Statistica Sinica Preprint No: SS-2017-0501	
Title	Universal and Efficient Tests for Homogeneity of Mean Directions of Circular Populations
Manuscript ID	SS-2017-0501
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0501
Complete List of Authors	Ashis SenGupta and Hemangi V. Kulkarni
Corresponding Author	Hemangi V. Kulkarni
E-mail	kulkarni.hemangi@gmail.com
Notice: Accepted version subject to English editing.	

Universal and Efficient Tests for Homogeneity of Mean Directions of Circular Populations

Ashis SenGupta and Hemangi V. Kulkarni

Indian Statistical Institute, Kolkata;Augusta University, Augusta, GA, USA. and Shivaji University, Kolhapur

Abstract: We aim to develop an efficient test for homogeneity of mean directions of several independent circular populations (ANOMED), which can be universally implemented. Current tests for ANOMED are available only under highly concentrated and/or large-sized groups. We intend to fill the gap for a usable test under highly dispersed and/or small to medium size groups. Focusing on the popular von Mises distribution, a simple elegant test statistic is derived under homogeneous concentrations across the groups. The hurdle of the non location-scale nuisance parameter κ is overcome by adopting a new approach through the integrated likelihood ratio test (ILRT). Second order accurate asymptotic Chi-square distribution of the ILRT is established. This test notably outperforms the existing ones under small to moderate size and highly dispersed (small κ) groups, which is precisely the parametric region of prime concern, where the earlier tests were unusable or unsatisfactory. It also outperforms the popular Watson-Williams test under highly concentrated small size groups and continues to compete well against its best competitors otherwise, and hence can be universally used in all situations. ILRT extends naturally under heterogeneous concentrations and is amenable to elegant generalizations to a rich variety of circular populations as well as to higher dimensions, i.e. to distributions on the sphere and hypersphere. The test is illustrated through three real-life data sets.

Key words and phrases: Circular ANOVA, Circular normal distribution, Generalized von Mises distribution, Integrated likelihood ratio tests, Watson-Williams test, Batschelet distribution.

1. Introduction

Observations on angular movements or displacements, and on directional propagations on the plane commonly constitute circular data. Strictly periodic occurrences, rhythmic activities and compositional data can also be cast in its arena. Analytically, any data that can be mapped uniquely into the circumference of a unit circle is defined as circular data. Analysis of such data differs markedly from those for linear ones due to the disparate topologies between the line and the circle. Readers are referred to Mardia and Jupp (2000) (MJ), Jammalamadaka and SenGupta (2001) (JS) and Fisher (1993) for further details.

Often a situation demands the comparison of the mean directions of several independent populations - see e.g. Lozano (2016) and Shay *et al.* (2016) for recent applications among numerous others. We will refer to such comparisons as Analysis of Mean Directions (ANOMED). The present work develops efficient test procedures for ANOMED under the very popular von Mises (vM) or circular normal distribution.

An overview (section 2) of the existing literature on ANOMED for vM reveals that the tests are available either for highly concentrated data (Watson, 1956) or for large samples (e.g. see the corresponding likelihood ratio test (LRT) in MJ). Under the same situations, some useful references on ANOMED are: Beran and Fisher (1998) for bootstrap based pairwise comparisons among mean directions, Larsen *et al.* (2002) for likelihood ratio based improved tests for the two sample problem, and SenGupta and Roy (2011) for analysis of deviance based approach with vM and wrapped Cauchy distributions.

In the present scenario, it appears that for vM, no satisfactory test for highly dispersed data (small concentration parameter) and small to moderate group sizes is available in the literature. However, there abound real-life data on such a situation, which possibly is more realistic, in diverse areas of applied research. An example attesting to this fact is also given in this paper. The present work attempts to fill this gap by developing an integrated likelihood ratio test (ILRT), which eliminates the nuisance concentration parameter κ by integrating it out from the likelihood function through a suitably chosen weight function. The second order accurate asymptotic Chi-square distribution for the ILRT is derived. Extensive simulation based comparison reveals its notable out-performance particularly under small concentration parameters as desired and equally well performance over its best competitors otherwise, rendering it universally applicable. Tests for ANOMED under Generalized von Mises (GvM) and Batschelet distributions are outlined. A version of ILRT for heterogeneous concentrations across the groups is also developed. The new test is illustrated with real data sets representing varied real situations.

