A Likelihood Ratio Test for Homogeneity in Circular Data

Juan José Fernández-Durán^{1*} and María Mercedes Gregorio-Domínguez²

¹Department of Statistics, División Académica de Actuaría, Estadística y Matemáticas, Instituto Tecnológico Autónomo de México ²Department of Actuarial Science, División Académica de Actuaría, Estadística y Matemáticas, Instituto Tecnológico Autónomo de México

Abstract

Testing for the homogeneity of density functions of circular random variables is useful in many settings including the study of wind patterns, paleocurrents trends, the seasonality in human-related events such as homicides and suicides and the seasonality in the appearance of diseases. In this paper, we considered that the density functions are members of the flexible family of circular distributions based on nonnegative trigonometric (Fourier) sums (series) developed by Fernandez-Duran [1]. We constructed a test based on the likelihood ratio and we applied the proposed test to simulated and real datasets.

Keywords: Circular data; Nonnegative fourier series

Introduction

If Θ is a circular random variable, $\Theta \in (0,2\pi)$, then the circular density function based on nonnegative trigonometric sums (the NNTS density), developed by Fernandez-Duran [1] and based on Fejer [2], is expressed as

$$f(\theta; M, \underline{c}) = \left\| \sum_{k=0}^{M} c_k e^{ik\theta} \right\|^2 = \sum_{k=0}^{M} \sum_{l=0}^{M} c_k \overline{c_l} e^{i(k-l)\theta}$$
(1)

where $i = \sqrt{-1}$, C_k are the complex numbers $C_k = C_{rk} + Ic_{ck}$ for k = 0, ..., M,

and $\overline{c}_k = c_{rk} - ic_{ck}$ is the conjugate of C_k . The following constraint in the *c* parameter is imposed to $f(\theta; M, \underline{c})$ to integrate one:

$$\sum_{k=0}^{M} ||c_k||^2 = \frac{1}{2\pi}$$
(2)

With $c_{c0} = 0$ and $c_{r0} \ge 0$, i.e., c_0 is a nonnegative real number. The total number of *c* free parameters is equal to 2*M*. Also, *M* is the total number of terms in the sum that defines the NNTS density, which is equal to the maximum number of the modes of the density and is an additional unknown parameter.

Equivalently, the NNTS density can be expressed as

$$f(\theta; M, \underline{c}) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{M} \left(a_k \cos(k\theta) + b_k \sin(k\theta) \right)$$
(3)

where $a_k - ib_k = 2\sum_{\nu=0}^{M-k} c_{\nu+k}\overline{c_{\nu}}$ for k = 1, 2, ..., M. Note that for the NNTS family, the *k*-th trigonometric moment is equal to $E(e^{ik\theta}) = a_k + ib_k$. The case M = 0 corresponds to a uniform circular density on $(0, 2\pi)$. The NNTS family of circular distributions is very flexible to model datasets that present multimodality and/or skewness.

The accumulated distribution function of an NNTS density is easily calculated as:

$$F(\theta; M, \underline{c}) = \int_{0}^{\theta} f(s; M, \underline{c}) ds = \frac{\theta}{2\pi} + \sum_{k=1}^{M} \frac{a_{k}}{k} \sin(k\theta) + \frac{b_{k}}{k} (1 - \cos(k\theta)).$$
(4)

The original support of the NNTS circular distribution is the interval $(0,2\pi)$, but in applications that are related to the occurrence of events over time, it is more common to use as support the interval (0,1). In that case, we transform the variable Θ to the variable *T* by $T = \frac{\Theta}{2\pi}$ with the density function

$$f_{T}(t; M, \underline{c}) = 1 + 2\sum_{k=1}^{M} a_{k} \cos(2\pi kt) + b_{k} \sin(2\pi kt)$$
(5)
and the distribution function

 $F_{T}(t; M, \underline{c}) = t + \frac{1}{\pi} \sum_{k=1}^{M} \frac{a_{k}}{k} \sin(2\pi kt) + \frac{b_{k}}{k} (1 - \cos(2\pi kt))$ (6) For $t \in (0, 1)$. In the case that M = 0 the NNTS distribution is

For $t \in (0,1)$. In the case that M = 0 the NNTS distribution is equivalent to a uniform distribution and $f_T(t;M=1) = 1$ and $f_T(t;M=0) = t$ for $t \in (0,1)$.

For the case of grouped data $n_1, n_2, ..., n_R$ where n_i is the total number of the occurrences of the event of interest in the *i*-th interval, $(R_{i,i}, R_i)$ in which (0,1) has been partitioned, i = 0, 1, ..., R, we consider that there are R such intervals. For example, in the case of using a year as the unit of time and months as the intervals, $R_{i,1}$ and R_i are the beginning and end, respectively, of the *i*-th month in yearly terms and R = 12. In the case of grouped data, the likelihood function, $L(M, \underline{c} | n_1, n_2, ..., n_R)$, is given by

$$L(M, \underline{c}|n_1, n_2, ..., n_R) = \prod_{k=1}^{m} \left(F_T(R_k) - F_T(R_{k-1}) \right)^{n_k}$$
(7)
$$\left(P_{k-1} - \frac{1}{2} \sum_{k=1}^{M} a_k \left(\sin(2\pi k P_k) - \sin(2\pi k P_k) \right) \right)^{n_k}$$

$$=\prod_{k=1}^{R} \left(\frac{R_{k} - R_{k-1} + \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{a_{k}}{k} (\sin(2\pi kR_{k}) - \sin(2\pi kR_{k-1})) + \frac{b_{k}}{k} (\cos(2\pi kR_{k-1}) - \cos(2\pi kR_{k})) \right).$$

Additional properties of the NNTS circular models are presented by Fernandez-Duran [1,3]. The maximum likelihood estimates of the parameters of the NNTS model are obtained using an efficient Newtonlike algorithm that was developed by Fernandez-Duran [4] and implemented in the statistical software R [5] in the library CircNNTSR [6,7].

