

# A GENERALIZATION OF THE MOORE–RAYLEIGH TEST FOR TESTING SYMMETRY OF VECTOR DATA AND TWO-SAMPLE PROBLEMS

## TECHNICAL REPORT MI-2009-05

MICHAEL MUSKULUS, ALIZE SCHEENSTRA, AND SJOERD VERDUYN-LUNEL

**ABSTRACT.** The Rayleigh test is a popular one-sample test of randomness for directional data on the unit circle. Based on the Rayleigh test, Moore developed a nonparametric test for two-dimensional vector data that takes vector lengths into account as well. In this paper we generalize this test to arbitrary dimensions. In the important case of three-dimensional vector data the asymptotic distribution can be given in closed form as a finite combinatorial sum. This reduces the computational effort considerably. In particular, when analyzing deformation fields arising in nonlinear brain registration, the generalized Moore-Rayleigh test offers an efficient alternative to conventional permutation testing for the initial screening of voxels.

Simulation results for a few multivariate distributions are given. Compared with the permutation version of Hotelling's  $T^2$  test its increased power should allow for improved localization of brain regions with significant deformations.

### CONTENTS

1. Introduction	1
2. The Moore-Rayleigh test	2
2.1. The one-dimensional case	4
2.2. The three-dimensional case	5
2.3. Power estimates	7
3. The two-sample test	10
3.1. Testing for symmetry	10
3.2. Further issues	12
4. Simulation results	13
5. Discussion	14
Acknowledgments	15
Appendix A. The Fisher distribution	15
References	15

### 1. INTRODUCTION

Consider the following illustrating example. In the voxel-based analysis of brain deformations, individual brain volumes are mapped to a reference brain image by a nonlinear transformation (Kovacevic et al., 2004). This process of image registration results in a three-dimensional vector field of displacement vectors. The significance of local deformations between groups of subjects, usually a treatment and a control group, can be tested by either considering the Jacobian of the deformation field, or testing the displacement vectors directly (Chung et al., 2001). In the latter case, if one assumes that variations between

---

*Date:* April 7, 2009 (Version 1.2).

*Key words and phrases.* Nonparametric; Rayleigh test; Vector data; Permutation testing; Multivariate symmetry testing; Alzheimer's disease; Image analysis.

subjects are given by a Gaussian random field, Hotelling’s  $T^2$  statistic can be used to test for significant differences between groups (Cao & Worsley, 1999). Its value is the squared sample Mahalanobis distance, estimated from the pooled covariance matrix, and the test assumes normality of the population of deformation vectors and equal covariances for the two groups. If these assumptions are not met, the  $T^2$  test is known to fail gracefully, i.e., it will still be approximately conservative and the loss in power for the alternative will not be too dramatic for moderate violations of the assumptions. However, it is preferable to analyze deformation fields nonparametrically.

Permutation tests, with their minimal assumptions, are the usual method of choice for this two-sample problem (Chen et al., 2005; Nichols & Holmes, 2007). However, they also rely on a test statistic that is evaluated for each labelling (“permutation”), and the null hypothesis is that this statistic is distributed symmetrically around zero. The usual choice for the statistic is again Hotelling’s  $T^2$ , so permutation tests are not nonparametric, but rather result in *adjusted* significance probabilities (Davison & Hinkley, 1997, Chap. 4.5). For example, as shown in Lehmann & Romano (2005, Chap. 5.9), the adjusted one-dimensional version of the  $T^2$  test, i.e., the permutation version of the classic  $t$ -test, is the uniformly most powerful test for the Gaussian alternatives with fixed variance, but fails to be uniformly most powerful against other alternatives.

A more serious practical problem is that, even for small sample sizes, the number of permutations to consider for an exact test is prohibitively large. Especially so, if the number of voxels, i.e., the number of tests, is on the order of hundreds of thousands, as common in neuroimaging applications. Therefore, in current analyses one often limits the data to only 10000 or less random labellings per voxel, at the expense of increasing the simulation error. Moreover, correcting for multiple comparisons imposes severe lower bounds on the numbers of labellings needed per voxel for testing at realistic significance levels, i.e., on the sample size and simulation time. Particularly for small sample sizes that occur in prospective studies, permutation tests cannot resolve low enough significance probabilities to allow for strong control of the family-wise error. Even the modern, liberal approach of limiting the False Discovery Rate (Benjamini & Hochberg, 1995; Schwartzman et al., 2009) does often not lead to useful results in these datasets. This implies that although permutation tests are elegant and theoretically well understood, they can not be used on a routine basis (e.g. in a clinical setting) to assess and quantify brain changes.

For these reasons, in the analysis of MR images classical hypothesis testing is still unmatched in its efficiency and speed. In this paper we describe a new nonparametric statistical test that allows to efficiently perform a large number of such tests on vector data. The two-sample version of the test is not provably conservative, but its advantage is that it can be used for the initial screening of voxels. It is sensitive enough to work even under the conservative Bonferroni correction. Voxels where the null hypothesis is rejected can then be analyzed further by this test under the permutation distribution of the data; alternatively a different test statistic can be employed.

This problem of testing one or more groups of vectors for distributional differences does not only arise in neuroimaging, but also in a number of other disciplines and diverse contexts, e.g. in geostatistics, human movement sciences, astronomy and biology. In the two-dimensional case, a natural nonparametric test for such problems has been given by Moore (1980), which we describe next. After generalizing this test to arbitrary dimensions, in Section 2.2 we focus on the three-dimensional case, being the most important one for applications.

## 2. THE MOORE-RAYLEIGH TEST

Let  $X = (X_1, \dots, X_N)$  be a finite sample of real  $k$ -vector-valued random variables  $X_i = (X_{i,1}, \dots, X_{i,k})$ . If we assume that the  $X_i$  are independently drawn from a common absolutely continuous probability distribution with density  $f : \mathbb{R}^k \rightarrow [0, \infty)$ , then the null

hypothesis is:

$H_0$  : The probability density  $f$  is *spherically symmetric*.

Consequently, this implies that the density  $f$  is spherically decomposable. It factors into the product of a radial density  $p_r : [0, \infty) \rightarrow [0, \infty)$  and the uniform distribution on each hypersphere  $rS^{k-1} = \{x \in \mathbb{R}^k \mid \|x\| = r\}$ , such that  $f(x) = p_r(\|x\|)/\text{vol}(\|x\|S^{k-1})$ . We can then write  $X_i = R_i U_i$ , where  $R_i \sim p_r$  and  $U_i$  is distributed uniformly on the  $k$ -dimensional unit sphere  $S^{k-1}$ . The latter distribution can be realized as the projection of a  $k$ -dimensional diagonal Gaussian distribution with equal variance in each dimension. The sum  $\sum_{i=1}^N X_i$ , where the  $X_i$  are independently distributed according to a common, spherically symmetric distribution, is easy to interpret. It corresponds to a Rayleigh random flight (Dutka, 1985) with  $N$  steps, whose lengths are distributed according to  $p_r$ .

Scaling the vector-valued random variables  $X$  by the ranks of their lengths, the distribution of the resultant vector

$$(2.1) \quad S_N = \sum_{i=1}^N \frac{iX_{(i)}}{\|X_{(i)}\|},$$

where  $X_{(i)}$  denotes the  $i$ -th largest vector in the sample (with ties being arbitrarily resolved), is independent of  $p_r$ ; consequently, a test based on  $S_N$  is nonparametric. The test statistic of interest here is the asymptotically scaled length of the resultant,

$$(2.2) \quad R_N^* = \frac{\|S_N\|}{N^{3/2}}.$$

A large value of  $R_N^*$  for a given sample  $X$  from an unknown distribution (not necessarily absolutely continuous) indicates a deviation from spherical symmetry. This test was introduced by Moore (1980), who treated the two-dimensional case numerically, and has been used in neuroscience (Kajikawa & Hackett, 2005; Tukker et al., 2007; Richardson et al., 2008), human movement science (van Beers et al., 2004) and avian biology (Able & Able, 1997; Mcnaught & Owens, 2002; Burton, 2006; Chernetsov et al., 2006). In contrast to the Rayleigh test of uniformity (Mardia & Jupp, 2000, Chap. 10.4.1), where the  $X_i$  are constrained to lie on (alternatively, are projected onto) the unit sphere, in the Moore-Rayleigh test also the vector length influences the test statistic. This follows the observation of Gastwirth (1965), that differences in scale between two distributions will be mostly evident in their (radial) tails, i.e., when moving away from the mean. The interpretation of  $R_N^*$  is not so easy as in the Rayleigh test, however, where the test statistic is a measure of *spherical variance*.