In the remainder of the paper, Section 2 summarizes the existing methods. The method ILRT and its asymptotic distribution is introduced in section 3. Section 4 develops the analogue of ILRT under unequal concentrations, derives its asymptotic distribution and briefs its extensions to other distributions. Section 5 illustrates the tests with real data sets. Section 6 presents concluding remarks.

2. Preliminaries and review

2.1 Preliminaries

The angular observations θ_{ij}, $i = 1, ..., p$, $j = 1, ..., n_i$; $\theta_{ij} \in (0, 2\pi]$ are assumed to have come from the von-Mises or also termed circular normal distribution, with pdf

$$f(\theta_{ij}) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa . \cos(\theta_{ij} - \mu_i)\},$$

where $\kappa > 0$ is the concentration parameter, $\mu_i \in (0, 2\pi]$ is the mean direction for the i^{th} group and $I_0(\kappa)$ is the modified Bessel function of the first kind with order zero. The maximum likelihood estimate (MLE) of the i^{th} group mean direction μ_i is given by $\theta_{i.}$, the quadrant specific sample mean direction (JS, p. 13). Let $C_i = \sum_{j=1}^{n_i} \cos(\theta_{ij})$, $S_i = \sum_{j=1}^{n_i} \sin(\theta_{ij})$ and $n = \sum_{i=1}^{p} n_i$. The length of the resultant vector for the i^{th} group in its two equivalent forms is $R_i = (C_i^2 + S_i^2)^{\frac{1}{2}} = \sum_{j=1}^{n_i} \cos(\theta_{ij} - \theta_{i.})$. The MLE $\hat{\kappa}_1$ of κ is a solution of the equation $A(\hat{\kappa}_1) = (\sum_{i=1}^{p} R_i)/n$, where $A(.) = I_1(.)/I_0(.)$.

The mean direction $\theta_{..}$ of the combined sample and its corresponding resultant length R are obtained similarly based on the combined sample, on replacing C_i and S_i by $C = \sum_{i=1}^{p} C_i$ and $S = \sum_{i=1}^{p} S_i$ respectively. The standardized lengths for i^{th} group and combined sample are respectively given by $\bar{R}_i = R_i/n_i$ and $\bar{R} = R/n$.

The problem is to test $H_0 : \mu_1 = ... = \mu_p$, versus at least one inequality in mathematical terms. Under H_0, the MLE of μ_0 is $\theta_{..}$ while that of κ is $\hat{\kappa}_0$, where $A(\hat{\kappa}_0) = \bar{R}$. The existing tests for ANOMED under vM are described below.

2.2 Existing methods

Existing literature addresses the problem of ANOMED for high concentration (large κ) or for large sample sizes. Correspondingly, there are familiar tests of four types (see e.g. MJ, Chapter 10): The two high concentration tests, namely Watson-Williams (WW) and Harrison, Kanji, Gadsden (HKG) tests and two LRT based large sample tests.

1. WW test with a multiplicative correction:

$$T_{WW} \equiv (1 + \frac{3}{8\hat{\kappa}_0})\frac{(n-p)SS_B}{(p-1)SS_W} \quad \sim F_{p-1,n-p},$$

for large κ, where $SS_W = 2\kappa(n - \sum_{i=1}^{p} R_i)$, and $SS_B = 2\kappa(\sum_{i=1}^{p} R_i - R)$. The corrective adjustment $(1 + \frac{3}{8\hat{\kappa}_0})$ is suggested by Stephens (1972) and is recommended to be used for $\hat{\kappa}_0 > 2$.

2. HKG with a multiplicative correction:

$$T_{HKG} \equiv (1 - 1/(5\hat{\kappa}_0) - 1/(10\hat{\kappa}_0^2))\frac{(n-p)SS_{Tr}}{(p-1)SS_E} \quad \sim F_{p-1,n-p},$$

for large κ, where $SS_{Tr} = (\sum_i n_i\bar{R}_i^{\,2} - n\bar{R}_{..}^{\,2}), SS_E = (n - \sum_i n_i\bar{R}_i^{\,2})$,

3. LRT:

$$T_{LRT} = 2[n\{\log I_0(\hat{\kappa}_0) - \log I_0(\hat{\kappa}_1)\} + \hat{\kappa}_1 \sum_i R_i - \hat{\kappa}_0 R] \quad \overset{a}{\sim} \chi^2_{p-1}.$$

4. Anderson and Wu test (AW):

Anderson and Wu (1995) suggested to use $\hat{\kappa}_0$ in place of $\hat{\kappa}_1$ in T_{LRT} with the same asymptotic chi-square distributional assumption.

