On the distribution-freeness of a test of angular symmetry based on halfspace depth

Alexander Dürre[a], Davy Paindaveine[b]

[a] *Leiden University, Mathematical Institute, Niels Bohrweg 1, Leiden, 2333 CA, The Netherlands*

[b] *Universite libre de Bruxelles, ECARES and Department of Mathematics, 50, Av. Roosevelt, CP114/04, Brussels, 1050, Belgium*

Abstract

The problem of testing the null hypothesis of angular symmetry about a specified location in $\mathbb{R}^d$ is considered, with the focus being on a well-known test based on halfspace depth. In the bivariate case $d = 2$, the exact null distribution of the corresponding test statistic is explicitly known and turns out not to depend on the underlying angularly symmetric distribution, so that the test is *distribution-free* under the null hypothesis. Distribution-freeness, which is of course crucial to make this test applicable in practice, is further investigated here. In dimension $d = 2$, the reason *why* distribution-freeness holds is explained and it is shown through Monte Carlo exercises that distribution-freeness does not hold in dimension $d = 3$. Through suitable concepts of hyperplane arrangements, it is then investigated why the test behaves differently for $d = 2$ and $d \geq 3$. The results reveal why distribution-freeness fails, and, for a particular sample size considered to ease the presentation, they show that deviations with respect to distribution-freeness, still for $d = 3$, will actually remain very small. That these deviations will remain very small for other sample sizes is supported by a broader Monte Carlo exercise. Finally, a feasible conditional version of the test is proposed and asymptotic distribution-freeness is briefly discussed.

Keywords: angular symmetry, distribution-freeness, halfspace depth, hyperplane arrangements, symmetry testing

---

1. Introduction

A probability measure P on $\mathbb{R}$ is said to be symmetric about a real number θ if and only if $P[\theta+B] = P[\theta-B]$ for any Borel set B; throughout, $x+\lambda C$ will stand for $\{x+\lambda y : y \in C\}$. If P is absolutely continuous with respect to the Lebesgue measure, with corresponding density f, then this holds if and only if $f(\theta+x) = f(\theta-x)$ for almost any real number x. For probability measures on $\mathbb{R}^d$ with $d \geq 2$, several symmetry concepts are available in the literature; see, e.g., Serfling (2006). Arguably, the most natural extension is the concept of *centro-symmetry*: a probability measure P on $\mathbb{R}^d$ is said to be centro-symmetric about a d-vector θ if and only $P[\theta+B] = P[\theta-B]$ for any d-dimensional Borel set B. If P is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, still with corresponding density f, then P is centro-symmetric about θ if and only if $f(\theta+x) = f(\theta-x)$ for almost any $x \in \mathbb{R}^d$. Other classical symmetry concepts on $\mathbb{R}^d$ include the (stronger) elliptical symmetry and spherical symmetry; see Serfling (2006).

Symmetry has played a key role in multivariate nonparametric statistics. A prototypical example is robust location testing, that is, testing the null hypothesis $\mathcal{H}_0 : \theta = \theta_0$ against the alternative hypothesis $\mathcal{H}_1 : \theta \neq \theta_0$, where θ is a suitable location parameter and θ_0 is a fixed d-vector; this of course is to be tested on the basis of a random sample of size n from the distribution P at hand. A classical parametric approach defines θ as the mean vector of P, but this has the disadvantage of requiring a priori finite first-order moments. Many works in the literature therefore rather defined θ as the centro-symmetry center or elliptical symmetry center, which allows one to avoid any moment assumption; see, among many others, Peters and Randles (1990), Hettmansperger et al. (1994), Hettmansperger et al. (1997), Hallin and Paindaveine (2002), Ollila et al. (2002), and Oja (2010). Obviously, such a nonparametric approach can deal with arbitrarily heavy tails, hence is robust to some extent, yet the price to pay to achieve this is not negligible since centro-symmetry, and even more so elliptical symmetry, is a rather stringent assumption.

This motivates considering a weaker multivariate symmetry concept, namely the concept of *angular symmetry* (see Liu (1988, 1990)): a probability measure P on $\mathbb{R}^d$ is said to be angularly symmetric about a d-vector θ if and only if $P[\theta+C] = P[\theta-C]$ for any d-dimensional cone C with apex at the origin of $\mathbb{R}^d$—that is, for any subset C of $\mathbb{R}^d$ such that if $x \in C$, then $\lambda x \in C$ for any $\lambda \geq 0$. Figure 1 shows some level sets of the bivariate density (expressed in polar coordinates, with $r \in [0, \infty)$ and $\psi \in [0, 2\pi)$)

$$f(z) = f(r\cos\psi, r\sin\psi) = \frac{2\sqrt{2+\sin\psi}}{\pi^2(1 + r^4(2+\sin\psi))}, \tag{1}$$

for which the corresponding probability measure is angularly symmetric about the origin of $\mathbb{R}^2$. Clearly, angular symmetry is a much weaker concept than centro-symmetry; in particular, any probability measure on $\mathbb{R}$ is angularly symmetric about any of its medians (for $d > 1$, any probability measure on $\mathbb{R}^d$ may have at most one angular symmetry center). In the context of multivariate location testing, it is thus advantageous to define the location functional θ as the possible angular symmetry center.

As a consequence, angular symmetry is often considered when comparing various location functionals (Zuo and Serfling (2000b)) and is often assumed when performing nonparametric hypothesis testing for multivariate location (see, e.g., Oja and Randles (2004) for conditional testing, or Larocque (2003) and Larocque et al. (2007) in the special case of cluster-correlated data). Angular symmetry is also a natural assumption to tests for randomness based on multivariate runs concepts (Van Bever (2016)), and it is relevant in the context of shrinkage estimation (Perlman and Chaudhuri (2012)). Obviously, due to its very tight connection with *halfspace symmetry* (Zuo and Serfling (2000b)), angular symmetry has played an important role in the halfspace depth literature; see, e.g., Liu et al. (1999) and Zuo and Serfling (2000a). In the same vein, angular symmetry was also assumed, e.g., in Chen and Tyler (2002) when deriving the influence function and maximum contamination bias function of the Tukey median.

While the angular symmetry assumption is much weaker than its centro-symmetry counterpart, it is still a non-trivial assumption for $d \geq 2$, which makes it desirable to be able *to test for angular symmetry in* $\mathbb{R}^d$—for $d \geq 2$ only, since angular symmetry always holds over the real line. Few tests of angular symmetry are available in the literature, and the present note focuses on the test defined in Rousseeuw and Struyf (2002), which we now recall.

To this end, we need to define the Tukey (1975) *halfspace depth* concept : if P is a probability measure on $\mathbb{R}^d$, then the halfspace depth of $\theta(\in \mathbb{R}^d)$ with respect to P is

$$HD(\theta, P) = \inf_{u \in \mathcal{S}^{d-1}} P[H_{\theta,u}],$$

where $\mathcal{S}^{d-1} := \{z \in \mathbb{R}^d : \|z\|^2 = z'z = 1\}$ is the unit sphere in $\mathbb{R}^d$ and $H_{\theta,u} := \{z \in \mathbb{R}^d : u'(z - \theta) \geq 0\}$ is the closed halfspace with θ on

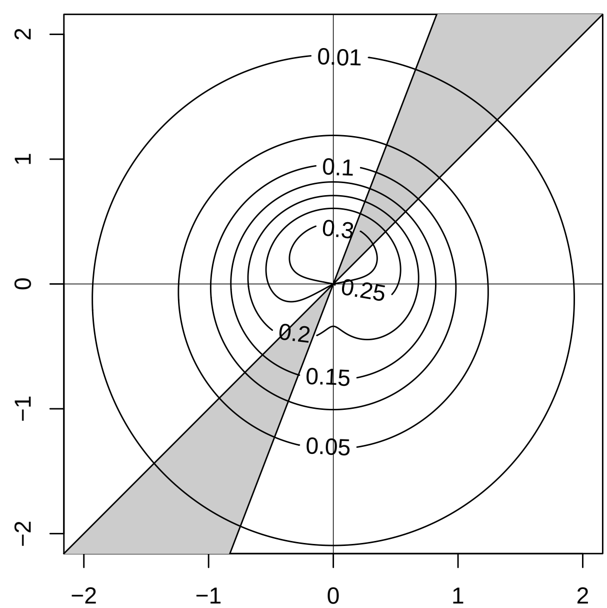

Figure 1: Some level sets of the density in (1). The corresponding probability measure is angularly symmetric (but of course not centro-symmetric) about the origin of $\mathbb{R}^2$. Both grey cones with apex at the origin have a probability equal to 1/15.

its boundary hyperplane and whose normal "direction" is u. This measure of *statistical depth* (see, e.g., Zuo and Serfling (2000a)) indicates how central the fixed location θ is with respect to the probability measure P. It is known that, for any probability measure P on $\mathbb{R}^d$, there exists $\theta \in \mathbb{R}^d$ that maximizes $HD(\theta, P)$ (see Proposition 7 in Rousseeuw and Ruts (1999)), and that the corresponding maximal depth satisfies

$$\frac{1}{d+1} \leq \max_{\theta \in \mathbb{R}^d} HD(\theta, P) \leq \frac{1}{2} + \frac{1}{2} \max_{\theta \in \mathbb{R}^d} P[\{\theta\}];$$

see Proposition 9 in Rousseeuw and Ruts (1999) (for the lower bound) and Lemma 1 in Rousseeuw and Struyf (2004) (for the upper bound). Remarkably, the upper bound holds pointwise in θ and allows one to characterize angular symmetry; more precisely,

$$HD(\theta, P) \leq \frac{1}{2} + \frac{1}{2} P[\{\theta\}] \quad \text{for any } \theta \in \mathbb{R}^d \tag{2}$$

and equality holds at some $\theta_0 \in \mathbb{R}^d$ *if and only if* P is angularly symmetric about θ_0 (and $HD(\theta_0, P)$ is the maximal depth achieved by P); see Theorem 30.2.1 in Rousseeuw and Struyf (2002), as well as Theorems 1–2 in Rousseeuw and Struyf (2004). In particular, restricting for the sake of simplicity to absolutely continuous probability measures, P is angularly symmetric about θ_0 if and only if $HD(\theta_0, P) = 1/2$.

