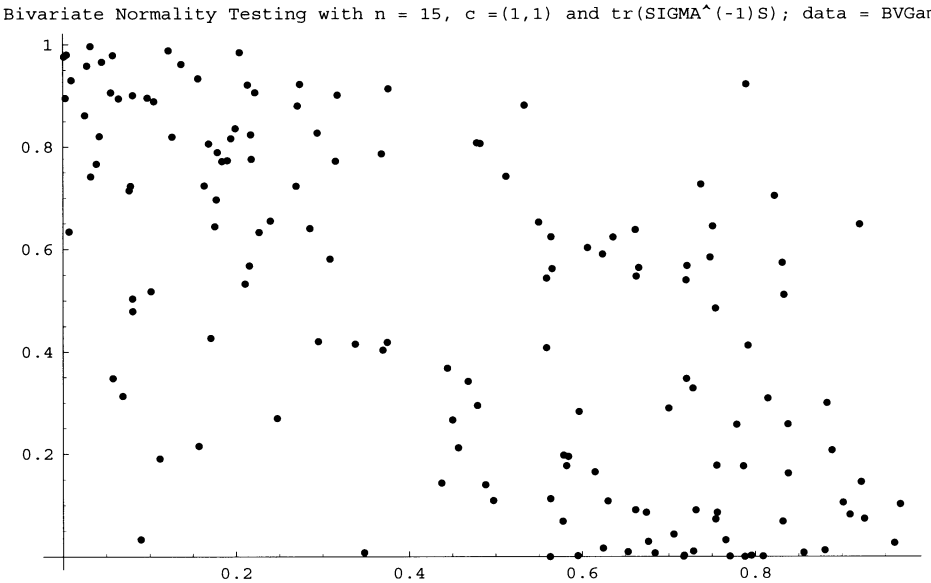
We return to the univariate case and again focus on the independence of the sample mean and sample variance; however, in this section, we will consider the subsampling methodology—see e.g., Politis, Romano and Wolf (1999). Denote by $\mathcal{B}_{b,1}, \ldots, \mathcal{B}_{b,Q}$ the $Q = \binom{n}{b}$ subsamples of size b that can be extracted from the sample $X_1, \ldots, X_n$. The subsamples are ordered in an arbitrary fashion except that, for convenience, the first $q = [n/b]$ subsamples will be taken to be the non-overlapping stretches, i.e., $\mathcal{B}_{b,1} = (X_1, \ldots, X_b)$, $\mathcal{B}_{b,2} = (X_{b+1}, \ldots, X_{2b}), \ldots$, $\mathcal{B}_{b,q} = (X_{(q-1)b+1}, \ldots, X_{qb})$. In the above, b is an integer in $(1, n)$ and $[\cdot]$ denotes integer part.

Let $\bar{X}_{b,i}$ and $s^2_{b,i}$ denote the sample mean and sample variance as calculated from subsample $\mathcal{B}_{b,i}$ alone. Similarly, let $U_{b,i} = \Phi(\frac{\sqrt{b}\ (\bar{X}_{b,i}-\mu)}{\sigma})$, and $V_{b,i} =$

Bivariate Normality Test using n = 10, c = (1,1), and tr(SIGMA^(-1)S); data = BVN(0,I)



Bivariate Normality Testing with n = 15, c =(1,1) and tr(SIGMA^(-1)S); data = BVGamma



$G_{b-1}(\frac{\sqrt{b-1}\ s^2_{b,i}}{\sigma^2})$. Thus, if b were n, these would just be U_n and V_n as defined in subsection 3.1. Note that $U_{b,i}$ and $V_{b,i}$ are not proper statistics since μ and σ are unknown; our proxies for $U_{b,i}$ and $V_{b,i}$ will be $\hat{U}_{b,i} = \Phi(\frac{\sqrt{b}\ (\bar{X}_{b,i}-\bar{X})}{s})$ and $\hat{V}_{b,i} = G_{b-1}(\frac{\sqrt{b-1}\ s^2_{b,i}}{s^2})$ respectively.

Let $H_b(x,y) = P(U_{b,1} \le x, V_{b,1} \le y)$. Recall that, under normality, H_b is uniform on the unit square. However, using subsampling we can consistently estimate H_b (or its limit H given in Theorem 4) whether normality holds or not. As in Politis et al. (1999), we define the subsampling distribution estimator by

$$\hat{L}_b(x,y) = \frac{1}{Q}\sum_{i=1}^{Q} \mathbf{1}\{\hat{U}_{b,i} \le x, \hat{V}_{b,i} \le y\}. \tag{2}$$

Then the following consistency result ensues.

