Les Cahiers du GERAD ISSN: 0711–2440

A One-Sided Conditionally Distribution-Free Sign Test for Bivariate Data

Denis Larocque
Mélanie Labarre

G-2002-55

October 2002

A One-Sided Conditionally Distribution-Free Sign Test for Bivariate Data

Denis Larocque*

Mélanie Labarre

Department of Management Science, HEC Montréal
3000 chemin de la Côte-Sainte-Catherine
Montréal (Québec), Canada, H3T 2A7
**and GERAD*

October, 2002

Les Cahiers du GERAD

G–2002–55

Abstract

The one-sided one sample problem with bivariate data is considered. A conditionally distribution-free sign test is proposed for that problem. This test is related to Hodges test and can be seen as a union-intersection test. Moreover, it is valid under very mild assumptions and it can be easily implemented. An explicit formula for the exact null conditional distribution of the test statistic is derived. This conditional distribution can be used to compute exact conditional p-values. A simulation study compares the new test to some competitors including the likelihood ratio test. The results show that the new test is very competitive for a wide variety of distributional models.

Key Words: Positive orthant alternative, One-sided alternative, Bivariate data, Conditionally distribution-free, Sign test, Hodges test, Union-intersection, Random walk.

Résumé

Nous considérons le problème de position unilatéral bidimensionnel. Nous proposons un test du signe conditionnellement “distribution-free”. Ce test, basé sur le principe de l’union-intersection, est apparenté au test de Hodges. De plus, sa validité ne requiert que des conditions très peu restrictives. Le test peut également être implémenté facilement. Nous donnons une formule explicite donnant sa loi conditionnelle exacte sous l’hypothèse nulle. Cette loi conditionnelle peut être utilisée pour calculer le seuil descriptif . Le nouveau test est comparé à certains compétiteurs, dont le test du rapport de vraisemblance, à l’aide d’une étude par simulation. Les résultats de cette étude démontrent que le nouveau test est très compétitif, et ce, pour une grande variété de modèles.

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent bivariate random vectors. For each $1 \leq i \leq n$, it is assumed that, for any line passing through the origin, the conditional univariate distribution of $(X_i - \mu_1, Y_i - \mu_2)$ given that it is on that line has median 0. However, we do not assume that the observations have the same distribution. These assumptions hold for a random sample from an elliptical distribution, or for any symmetric distribution, centered at (μ_1, μ_2) but they are a lot more general allowing for different kinds of skewed models. We wish to test the hypotheses

$$H_0 : (\mu_1, \mu_2) = (0, 0) \quad \text{vs} \quad H_1 : (\mu_1, \mu_2) \geq (0, 0) \quad \text{with at least one strict inequality.} \tag{1}$$

For the more general case of p dimensional vectors and assuming that the observations come from a multivariate normal population $N_p(\mu, \mathbf{\Sigma})$, different testing procedures have been proposed. Perlman (1969) derived the likelihood ratio test (LRT) when the covariance matrix is unknown. The null distribution of the LRT depends on the unknown covariance matrix but Perlman gave a bound for the critical point that can be used to perform a conservative test at a given level. However, this results in a loss of power and the LRT have been criticized over the years. In an interesting paper, Perlman and Wu (1999) came to its defense but judging from the discussions that follow their article, it seems that the matter is not completely resolved right now. Silvapulle (1995) proposed a Hotelling's type test that is asymptotically equivalent to the LRT. Wang and McDermott (1998a) proposed a conditional version of the LRT test. This test can be difficult to use in practice because it requires the numerical evaluation of integrals. Wang and McDermott (1998b) used a similar approach to derive a conditional Hotelling's test. Perlman and Wu (2002) proposed another conditional version of the LRT that is simpler to implement. Schucany, Frawley, Gray and Wang (1999) proposed to use the bootstrap to estimate the p-value of the LRT. As long as the computation of the LRT statistic is practical, this is a good way of avoiding the use of the conservative critical point and it also relaxes the normality assumption by letting the test adapt itself to the data. Sen and Tsai (1999) proposed a union-intersection test related to the LRT. As for the LRT, the direct application of this method results in a conservative test. However, Sen and Tsai proposed a two-stage version of their test and of the LRT which are unbiased. A possible disadvantage of this approach is that it requires an arbitrary choice of two quantities, a first stage sample size and a $p \times p$ matrix. With the exception of the bootstrap LRT test that can adapt itself to other distributions, all the tests above rely on the normality assumption. In order to make the procedure less dependent on the normality assumption, Mudholkar, Kost and Subbaiah (2001) proposed stepwise tests and robust stepwise tests based on trimmed means.

Another approach consists in utilizing a procedure to test H_0 against an alternative different but related to H_1 and then see how it performs for the one-sided problem. In that vein, Tang (1994) investigated tests for half space alternatives and Follmann (1996) proposed a simple an intuitive procedure to test H_0 against another related alternative specifying that the sum of the means is greater than 0.

Sen and Silvapulle (2002) give an interesting an up to date review of some aspects of inference under inequality constraints which includes the one-sided problem.

It seems that no tests exhibiting a finite sample distribution-free property, under H_0, over a large class of models have been proposed in the literature for the one-sided problem. This paper aims to fill that gap. We propose a bivariate sign test for the one-sided alternative. Under the null hypothesis, the test statistic is conditionally distribution-free under very mild assumptions. The conditional distribution can be used to compute exact conditional p-values. We provide a simple way to compute the test statistic and give an explicit formula for its conditional null distribution. The test procedure can then be implemented easily. The performance of this test against some competitors is investigated with an extensive simulation study.