With the motivation provided earlier, consider then the problem of testing the null hypothesis that the probability measure P at hand (which we will throughout tacitly assume to be absolutely continuous) is angularly symmetric about θ_0 (a fixed d-vector), based on a random sample $X_1, \ldots, X_n$ from P. The previous paragraph suggests considering the test that rejects the null hypothesis for small values of

$$T_n := HD(\theta_0, P_n),$$

where P_n is the empirical probability measure associated with $X_1, \ldots, X_n$, so that T_n is the sample halfspace depth of θ_0 with respect to the sample at hand. This test was proposed in Section 30.3 of Rousseeuw and Struyf (2002). Implementation of this test of course requires distribution-freeness of T_n under the null hypothesis. As mentioned in Rousseeuw and Struyf (2002), a result from Daniels (1954) actually states that, for $d = 2$, the exact null distribution of T_n is given by

$$P\big[T_n \leq \tfrac{k}{n}\big] = \begin{cases} 0 & \text{if } k < 0 \\ \frac{n-2k}{2^{n-1}} \sum_{j=0}^{\lfloor k/(n-2k) \rfloor} \binom{n}{n-k+j(n-2k)} & \text{if } 0 \leq k \leq \lfloor \frac{n-1}{2} \rfloor \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

Clearly, this null distribution does not depend on the particular angularly symmetric distribution at hand, so that, in the bivariate case $d = 2$, the test statistic T_n is exactly distribution-free under the null hypothesis, which allows one to perform the test proposed in Rousseeuw and Struyf (2002).

Daniels' result establishes distribution-freeness in dimension $d = 2$, yet it leaves open a number of natural key questions we intend to address in the present paper. First and foremost, it would be nice to understand *why* T_n is distribution-free under the null hypothesis for $d = 2$. We argue that this is actually even more fundamental than the explicit formula in (3) (would one know that the test statistic is distribution-free under the null hypothesis, then one could approximate arbitrarily well, for any given significance level α, the required critical value by picking the corresponding sample quantile in a large collection of simulated values of T_n computed within samples generated from an arbitrary distribution that is angularly symmetric with respect to θ_0, such as, e.g., the multinormal distribution with mean vector θ_0 and identity covariance matrix). Second, it is of interest to know whether or not distribution-freeness under the null hypothesis extends to higher dimensions $d \geq 3$ and, if not, to understand why distribution-freeness fails. We aim at answering these questions in the present work.

The outline of the paper is as follows. In Section 2, we explain how distribution-freeness in the bivariate case $d = 2$ actually results from suitable invariance arguments. In Section 3, we explain why similar arguments cannot be used to establish distribution-freeness in higher dimensions and we show through simulations that distribution-freeness actually does not hold in general for $d = 3$. In Section 4, we investigate distribution-freeness through another approach that links the null distributional properties of T_n to a concept of *hyperplane arrangements*. This allows us to establish in an alternative way distribution-freeness for $d = 2$ and also identifies why distribution-freeness fails in higher dimensions. For a particular sample size, we provide theoretical support why deviations from distribution-freeness will actually remain very small for $d = 3$. In Section 5, we perform a Monte Carlo study showing that departures from distribution-freeness remain very small for larger sample sizes, too. We conclude by providing final comments in Section 6, where we describe a feasible conditional version of the test considered in this work and shortly discuss asymptotic distribution-freeness.

2. Distribution-freeness in the bivariate case

Throughout the paper, we will consider the problem of testing for angular symmetry about the origin of $\mathbb{R}^d$, that is, we restrict to $\theta_0 = 0$. This is of course without any loss of generality (to test for angular symmetry about a fixed $\theta_0 \neq 0$, one can apply a test for angular symmetry about the origin to the centered observations $X_1 - \theta_0, \ldots, X_n - \theta_0$). The following result states that, in the bivariate case $d = 2$, the test statistic $T_n = HD(0, P_n)$ is distribution-free under the null hypothesis as soon as absolute continuity is assumed.

Theorem 2.1. *Let $\mathcal{P}_2$ be the collection of probability measures on $\mathbb{R}^2$ that are angularly symmetric about the origin of $\mathbb{R}^2$ and admit a density with respect to the Lebesgue measure. Let $X_1, \ldots, X_n$ be a random sample from $P \in \mathcal{P}_2$. Then, the distribution of $T_n = T_n(X_1, \ldots, X_n)$ does not depend on P.*

Proof of Theorem 2.1. We will prove the result by showing that, for any real number t and any $P \in \mathcal{P}_2$,

$$P[T_n(X_1, \ldots, X_n) \leq t] = P_0[T_n(X_1, \ldots, X_n) \leq t],$$

where P_0 is the probability measure associated with the uniform distribution over $\mathcal{S}^1$. To this end, we consider the transformation $g = g_b \circ g_a$, where

$$g_a(x_1, \ldots, x_n) = \left(\frac{x_1}{\|x_1\|}, \ldots, \frac{x_n}{\|x_n\|} \right)$$

projects observations radially onto $\mathcal{S}^1$, and where

$$g_b(x_1, \ldots, x_n) = (h_P(x_1), \ldots, h_P(x_n))$$

involves the transformation h_P defined (with obvious notation, in polar coordinates) through

$$h_P(x) = h_P(\cos\psi, \sin\psi) = (\cos F_P(\psi), \sin F_P(\psi)),$$

with $F_P(\psi) = 2\pi P[\arg(X_1) \leq \psi]$; here, $\arg(X_1)$ denotes the random angle Ψ_1 such that $X_1/\|X_1\| = (\cos\Psi_1, \sin\Psi_1)$. Since $P \in \mathcal{P}_2$, the random variable $\arg(X_1)$ admits a density with respect to the Lebesgue measure over $[0, 2\pi)$, so that $F_P(0) = 0$. Angular symmetry of P about the origin also implies that

$$F_P(\psi + \pi) = F_P(\psi) + \pi \qquad \text{for any } \psi \in [0, \pi). \tag{4}$$

From the usual probability integral transform, $F_P(\arg(X_1))$ is uniformly distributed over $[0, 2\pi)$; thus, if $(X_1, \ldots, X_n)$ is a random sample from P, then $g(X_1, \ldots, X_n)$ is a random sample from P_0.

Now, would the test statistic T_n be invariant under the transformation g, in the sense that

$$T_n(g(x_1, \ldots, x_n)) = T_n(x_1, \ldots, x_n)$$

for any $x_1, \ldots, x_n \in \mathbb{R}^2$, then we would have

$$P[T_n(X_1, \ldots, X_n) \leq t] = P[T_n(g(X_1, \ldots, X_n)) \leq t] = P_0[T_n(X_1, \ldots, X_n) \leq t]$$

for any t, which would establish the result. It therefore only remains to prove the aforementioned invariance of T_n. To do so, note that, for any $x_1, \ldots, x_n \in \mathbb{R}^2$, we have (with obvious notation)

$$\begin{aligned} T_n(g_a(x_1, \ldots, x_n)) &= HD(0, g_a(x_1, \ldots, x_n)) \\ &= HD(0, (x_1, \ldots, x_n)) \\ &= T_n(x_1, \ldots, x_n). \end{aligned} \tag{5}$$

Now, consider arbitrary $x_1, \ldots, x_n \in \mathcal{S}^1$, and note that

$$\begin{aligned} &T_n(x_1, \ldots, x_n) = HD(0, (x_1, \ldots, x_n)) \\ &\quad = \min_{\psi \in [0,\pi)} \min \left(\sum_{i=1}^n \mathbb{I}[\psi_i \in [\psi, \psi + \pi]], n - \sum_{i=1}^n \mathbb{I}[\psi_i \in (\psi, \psi + \pi)] \right), \end{aligned}$$

where $\mathbb{I}$ is an indicator function. Since $F_P(0) = 0$ and $F_P(\pi) = F_P(0) + \pi = \pi$ (this follows from (4)), continuity of F_P implies that $F_P : [0, \pi) \to [0, \pi)$ is surjective, so that

$$
\begin{aligned}
&T_n(g_b(x_1, \ldots, x_n)) \\
&\quad = \min_{\psi \in [0,\pi)} \min \left(\sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in [\psi, \psi+\pi]], n - \sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in (\psi, \psi+\pi)] \right) \\
&\quad = \min_{\psi \in [0,\pi)} \min \left(\sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in [F_P(\psi), F_P(\psi)+\pi]], \right. \qquad (6) \\
&\qquad\qquad \left. n - \sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in (F_P(\psi), F_P(\psi)+\pi)] \right).
\end{aligned}
$$

Using (4) then the monotonicity of F_P thus yields

$$
\begin{aligned}
&T_n(g_b(x_1, \ldots, x_n)) \\
&\quad = \min_{\psi \in [0,\pi)} \min \left(\sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in [F_P(\psi), F_P(\psi+\pi)]], \right. \qquad (7) \\
&\qquad\qquad \left. n - \sum_{i=1}^n \mathbb{I}[F_P(\psi_i) \in (F_P(\psi), F_P(\psi+\pi))] \right) \\
&\quad = \min_{\psi \in [0,\pi)} \min \left(\sum_{i=1}^n \mathbb{I}[\psi_i \in [\psi, \psi+\pi]], n - \sum_{i=1}^n \mathbb{I}[\psi_i \in (\psi, \psi+\pi)] \right) \\
&\quad = T_n(x_1, \ldots, x_n). \qquad (8)
\end{aligned}
$$

Since invariance of T_n follows from (5) and (8), the proof is complete. □

It is worth noting that the key ingredient bringing distribution-freeness in the proof of Theorem 2.1 is the probability integral transform used in g_b. The use of this, by nature *univariate*, transform is possible since the unit circle, or equivalently its angle parametrization $[0, 2\pi)$, is a one-dimensional space; this of course is not the case for higher-dimensional unit spheres $\mathcal{S}^{d-1}$, with $d \geq 3$, which would call for higher-dimensional integral probability transforms; see below.