Here $\overset{a}{\sim}$ refers to an asymptotic distribution.

While the situations of highly dispersed and/or small to medium size groups are commonly encountered in practice, unfortunately existing tests either are not applicable or fail to perform well there. Please see section 3.3 for a rigorous discussion of this point. The main emphasis of the current work is to fill this gap. Our attempt is to develop a test that should work reasonably well uniformly for all situations. Elimination of the nuisance concentration parameter to improve the quality of LRT based tests was felt to be a reasonable step in that direction.

In the next section we develop an integrated likelihood test, ILRT, for ANOMED. Its second order asymptotic Chi-square distribution is derived. A detailed assessment of the test reported in section 3.3 exhibited uniformly satisfactory performance and out-performance over the tests listed above in the aforementioned regions.

3. THE PROPOSED INTEGRATED LIKELIHOOD RATIO TEST (ILRT)

3.1 The integrated likelihood ratio approach

Consider a likelihood function $L(\boldsymbol{\psi}, \boldsymbol{\lambda})$ where $\boldsymbol{\psi}$ is the parameter of interest and $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ is the nuisance parameter. The likelihood inference about $\boldsymbol{\psi}$ is often based on a pseudo-likelihood function $L_{\boldsymbol{\psi}}$ obtained by eliminating $\boldsymbol{\lambda}$ in a suitable way which maintains the properties similar to those of a regular likelihood. The most popular one is the profile likelihood (PL) L_p (and its modifications) which replaces the nuisance parameter by $\hat{\boldsymbol{\lambda}}_{\boldsymbol{\psi}}$, the maximizer of L with respect to $\boldsymbol{\lambda}$ under fixed $\boldsymbol{\psi}$. However the PL has some drawbacks. Maximization over $\boldsymbol{\Lambda}$, can often be formidable under large number of nuisance parameters. See also example 2 (yielding $'0'$ as the PL based MLE for population variance under every observable data set), example 3 (yielding a strange profile likelihood, rapidly growing to ∞ as the parameter $\theta \to \infty$ or $-\infty$ depending on the sign of the sample mean) and example 4 (PL is nearly useless for inference being nearly constant over a huge range of the parameter space) of Berger *et al.* (1999) for other undesirable situations.

Under this background, the "averaging" effect produced by an integrated likelihood (IL) to be discussed next is expected to produce a better summary of the original likelihood than the "maximization" involved in the profile likelihood. We refer to Berger *et al.* (1999) for a critical discussion about pseudo likelihoods where use of IL is strongly recommended from several perspectives, including accounting for nuisance parameter uncertainty. For further insights, the reader is also referred to e.g. Kalbfleisch and Sprott (1970) and Liseo (1993) among others. An IL is of the form

$$\bar{L}(\boldsymbol{\psi}) = \int_{\boldsymbol{\Lambda}} L(\boldsymbol{\psi}, \boldsymbol{\lambda}) \cdot \Pi(\boldsymbol{\lambda}|\boldsymbol{\psi}) \, d\boldsymbol{\lambda}. \tag{1}$$

Here Π is a non-negative weight function on $\boldsymbol{\Lambda}$ making the above integral convergent for every fixed $\boldsymbol{\psi}$.

As $\bar{L}$ depends only on the data and the parameter of interest $\boldsymbol{\psi}$, it can be used like a standard likelihood function for all likelihood based inference procedures. However, a proper choice of Π to produce good inference procedures is an important issue which remains unresolved under multiple parameters of interest like the present inference problem of ANOMED.