Consider the projections

$$(2.3) \quad S_{N,j} = \sum_{i=1}^N \frac{iX_{(i),j}}{\|X_{(i)}\|}, \quad (j = 1, \dots, k).$$

A direct calculation shows that under the null hypothesis the variance of  $X_{(i),j}/\|X_{(i)}\|$  is  $1/k$ , and that

$$(2.4) \quad \sigma^2 = \text{var}(S_{N,j}) = N(N+1)(2N+1)/(6k).$$

As  $E(S_{N,j})^3 = 0$  and  $\sigma^2 < \infty$ , the Lyapunov version of the Central Limit Theorem implies that the random variables  $S_{N,j}$  approach Gaussian  $\mathcal{N}(0, \sigma^2)$  distributions for large sample sizes  $N$ . Although the random variables  $\|S_{N,j}\|$  are obviously not independent, by the same argument as in Stephens (1962) the corresponding distribution of  $\|S_N\|^2/\sigma^2$  asymptotically approaches a  $\chi_k^2$  distribution.

Let  $\alpha_N = N^{3/2}$ . The exact null distribution of  $R_N = \alpha_N R_N^*$  in  $k$  dimensions,  $k \geq 2$ , is given by

$$(2.5) \quad \text{pr}(R_N \leq \alpha_N r; k) = r \left[ \Gamma\left(\frac{k}{2}\right) \right]^{N-1} \int_0^\infty \left(\frac{rt}{2}\right)^{\frac{k-2}{2}} J_{\frac{k}{2}}(rt) \prod_{n=1}^N \frac{J_{\frac{k-2}{2}}(nt)}{(nt/2)^{\frac{k-2}{2}}} dt,$$

where  $J_l$  denotes the Bessel function of order  $l$ ; see (Lord, 1954).

**2.1. The one-dimensional case.** In one dimension, the Moore–Rayleigh statistic for the null hypothesis corresponds to a symmetric random walk with linearly growing steps,

$$(2.6) \quad S_N = \sum_{i=1}^N \gamma_i i, \quad \text{where } \gamma_i = \pm 1 \text{ with equal probability.}$$

**Proposition 1.** *The probability mass function  $\text{pr}(S_N = r) \stackrel{\text{def}}{=} p(r, N)/2^N$  is given by the recurrence*

$$(2.7) \quad p(r, N) = p(r - n, N - 1) + p(r + n, N - 1)$$

with initial condition  $p(0, 0) = 1$  and  $p(r, 0) = 0$  for  $r \neq 0$ .

Rewriting Eq. 2.6 as

$$(2.8) \quad \sum_{\{\gamma_i = +1\}} i = \frac{1}{2} \left( S_N + \frac{1}{2} N(N + 1) \right),$$

where the sum runs over all step sizes  $i \in \{1, \dots, N\}$  that have positive sign  $\gamma_i$ , shows that the numbers  $p(r, N)$  have a well-known combinatorial interpretation.

**Proposition 2.** *The numbers  $p(r, N)$  count the number of partitions of  $\frac{1}{2}(r + \frac{1}{2}N(N + 1))$  with distinct parts less or equal to  $N$ .*

As before, denote the length of the resultant by  $R_N = \|S_N\|$ . Its probability function  $\text{pr}(R_N = r)$  is given by

$$(2.9) \quad \text{pr}(R_N = r) = \begin{cases} p(r, N)/2^{N-1} & \text{if } r > 0, \\ p(0, N)/2^N & \text{if } r = 0, \\ 0 & \text{otherwise.} \end{cases}$$

In the sequel, we also need the random signs defined by

$$(2.10) \quad \epsilon_N = \prod_{i=1}^N \gamma_i,$$

conditional on the resultant  $S_N$ : Let  $\epsilon_{r, N}$  denote the average sign of the partitions of  $\frac{1}{2}(r + \frac{1}{2}N(N + 1))$  with distinct terms less or equal to  $N$ , i.e.,

$$(2.11) \quad \epsilon_{r, N} \stackrel{\text{def}}{=} E(\epsilon_N \mid S_N = r).$$

Anticipating the two-sample Moore–Rayleigh test discussed in Section 3, we note the following:

**Remark 1** (Relation to the Wilcoxon signed-rank test). In the Wilcoxon signed-rank test for two paired samples  $X$  and  $Y$  of equal size  $|X| = |Y| = N$ , the null hypothesis is that the paired differences  $Z_i = Y_i - X_i$  are distributed (independently and identically) symmetrically around zero (Wilcoxon, 1945). The test statistic is the sum  $W_+ = \sum_{i=1}^N iI(Z_i > 0)$ , where  $I(\cdot)$  is an indicator function. Under the null hypothesis we have that  $\text{pr}(Z_i > 0) = \text{pr}(Z_i < 0) = \frac{1}{2}$ . Assuming that  $\text{pr}(X_i = Y_i) = 0$ , which is fulfilled

with probability 1 for continuous distributions, we can then identify  $I(Z_i > 0) - I(Z_i < 0)$  with a random sign  $\gamma_i$ , such that

$$\begin{aligned} \sum_{i=1}^N \gamma_i i &= \sum_{i=1}^N i I(Z_i > 0) - \sum_{i=1}^N (1 - I(Z_i > 0))i \\ &= 2W_+ - \frac{1}{2}N(N+1). \end{aligned}$$

Therefore, testing for symmetry of the  $Z_i$  under the one-dimensional Moore-Rayleigh test is equivalent to the signed-rank Wilcoxon two-sample test of  $X$  and  $Y$ , with

$$\text{pr}(W_+ = r) = \text{pr}(S_N = 2r - \frac{1}{2}N(N+1), N).$$

This approach easily generalizes to more than one dimension.

**Remark 2** (Testing for radial dependence). Assume the density  $f$  decomposes spherically, such that  $X_i = R_i U_i$ , with  $R_i \sim p_r$  and  $U_i \sim u$ , where  $p_r(r) = \text{pr}(|X_i| = r)$  and  $u(x) = \text{pr}(X_i/|X_i| = x)$ . In one dimension,  $u$  can only attain the values  $\{-1, +1\}$  and  $u(\mp 1) = \text{pr}(X_i \leq 0)$ . If the mean of  $f$  is zero, i.e.,  $E(X_i) = 0$ , then  $\text{pr}(X_i > 0) = \text{pr}(X_i < 0) = 1/2$ , and this implies that  $f$  is (spherically) symmetric. The Moore-Rayleigh test, under the assumption that  $X_i = R_i U_i$ , therefore tests the null hypothesis that  $E(X_i) = 0$ .

On the other hand, assume that  $E(X_i) = 0$ . If the Moore-Rayleigh test finds a significant departure from uniformity, then this leads to the rejection of the hypothesis that the density  $f$  decomposes in such way, i.e., to accept the alternative that the common distribution of the random variables  $X_i$  is conditional on the length  $|X_i|$ . In practice, centering  $X = (X_1, \dots, X_N)$  by the sample mean, the Moore-Rayleigh test could be used to detect such radial dependence. However, its power would be quite limited and it seems likely that directly testing for differences in the two tails  $\{X_i > x\}$  and  $\{X_i < -x\}$  will be more powerful.

