Nonlinear Analysis: Modelling and Control, 2010, Vol. 15, No. 1, 15–28

N-distance tests of uniformity on the hypersphere

A. Bakshaev

Institute of Mathematics and Informatics
Akademijos str. 4, LT-08663 Vilnius, Lithuania
aleksej.bakshaev@gmail.com

Received: 2009-10-30 **Revised:** 2010-01-06 **Published online:** 2010-03-04

Abstract. In this paper we propose an application of N-distance theory for testing the hypothesis of uniformity on hypersphere S^{p-1}. The work is a continuation of our research started in [1, 2]. Particular attention is devoted to $p = 2, 3$ cases. A brief comparative Monte Carlo power study for proposed criteria is provided.

Keywords: tests of uniformity, N-distances.

1 Introduction

Several invariant tests for uniformity of a distribution on the circle, the sphere and the hemisphere have been proposed by Rayleigh [3, 4], Watson [5, 6], Ajne [7], Beran [8] and others. In this paper we propose an application of N-distance theory for testing the hypothesis of uniformity of spherical data. The proposed procedures have a number of advantages: consistency against all fixed alternatives, invariance of the test statistics under rotations of the sample, computational simplicity and ease of application even in high-dimensional cases.

We start from a brief review of N-distance theory. Then some new criteria of uniformity on S^{p-1} based on N-metrics are introduced. Particular attention is devoted to $p = 2$ (circular data) and $p = 3$ (spherical data). In these cases the asymptotic behavior of proposed tests under the null hypothesis is established using two approaches: first is based on an adaptation of methods of goodness of fit tests described in [1, 2], and second using Gine theory based on Sobolev norms [9, 10].

At the end of the paper we present a brief comparative Monte Carlo power study for proposed uniformity criteria. S^1 and S^2 cases are considered. Analyzed tests are compared with classical criteria: Rayleigh, Giné and Ajne using a variety of alternative hypotheses (see also [3]). Results of simulations show that the proposed tests are powerful competitors to existing classical criteria.

2 N-distances

Tests, proposed in this article are based on a class of probability metrics – N-distances, introduced by Klebanov in [11]. These metrics, generated by negative definite kernels are very convenient and allow obtaining new statistical criteria for testing parametric and nonparametric hypothesis in arbitrary dimension.

Let (X, U) be a measurable space and B the set of all probability measures μ on it. Suppose that L is a real continuous function, and denote by B_L the set of all probability measures μ on (X, U) under condition

$$\int\limits_X \int\limits_X L(x, y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) < \infty.$$

Denote by

$$\begin{aligned} N(\mu, \nu) := 2 \int\limits_X \int\limits_X L(x, y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) - \int\limits_X \int\limits_X L(x, y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\ - \int\limits_X \int\limits_X L(x, y)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y), \end{aligned} \tag{1}$$

where $\mu, \nu \in B_L$.

The theorem, proved by Klebanov [11], says that if $L(x, y) = L(y, x)$ and $L(x, x) = 0\ \forall x, y \in X$ the inequality

$$N(\mu, \nu) \geq 0$$

holds for all measures $\mu, \nu \in B_L$ with equality in the case $\mu = \nu$ only, if and only if L is a strongly negative definite kernel. This fact allows us to obtain consistent tests against all fixed alternatives.

Some examples of strongly negative definite kernels for practical usage can be found in Section 4 or in [1, 2, 11].

3 Tests of uniformity on the hypersphere

3.1 Statement of the problem

Consider the sample $X_1, \dots, X_n$ of observations of random variable X, where $X_i \in \mathbb{R}^p$ and $\|X_i\| = 1$, $i = 1, \dots, n$. Let us test the hypothesis H_0 that X has a uniform distribution on S^{p-1}.

The statistics for testing H_0 based on N-distance with the kernel $L(x, y)$ have the form

$$T_n = n\left[\frac{2}{n}\sum_{i=1}^{n} \mathbf{E}_{\mathbf{Y}}\{L(X_i, Y)\} - \frac{1}{n^2}\sum_{i,j=1}^{n} L(X_i, X_j) - \mathbf{E}\big\{L\big(Y, Y'\big)\big\}\right], \tag{2}$$

where X, Y, Y' are independent random variables from the uniform distribution on S^{p-1} and $\mathbf{E}_{\mathbf{Y}}\{L(X_i, Y)\} = \int L(X_i, y)\,\mathrm{d}F_Y(y)$ is a mathematical expectation calculated by Y with fixed X_i, $i = 1, \ldots, n$.