Inspection of the proof reveals that distribution-freeness extends to the collection of probability measures P over $\mathbb{R}^2$ such that P is angularly symmetric about the origin, does not charge the origin (i.e., $P[\{0\}] = 0$), and is such that if X has distribution P, then $X/\|X\|$ admits a density with respect to the surface area measure on $\mathcal{S}^1$ (the latter absolute continuity is indeed sufficient to allow for the use of the probability integral transform in g_b). This is of course a wider class of distributions than the one considered in Theorem 2.1 since it also contains, e.g., the uniform distribution on $\mathcal{S}^1$.

3. Towards higher dimensions

To the best of our knowledge, whether or not the distribution-freeness result from Theorem 2.1 extends to higher dimensions remains an open question. The proof of Theorem 2.1, that relies on invariance arguments, unfortunately does not extend to higher dimensions d, as we now explain. For the sake of simplicity, let us focus on $d = 3$. Of course, the halfspace depth of the origin is still invariant when observations are projected radially through g_a onto the unit sphere $\mathcal{S}^2$ of $\mathbb{R}^3$, that is, (5) still holds for $d = 3$. Now, the transformation g_b needs to be generalized to a transformation from $\mathcal{S}^2$ to $\mathcal{S}^2$ that maps an arbitrary angularly symmetric (equivalently, centro-symmetric) absolutely continuous distribution on $\mathcal{S}^2$ onto the uniform distribution on $\mathcal{S}^2$. This requires

- either a bivariate probability integral transform such as, e.g., the usual Rosenblatt transformation, that, in spherical coordinates (ψ, ζ), with longitude $\psi \in [0, 2\pi)$ and colatitude $\zeta \in [0, \pi]$, is of the form $(\psi, \zeta) \mapsto (F_{P,\psi}(\psi), G_{P,\psi,\zeta}(\psi, \zeta))$, where $F_{P,\psi}$ is still (2π times) the cumulative distribution function of the longitude under P and where $G_{P,\psi,\zeta}$ is (π times) the cumulative distribution function of the colatitude conditional on the value ψ of the longitude,

- or an empirical concept of bivariate center-outward distribution function, such as the one relying on measure transportation ideas proposed in Hallin et al. (2021).

While both resulting transformations g_b would indeed yield the uniform distribution on $\mathcal{S}^2$, it is not so that the halfspace depth of the origin will be invariant under such transformations for any P (measure transportation, however, would naturally lead to alternative distribution-free tests). The proof therefore collapses in dimension $d = 3$ (and more generally in dimensions $d \geq 3$), and distribution-freeness in such dimensions therefore remains an open question.

In order to explore distribution-freeness in dimension $d = 3$, we conducted the following Monte Carlo exercise. We generated $M = 10^7$ mutually independent random samples of size $n = 6$ from both following probability measures on $\mathbb{R}^3$: P_1, the standard trivariate normal distribution, and P_2, an equal-weight mixture with one mixture component being the trivariate normal distribution with mean vector $-5e_1 = (-5, 0, 0)'$ and identity covariance matrix, and the other mixture component being the trivariate normal distribution with mean vector $5e_1$ and still identity covariance matrix. Note that P_1 and P_2 are centro-symmetric—hence also angularly symmetric—about the origin of $\mathbb{R}^3$. We evaluated T_n in each sample through the R package `depth` (Genest et al. (2019)), which uses the fast exact algorithm from Rousseeuw and Struyf (1998). When generating random samples of size $n = 6$ from an absolutely continuous distribution (which will almost surely provide observations in general position), only the values 0, 1/6 and $2/6 = 1/3$ of T_n may arise with positive probability (this easily follows by applying (2) to the empirical distribution P_n at hand). Table 1 then provides the resulting approximations of the probabilities $P[T_n = 0]$, $P[T_n = 1/6]$, and $P[T_n = 1/3]$ for both P_1 and P_2. The results are compatible with the fact that $P[T_n = 0]$ actually does not depend on the underlying (absolutely continuous) angularly symmetric distribution P (see Wendel (1962)), yet they suggest that the distribution of T_n is not the same under P_1 and under P_2. To investigate this, we performed a formal test for the null hypothesis that T_n has the same distribution under both probability measures, that is, we performed a chi-squared test of homogeneity for the vectors of probabilities $(P_j[T_n = 0], P_j[T_n = \frac{1}{6}], P_j[T_n = \frac{1}{3}])$, $j = 1, 2$. The resulting p-value, for $(3-1)\times(2-1) = 2$ degrees of freedom, is below 2.2×10^{-16}; thus, although differences are small in Table 1, we may conclude, at all usual significance levels, that distribution-freeness does not hold in dimension $d = 3$.

4. Distribution-freeness and hyperplane arrangements

In this section, we investigate why the cases $d = 2$ and $d \geq 3$ behave differently, by making use of hyperplane arrangements. Unless it coincides with the origin of $\mathbb{R}^d$, each observation X_i defines the hyperplane $h_{X_i} := \{x \in \mathbb{R}^d : x'X_i = 0\}$ that contains the origin of $\mathbb{R}^d$ and is orthogonal to $u_{X_i} := X_i/\|X_i\|$. The set of hyperplanes $h_{X_1,\ldots,X_n} := \{h_{X_1}, \ldots, h_{X_n}\}$ is what we will call *a hyperplane arrangement*. It cuts $\mathbb{R}^d$ into $n_{X_1,\ldots,X_n}$ connected open sets,

	P_1	P_2
$P[T_n = 0]$	0.49995 [0.49964 0.50026]	0.50012 [0.49981 0.50043]
$P[T_n = \frac{1}{6}]$	0.49416 [0.49385 0.49447]	0.49364 [0.49333 0.49395]
$P[T_n = \frac{1}{3}]$	0.00589 [0.00584 0.00594]	0.00624 [0.00619 0.00629]

Table 1: Empirical probabilities of T_n for $n = 6$ (together with the corresponding 95% confidence intervals) under both probability measures P_1 and P_2 considered in Section 2; for each probability measure, estimation is based on $M = 10^7$ mutually independent samples. When performing a chi-squared test for the null hypothesis of distribution-freeness, the p-value is below 2.2×10^{-16}.

in the sense that

$$\mathbb{R}^d \backslash h_{X_1,\ldots,X_n} = A_1 \cup \ldots \cup A_{n_{X_1,\ldots,X_n}},$$

where the A_j's are connected (pairwise disjoint) open cones (note that these cones are defined up to a permutation only, which will need to be handled in the sequel). A two-dimensional hyperplane arrangement resulting from a random sample of size $n = 4$ is shown in Figure 2. There, the plane $\mathbb{R}^2$ is cut into 8 cones. If all observations are different from the origin and no pair $\{i, i'\}$ with $i \neq i'$ provides $u_{X_i} = u_{X_{i'}}$ or $u_{X_i} = -u_{X_{i'}}$ (which will be the case almost surely if $X_1, \ldots, X_n$ form a random sample from an absolutely continuous distribution), then the plane is cut into $n_{X_1,\ldots,X_n} = 2n$ cones.

Surprisingly, it is also the case in dimension d that, if observations are randomly sampled from an absolutely continuous distribution, then $n_{X_1,\ldots,X_n}$ is almost surely equal to a constant that only depends on n and d. This readily follows from the following result, in which, by definition, we say that n hyperplanes are *in general position* if and only if the intersection of every selection of $k(\leq d)$ hyperplanes has dimension $d - k$.

Proposition 4.1 (Winder (1966)). *If n hyperplanes of $\mathbb{R}^d$ go through a common point and are in general position, then these hyperplanes divide $\mathbb{R}^d$ into*

$$n_{n,d} := 2 \sum_{i=0}^{d-1} \binom{n-1}{i}$$

regions.

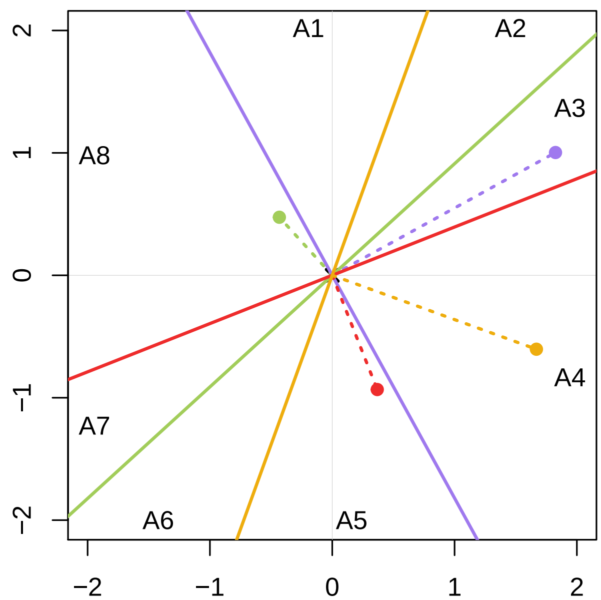

Figure 2: The hyperplane arrangement associated with $n = 4$ observations randomly sampled from the bivariate normal distribution. The coloured dots represent the observations X_i, $i = 1, 2, 3, 4$, whereas the solid lines are the corresponding hyperplanes h_{X_i}. The hyperplane arrangement determines $2n = 8$ open cones $A_1, \ldots, A_8$.