**2.2. The three-dimensional case.** Taking derivatives, the distribution function of  $R_N = \alpha_N R_N^*$ , given in Eq. (2.5), reduces to the density

$$(2.12) \quad \text{pr}(R_N = r) = \frac{2r}{\pi} \int_0^\infty t \frac{\sin rt/\alpha_N}{r} \prod_{n=1}^N \frac{\sin nt}{nt} dt$$

in the three-dimensional case ( $k = 3$ ). This formula can alternatively be derived by using characteristic functions (see Eq. 16 in Dutka (1985)). The oscillating integral in Eq. (2.12) can be evaluated by numerical quadrature, but it is difficult to calculate its tail accurately. Another approach to evaluate this integral is based on a finite series representation, following an idea originally due to G. Pólya. Let  $N_{\max} = N(N+1)/2$ . If we expand  $\sin(nt) = (e^{nt} - e^{-nt})/2i$  and integrate the oscillating integral in Eq. (2.12) by parts  $N-2$  times as in Borwein & Borwein (2001), a simple but tedious calculation (which we omit) results in the following representation:

**Theorem 1.** *The probability density of  $R_N^*$  under the null hypothesis can be evaluated as*

$$(2.13) \quad \text{pr}(R_N^* = r) = \frac{2rN^3}{N!(N-2)!} \sum_{\substack{k \in \mathbb{N}: \\ \alpha_N r < k \leq N_{\max}}} \epsilon_{k,N} (\alpha_N r - k)^{N-2},$$

where  $\epsilon_{k,N}$  is given by Eq. 2.11.

This is a generalization of Treolar's representation for the random flight with equal step sizes (Dvorák, 1972). We see that, interestingly, the density of the three-dimensional case can be expressed in terms of statistical properties of the one-dimensional case. Integrating Eq. 2.13 term-by-term from  $r$  to infinity, we have the following corollary.

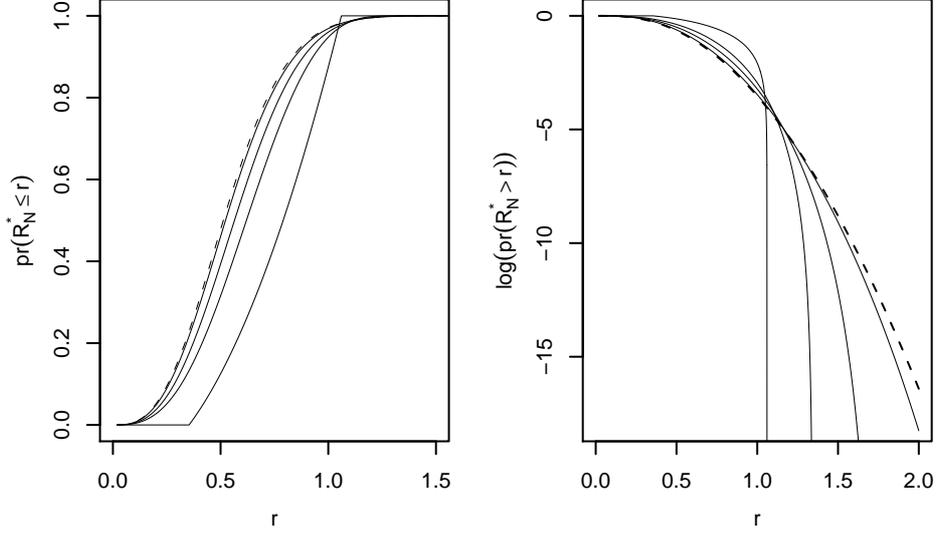


FIGURE 1. Left: Distribution function for  $N = 2, 5, 10$  and  $50$  (solid, from right to left), compared to the asymptotic  $\chi_3^2$  distribution (dashed). Right: Logarithm of significance probabilities.

**Corollary 1.** *The cumulative distribution function of  $R_N^*$  under the null hypothesis can be evaluated as*

$$(2.14) \quad \text{pr}(R_N^* \leq r) = 1 - \frac{2}{N!N!} \sum_{\substack{k \in \mathbb{N}: \\ \alpha_N r < k \leq N_{\max}}} \epsilon_{k,N} (\alpha_N r - k)^{N-1} (\alpha_N r (1 - N) - k).$$

In particular,  $\text{pr}(R_N^* > (N + 1)/(2\sqrt{N})) = 0$ .

Note that because of the representation (2.14) for smaller  $r$  successively more and more terms enter the sum in the calculation of  $\text{pr}(R_N^* > r)$ . The numerical accuracy is therefore higher for larger  $r$ , i.e., in the tail of the distribution.

The representations (2.13) and (2.14) therefore allow the efficient computation of exact significance probabilities for the test statistic  $R_N^*$  for small to moderately large sample sizes  $N$  (e.g., for  $N \lesssim 60$  under double precision IEEE 754 arithmetic). This restriction on the sample size is only due to numerical accuracy; for larger  $N$  approximations of the Gamma function can be used.

In Figure 1 the distribution of  $R_N^*$ , for some values of  $N$ , is plotted and compared with the asymptotic  $\chi_3^2$  distribution. In Table 1 the values of the quantile function are listed. These values have been calculated by numerically inverting Eq. (2.14) with a bisection method and are conservatively rounded.

**Remark 3** (What is tested by the Moore-Rayleigh test?). As in Remark 2, assume that  $X_i = R_i U_i$ , with  $R_i \sim p_r$  and  $U_i \sim u$ , where  $p_r(r) = \text{pr}(|X_i| = r)$  and  $u(x) = \text{pr}(X_i/|X_i| = x)$  are arbitrary. If  $E(X_i) = 0$ , this implies  $E(U_i) = 0$ , and suggests that  $\sum_i U_i \approx 0$  for a sample. More precisely, an upper bound for the variance of the test statistic  $R_N^*$  is realized by the one-dimensional Moore-Rayleigh null hypothesis, whose distribution is similar to the null hypothesis of the three-dimensional case (confer Figure 6). Therefore, as in the one-dimensional case, the Moore-Rayleigh test under the assumption of radial decomposability tests mostly for differences in location.

TABLE 1. Critical values of Moore-Rayleigh statistic in 3D

Sample size	Probability			-Log(Probability)							
	0.100	0.010	0.001	4	5	6	9	12	15	18	
2	1.013	1.056	1.061								
3	0.973	1.100	1.138	1.150	1.153	1.155					
4	0.948	1.116	1.189	1.222	1.237	1.244	1.250				
5	0.930	1.124	1.221	1.275	1.304	1.321	1.338	1.341	1.342		
6	0.916	1.129	1.245	1.314	1.357	1.384	1.418	1.427	1.429		
7	0.905	1.132	1.262	1.344	1.398	1.435	1.488	1.505	1.510	1.511	
8	0.897	1.133	1.275	1.368	1.432	1.477	1.549	1.576	1.586	1.588	
9	0.890	1.134	1.284	1.387	1.460	1.513	1.603	1.640	1.656	1.659	
10	0.885	1.134	1.292	1.402	1.483	1.543	1.649	1.698	1.720	1.726	
12	0.877	1.133	1.303	1.426	1.519	1.590	1.727	1.797	1.834	1.844	
14	0.871	1.133	1.310	1.443	1.545	1.626	1.788	1.879	1.931	1.946	
16	0.866	1.132	1.316	1.455	1.565	1.654	1.838	1.947	2.013	2.033	
18	0.863	1.132	1.320	1.464	1.580	1.675	1.878	2.003	2.083	2.108	
20	0.860	1.131	1.323	1.472	1.593	1.693	1.911	2.051	2.144	2.174	
30	0.851	1.129	1.331	1.493	1.629	1.746	2.016	2.209	2.350	2.399	
40	0.847	1.128	1.335	1.503	1.647	1.771	2.071	2.294	2.467	2.529	
50	0.844	1.127	1.337	1.509	1.657	1.787	2.103	2.347	2.540	2.612	
60	0.843	1.126	1.338	1.513	1.664	1.797	2.125	2.382	2.590	2.668	
$\infty$	0.834	1.123	1.345	1.532	1.697	1.846	2.233	2.559	2.847		

Note that symmetry of the  $U_i$ , i.e.,  $\text{pr}(U_i = u) = \text{pr}(U_i = -u)$ , implies that  $E[\sum_i U_i] = 0$ . Thus, under the assumption of decomposability, testing for spherical symmetry and testing for symmetry are approximately equivalent, i.e., the Moore-Rayleigh test will not be sensitive to deviations from spherical uniformity if the underlying distribution is merely symmetric or mean-centered. This is actually an advantage when the Moore-Rayleigh test is considered as a two-sample test (see below).