We should reject the null hypothesis in case of large values of our test statistics, that is if $T_n > c_\alpha$, where c_α can be found from the equation:

$$P_0(T_n > c_\alpha) = \alpha,$$

where P_0 is the probability distribution corresponding to the null hypothesis and α is the size of the test.

For our further research let us consider a strongly negative definite kernels of the form $L(x, y) = G(\|x - y\|)$, where $\|\cdot\|$ is the Euclidean norm. In other words, $G(\cdot)$ depends on the length of the chord between two points on hypersphere. As an example of such kernels we propose the following ones

$$L(x, y) = \|x - y\|^\alpha, \quad 0 < \alpha < 2,$$
$$L(x, y) = \frac{\|x - y\|}{1 + \|x - y\|},$$
$$L(x, y) = \log\left(1 + \|x - y\|^2\right).$$

Note, that considered kernels are rotation-invariant. This property implies that the mathematical expectation of the length of the chord between two independent uniformly distributed random variables Y and Y' on S^{p-1} is equal to the mean length of the chord between a fixed point and a uniformly distributed random variable Y on S^{p-1}. Thus, we can rewrite (2) in the form

$$T_n = n\left[\mathbf{E}\big\{G\big(\|Y - Y'\|\big)\big\} - \frac{1}{n^2}\sum_{i,j=1}^{n} G(\|X_i - X_j\|)\right]. \tag{3}$$

In practice statistics T_n with the kernel $L(x, y) = \|x - y\|^\alpha$, $0 < \alpha < 2$ can be calculated using the following proposition.

Proposition 1. *In cases of $p = 2, 3$ statistics T_n have the form:*

$$T_n = \frac{(2R)^\alpha \Gamma(\frac{\alpha+1}{2})\Gamma(\frac{1}{2})}{\pi\Gamma(\frac{\alpha+2}{2})} n - \frac{1}{n}\sum_{i,j=1}^{n} \|X_i - X_j\|^\alpha \quad (p = 2),$$
$$T_n = (2R)^\alpha \frac{2n}{\alpha + 2} - \frac{1}{n}\sum_{i,j=1}^{n} \|X_i - X_j\|^\alpha \quad (p = 3),$$

where R is the radius of hypersphere and $\alpha \in (0, 2)$.

The proof of the Proposition 1 is presented in Section 5.

In case of $L(x, y) = \|x - y\|$, the test statistic (3) is very similar to Ajne's statistic A, where instead of chord is taken the length of the smaller arc

$$A = \frac{n}{4} - \frac{1}{\pi n} \sum_{i,j=1}^{n} \psi_{ij},$$

where ψ_{ij} is the smaller of two angles between X_i and X_j, $i, j = 1, 2, \ldots, n$.

One can see, that the Ajne's test is not consistent against all alternatives, as an example consider the distribution on the circle concentrated in two diametrically opposite points with equal probabilities. Taking instead of arc the length of the chord lead to a consistency of the N-distance test against all fixed alternatives.

$$\frac{T_n}{n} \xrightarrow{P} N(X, Y), \quad n \to \infty,$$

where $N(X, Y)$ is the N-distance given by (1) between probability distributions of random variables X and Y. If $X \neq^d Y$, then $N(X, Y) > 0$ and $T_n \to \infty$, as $n \to \infty$.

Further we consider the asymptotic distribution of statistics T_n given by (2) under the null hypothesis. Particular attention is devoted to circular and spherical data ($p = 2, 3$). In these cases the asymptotic behavior of proposed tests under the null hypothesis is established using two approaches. First is based on an adaptation of methods of goodness of fit tests described in [1,2], and second using Giné theory based on Sobolev norms [9,10].

For an arbitrary dimension ($p \geq 3$) it is rather difficult from the computational point of view to establish the distribution of test statistics T_n analytically, in this case the critical region of our criteria can be determined with the help of simulations of independent samples from the uniform distribution on S^{p-1}.