Theorem 8. *Assume the conditions of Theorem 4. Then*

(i) *For any fixed integer $b > 1$, we have $\hat{L}_b(x,y) \xrightarrow{P} H_b(x,y)$ as $n \to \infty$ for all points (x,y) of continuity of H_b.*

(ii) *If $\min(b, n/b) \to \infty$, then $\sup_{x,y} |\hat{L}_b(x,y) - H(x,y)| \xrightarrow{P} 0$.*

Proof. (i) Let (x,y) be a point of continuity of H_b, and define

$$L_b(x,y) = \frac{1}{Q}\sum_{i=1}^{Q} \mathbf{1}\{U_{b,i} \le x, V_{b,i} \le y\}. \tag{3}$$

Note that by an argument similar to that in the proof of Theorem 2.2.1 in Politis, Romano and Wolf (1999), we have that

$$\hat{L}_b(x,y) - L_b(x,y) \to 0$$

on a set whose probability tends to one. Thus it suffices to show that $L_b(x,y) \xrightarrow{P} H_b(x,y)$. But note that $EL_b(x,y) = H_b(x,y)$; hence, it suffices to show that $Var(L_b(x,y)) = o(1)$.

Let

$$\tilde{L}_b(x,y) = \frac{1}{q}\sum_{i=1}^{q} \mathbf{1}\{U_{b,i} \le x, V_{b,i} \le y\}.$$

By a Cauchy–Schwartz argument, it can be shown that $Var(L_b(x,y)) \le Var(\tilde{L}_b(x,y))$; in other words, extra averaging will not increase the variance.

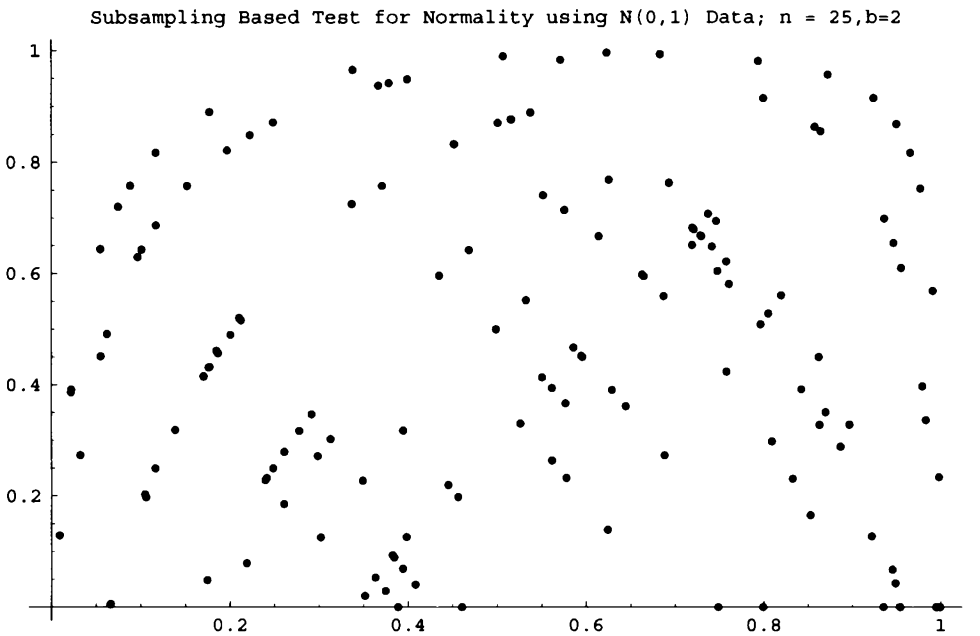
But $Var(\tilde{L}_b(x,y)) = O(1/q) = O(b/n)$ since $\tilde{L}_b(x,y)$ is an average of q i.i.d. random variables. Hence $Var(L_b(x,y)) = O(b/n) = o(1)$ and part (i) is proven. Part (ii) follows by a similar argument; the uniform convergence follows from the continuity of H given in Theorem 4 and a version of Polya's theorem for random cdfs. □