The test statistic is presented in Section 2 where its null conditional distribution is derived. The results from a simulation study are presented in Section 3 followed by concluding remarks.

2 Test statistic

Let θ be in the interval $[0, 2\pi]$. For $i = 1, \ldots, n$, define

$$P_i(\theta) = X_i \cos(\theta) + Y_i \sin(\theta) \tag{2}$$

to be the projection of (X_i, Y_i) on the directed line passing through the origin with angle θ. Let ψ be the function defined by $\psi(u) = 1$ or 0 as $u > 0$ or $u \leq 0$. Define

$$S(\theta) = \sum_{i=1}^{n} \psi(P_i(\theta)) \tag{3}$$

as the number of positive observations among the projected points on the axis with angle θ. $S(\theta)$ is thus equivalent to the sign statistic computed on the projected points. Let's assume throughout that, for any $\theta \in [0, \pi/2]$, $P(P_i(\theta) = 0) = 0$ for any i. Then, under the hypothesis H_0, $\psi(P_1(\theta)), \ldots, \psi(P_n(\theta))$ are independent random variables taking the values 0 or 1, each with probability 1/2.

Let's consider first the problem of testing H_0 against the unrestricted alternative

$$H_1^* : (\mu_1, \mu_2) \neq (0, 0). \tag{4}$$

Hodges (1955) was the first to proposed a sign test for that problem. His test is based on

$$\sup_{0 \leq \theta \leq 2\pi} S(\theta). \tag{5}$$

This statistic is strictly distribution-free under H_0. The basic idea behind this approach is that, for a given angle θ a high (or low) value of $S(\theta)$ provides evidence against the null hypothesis. The test statistic simply takes the maximum of those values over all angles.

Larocque, Tardif and van Eeden (2000) studied different ways of averaging the values of $S(\theta)$, as θ varies, to construct a test statistic.

Let's go back to the one-sided problem (1). Roy's (1953) union-intersection (UI) principle can be used to motivate our test statistic; see also Sen and Silvapulle (2002). Following Sen and Tsai (1999), let $(\delta_1, \delta_2) \in \Delta = \{(x, y) : x \geq 0,\ y \geq 0,\ \max(x, y) > 0\}$. If we define

$$H_{0,(\delta_1,\delta_2)} : \delta_1\mu_1 + \delta_2\mu_2 = 0 \quad \text{and} \quad H_{1,(\delta_1,\delta_2)} : \delta_1\mu_1 + \delta_2\mu_2 > 0,$$

then

$$H_0 = \bigcap_{(\delta_1,\delta_2)\in\Delta} H_{0,(\delta_1,\delta_2)} \quad \text{and} \quad H_1 = \bigcup_{(\delta_1,\delta_2)\in\Delta} H_{1,(\delta_1,\delta_2)}.$$

For each $(\delta_1, \delta_2) \in \Delta$, a one-sided univariate test statistic $T(\delta_1, \delta_2)$ can be used to test $H_{0,(\delta_1,\delta_2)}$ against $H_{1,(\delta_1,\delta_2)}$ and then, a UI test for testing H_0 against H_1 can be based on $\sup_{(\delta_1,\delta_2)\in\Delta} T(\delta_1, \delta_2)$. If the univariate statistic is scale-invariant, we can restrict the search of the supremum to the vectors satisfying $\delta_1^2 + \delta_2^2 = 1$ and then, the test statistic reduces to $\sup_{0\leq\theta\leq\pi/2} T(\theta)$ where $T(\theta)$ is the univariate test statistic computed with the projected observations $P_1(\theta), \ldots, P_n(\theta)$.

Sen and Tsai (1999) proposed such a UI test by using the univariate t-test. Here, we propose using the sign test instead. Our UI test can also be seen as an analogue of Hodges test for the one-sided problem and is based on

$$S = \sup_{0\leq\theta\leq\pi/2} S(\theta). \tag{6}$$

The hypothesis H_0 will be rejected for large values of S. The statistic S is scale and permutation invariant. This means that the value of S remains unchanged if we compute it using the sample $(Y_1, X_1), \ldots, (Y_n, X_n)$ or any sample of the type $(aX_1, bY_1), \ldots, (aX_n, bY_n)$ where $a > 0$ and $b > 0$. The statistics S can also be quickly and easily calculated. A straightforward way of doing so is given in the Appendix.

Let's partition the two-dimensional plane in four mutually exclusive regions in the usual way:

$$\begin{aligned} \text{Quadrant I} &= \{(x, y) : x \geq 0, y \geq 0\} \\ \text{Quadrant II} &= \{(x, y) : x < 0, y > 0\} \\ \text{Quadrant III} &= \{(x, y) : x \leq 0, y \leq 0\} \\ \text{Quadrant IV} &= \{(x, y) : x > 0, y < 0\}. \end{aligned} \tag{7}$$

For each $1 \leq i \leq n$, let

$$\phi_i = -\arctan(X_i/Y_i) \tag{8}$$

be the angle ($\in [-\pi/2, \pi/2]$) such that the projection of the point (X_i, Y_i) on the line with that angle is 0, i.e. $P_i(\phi_i) = 0$. Since $\phi_i \in (-\pi/2, 0)$ for any observations in quadrant I or III, we have $\psi(P_i(\theta)) = 1$ (0) for all $\theta \in [0, \pi/2]$ for any observations in quadrant I (III). For the observations in quadrant II or IV, $\phi_i \in [0, \pi/2]$. Thus, for an observation in quadrant IV, $\psi(P_i(\theta)) = 1$ or 0 according to $\theta \in [0, \phi_i)$ or $\theta \in [\phi_i, \pi/2)$. Similarly, for an observation in quadrant II, $\psi(P_i(\theta)) = 0$ or 1 according to $\theta \in [0, \phi_i]$ or $\theta \in (\phi_i, \pi/2)$.