Under absolute continuity, there are thus almost surely $n_{n,d}$ cones A_j, $j = 1, \ldots, n_{n,d}$. To each A_j, we associate its *halfspace sign*

$$s_{j,X_1,\ldots,X_n} = \begin{pmatrix} \mathbb{I}[A_j \subset H^+_{X_1}] - \mathbb{I}[A_j \subset H^-_{X_1}] \\ \vdots \\ \mathbb{I}[A_j \subset H^+_{X_n}] - \mathbb{I}[A_j \subset H^-_{X_n}] \end{pmatrix}, \tag{9}$$

where $H^{\pm}_{X_i} := \{x \in \mathbb{R}^d : \pm x'X_i > 0\}$ are the open halfspaces "above" and "below" the hyperplane H_{X_i}, respectively (we refer to Mozharovskyi (2016) for a related concept of hyperplane arrangements and similar encoding). Writing $\underline{1} := (1, \ldots, 1)' \in \mathbb{R}^d$ and $\|v\|_1 := \sum_{i=1}^d |v_i|$ for $v \in \mathbb{R}^d$, and assuming that the observations $X_1, \ldots, X_n$ form a random sample from an absolutely continuous distribution, we almost surely have

$$\begin{aligned} T_n = HD(0, P_n) &= \frac{1}{n} \min_{j=1,\ldots,n_{n,d}} \sum_{i=1}^n \mathbb{I}[A_j \subset H^+_{X_i}] \\ &= \frac{1}{2n} \min_{j=1,\ldots,n_{n,d}} \|s_{j,X_1,\ldots,X_n} + \underline{1}\|_1, \end{aligned} \tag{10}$$

which shows that T_n can be obtained from the collection of halfspace signs. The next result decomposes an angularly symmetric random vector into three components, which will give further insight into the distribution of halfspace signs (see the appendix for a proof).

Lemma 4.1. *Let the random d-vector X admit a density with respect to the Lebesgue measure. Denote as $e_d = (0, \ldots, 0, 1)'$ the last vector of the canonical basis of $\mathbb{R}^d$. Then, X is angularly symmetric about the origin of $\mathbb{R}^d$ if and only if X can be decomposed into*

$$X = O \times S \times M := \mathrm{Sign}(e_d'X) \times \left(\mathrm{Sign}(e_d'X) \frac{X}{\|X\|} \right) \times \|X\|,$$

where O satisfies $P[O = 1] = P[O = -1] = \frac{1}{2}$ and is independent of S.

If X is angularly symmetric about the origin of $\mathbb{R}^d$, then we will refer to O, S and M as the *orientation*, the *sign* and the *magnitude* of X, respectively. Note that the distribution of O is completely specified, whereas those of S and M depend on the angularly symmetric distribution at hand. Under the null hypothesis of angular symmetry about the origin of $\mathbb{R}^d$, any distribution-free test statistic can then in principle depend on $X_1, \ldots, X_n$ only through the corresponding orientations $O_1, \ldots, O_n$. Note that the signs $S_1, \ldots, S_n$ fully determine the halfspace arrangement associated with $X_1, \ldots, X_n$, hence also the halfspace signs $s_{j,X_1,\ldots,X_n} \in \{-1, 1\}^n$ that could be obtained when considering all possible orientations. The orientations $O_1, \ldots, O_n$ flip randomly signs, whereas the magnitudes $M_1, \ldots, M_n$ have no influence on halfspace signs at all. This will play a key role in our analysis.

Now, we associate with the sample $X_1, \ldots, X_n$ at hand the sign matrix

$$(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{d,n},S_1,\ldots,S_n}), \tag{11}$$

where $S_1, \ldots, S_n$ are the signs of $X_1, \ldots, X_n$ from the decomposition given in Proposition 4.1. Since labelling of the regions A_j, $j = 1, \ldots, n_{n,d}$, is arbitrary, this matrix, however, is properly defined up to permutations of its columns only. Also, the test statistic $T_n = HD(0, P_n)$ is invariant under permutations of the observations $X_1, \ldots, X_n$, which corresponds to permuting the rows of the sign matrix in (11). Consequently, rather than this matrix, we will consider the corresponding *halfspace arrangement sign set* of $X_1, \ldots, X_n$, defined as

$$\text{HASS}(X_1, \ldots, X_n) = \cup_{K \in \mathcal{M}_n} \cup_{L \in \mathcal{M}_{n_{n,d}}} \big\{ K(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{d,n},S_1,\ldots,S_n}) L \big\},$$

where $\mathcal{M}_k$ is the collection of $k \times k$ permutation matrices (that is, the collection of matrices obtained by permuting arbitrarily the columns of the $k \times k$ identity matrix). By definition, the HASS of a sample only depends on the signs $S_1, \ldots, S_n$ but not on the orientations $O_1, \ldots, O_n$, which will provide the following result.

Theorem 4.1. *Let the random d-vector X be angularly symmetric about the origin of $\mathbb{R}^d$ and admit a density with respect to the Lebesgue measure, and let $X_1, \ldots, X_n$ be mutually independent copies of X. Let $g : \mathbb{R}^{n \times n_{n,d}} \to \mathbb{R} : C \mapsto g(C)$ be a function that is invariant under permutations of the rows/columns of its matrix argument C. Then, conditional on the HASS of $X_1, \ldots, X_n$, the statistic*

$$R = R(X_1, \ldots, X_n) = g((s_{1,X_1,\ldots,X_n} \ \ldots \ s_{n_{n,d},X_1,\ldots,X_n}))$$

is distribution-free. In particular, conditional on the HASS of $X_1, \ldots, X_n$, the statistic $T_n = HD(0, P_n)$ is distribution-free.

Proof of Theorem 4.1. By definition of halfspace signs, we have

$$s_{j,X_1,\ldots,X_n} = \begin{pmatrix} O_1(s_{j,S_1,\ldots,S_n})_1 \\ \vdots \\ O_n(s_{j,S_1,\ldots,S_n})_n \end{pmatrix} = \begin{pmatrix} O_1 & & 0 \\ & \ddots & \\ 0 & & O_n \end{pmatrix} s_{j,S_1,\ldots,S_n},$$

for $j = 1, \ldots, n_{n,d}$, which rewrites

$$\big(s_{1,X_1,\ldots,X_n} \ \ldots \ s_{n_{n,d},X_1,\ldots,X_n}\big) = \begin{pmatrix} O_1 & & 0 \\ & \ddots & \\ 0 & & O_n \end{pmatrix} \big(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{n,d},S_1,\ldots,S_n}\big). \tag{12}$$

For any K, L as in the definition of HASS, we then have

$$\begin{aligned} &K\big(s_{1,X_1,\ldots,X_n} \ \ldots \ s_{n_{n,d},X_1,\ldots,X_n}\big)L \\ &\quad = K \begin{pmatrix} O_1 & & 0 \\ & \ddots & \\ 0 & & O_n \end{pmatrix} K^{-1}K\big(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{n,d},S_1,\ldots,S_n}\big)L \\ &\quad = \begin{pmatrix} O_{\pi(1)} & & 0 \\ & \ddots & \\ 0 & & O_{\pi(n)} \end{pmatrix} K\big(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{n,d},S_1,\ldots,S_n}\big)L \end{aligned}$$

for some permutation π of $\{1, \ldots, n\}$ (that depends on K). By assumption, it follows that, for any such K, L,

$$\begin{aligned} R &= g((s_{1,X_1,\ldots,X_n} \ \ldots \ s_{n_{n,d},X_1,\ldots,X_n})) \qquad (13) \\ &= g\left(\begin{pmatrix} O_{\pi(1)} & & 0 \\ & \ddots & \\ 0 & & O_{\pi(n)} \end{pmatrix} K\big(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{n_{n,d},S_1,\ldots,S_n}\big)L\right). \end{aligned}$$

Now, conditional on $(S_1, \ldots, S_n)$, the distribution of $(O_1, \ldots, O_n)$ is uniform over $\{-1, 1\}^n$, hence so is the distribution of $(O_{\pi(1)}, \ldots, O_{\pi(n)})$. Since the fact that (13) holds for any K, L implies that the distribution of R conditional on the HASS of $X_1, \ldots, X_n$ coincides with the distribution of R conditional on $(S_1, \ldots, S_n)$, the result follows. □

Theorem 4.1 shows that T_n is distribution-free under the null hypothesis conditional on the HASS of $X_1, \ldots, X_n$, but this does not entail that T_n is (unconditionally) distribution-free under the null hypothesis, as the distribution of the HASS of $X_1, \ldots, X_n$ might not be distribution-free itself. In dimension $d = 2$, however, distribution-freeness of T_n now readily follows from the next result.

Theorem 4.2. *Assume that the bivariate observations $X_1, \ldots, X_n$ are randomly sampled from an absolutely continuous distribution. Then, with probability one, the HASS of $X_1, \ldots, X_n$ is $\big\{KJL : K \in \mathcal{M}_n, L \in \mathcal{M}_{2n}\big\}$, with*

$$J := \begin{pmatrix} 1 & -1 & -1 & \ldots & -1 & 1 & 1 & \ldots & 1 \\ 1 & 1 & -1 & \ldots & -1 & -1 & 1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & -1 & -1 & -1 & \ldots & 1 \\ & & & & & & & & \\ 1 & 1 & 1 & \ldots & -1 & -1 & -1 & \ldots & 1 \\ 1 & 1 & 1 & \ldots & -1 & -1 & -1 & \ldots & -1 \end{pmatrix}.$$

In particular, the HASS of $X_1, \ldots, X_n$ is almost surely unique.