**2.3. Power estimates.** To evaluate the performance of the three-dimensional Moore-Rayleigh test (MR3), power functions for a number of distributions were obtained by Monte-Carlo simulation. These show the fraction of rejections of the null hypothesis for a specific distribution, significance level  $\alpha$ , and sample size  $N$ . The left panel of Figure 2 shows the power function for a family of diagonal Gaussian distributions with unit variances, shifted away from zero (along the  $z$ -axis) a constant distance  $\mu \geq 0$ . Each point power estimate was obtained by 1000 realizations of the distributions and represents the fraction of significance probabilities (“p-values”) less than the nominal significance level  $\alpha$ . The test was performed on  $N = 10$  randomly drawn samples, and is compared to Hotelling’s (non-randomized)  $T^2$  one-sample test of location (Hotelling, 1931), as implemented in the R package ICSNP (Nordhausen et al., 2007), and to the spherical uniformity permutation test of Diks & Tong (1999), under  $10^4$  resamplings. The test statistic of the latter is an U-estimator of the difference between two probability distributions of vectors, calculated by a Gaussian kernel with a bandwidth parameter. The choice of the proper bandwidth is the subject of ongoing research; we show results for the two bandwidths  $b_1 = 0.25$  and  $b_2 = 2.5$ , and denote the corresponding tests by “Diks1” and “Diks2”, respectively.

In comparison with the  $T^2$  test, MR3 shows larger power, an effect that is more pronounced for lower significance levels. It is thus a more sensitive measure of changes in location. Note that this does not contradict the well-known optimality of Hotelling’s  $T^2$  test for the family of multivariate Gaussian distributions, since in the calculation of  $T^2$  the covariance matrix needs to be estimated from the data. In the special case of equal covariances considered here, the Moore-Rayleigh test can therefore exhibit larger power. Also

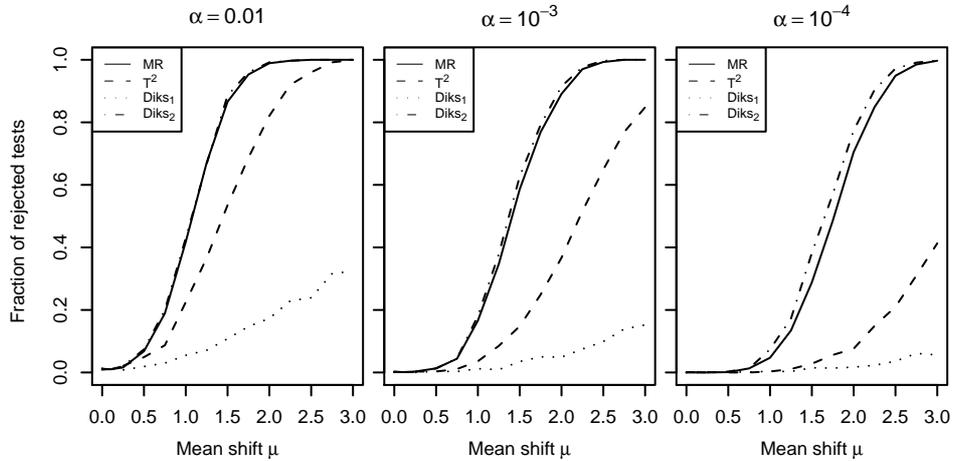


FIGURE 2. Estimated power functions for the family of Gaussian distributions with covariance matrix the identity and mean shifted a distance  $\mu$  away from zero. Sample size  $N = 10$ .

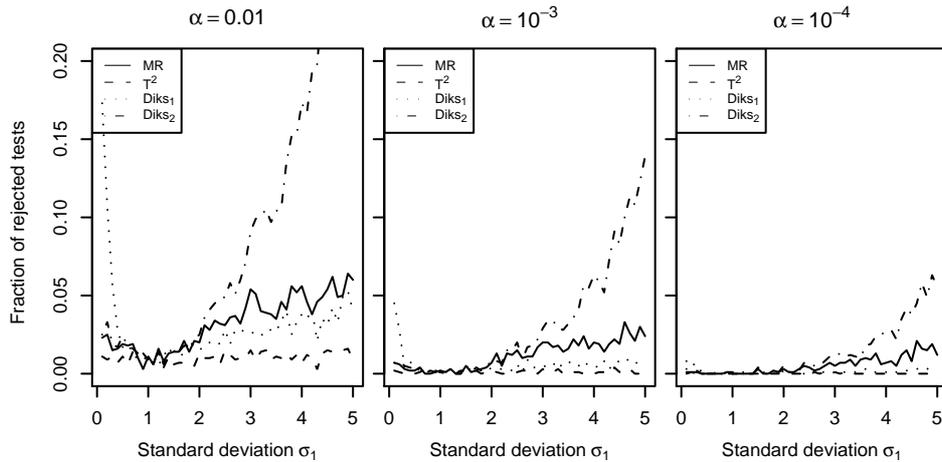


FIGURE 3. Estimated power functions for the family of Gaussian distributions, varying the standard deviation  $\sigma$  of a single axis. Sample size  $N = 10$ . Note the small range of the power.

note that the test of Diks & Tong can be more powerful than the MR3 test, but as its results depend strongly on the bandwidth parameter, it is difficult to apply it routinely.

In Figure 3, power functions are shown for a family of diagonal Gaussian distributions where the standard deviation of one axis was varied from  $\sigma = 0.1$  to  $\sigma = 5.0$  in steps of 0.1, the other standard deviations were kept at unity. As expected from Remark 3, the MR3 test performs poorly for this specific violation of spherical symmetry. The remaining symmetry in the distribution means that although sample points are now increasingly less concentrated on one axis, on average their contributions to the resultant length still mostly cancel each other. Analogously, the  $T^2$  test has only nominal power for the anisotropic multivariate Gaussian, being a test of location only. Note that MR3 shows slightly more power than the nominal significance levels  $\alpha$  for  $\sigma \neq 1$ , as do the tests “Diks1” and “Diks2”.

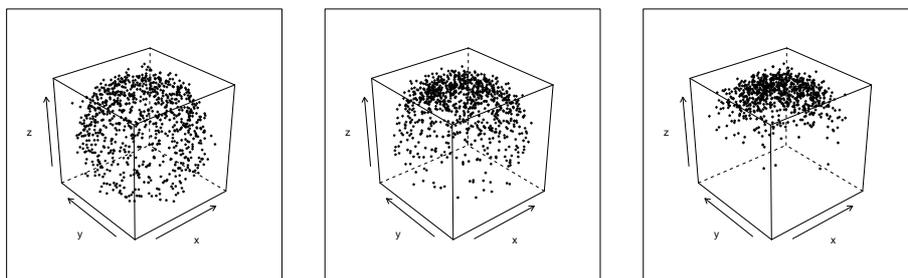


FIGURE 4. Scattered Fisher distribution, visualized by 1000 randomly drawn points in the unit-cube. Left: Concentration parameter  $\lambda = 1$ . Middle: Concentration parameter  $\lambda = 2.5$ . Right: Concentration parameter  $\lambda = 5$ .

