Journal of Sciences, Islamic Republic of Iran 34(1): 43 - 50 (2023) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Hypotheses Tests for Circular Data in Weighted Sampling

F. Shahsanaei and R. Chinipardaz*

Department of Statistics, Shahid Chamran University of Ahvaz, Ahvaz, Ahvaz, Islamic Republic of Iran

Received: 20 September 2022 / Revised: 2 January 2023 / Accepted: 30 January 2023

Abstract

This paper is concerned with the problem of statistical hypotheses testing in circular data under weighted sampling. The most powerful test has been obtained when the sampling is subjected to a weight function. Different weight functions are examined for the von Misses distribution. For the same weight function, the critical values and the power of the test can be calculated analytically, and for some, we need to use a numerical method. The simulation study shows that the power of the test increased as the weighted circular distribution is considered in replace of the original circular distribution and the sample data increased. A real-data example has been carried out to show the performance of our method.

Keywords: Circular data; Weight function; Weighted sampling; MP test; Monte Carlo simulation.

Introduction

Angles and directions are used to measure circular data. Typically, they are observed in scientific fields as diverse as life sciences (1), behavioral biology (2), cognitive psychology (3), bioinformatics (4), and political sciences (5). Circular data are frequently encountered in the study of motor behavior (6-9), and also in the application of circumplex models (10-12). The difference between circular data and linear data is that circular data are measured in a periodical sample space. For example, an angle of 10° is quite close to an angle 350°, despite linear intuition suggesting otherwise.

Weighted sampling is a generalization of random, in which the sampling mechanism records units with a probability based on a non-negative weight function. The recorded data called a weighed sample is not a random sample from the original population. This sampling approach is called weighted sampling.

The usual methods of inferring unknown parameters from a weighted sample are not useful and should be adjusted. Suppose that the circular random variable θ is distributed with a probability density function, pdf, $f(\theta;\eta)$ where the natural parameter is η. Suppose a realization θ of Θ records or enters the sample with a probability proportional to an arbitrary non-negative weight function $w(\theta;\lambda)$, where λ is the weight parameter. In this case, the observation θ is not based on the circular random variable Θ, but on its weighted version instead. This random variable which is denoted by Θ^w has a weighted distribution with the following probability density function:

$$f^w(\theta;\eta) = \frac{w(\theta;\lambda)f(\theta;\eta)}{E[w(\theta;\lambda)]} \quad (1)$$

where $E(w(\theta;\lambda))$ is required to be finite as a normalizing constant. The $f^w(\theta;\eta)$ is the weighted probability density function of Θ under the weight function $w(\theta;\lambda)$. When a random sample of an interest population is unavailable, weighted sampling is conducted to represent the unequal probability of samples entering the population. A carefully chosen biased sample may prove more informative than a

* Corresponding author: Tel: +989161117832; Email: chinipardaz_r@scu.ac.ir

Table 1. Some weight functions.

	Weight function ($w(\theta)$)	Name
1	$w(\theta) = c$	Constant
2	$w(\theta)=\exp\{\lambda_1 \cos j\theta + \lambda_2 \sin j\theta\}$	Exponential
3	$w(\theta)=1+\lambda_1 \sin j\theta + \lambda_2 \cos j\theta$	linear

sample obtained even if the experimenter is able to obtain one (13) introduced the idea of weighted distribution by demonstrating that models must be adjusted based on how data are collected (14, 15) was the first to unify this concept. Weighted distributions have been used to select appropriate models from data sampled without a proper frame. In this study, the hypothesis testing problem is considered for circular data when the sampled data is subjected to a weight function. The main study's motivation is that in weighted samplings, the parameter, η, may be represented by more or less information than under random samplings. Many weight functions have been introduced by authors in linear statistics. The different weight functions can be considered based on the nature of the data when analyzing circular data. A set of suitable weight functions are proposed in Table 1 that we believe can lead to a tractable analysis of circular data. The paper is provided in five sections. In Section 2 the effect of the weight function on the test function and power of the test is studied for von Mises distribution under some weight functions. In Section 3, selecting an appropriate weight function from two or a set of candidates of weight functions based on the Neyman–Pearson Lemma is discussed for von Mises and cardioid distributions with different sample sizes. In Section 4, we use Monte Carlo simulation to determine the powers of these tests when the distributions of test statistics are complicated. Finally, in Section 5, using the soil samples of Ahvaz in Iran, the weighted sampling is tested against random sampling.

Hypothesis testing for original parameters of population

In this section, we look at the problem of hypothesis testing under weighted sampling for parameter η of $\Theta \sim f(\theta;\eta)$. Suppose that the parameters of the weight function, $\boldsymbol{\lambda}$, are known and we wish to test $H_0{:}\,\eta = \eta_0$ versus $H_1{:}\,\eta = \eta_1$ where η_0 and η_1 are known. Consider $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)'$ is realized out of the weighted sample $\boldsymbol{\Theta}^w \sim (\Theta_1^w, \Theta_2^w, \ldots, \Theta_n^w)'$ from Θ under $w(\theta;\lambda)$.