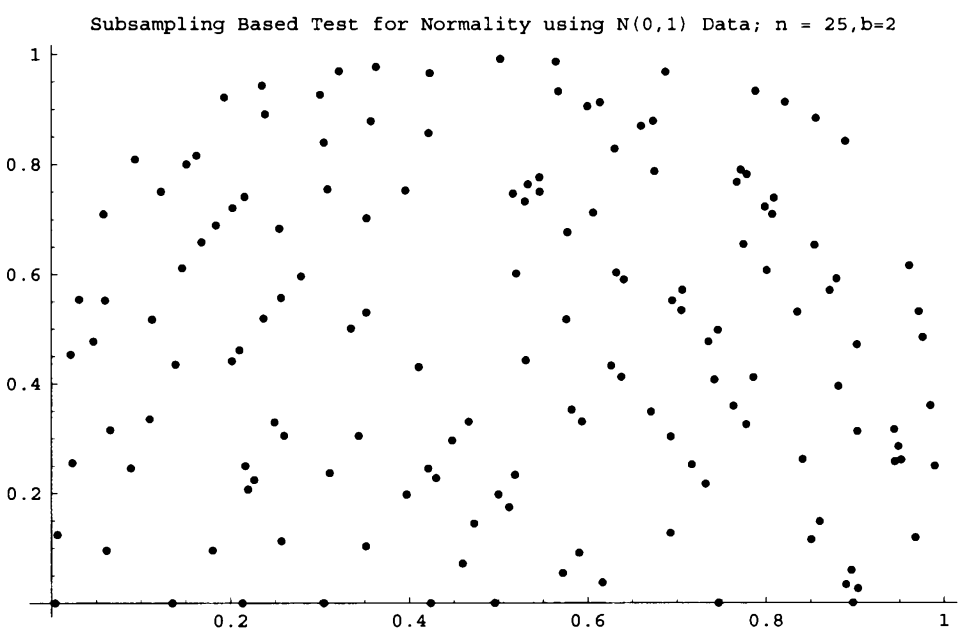
5.2. Subsampling based scatterplots

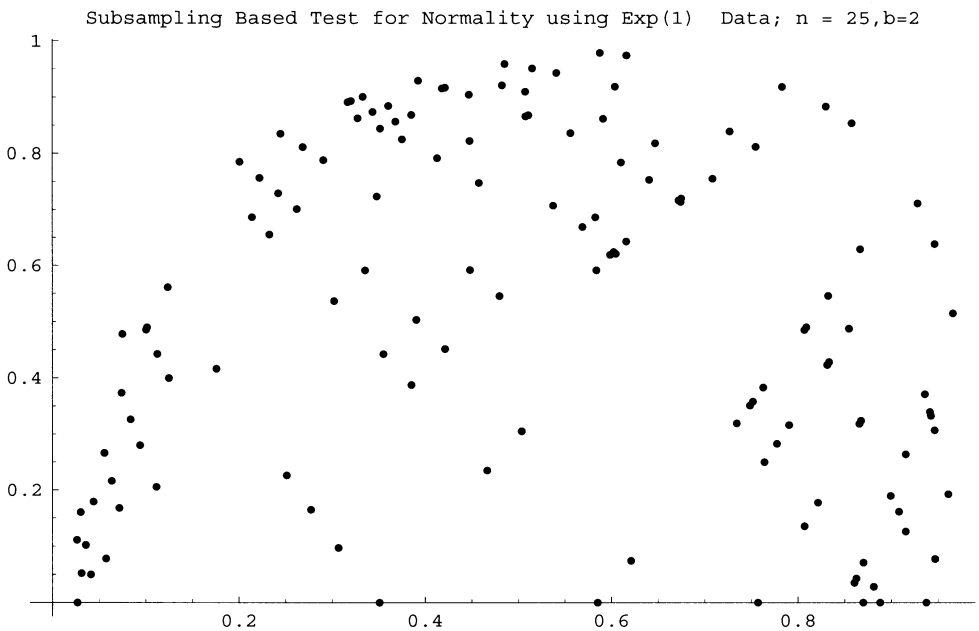
Theorem 8 suggests looking at a scatterplot of the pairs $\hat{w}_{b,i} = (\hat{U}_{b,i}, \hat{V}_{b,i})$ to detect non-normality since (under normality) the points should look uniformly scattered over the unit square, in a fashion analogous to the pairs $\hat{w}^*_i$ in Sections 3 and 4.