3.2 Asymptotic distribution

3.2.1 *Uniformity on the circle S^1*

For our further research, without loss of generality, we consider the circle S^1 with unit length, that is with $R = \frac{1}{2\pi}$. Let us transform the circle, and therefore our initial sample $X_1, \ldots, X_n$, $X_i = (X_{i1}, X_{i2})$, $X_{i1}^2 + X_{i2}^2 = R^2$ to the interval $[0, 1)$ by making a cut in arbitrary point x_0 of the circle

$$x \leftrightarrow x^*, \quad x \in S^1, \;\; x^* \in [0, 1),$$

where x^* is the the length of the smaller arc $x_0 x$. It is easy to see, that if X has a uniform distribution on S^1, after described transformation we will get the random variable X^* with uniform distribution on $[0, 1)$.

Let $L(x, y)$ be a strongly negative definite kernel in $\mathbb{R}^2$, then the function $H(x^*, y^*)$ on $[0, 1)$ defined as

$$H(x^*, y^*) := L(x, y) \tag{4}$$

is a strongly negative definite kernel on $[0,1)$. In this case N-distance statistic T_n^*, based on $H(x^*,y^*)$, for testing the uniformity on $[0,1)$ has the form (see [1,2])

$$T_n^* = -n \int_0^1 \int_0^1 H(x^*,y^*)\,\mathrm{d}\big(F_n(x^*) - x^*\big)\,\mathrm{d}\big(F(y^*) - y^*\big),$$

where $F_n(x^*)$ is the empirical distribution function, based on the sample $X_1^*,\dots,X_n^*$, $X_i^* \in [0,1)$, $i = 1,\dots,n$.

Due to (4) the following equality holds

$$T_n = T_n^*, \tag{5}$$

where T_n is defined by (2).

Thus, instead of testing the initial hypothesis on S^1 using T_n, we can test the uniformity on $[0,1)$ for X^* on the basis of statistics T_n^* with the same asymptotic distribution. The limit distribution of T_n^* is established in Theorem 1 in [1] and leads to the result:

Theorem 1. *Under the null hypothesis, statistics T_n have the same asymptotic distribution as the quadratic form*

$$T = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{a_{kj}}{\pi^2 kj}\zeta_k\zeta_j, \tag{6}$$

where ζ_k are independent random variables from the standard normal distribution and

$$a_{kj} = -2\int_0^1\int_0^1 H(x^*,y^*)\,\mathrm{d}\sin(\pi k x^*)\,\mathrm{d}\sin(\pi j y^*).$$

It is easy to see, that in case $L(x,y)$ is a rotation-invariant function on the circle, the considered transformation of S^1 to $[0,1)$ does not depend on the choice of the point of cut.

Proposition 2. *If strongly negative definite kernel $L(x,y) = \|x-y\|^\alpha$, where $0<\alpha<2$, $x,y \in S^1$, then*

$$H(x^*,y^*) = \left[\frac{\sin \pi d}{\pi}\right]^\alpha,$$

where $d = \min(|x^*-y^*|, 1-|x^*-y^*|)$, $x^*,y^* \in [0,1)$.

The proof of the Proposition 2 is presented in Section 5.

3.2.2 *Uniformity on the sphere S^2*

In case of the sphere we also first try to substitute the initial hypothesis of uniformity on S^2 by testing the uniformity on the unit square. Consider sphere S^2 with unit surface area, that is $R^2 = \frac{1}{4\pi}$.

Note, that if $X^* = (X_1^*, X_2^*)$ has the uniform distribution on $[0,1)^2$ then random variable $X = (X_1, X_2, X_3)$

$$X_1 = R\cos\theta_1, \quad X_2 = R\sin\theta_1\cos\theta_0, \quad X_3 = R\sin\theta_1\sin\theta_0, \tag{7}$$

where

$$\theta_0 = 2\pi X_1^*, \quad \theta_1 = \arccos(1 - 2X_2^*)$$

has the uniform distribution on S^2.

Consider the strongly negative definite kernel $H(x^*, y^*)$ on $[0,1)^2$ defined by

$$H(x^*, y^*) := L(x, y), \tag{8}$$

where $L(x,y)$ is a strongly negative definite kernel in $\mathbb{R}^3$, $x^*, y^* \in [0,1)^2$, $x, y \in S^2$ and the correspondence between x and x^* follows from (7).

N-distance statistics, based on $H(x^*, y^*)$, for testing the uniformity on $[0,1)^2$ has the form (see [1, 2])

$$T_n^* = -n \int\limits_{[0,1)^2}\int\limits_{[0,1)^2} H(x^*, y^*)\,\mathrm{d}\big(F_n(x^*) - x_1^* x_2^*\big)\,\mathrm{d}\big(F(y^*) - y_1^* y_2^*\big),$$

where $F_n(x^*)$, $x^* \in \mathbb{R}^2$ is the empirical distribution function based on the transformed sample X^*.