The most powerful (MP) test with size α based on the Neyman-Pearson Lemma is given by

$$\varphi(\boldsymbol{\theta}^w) = \begin{cases} 1 & f^w(\boldsymbol{\theta};\eta_1) \geq c_1 f^w(\boldsymbol{\theta};\eta_0) \\ 0 & otherwise \end{cases}$$

where the critical value, c_1, is obtained with $E_{\eta_0}(\varphi(\boldsymbol{\theta}^w)) = \alpha$. Then

$$\frac{f^w(\boldsymbol{\theta};\eta_1)}{f^w(\boldsymbol{\theta};\eta_0)} = \frac{\prod_{i=1}^n f(\theta_i;\eta_1)(E_{\eta_0} w(\theta;\lambda))^n}{\prod_{i=1}^n f(\theta_i;\eta_0)(E_{\eta_1} w(\theta;\lambda))^n}$$
$$= v(\eta_0,\eta_1)\frac{f(\boldsymbol{\theta};\eta_1)}{f(\boldsymbol{\theta};\eta_0)},$$

where $v(\eta_0,\eta_1)$ is independent of the test statistic. As $v(\eta_0,\eta_1)$ is positive,

$$\frac{f(\boldsymbol{\theta};\eta_1)}{f(\boldsymbol{\theta};\eta_0)} > c_2 \Leftrightarrow \frac{f^w(\boldsymbol{\theta};\eta_1)}{f^w(\boldsymbol{\theta};\eta_0)} > c_1,$$

where

$$c_1 = c_2((E_{\eta_0} w(\theta;\lambda))^n / (E_{\eta_1} w(\theta;\lambda))^n).$$

It means that the test structure is the same for both weighted and random sampling, although the critical value may differ. The following sections illustrate this.

Test for mean direction

Suppose that $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)'$ is a circular random sample from $\theta \sim \mathrm{vM}(\mu,\kappa)$ where κ is known and the notation vM represents von Mises distribution and has density

$$f_{vM}(\theta;\mu,\kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)}, \tag{2}$$

for $\theta \in (-\pi,\pi)$, $\mu \in (-\pi,\pi)$, $\kappa \geq 0$ and where $I_r(\kappa)$ is the modified Bessel function of the first kind of order r, defined as

$$I_r(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos r\theta \, e^{\kappa\cos\theta} d\theta,$$
$$r = 0, \pm 1, \pm 2, \ldots.$$

The MP test for testing $H_0{:}\,\mu = 0$ versus $H_1{:}\,\mu = \mu_1$ $(\mu_1 > 0)$ is

$$\varphi(\boldsymbol{\theta}) = \begin{cases} 1 & \sin(\bar{\theta} - \mu_1/2) \geq d_1 \\ 0 & \sin(\bar{\theta} - \mu_1/2) < d_1 \end{cases}$$

Equivalently

$$\varphi(\boldsymbol{\theta}) = \begin{cases} 1 & \bar{\theta} \in g \\ 0 & \bar{\theta} \notin g \end{cases}$$

where

$$g = \mathrm{arc}\left(\left\{\frac{\pi}{2} + \frac{\mu_1}{2} - \delta\right\}(\mathrm{mod}\ 2\pi), \left\{\frac{\pi}{2} + \frac{\mu_1}{2} + \delta\right\}(\mathrm{mod}\ 2\pi)\right),$$

$0<\delta<\pi/2$.

Let $\bar{\theta}|R$ distributed as $vM(\mu,\kappa R)$, where $R=n(\bar{C}^2+\bar{S}^2)^{\frac{1}{2}}$, $\bar{C}=\frac{1}{n}\sum_{i=1}^{n}\cos\theta_i$ and $\bar{S}=\frac{1}{n}\sum_{i=1}^{n}\sin\theta_i$. When $R>0$, the mean direction $\bar{\theta}$ is given by

$$\bar{\theta}=\begin{cases}\tan^{-1}(\bar{S}/\bar{c}) & \bar{C}\geq 0\\ \tan^{-1}(\bar{S}/\bar{c})+\pi & \bar{C}<0.\end{cases}$$

The constant δ is then determined from

$$\alpha=\int_{\frac{\pi}{2}+\frac{\mu_1}{2}-\delta}^{\frac{\pi}{2}+\frac{\mu_1}{2}+\delta}\frac{e^{\kappa R\cos\theta}}{2\pi I_0(\kappa R)}\,d\theta, \qquad (3)$$

similarly, the power of the test is

$$\beta=\int_{\frac{\pi}{2}+\frac{\mu_1}{2}-\delta}^{\frac{\pi}{2}+\frac{\mu_1}{2}+\delta}\frac{e^{\kappa R\cos(\theta-\mu_1)}}{2\pi I_0(\kappa R)}\,d\theta,$$

See Mardia (16).