On the contrary to Hodges statistic, S is not unconditionally distribution-free under H_0. This can be seen easily from the following example. Suppose that the observations arise from a degenerate distribution on the main diagonal, i.e. the line x=y. Then under H_0, S is distributed as a binomial random variable with parameters n and 1/2, $B(n, 1/2)$. This comes from the fact that each observation is either in quadrant I or III with probability 1/2. On the other hand, if the observations arise from a degenerate distribution on the line perpendicular to the main diagonal, then S has the same distribution as $\max(X, n - X)$ where X has the $B(n, 1/2)$ distribution. This comes from the fact that each observations is either in quadrant II or IV with probability 1/2 and that $\sup_{0 \leq \theta \leq \pi/2} S(\theta)$ takes its maximal value when either $\theta = 0$ or $\pi/2$. This also shows that we can not expect S to be unconditionally distribution-free even as $n \to \infty$.

In the rest of this section, we obtain a conditional distribution of S under H_0. Define M to be the number of observations in quadrants II and IV. We will show that, conditionally on M, S is distribution-free under H_0 and we will derive its exact null distribution. When applying the test to data, this conditional distribution will permit the computation of exact conditional p-values.

From now on, suppose that $M = m$ is observed and let $\phi_1^*, \ldots, \phi_m^*$ be the ϕ angles, as defined by (8), of the m observations in quadrant II and IV. For simplicity, we will assume for the rest of the paper that those m angles are distinct with probability one.

Let Y be the number of observations in quadrant I. Given $M = m$, the null distribution of Y is $B(n - m, 1/2)$. Define A and B to be the number of observations in quadrant IV and II respectively. Given $M = m$, the null distribution of A, and also of B, is $B(m, 1/2)$ and we have $A + B = m$. When θ goes from 0 to $\pi/2$, the projection of any point in quadrant I is always positive while the one of any point in quadrant III is always negative. The contribution of those points to the statistic S is thus simply Y. The contribution of any point in quadrant IV is 1 at first (when $\theta = 0$), it changes one time along the way (at the corresponding ϕ angle) and is 0 in the end (when $\theta = \pi/2$). The opposite is true for any point in quadrant II, i.e., its contribution is 0 at first and 1 in the end. Let

$$X = \sup_{0 \leq \theta \leq \pi/2} X(\theta)$$

where

$$X(\theta) = \text{card}\,\{(X_i, Y_i) \in \text{ quadrant II or IV } : P_i(\theta) > 0\}.$$

X is thus the contribution of the points in quadrant II and IV to the statistic S. Consequently, the distribution on S is the same as that of $X + Y$.

The process $\{X(\theta) : \theta \in [0, \pi/2]\}$ starts at A and finishes at B in m steps. The steps in the process $X(\theta)$ occur at $\phi_1^*, \ldots, \phi_m^*$. At each of these steps, the value of $X(\theta)$ increases or decreases by 1. Under the null hypothesis, each of those occurrences are equally likely.

Thus the probability that the process moves up by one is 1/2 and the probability that it goes down by one is also 1/2. Consequently, the null distribution of X is the same as that of the maximum of a random walk that starts at A and finishes at B in m steps. We can summarize all of the above in the next theorem.

Assumptions 2.1:

1. $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent bivariate random vectors not necessarily having the same distribution.
2. For each $1 \leq i \leq n$, it is assumed that, for any line passing through the origin, the conditional univariate distribution of $(X_i - \mu_1, Y_i - \mu_2)$ given that it is on that line has median 0.
3. The angles $\phi_1^*, \ldots, \phi_m^*$, as defined by (8), of the observations in quadrant II and IV are distinct with probability one.
4. For each $i = 1, \ldots, n$, $P(P_i(\theta) = 0) = 0$ for any $\theta \in [0, \pi/2]$ where $P_i(\theta)$ is the projection defined by (2).

These assumption are very mild and hold for any elliptical or symmetric distributions. But they also hold for certain types of skewed distributions and they allow non identically distributed observations.

Theorem 2.1 *Under the assumptions (2.1) and conditionally on $M = m$, the null distribution of S is the same as that of $X+Y$ where X and Y are independent random variables. Y has a $B(n-m, 1/2)$ distribution and X has the distribution of the maximum of a random walk that starts at A and finishes at B in m steps where $A + B = m$ and where A is distributed as a $B(m, 1/2)$ variate.*

Once assumptions (2.1) are met, the conditional distribution of S does not depend on the specific distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$. S is thus conditionally distribution-free under H_0. To use the test in practice, we need an expression for the exact conditional null distribution of S. The next theorem, proved in the Appendix, gives such an expression.

Theorem 2.2 *Under the assumptions (2.1) and conditionally on $M = m$, let $x \in \{m/2, \ldots, n\}$ if m is even and $x \in \{(m+1)/2, \ldots, n\}$ if m is odd, then*

$$P(S \geq x | M = m) = \frac{1}{2^n} \sum_{y=0}^{n-m} F(x-y) \binom{n-m}{y}$$

where

$$F(x) = \sum_{\substack{a=0 \\ \{a : x < \max(a, m-a)\}}}^{m} \binom{m}{a} + \sum_{\substack{a=0 \\ \{a : \max(a, m-a) \leq x \leq m\}}}^{m} \binom{m}{x}.$$

In practice, having observed $M = m$ and the value of S, the formula above can be used to compute the exact conditional p-value very easily . For example, suppose that the sample size is $n = 40$ and that we observe $M = m = 18$ and $S = 23$, then $P(S \geq 23 | M = 18) = 0.559$.