The equations (7) and (8) implies that

$$T_n = T_n^*. \tag{9}$$

Thus, the asymptotic distribution of T_n coincides with the limit distribution of T_n^*, established in Theorem 2 in [1].

Theorem 2. *Under the null hypothesis statistics T_n will have the same asymptotic distribution as quadratic form*

$$T = \sum_{i,j,k,l=1}^{\infty} a_{ijkl}\sqrt{\alpha_{ij}\alpha_{kl}}\,\zeta_{ij}\zeta_{kl}, \tag{10}$$

where ζ_{ij} are independent random variables from the standard normal distribution,

$$a_{ijkl} = -\int\limits_{[0,1]^4} H(x,y)\,\mathrm{d}\psi_{ij}(x)\,\mathrm{d}\psi_{kl}(y), \quad x, y \in \mathbb{R}^2,$$

α_{ij} *and* $\psi_{ij}(x,y)$ *are eigenvalues and eigenfunctions of the integral operator* A

$$Af(x) = \int\limits_{[0,1]^2} K(x,y)f(y)\,\mathrm{d}y \tag{11}$$

with the kernel

$$K(x,y) = \prod_{i=1}^{2}\min(x_i,y_i) - \prod_{i=1}^{2} x_i y_i.$$

Note, that if $L(x,y)$ is a rotation-invariant function on the sphere then the values of statistics T_n and T_n^* does not depend on the choice of coordinate system on S^2.

The main difficulties in application of the Theorem 2 are connected with calculations of eigenfunctions of the integral operator (11). One of the possible solutions of these problems is in detail discussed in [1]. Another approach is considered in the next subsection, where the asymptotic distribution of proposed statistics for some strongly negative definite kernels is established with the help of Giné theory based on Sobolev tests.

3.2.3 *Alternative approach to limit distribution of* T_n

In this section we propose an application of Giné theory of Sobolev invariant tests for uniformity on compact Riemannian manifolds M to establish the null limit distribution of some N-distance statistics on the circle and sphere. A detailed review of Giné theory can be found in [9, 12].

Let M be the circle $x_1^2 + x_2^2 = 1$ in $\mathbb{R}^2$. Giné showed (see [9]) that in general case Sobolev test statistics $S_n(\{a_k\})$ on M has the form

$$S_n(\{a_k\}) = 2n^{-1}\sum_{k=1}^{\infty} a_k^2 \sum_{i,j=1}^{n} \cos k(X_i - X_j), \tag{12}$$

where $\{a_1, a_2, \ldots\}$ is a sequence of real numbers such that $\sum_{i=1}^{\infty} a_k^2 < \infty$.

The limit null distribution of (12) is established in Theorem 4.1 in [9] and coincides with the distribution of random variable

$$\sum_{k=1}^{\infty} a_k^2 \chi_k,$$

where χ_k are independent random variables with chi-square distribution with two degrees of freedom.

Consider statistics T_n on M with strongly negative definite kernel $L(x,y) = \|x-y\|,\ \ x,y \in \mathbb{R}^2$. From Proposition 1 we have

$$T_n = \frac{4n}{\pi} - \frac{1}{n}\sum_{i,j=1}^{n}\|X_i - X_j\| = \frac{4n}{\pi} - \frac{2}{n}\sum_{i,j=1}^{n}\sin\frac{X_i - X_j}{2}, \tag{13}$$

where $X_i - X_j$ and $\|X_i - X_j\|$ denotes the length of the arc and chord between X_i and X_j respectively.

Under the null hypothesis the limit distribution of T_n is established by the theorem

Theorem 3. *If $X_1, \dots, X_n$ is a sample of independent observations from the uniform distribution on the circle with unit radius, then*

$$\frac{\pi}{4} T_n \xrightarrow{d} \sum_{k=1}^{\infty} a_k^2 \chi_k^2, \tag{14}$$

where χ_k^2 are independent random variables with chi-square distribution with two degrees of freedom and

$$a_k^2 = \frac{1}{2\pi} \int_0^{2\pi} \left(1 - \frac{\pi}{2} \sin \frac{x}{2}\right) \cos kx \, \mathrm{d}x.$$

The proof of Theorem 3 is presented in Section 5.