Proof of Theorem 4.2. By definition, the signs $S_1, \ldots, S_n$ belong to the upper half-circle $\{x \in \mathcal{S}^1 : e_2'x \geq 0\}$, and, from absolute continuity, they belong almost surely to $\{x \in \mathcal{S}^1 : e_2'x > 0\}$. Incidentally, note that this implies that, with probability one,

$$e_2 \in H^+_{S_1} \cap \ldots \cap H^+_{S_n}. \tag{14}$$

From absolute continuity again, the signs are pairwise different with probability one. Now, the concept of HASS being invariant under permutation of the observations, we may assume without any loss of generality that the signs $S_1, \ldots, S_n$ are ordered clockwise (equivalently, that their first coordinates are in ascending order). Similarly, since the concept of HASS does not depend on the labelling of the open cones A_j, we will define A_1 as the cone containing e_2 (note that (14) states that e_2 belongs to an open cone with probability one) and we define $A_2, \ldots, A_{2n}$ as the next cones that are found when turning in the clockwise direction (this cone labelling is actually the one that was adopted in Figure 2). Note that (14) implies that, with probability one,

$$s_{1,S_1,\ldots,S_n} = (1, 1, \ldots, 1)'$$

When going from one region A_j to the next region A_{j+1}, exactly one entry of the corresponding halfspace sign will change (from -1 to 1, or from 1 to

-1). Since $A_{n+j} = -A_j$ for any $j = 1, \ldots, n$, we must in particular have

$$s_{n+1,S_1,\ldots,S_n} = -s_{1,S_1,\ldots,S_n} = (-1, -1, \ldots, -1)'.$$

Consequently, the n successive entry changes in the halfspace signs incurred when going from A_1 to A_{n+1} must all be a 1 becoming a -1. It also follows that the n successive entry changes in the halfspace signs incurred when going from A_{n+1} to A_{2n} must all be a -1 becoming a 1. It remains to figure out which entry changes in which step in the corresponding halfspace signs. Let us start by going from $s_{1,S_1,\ldots,S_n}$ to $s_{2,S_1,\ldots,S_n}$. Since A_1 and A_2 as well as $S_1, \ldots, S_n$ are ordered clockwise and since all signs are on the upper half-circle, it must be $H^+_{S_1}$ that does not overlap A_2, so that $s_{1,S_1,\ldots,S_n}$ and $s_{2,S_1,\ldots,S_n}$ differ in their first entry only. Repeating this argument, we obtain that, for any $j = 1, \ldots, n$, the halfspace signs $s_{j,S_1,\ldots,S_n}$ and $s_{j+1,S_1,\ldots,S_n}$ differ in their jth entry only. This fully determines $s_{j,S_1,\ldots,S_n}$, $j = 1, \ldots, n$, and we have

$$(s_{1,S_1,\ldots,S_n} \ \ldots \ s_{2n,S_1,\ldots,S_n}) = J, \tag{15}$$

where J is the matrix defined in the statement of the theorem (the last n halfspace signs in (15) were obtained from the relation $s_{j+n,S_1,\ldots,S_n} = -s_{j,S_1,\ldots,S_n}$, $j = 1, \ldots, n$, that results from $A_{j+n} = -A_j$). This establishes the result. □

Theorem 4.2 does not extend to higher dimensions (indeed, would it hold in dimension $d = 3$, then T_n would be distribution-free under the null hypothesis for $d = 3$, which, as we showed at the end of Section 2, is not the case). In other words, HASS are not unique for $d \geq 3$, at least not for any sample n. Actually, for $n = 6$, there exist four different HASS in dimension $d = 3$; see pages 164–165 in Miles (1971) or pages 394–395 in Grünbaum, 2003 (for $n \leq 5$, HASS are unique in dimensions $d = 3$ as long as hyperplanes are in general position, so that $n = 6$ is the simplest case where different HASS occur, hence where distribution-freeness may fail in the absolutely continuous case). The top panels of Figure 3 offer a graph representation of two of these four HASS. Graph representations are obtained as follows: each vertex stands for a region A_j and is labelled with its halfspace sign $s_{j,S_1,\ldots,S_n} \in \{-1, 1\}^n$ (according to Proposition 4.1, each graph contains $n_{n,d} = n_{6,3} = 32$ vertices), whereas an edge between two vertices/regions is drawn for neighbouring regions only, i.e., for pairs of regions that are separated by a two-dimensional cone (this happens if and only if the corresponding halfspace signs differ in exactly one entry). The HASS in

Figure 3(a_ℓ), HASS_ℓ say, involves six regions A_j having exactly five neighbouring regions (these six regions are marked in red), whereas the HASS in Figure 3(a_r), HASS_r say, shows no region with exactly five neighbouring regions, but offers two regions with six neighbouring regions (which is the maximum possible).

As we will show, both these HASS yield different conditional null distributions of T_n. To this end, it will be useful to also consider the bottom panels of Figure 3, which, for both HASS_ℓ and HASS_r, show the 32 halfspace signs that are *missing* in the respective HASS; each HASS was involving $n_{n,d} = 32$ halfspace signs out of the full collection of $2^n = 64$ possible halfspace signs, so that also 32 halfspace signs were missing in each case (in the present case $(n, d) = (6, 3)$, the top and bottom panels in Figures 3 thus show a similar structure, but, for n large, the proportion of all possible signs showing in a HASS—namely, $n_{n,d}/2^n$—would of course be much smaller, so that the graph representations associated with the top panels of Figure 3 would then be much sparser than those associated with its bottom panels).

Let us start with HASS_ℓ. Consider first $P[T_n = 0|\text{HASS} = \text{HASS}_\ell]$. In view of (10), $T_n = 0$ is zero if and only if at least one of the halfspace signs

$$s_{j,X_1,\ldots,X_n} = \begin{pmatrix} O_1(s_{j,S_1,\ldots,S_n})_1 \\ \vdots \\ O_n(s_{j,S_1,\ldots,S_n})_n \end{pmatrix}, \quad j = 1, \ldots, 32, \tag{16}$$

is equal to $-\underline{1} := (-1, \ldots, -1) \in \mathbb{R}^6$. Conditional on $\text{HASS} = \text{HASS}_\ell$, randomness is only associated with the uniform distribution of the combination vector $(O_1, \ldots, O_n)'$ over its 64 possible values in $\{-1, 1\}^6$. From Figure 3(a_ℓ), it is then clear that 32 out of these 64 equally likely values of $(O_1, \ldots, O_n)'$ will make one of the resulting halfspace signs $s_{j,X_1,\ldots,X_n}$ equal to $-\underline{1}$, hence will provide $T_n = 0$. Consequently,

$$P[T_n = 0|\text{HASS} = \text{HASS}_\ell] = \frac{32}{64} = \frac{1}{2}.$$

Let us then turn to $P[T_n = 1/6|\text{HASS} = \text{HASS}_\ell]$. From (10) again, $T_n = 1/6$ if and only if none of the halfspace signs in (16) is equal to $-\underline{1}$ but at least one contains exactly one entry equal to 1. We have just seen that among the 64 possible combination vectors, 32 would make at least one of the halfspace signs in (16) equal to $-\underline{1}$, so that the remaining 32 possible ones will provide no halfspace signs equal to $-\underline{1}$. Among the latter 32 possible combination

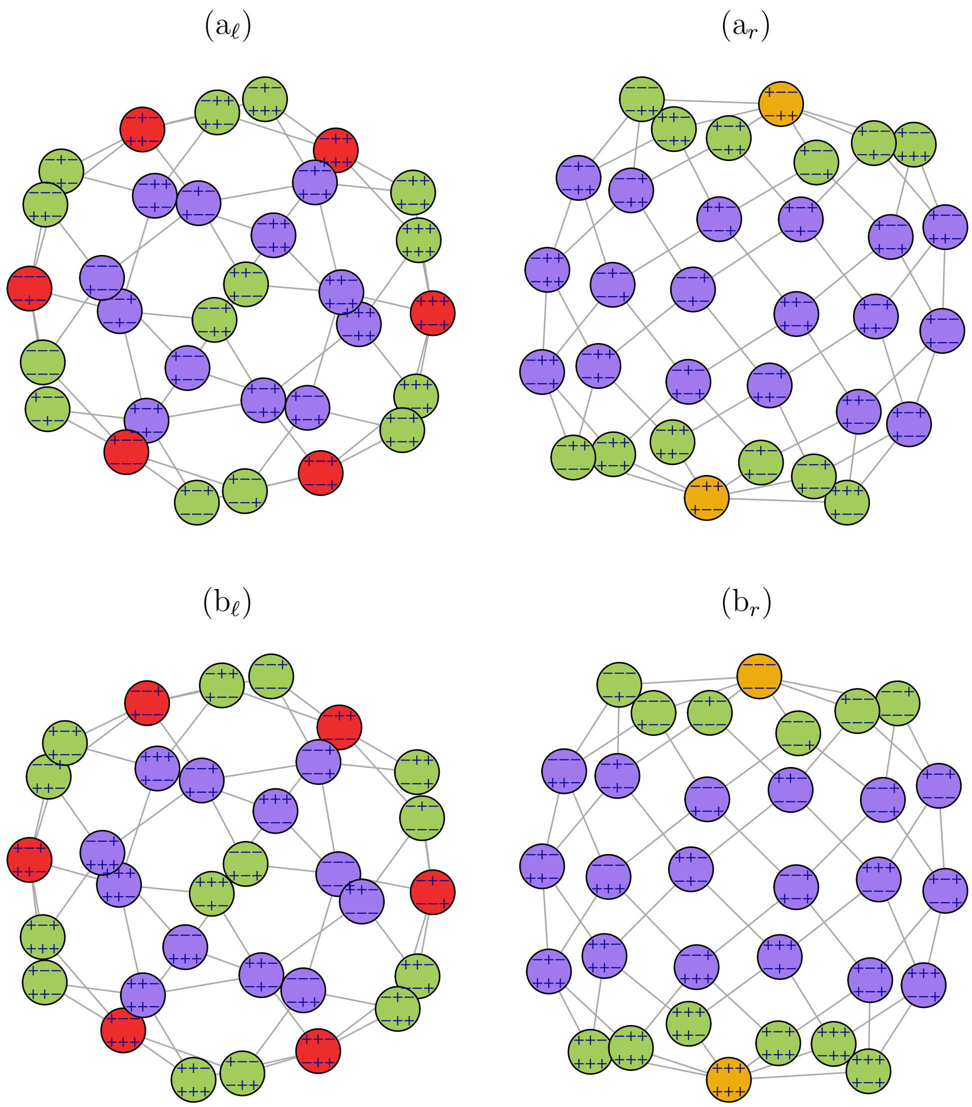

Figure 3: (Top:) Two different HASS for $d = 3$ and $n = 6$ plotted as graphs where each vertex represents a region A_j and is labelled with its halfspace sign $s_{j,S_1,\ldots,S_n} \in \{-1, 1\}^n$. An edge between two regions A_j and $A_{j'}$ is drawn for neighbouring regions, i.e., for pairs of regions that are separated by a two-dimensional cone (this happens if and only if the corresponding halfspace signs differ in exactly one entry). Vertices are coloured according to their number of neighbours. (Bottom:) The corresponding *missing* halfspace signs in the respective HASS. An edge is drawn between two vertices if and only if the corresponding halfspace signs differ by exactly one entry. Vertices are still coloured according to their number of neighbours.