Now, suppose that $\boldsymbol{\theta}^{w}=(\theta_1^{w},\theta_2^{w},\ldots,\theta_n^{w})'$ is a weighted sample from $\theta\sim vM(\mu,\kappa)$ with weight function $w(\theta)=e^{\lambda\cos(\theta-\mu)}$ where λ is known, then $\theta^{w}\sim vM(\mu,\kappa+\lambda),\kappa+\lambda>0$. Then the MP test is again

$$\varphi(\theta^{w})=\begin{cases}1 & \sin(\bar{\theta}^{w}-\mu_1/2)\geq d_2\\ 0 & \sin(\bar{\theta}^{w}-\mu_1/2)<d_2\end{cases}$$

Equivalently,

$$\varphi(\theta^{w})=\begin{cases}1 & \bar{\theta}^{w}\in g\\ 0 & \bar{\theta}^{w}\notin g\end{cases}$$

Proof. See the Appendix.

Let $\bar{\theta}^{w}|R$ distributed as $vM(\mu,\kappa R+\lambda)$. The constant δ is then determined by

$$\alpha=\int_{\frac{\pi}{2}+\frac{\mu_1}{2}-\delta}^{\frac{\pi}{2}+\frac{\mu_1}{2}+\delta}\frac{e^{(\kappa R+\lambda)\cos\theta}}{2\pi I_0(\kappa R+\lambda)}\,d\theta,$$

similarly, the power of the test is

$$\beta^{w}=\int_{\frac{\pi}{2}+\frac{\mu_1}{2}-\delta}^{\frac{\pi}{2}+\frac{\mu_1}{2}+\delta}\frac{e^{(\kappa R+\lambda)\cos(\theta-\mu_1)}}{2\pi I_0(\kappa R+\lambda)}\,d\theta.$$

Table 2 shows the power of the test $H_0:\mu=0$ versus $H_1:\mu=\frac{\pi}{2}$ for $\kappa R=1$ and various α and λ. According to Table 2, the power of the test increases when λ increases and also the power of the test of weighted sampling ($\lambda\neq 0$) is more than random sampling ($\lambda=0$). The last column of the table shows the nominal significance level (nsl) obtained from the simulation study (n=1000), which is close to the pre-assigned significance level.

Test for the concentration parameter

Suppose that $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_n)'$ is a circular random sample from $\theta\sim vM(0,\kappa)$ where $\mu=0$. The MP test for testing $H_0:\kappa=\kappa_0$ versus $H_1:\kappa=\kappa_1(\kappa_0<\kappa_1)$ is

$$\varphi(\theta)=\begin{cases}1 & C\geq d_1\\ 0 & C<d_1\end{cases}$$

where $C=\sum_{i=1}^{n}\cos\theta_i$ and the probability density function of C is given by

$$g(c;\kappa)=\{\pi I_0^{n}(\kappa)\}^{-1}e^{\kappa c}\int_0^{\infty}cos(ct)J_0^{n}(t)\,dt,$$
$$-n<c<n.$$

where $J_0(x)$ is the standard Bessel function of zero order, it is given by

$$J_0(x)=\sum_{k=0}^{\infty}\frac{(-1)^k}{(k!)^2}\left(\frac{x}{2}\right)^{2k},$$

See Mardia (16).

The constant d_1 is then determined from

$$\alpha=P(C>d_1|\kappa=\kappa_0),$$

with the power of the test

Table 2. The power of the test and the nominal significance level $H_0:\mu=0$ versus $H_1:\mu=\frac{\pi}{2}$ for the several of α and λ.

	$\boldsymbol{\alpha=0.01}$			$\boldsymbol{\alpha=0.05}$			$\boldsymbol{\alpha=0.1}$		
$\boldsymbol{\lambda}$	δ	β^{w}	*nsl*	δ	β^{w}	*nsl*	δ	β^{w}	*nsl*
0	0.087	0.044	0.012	0.393	0.199	0.046	0.785	0.390	0.105
1	0.288	0.166	0.011	0.925	0.542	0.055	1.287	0.723	0.114
2	0.742	0.458	0.017	1.285	0.788	0.048	1.545	0.884	0.099
5	1.344	0.906	0.01	1.654	0.977	0.054	1.815	0.990	0.118
10	1.641	0.996	0.01	1.850	0.999	0.049	1.972	0.999	0.109
20	1.841	0.999	0.013	1.985	0.999	0.052	2.075	1	0.107

$$\beta = P(C > d_1 | \kappa = \kappa_1).$$

Also, for $\kappa > 2$, we can use Stephen's approximation (17)

$$2\gamma(n - C) \simeq \chi^2_{(n)}, \qquad \gamma^{-1} = \kappa^{-1} + \frac{3}{8}\kappa^{-2}.$$

Therefore, $d_1 = n - \frac{\chi^2_{(n,\alpha)}}{2\gamma_0}$ where $\gamma_0^{-1} = \kappa_0^{-1} + \frac{3}{8}\kappa_0^{-2}$ and $\chi^2_{(n,\alpha)}$ is the αth quantile of chi-square distribution with n degree of freedom.