Effective IL based inference procedures are considered by Chamberlain (2007), Ghosh *et al.* (2006), Malley *et al.* (2003), among others. Severini (2007, 2010, 2011) gives a thorough development of inference procedures about a scalar parameter of interest ψ, particularly when the nuisance parameters $\boldsymbol{\lambda}$ and a scalar parameter of interest ψ are orthogonal, i.e. the expected values of the mixed derivatives of the log likelihood function with respect to $\boldsymbol{\lambda}$ and ψ are zeros. In this case the impact of the choice of Π is quite low more effectively when Π does not depend on $\boldsymbol{\psi}$, for moderate to large samples. However, parameter orthogonality is not a necessary condition for ILRT to produce good inference procedures.

In the following, the ILRT statistic is developed for ANOMED under equally dispersed vM distributions. Its second order asymptotic Chi-square distribution is derived. An extensive simulation-based assessment of its performance is carried out in section 3.3.

3.2. ILRT under equal concentration parameters:

Referring to section 2.1, the likelihood function is

$$L(\boldsymbol{\mu}|\boldsymbol{\theta},\kappa) \ = \ \frac{1}{I_0(\kappa)^n}\exp\left[\kappa\{\textstyle\sum_{i=1}^{p}\sum_{j=1}^{n_i}(\cos(\theta_{ij}-\mu_i))\}\right], \quad \kappa>0,\ \mu_i\in[0,\ 2\pi), \quad \forall i.$$

Here $\boldsymbol{\theta} = (\theta_{11},\theta_{12},...,\theta_{1n_1},..,\theta_{p1},...,\theta_{pn_p})$ is the vector of all observations and $\boldsymbol{\psi} = \boldsymbol{\mu} = (\mu_1,...,\mu_p)$ is the vector parameter of interest. The choice of the prior

$$\Pi(\kappa) = I_0(\kappa)^n \exp(-n\kappa)\ \kappa^{a_n-1},\ \kappa>0, \tag{2}$$

is motivated by its success in attaining a simple closed form of the IL, after eliminating the normalizing constant $I_0(\kappa)^{-n}$ and choosing the exponent $\exp(-n\kappa)$ to make the resulting integral convergent for all observed data sets and mean directions. Nevertheless, we are keen to keep it free from the parameter of interest $\boldsymbol{\psi}$. This choice, together with the parameter orthogonality between $\boldsymbol{\psi}$ and κ, facilitate to bestow the resulting ILRT with the desired second order properties as seen in the proof of Theorem 1 (see also Severini (2007)). The term $\kappa^{a_n/2-1}$, whose exponent produces the scaling factor a_n in the resulting ILRT statistic was invoked to attain a non degenerate limiting distribution. In line with the Welch-Satterthwaite technique, an initial guess of $a_n = n-1$ was based on matching the simulated means (first moments) of T_{ILRT} to $p-1$, the expected values of the desired asymptotic χ^2 distribution, under large group size and large κ. (More precisely, $n_1 = 100$ and $\kappa = 15$ were taken as representatives of large group sizes and large concentrations respectively. Then, the ratio of $p-1$ to the simulated mean (based on 500000 simulations) of the RHS of (3), excluding the $(n-1)$ term (which is the simulated value of the a_n term), was regressed on the total sample size n for $p = 2(1)8$. The value of a_n was further fine-tuned for its modest dependence on the unknown κ through multiplicative adjustments as suggested in section A.1 of Appendix-A. Finally, integrating $L.\Pi$ over $\kappa \in (0, \infty)$ results in the integrated likelihood function

$$\bar{L}(\boldsymbol{\mu}|\boldsymbol{\theta}) \propto \left[n - \sum_{i=1}^{p}\sum_{j=1}^{n_i}\cos(\theta_{ij}-\mu_i)\right]^{-(n-1)/2}.$$