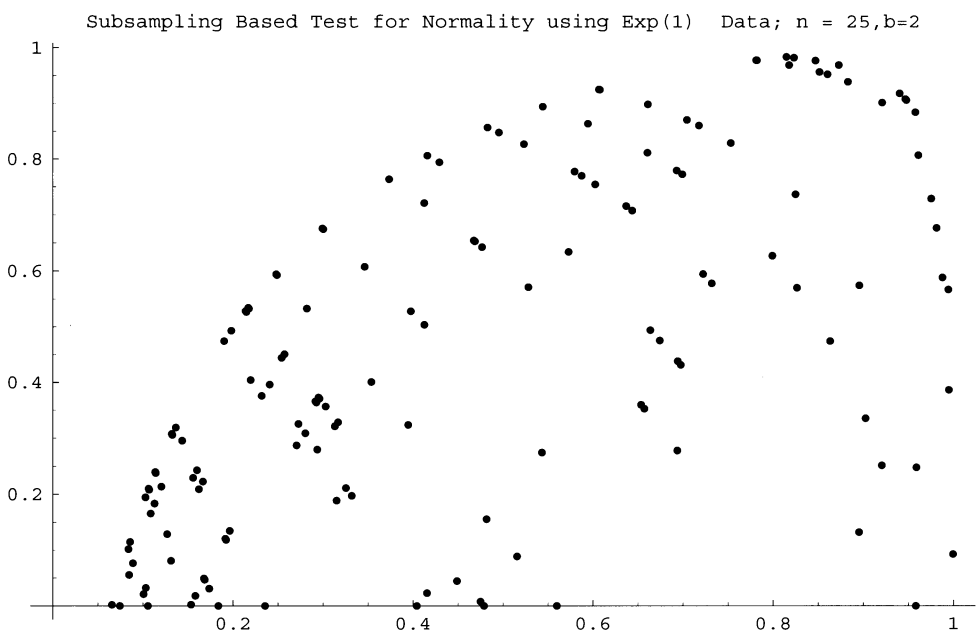
Below, we present a few of these scatterplots and then discuss the plots. The subsample size b in the plots is taken to be 2.

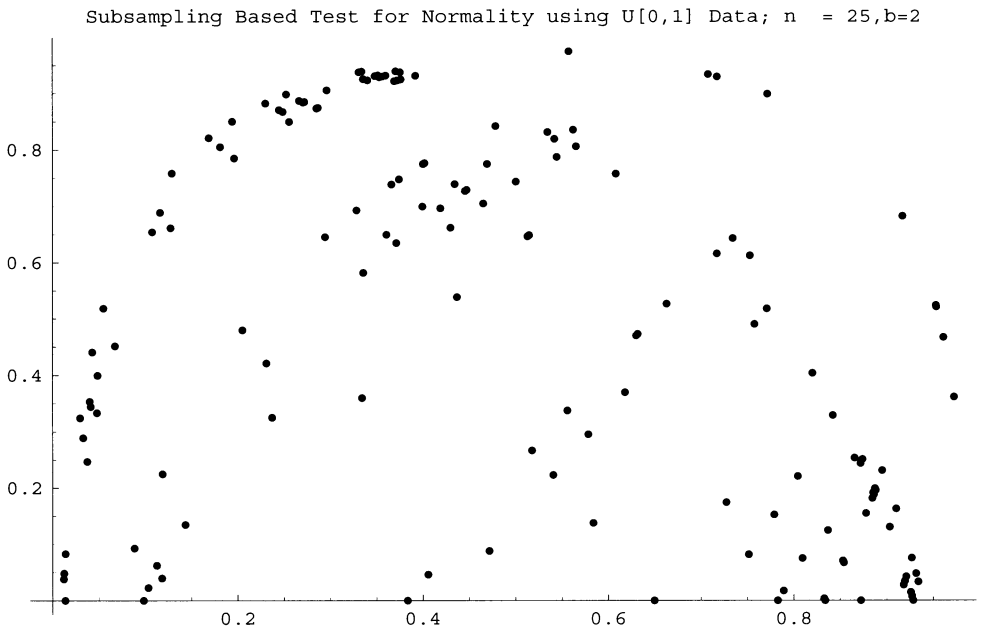
For each distribution, two separate plots are presented to illustrate the quite dramatic nonuniform structure for the nonnormal cases.