3 Simulation study and concluding remarks

In this section, we compare the new test to five competitors in a simulation study. The first two, that are valid against the unrestricted alternative (4), are Blumen's (1958) and Hotelling's tests. Blumen's test is an affine-invariant bivariate sign test. There are a lot of different bivariate sign tests for the unrestricted alternative that have been proposed in the literature but Oja and Nyblom (1989) showed that, for elliptical distributions, Blumen's test is a locally most powerful invariant sign test. This is why it was selected to represent the class of bivariate sign tests in this simulation study. Hotelling's test was selected to represent the class of tests for the unrestricted alternative under normality since it is uniformly most powerful among affine-invariant tests in that case; see Anderson (1984). The three other tests are tailored for the one-sided alternative. The first one is the likelihood ratio test (LRT) of Perlman (1969). The second one is the bootstrap version of the LRT test proposed by Schucany, Frawley, Gray and Wang (1999). Since this paper is not yet published, the bootstrap procedure is reproduced in the Appendix for completeness. The last one has not been proposed explicitly anywhere but it is simply the bootstrap version of the UI test of Sen and Tsai (1999), given by their equation (2.5), and discussed in the preceding section. In that case, the resampling is done exactly as for the bootstrap LRT and we call this test bootstrap UIT. We chose to do the bootstrap version of the Sen and Tsai test because the critical point of this test is difficult to obtain. This also gives an advantage to this approach because the use of the critical point produces a conservative test like it does for the LRT test.

All tests were performed at the 5% level and the sample size is $n = 40$. For Blumen's test, the asymptotic (χ^2_2) critical point is used. The exact (under normality) critical point from the F distribution is used for Hotelling's test. For the two bootstrap tests, the p-value is approximated using 499 bootstrap samples. For the LRT test the conservative critical point of Perlman (1969) is used. Since S has a discrete distribution, the critical points were chosen such that the probability of falsely rejecting H_0 is as close to but less than 0.05. For $n = 40$ this test can be written as, reject H_0 if:

$$S \geq \begin{cases} 26 & \text{if} \quad m = 0 \\ 27 & \text{if} \quad 1 \leq m \leq 6 \\ 28 & \text{if} \quad 7 \leq m \leq 18 \\ 29 & \text{if} \quad 19 \leq m \leq 40 \end{cases}$$

By using these critical points, the actual level of the test, as a function of m, varies between 0.0223 and 0.0494. Thus, this way of using S gives rise to a conservative test. It is important to note that the conservative nature of the test comes from the fact that the exact null conditional distribution is discrete and not because it depends on a nuisance parameter as it is the case for the LRT. Consequently, as n becomes larger, the test becomes less conservative.

The number of replications is 10000. The computations were performed using Ox version 3.20; Doornik (1999).

Six distributions were chosen. The first one is the standard bivariate normal distribution $N_2(0, I)$ where I is the identity matrix. The next two are contaminated normal distributions generated the following way: with probability p, a $N_2(0, I)$ random variable is generated and with probability $1 - p$, a $N_2(0, 16I)$ is generated. Two values of p, 0.1 and 0.2 are used. The third distribution is the standard (elliptic) Cauchy distribution also known as the t distribution with 1 df. Those first four distributions are elliptical. The fifth one is a symmetric but non-elliptical distribution generated the following way. First we generate an angle ($\in [0, 2\pi]$) by $\theta_1 = \pi(Bet + Ber)$ where Bet is distributed as a Beta random variable with parameters .2 and .2 and Ber is a Bernoulli random variable (independent of Bet) with probability of success 0.5. Second, a radius R_1 is generated as a uniform random variable on the interval [0,10]. The generated random variable is then $R_1(\cos(\theta_1), \sin(\theta_1))$. We denote this distribution by beta-angle for simplicity. The last one is a non-symmetric distribution, that satisfies assumptions (2.1), generated the following way: let θ_2 be an angle distributed uniformly on $[0, \pi]$ and let R_2 be a radius distributed as (0.6931–E) where E is an exponential random variable with parameter 1. The observation generated is $R_2(\cos(\theta_2), \sin(\theta_2))$. We denote this distribution by exponential radius. Note that the mean vector on this distribution is not (0,0) under H_0 so the normal-theory tests based on averages (LRT, bootstrap LRT and UIT, Hotelling) are not valid in that case but the two sign tests are valid.

Three different dependence structure are considered for each distribution. First, the original observations are used. This case is denoted by $\rho = 0$ in the tables that follows. Then each bivariate observation is multiplied by the matrix ((0.97891, 0.20431),(0.20431, 0.97891)). This produces a correlation of 0.4 for the normal distributions and is denoted by $\rho = 0.4$ in the tables. Then each observation is multiplied by the matrix ((–0.97891, 0.20431),(0.20431, –0.97891)). This produces a correlation of –0.4 for the normal distributions and is denoted by $\rho = -0.4$ in the tables. We thus have eighteen (6×3) different distributions in all. For each of the eighteen distributions, the observed probability of rejection was recorded for the null hypothesis and for nine different shift alternatives, three of them are on the X axis, three others are on the line with angle $\pi/8$ and the last three are on the main diagonal (the line with angle $\pi/4$). The amounts of shift were selected so that the power on the S test is about 0.2, 0.5 and 0.8 respectively for the three shifts in each of the directions.

The results are presented in tables 1 to 6. Before examining each distribution separately, here are some general remarks. The observed level of the test S is always below the nominal level 0.05 while the one of Blumen's test is always close to 0.05. In the situation where they are valid, the observed levels of the average based tests are close to the nominal value or below it for the conservative LRT test. For the Cauchy and exponential radius distributions, those tests are not valid and their observed levels are either too high or too low.