The integrated MLE's obtained by maximizing $\bar{L}$ with respect to $\boldsymbol{\mu}$ under the null and the alternative hypotheses coincide with the usual ones $\bar{\psi}_0 = \bar{\mu}_0 = \theta_{..}$ and $\bar{\boldsymbol{\psi}}_1 = \bar{\boldsymbol{\mu}}$; $\bar{\mu}_i = \theta_{i.}$, $i = 1, ..., p$ respectively (see section 2.1). The resulting IL ratio is

$$\bar{\lambda} = \frac{sup_{\Theta_0}\bar{L}(\mu|\boldsymbol{\theta})}{\sup_{\Theta_1}\bar{L}(\boldsymbol{\mu}|\boldsymbol{\theta})} \quad = [(n - \sum_{i=1}^{p} R_i)/(n-R)]^{(n-1)/2},$$

where $\Theta_1 = \{\boldsymbol{\mu} : \mu_i \in (0, 2\pi], i = 1, ..., p\}$ and $\Theta_0 = \{\mu(1, 1, ..., 1)_{p\times 1}, \quad \mu \in (0, 2\pi]\ \}$ is the subset of Θ_1, where all components of $\boldsymbol{\mu}$ are equal to μ. The proposed ILRT statistic $-2\log\bar{\lambda}$ equals

$$T_{ILRT} = -(n-1)\ \log[(n - \sum_{i=1}^{p} R_i)/(n-R)]. \tag{3}$$

The asymptotic χ^2 distribution of T_{ILRT} is stated in Theorem 1 below. Throughout this work the parameter space for κ is assumed to be $(0,\infty)$. (The case $\kappa = 0$ is excluded, being a uniform distribution over $[0,\, 2\pi)$).

Theorem 1: The asymptotic distribution of T_{ILRT} is χ^2_{p-1}.

Proof of Theorem 1. Let $C(\boldsymbol{\psi}) = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\cos(\theta_{ij}-\mu_i)$. Writing $L(\kappa)$ for $L(\kappa|\boldsymbol{\mu};\boldsymbol{\theta})$, let

$$\begin{aligned} h(\kappa) &= -\frac{1}{n}\log(L(\kappa)) \\ &= \log(I_0(\kappa)) - \frac{\kappa}{n}C(\boldsymbol{\psi}). \end{aligned}$$

It is easily seen that the partial derivatives of h with respect to κ are

$$\begin{aligned} h'(\kappa) &= -\frac{C(\boldsymbol{\psi})}{n} + A(\kappa); \\ h^{(j)}(\kappa) &= A^{(j-1)}(\kappa); \quad j = 2,3,.... \quad , \end{aligned} \tag{4}$$

$A^{(j)}$ being the j^{th} derivative of $A(\kappa)$ with respect to κ.

First consider the case of large concentrations, $\kappa > 1$. Here $A(\kappa)$ and its j^{th} derivative can be shown to be piece-wise well approximated to $O(10^{-3})$ by

$$\begin{aligned} &A(\kappa) \approx c + b/\kappa; \\ &A^{(j)} \approx b\,(-1)^{(j)}\ j!/\kappa^{j+1};\ j = 1,2,\ldots\,. \end{aligned} \tag{5}$$

The constant c varies slightly across the pieces and is almost zero for large concentrations, while the slope b is very close to 2. (see Table 2 of section A.2, Appendix-A for details. See also A.13 of Appendix 1 of JM for another approximation). Taylor expansion of h about $\hat{\kappa}_{\boldsymbol{\psi}} \equiv \hat{\kappa}$, with equations (4) and (5), and the fact that $h'(\hat{\kappa}) = 0$, give

$$nh(\kappa) = -\log(L(\hat{\kappa})) + \frac{1}{2}A'(\hat{\kappa})u^2 + \left\{\frac{-2\ u^3}{\sqrt{n}3(\hat{\kappa})^3} + \frac{2\ u^4}{n4(\hat{\kappa})^4}\right\} + r_n(u),$$

where $u = \sqrt{n}(\kappa - \hat{\kappa})$ and $r_n(u)$ is $O(n^{-1.5})$.

Note that the prior Π is continuously differentiable. Then, applying the expansion of e^{-x} to the 3^{rd} term (inside the curly bracket) and using Taylor's expansion of $\Pi(\kappa)$ about $\hat{\kappa}$, gives