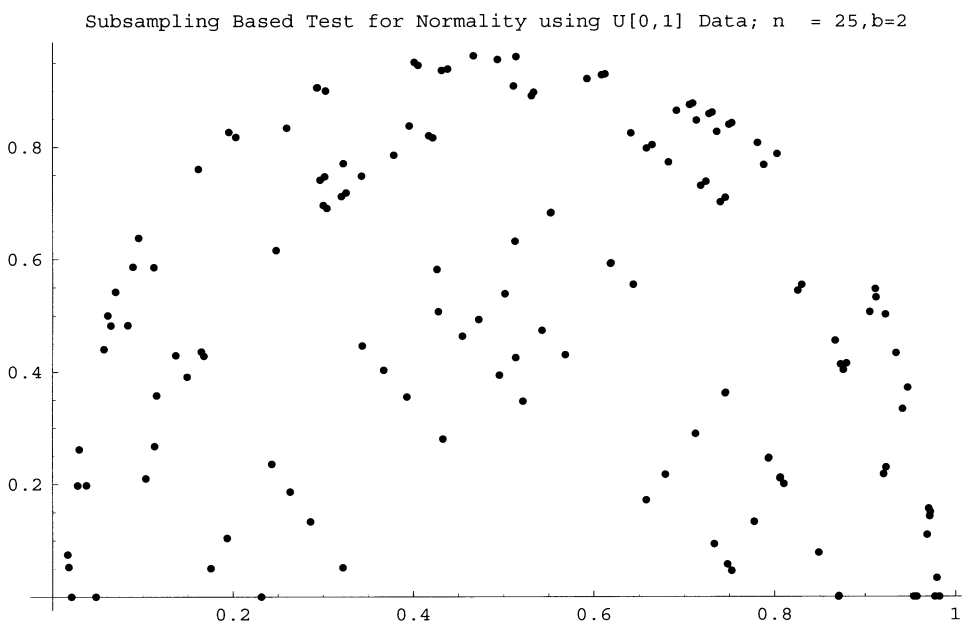

Subsampling Based Test for Normality using N(0,1) Data; n = 25,b=2
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


Subsampling Based Test for Normality using N(0,1) Data; n = 25,b=2
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


Subsampling Based Test for Normality using Exp(1) Data; n = 25,b=2
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8


Subsampling Based Test for Normality using Exp(1) Data; n = 25,b=2
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1


Subsampling Based Test for Normality using U[0,1] Data; n = 25,b=2
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8


Subsampling Based Test for Normality using U[0,1] Data; n = 25,b=2
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1

5.3. Discussion of the plots

Again, we are forced to present a limited number of plots due to space considerations. The plots corresponding to the Exponential and the uniform case show obvious *nonuniform* structure; they also show significant amounts of empty space. In fact, compared to the corresponding scatterplots for uniform data for the bootstrap based test in Section 3.3, the structured deviation from a uniform scattering is more evident in these plots. Subsampling seems to be working rather well in detecting nonnormality in the way we propose here! But there is also a problem. The problem seems to be that even for normal data, the scatterplots exhibit structured patterns, much in the same way for uniform data, but to *a lesser extent.* Additional theoretical justification for these very special patterns in the plots is needed.

We do not address other issues such as choice of the subsample size due to space considerations and for our focus in this article on just the resampling part.

6. Scope of other applications

The main merits of our proposal in this article are that they give a user something of credibility to use in small samples, and that the proposal has scope for broad applications. To apply our proposal in a given problem, one only has to look for an effective characterization result for the null hypothesis. If there are many characterizations available, presumably one can choose which one to use. We give a very brief discussion of potential other problems where our proposal may be useful. We plan to present these ideas in the problems stated below in detail in a future article.

1. Testing for sphericity

Suppose $X_1, X_2, \ldots, X_n$ are iid p-vectors and we want to test the hypothesis H_0: the common distribution of the X_i is spherically symmetric. For simplicity of explanation here, consider only the case $p = 2$. Literature on this problem includes Baringhaus (1991), Koltchinskii and Li (1998) and Beran (1979).