vectors, how many will provide at least one halfspace sign in (16) with exactly one entry equal to 1? Figure 3(b_ℓ) allows us to answer this question. The 32 considered combination vectors are such that $-\underline{1}$ is missing. If we fix one of these combination vectors, then $-\underline{1}$ will be part of the halfspace signs

$$s_{j,X_1,\ldots,X_n} = \begin{pmatrix} O_1(t_{j,S_1,\ldots,S_n})_1 \\ \vdots \\ O_n(t_{j,S_1,\ldots,S_n})_n \end{pmatrix}, \quad j = 1,\ldots,32, \tag{17}$$

associated with the halfspace signs $t_{j,S_1,\ldots,S_n}$ from Figure 3(b_ℓ) (we thus use the notation $t_{j,S_1,\ldots,S_n}$ for halfspace signs that are missing in the considered HASS). Now, none of the vertices of Figure 3(b_ℓ) has six neighbours, so that at least one of the six neighbours of $-\underline{1}$ must not be missing among the halfspace signs in (16). Therefore, each of the 32 combination vectors that make no halfspace signs in (16) equal to $-\underline{1}$ provides at least one halfspace sign in (16) containing exactly one entry equal to 1. This entails that

$$P[T_n = \tfrac{1}{6} | \mathrm{HASS} = \mathrm{HASS}_\ell] = \frac{32}{64} = \frac{1}{2},$$

and we conclude that

$$P[T_n = x | \mathrm{HASS} = \mathrm{HASS}_\ell] = \begin{cases} \frac{1}{2} & \text{if } x \in \{0, \frac{1}{6}\} \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

The situation is actually different for HASS_r. In the present paragraph, when we refer to the halfspace signs in (16) (resp., to those in (17)), this will of course be relative to the $s_{j,S_1,\ldots,S_n}$'s that are in Figure 3(a_r) (resp., to the $s_{j,S_1,\ldots,S_n}$'s that are missing in Figure 3(a_r), equivalently, to the $t_{j,S_1,\ldots,S_n}$'s that are present in Figure 3(b_r)). Now, the exact same argument as for HASS_ℓ shows that

$$P[T_n = 0 | \mathrm{HASS} = \mathrm{HASS}_r] = \frac{32}{64} = \frac{1}{2};$$

indeed, it is still so that among the 64 possible combination vectors, 32 will make at least one of the halfspace signs in (16) equal to $-\underline{1}$ and the remaining 32 possible ones will provide none of these halfspace signs equal to $-\underline{1}$. As above, $T_n = 1/6$ if and only if none of the halfspace signs in (16) is equal to $-\underline{1}$ but at least one contains exactly one entry equal to 1. Here, however, among the 32 combination vectors that are such that $-\underline{1}$ is missing among the halfspace signs in (16), equivalently, that are such that $-\underline{1}$ is among the halfspace signs in (17), only 30 out of the 32 have strictly less than six neighbours, which ensures that the combinations make at least one halfspace sign in (16) contain exactly one entry equal to 1. Thus,

$$P[T_n = \tfrac{1}{6} | \text{HASS} = \text{HASS}_r] = \frac{30}{64} = \frac{15}{32}.$$

Both the remaining 2 combination vectors not only are such that the halfspace signs in (16) do not contain $-\underline{1}$ but they are also such that no halfspace sign in (16) contains exactly one entry equal to 1; this corresponds to both cases where one of the halfspace signs in (17) is equal to $-\underline{1}$ and is a yellow vertex in Figure 3(b_r). Because, in this panel, none of the green vertices that are at distance one from a yellow vertex has six neighbours, there must be among the halfspace signs in (16) a vector that contains exactly two entries equal to 1, which then provides $T_n = 2/6 = 1/3$. Thus,

$$P[T_n = \tfrac{1}{3} | \text{HASS} = \text{HASS}_r] = \frac{2}{64} = \frac{1}{32},$$

and we finally conclude that

$$P[T_n = x | \text{HASS} = \text{HASS}_r] = \begin{cases} \frac{1}{2} & \text{if } x = 0 \\ \frac{15}{32} & \text{if } x = \frac{1}{6} \\ \frac{1}{32} & \text{if } x = \frac{1}{3} \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

We have thus shown that the null distribution of T_n conditional on HASS = HASS_ℓ is different from the null distribution of T_n conditional on HASS = HASS_r. Since it can be checked similarly that the remaining two HASS for $d = 3$ and $n = 6$ provide the conditional distribution in (18), the small difference between the conditional distributions in (18)–(19) now explains why departures from distribution-freeness in Table 1 was so modest. In the case $d = 3$ and $n = 6$, we have thus provided theoretical support for the fact that deviations from distribution-freeness will remain very small.

Table 2 reports the frequencies according to which the conditional distributions in (18) and (19) showed up, under both probability measures P_1 and P_2, when conducting the Monte Carlo exercise that led to Table 1. It is seen that the conditional distribution in (18) was more frequent under P_1 than under P_2, which explains the lack of distribution-freeness in Table 1.

	P_1	P_2
$P[T_n = x \vert \text{HASS} = \text{HASS}_\ell]$	0.8102 [0.8099,0.8104]	0.8006 [0.8004,0.8009]
$P[T_n = x \vert \text{HASS} = \text{HASS}_r]$	0.1898 [0.1896,0.1901]	0.1994 [0.1991,0.1996]

Table 2: Empirical probabilities (together with the corresponding 95% confidence intervals) that the conditional empirical distributions in (18) and (19) showed up, under both probability measures P_1 and P_2, when conducting the Monte Carlo exercise that led to Table 1; recall that, for each probability measure, estimation is based on $M = 10^7$ mutually independent samples.

5. A Monte Carlo study

In this section, we conduct a broader Monte Carlo study, involving 10 different probability measures and various sample sizes, to show that departures from distribution-freeness remain small in general, using the R package `ddalpha` (Pokotylo et al., 2019) to evaluate T_n. Beyond the probability measures P_1 and P_2 already considered above, we considered the following eight (angularly symmetric) ones:

- P_3 : the uniform distribution over the cube with vertices at $(\pm 1, \pm 1, \pm 1)'$;
- P_4 : the distribution of $(Z_1, Z_2, Z_3)'$, where the Z_i's are i.i.d. Cauchy;
- P_5 : the distribution of $S(Z_1, Z_2, Z_3)'$, where S, Z_1, Z_2, Z_3 are independent, S is Rademacher and the Z_i's are exponential with mean one;
- P_6 : the distribution of

$$W \begin{pmatrix} \sin(U)\sqrt{1-V^2} \\ \cos(U)\sqrt{1-V^2} \\ V \end{pmatrix},$$

  where U, V, W are independent, $U \sim \mathcal{N}(0, \frac{1}{4})$, V is uniformly distributed over $[-1, 1]$, and $W \sim \mathcal{N}(0, 1)$;
- P_7 : the distribution of SX, where S is Rademacher and X is a mixture distribution with equal mixture weights and mixture components

$$X_1 \sim N\left(\begin{pmatrix} 2 \\ -2 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \quad X_2 \sim N\left(\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .7 & .7 \\ .7 & 1 & .7 \\ .7 & .7 & 1 \end{pmatrix} \right),$$

and

$$X_3 \sim N\left(\begin{pmatrix}3\\0\\0\end{pmatrix}, \begin{pmatrix}1 & 0 & 0\\0 & 5 & 0\\0 & 0 & 1\end{pmatrix}\right);$$

- P_8 : the distribution of SX, where S is Rademacher and X is a mixture distribution with equal mixture weights and (four) mixture components being trivariate normal distributions with mean

$$\begin{pmatrix}-4\\-8\\12\end{pmatrix}, \begin{pmatrix}0\\-12\\0\end{pmatrix}, \begin{pmatrix}0\\12\\-12\end{pmatrix}, \begin{pmatrix}-2\\2\\12\end{pmatrix}$$

and common covariance matrix

$$\begin{pmatrix}10 & 9 & 9\\9 & 10 & 9\\9 & 9 & 10\end{pmatrix};$$

- P_9 : the distribution of SX, where S is Rademacher and X is a mixture distribution with equal mixture weights and (thirteen) mixture components being trivariate normal distributions with mean

$$\begin{pmatrix}2\\0\\0\end{pmatrix}, \begin{pmatrix}0\\2\\0\end{pmatrix}, \begin{pmatrix}0\\0\\2\end{pmatrix}, \begin{pmatrix}2\\2\\0\end{pmatrix}, \begin{pmatrix}2\\0\\2\end{pmatrix}, \begin{pmatrix}0\\2\\2\end{pmatrix},$$

$$\begin{pmatrix}2\\-2\\0\end{pmatrix}, \begin{pmatrix}2\\0\\-2\end{pmatrix}, \begin{pmatrix}0\\2\\-2\end{pmatrix}, \begin{pmatrix}2\\2\\2\end{pmatrix}, \begin{pmatrix}2\\-2\\2\end{pmatrix}, \begin{pmatrix}2\\2\\-2\end{pmatrix} \begin{pmatrix}2\\-2\\-2\end{pmatrix}$$

and common covariance matrix diag(.1, .1, .1);