$$\begin{aligned} L(\kappa).\Pi(\kappa) &= \exp\{-nh(\kappa)\}\Pi(\kappa) \\ &= \frac{L(\hat{\kappa})}{\sqrt{n\ |A'(\hat{\kappa})|}}.\ \sqrt{n\ |A'(\hat{\kappa})|}\exp\{-A'(\hat{\kappa})u^2/2\} \\ &.\left\{1 - \left[\frac{-2\ u^3}{3\sqrt{n}(\hat{\kappa})^3} + \frac{2\ u^4}{4n(\hat{\kappa})^4}\right] + \frac{1}{2}\left[\frac{-2\ u^3}{3\sqrt{n}(\hat{\kappa})^3} + \frac{2\ u^4}{4n(\hat{\kappa})^4}\right]^2 +\ R_{1n}(\kappa, \hat{\kappa})\right\} \\ &\left\{\Pi(\hat{\kappa}) + \frac{1}{\sqrt{n}}\Pi'(\hat{\kappa})u + \frac{1}{2n}\Pi''(\hat{\kappa})u^2 + \frac{1}{6n\sqrt{n}}\Pi^{(3)}(\hat{\kappa})u^3\ +\ R_{2n}(\kappa, \hat{\kappa})\right\}, \end{aligned} \tag{6}$$

where R_{1n} and R_{2n} are of $O(n^{-2})$. First compute the product of the two bracketed terms in the RHS of (6) and multiply the resulting terms by $\sqrt{n\ |A'(\hat{\kappa})|}\exp\{-\frac{1}{2}A'(\hat{\kappa})u^2\}$, which is proportional to a normal density with mean zero and standard deviation $A'(\hat{\kappa})^{-1/2}$. Next, integrate term by term with respect to κ and note that $d\kappa = \sqrt{n}\ du$.

It is then seen that in the RHS of (6), the integrals involving powers of u are proportional to the raw moments of a normal distribution with mean zero and standard deviation $A'(\hat{\kappa})^{-1/2}$. Use of (5) with the approximation $b \approx 2$ makes the $(2j)^{th}$ raw moment $\mu'_{2j} = 2\hat{\kappa}^{2j}(2j)!/(2^j j! 2^j)$ - a constant multiple of $\hat{\kappa}^{2j}$, while the odd order moments vanish. On ignoring the $O(n^{-2})$ terms, the RHS of (6) becomes

$$\int L(\kappa).\Pi(\kappa)d\kappa \propto L_A(\hat{\kappa})\Bigg\{\Pi(\hat{\kappa}) + \frac{\Pi''(\hat{\kappa})\mu'_2}{2n} + \frac{\Pi'(\hat{\kappa})\mu'_4}{3n\hat{\kappa}^3} - \frac{\Pi(\hat{\kappa})\mu'_4}{2n\hat{\kappa}^4} + \frac{\Pi(\hat{\kappa})\mu'_6}{9n\hat{\kappa}^6} + O(n^{-2})\Bigg\}, \tag{7}$$

where $L_A(\boldsymbol{\psi}) \propto L(\boldsymbol{\psi}, \hat{\kappa}_{\boldsymbol{\psi}})|l_{\kappa\kappa}|^{-1/2}_{\{\kappa=\hat{\kappa}\}}$ is the Cox and Reid (1987) adjusted profile likelihood; $|l_{\kappa\kappa}|_{\{\kappa=\hat{\kappa}\}} = nA'(\hat{\kappa})$, $l_{\kappa\kappa}$ being the second order partial derivative of the log likelihood l with respect to κ. Furthermore, on account of (5), $\Pi'(\kappa) \approx n\Pi(\kappa)(2.5/\kappa - \theta)$ and $\Pi''(\kappa) \approx \Pi(\kappa)n^2(2.5/\kappa - \theta)^2$ where $\theta = 1 - c$ is close to 1 (for large κ, c is very close to 0, see Table 2, section A.2, Appendix-A for more details). Consequently, ignoring the $O(n^{-2})$ terms, all the above observations finally lead to

$$\bar{L}(\boldsymbol{\psi}) = \int L\ .\Pi(\kappa)d\kappa \propto L_A(\hat{\kappa}).\Pi(\hat{\kappa})g(\hat{\kappa})\{1 + O(n^{-1.5})\} \tag{8}$$

where for every fixed n, both g and Π are finite and continuous in κ. Taking logarithms and denoting the log likelihoods by l, give,

$$\bar{l}(\boldsymbol{\psi}) = l_A(\boldsymbol{\psi}) + log(\Pi(\hat{\kappa}_{\boldsymbol{\psi}})) + log(g(\hat{\kappa}_{\boldsymbol{\psi}})) + log(\{1 + O(n^{-1.5})\}).$$