Transforming each X to its polar coordinates r, θ, under H_0, r and θ are independent. Thus, we can test H_0 by testing for independence of r and θ. The data we will use is a sample of n pairs of values $(r_i, \theta_i), i = 1, 2, \ldots, n$. Although the testing can be done directly from these pairs without recourse to resampling or subsampling, for *small* n, re or subsampling tests may be useful, as we witnessed in the preceding sections in this article.

There are several choices on how we can proceed. A simple correlation based test can be used. Specifically, denoting D_i as the difference of the ranks of the r_i and θ_i (respectively among all the r_i and all the θ_i), we can use the well known Spearman coefficient:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}.$$

For small n, we may instead bootstrap the (r_i, θ_i) pairs and form a scatterplot of the bootstrapped pairs *for each bootstrap replication.* The availability of replicated scatterplots gives one an advantage in assessing if any noticeable correlation between r and θ seems to be present. This would be an easy, although simple, visual method. At a slightly more sophisticated level, we can bootstrap the r_S statistic and compare percentiles of the bootstrap distribution to the theoretical percentiles under H_0 of the r_S statistic. We are suggesting that we break ties just by halving the ranks. For small n, the theoretical percentiles are available exactly; otherwise, we can use

the percentiles from the central limit theorem for r_S as (hopefully not too bad) approximations.

We should mention that other choices exist. An obvious one is Hoeffding's D-statistic for independence. Under H_0, $nD_n + \frac{1}{36}$ has a known (nonnormal) limit distribution. Although an exact formula for its CDF appears to be unknown, from the known formula for its characteristic function (see Hoeffding (1948)), we can pin down any specified percentile of the limit distribution. In addition, for small n, the exact distribution of D_n under H_0 is available too. We can thus find either the exact or approximate percentiles of the sampling distribution of $nD_n + \frac{1}{36}$, and compare percentiles of the bootstrap distribution to them. If we prefer a plot based test, we can construct a Q-Q plot of bootstrap percentiles against the theoretical percentiles under H_0 and interpret the plot in the standard manner a Q-Q plot is used.

2. Testing for Poissonity

This is an important problem for practitioners and has quite a bit of literature, e.g., Brown and Zhao (2002), and Gurtler and Henze (2000). Both articles give references to classic literature. If $X_1, X_2, \ldots, X_n$ are iid from a $Poisson(\lambda)$ distribution, then obviously $\sum_{i=1}^n X_i$ is also Poisson-distributed, and therefore every cumulant of the sampling distribution of $\sum_{i=1}^n X_i$ is $n\lambda$. We can consider testing that a set of specified cumulants are equal by using re or subsampling methods. Or, we can consider a fixed cumulant, say the third for example, and inspect if the cumulant estimated from a bootstrap distribution behaves like a linear function of n passing through the origin. For example, if the original sample size is $n = 15$, we can estimate a given order cumulant of $\sum_{i=1}^m X_i$ for each $m = 1, 2, \ldots, 15$, and visually assess if the estimated values fall roughly on a straight line passing through the origin as m runs through 1 to 15. The graphical test can then be repeated for a cumulant of another order and the slopes of the lines compared for approximate equality too. Using cumulants of different orders would make the test more powerful, and we recommend it.

The cumulants can be estimated from the bootstrap distribution either by differentiating the empirical cumulant generating function $\log(\sum_s e^{ts} P_*(S_n^* = s))$ or by estimating instead the moments and then using the known relations between cumulants and moments (see, e.g., Shiryaev (1980)).

3. Testing for exponentiality

Testing for exponentiality has a huge literature and is of great interest in many areas of application. We simply recommend Doksum and Yandell (1984) as a review of the classic literature on the problem. A large number of characterization results for the family of Exponential distributions are known in the literature. Essentially any of them, or a combination, can be used to test for exponentiality. We do not have reliable information at this time on which characterizations translate into better tests. We mention here only one as illustration of how this can be done.

One possibility is to use the spacings based characterization that $(n - i + 1)R_i$ are iid $Exponential(\lambda)$ where λ is the mean of the population under H_0, and R_i are the successive spacings. There are a number of ways that our general method can be used. Here are a few. A simple plot based test can select two values of i, for example $i = [n/2]$, and $[n/2] + 1$, so that the ordinary bootstrap instead of a m-out of-n bootstrap can be used, and check the pairs for independence. For example, a scatterplot of the bootstrapped pairs can be constructed. Or, one can standardize

the bootstrapped values by $\overline{X}$, so that we will then have pairs of approximately iid *Exponential*(1) values. Then we can use the quantile transformation on them and check these for uniformity in the unit square as in Section 3. Or, just as we described in the section on testing for sphericity, we can use the Hoeffding D-statistic in conjunction with the bootstrap with the selected pairs of $(n-i+1)R_i$. One can then use two other values of i to increase the diagnostic power of the test. There are ways to use all of the $(n-i+1)R_i$ simultaneously as well, but we do not give the details here.