Previous work on homogeneity tests for circular data includes that of Mardia and Spurr [8] for *l*-modal and axial von Mises distributions with modes separated by a constant arc. In addition, Mardia [9] developed tests against shift-type alternatives, based on uniform scores as an extension of Wheeler and Watson [10]. In particular, Rao's test of homogeneity [11,12] tests for the equality of the mean directions and dispersions of *N* circular populations by considering the means of the

*Corresponding author: Fernandez-Duran JJ, Department of Statistics, División Académica de Actuaría, Estadística y Matemáticas, Instituto Tecnológico Autónomo de México, Rio Hondo No. 1, Col. Progreso Tizapan, C.P.01080, México D.F., México, Tel: (5255) 56 28 38 39; Fax: (5255) 56 28 40 86; E-mail: jfdez@itam.mx

Received October 14, 2010; Accepted November 22, 2010; Published November 24, 2010

Citation: Fernandez-Duran JJ, Gregorio-Dominguez MM (2010) A Likelihood Ratio Test for Homogeneity in Circular Data. J Biomet Biostat 1:107. doi:10.4172/2155-6180.1000107

Copyright: © 2010 Fernandez-Duran JJ, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

cosine and sine values, $\overline{C}_j = n_j^{-1} \sum_{k=1}^{n_j} \cos(\theta_{kj})$ and $\overline{S}_j = n_j^{-1} \sum_{k=1}^{n_j} \sin(\theta_{kj})$

for j = 1,...,N, by comparing the values of $T_j = \frac{\overline{s}_j}{\overline{c}_j}$ for testing the equality of the mean directions, and $U_j = \overline{c}_j^2 + \overline{s}_j^2$ for testing the equality of the dispersions [13]. Watson's two-sample test of homogeneity uses the squared differences between the sample distribution functions as test statistic [14]. Tests for the equality of the mean directions and dispersion parameters of von Mises populations that were developed by Mardia [9], Watson and Williams [15], Stephens [16], Watson [17], and Fisher [18] are similar to the homogeneity tests. Similarly, Harrison et al. [19,20], Stephens [21,22] have developed two-way and multi-way ANOVAs for circular data. Recently, Grimshaw et al. [23], Chen et al. [24], and Fu et al. [25] developed homogeneity tests for mixtures of von Mises distributions using likelihood ratio tests. In contrast to previous work that used von Mises or mixtures of von Mises densities as the population densities, the main objective of the present paper was to develop a likelihood ratio test for the homogeneity of circular populations by considering their density functions as members of the flexible NNTS family for ungrouped and grouped observations.

This paper is divided into five sections, including this introduction. The second section outlines the development of the likelihood ratio test for homogeneity. In the third section, the proposed test is applied to simulated and real datasets. The fourth section includes a simulation exercise to study the type I error and power of the proposed test. Lastly, in the fifth section, the conclusions of the present work are presented.

Likelihood ratio homogeneity test

be independent random samples from N different continuous circular populations, and let $F_k(\theta)$ and $f_k(\theta)$ be their respective distribution (density) functions. A test of homogeneity has a null hypothesis given as $H_0: F_1 = F_2 = \ldots = F_N$ ($H_0: f_1 = f_2 = \ldots = f_N$) and an alternative hypothesis given as $H_a: F_1, \ldots, F_N$ are not all the same ($H_a: f_1, \ldots, f_N$ are not all the same). By considering that for $k = 1, \ldots, N, F_k(f_k)$ is a member of the NNTS family of the circular distributions with parameters M_k and c_k , a test for homogeneity can be constructed by considering the likelihood ratio,. This ratio is defined as

Page 2 of 9

$$\Lambda = \frac{\max\{L_0(M_0, \underline{c}_0|\underline{\theta}_1, \dots, \underline{\theta}_N)\}}{\max\{L(M_0, \underline{c}|\underline{\theta}_1, \dots, \underline{\theta}_N)\}} = \frac{L(M_0, \underline{c}_0|\underline{\theta}_1, \dots, \underline{\theta}_N)}{L(M_0, \underline{c}|\underline{\theta}_1, \dots, \underline{\theta}_N)}$$
(8)

where $L(M_0, \underline{c}_0 | \underline{\theta}_1, ..., \underline{\theta}_N)$ is the maximum of the likelihood function over the restricted parameter space specified by H_0 , and \underline{c}_0 is the maximum likelihood estimate under the null hypothesis. Note that we are considering that $M_k = M_0$ for k = 1, ..., N and that the value of M_0 is specified before applying the test. The researcher can select a value for M_0 by considering the maximum number of modes in all of the considered populations (taking into account the sample sizes in each population) because the NNTS models are nested with respect to the increasing values of the parameter M. Now, $L(M_0, \underline{c} | \underline{\theta}_1, ..., \underline{\theta}_N)$ is the maximum of the likelihood function over the complete parameter space, and $\underline{c} = (\underline{c}_1, ..., \underline{c}_N)$ are the unrestricted maximum likelihood estimates that corresponds to the unrestricted maximum likelihood estimate for the parameter \underline{c} in each of the N considered circular populations. The likelihood function for the ungrouped data under the null hypothesis is calculated as

$$L_0(M_0, \underline{c}_0 | \underline{\theta}_1, \dots, \underline{\theta}_N) = \prod_{k=1}^N \prod_{i=1}^{n_k} f(\theta_{ik} | M_0, \underline{c}_0)$$
(9)

Ungrouped observation: Let $\underline{\theta}_k = (\theta_{k1}, \theta_{k2}, ..., \theta_{kn_k})$ for k = 1, ..., N

where $f(\theta_{ik}|M_0, \underline{c}_0)$ is the common NNTS circular density function

M	Population 1	Population 2	Populations 1 and 2	-2ln(Λ)	d.f.	χ²p-value
0	-91.8939	-91.8939	-183.7877			
1	-72.8523	-71.1008	-144.1506	0.3949	2	0.8208
2	-68.7839	-66.1711	-135.5405	1.1710	4	0.8829
3	-68.2188	-65.0110	-134.3438	2.2281	6	0.8976
4	-65.6811	-64.7347	-133.2813	5.7310	8	0.6773
5	-64.6047	-64.6029	-132.6212	6.8271	10	0.7417
6	-63.9279	-64.4850	-131.9476	7.0695	12	0.8530

Table 1: Results of the NNTS likelihood ratio test for homogeneity when applied to two samples of 50 data points from a von Mises population with μ = 3.1962 and κ = 1.5510.