The power of the test is

$$\beta=P(C>n-\frac{\chi^2_{(n,\alpha)}}{2\gamma_0})=F_{\chi^2_{(n)}}(\frac{\gamma_1}{\gamma_0}\chi^2_{(n,\alpha)}),$$

where $F_{\chi^2_{(n)}}$ is comulative chi-square distribution function and $\gamma_1^{-1} = \kappa_1^{-1} + \frac{3}{8}\kappa_1^{-2}$.

Now suppose that $\boldsymbol{\theta}^{\mathbf{w}}=(\theta_1^w,\theta_2^w,...,\theta_n^w)'$ is a weighted sample from $vM(0,\kappa)$ with weight function $w(\theta) = e^{\lambda\cos\theta}$ where λ is known, then $\theta^w \sim vM(0, \kappa + \lambda)$. For testing $H_0: \kappa = \kappa_0$ versus $H_1: \kappa = \kappa_1 (\kappa_0 < \kappa_1)$ under the weighted sample, the MP test is

$$\varphi(\theta^w) = \begin{cases} 1 & C^w \geq d_2 \\ 0 & C^w < d_2 \end{cases}$$

where $C^w=\sum_{i=1}^n \cos\theta_i^w$ and $d_2=n-\frac{\chi^2_{(n,\alpha)}}{2\gamma_0}$, $\gamma_0^{-1}=(\kappa_0+\lambda)^{-1}+\frac{3}{8}(\kappa_0+\lambda)^{-2}$. The power of test under weighted sample is

$$\beta^w=E_{\kappa_1}(\varphi(\theta^w))=F_{\chi^2_{(n)}}(\frac{\gamma_1}{\gamma_0}\chi^2_{(n,\alpha)}),$$

where $\gamma_1^{-1}=(\kappa_1+\lambda)^{-1}+\frac{3}{8}(\kappa_1+\lambda)^{-2}$.

Hypothesis testing for weight functions

There is no common usage method to find an appropriate weight function. As the selection of the weight function is a very important issue in data analysis, especially in testing and model selection problems, as was shown before, very careful efforts have to be made when selecting the weight function from many candidates.

Several weight functions could be considered. In the case of the weight function $w(\theta) = c$, i.e. the sample data is random, there is no change in the results because the weighted distribution is the same as the original distribution. Assume we are interested in finding out what weight function is satisfied by the recorded data:

H_0: Data is the weighted sample with weight function $w_0(\theta)$

H_1: Data is the weighted sample with weight function $w_1(\theta)$

The following hypotheses are equivalent to these statements: $H_0: \boldsymbol{\theta}^w \sim f^{w_0}(\theta)$ versus $H_1: \boldsymbol{\theta}^w \sim f^{w_1}(\theta)$. If $f^{w_i}(\theta) (i = 0,1)$ is completely specified under $H_j (j = 0,1)$. According to the Neyman–Pearson Lemma, the MP test is given by

$$\varphi(\boldsymbol{\theta})=\begin{cases} 1 & \frac{f^{w_1}(\boldsymbol{\theta})}{f^{w_0}(\boldsymbol{\theta})}>d_1 \\ 0 & \frac{f^{w_1}(\boldsymbol{\theta})}{f^{w_0}(\boldsymbol{\theta})}\leq d_1 \end{cases} = \begin{cases} 1 & T^w>d_2 \\ 0 & T^w\leq d_2 \end{cases}$$

where test statistic is $T^w=\prod_{i=1}^n w_1(\theta_i)/\prod_{i=1}^n w_0(\theta_i)$ and d_2 is obtained from

$$P^{w_1}(T^w > d_2) = \alpha,$$

and P^{w_1} represents for the probability under weight function $w_1(\theta)$.

Test for two exponential weight functions

Suppose that $\boldsymbol{\theta}^{\mathbf{w}} = (\theta_1^w, \theta_2^w, \ldots, \theta_n^w)'$ is a circular weighted sample from $\theta \sim vM(0, \kappa)$. We want to test $H_0: w(\theta) = e^{\lambda_0\cos\theta}$ versus $H_1: w(\theta) = e^{\lambda_1\cos\theta}$, where $\lambda_1 > \lambda_0$ and under $H_i(i = 0,1)$, $\theta^w \sim vM(0, \kappa + \lambda_i)$. Therefore, the MP test level α is

$$\varphi(\boldsymbol{\theta}) = \begin{cases} 1 & C^w \geq d_2 \\ 0 & C^w < d_2 \end{cases}$$

where $C^w = \sum_{i=1}^n \cos\theta_i{}^w$ and the probability density function of C^w under H_i for $i = 0,1$ is given by

$$g(c^w; \kappa + \lambda_i) = \frac{e^{(\kappa+\lambda_i)c^w}}{\pi I_0^n(\kappa+\lambda_i)} \int_0^\infty cos(c^w x) J_0^n(x) dx,$$

$$n < c^w < n.$$

The constant d_2 is then determined from

$$\alpha = P(C^w > d_2 | \lambda = \lambda_0),$$

and the power of the test is

$$\beta^w = P(C^w > d_2 | \lambda = \lambda_1).$$

Also, for $\kappa + \lambda_i > 2$, we can use Stephen's approximation (17) for $i = 0,1$

$$2\gamma_i(n - C^w) \simeq \chi^2_{(n)},$$

$$\gamma_i^{-1} = (\kappa + \lambda_i)^{-1} + \frac{3}{8}(\kappa + \lambda_i)^{-2}.$$