Let's examine the observed powers. For each alternative, an asterisk has been added to the test with the highest power and to any other that is not significantly different from it at the 5% level. Overall, the test S is among the most powerful tests for 99 out of the

162 different alternatives considered. The bootstrap tests are among the most powerful tests for 45 of those alternatives and Blumen's test is among the best for 36 of those alternatives. The S test is at its best for alternatives that are on the main diagonal or on the line with angle $\pi/8$ but is also generally good when the alternative is on the X-axis. We see that the bootstrap LRT and bootstrap UIT tests are indistinguishable from each other. Also, most of the time, but not always, the bootstrap tests are more powerful than the LRT and Hotelling's test.

Let's look at each distribution in details. For the normal distribution (Table 1), the two bootstrap tests are the most powerful everywhere. Among the two sign tests, the S test is better than Blumen's test when the alternative is either on the main diagonal or on the line with angle $\pi/8$. Blumen's test is better (ρ=0 or .4) or the two are equivalent (ρ=–.4) when the alternative is on the X-axis. For the contaminated normal distribution with $p = 0.1$, the bootstrap tests are better when the alternative is close to H_0 but the two sign tests are better otherwise, the test S being most powerful when the alternative is on the main diagonal or on the line with angle $\pi/8$. The two sign tests are generally preferable for the contaminated normal distribution with $p = 0.2$ with an edge for the S test when the alternative is on the main diagonal or on the line with angle $\pi/8$. When the alternative is on the X-axis, Blumen's test is better when ρ=0 or .4 but the two sign tests are close to each other (with a slight edge for the S test) when $\rho = -.4$.

For the Cauchy distribution, the S test is the best everywhere except when the alternative is on the X-axis and $\rho = .4$ where Blumen's test is better. For the beta-angle distribution, the S test is most powerful when $\rho = -.4$ and Blumen's test is preferable when $\rho = .4$. When $\rho = 0$, the bootstrap tests are the best when the alternative is on the X-axis but the S test is the best otherwise. This is an example that shows that sign tests can be competitive even for bounded distributions. Finally, for the exponential radius distribution, the average based tests are not valid, the S test is the most powerful when the alternative is on the main diagonal or on the line with angle $\pi/8$ and Blumen's test is the best when the alternative is on the X-axis (although tied with the S test when $\rho = -.4$). This simulation study shows that the new test is very competitive over a wide range of situations.

In summary, we proposed a conditionally distribution-free bivariate sign test for the one-sided location problem. The test statistic can be calculated easily and we provide a simple formula to compute the exact conditional p-value. The simulation study showed that the new test can be very competitive over a wide variety of situations. Moreover, the test is valid under very mild assumptions and can be used with certain skewed models.

Acknowledgement

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by Le Fonds Québécois de la Recherche sur la Nature et les Technologies (NATEQ). We wish to thank William Schucany for providing a copy of an unpublished technical report.

Table 1: Observed probability of rejecting H_0 for the normal distribution (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
	ρ=0									
		Alternative is on the								
S	.032	.199	.481	.786	.188	.5	.785	.195	.507	.811
Blumen	.049	.197	.502	.818	.148	.436	.733	.146	.422	.741
LRT	.032	.267	.65	.926	.225	.614	.89	.226	.597	.894
Bootstrap LRT	.046	.331*	.717*	.95*	.285*	.682*	.92*	.289*	.663*	.923*
Bootstrap UIT	.046	.332*	.716*	.947*	.279*	.679*	.918*	.281*	.659*	.921*
Hotelling	.05	.248	.616	.913	.185	.545	.849	.182	.523	.851

Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
	ρ=.4									
		Alternative is on the								
S	.035	.215	.505	.807	.206	.508	.798	.209	.525	.81
Blumen	.043	.247	.627	.913	.153	.437	.759	.13	.39	.7
LRT	.035	.337	.777	.976	.23	.619	.902	.206	.557	.856
Bootstrap LRT	.045	.381*	.807*	.983*	.268*	.659*	.918*	.24*	.604*	.879*
Bootstrap UIT	.044	.387*	.807*	.983*	.267*	.656*	.917*	.237*	.597*	.876*
Hotelling	.046	.309	.748	.97	.185	.551	.867	.157	.479	.806

Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
	ρ=−.4									
		Alternative is on the								
S	.024	.213	.508	.797	.228	.527	.802	.218	.523	.806
Blumen	.043	.196	.489	.788	.194	.483	.778	.182	.468	.776
LRT	.023	.276	.647	.905	.294	.666	.914	.282	.656	.912
Bootstrap LRT	.041	.37*	.734*	.942*	.393*	.749*	.947*	.377*	.744*	.947*
Bootstrap UIT	.041	.369*	.73*	.94*	.391*	.747*	.945*	.373*	.74*	.946*
Hotelling	.047	.257	.617	.89	.246	.604	.881	.231	.583	.878

Table 2: Observed probability of rejecting H_0 for the contaminated normal distribution with p=0.1 (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

Test	ρ=0									
		Alternative is on the								
	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.03	.204	.498*	.802	.199	.508*	.804*	.219	.538*	.802*
Blumen	.048	.199	.511*	.833*	.158	.434	.756	.168	.447	.73
LRT	.027	.186	.455	.723	.169	.421	.681	.179	.434	.667
Bootstrap LRT	.052	.244*	.489	.715	.225*	.463	.681	.238*	.475	.666
Bootstrap UIT	.051	.24*	.487	.716	.221*	.456	.679	.235*	.468	.66
Hotelling	.04	.171	.431	.706	.133	.362	.628	.14	.366	.604