Acknowledgement

Peter Bickel mentioned to one of the authors that uniform data look like normal on a Q-Q plot and suggested a study. Len Haff and David Moore made helpful comments. J. K. Ghosh, Bimal Sinha and Malay Ghosh made comments on the results in Section 2. The work was partially funded by NSF grant DMS 00-71757.

References

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes, Ann. Math. Stat. 23 193–212. MR50238

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit, JASA 49 765–769. MR69459

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York. MR771294

Babu, G. J. and Rao, C. R. (2004). Goodness of fit tests when parameters are estimated, Sankhyá, 66, To appear. MR2015221

Basu, D. (1955). On statistics independent of a complete and sufficient statistic, Sankhyá, 377–380. MR74745

Baringhaus, L. (1991). Testing for spherical symmetry of a multivariate distribution, Ann. Stat. 19(2) 899–917. MR1105851

Beran, R. (1979). Testing for ellipsoidal symmetry of a multivariate density, Ann. Stat. 7(1) 150–162. MR515690

Brown, L. and Zhao, L. (2002). A test for the Poisson distribution, Sankhyá, A Special Issue in Memory of D. Basu, A. DasGupta Eds., 64, 3, Part I, 611–625. MR1985402

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness of Fit Techniques*, Marcel Dekker Inc., New York. MR874534

DasGupta, A. and Haff, L. R. (2003). Asymptotic values and expansions for correlations between different measures of spread, Invited article for Special issue in Memory of S.S. Gupta, Jour. Stat. Planning and Inf.

Doksum, K. and Yandell, B. (1984). Tests for Exponentiality, In Handbook of Statistics, 4, P.R. Krishnaiah and P.K. Sen Eds., North-Holland, Amsterdam, 579–612. MR831730

Gurtler, N. and Henze, N. (2000). Recent and classical goodness of fit tests for the Poisson distribution, Jour. Stat. Planning and Inf., 90(2) 207–225. MR1795597

Hoeffding, W. (1948). A nonparametric test of independence, Ann. Math. Stat., 19, 546–557. MR29139

Johnson, R. and Wichern, D. (1992). *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, New Jersey. MR1168210

Kagan, A. M., Linnik, Yu. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, John Wiley, New York. MR346969

Koltchinskii, V. and Li, L. (1998). Testing for spherical symmetry of a multivariate distribution, Jour. Mult. Analysis, 65(2) 228–244. MR1625889

Li, Xuefeng (2003). *Infinitely Divisible Time Series Models*, Ph.D. Thesis, University of Pennsylvania, Department of Statistics.

McDonald, K. L. and Katti, S. K. (1974). Test for normality using a characterization, in: *Proc. of Internat. Conf. on Characterizations of Stat. Distributions with Applications*, pp. 91–104.

Mudholkar, G. S., McDermott, M. and Srivastava, D. (1992). A test of p-variate normality, Biometrika **79**(4) 850–854. MR1209484

Mudholkar, G. S., Marchetti, C. E. and Lin, C. T. (2002). Independence characterizations and testing normality against restricted skewness—kurtosis alternatives, Jour. Stat. Planning and Inf., **104**(2) 485–501. MR1906268

Parthasarathy, K. R. (1976). Characterisation of the normal law through the local independence of certain statistics, Sankhyá, Ser. A, **38**(2) 174–178. MR461747

Politis, D., Romano, J. and Wolf, M. (1999). *Subsampling*, Springer-Verlag, New York. MR1707286

Serfling, R. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York. MR595165

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality: Complete samples, Biometrika **52** 591–611. MR205384

Shiryaev, A. (1980). Probability, Springer, New York. MR1368405

Stephens, M. A. (1976). Asymptotic results for goodness of fit statistics with unknown parameters, Ann. Stat. **4** 357–369. MR397984

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, New York. MR1652247