We now pass over to N-distance and Sobolev tests on the sphere. If $M = S^2$ is the unit sphere $x_1^2 + x_2^2 + x_3^2 = 1$ in $\mathbb{R}^3$, then the general expression of Sobolev test statistic on the sphere has the form (see [9])

$$S_n(\{a_k\}) = n^{-1} \sum_{k=1}^{\infty} (2k+1) a_k^2 \sum_{i,j=1}^{n} P_k\big(\cos(\widehat{X_i, X_j})\big), \tag{15}$$

where $\{a_1, a_2, \dots\}$ is a sequence of real numbers under condition $\sum_{i=1}^{\infty} (2k+1) a_k^2 < \infty$, $\widehat{X_i, X_j}$ is the smaller angle between X_i and X_j, $P_k(\cdot)$ are Legendre polynomials

$$P_k(x) = (k! 2^k)^{-1} \big(\mathrm{d}^k / \mathrm{d}x^k\big) \big(x^2 - 1\big)^k.$$

Under the null hypothesis the limit distribution of $S_n(\{a_k\})$ coincides with the distribution of random variable

$$\sum_{k=1}^{\infty} a_k^2 \chi_{2k+1}^2, \tag{16}$$

where χ_{2k+1}^2 are independent random variables with chi-square distribution with $2k+1$ degrees of freedom.

Consider statistics T_n on S^2 with strongly negative definite kernel $L(x, y) = \|x - y\|$, $x, y \in \mathbb{R}^3$. From Proposition 1 we have

$$T_n = \frac{4n}{3} - \frac{1}{n} \sum_{i,j=1}^{n} \|X_i - X_j\| = \frac{4n}{3} - \frac{2}{n} \sum_{i,j=1}^{n} \sin \frac{\widehat{X_i, X_j}}{2}, \tag{17}$$

where $\widehat{X_i, X_j}$ and $\|X_i - X_j\|$ denotes the smaller angle and the chord between X_i and X_j respectively.

The asymptotic distribution of T_n is established by the Theorem 4.

Theorem 4. *If $X_1, \dots, X_n$ is a sample of independent observations from the uniform distribution on S^2, then*

$$\frac{3}{4}T_n \xrightarrow{d} \sum_{k=1}^{\infty} a_k^2 \chi^2_{2k+1}, \tag{18}$$

where χ^2_{2k+1} are independent random variables with chi-square distribution with $2k+1$ degrees of freedom and

$$a_k^2 = \frac{1}{2}\int_0^{\pi} \left(1 - \frac{3}{2}\sin\frac{x}{2}\right) \sin x P_k(\cos x)\,\mathrm{d}x, \tag{19}$$

where $P_k(x)$ are Legendre polynomials.

The proof of the Theorem 4 is presented in Section 5.

The inverse values to the largest coefficients a_k^2 given by (19) are calculated below:

5	35	105	231	429
715	1105	1615	2261	3059
4025	5175	6525	8091	9889

4 Empirical power results

Let us switch to a comparative Monte Carlo power study of proposed uniformity criteria. N-distance tests with strongly negative definite kernel $L(x, y) = \|x - y\|$ are compared with classical criteria: Rayleigh (R) [3, 4], Watson (W) [5, 6], Giné (G) [3] and Ajne (A) [7, 8] for circular S^1 and spherical S^2 cases.

4.1 Simulation design

In all the cases we investigate the behavior of above mentioned tests for sample sizes $n = 30,\ 50,\ 100$ and significance level $\alpha = 0.05$. All the empirical results were produced by the means of Monte Carlo simulations done with the help of R statistical package. The first part of simulations (Table 1) is devoted to the circular case. In the second part of our study (Table 2) we consider the uniformity test on the sphere S^2. In both cases for N-distance statistics we used the critical values obtained from the asymptotic null distribution established in Theorems 3, 4.

The power of the tests was estimated from a simulation of 200 samples Z of alternative distributions on the circle and sphere, which were modeled using the formulas:

- Circular data

  $$Z = (\cos 2\pi X,\ \sin 2\pi X),$$

  where X is a random variable with the distributions from the first column of Table 1.

- Spherical data

$$Z = \left(\cos(2\pi X),\ \sin(2\pi X)(1-2Y),\ \sin(2\pi X)\sin\left(\arccos(1-2Y)\right)\right),$$

  where X, Y are independent random variables with the distributions from the first column of Table 2.

Proposed alternatives gave us a wide variety of types of departure from null hypothesis and allowed to test the sensitivity of criteria to each of them.