Test	ρ=.4									
		Alternative is on the								
	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.037	.212	.497	.795	.218*	.505*	.82*	.219*	.526*	.804*
Blumen	.048	.239*	.616*	.905*	.165	.442	.787	.144	.397	.693
LRT	.03	.225	.54	.792	.174	.428	.711	.156	.394	.631
Bootstrap LRT	.042	.25*	.537	.761	.201	.441	.683	.188	.408	.611
Bootstrap UIT	.041	.247*	.536	.759	.197	.433	.678	.182	.404	.606
Hotelling	.037	.207	.516	.773	.137	.367	.659	.114	.323	.563

Test	ρ=−.4									
		Alternative is on the								
	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.033	.225	.507*	.799*	.22	.502*	.804*	.221	.518*	.809*
Blumen	.051	.211	.484	.792*	.194	.463	.771	.188	.475	.784
LRT	.022	.187	.428	.689	.189	.434	.697	.19	.451	.706
Bootstrap LRT	.055	.276*	.497*	.708	.283*	.507*	.714	.282*	.52*	.72
Bootstrap UIT	.054	.273*	.498*	.708	.278*	.504*	.713	.28*	.514*	.718
Hotelling	.039	.172	.399	.666	.156	.38	.646	.156	.391	.647

Table 3: Observed probability of rejecting H_0 for the contaminated normal distribution with p=0.2 (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

	ρ=0									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.035	.21*	.501	.804	.207*	.516*	.811*	.207*	.514*	.799*
Blumen	.049	.209*	.52*	.831*	.163	.452	.763	.162	.426	.733
LRT	.03	.149	.339	.591	.137	.327	.556	.137	.318	.538
Bootstrap LRT	.056	.202*	.39	.599	.192	.38	.574	.193	.37	.558
Bootstrap UIT	.056	.2*	.388	.601	.188	.376	.574	.193	.363	.554
Hotelling	.042	.138	.324	.565	.111	.276	.499	.108	.261	.475
	ρ=.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.038	.216	.51	.793	.228*	.512*	.831*	.229*	.535*	.811*
Blumen	.053	.245*	.628*	.891*	.172	.438	.787	.148	.402	.714
LRT	.031	.179	.422	.672	.146	.327	.588	.133	.302	.519
Bootstrap LRT	.046	.199	.433	.652	.171	.346	.576	.162	.328	.514
Bootstrap UIT	.045	.198	.429	.652	.169	.343	.57	.157	.322	.509
Hotelling	.042	.168	.397	.647	.119	.274	.526	.101	.241	.45
	ρ=−.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.033	.222*	.514*	.807*	.225*	.512*	.805*	.216	.495*	.827*
Blumen	.054	.204	.493	.798*	.191	.466	.779	.184	.452	.797
LRT	.022	.145	.333	.565	.151	.335	.576	.149	.325	.602
Bootstrap LRT	.061	.225*	.411	.606	.235*	.42	.616	.23*	.413	.634
Bootstrap UIT	.061	.225*	.411	.604	.232*	.415	.614	.224*	.405	.63
Hotelling	.043	.137	.308	.546	.135	.286	.523	.127	.274	.538

Table 4: Observed probability of rejecting H_0 for the Cauchy distribution (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

	ρ=0									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.032	.211*	.492*	.803*	.21*	.501*	.806*	.2*	.495*	.803*
Blumen	.049	.184	.431	.724	.167	.411	.737	.155	.423	.767
LRT	.01	.034	.08	.161	.037	.082	.167	.038	.083	.165
Bootstrap LRT	.066	.105	.147	.212	.109	.157	.221	.109	.159	.22
Bootstrap UIT	.049	.086	.126	.183	.091	.133	.192	.09	.135	.193
Hotelling	.018	.034	.07	.145	.03	.066	.134	.03	.064	.134
	ρ=.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.036	.201	.496	.827	.208*	.507*	.84*	.211*	.502*	.817*
Blumen	.051	.247*	.602*	.89*	.181	.488	.823	.17	.47	.826*
LRT	.017	.05	.118	.239	.046	.101	.212	.045	.095	.19
Bootstrap LRT	.017	.043	.091	.167	.039	.081	.153	.04	.078	.14
Bootstrap UIT	.012	.036	.077	.149	.033	.069	.133	.033	.064	.12
Hotelling	.017	.041	.103	.217	.034	.078	.174	.031	.072	.152
	ρ=−.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.031	.188*	.524*	.818*	.213*	.514*	.831*	.214*	.521*	.829*
Blumen	.049	.153	.439	.737	.17	.454	.796	.175	.481	.813
LRT	.008	.033	.079	.156	.038	.083	.173	.038	.09	.177
Bootstrap LRT	.088	.134	.188	.249	.144	.199	.27	.145	.203	.273
Bootstrap UIT	.084	.131	.184	.242	.137	.194	.261	.14	.193	.264
Hotelling	.015	.028	.069	.142	.031	.069	.147	.031	.072	.148