Therefore, $d_2 = n - \frac{\chi^2_{(n,\alpha)}}{2\gamma_0}$ where $\gamma_0^{-1} = (\kappa +$

$\lambda_0)^{-1} + \frac{3}{8}(\kappa + \lambda_0)^{-2}$ and $\chi^2_{(n,\alpha)}$ is the αth quantile of chi-square distribution.

The power of the test is

$$\beta^w = P(C^w > n - \frac{\chi^2_{(n,\alpha)}}{2\gamma_0} | \lambda = \lambda_1) = F_{\chi^2_{(n)}}\left(\frac{\gamma_1}{\gamma_2}\chi^2_{(n,\alpha)}\right),$$

where $F_{\chi^2_{(n)}}(u)$ is comulative chi-square distribution function at u.

Test for the linear versus exponential weight function

Suppose that $\boldsymbol{\theta}^{\mathbf{w}}=(\theta_1^w,\theta_2^w,...,\theta_n^w)'$ is a sample from $\theta \sim vM(0,\kappa)$. We want to test $H_0: w(\theta) = 1 + \lambda\sin\theta$ versus $H_1: w(\theta) = e^{\gamma\cos\theta}$, where λ and γ are specified.

Under H_1, $\theta^w \sim vM(0, \kappa + \gamma)$.

The MP test is given by

$$\varphi(\theta^w) = \begin{cases} 1 & T^w > k^* \\ 0 & T^w \leq k^* \end{cases}$$

where test statistics is $T^w = \gamma \sum_{i=1}^n \cos\theta_i - \sum_{i=1}^n \ln(1 + \lambda\sin\theta_i)$ and k^* is obtained from

$$P_{H_0}(\gamma\sum_{i=1}^{n}\cos\theta_i - \sum_{i=1}^{n}\ln(1+\lambda\sin\theta_i) > k^*) = \alpha.$$

The distribution of T^w is complicated. However, k^* and β^w can be calculated using Monte Carlo simulation. Table 3 represents the values of the simulated power of this test for $n = 5, 10, 50$ and 100 using the simulation with $m_1 = m_2 = 10000$ iterations.

Test for the constant versus linear weight function

Clearly, for random sampling, $w_i(\theta) = 1$. In this case, the researcher wants to know if the sampling is random or weighted with the weight function such as $w(\theta)$:

H_0: Data is a random sample

H_1: Data is a weighted sample with weight function $w(\theta)$.

Therefore, H_0 is rejected if

$$T^w = \prod_{i=1}^{n} w(\theta_i^w) > k^*,$$

where k^* is given by $P_{Random}(T^w > k^*) = \alpha$.

Suppose that $\boldsymbol{\theta}^{\mathbf{w}}=(\theta_1^w,\theta_2^w,...,\theta_n^w)'$ is a sample from $\Theta \sim f(\theta) = \frac{1}{2\pi}(1+2\rho\cos\theta)$, $|\rho| \leq \frac{1}{2}$. To test $H_0: w(\theta) = 1$ versus $H_1: w(\theta) = 1 + \lambda\sin\theta$, when λ is specified, we have:

$$\varphi(\theta^w) = \begin{cases} 1 & \prod_{i=1}^{n}(1 + \lambda\sin\theta_i{}^w) > k^* \\ 0 & \prod_{i=1}^{n}(1 + \lambda\sin\theta_i{}^w) \leq k^* \end{cases}$$

where k^* is obtained from $P_{H_0}(\prod_{i=1}^n (1 + \lambda\sin\theta_i) > k^*) = \alpha$, because under H_0, $\theta^w \equiv \theta$.

For $n = 1$, k^* is obtained from

$$P(1 + \lambda\sin\theta > k^*) = \frac{1}{2\pi}(\pi - \sin^{-1}\left(\frac{k^* - 1}{\lambda}\right) - 2\rho\left(\frac{k^* - 1}{\lambda}\right)) = \alpha.$$

This equation is not solvable and therefore, no value can be obtained for k^*.

For $n > 1$, the distribution of T^w is complicated. However, k^* and β^w can be calculated using Monte Carlo simulation. Table 4 represents the values of the simulated power of this test for $n = 10, 20, 50$ and 100 using simulation with $m_1 = m_2 = 10000$ iterations.