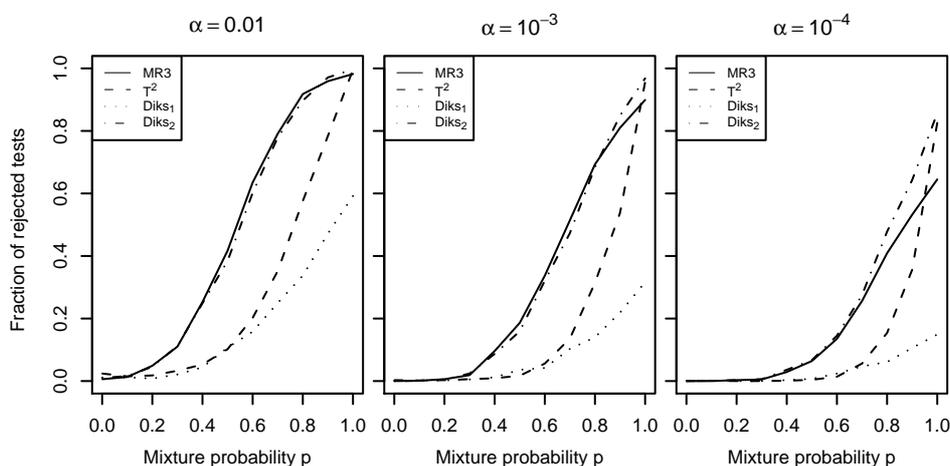


FIGURE 5. Estimated power functions for the mixture of the scattered Fisher distribution  $2F3_5$  with the uniform distribution on the sphere  $S^2$ , varying the mixture probability  $p$  that a sample vector arises from the first distribution. Sample size  $N = 10$ .

To assess the effect of asymmetry of the sample distribution, we employ the Fisher distribution, also known as the Fisher–Bingham three-parameter distribution. This is the  $k = 3$  case of the  $k$ -dimensional von–Mises Fisher distributions commonly used in directional statistics (Mardia & Jupp, 2000, Chap. 9.3.2). Details of its computation are given in the Appendix.

We denote the Fisher distribution with concentration parameter  $\lambda$  (and with the choice  $\xi = e_3$ ) by  $F3_\lambda$ . To avoid degeneracies due to its singular character, the  $F3_\lambda$  distribution is multiplied by  $1 - Z$ , where  $Z \sim \mathcal{N}(0, 0.1)$ . Figure 4 shows three examples of  $N = 1000$  random variates obtained from these “scattered” Fisher distributions for distinct values of the concentration parameter  $\lambda$ , with increasingly larger deviation from the uniform distribution.

The power of MR3 for the family of scattered Fisher distributions, varying the concentration parameter, is comparable to the power of the other tests (not shown). Let us now consider a mixture, where the samples are chosen either (i) from the uniform distribution on the unit sphere, or (ii) from the scattered Fisher distribution  $2F3_5$ . The probability

$0 \leq p \leq 1$  for each sample vector to be chosen from the second distribution is the parameter of this family of mixture distributions, with larger  $p$  indicating stronger deviations in uniformity for the larger vectors. Figure 5 depicts the estimated power for this family under variation of the mixture probability  $p$ . Compared to the  $T^2$  test, the MR3 test is seen to be more sensitive to these specific departures from uniformity. It should be noted that reversing the situation, e.g., by considering  $F3_5/2$  instead of  $2F3_5$ , such that the smaller vectors exhibit deviations from uniformity, the power of MR3 becomes less than that of the  $T^2$  test (not shown).

### 3. THE TWO-SAMPLE TEST

The most interesting application of the Moore-Rayleigh test is the two-sample problem. There, we are given two vector-valued random variables

$$(3.1) \quad X = (X_1, \dots, X_N) \quad \text{and} \quad Y = (Y_1, \dots, Y_N),$$

and we assume that they are identically and independently distributed with densities  $f$  and  $g$ , respectively. The differences  $Y_j - X_i$  are then distributed according to the convolution  $g * (-f)$ , whose density is

$$(3.2) \quad \text{pr}(Y - X = x) = \int \text{pr}(Y = u) \text{pr}(X = u + x) \, d^k u.$$

Under the null hypothesis that the  $X_i$  and  $Y_j$  come from a common probability density  $f$ , this reduces to the *symmetrization* of  $f$ , with density

$$(3.3) \quad \text{pr}(Y - X = x) = \int \text{pr}(X = u) \text{pr}(X = u + x) \, d^k u.$$

If the probability density  $f$  is *spherically symmetric* around its mean  $\mu$ , i.e., uniform on each hypersphere  $\{x \mid \|x - \mu\| = r\}$ , then Eq. (2.14) gives the significance probability of a deviation from the null hypothesis. In particular, this applies when  $f$  is assumed to arise from a multivariate normal distribution, justifying the use of the Moore-Rayleigh statistic in many practical situations.

**3.1. Testing for symmetry.** In general, however, the distribution of  $h = f * (-f)$  is merely *symmetric*, i.e.,  $h(x) = h(-x)$  for all  $x \in \mathbb{R}^k$ . This follows from

$$(3.4) \quad \int \text{pr}(X = u) \text{pr}(X = u + x) \, d^k u = \int \text{pr}(X = u) \text{pr}(X = u - x) \, d^k u.$$

The following example demonstrates the difference.

**Example 1.** Consider the symmetric singular distribution  $B_x \stackrel{\text{def}}{=} \frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x}$ , where  $\delta_x$  is the Dirac measure concentrated at the point  $x \in \mathbb{R}^k$ . The distribution  $B_x$  leads to an embedding of the one-dimensional Moore-Rayleigh null distribution in three dimensional space. Its realizations take values  $x$  and  $-x$  with equal probability, and it is not spherically symmetric. As it is,  $B_x$  is neither absolutely continuous, nor can it arise as the symmetrization of a distribution. Nevertheless, it is a model for a distribution that can arise in practice: First, the delta distributions can be approximated, e.g., by a series of Gaussian distributions with decreasing variance. Secondly, consider the singular distribution  $B^x$  that is concentrated on a line  $\{\lambda x \mid \lambda \in \mathbb{R}\} \subseteq \mathbb{R}^k$  through the origin. Applying the Moore-Rayleigh test to  $B^x$  is equivalent to calculating the test statistic from  $B_1$ , since  $B^x$  is invariant under symmetrization and is projected, before ranking, to the sphere  $S^0 = \{-1, +1\}$ .

The distribution  $B_1$  is a representative of the class of “fastest growing” random flights in three dimensions, since any other distribution of increments has less or equal probability to reach the highest values of the test statistic. On the other hand, the uniform distribution on the sphere, which represents the null hypothesis of the Moore-Rayleigh test statistic  $R_N^*$ , will attain lower values of  $R_N^*$  with higher probability, as the uniform random walk

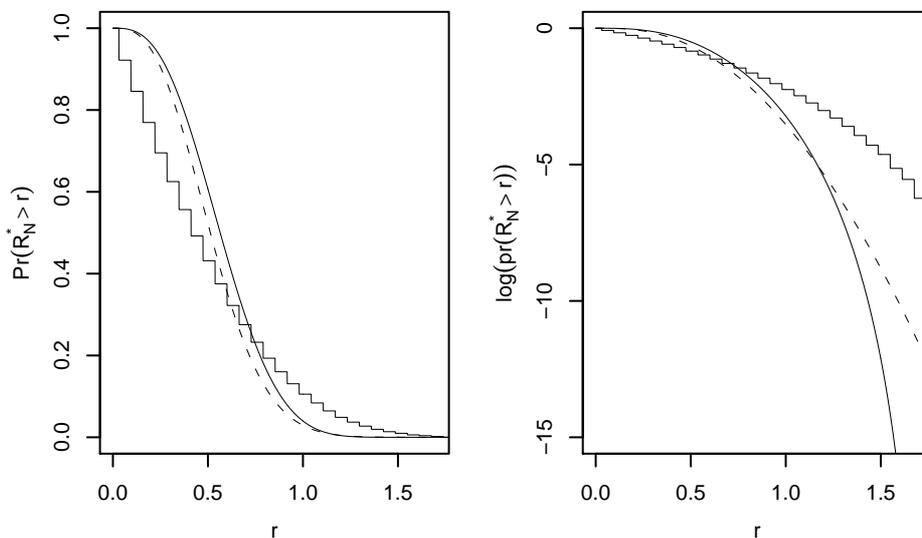


FIGURE 6. Comparison of significance probabilities for resultant lengths of spherically symmetric (smooth curve) and one-dimensional symmetric random walk (piecewise-linear curve) in three dimensions for  $N = 10$  steps. Dotted curve shows the asymptotic case.

can do “orthogonal” steps that increase the distance from the origin faster than in  $B_1$  (on the average). To be specific, if the finite sample  $X$  is distributed according to  $B_1$ , the  $n$ -th step of the scaled random walk either increases or decreases the distance from the origin by  $n$  (when crossing the origin, there is an obvious correction to this). However, if the  $n$ -th step were taken in a direction that is orthogonal to the resultant obtained so far, the distance will increase from  $R$  to  $\sqrt{R^2 + n^2} \approx R + n/(2R)$ , with probability 1 (conditional on the orthogonality).