Table 5: Observed probability of rejecting H_0 for the beta-angle distribution (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
	ρ=0									
		Alternative is on the								
S	.031	.204	.494	.77	.22*	.526*	.796*	.198*	.499*	.822*
Blumen	.049	.186	.454	.735	.103	.276	.562	.096	.259	.602
LRT	.032	.343	.815	.986	.035	.054	.136	.034	.046	.115
Bootstrap LRT	.055	.403*	.85*	.989*	.061	.086	.189	.058	.074	.164
Bootstrap UIT	.055	.402*	.853*	.991*	.058	.084	.186	.056	.074	.164
Hotelling	.05	.319	.79	.983	.051	.058	.108	.051	.054	.099
	ρ=.4									
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.036	.21	.484	.78	.219	.52	.812	.205	.496	.803
Blumen	.048	.735*	.949*	.997*	.505*	.795*	.938*	.666*	.873*	.955*
LRT	.04	.42	.905	.998*	.201	.551	.877	.179	.447	.729
Bootstrap LRT	.044	.438	.918	.999*	.217	.568	.882	.206	.473	.739
Bootstrap UIT	.045	.435	.918	.998*	.214	.566	.88	.198	.47	.738
Hotelling	.051	.394	.885	.997*	.156	.471	.829	.145	.381	.671
	ρ=−.4									
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.036	.223*	.513*	.817*	.242*	.513*	.816*	.217*	.504*	.819*
Blumen	.047	.099	.27	.613	.105	.259	.56	.096	.254	.557
LRT	.024	.027	.06	.267	.026	.041	.125	.025	.038	.107
Bootstrap LRT	.056	.064	.112	.377	.062	.088	.217	.06	.082	.188
Bootstrap UIT	.056	.064	.113	.377	.06	.088	.212	.059	.082	.185
Hotelling	.044	.045	.065	.249	.044	.052	.112	.045	.05	.098

Table 6: Observed probability of rejecting H_0 for the exponential radius distribution (n=40; 10000 replications; in each column, an * indicates that the corresponding test is either the one with the highest power or any other that is not significantly different from it at the 5% level)

	ρ=0									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.029	.222*	.514	.823	.213*	.503*	.819*	.211*	.51*	.808*
Blumen	.044	.223*	.541*	.863*	.17	.422	.757	.156	.402	.718
LRT	.004	.034	.15	.435	.025	.101	.309	.019	.075	.225
Bootstrap LRT	.003	.022	.095	.291	.018	.068	.211	.015	.054	.159
Bootstrap UIT	.003	.026	.103	.313	.018	.068	.211	.015	.052	.156
Hotelling	.18	.248	.429	.721	.154	.213	.391	.109	.114	.216

	ρ=.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.042	.242	.555	.844	.24*	.527*	.81*	.252*	.537*	.822*
Blumen	.05	.279*	.682*	.951*	.188	.476	.8*	.167	.407	.732
LRT	.008	.079	.344	.758	.049	.17	.447	.034	.089	.246
Bootstrap LRT	.006	.04	.183	.515	.026	.093	.268	.019	.053	.152
Bootstrap UIT	.006	.045	.201	.549	.028	.096	.271	.02	.053	.151
Hotelling	.189	.339	.646	.92	.195	.294	.519	.113	.118	.222

	ρ=−.4									
		Alternative is on the								
Test	H_0	X axis			line with angle $\pi/8$			main diagonal		
S	.03	.197*	.515*	.792*	.201*	.497*	.809*	.178*	.488*	.785*
Blumen	.047	.187*	.509*	.796*	.176	.459	.795	.158	.45	.764
LRT	.186	.341	.592	.821	.388	.658	.898	.391	.702	.914
Bootstrap LRT	.355	.522	.728	.874	.567	.776	.927	.569	.805	.938
Bootstrap UIT	.349	.513	.718	.865	.56	.766	.921	.557	.799	.932
Hotelling	.183	.278	.491	.745	.331	.571	.848	.348	.632	.876

A APPENDIX

Calculation of S: Let m be the number of observations in quadrant II and IV as defined by (7). Let $0 \leq \phi^*_{[1]} < \ldots < \phi^*_{[m]} \leq \pi/2$ be the ordered ϕ angles, as defined by (8), of those m observations. Define $\phi^*_{[0]} = 0$ and $\phi^*_{[m+1]} = \pi/2$. For each $i \in \{1, \ldots, m+1\}$, let θ_i be any number in the interval $(\phi^*_{[i-1]}, \phi^*_{[i]})$. Then

$$S = \max\{S(\theta_1), \ldots, S(\theta_{m+1})\}$$

where $S(\theta)$ is given by (3).

Proof of Theorem 2.2: First, let's recall some basic facts about random walks. Let $Z_1, Z_2, \ldots$ be independent and identically distributed random variables such that $P(Z_i = -1) = P(Z_i = 1) = 1/2$. Let $\{R_i : i = 0, 1, 2, \ldots\}$ be the random walk defined by $R_0 = 0$ and $R_i = Z_1 + \ldots Z_i$. Define $M_n = \max_{0 \leq i \leq n} R_i$ to be the maximum attained by the random walk after n steps. The joint distribution of (M_n, R_n), given on page 86 of Billingsley (1968), is

$$P(M_n \geq b, R_n = v) = \begin{cases} p_n(v) & \text{if} \quad v \geq b \\ p_n(2b - v) & \text{if} \quad v \leq b \end{cases}$$

for $v \in \{-n, -n+2, \ldots, n-2, n\}$ and $b \in \{\max(0, v), \ldots, (n+v)/2\}$ where $p_n(v) = \binom{n}{(n+v)/2}(1/2)^n = P(R_n = v)$. From this, the conditional distribution of M_n given $R_n = v$ is

$$P(M_n \geq b | R_n = v) = \frac{p_n(2b - v)}{p_n(v)}. \tag{9}$$

These results are sufficient to prove Theorem 2.2. Recall that we are conditioning on $M = m$ and that A and $B = m - A$ are the number of observations in quadrant IV and II respectively. Now, we condition further on A. Given $A = a$, we have $B = m - a$ and X has the same distribution as $a + M_m$ given $R_m = b - a$. Using (9), we obtain

$$\begin{aligned} P(X \geq x | A = a, M = m) &= P(M_m \geq x - a | R_m = m - 2a) \\ &= \begin{cases} 0 & \text{if} \quad x > m \\ \frac{\binom{m}{x}}{\binom{m}{m-a}} & \text{if} \quad \max(a, m-a) \leq x \leq m \\ 1 & \text{if} \quad x < \max(a, m-a) \end{cases} \end{aligned}$$