4.2 Simulation results

Empirical results summarized in Tables 1, 2 illustrate that none of the tests are universally superior. In S^1 case proposed N-distance criteria, together with Watson test, showed one of the best results against all considered alternatives for all sample sizes.

The empirical results for spherical data are summarized in Table 2. In comparison with circular case, where all the criteria, except possibly Giné test, showed more or less similar results, the performance of N-distance test was really good for all sample sizes against truncated uniform and von Mises distributions. Giné test, which was not so powerful against considered alternatives in S^1 case, was really sensitive to contamination of hypothesized distribution with truncated uniform in case of spherical data.

Table 1. Empirical power of tests of uniformity on the circle.

$Alternative$	n	W	A	R	G	T_n
$U(0, 0.9)$[1]	30	9	8	8	9	9
$U(0, 0.9)$	50	13	13	13	12	13
$U(0, 0.9)$	100	30	30	28	23	30
$U(0, 0.8)$	30	47	43	42	24	46
$U(0, 0.8)$	50	74	60	57	45	70
$U(0, 0.8)$	100	99	93	91	72	98
$0.9U(0,1) + 0.1U(0, 0.1)$	30	9	8	9	9	9
$0.9U(0,1) + 0.1U(0, 0.1)$	50	15	13	13	15	15
$0.9U(0,1) + 0.1U(0, 0.1)$	100	30	28	23	27	29
$0.8U(0,1) + 0.2U(0, 0.1)$	30	25	20	20	24	24
$0.8U(0,1) + 0.2U(0, 0.1)$	50	54	40	40	47	54
$0.8U(0,1) + 0.2U(0, 0.1)$	100	94	82	74	85	92
$0.8U(0,1) + 0.2U(0, 0.25)$	30	20	19	20	8	20
$0.8U(0,1) + 0.2U(0, 0.25)$	50	44	39	40	20	45
$0.8U(0,1) + 0.2U(0, 0.25)$	100	73	67	66	32	71

[1] $U(a, b)$ is a uniform distirbution on $[a, b]$

$Alternative$	n	W	A	R	G	T_n
$0.8U(0,1) + 0.2U(0,0.5)$	30	11	11	12	5	11
$0.8U(0,1) + 0.2U(0,0.5)$	50	16	15	15	6	17
$0.8U(0,1) + 0.2U(0,0.5)$	100	41	42	42	9	41
$vonMises(0,0.5)$[2]	30	36	36	38	5	37
$vonMises(0,0.5)$	50	58	59	58	7	59
$vonMises(0,0.5)$	100	88	88	88	10	88
$vonMises(0,0.3)$	30	14	15	15	5	15
$vonMises(0,0.3)$	50	27	26	27	7	29
$vonMises(0,0.3)$	100	50	51	50	10	51
$0.5U(0,1) + 0.5vonMises(0,0.5)$	30	15	14	15	5	15
$0.5U(0,1) + 0.5vonMises(0,0.5)$	50	19	19	19	9	21
$0.5U(0,1) + 0.5vonMises(0,0.5)$	100	31	34	33	10	32
$0.5U(0,1) + 0.5vonMises(0,0.8)$	30	19	21	22	5	21
$0.5U(0,1) + 0.5vonMises(0,0.8)$	50	40	40	40	6	42
$0.5U(0,1) + 0.5vonMises(0,0.8)$	100	65	67	67	9	65

Table 2. Empirical power of tests of uniformity on the sphere.

$Alternative$	n	A	R	G	T_n
$U(0,0.9)$	30	16	15	10	18
$U(0,0.9)$	50	23	20	18	24
$U(0,0.9)$	100	53	50	43	58
$U(0,0.8)$	30	61	61	34	69
$U(0,0.8)$	50	86	85	57	93
$U(0,0.8)$	100	99	99	91	100
$vonMises(0,0.5)$	30	27	27	17	31
$vonMises(0,0.5)$	50	38	32	30	42
$vonMises(0,0.5)$	100	83	83	73	90
$vonMises(0,0.3)$	30	13	13	8	14
$vonMises(0,0.3)$	50	14	14	13	15
$vonMises(0,0.3)$	100	39	38	28	44
$0.9U(0,1) + 0.1U(0,0.1)$	30	9	8	7	10
$0.9U(0,1) + 0.1U(0,0.1)$	50	13	12	16	14
$0.9U(0,1) + 0.1U(0,0.1)$	100	35	30	41	36
$0.8U(0,1) + 0.2U(0,0.1)$	30	11	11	11	12
$0.8U(0,1) + 0.2U(0,0.1)$	50	54	41	81	66
$0.8U(0,1) + 0.2U(0,0.1)$	100	96	92	99	99