M ₀	Pop. 1	Pop. 2	Pop. 3	Pop. 4	Pop. 5	Pop. 6	Pop. 7	Pops. 1 to 7	$-2\ln(\Lambda)$	d.f.	χ²p-value
0	-91.8939	-91.8939	-91.8939	-91.8939	-91.8939	-91.8939	-91.8939	-643.2570			
1	-90.7435	-90.8332	-91.8541	-91.8634	-91.2191	-91.6372	-91.8033	-642.4010	4.8942	12	0.9614
2	-89.6939	-89.4777	-91.3742	-89.0410	-91.2099	-91.2177	-88.7269	-641.9138	22.3450	24	0.5587
3	-88.4427	-89.3987	-89.9833	-88.9220	-90.0088	-88.3413	-86.7960	-640.5025	37.2196	36	0.4127
4	-86.8474	-88.8507	-89.9685	-86.4008	-89.3175	-87.7075	-86.7055	-639.4287	47.2615	48	0.5030
5	-84.6353	-88.3876	-86.1472	-85.5370	-87.7413	-87.0995	-86.5498	-637.6810	63.1670	60	0.3651
6	-84.1664	-83.5998	-85.8473	-85.5001	-87.7211	-85.1840	-83.3282	-636.9619	83.2302	72	0.1720

Table 2: Results of the NNTS likelihood ratio test for homogeneity when applied to seven samples of 50 data points from a circular uniform distribution.

Mo	1951-1970	1971-2008	1951-2008	-2ln(Λ)	d.f.	χ² p-value
0	-360.2239	-766.3947	-1126.6190			
1	-250.9487	-530.7715	-781.7269	0.0134	2	0.9933
2	-213.7020	-456.5898	-671.3008	2.0180	4	0.7324
3	-203.8207	-431.7584	-636.8067	2.4552	6	0.8734
4	-196.9491	-424.7851	-626.3378	9.2072	8	0.3251
5	-191.2163	-419.8673	-615.7329	9.2986	10	0.5040
6	-191.1102	-415.6410	-612.2737	11.0450	12	0.5251
7	-190.5867	-415.3392	-611.7719	11.6920	14	0.6310
8	-189.7132	-414.9694	-611.6512	13.9372	16	0.6034
9	-189.3480	-413.3788	-610.6316	15.8096	18	0.6058
10	-188.1896	-413.3687	-610.3693	17.6220	20	0.6123

Table 3: Results of the NNTS homogeneity test for the hurricane data.

Season													
Winter	50	120	190	210	220	250	260	290	290	320	320	340	
Spring	0	20	40	60	160	170	200	220	270	290	340	350	
Summer	10	10	20	20	30	30	40	150	150	150	170	190	290
Autumn	30	70	110	170	180	100	240	250	260	260	200	350	

Table 4: Wind directions in degrees at Gorleston, England [26].

for all of the observations in all of the considered populations. The likelihood under the complete parameter space corresponds to

$$L(M_0, \underline{c} | \underline{\theta}_1, \dots, \underline{\theta}_N) = \prod_{k=1}^N L_k(M_0, \underline{c}_k | \underline{\theta}_k) = \prod_{k=1}^N \prod_{i=1}^n f(\theta_{ik} | M_0, \underline{c}_k)$$
(10)

where $f(\theta_{ik} | M_0, \underline{c}_k)$ corresponds to the NNTS circular density of the *k*-th population, with parameters M_0 and \underline{c}_k . Maximizing $L(M_0, \underline{c} | \underline{\theta}_1, ..., \underline{\theta}_N)$ is equivalent to maximizing each of the $L_k(\underline{\theta}_k | M_0, \underline{c}_k)$ likelihood functions for each population, with respect to the corresponding \underline{c}_k parameters for k = 1, ..., N.

By the likelihood asymptotic theory, $-2\ln(\Lambda)$ converges to a chi-squared distributed random variable with a number of degrees of freedom that is equal to the difference in the number of the free parameters in the unrestricted and restricted parameter spaces.

Grouped observations: In the case that the observed data consists of grouped observations, i.e., only the total number of occurrences of the circular random variables in the different intervals is observed, then, the likelihood function for the null hypothesis of homogeneity is calculated as

$$L(M_0, \underline{c}_0 | \underline{n}_1, \dots, \underline{n}_N) = \prod_{k=1}^N \prod_{i=1}^{I_k} \begin{pmatrix} F(R_{ki} | M_0, \underline{c}_0) - \\ F(R_{k,i-1} | M_0, \underline{c}_0) \end{pmatrix}^{n_{ki}}$$
(11)

where n_{ki} is the total number of observations of population k in the *i*-th interval in population k, $(R_{k,i-1}, R_{ki})$. The total number of grouping intervals for population k is denoted by I_k . Note that the grouping intervals may be different for each population. The likelihood function for the whole parameter space is given by

$$L(M_0, \underline{c} | \underline{n}_1, \dots, \underline{n}_N) = \prod_{k=1}^N L_k(M_0, \underline{c}_k | \underline{n}_k)$$
$$= \prod_{k=1}^N \prod_{i=1}^{I_k} \left(F(R_{ki} | M_0, \underline{c}_k) - F(R_{k,i-1} | M_0, \underline{c}_k) \right)^{n_k}$$
(12)

In practice, the grouping intervals are commonly the same for all of the observed populations.

Examples

Ungrouped simulated data

Simulated von mises data: A random sample of 100 observations from a von Mises density, with a location parameter μ =3.1962 and a dispersion parameter =1.5510, was generated. We divided the dataset in two. The first dataset included the first 50 observations, and the second included the last 50 observations. The results of applying the NNTS

homogeneity likelihood ratio test to these datasets are included in Table 1. The NNTS models of order $M_0 = 0,1,...,6$ were fitted to each dataset of 50 observations (Population 1 and Population 2) and to the complete dataset of 100 observations (Populations 1 and 2). The columns in Table 1 contain the values of the maximized log-likelihood, the observed value of the likelihood ratio statistic $(-2\ln(\Lambda))$, the number of the degrees of freedom $(d_f.)$, and the corresponding p-values when comparing the observed test statistic with the distribution of a chi-squared random variable. For $M_0 = 1,2,...,6$, the observed values of the NNTS likelihood ratio statistic resulted, as expected, in the decision of not rejecting the null hypothesis of homogeneity.