- P_{10} : the distribution of SX, where S is Rademacher and X is a mixture distribution with equal mixture weights and mixture components

$$X_1 \sim N\left(\begin{pmatrix}0\\3\\0\end{pmatrix}, \begin{pmatrix}10 & 0 & 0\\0 & 1 & 0\\0 & 0 & .2\end{pmatrix}\right) \quad \text{and} \quad X_2 \sim N\left(\begin{pmatrix}0\\3\\0\end{pmatrix}, \begin{pmatrix}.2 & 0 & 0\\0 & 1 & 0\\0 & 0 & 10\end{pmatrix}\right).$$

Visualizing these distributions is difficult due to their three-dimensional nature. Since the test statistic T_n is invariant when observations are projected radially onto the unit sphere, we rather show density estimates of the

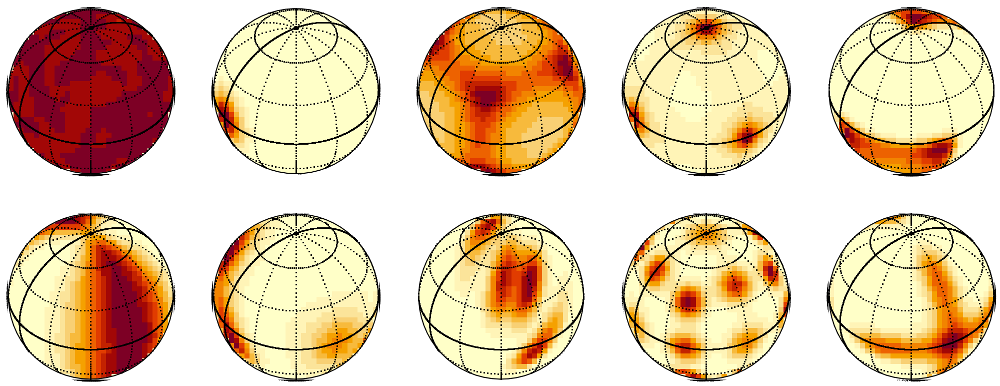

Figure 4: Density estimates of the radial projections of $P_1, \ldots, P_{10}$ onto the unit sphere. The first row shows $P_1, \ldots P_5$ (from left to right), whereas the second row shows $P_6, \ldots, P_{10}$ (still from left to right).

projected distributions. Figure 4 plots such density estimates, based in each case on a random sample of size $n = 10^5$; this was done by using the R function *sm.spherical* from the R-package `sm` (Bowman and Azzalini, 2021) with smoothing parameter $\kappa = 100$.

To assess how small departures from distribution-freeness are, we choose P_1 as a reference distribution, which is natural since its projection onto the unit sphere is uniformly distributed. For each combination of a sample size $n \in \{6, 12, 24, 48, 96, 192, 384, 768\}$ and of a probability measure P_i among $P_2, \ldots, P_{10}$, we then evaluated the estimated Kolmogorov–Smirnov distance

$$\hat{d}_n(P_i, P_1) := \sup_{x \in [0, \frac{1}{2}]} |\hat{F}_n^{(i)}(x) - \hat{F}_n^{(1)}(x)|, \tag{20}$$

where $\hat{F}_n^{(i)}$ is the empirical distribution function of T_n obtained from $m = 10^7$ mutually independent random samples of size n from P_i. Table 3, which provides these distances multiplied by 10^3 for better readability, reveals that distances remain very small. Note that the critical value for a two-sided Kolmogorov–Smirnov two-sample test using a significance level of 5% and a sample size of 10^7 is 0.6×10^{-3}. While distances indeed remain small in Table 3, most of them are thus significant, which supports again the lack of distribution-freeness. Notably $\hat{d}_n(P_6, P_1)$ remains small and non-significant for all sample sizes. It turns out that the projected density of P_6 in Figure 4 can be transformed to the uniform distribution on the sphere by applying a monotone transformation on the longitude (and no transformation at all on the latitude) of the spherical coordinates. Since T_n is invariant under such a transformation, the distributions of T_n under P_1 and P_6 do coincide, which explains that the corresponding distances in Table 3 are small.

n	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
6	0.28	0.13	0.30	0.81	0.24	0.20	0.44	0.25	0.91
12	0.53	0.16	0.40	1.30	0.33	0.38	0.75	0.43	1.30
24	1.08	0.50	0.46	2.07	0.35	0.49	1.51	0.85	1.46
48	1.10	0.51	0.92	1.99	0.29	0.75	1.88	0.87	1.66
96	1.26	0.41	0.75	2.24	0.54	0.79	1.88	1.32	1.94
192	1.33	0.47	0.97	2.31	0.55	0.67	1.98	1.30	2.06
384	1.23	0.96	1.24	2.29	0.50	1.20	1.78	1.42	2.09
768	1.10	0.31	0.94	2.10	0.24	0.74	1.83	1.27	1.91

Table 3: For each combination of a sample size $n \in \{6, 12, 24, 48, 96, 192, 384, 768\}$ and of a probability measure P_i among $P_2, \ldots, P_{10}$, this provides the estimated Kolmogorov–Smirnov distance $\hat{d}_n(P_i, P_1)$ in (20), multiplied by 1000 for better readability.

6. Final comments

Since the test statistic T_n is not distribution-free under the null hypothesis in dimension $d \geq 3$, it is not possible to estimate critical values through simulations. However, one may still apply the test conditionally. More precisely, when testing for angular symmetry about the origin of $\mathbb{R}^d$ based on the sample $X_1, \ldots, X_n$, this consists in evaluating the 2^n test statistic values

$$T_n^o = T_n(o_1 S_1, \ldots, o_n S_n), \quad o = (o_1, \ldots, o_n) \in \{-1, 1\}^n,$$

and in rejecting the null hypothesis at level α if $T_n = T_n(X_1, \ldots, X_n)$ is below the sample α-quantile in $\{T_n^o : o = (o_1, \ldots, o_n) \in \{-1, 1\}^n\}$ (as usual, for most values of α, achieving exactly significance level α will require a suitable randomization). This test, which is thus conditional on the sign vector $(S_1, \ldots, S_n)$, is a perfectly valid procedure in any dimension $d \geq 2$, including dimensions for which T_n is not distribution-free under the null.

The conditional test just described requires evaluating 2^n values of the test statistic, hence can be applied for small to moderate sample sizes n

only. For large sample sizes, one would typically want to obtain a critical value from the asymptotic null distribution of T_n. In dimension $d = 2$, Daniels (1954) actually also provided a large-n approximation of the exact distribution in (3), namely

$$P\big[T_n \leq \tfrac{k}{n}\big] \sim \frac{4(n-2k)}{\sqrt{2\pi n}} \sum_{j=0}^{\infty} e^{-(2j+1)^2(n-2k)^2/(2n)} =: F_{\text{approx},n}(k) \tag{21}$$

(no explanation on how to obtain this approximation is given in Daniels (1954), but a proof, relying on an alternative expression for (3), can be found in Joffe and Klotz (1962)). The maximal absolute error

$$\max_{k=0,\ldots,\lfloor (n-1)/2 \rfloor} \Big| P\big[T_n \leq \tfrac{k}{n}\big] - F_{\text{approx},n}(k) \Big| \tag{22}$$

is shown in Figure 5 as a function of the sample size n and indicates that the approximation works reliably even for relatively moderate values of n. Obviously, exact distribution-freeness for $d = 2$ entails asymptotic distribution-freeness (under the null hypothesis in both cases), and the approximation in (22) indeed does not involve the underlying angular symmetric distribution. Given the finite-sample analysis conducted in the present work, a natural question is whether or not the test statistic T_n is *asymptotically* distribution-free under the null hypothesis. In view of the very small departures from distribution-freeness we encountered in this work, it is not easy to explore through Monte Carlo exercises whether or not asymptotic distribution-freeness holds, yet results from Table 3 seem to indicate that asymptotic distribution-freeness does not hold in dimension 3.

The testing problem considered in the paper is the problem of testing the null hypothesis of angular symmetry about a given location θ_0 of $\mathbb{R}^d$, so that location shifts, that will induce angular symmetry about $\theta(\neq \theta_0)$, are part of the alternative hypothesis. While we focused in the present work on Type I risks, hence did not touch such issues, we stress here that tests of multivariate symmetry may be biased as particular combinations of location shifts and structural asymmetry may lead to cancel each other, hence lead to alternatives that cannot be detected; we refer to Babić et al. (2022) for a precise description of this in the context of testing for elliptical symmetry.

Appendix A.

It only remains to prove Lemma 4.1.