Figure 6 compares significance probabilities for  $B_1$  with those of the uniform random flight that represents the null hypothesis of the Moore-Rayleigh test, for  $N = 10$  sample points. There exists a value of the test statistic where the two curves cross (at about  $p = 0.20$ ), and after which the distribution function (significance probability) of the one-dimensional random walk  $B_1$  lies below (above) the one for the uniform random flight.

The two-sample Moore-Rayleigh test, interpreted as a goodness-of-fit test, is therefore liberal, which has escaped Moore (1980) from his attention and casts doubt on the applicability of the test in this setting. The optimal upper bound for a conservative significance probability would be

$$(3.5) \quad G_N^*(r) = \sup_{\Psi_N} \Pr(|S_N| \geq r),$$

where the supremum is taken over the set  $\Psi_N$  of all possible symmetric probability distributions for  $N$  increments. More precisely, these increments are not independent but arise from a mixture of independent distributions by the order distribution (due to the ranking of vector lengths) of their radial projections. Even if one restricts this to the class where only independent, not necessarily identical symmetric probability distributions for each step are considered, this is a difficult problem. First steps in this direction have been made by Kingman (1963), where the three-dimensional problem is reduced to a similar problem in one dimension by the familiar tangent-normal decomposition of the sphere. Apart from that, there has not been much progress in determining the envelope in Eq. 3.5. Even in the

one-dimensional case it is not clear what the “fastest” random flight with linearly bounded increments is.

If a liberal test is admissible for the specific problem at hand, e.g., in exploratory data analysis, MR3 offers an efficient two-sample test. Moreover, the Remarks and Figure 3 suggest that the significance probabilities are only liberal for relatively large values of the test statistic. Studies with synthetic data seem to confirm that the MR3 test fails gracefully, if at all, for distributions expected in biomedical imaging practice (Scheenstra et al., accepted).

Since the assumed null hypothesis is stronger than mere symmetry, MR3 can also be used for negative testing, i.e., if the null hypothesis of the uniform random flight cannot be rejected for a sample of difference vectors, then the modified null hypothesis that  $g * (-f)$  is symmetric, not necessarily spherically symmetric, cannot be rejected. For the null hypothesis of mere symmetry, there does not exist an accessible sufficient statistic and existing tests are either only asymptotically nonparametric or require further randomization of the underlying distribution (Aki, 1987; Jupp, 1987; Diks & Tong, 1999; Henze et al., 2003; Fernández et al., 2008; Ngatchou-Wandji, 2009), so the MR3 test offers a simpler and much more efficient alternative, albeit with the disadvantage that it is potentially liberal.

**3.2. Further issues.** For a truly conservative test it is possible to adjust the  $p$ -values of the MR3 test by bootstrapping the distribution of  $p$  as in Davison & Hinkley (1997). In practice this makes use of the exchangeability of the vectors from  $X$  and  $Y$ , assuming that they both arise from the same distribution. For each pair  $Y_i - X_i$  we can therefore introduce a random sign  $\epsilon_i \in \{-1, +1\}$ . The fraction of the test statistics  $R_N^*(\epsilon)$  under all  $2^N$  possibilities of the signs  $\epsilon = (\epsilon_1, \dots, \epsilon_N)$  that result in a larger value than the one for the trivial signs (all positive) results in an exact  $p$ -value; confer Diks & Tong (1999) for elaboration and also Lehmann & Romano (2005) for general background on symmetries and invariance in hypothesis testing. The drawback of this adjustment is that it is not routinely feasible, as it suffers from the same computational complexity problems that affect conventional permutation tests.

A different issue with the Moore-Rayleigh test arises in the (usual) case of unpaired samples. The two sample test we have presented up to now assumes paired vectors, and this approach reduces the symmetry group of the null hypothesis from the group of permutations to the much smaller group of reflection symmetry of the given pairs. The main reason here is simplicity in applications and reproducibility of the test statistic. If there is no natural pairing, it seems advisable to randomly pair samples, as e.g. Moore (1980) advocates. However, a drawback is that the test statistic then becomes a random variable, and replications of the test will result in distinct significance probabilities. This is undesirable, for example, in a clinical context. Bootstrapping the test, i.e., considering the mean of the test statistic  $R_N^*$  obtained during a large enough number of resamples from the empirical distributions, is a natural way to obtain more or less replicable significance probabilities, but on the expense of computational time. It is also not precisely known at present what the convergence properties of such an estimator are.

A different approach would be to pair samples based on a measure of optimality. This seems natural enough, but has the undesirable feature that the test might become biased, e.g., too sensitive in case the sample points are matched by the method of least-squares or the Wasserstein distance. Therefore, as a practical solution in a context where reproducibility is desired, we propose to pair samples based on their ranks, such that  $X_{(i)}$  is matched with  $Y_{(i)}$ ,  $i = 1, 2, \dots, N$  (with ties resolved arbitrarily). Under the null hypothesis, the decomposability of the common distribution of  $X$  and  $Y$  guarantees the asymptotic unbiasedness of this approach, although for finite samples a slight bias is expected.

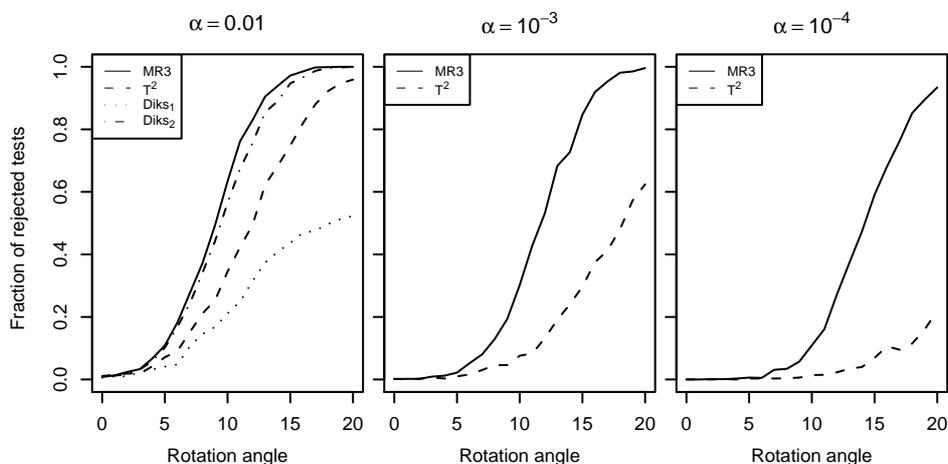


FIGURE 7. Estimated power for translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Sample size  $N = 10$ .

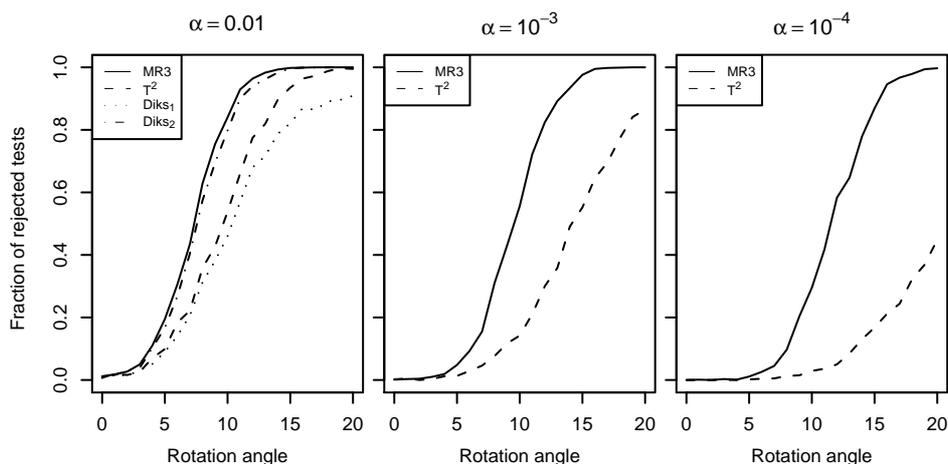


FIGURE 8. Estimated power functions for translated (10 standard deviations), then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size  $N = 10$ . Note the small power at angle zero.