Simulation study

Frequently, testing $H_0: w(\theta) = w_0(\theta)$ versus $H_1: w(\theta) = w_1(\theta)$, the distributions of T^w under H_0 and H_1 are complicated. Using Monte Carlo simulation, the critical value, k^*, and the power test, β^w, can be calculated. We employed R Software to generate random numbers. The procedure of our desired algorithm is given by the following steps:

1. Simulate m_1 times from a sample of size n from the weighted distribution specified in the null hypothesis, namely $f^{w_0}(\theta)$, and compute the test statistic for each sample. Then k^* is introduced as $(1 - \alpha)$th quantile of these m_1 values.
2. Simulate m_2 times from a sample of size n from the weighted distribution specified in H_1, namely $f^{w_1}(\theta)$, and compute the test statistic, T_i^w, for each sample. Then the power of the test, β^w, will be

$$\beta^w = \frac{\text{Number of simulated } T_i^w \text{ less than } k^*}{m_2}.$$

According to Table 3, the power of the test increases when n and γ increase for fixed α and from Table 4, we can see that the power of test increases when λ increases for fixed α. The nsl columns of tables 3 and 4 show the nominal significance level obtained from the simulation study, which is close to the pre-assigned significance level.

Table 3. Simulated power of test and nominal significance level $H_0: w(\theta) = 1 + 0.5sin\theta$ versus $H_1: w(\theta) = e^{\gamma cos\theta}$ using $m_1 = m_2 = 10000$ iterations and several sample sizes when $\kappa = 1$ and γ=1 and 2.

		$\gamma = 1$			$\gamma = 2$		
Sample size	α	k^*	β^w	nsl	k^*	β^w	nsl
	0.01	5	0.105	0.0096	5.002	0.161	0.0115
	0.025	4.657	0.215	0.0286	4.680	0.327	0.0286
n=5	0.05	4.378	0.312	0.0504	4.352	0.498	0.0546
	0.1	3.919	0.483	0.1	3.929	0.691	0.0987
	0.15	3.580	0.601	0.1524	3.579	0.804	0.1519
	0.2	3.302	0.682	0.2005	3.305	0.866	0.2022
	0.01	8.563	0.253	0.0103	8.516	0.526	0.011
	0.025	8.002	0.395	0.0233	7.953	0.700	0.0268
n=10	0.05	7.368	0.557	0.0522	7.402	0.834	0.0524
	0.1	6.630	0.718	0.1056	6.684	0.924	0.1035
	0.15	6.158	0.800	0.1565	6.218	0.959	0.1528
	0.2	5.770	0.854	0.2044	5.822	0.977	0.2004
	0.01	30.430	0.969	0.0099	30.711	0.999	0.0081
	0.025	28.768	0.988	0.0256	28.895	1	0.0227
n=50	0.05	27.264	0.996	0.0523	27.598	1	0.0438
	0.15	24.583	0.999	0.1512	24.575	1	0.145
	0.2	23.645	0.999	0.2009	23.596	1	0.202
	0.01	54.315	1	0.01	55.073	1	0.0077
	0.025	52.600	1	0.0228	52.290	1	0.0231
n=100	0.05	50.059	1	0.0537	50.115	1	0.0456
	0.15	46.012	1	0.1565	46.203	1	0.1377
	0.2	44.804	1	0.1984	44.655	1	0.1952

Table 4. Simulated power of test and nominal significance level $H_0: w(\theta) = 1$ versus $H_1: w(\theta) = 1 + \lambda sin\theta$ using $m_1 = m_2 = 10000$ iterations and several sample sizes when $\rho = 0.1$.

		$\lambda = 0.25$			$\lambda = 0.5$			$\lambda = 0.75$		
Sample size	α	k^*	β^w	nsl	k^*	β^w	nsl	k^*	β^w	nsl
	0.01	1.117	0.044	0.0125	1.997	0.098	0.0101	2.609	0.242	0.0101
	0.025	0.932	0.068	0.0264	1.589	0.203	0.0255	2.026	0.410	0.0251
n=10	0.05	0.783	0.137	0.0462	1.246	0.314	0.0523	1.477	0.561	0.0501
	0.15	0.426	0.319	0.1503	0.561	0.549	0.1563	0.273	0.785	0.1501
	0.2	0.319	0.400	0.2023	0.335	0.628	0.2051	-0.086	0.847	0.2
	0.01	1.515	0.058	0.0099	2.401	0.213	0.0105	2.831	0.559	0.0105
	0.025	1.218	0.121	0.0271	1.791	0.362	0.0275	1.823	0.719	0.0258
n=20	0.05	1.017	0.203	0.0481	1.377	0.482	0.0536	1.016	0.830	0.0513
	0.15	0.528	0.404	0.1455	0.379	0.726	0.1568	-0.675	0.944	0.1504
	0.2	0.375	0.487	0.1968	0.069	0.790	0.199	-1.152	0.965	0.1997
	0.01	2.157	0.149	0.0113	2.332	0.603	0.014	1.191	0.964	0.0108
	0.025	1.686	0.247	0.0266	1.423	0.735	0.0242	-0.235	0.983	0.0265
n=50	0.05	1.255	0.337	0.0543	0.572	0.837	0.0577	-1.793	0.994	0.0528
	0.15	0.491	0.587	0.1584	-0.635	0.937	0.1495	-4.347	0.999	0.1485
	0.2	0.268	0.653	0.2041	-1.186	0.960	0.2019	-5.437	0.999	0.2063
	0.01	2.507	0.297	0.0114	1.885	0.915	0.0104	-3.586	0.999	0.0098
	0.025	1.874	0.435	0.0278	0.493	0.962	0.0259	-5.590	1	0.0256
n=100	0.05	1.347	0.555	0.0531	-0.741	0.976	0.0575	-7.606	1	0.0527
	0.15	0.252	0.773	0.1545	-3.001	0.995	0.154	-11.613	1	0.1462
	0.2	-0.088	0.835	0.2046	-3.689	0.997	0.1999	-12.999	1	0.1977