Simulated circular uniform data: We generated seven random samples of 50 data points from a circular uniform distribution. Table 2 contains the results of the NNTS likelihood ratio test. For all of the considered values of M_0 , the null hypothesis of homogeneity was not rejected.

Ungrouped real data

Hurricane occurrence data: Recently, there has been debate about whether an increase exists in the number of tropical storms and hurricanes that occur yearly in the North Atlantic Ocean. Instead of analyzing the possibility of an increase in the number of tropical storms and hurricanes, we analyzed whether there is evidence of a change in the start date of the tropical storms and hurricanes. The starting dates of the tropical storms and hurricanes were obtained from http://weather. unisys.com/archive/index.html. This website includes tropical storms and hurricanes that occurred between 1851 and 2008 and is based on the HURDAT database of the National Oceanic and Atmospheric Administration (NOAA). The dates were converted to numbers between 0 and 1 to represent the fraction of the year that elapsed at the starting date of the tropical storm or hurricane. These values were then multiplied by 2π to convert them to circular data. Then, we compared the starting dates of the storms that occurred between 1951 and 1970 to those that occurred between 1971 and 2008. Table 3 contains the results of the NNTS homogeneity test. For all of the considered values of M_{o} , we did not reject the null hypothesis of homogeneity.

Wind direction data: Mardia and Jupp [26] analyzed a dataset of the wind directions in degrees at Gorleston, England, between 11 a.m. and 12 p.m. on Sundays in 1968, as classified by seasons. The data is shown in Table 4.

The authors applied a homogeneity test to the data based on the ranks of the angles and the uniformity test of Rayleigh, known as the uniform scores test [27], to this dataset. They concluded that the null hypothesis was rejected at the 5% level (p-value=0.046). We applied the NNTS homogeneity test to this dataset, and the results are shown in Table 5.

M	Winter	Spring	Summer	Autumn	Combined	-2ln(Λ)	d.f.	χ^2 p-value
0	-22.0545	-22.0545	-23.8924	-22.0545	-90.0560			
1	-19.9543	-21.6284	-22.1090	-21.3295	-89.1455	8.2486	6	0.2205
2	-19.5298	-19.5591	-18.7057	-20.6901	-86.5384	16.1075	12	0.1864
3	-18.9205	-19.3480	-15.0076	-20.5220	-83.9955	20.3948	18	0.3110
4	-17.3523	-18.3262	-14.5950	-18.0935	-83.5365	30.3392	24	0.1738
5	-16.9660	-17.7181	-11.9490	-17.0136	-83.0619	38.8305	30	0.1296
6	-16.9440	-16.5327	-11.7001	-16.5227	-82.8683	42.3377	36	0.2163

M _o	set 1	set2	combined set (1 and 2)	-2ln(Λ)	d.f.	χ ² p-value
0	-45.9469	-191.1392	-237.0861			
1	-41.4898	-176.0237	-218.5734	2.1199	2	0.3465
2	-38.9377	-169.1353	-215.9945	15.8430	4	0.0032
3	-38.3024	-168.9135	-215.8106	17.1893	6	0.0086
4	-37.0521	-168.4393	-215.4071	19.8314	8	0.0110
5	-36.9842	-168.2582	-215.3776	20.2705	10	0.0268
6	-36.3596	-168.2576	-215.3653	21.4962	12	0.0436

Table 6: NNTS homogeneity tests results for the cross-bed measurements from the Himalayan molasse dataset [29].

Citation: Fernandez-Duran JJ, Gregorio-Dominguez MM (2010) A Likelihood Ratio Test for Homogeneity in Circular Data. J Biomet Biostat 1:107. doi:10.4172/2155-6180.1000107

Page 4 of 9

	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Suicides	2005	318	335	370	360	428	367	366	386	390	342	331	320
	%	7.37	7.77	8.58	8.35	9.92	8.51	8.49	8.95	9.04	7.93	7.67	7.42
	2007	284	303	416	391	413	392	383	383	350	367	354	355
	%	6.47	6.90	9.47	8.90	9.41	8.93	8.72	8.72	7.97	8.36	8.06	8.08
Homicides	2005	776	776	819	859	822	889	817	792	837	814	825	881
	%	7.83	7.83	8.27	8.67	8.30	8.97	8.25	7.99	8.45	8.22	8.33	8.89
	2007	595	508	781	775	935	757	740	764	713	768	708	797
	%	6.73	5.75	8.83	8.77	10.58	8.56	8.37	8.64	8.06	8.69	8.01	9.01

ble 7: Suicides and homicides in Mexico for the years of 2005 and 2007.

Mo	Suicides 2005	Homicides 2005	Suicides and Homicides 2005	-2ln(Λ)	d.f.	χ² p-value
0	-10717.05	-24620.57	-35337.62			
1	-10707.97	-24620.20	-35333.37	10.40	2	0.0055
2	-10706.07	-24618.29	-35331.64	14.56	4	0.0057
3	-10705.13	-24618.17	-35331.44	16.28	6	0.0123
4	-10702.36	-24614.96	-35327.69	20.74	8	0.0079
5	-10701.89	-24611.23	-35324.31	22.38	10	0.0133
6	-10701.56	-24609.14	-35323.52	25.64	12	0.0121

Table 8: NNTS homogeneity test results: Suicides and homicides in Mexico in 2005.

M _o	Suicides 2007	Homicides 2007	Suicides and Homicides 2007	-2ln(Λ)	d.f.	χ ² p-value
0	-10907.41	-21949.65	-32857.06			
1	-10896.88	-21931.98	-32829.40	1.08	2	0.5827
2	-10891.65	-21903.18	-32798.34	7.02	4	0.1348
3	-10888.55	-21895.23	-32790.39	13.22	6	0.0397
4	-10886.79	-21892.36	-32785.85	13.40	8	0.0988
5	-10886.16	-21883.11	-32777.17	15.80	10	0.1055
6	-10886.10	-21882.93	-32777.14	16.22	12	0.1814

Table 9: NNTS homogeneity test results: Suicides and homicides in Mexico in 2007.