[2] $vonMises(\mu, \kappa)$ is a von Mises distribution (also known as the circular normal distribution) with location μ and concentration κ parameters

5 Proofs

5.1 The proof of Proposition 1

The stated formulas follow directly from (3) and the property

$$\mathbf{E}\|Y - Y'\|^\alpha = \mathbf{E}\|Y - a\|^\alpha,$$

where Y, Y' are independent random variables from the uniform distribution on S^{p-1} and a is a fixed arbitrary point on S^{p-1}.

In the two-dimensional case, let us calculate the mathematical expectation of the length of the chord between fixed point $a = (0, R)$ and an uniformly distributed random variable Y

$$\begin{aligned}
\mathbf{E}\|a - Y\|^\alpha &= \frac{1}{2\pi R} \int_0^{2\pi} R\big(R^2 \cos^2\phi + \big(R\sin^2\phi - R\big)^2\big)^{\frac{\alpha}{2}} \,\mathrm{d}\phi \\
&= \frac{2^{\frac{\alpha}{2}-1}R^\alpha}{\pi} \int_0^{2\pi} (1-\cos\phi)^{\frac{\alpha}{2}} \,\mathrm{d}\phi = \frac{2^{\alpha+1}R^\alpha}{\pi} \int_0^{\frac{\pi}{2}} \sin^\alpha\phi \,\mathrm{d}\phi \\
&= \frac{(2R)^\alpha \Gamma(\frac{\alpha+1}{2})\Gamma(\frac{1}{2})}{\pi\Gamma(\frac{\alpha+2}{2})}.
\end{aligned}$$

In case $p = 3$ let us fix point $a = (0, 0, R)$ and calculate the average length of the chord

$$\begin{aligned}
&\mathbf{E}\|a - Y\|^\alpha \\
&= \frac{1}{4\pi R^2} \int_{-\pi}^{\pi}\int_0^{\pi} R^2 \sin\theta \big(R^2\big(\sin^2\theta\cos^2\phi + \sin^2\theta\sin^2\phi + (\cos\theta - 1)^2\big)\big)^{\frac{\alpha}{2}} \mathrm{d}\theta\,\mathrm{d}\phi \\
&= \frac{2^{\frac{\alpha}{2}}R^\alpha}{4\pi} \int_{-\pi}^{\pi}\int_0^{\pi} (1-\cos\theta)^{\frac{\alpha}{2}} \sin\theta \,\mathrm{d}\theta\,\mathrm{d}\phi = 2^{\alpha+1}R^\alpha \int_0^{\frac{\pi}{2}} \sin^{\alpha+1}\theta \,\mathrm{d}\sin\theta \\
&= (2R)^\alpha \frac{2}{\alpha+2}.
\end{aligned}$$

5.2 The proof of Proposition 2

Kernel $L(x, y)$ in the circle equals to the length of the chord between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in α power. After proposed transformation, the length of the smaller arc between x and y equals to $d = min(|x^* - y^*|, 1 - |x^* - y^*|)$. The length of the chord in the circle with $R = \frac{1}{2\pi}$ based on the angle $2\pi d$ equals to $\frac{\sin \pi d}{\pi}$, and this completes the proof of the statement.

5.3 The proof of the Theorem 3

Let us express statistics T_n given by (13) in the form

$$T_n = \frac{4}{\pi} n^{-1} \sum_{i,j=1}^{n} h(X_i - X_j),$$

where $h(x) = 1 - \frac{\pi}{2} sin \frac{x}{2}$.