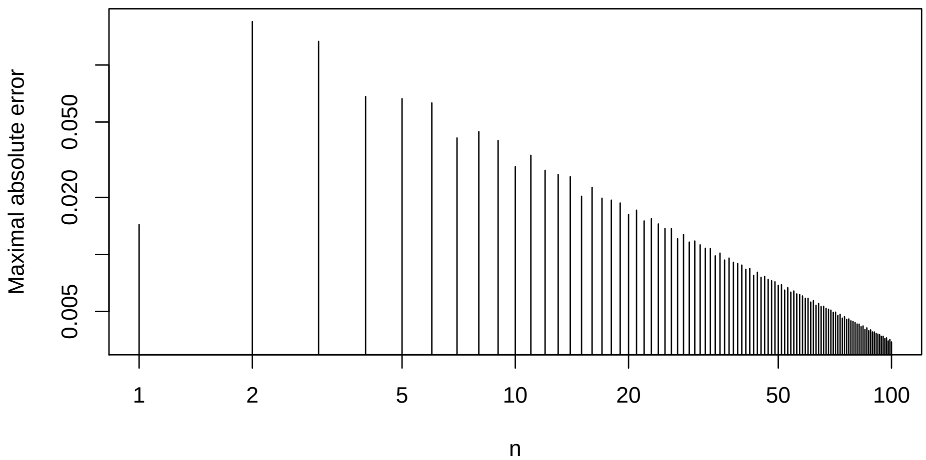

Figure 5: The maximal absolute error in (22) with respect to sample size n (in log scales).

Proof of Lemma 4.1. ($\Rightarrow$) Letting $H_d := \{x \in \mathbb{R}^d : e_d' x > 0\}$, angular symmetry of X about the origin of $\mathbb{R}^d$ (and absolute continuity) entails that

$$\frac{1}{2} = P[X \in H_d] = P[OSM \in H_d] = P[OS \in H_d \cap \mathcal{S}^{d-1}] = P[O = 1], \quad \text{(A.1)}$$

where we used the fact that $S \in H_d \cap \mathcal{S}^{d-1}$ with probability one. Since the absolute continuity assumption ensures that $P[O = 0] = 0$, this implies that $P[O = -1] = 1 - P[O = 1] = 1/2$. It remains to show that O and S are mutually independent. To this end, let B be an arbitrary Borel set on $\mathcal{S}^{d-1}$ and $A_B := \{x \in \mathbb{R}^d : x/\|x\| \in B\}$ be the corresponding cone with apex at the origin of $\mathbb{R}^d$. Then,

$$P[X \in A_B \cap H_d] = P[OSM \in A_B \cap H_d] = P[SM \in A_B \cap H_d, O = 1]$$
$$= P[S \in B \cap H_d, O = 1] = P[S \in B, O = 1]$$

and, similarly,

$$P[X \in -(A_B \cap H_d)] = P[OSM \in -(A_B \cap H_d)]$$
$$= P[SM \in A_B \cap H_d, O = -1] = P[S \in B \cap H_d, O = -1] = P[S \in B, O = -1].$$

From angular symmetry, we thus have $P[S \in B, O = 1] = P[S \in B, O = -1]$, so that

$$P[S \in B, O = 1] = P[S \in B] - P[S \in B | O = -1] = P[S \in B] - P[S \in B | O = 1],$$

which yields

$$P[S \in B, O = 1] = \frac{1}{2} P[S \in B] = P[S \in B] P[O = 1].$$

Since the same argument shows that $P[S \in B, O = -1] = P[S \in B]P[O = 1]$ and since B is arbitrary, we conclude that O and S are mutually independent.

($\Leftarrow$) Let A be an arbitrary cone whose apex is at the origin of $\mathbb{R}^d$. Then,

$$P[X \in A] = P[OSM \in A] = P[OS \in A \cap \mathcal{S}^{d-1}],$$

where $\mathcal{S}^{d-1}$ denotes the unit sphere of $\mathbb{R}^d$. The law of total probability then provides

$$\begin{aligned} P[X \in A] &= P[OS \in A \cap \mathcal{S}^{d-1} | O = 1] P[O = 1] \\ &\quad + P[OS \in A \cap \mathcal{S}^{d-1} | O = -1] P[O = -1] \\ &= \big(P[S \in A \cap \mathcal{S}^{d-1}] + P[-S \in A \cap \mathcal{S}^{d-1}]\big)/2 \\ &= \big(P[S \in A \cap \mathcal{S}^{d-1}] + P[S \in -A \cap \mathcal{S}^{d-1}]\big)/2, \end{aligned}$$

where we used the fact that O satisfies $P[O = 1] = P[O = -1] = \frac{1}{2}$ and is independent of S. Since the last expression is unchanged when substituting $-A$ for A, we have that $P[X \in -A] = P[X \in A]$, which establishes that X is angularly symmetric about the origin of $\mathbb{R}^d$. $\square$

References

Babić, S., Gelbgras, L., Hallin, M., Ley, C., 2022. Optimal tests for elliptical symmetry. Bernoulli 27, 2189–2216.

Bowman, A.W., Azzalini, A., 2021. R package `sm`: nonparametric smoothing methods (version 2.2-5.7). University of Glasgow, UK and Università di Padova, Italia. URL: http://www.stats.gla.ac.uk/~adrian/sm/.

Chen, Z., Tyler, D., 2002. The influence function and maximum bias of Tukey's median. Ann. Statist. 30, 1737–1759.

Daniels, H., 1954. A distribution-free test for regression parameters. Ann. Math. Statist. 25, 499–513.

Genest, M., Masse, J.C., src/depth.f contains eigen, J.F.P., tql2, tred2 written by the EISPLACK authors, dgedi, dgefa from LINPACK written by Cleve Moler, daxpy, dscal, dswap, idamax from LINPACK written by Jack Dongarra, from NAPACK, V., written by J. C. Gower, A.., written by F. K. Bedall, A.., Zimmermann, H., written by P.J. Rousseeuw, A.., Ruts., I., 2019. depth: Nonparametric Depth Functions for Multivariate Analysis. R package version 2.1-1.1.

Grünbaum, B., 2003. Convex Polytopes. 2nd ed., Springer, New York.

Hallin, M., del Barrio, E., Cuesta-Albertos, J., Matrán, C., 2021. Center-outward distribution and quantile functions, ranks, and signs in $\mathbb{R}^d$: a measure transportation approach. Ann. Statist. 49, 1139–1165.

Hallin, M., Paindaveine, D., 2002. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. Ann. Statist. 30, 1103–1133.

Hettmansperger, T., Möttönen, J., Oja, H., 1997. Affine-invariant multivariate one-sample signed-rank tests. J. Amer. Statist. Assoc. 92, 1591–1600.

Hettmansperger, T., Nyblom, J., Oja, H., 1994. Affine invariant multivariate one-sample sign tests. J. R. Statist. Soc. Ser. B 56, 221–234.

Joffe, A., Klotz, J., 1962. Null distribution and Bahadur efficiency of the Hodges bivariate sign test. The Annals of Mathematical Statistics , 803–807.

Larocque, D., 2003. An affine-invariant multivariate sign test for cluster correlated data. Canad. J. Statist. 31, 437–455.

Larocque, D., Nevalainen, J., Oja, H., 2007. A weighted multivariate sign test for cluster-correlated data. Biometrika 94, 267–283.

Liu, R.Y., 1988. On a notion of simplicial depth. Proc. Natl. Acad. Sci. USA 85, 1732–1734.

Liu, R.Y., 1990. On a notion of data depth based on random simplices. Ann. Statist. 18, 405–414.

Liu, R.Y., Parelius, J.M., Singh, K., 1999. Multivariate analysis by data depth: descriptive statistics, graphics and inference. Ann. Statist. 27, 783–840.

Miles, R., 1971. Random points, sets and tessellations on the surface of a sphere. Sankhyā: The Indian Journal of Statistics, Series A , 145–174.

Mozharovskyi, P., 2016. Tukey depth: linear programming and applications. Technical Report. arXiv:1603.00069.

Oja, H., 2010. Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks. Springer-Verlag, New York.

Oja, H., Randles, R.H., 2004. Multivariate nonparametric tests. Statist. Sci. 19, 598–605.

Ollila, E., Oja, H., Hettmansperger, T., 2002. Estimates of regression coefficients based on the sign covariance matrix. J. R. Stat. Soc. Ser. B 64, 447–466.

Perlman, M., Chaudhuri, S., 2012. Reversing the Stein effect. Statist. Sci. 27, 135–143.

Peters, D., Randles, R.H., 1990. A multivariate signed-rank test for the one-sample location problem. J. Amer. Statist. Assoc. , 552–557.

Pokotylo, O., Mozharovskyi, P., Dyckerhoff, R., 2019. Depth and depth-based classification with R package ddalpha. Journal of Statistical Software 91, 1–46. doi:10.18637/jss.v091.i05.

Rousseeuw, P.J., Ruts, I., 1999. The depth function of a population distribution. Metrika 49, 213–244.

Rousseeuw, P.J., Struyf, A., 1998. Computing location depth and regression depth in higher dimensions. Stat. Comput. 8, 193–203.

Rousseeuw, P.J., Struyf, A., 2002. A depth test for symmetry, in: Goodness-of-fit tests and model validity (Paris, 2000). Birkhäuser Boston, Boston, MA. Stat. Ind. Technol., pp. 401–412.

Rousseeuw, P.J., Struyf, A., 2004. Characterizing angular symmetry and regression symmetry. J. Statist. Plann. Inference 122, 161–173.

Serfling, R., 2006. Multivariate symmetry and asymmetry, in: Kotz, S., Balakrishnan, N., Read, C., Vidakovic, B. (Eds.), Encyclopedia of Statistical Sciences. second ed.. Wiley. volume 8, pp. 5338–5345.

Tukey, J.W., 1975. Mathematics and the picturing of data, in: Proceedings of the International Congress of Mathematicians, Vancouver, 1975, pp. 523–531.

Van Bever, G., 2016. Simplicial bivariate tests for randomness. Statist. Probab. Lett. 112, 20–25.

Wendel, J.G., 1962. A problem in geometric probability. Math. Scand. 11, 109–111.

Winder, R.O., 1966. Partitions of n-space by hyperplanes. SIAM J. Appl. Math. 14, 811–818.

Zuo, Y., Serfling, R., 2000a. General notions of statistical depth function. Ann. Statist. 28, 461–482.

Zuo, Y., Serfling, R., 2000b. On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. Journal of Statistical Planning and Inference 84, 55–79.