In contrast to the uniform score test results of Mardia and Jupp [26], we did not reject the null hypothesis of homogeneity for the considered values of M_0 . This contrast may be due to the small sample sizes. Mardia and Jupp [26] also analyzed this dataset using the von Mises model and concluded that it was not possible to reject the null hypotheses of equal mean directions and equal concentration parameters. Fisher [28] also analyzed this dataset. By applying a uniform score test for the Winter, Spring, and Autumn data, Fisher concluded that there was little evidence that the wind distributions in these three seasons differed.

Himalayan molasse data: Fisher [28] analyzed two samples of cross-bed measurements that were collected from the Himalayan molasse in Pakistan [29]. The first sample consisted of 35 measurements of the Rakhi Nala ripple cross-beds, and the second sample consisted of 104 measurements of the Chaudhwan Zam large bedforms. Fisher suggested that the samples were drawn from von Mises distributions with different shapes but not necessarily with different mean directions and applied a bootstrap method test for different means. Fisher did not find evidence that the mean directions differed. The results of the NNTS homogeneity test are shown in Table 6. For the selected M_0 values equal to 2 and 3, the null hypothesis of homogeneity was rejected at the 1% significance level. For $M_0 = 4,5$ and 6, the null hypothesis was rejected at the 5% significance level, and for $M_0 = 1$, the null hypothesis was not rejected because the p-value was equal to 0.3465. This example illustrates the necessity to select a sensible value for M_0 before applying the test.

Grouped real data

Suicides and homicides data: The Mexican Statistical Agency (INEGI) reports the monthly number of suicides and homicides in Mexico every year. To compare the monthly patterns of the number of suicides and homicides, we applied the NNTS homogeneity test for the monthly numbers of suicides and homicides for the years of 2005 and 2007. Table 7 shows the monthly grouped raw data.

Table 8 shows the results of applying the NNTS homogeneity test to the number of suicides and homicides in Mexico in 2005. The null hypothesis of homogeneity was rejected at the 5% significance level when using $M_0 = 1,2,3,4,5$ and 6. Note that the case $M_0 = 6$ corresponds to the saturated model.

For the data from the year 2007, Table 9 shows the results of the NNTS homogeneity test where, for $M_0 = 1,2,4$, and 6 the test did not reject the null hypothesis of homogeneity. Only for $M_0 = 3$ was the null hypothesis of homogeneity rejected at the 5% significance level. The change in the seasonal patterns of suicides and homicides from 2005 to 2007 may be the result of the mis-reporting, over-reporting, or underreporting of these crimes.

Power and type I error of the NNTS test of homogeneity

To study the power of the NNTS test of homogeneity and its type I error rate, we simulated data from three different models: a uniform



Figure 1: Considered models in the simulation study (Uniform (solid), von Mises (dot dash) and NNTS with M = 4 (dash)).

Citation: Fernandez-Duran JJ, Gregorio-Dominguez MM (2010) A Likelihood Ratio Test for Homogeneity in Circular Data. J Biomet Biostat 1:107. doi:10.4172/2155-6180.1000107

Page 5 of 9

									NNTS	test							
Model 1	n,	Model 2	n,	α	$M_0 = 1$	2	3	4	5	6	7	8	9	10	Watson	Rao	Rao
Uniform	20	von Mises	20	0.01	100	100	100		-	-		-	-		100	Mean 25	Disp.
Onnorm	20	Von Mises	20	0.01	100	100	100								100	51	100
				0.1	100	100	100								100	59	100
			50	0.01	100	100	100								100	60	100
				0.03	100	100	100								100	91	100
			100	0.01	100	100	100								100	88	100
				0.05	100	100	100								100	94	100
		NNTS	20	0.1	100	100	100								100	95 56	79
				0.05	100	100	100								100	75	95
				0.1	100	100	100								100	81	97
			50	0.01	100	100	100								100	01	100
				0.03	100	100	100								100	94	100
			100	0.01	100	100	100								100	96	100
				0.05	100	100	100								100	99	100
Uniform	50	von Misos	20	0.1	100	100	100								100	20	100
Unitoni		V011 1VIISES	20	0.01	100	100	100								100	50	100
				0.1	100	100	100								100	62	100
			50	0.01	100	100	100	100	100	100					100	67	100
				0.05	100	100	100	100	100	100					100	89	100
			100	0.01	100	100	100	100	100	100					100	91	100
				0.05	100	100	100	100	100	100					100	98	100
		NNTS	20	0.1	100	100	100	100	100	100					100	100	100
		ININIS	20	0.01	100	100	100								100	76	96
				0.1	100	100	100								100	83	98
			50	0.01	100	100	100	100	100	100					100	81	100
				0.05	100	100	100	100	100	100					100	93	100
			100	0.01	100	100	100	100	100	100					100	100	100
				0.05	100	100	100	100	100	100					100	100	100
Liniform	100	Van Misso	20	0.1	100	100	100	100	100	100					100	100	100
Unitorni	100	von mises	20	0.01	100	100	100								100	54	100
				0.00	100	100	100								100	63	100
			50	0.01	100	100	100	100	100	100					100	73	100
				0.05	100	100	100	100	100	100					100	85	100
			100	0.01	100	100	100	100	100	100	100	100	100	100	100	95	100
				0.05	100	100	100	100	100	100	100	100	100	100	100	97	100
		NNTO	00	0.1	100	100	100	100	100	100	100	100	100	100	100	100	100
		NNIS	20	0.01	100	100	100								100	63 77	/8 95
				0.00	100	100	100								100	82	98
			50	0.01	100	100	100	100	100	100					100	87	100
				0.05	100	100	100	100	100	100					100	95	100
			100	0.1	100	100	100	100	100	100	100	100	100	100	100	90	100
				0.05	100	100	100	100	100	100	100	100	100	100	100	100	100
				0.1	100	100	100	100	100	100	100	100	100	100	100	100	100
von Mises	20	NNTS	20	0.01	0	2	<u> </u>								12	<u>0</u> 3	6
			<u> </u>	0.1	7	4	10								23	9	13
			50	0.01	1	0	3								5	0	1
				0.05	3	1	10								12	4	9
			100	0.01	0	0	2								3	1	2
				0.05	1	2	15								15	4	5
	=		00	0.1	7	7	29								20	5	17
von Mises	50	NNTS	20	0.01	0	1 6	16								2	<u> </u>	10
				0.1	18	12	27					1			20	13	15
			50	0.01	3	2	11	18	14	12					5	0	1
				0.05	10	14	29	34	27	30					23	5	6
			100	0.1	3	6	24	32	38	17					20 12	10	5
				0.05	11	27	45	59	55	48					28	5	7
				0.1	24	33	64	68	66	59					41	12	14
von Mises	100	NNTS	20	0.01	11	1 7	3								4	0	6
				0.05	19	12	25								22	8	13
			50	0.01	4	8	17	19	17	19					7	2	1
				0.05	17	19	41	47	38	30					26	5	5
			100	0.1	28	30	49	59	53	49	20	27	24	10	32	13	10
			100	0.01	20	30	59	69	4 2 67	65	60	57	47	46	29	5	6
				0.1	28	40	72	81	79	74	68	66	65	59	42	12	12