From this and from the fact that A has a $B(m, 1/2)$ distribution, we have

$$\begin{aligned} P(X \geq x | M = m) &= \sum_{a=0}^{m} P(X \geq x | M = m, A = a) \\ &= \frac{1}{2^m} \left(\sum_{\substack{a=0 \\ \{a : x < max(a, m-a)\}}}^{m} \binom{m}{a} + \sum_{\substack{a=0 \\ \{a : max(a, m-a) \leq x \leq m\}}}^{m} \frac{\binom{m}{a}\binom{m}{x}}{\binom{m}{m-a}} \right). \end{aligned}$$

for $x \in \{m/2, \ldots, m\}$ if m is even and $x \in \{(m+1)/2, \ldots, m\}$ if m is odd.

Finally, since $S = X + Y$ where X and Y are independent and Y has a $B(n-m, 1/2)$ distribution,

$$P(S \geq x | M = m) = P(X + Y \geq x | M = m) = \sum_{y=0}^{n-m} P(X \geq x | M = m) P(Y = y)$$

and this concludes the proof.

Bootstrap procedure for the LRT: Schucany, Frawley, Gray and Wang (1999)

1. Compute LRT(original) = value of the LRT statistic for the original set of observations.
2. Subtract the mean of the sample from each of the observations to get the n residuals.
3. For $b = 1, \ldots, B$, draw with replacement a random sample of size n from the residuals and compute LRT(b) = value of the LRT statistic for that sample.
4. Let K = number of times that LRT(b) $>$ LRT(original) for the B bootstrap samples.
5. Estimated p-value = $(K{+}1)/(B{+}1)$.

References

◇ Anderson, T. W. (1984). *Multivariate Statistical Analysis.* Wiley, New York.

◇ Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

◇ Blumen, I. (1958). A New Bivariate Sign Test. *Journal of the American Statistical Association*, **53**, 448-456.

◇ Doornik, J. A. (1999). *Object-Oriented Matrix Programming Using Ox*, 3rd ed. Timberlake Consultants Press and Oxford, London.

◇ Follmann, D. (1996). A Simple Multivariate Test for One-Sided Alternatives. *Journal of the American Statistical Association* **91**, 854-861.

◇ Hodges, J. L. (1955). A Bivariate Sign Test. *The Annals of Mathematical Statistics*, **26**, 523-527.

◇ Larocque, D., Tardif, S. and van Eeden, C. (2000). Bivariate Sign Tests Based on the Sup, L_1 and L_2 Norms. *Annals of the Institute of Statistical Mathematics* **52**, 488-506.

◇ Mudholkar, G. S., Kost, J. and Subbaiah, P. (2001). Robust Tests for the Significance of Orthant Restricted Mean Vector. *Communication in Statistics - Theory and Methods* **30**, 1789-1810.

◇ Oja, H. and Nyblom, J. (1989). Bivariate Sign Tests. *Journal of the American Statistical Association*, **84**, 249-259.

◇ Perlman, M. D. (1969). One-Sided Testing Problems in Multivariate Analysis. *The Annals of Mathematical Statistics* **40**, 549-567.

◇ Perlman, M. D. and Wu, L. (1999). The Emperor's New Tests. *Statistical Science* **14**, 355-381. (with discussions)

◇ Perlman, M. D. and Wu, L. (2002). A Class of Conditional Tests for a Multivariate One-sided Alternative. *Journal of Statistical Planning and Inference* **107**, 155-171.

◇ Roy, S. N. (1953). On a Heuristic Method of Test Construction and its Use in Multivariate Analysis. *The Annals of Mathematical Statistics* **24**, 220-238.

◇ Schucany, W. R., Frawley, W. H., Gray, H. L. and Wang, S. (1999). Bootstrap Testing for Ordered Multivariate Means. Unpublished Southern Methodist University Technical Report.

◇ Sen, P. K. and Silvapulle, M.-J. (2002). An Appraisal of Some Aspects of Statistical Inference Under Inequality Constraints. *Journal of Statistical Planning and Inference* **107**, 3-43.

⋄ Sen, P. K. and Tsai, M.-T. (1999). Two-Stage Likelihood Ratio and Union-Intersection Tests for One-Sided Alternatives Multivariate Mean with Nuisance Dispersion Matrix. *Journal of Multivariate Analysis* **68**, 264-282.

⋄ Silvapulle, M.-J. (1995). A Hotelling's T^2-Type Statistic for Testing against One-Sided Hypotheses. *Journal of Multivariate Analysis* **55**, 312-319.

⋄ Tang, D.-I. (1994). Uniformly More Powerful Tests in a One-Sided Multivariate Problem. *Journal of the American Statistical Association* **89**, 1006-1011.

⋄ Wang, Y. and McDermott, M. P. (1998a). Conditional Likelihood Ratio Test for a Nonnegative Normal Mean Vector. *Journal of the American Statistical Association* **93**, 380-386.

⋄ Wang, Y. and McDermott, M. P. (1998b). A Conditional Test for a Non-negative Mean Vector Based on a Hotelling's T^2-Type Statistic. *Journal of Multivariate Analysis* **66**, 64-70.