Now, suppose that $\theta_1, \theta_2, \ldots, \theta_{20}$ are observed values of a sample from $\Theta \sim f(\theta) = \frac{1}{2\pi}(1 + 0.1\cos\theta)$. Thus, H_0: 'the sample is random' versus H_1: 'the sample is weighted with weight function $w(\theta) = 1 + 0.75\sin\theta$' is rejected at the significant level 2.5%, when $\prod_{i=1}^{20}(1 + 0.75\sin{\theta_i}^w) > 1.823$. Note that the power of this test is 0.719 in Table 4.

Application

Directional data is an important category of geologic information. The direct shear test is a laboratory test performed on samples taken from drilling and excavating operations in the field and after transfer to the laboratory. The internal friction angle of the soil particles is one of soil shear strength parameters and the important outputs of this test and depends directly on soil type, wet percent, and density of undisturbed soil. Considering the mechanism for selecting the samples

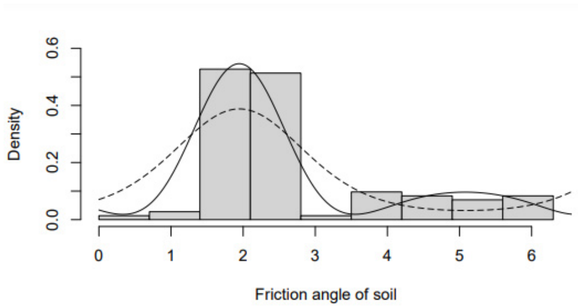

Figure 1. Histogram of the Friction angle of soil data (in radians), and fitted densities for the weighted von Mises distribution under H_0 (long dashed) and H_1 (soild).

tested on the soil type, the harvested samples do not have the same chances in the study area. The data set of this paper, which is obtained from direct shear tests of samples taken from Ahvaz in the southwest of Iran, is the internal friction angle of soil particles. Considering clay and sand soil have a better chance of being selected in Ahvaz city, it can be assumed that the chance of chosen soil samples is proportional to the weight functions.

Suppose that $\boldsymbol{\theta^w} = (\theta_1^w, \theta_2^w, \ldots, \theta_n^w)'$ is a sample from $\Theta \sim vM(\mu, \kappa)$, we want to test $H_0: w(\theta) = 1$ versus $H_1: w(\theta) = 1 + \lambda_1 \sin(2(\theta - \mu)) + \lambda_2 \cos(2(\theta - \mu))$, where $|\lambda_1| + |\lambda_2| < 1$. The MP test is

$$\varphi(\theta^w) = \begin{cases} 1 & T^w > k^* \\ 0 & T^w \le k^* \end{cases}$$

where $T^w = \sum_{i=1}^{n} \ln(1 + \lambda_1 \sin(2(\theta_i - \mu)) + \lambda_2 \cos(2(\theta_i - \mu)))$ is a test statistic. The maximum likelihood estimations of μ, λ_1 and λ_2 based on the weighted sample are 1.95, -0.01 and 0.85 (see (18)). As the distribution of T^w is complicated, the critical value, k^*, and the power test, β^w, are calculated using Monte Carlo simulation with 10000 iterations.

The observed value of the test statistic for our sample with size 103 is

$$T^w = \sum_{i=1}^{103} \ln(1 - 0.01\sin(2(\theta_i - 1.95)) + 0.85\cos(2(\theta_i - 1.95))) = -0.032.$$

Since the test statistic is more than -0.229, the simulated critical value at the significance level of 0.01, so the random sample versus the weighted sample (with the weight function $w(\theta) = 1 - 0.01 \sin(2(\theta - 1.95)) + 0.85 \cos(2(\theta - 1.95)))$ at the significance level of 0.01 is rejected. It means that the analyzing internal friction angle of soil particles data without considering a proportional weight function may lead to misleading results. As the internal friction angle of soil particles depends on different factors, such as wet percent, the density of undisturbed soil, and soil type, different weight functions should be investigated for other areas.

Figure 1 shows the histogram of the data and fitted densities for the weighted von Mises distribution under H0 and H1.

Results

In this paper, we considered testing hypotheses in directional data when the sampling is subjected to a weight function rather than random sampling.

We showed that the most powerful test might be invariant from some appropriate weight functions; however, the power of the test is crucially changed with the weight function parameters.