Table 10: NNTS homogeneity test: Statistical power.

Page 6 of 9

							NN	ITS test	1							
Model	n,	n,	α	$M_{0} = 1$	2	3	4	5	6	7	8	9	10	Watson	Rao Mean	Rao Disp.
Uniform	20	20	0.01	0	0	0								3	1	2
			0.05	0	0	0								6	3	9
			0.1	0	0	0								10	8	13
	20	50	0.01	0	0	0								2	2	3
			0.05	0	0	0								4	5	9
			0.1	0	0	0								10	11	14
	20	100	0.01	0	0	0								2	0	5
			0.05	0	0	0								6	3	10
	50	50	0.1	0	0	0	0	0	0					11	8	14
	50	50	0.01	0	0	0	0	0	0					0	7	0
			0.05	0	0	0	0	0	0					4	1/	0
	50	100	0.1	0	0	0	0	0	0					1	2	2
	00	100	0.05	0	0	0	0	0	0					2	5	4
			0.1	0	0	0	0	0	0					7	8	9
	100	100	0.01	0	0	0	0	0	0	0	0	0	0	1	1	1
			0.05	0	0	0	0	0	0	0	0	0	0	4	8	4
			0.1	0	0	0	0	0	0	0	0	0	0	10	13	10
von Mises	20	20	0.01	0	1	0								1	2	0
			0.05	2	4	1								8	5	7
			0.1	2	6	5								11	7	13
	20	50	0.01	0	1	0								1	2	0
			0.05	4	2	2								3	5	5
	20	100	0.1	9	5	5								9	6	<u>11</u>
	20	100	0.01	1		0								2	2	2
			0.05	2	2	3								11	5	0
	50	50	0.1	0	0		0	0	1					0	0	5
	- 50	- 50	0.01	1	3	3	5	3	2					7	2	8
			0.1	6	3	9	8	5	9					13	5	9
	50	100	0.01	2	3	1	1	1	0					1	1	2
			0.05	6	8	9	3	5	4					9	4	7
			0.1	11	10	13	14	14	10					13	10	14
	100	100	0.01	1	0	0	1	0	0	0	1	4	4	2	0	1
			0.05	6	7	5	3	4	4	3	5	6	9	8	5	3
			0.1	11	13	14	11	11	7	10	8	10	10	12	11	11
NNIS	20	20	0.01	1	0	1									0	2
			0.05	b 12	3	2								0	b 11	8 15
	20	50	0.1	1	0	0								9		10
	20	- 50	0.01	6	1	7								8	1	12
			0.05	7	5	11								10	8	16
	20	100	0.01	0	0	0								0	0	5
			0.05	3	1	1								2	2	9
			0.1	7	3	5								8	12	17
	50	50	0.01	4	0	0	1	1	1					3	0	3
			0.05	9	6	5	7	4	2					7	6	11
			0.1	17	14	14	12	10	5					14	14	15
	50	100	0.01	3	0	0	0	0	1					0	0	1
			0.05	5	6	8	7	6	2					6	7	10
	100	400	0.1	10	16	11	12	8	9			_		11	13	13
	100	100	0.01	1	0	1	0	2	1	3	3	1	3		0	0
			0.05	4	2	3	4		5	5	8	4	5	10	2	8
	1		0.1	TI	0	1	9	14	12	15	14	12		12	ŏ	15

Table 11: NNTS homogeneity test: Type I errors.

circular distribution, a von Mises distribution with μ = 3.1962, and κ = 1.5510, and NNTS density with *M* = 4. Figure 1 presents the graphs for the considered models. We generated 100 simulated datasets of sample sizes equal to 20, 50, and 100 for each of the three models. When comparing the simulated datasets from the different models, we studied the power of the test (the probability that the test will reject a false null hypothesis of homogeneity), and when comparing the simulated datasets for the test (the probability that the type I error of the test (the probability that the test will reject a true null hypothesis of homogeneity).

Table 10 contains the values of the power of the test that were obtained from simulating 100 different datasets for the considered models, applying the NNTS homogeneity likelihood ratio test and counting the number of times that the NNTS test correctly did not reject the null hypothesis of homogeneity for the different values of M_0 and the statistical significance of 1%, 5% and 10%. For comparison purposes, we also applied Watson's two-sample test of homogeneity and Rao's test of homogeneity for the mean directions and dispersions using the R library circular [30]. When we tested a uniform population

against a von Mises or NNTS population, the NNTS and Watson's tests gave 100% correct decisions, but Rao's test had a lower level of power specifically when testing for the equality of mean directions. The results when using Rao's test are a consequence of the fact that the mean values of the simulated models are very similar (see Figure 1); however, there are also cases with low power when testing for the equality of dispersions using Rao's test. When comparing von Mises and NNTS populations in two datasets that have a sample size equal to 20, Watson and Rao's tests have higher power than the NNTS test. By increasing the sample sizes, the power of the NNTS test increases much more quickly than the Watson's test and Rao's test. When both sample sizes are equal to 100, the NNTS has a much higher power than Watson's test and Rao's test. Generally, the maximum power the NNTS test is obtained when $M_0 = 4$, which corresponds to the true M value.