#### 4. SIMULATION RESULTS

In this section we show the results of a number of numerical simulations for the two-sample problem and compare them with Hotelling's  $T^2$  test and the "Diks1" and "Diks2" tests. Throughout, we use random matching of samples and  $N = 10$ .

Figure 7 shows the results for two standard Gaussian distributions that were first translated in the same direction by ten standard deviations, and then one of them was rotated against the other (with the origin as the center of rotation), for 1000 realizations. The Moore-Rayleigh test performs well: Its power for the trivial rotation is nominal, and for larger rotation angles higher than the power of the  $T^2$  test. Similar results are obtained

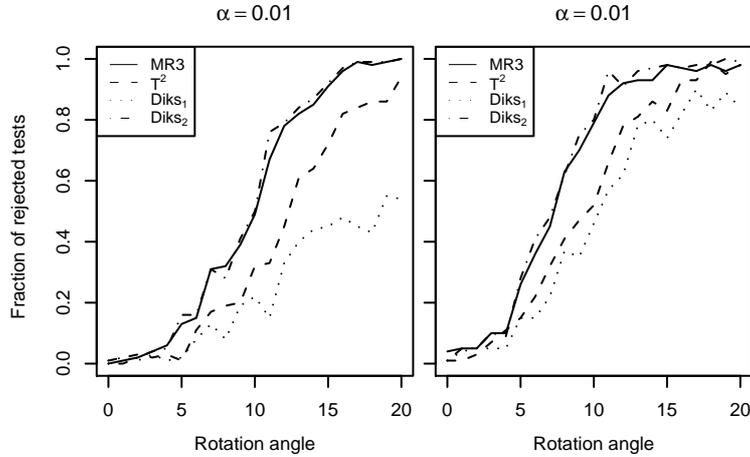


FIGURE 9. Adjusted estimated power functions. Left: translated (10 standard deviations), then rotated diagonal Gaussians with unit variances as a function of relative rotation angle. Right: translated (10 standard deviations) then rotated diagonal Gaussians with unit variance against the scattered Fisher distribution (scaled by a unit Gaussian) as a function of relative rotation angle. Sample size  $N = 10$ . Results based on 100 realizations of  $10^4$  permutations each.

when rotating Fisher distributions (not shown). Note that the “Diks1/Diks2” tests are performed on the group of symmetric sign changes (of order  $2^{10}$ ), in contrast to the previous section where the full symmetry group of all rotations (of infinite order) was used, and do not resolve significance probabilities smaller than  $1/1000$ , i.e., their power is zero for the lower significance levels, and therefore not indicated.

Figure 8 compares the Gaussian distribution with the distribution  $R \cdot F_{3\lambda}$ ,  $R \sim \mathcal{N}(0, 1)$ , when both distributions are first translated and then rotated against each other, with similar results.

Finally, Figure 9 shows adjusted  $p$ -values, for  $10^4$  permutations and 100 realizations each. The Moore-Rayleigh test again shows slightly better power than the  $T^2$  test. More importantly, there is not much difference with the unadjusted power functions. These results are based on 100 realizations only, to speed up the considerable amount of computations, which accounts for the visible fluctuations.

## 5. DISCUSSION

It is possible to test spherical symmetry in three dimensions with high numerical accuracy by using the combinatorial sum representation given in Eq. (2.14). In combination with the Kahan summation (Goldberg, 1991), this representation makes it feasible to routinely calculate  $p$ -values for finite sample sizes that allow to assess statistical significance. Even for hundreds of thousands of multiple comparisons with a Bonferroni correction, as is common practice in neuroscientific imaging applications, the proposed approach is effective. Permutation methods, although theoretically preferred, are difficult to use in this setting due to practical limitations. The standard approaches to cope with these limitations, based on either saddle-point approximations to permutation tests (Robinson, 1982) or on permutation tests for linear test statistics, where the conditional characteristic function can be rewritten as a convergent approximating series (Gill, 2007), are not directly applicable because these statistics usually do not arise in these practical problems or are too involved in the multivariate case. An alternative might be the use of optimal (Bayesian) stopping

rules in the resampling process (Besag & Clifford, 1991; Fay et al., 2007). However, small sample sizes can still seriously restrict the possible range of the significance probabilities.

In the special case of the two-sample problem, the distribution of the null hypothesis is conditional on the unknown distribution of the data, and the generalized Moore-Rayleigh test is only approximately valid, a feature that all other (non-randomized) tests of symmetry exhibit. In Scheenstra et al. (accepted), we evaluated the properties of this generalized Moore-Rayleigh test empirically with simulated imaging data of known ground-truth and by comparison with other nonparametric tests. Even though the test is theoretically liberal, it seems to work well in practice, as it is not overtly sensitive to the difference between symmetry and spherical symmetry. An exact test is furthermore available by the permutation variant of the Moore-Rayleigh test, with slightly improved power when compared with conventional permutation testing. This can be used in a second stage after initial screening with the fast, unadjusted Moore-Rayleigh test. Although such screening could also be realized by the  $T^2$  test, the MR3 test seems better suited to this problem due to its enhanced power, which allows for strong control of the family-wise error. In contrast, the  $T^2$  test does often not allow the localization of individual voxels. It should be noted that we have only considered the conservative Bonferroni correction here, for simplicity, but it is expected that the MR3 test remains a more sensitive instrument also under modern step-down multiple comparison procedures (as described in, e.g., Nichols & Holmes (2007)).

An implementation of the Moore–Rayleigh test as a package for the statistical computing environment R is available on request.

#### ACKNOWLEDGMENTS

M. Muskulus was partially funded by the Netherlands Organization for Scientific Research (NWO) under grant #635.100.006. *Conflict of Interest*: None declared.

#### APPENDIX A. THE FISHER DISTRIBUTION

The Fisher distribution is a singular distribution on the hypersphere  $S^{k-1}$  whose density  $f(x)$ ,  $x \in \mathbb{R}^k$ , is proportional to  $e^{\lambda \xi^t x}$ , where  $\xi^t$  denotes the transpose of  $\xi$ . The mean direction  $\xi$  is constrained to be a unit vector, and  $\lambda \geq 0$  is a concentration parameter. Without restricting generality, we let  $\xi = e_k$  be the unit vector in the  $k$ -th dimension, so  $f \sim e^{\lambda x_k}$  only depends on the last coordinate, and we are left with a one-parameter family of distributions.

Following Ulrich (1984) and Wood (1994), a random variate distributed according to the von–Mises Fisher distribution is obtained by generating a random variate  $W$  for the last coordinate, by the density proportional to

$$e^{\lambda w} (1 - w^2)^{(k-3)/2}, \quad w \in (-1, 1), \quad k \geq 2,$$

and a  $k - 1$  dimensional variate  $V$  uniformly distributed on the hypersphere  $S^{k-2}$ . The vector

$$X = (\sqrt{1 - W^2} \cdot V^t, W) \in \mathbb{R}^k$$

then has the desired density. In  $k = 3$  dimensions the former can be achieved by integrating the distribution function of  $W$  directly. Choosing a uniform variate  $U$  on the interval  $[-1, 1]$ , a random variate  $W$  is clearly given by

$$W = \frac{1}{\lambda} \log(2U \sinh \lambda + e^{-\lambda}).$$

#### REFERENCES

- ABLE, K. P. & ABLE, M. A. (1997). Development of sunset orientation in a migratory bird: no calibration by the magnetic field. *Anim. Behav.* **53**, 363–368.
- AKI, S. (1987). On nonparametric tests for symmetry in  $R^m$ . *Ann. Inst. Statist. Math.* **39**, 457–472.