Recall from section 2.1 that $\bar{\boldsymbol{\psi}}_{\mathbf{1}} = \hat{\boldsymbol{\mu}}$, $\bar{\psi}_0 = \hat{\mu}_0$, are the usual MLEs of $\boldsymbol{\mu}$, while those of κ are $\hat{\kappa}_1 = \sup_{\boldsymbol{\psi}} \hat{\kappa}_{\boldsymbol{\psi}}$ and $\hat{\kappa}_0$ under H_i, $i = 1, 0$ respectively. The resulting ILRT statistic is,

$$-2\log\bar{\lambda} = -2(l_A(\hat{\boldsymbol{\mu}}) - l_A(\hat{\mu}_0)) + 2log(\Pi(\hat{\kappa}_1)/\Pi(\hat{\kappa}_0)) + 2log(g(\hat{\kappa}_1)/g(\hat{\kappa}_0)) + O_P(n^{-1.5}). \tag{9}$$

Note that Π (by our choice) and g depend on the parameter of interest $\boldsymbol{\psi}$ only through $\kappa_{\boldsymbol{\psi}}$. Additionally, $E(\partial^2 l/\partial\kappa\partial\mu_i) = 0$, $i = 1, ..., p$, (see MJ) so that the nuisance parameter κ is orthogonal to the parameter of interest $\boldsymbol{\psi} = \boldsymbol{\mu}$. Consequently, (see section 2.2, result (iv) of Cox and Reid (1987)), $\hat{\kappa}_{\boldsymbol{\psi}}$, is less sensitive to the variation in $\boldsymbol{\psi}$ under the null and the alternative hypotheses. Additionally, both $\hat{\kappa}_1$ and $\hat{\kappa}_0$ being consistent for the same parameter κ, $|\hat{\kappa}_1 - \hat{\kappa}_0| = O_P(n^{-1})$. Together with the continuity of Π and g, this makes the terms $\log(\Pi(\hat{\kappa}_1)/\Pi(\hat{\kappa}_0))$, $\log(g(\hat{\kappa}_1)/g(\hat{\kappa}_0))$ and the middle term in equation (9), all of $O_P(n^{-1})$. These arguments finally lead to

$$T_{ILRT} = -2\log\bar{\lambda} = -2(l_A(\hat{\boldsymbol{\mu}}_1) - l_A(\hat{\mu}_0)) + O_P(n^{-1}),$$

with computational form given in equation (3). The asymptotic distribution of T_{ILRT} is as that of the adjusted profile log likelihood ratio, which is χ^2_{p-1}. The approximations involving κ in (5) (leading to the χ^2 distributional approximation) are very sharp for large κ, say $\kappa > 9$ but not so for $\kappa < 9$. A slight fine-tuning in the form of subtle multiplicative adjustments given in section A.1 of Appendix-A, which is based on the piece-wise partition for approximation of $A(\kappa)$, significantly improved the χ^2_{p-1} approximation for this case.

The case of small concentration can be dealt with on the similar lines on noting that in this case the function $A(\kappa)$ can be well approximated by $.107 + .46\kappa$, with maximum deviation of order 10^{-3} for $\kappa \in (0.1,\ 0.9)$. For $\kappa < 0.1$, the circular uniform distribution is recommended. Here the derivatives of $A(\kappa)$ and hence of $h(\kappa)$ of order greater than 1 all vanish, simplifying the RHS of (6) to a great extent. However, in this region the estimates of κ are likely to be more sensitive, moreover the derivatives of Π involve reciprocals of $\hat{\kappa}$. Hence, the RHS of (9) is expected to be more unstable and resulted in large observed sizes as revealed in a simulation study (not further discussed here for brevity). This problem was handled by ad-hoc multiplicative adjustments to the resulting ILRT, as developed in section A.1 of Appendix-A. $\square$

Remarks : (i) The aforementioned multiplicative adjustments controlled the sizes of the resulting tests very well without affecting its power function, as can be seen by the simulation study reported in the next subsection. These adjustments are used throughout in the sequel, including the performance assessment in section 3.3 as well as the real data analysis in section 5, and are strongly recommended in practice.