Table 11 contains the type I errors that were obtained after simulating 100 different datasets of the considered models, applying the NNTS homogeneity likelihood ratio test, and counting the number of times that the NNTS test erroneously rejected the null hypothesis of homogeneity for the different values of M_0 . For comparison, we

Page 7 of 9

	Probabilities			
Subinterval	Model 1	Model 2		
(0.000,0.833)	0.08549016	0.12739726		
(0.833,0.167)	0.08922589	0.14314715		
(0.167,0.250)	0.09138272	0.16986301		
(0.250,0.333)	0.09138272	0.15337195		
(0.333,0.417)	0.08922589	0.12739726		
(0.417,0.500)	0.08549016	0.08219178		
(0.500,0.583)	0.08117651	0.04246575		
(0.583,0.667)	0.07744078	0.01137866		
(0.667,0.750)	0.07528395	0.0000000		
(0.750,0.833)	0.07528395	0.01137866		
(0.833,0.917)	0.07744078	0.04109589		
(0.917,1.000)	0.08117651	0.08493151		

 Table 12: Probabilities of the simulated models for the grouped data.

		NNTS test								
<i>n</i> ₁	n ₂	α	<i>M</i> ₀ = 1	2	3	4	5	6		
20	20	0.01	30	21	14	10	8	3		
		0.05	51	42	38	23	20	15		
		0.1	66	53	48	46	35	26		
	50	0.01	56	44	33	30	20	16		
		0.05	76	66	57	52	46	40		
		0.1	87	71	69	65	59	54		
	100	0.01	75	62	52	50	37	30		
		0.05	91	81	79	74	66	58		
		0.1	93	88	89	83	78	72		
50	20	0.01	41	24	13	10	8	3		
		0.05	73	48	48	41	30	18		
		0.1	86	68	58	52	47	36		
	50	0.01	86	71	58	51	50	43		
		0.05	92	88	82	75	69	62		
		0.1	96	93	88	89	80	75		
	100	0.01	98	93	85	84	81	76		
		0.05	100	98	96	97	94	89		
		0.1	100	100	99	98	97	95		
400		0.01		00	00	45	40			
100	20	0.01	51	28	23	15	10	8		
		0.05	81	59	47	39	35	28		
		0.1	93	11	66	58	47	39		
	50	0.01	07	07	01	70	66	50		
	50	0.01	97	07	01	12	00	20		
		0.05	<u> </u>	9/	94	92	00			
		0.1	100	33	33	90	54	90		
	100	0.01	100	100	00	07	96	02		
	100	0.05	100	100	100	97	90	98		
		0.05	100	100	100	100	100	90		
	1	0.1	100	100	100	100	100	99		

Table 13: NNTS homogeneity test for grouped data: Statistical power.

also applied Watson's two-sample test of homogeneity and Rao's test of homogeneity to the mean directions and dispersions. When testing a uniform population against another uniform population, the NNTS yielded probabilities of a type I errors that were equal to zero. In contrast, the Watson and Rao's tests had higher probabilities of type I errors. When applying the homogeneity tests to the datasets that were simulated from the von Mises density, the type I errors for the NNTS tests were similar or smaller than the errors from the Watson and Rao's tests. Lastly, when testing for homogeneity in the NNTS populations, the type I errors of the NNTS test and Watson and Rao's tests are similar. Note that the type I errors of the NNTS test were congruent with the significance levels used to perform the test.

For the grouped data, we simulated 100 datasets of sizes 20, 50, and 100 for two circular populations in the interval (0,1) that were partitioned on subintervals of equal length $\frac{1}{12}$. Table 12 presents the probabilities for each model. For Model 1, the probabilities

were obtained using the formula for a simple sinusoidal curve [31] $P_k = \frac{1+0.1\sin(\frac{(2k-1)\pi}{12})}{12}$, and for Model 2, $P_k = \frac{1+0.25\sin(\frac{k\pi}{6})}{12}$ [32] for k = 1,...,12 were used. Based on the results in Tables 13 and 14, where n_1 and n_2 are the number of observations that were simulated from models 1 and 2, respectively, the NNTS test for homogeneity presented good power and type I errors that were congruent with the considered significance levels. As expected, the NNTS model with $M_0 = 1$ presented the best power because we were simulating from simple sinusoidal models.

Conclusions

The family of circular distributions that is based on nonnegative trigonometric sums (NNTS) is able to model circular populations that present multimodality and/or skewness. This flexibility of the NNTS models makes them suitable candidates for the construction of homogeneity tests for circular data when samples from different circular populations are used to test the equality of the populations'

Page	8	of 9	
------	---	------	--

	1	NNTS tost								
n	n	a	<i>M</i> = 1	2	3	4	5	6		
	n ₂	0.01		2	0	1	0	0		
20	20	0.01	1	2	1	0	0	0		
		0.05	0	10	5	8	2	1		
		0.1	9	10	12	12	9	4		
20	50	0.01	1	0	0	0	0	0		
20	50	0.01	1	0	3	3	0	0		
		0.05	6	11	11	8	6	5		
		0.1	0	11	11	0	0	5		
20	100	0.01	0	0	0	0	0	0		
20	100	0.05	1	2	1	0	0	0		
		0.00	3	5	3	3	3	0		
		0.1	•		Ŭ					
50	50	0.01	1	3	2	2	3	1		
		0.05	4	6	12	12	10	7		
		0.1	14	9	16	19	18	13		
50	100	0.01	1	0	0	0	0	0		
		0.05	4	3	1	2	2	1		
		0.1	10	10	8	7	6	3		
100	100	0.01	3	4	5	6	5	4		
		0.05	7	7	8	9	8	7		
		0.1	11	10	11	11	13	10		
20	20	0.01	0	1	0	0	0	0		
		0.05	1	4	2	2	1	1		
		0.1	3	5	3	2	2	1		
20	50	0.01	0	0	0	0	0	0		
		0.05	3	1	1	1	0	0		
		0.1	5	5	4	3	2	0		
	100			-			-			
20	100	0.01	0	0	0	0	0	0		
		0.05	1	2	1	0	0	0		
		0.1	3	5	3	3	3	U		
	50	0.01	0	0	0	0	0	0		
50	50	0.01	0	0	U 1	1	0	0		
		0.05	2	2			2	2		
		U. I	۷	4	3	3	3	3		
<u>50</u>	100	0.01	0	1	0	1	0	0		
	100	0.01	2	1	2		3	1		
		0.05	5	2	2	<u>ч</u> Л	1	1		
	1	0.1	5		4		4			
100	100	0.01	1	0	1	0	0	0		
100	100	0.01	1	0 3	1 4	0	0	0		

Table 14: NNTS homogeneity test for grouped data: Type I Errors.

circular distributions. In this paper, a likelihood ratio homogeneity test for circular data was constructed by considering that the distributions of the different populations are or can be approximated by members of the NNTS family. Importantly, this test can be applied to ungrouped or grouped observations from these populations. The test was applied to simulated data, confirming the suitability of the test. Also, the test was applied to interesting real datasets to obtain conclusions about the homogeneity of the circular populations. A simulation study showed that the type I error and the power of the proposed test are similar or better than those of the Watson's test and Rao's test for the tested populations (uniform, von Mises and NNTS).