Function $h(x)$ can be represented in the form of a series by complete orthonormal sequence of functions $\{\sqrt{2}\cos kx\}$ on $[0, 2\pi]$

$$h(x) = \sqrt{2} \sum_{k=1}^{\infty} \alpha_k \cos kx,$$

where $\alpha_k = \frac{\sqrt{2}}{2\pi} \int_0^{2\pi} (1 - \frac{\pi}{2} \sin \frac{x}{2}) \cos kx \, \mathrm{d}x$. Note, that $\alpha_k > 0$, $\forall k = 1, 2, \ldots$, really after some simple calculations we have

$$\int_0^{2\pi} \left(1 - \frac{\pi}{2} \sin \frac{x}{2}\right) \cos kx \, \mathrm{d}x = 4 \int_0^{\pi} \sin x \sin^2 kx \, \mathrm{d}x - 4,$$

$$\int_0^{\pi} \sin x \sin^2 kx \, \mathrm{d}x = -k^2 \int_0^{\pi k} \sin \left(\frac{1}{k} - 2\right) x \, \mathrm{d}x - \frac{k^2}{2k+1} \int_0^{\pi k} \sin \frac{x}{k} \, \mathrm{d}x$$
$$= \frac{4k^3}{(2k-1)(2k+1)} > 1 \quad \forall k = 1, 2, \ldots.$$

Thus statistics T_n can be rewritten in the form of Sobolev statistics (12)

$$\frac{4}{\pi} T_n = 2n^{-1} \sum_{k=1}^{\infty} a_k^2 \sum_{i,j=1}^{n} \cos k(X_i - X_j),$$

where $\sqrt{2} a_k^2 = \alpha_k$. After that the statement of the theorem follows directly from Theorem 4.1 in [9].

5.4 The proof of the Theorem 4

The proof of the theorem can be done in nearly the same way as that of Theorem 3. Let us first rewrite statistics T_n in the form

$$T_n = \frac{4}{3} n^{-1} \sum_{i,j=1}^{n} h\big(\widehat{X_i, X_j}\big),$$

where $h(x) = 1 - \frac{3}{2} \sin \frac{x}{2}$. And then decompose $h(x)$ to the series by orthonormal sequence of functions $\{\sqrt{2k+1} P_k(\cos x)\}$ for $x \in [0, \pi]$

$$h(x) = \sum_{k=1}^{\infty} \sqrt{2k+1} \, \alpha_k P_k(\cos x),$$

where

$$\alpha_k = \frac{\sqrt{2k+1}}{4\pi} \int_0^{2\pi} \int_0^{\pi} \left(1 - \frac{3}{2}\sin\frac{\theta}{2}\right) \sin\theta P_k(\cos\theta)\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

As a result statistics T_n can be expressed in the form of Sobolev statistics (15)

$$\frac{4}{3}T_n = n^{-1}\sum_{k=1}^{\infty}(2k+1)a_k^2 \sum_{i,j=1}^{n} P_k\big(\cos\widehat{X_i, X_j}\big),$$

where $\sqrt{2k+1}a_k^2 = \alpha_k$. Applying Theorem 4.1 in [9] the assertion of the theorem follows.

References

1. A. Bakshaev, Goodness of fit and homogeneity tests on the basis of n-distances, *J. Stat. Plan. Infer.*, **139**(11), pp. 3750–3758, 2009.

2. A. Bakshaev, Nonparametric tests based on n-distances, *Lith. Math. J.*, **48**(4), pp. 368–379, 2008.

3. A. Figueiredo, Comparison of tests of uniformity defined on the hypersphere, *Statistics and Probability Letters*, **77**, pp. 329–334, 2007.

4. A. Figueiredo, P. Gomes, Power of tests of uniformity defined on the hypersphere, *Communications in Statistics: Simulation and Computation*, **32**(1), pp. 87–94, 2003.

5. G.S. Watson, Goodness-of-fit tests on the circle, *Biometrika*, **48**, pp. 109–114, 1961.

6. G.S. Watson, Another test for the uniformity of a circular distribution, *Biometrika*, **54**, pp. 675–677, 1967.

7. B. Ajne, A simple test for uniformity of a circular distribution, *Biometrika*, **55**, pp. 343–354, 1968.

8. R.J. Beran, Testing for uniformity on a compact homogeneous space, *J. Appl. Probab.*, **5**, pp. 177–195, 1968.

9. E.M. Gine, Invariant tests for uniformity on compact riemannian manifolds based on sobolev norms, *Ann. Stat.*, **3**, pp. 1243–1266, 1975.

10. M. Hermans, J.P. Rasson, A new sobolev test for uniformity on the circle, *Biometrika*, **72**(3), pp. 698–702, 1985.

11. L.B. Klebanov, *N-distances and their applications*, Karolinum, Prague, 2005.

12. P.E. Jupp, Sobolev tests of goodness of fit of distributions on compact riemannian manifolds, *Ann. Stat.*, **33**(6), pp. 2957–2966, 2005.