- BENJAMINI, Y. & HOCHBERG, Y. (1995). Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**, 289–300.
- BESAG, J. & CLIFFORD, P. (1991). Sequential Monte Carlo p-values. *Biometrika* **78**, 301–304.
- BORWEIN, D. & BORWEIN, J. M. (2001). Some remarkable properties of sinc and related integrals. *Ramanujan J.* **5**, 73–89.
- BURTON, N. H. K. (2006). Nest orientation and hatching success in the tree pipit *Anthus trivialis*. *J. Avian Biol.* **37**, 312–317.
- CAO, J. & WORSLEY, K. J. (1999). The detection of local shape changes via the geometry of Hotelling's  $t^2$  fields. *Annals Stat.* **27**, 925–942.
- CHEN, X. J., KOVACEVIC, N., LOBAUGH, N. J., SLED, J. G., HENKELMAN, R. M. & HENDERSON, J. T. (2005). Neuroanatomical differences between mouse strains as shown by high-resolution 3D MRI. *NeuroImage* **29**, 99–105.
- CHERNETSOV, N., CHROMIK, W., DOLATA, P. T., PROFUS, P. & TRYJANOWSKI, P. (2006). Sex-related natal dispersal of white storks (*Ciconia Ciconia*) in Poland: How far and where to? *The Auk* **123**, 1103–1109.
- CHUNG, M. K., WORSLEY, K. J., PAUS, T., CHERIF, C., COLLINS, D. L., GIEDD, N., RAPOPORT, J. L. & EVANS, A. C. (2001). A unified statistical approach to deformation-based morphometry. *NeuroImage* **14**, 595–606.
- DAVISON, A. C. & HINKLEY, D. V. (1997). *Bootstrap Methods and their Applications*. Cambridge University Press.
- DIKS, C., TONG, H. (1999). A test for symmetries of multivariate probability distributions. *Biometrika* **86**, 605–614.
- DUDLEY, R. M. (2002). *Real analysis and probability*. Cambridge University Press.
- DUTKA, J. (1985). On the problem of random flights. *Arch. Hist. Exact Sci.* **32**, 351–375.
- DVORÁK, S. (1972). Treolar's distribution and its numerical implementation. *J. Phys. A: Gen. Phys.* **5**, 78–84.
- FAY, M. P., KIM, H.-J. & HACHEY, M. (2007). On using truncated sequential probability ratio test boundaries for Monte Carlo implementation of hypothesis tests. *J. Comput. Graph. Stat.* **16**, 946–967.
- FERNÁNDEZ, V. A., GARMERO, M. D. J. & GARÍA, J. M. (2008). A test for the two-sample problem based on empirical characteristic functions. *Comput. Stat. Data Anal.* **52**, 3730–3748.
- GASTWIRTH, J. L. (1965). Percentile modifications of two sample rank tests. *J. Amer. Stat. Assoc.* **60**, 1127–1141.
- GILL, P. M. W. (2007). Efficient calculation of p-values in linear-statistic permutation significance tests. *J. Stat. Comput. Sim.* **77**, 55–61.
- GOLDBERG, D. (1991). What every computer scientist should know about floating-point arithmetic. *ACM Comp.* **23**, 5–48.
- HENZE, N., KLAR, B. & MEINTANIS, S. G. (2003). Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *J. Multivariate Anal.* **87**, 275–297.
- HOTELLING, H. (1931). The generalization of Student's ratio. *Ann. Math. Statist.* **2**, 360–378.
- JUPP, P. E. (1987). A nonparametric correlation coefficient and a two-sample test for random vectors or directions. *Biometrika* **4**, 887–890.
- KAJIKAWA, Y. & HACKETT, T. A. (2005). Entropy analysis of neuronal spike train synchrony. *J. Neurosci. Meth.* **149**, 90–93.
- KINGMAN, J. F. C. (1963). Random walks with spherical symmetry. *Acta Math.* **63**, 11–53.
- KOVACEVIC, N., CHEN, J., SLED, J. G., HENDERSON, J. T. & HENKELMAN, R. M. (2004). Deformation based representation of groupwise average and variability. *Lect.*

- Notes Comput. Sci.* **3216**, 616–622.
- LEHMANN, E. L., ROMANO, J.P. (2005). *Testing statistical hypotheses*. Springer.
- LORD, R. D. (1954). The use of Hankel transforms in statistics. I. General theory and examples. *Biometrika* **41**, 44–55.
- MARDIA, K. V. & JUPP, P. E. (2000). *Directional statistics*. John Wiley & Sons.
- MCNAUGHT, M. K. & OWENS, I. P. F. (2002). Interspecific variation in plumage colour among birds: species recognition or light environment? *J. Evolution. Biol.* **15**, 505–514.
- MOORE, B. R. (1980). A modification of the Rayleigh test for vector data. *Biometrika* **67**, 175–180.
- NGATCHOU-WANDJI, J. (2009). Testing for symmetry in multivariate distributions. *Stat. Methodol.* **6**, 230–250.
- NICHOLS, T. & HOLMES, A. (2007). Non-parametric procedures. In *Statistical Parametric Mapping. The Analysis of Functional Brain Images*, K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols & W. D. Penny, eds., chap. 21. Elsevier, pp. 253–272.
- NORDHAUSEN, K., SIRKIA, S., OJA, H. & TYLER, D. E. (2007). *ICSNP: Tools for Multivariate Nonparametrics*. R package version 1.0-2.
- RICHARDSON, A. G., LASSI-TUCCI, G., PADOA-SCHIOPPA, C. & BIZZI, E. (2008). Neuronal activity in the cingulate motor areas during adaption to a new dynamic environment. *J. Neurophysiol.* **99**, 1253–1266.
- ROBINSON, J. (1982). Saddlepoint approximations for permutation tests and confidence intervals. *J. R. Statist. Soc. B* **44**, 91–101.
- SCHEENSTRA, A. E. H., MUSKULUS, M., STARING, M., VAN DEN MAAGDENBERG, A. M. J. V. et al. (accepted). The 3D Moore-Rayleigh test for the quantitative groupwise comparison of MR brain images. In *21st Conference on Information Processing in Medical Imaging (IPMI '09)*, Lect. Notes Comp. Sci.
- SCHWARTZMAN, A., DOUGHERTY, R. F., LEE, J., GHAHREMANI, D., TAYLOR, J. E. (2009). Empirical null and false discovery rate analysis in neuroimaging. *NeuroImage* **44**, 71–82.
- STEPHENS, M. A. (1962). Exact and approximate tests for directions. I. *Biometrika* **49**, 463–477.
- TUKKER, J. J., FUENTEALBA, P., HARTWICH, K., SOMOGYI, P. & KLAUSBERGER, T. (2007). Cell type-specific tuning of hippocampal interneuron firing during gamma oscillations *in vivo*. *J. Neurosci.* **27**, 8184–8189.
- ULRICH, G. (1984). Computer generation of distributions on the sphere. *Appl. Stat.* **33**, 158–163.
- VAN BEERS, R. J., HAGGARD, P. & WOLPERT, D. M. (2004). The role of execution noise in movement variability. *J. Neurophysiol.* **91**, 1050–1063.
- WILCOXON, F. (1945). Individual comparisons by ranking methods. *Biometrics Bull.* **1**, 80–83.
- WOOD, A. T. A. (1994). Simulation of the von Mises Fisher distribution. *Comm. Stat. - Sim. and Comp.* **23**, 157–164.

MATHEMATICAL INSTITUTE, LEIDEN UNIVERSITY, NIELS BOHRWEG 1, 2333 CA LEIDEN, THE NETHERLANDS

*E-mail address:* muskulus@math.leidenuniv.nl

*URL:* <http://www.math.leidenuniv.nl/~muskulus>

DEPARTMENT OF RADIOLOGY, LEIDEN UNIVERSITY MEDICAL CENTER, ALBINUSDREEF 2, 2333 ZA LEIDEN, THE NETHERLANDS

*E-mail address:* A.E.H.Scheenstra@lumc.nl

MATHEMATICAL INSTITUTE, LEIDEN UNIVERSITY, NIELS BOHRWEG 1, 2333 CA LEIDEN, THE NETHERLANDS

*E-mail address:* verduyn@math.leidenuniv.nl