Acknowledgements

The authors thank the referees for their useful comments and are grateful to the Asociación Mexicana de Cultura A.C. for its support.

References

- 1. Fernandez-Duran JJ (2004) Circular Distributions Based on Nonnegative Trigonometric Sums. Biometrics 60: 499-503.
- Fejer L (1916) Uber Trigonometrische Polynome. Journal fur die Reine und Angewandte Mathematik. 146: 53-82.
- Fernandez-Duran JJ (2007) Models for Circular-Linear and Circular-Circular Data Constructed from Circular Distributions Based on Nonnegative Trigonometric Sums. Biometrics 63: 579-585.
- 4. Fernandez-Duran JJ, Gregorio-Dominguez MM (2010) Maximum Likelihood

Estimation of Nonnegative Trigonometric Sum Models Using a Newton-like Algorithm on Manifolds. arXiV:1010.6098v1 [stat.CO].

- R Development Core Team (2010) R: A language and environment for statistical computing, ISBN 3-900051-07-0, R Foundation for Statistical Computing, Vienna, Austria.
- Fernandez-Duran JJ, Gregorio-Dominguez MM (2009) CircNNTSR: An R Package for the Statistical Analysis of Circular Data Using Nonnegative Trigonometric Sums (NNTS) Models v0.1. Working Paper, DE-C09-13, Department of Statistics, ITAM, Mexico.
- Fernandez-Duran JJ Gregorio-Dominguez MM (2010) CircNNTSR: An R package for the statistical analysis of circular data using nonnegative trigonometric sums (NNTS) models. R package version 1.0-1.
- Mardia KV, Spurr BD (1973) Multisample Tests for Multimodal and Axial Circular Populations. J R Stat Soc Series B Stat Methodol 35: 422-436.
- 9. Mardia KV (1972) Statistics of Directional Data. Academic Press, London
- Wheeler S, Watson GS (1964) A Distribution-Free Two-Sample Test on the Circle. Biometrika 51: 256-257.
- 11. Rao JS (1967) Large Sample Tests for the Homogeneity of Angular Data. Sankhya 28: 172-174.
- Rao JS (1984) Nonparametric Methods in Directional Data. Chapter 31 in Handbook of Statistics, Volume 4. Nonparametric Statistics, P.R. Krishnaiah and P.K. Sen (eds.), pp. 755-770. Amsterdam: Elsevier Science Publishers.
- Jammalamadaka SR, SenGupta A (2001) Topics in Circular Statistics.: World Scientific Publishing, Co., Singapore.

- 14. Watson GS (1962) Goodness-of-fit Tests on a Circle, II. Biometrika 49: 57-63.
- Watson GS, Williams EJ (1956) On the Construction of Significance Tests on the Circle and the Sphere. Biometrika 43: 344-352.
- Stephens MA (1969) Tests for the von Mises Distribution. Biometrika 56: 149-160.
- 17. Watson GS (1983) Statistics on Spheres. Wiley, New York.
- Fisher NI, Hall P (1990) New Statistical Methods for Directional Data I. Bootstrap Comparison of Mean Directions and the Fold Test in Palaeomagnetism. Geophysical Journal International 101: 305-313.
- Harrison D, Kanji GK, Gadsden RJ (1986) Analysis of Variance for Circular Data. J Appl Stat 13: 123-138.
- Harrison D, Kanji GK (1988) The Development of Analysis of Variance for Circular Data. J Appl Stat 15: 197-224.
- Stephens MA (1992) On Watson's ANOVA for Directions. In K.V. Mardia (ed.) The Art of Statistical Science: A Tribute to G.S. Watson, pp. 75-85, Chichester: Wiley.
- Anderson CM, Wu CFJ (1995) Measuring Location Effects from Factorial Experiments with a Directional Response. Int Stat Rev 63: 345-363.
- Grimshaw SD, Whiting DG, Morris TH (2001) Likelihood Ratio Tests for a Mixture of Two von Mises Distributions. Biometrics 57: 260-265.

24. Chen J, Li P, Fu Y (2008) Testing Homogeneity in a Mixture of von Mises Distributions with a Structural Parameter. Can J Stat 36: 129-142.

Page 9 of 9

- Fu Y, Chen J, Li P (2008) Modified Likelihood Ratio Test for Homogeneity in a Mixture of von Mises Distributions. J Stat Plan Inference 138: 667-681.
- 26. Mardia KV, Jupp PE (2000) Directional Statistics. (2Edition), John Wiley and Sons, Great Britain.
- 27. Mardia KV (1972) A Multi-Sample Uniform Scores Test on a Circle and its Parametric Competitor. J R Stat Soc Series B Stat Methodol 34: 102-113.
- Fisher NI (1993) Statistical Analysis of Circular Data. Cambridge: Cambridge University Press
- Wells NA (1990) Comparing Sets of Circular Orientations by Modified Chisquared Testing. Comput Geosci 16: 1155-1170.
- 30. Lund U, Agostinelli C (2010) circular: Circular Statistics. R package version 0.4.
- Marrero O (1983) The Performance of Several Statistical Tests for Seasonality in Monthly Data. Journal of Statistical Computation and Simulation 17: 275-296.
- Freedman LS (1979) The Use of a Kolmogorow-Smirnov Type Statistic in Testing Hypotheses about Seasonal Variation. J Epidemiol Community Health 33: 223-228.