We focused on von Misses distribution, the most popular directional data distribution, and weight functions considered for both location and concentrate parameters.

It has been shown that the proposed method works well; even the own weight function is considered a hypothesis to be tested.

For an application example, we considered the internal friction angle of the soil participles data, taken from the direct shear test in Ahvaz city southwest of Iran. The data has been tested to find any weight function that affected the data during collection.

The random sampling against a two-parameter weight function has been rejected with a 0.01 level test.

The given method can be further investigated for other directional distributions to find the weight function for the parameters in the testing problem.

Appendix

Proof of test for mean direction.

Suppose that $\boldsymbol{\theta^w} = (\theta_1^w, \theta_2^w, \ldots, \theta_n^w)'$ is a weighted sample with weight function $w(\theta) = e^{\lambda \cos(\theta - \mu)}$ where λ is known, then $\theta^w \sim vM(\mu, \kappa + \lambda), \kappa + \lambda > 0$ for

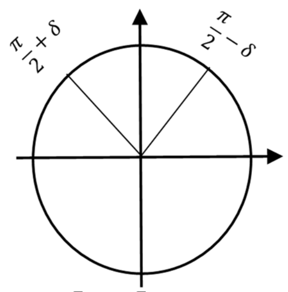

Figure 2. The arc $(\frac{\pi}{2}-\delta, \frac{\pi}{2}+\delta)$ is the critical region.

$H_0: \mu = 0$ versus $H_1: \mu = \mu_1$, we have

$$\frac{f_1(\boldsymbol{\theta^w})}{f_0(\boldsymbol{\theta^w})} = \frac{e^{(\kappa+\lambda)\sum_{i=1}^{n}\cos(\theta_i^w-\mu_1)}}{e^{(\kappa+\lambda)\sum_{i=1}^{n}\cos\theta_i^w}} = e^{(\kappa+\lambda)\sum_{i=1}^{n}[\cos(\theta_i^w-\mu_1)-\cos\theta_i^w]} > d$$

for some $d > 0$. Therefore, we have

$$\sum_{i=1}^{n}\sin(\theta_i^w - \frac{\mu_1}{2}) = R\sin\left(\bar{\theta}^w - \frac{\mu_1}{2}\right) > d$$

As $R > 0$, hence the critical region is given by (Fig. 2).

References

1. Mardia KV. How new shape analysis and directional statistics are advancing modern life-sciences. In: Int Stat Inst: Proc. 58th World Statistical Congress. 2011.
2. Bulbert MW, Page RA, Bernal XE. Danger comes from all fronts: predator-dependent escape tactics of Túngara frogs. PLoS One. 2015;10(4):e0120546.
3. Kaas AL, Van Mier HI. Haptic spatial matching in near peripersonal space. Exp Brain Res. 2006;170(3):403–413.
4. Mardia KV, Hughes G, Taylor CC, Singh H. A multivariate von Mises distribution with applications to bioinformatics. Canad J Stat. 2008;36(1):99–109.
5. Gill J, Hangartner D. Circular data in political science and how to handle it. Polit Anal. 2010;18(3): 316-336.
6. Mechsner F, Kerzel D, Knoblich G, Prinz W. Perceptual basis of bimanual coordination. Nature. 2001;414(6859):69-73.
7. Mechsner F, Stenneken P, Cole J, Aschersleben G, Prinz W. Bimanual circling in deafferented patients: Evidence for a role of visual forward models. J Neuropsychol. 2007;1(2):259-282.
8. Postma A, Zuidhoek S, Noordzij ML, Kappers AML. Keep an eye on your hands: on the role of visual mechanisms in processing of haptic space. Cogn Process. 2008;9(1):63-68.
9. Baayen C, Klugkist I, Mechsner F. A test of order-constrained hypotheses for circular data with applications to human movement science. J Mot Behav. 2012;44(5):351–363.
10. Gurtman MB, Pincus AL. The circumplex model: Methods and research applications. In: Handbook of Psychology. In: Handbook of Psychology. Wiley Online Library. 2003.
11. Gurtman MB. Exploring personality with the interpersonal circumplex. Soc Personal Psychol. 2009;3(4):601-619.
12. Leary T. Interpersonal Diagnosis of Personality. Ronald Press. New York. 1975.
13. Fisher RA. The effects of methods of ascertainment upon the estimation of frequencies. Ann Eugen. 1934;6:13-25.
14. Rao CR. On discrete distributions arising out of methods of ascertainment. Sankhya 1965;27:311-324.
15. Rao CR. Weighted distributions arising out of methods ascertainment: What population does a sample represent? in A Celebration of Statistics, eds. Atkinson AC, Fienberg SE, Springer-Verlag, New York. 1985.
16. Mardia KV. Statistics of Directional Data. Academic Press, New York. 1972.
17. Stephens MA. Tests for the von Mises distribution. Biometrika. 1969;56:149-60.
18. Shahsanaei F. Chinipardaz R. Statistical inference of weighted sampling in circular data, Submitted for publication. 2022.