(ii) Although the results from Cox and Reid (1987) under parameter orthogonality used here were originally developed for a real-valued parameter of interest, these remain valid for vector-valued parameter of interest as long as orthogonality between the vector parameter of interest and the nuisance parameters holds, as is for the present case.

The next subsection attempts an extensive simulation based comparison among the ILRT and the existing tests.

3.3 Performance assessment.

A study based on 50000 simulated observations from von-Mises distributions on the circle is conducted to compare the size and power performances of the ILRT with the two high concentration tests and the two likelihood based tests reported in section 2.2. A large number of situations, namely group sizes n_1 = 15,20,30,40,60; concentration parameters κ=.25, .3:.04:.44; .45, .5:.1:1.5; 2:10; 15, 20, 40, 70, 100; and number of groups p = 2:8; were considered to form a fair representation of practical situations. The level of significance was fixed at the commonly used 5% level. Note that the parametric space under H_1 for the mean vectors $\boldsymbol{\mu}$, viz. $\Theta_1 = [0, 2\pi)^p$ is p-dimensional while the power function is a surface in $p+1$ dimensions. A comparative study of power surfaces among a group of several tests, in $p+1$ dimensions, particularly when p is large is a little formidable task as well as may lack visual clarity. To avoid such complications, noting the periodic nature of μ_i so that farthest components of $\boldsymbol{\mu}$ can be at most π distance away from each other, a systematic subset of $[0, 2\pi)^p$ is selected, viz. $\{\boldsymbol{\mu} = h/(p-1) * (0, 1, ..., p-1)\}$ scaling the vector $(0, 1, 2, ..., p-1)/(p-1) \in \Theta_1$ by a real positive number h varying h over the grid $0 : (\pi/6) : \pi$ for computation of powers so that the resulting collection of powers can be plotted against h as a function in two dimensions, henceforth referred to as 'the power function' in the sequel. This enabled a visually clear picture of the power functions and more clear comparison across various tests.

(A) *Size performance.*

Note that the case of $h = 0$ corresponds to the observed sizes of the respective tests. Box plots of the simulated sizes for the aforementioned four tests and the ILRT are shown in Figure 1. A careful assessment of the simulated sizes based on various graphical tools (not reported here for brevity) revealed the following prominent features:

i) The large outliers in the box-plots for WW basically emerged from small concentrations ($\kappa < 1$). The magnitude of outliers increased with number of groups (p), but group sizes (n_1) had almost no impact - a behavior consistent with the role of the large concentration behind WW's construction.

ii) Observed sizes of the other large concentration test HKG revealed a similar impact of κ and p (not of n_1), but in the opposite directions, that is tiny sizes (often very close to zero) increased with κ and stabilized to the desired level after κ became as large as 40. This in turn resulted in reduced powers, as revealed by Figures 2-5.

iii) The whiskers and outliers for LRT notably emerged under all the three factors:

small κ, large p and small n_1. Under small concentrations, the group sizes required to stabilize the sizes around the desired level 0.05 were as large as 60. For large concentrations, the convergence was relatively fast.

iv) AW exhibited a pattern similar to LRT but in the opposite direction as HKG did.

Clearly, based on the size performance, WW and LRT were practically un-usable under small concentrations and/or small size groups.

(B) ***Power performance***

To have a fair comparison among the tests, two versions of the power function were simulated:

(I) For an unbiased comparison among all available tests, the normalized power function was generated by multiplying the original uncorrected tests by the ratio of the respective theoretical χ^2 or F quantile to the simulated quantile of the uncorrected statistics, for the particular parameter combination under concern. This guaranteed the size of all cases to be exactly .05 making the power comparison unbiased. Figure 2 presents the gain in power over other tests by ILRT under the normalized power function at $h = \pi$ as a function of κ for small concentrations at various combinations of p and n_1. Figure 3 displays similar plots for very small concentrations ($\kappa =$.25 and .3) as a function of p.

(II) Secondly, viewing AW as the 'size-corrected' version of LRT, the actual (non-normalized) power functions of AW and HKG were compared with ILRT. Representative power functions for $p = 2$ and $p = 8$ groups are presented in Figures 4-5 for small concentrations and in Figures 6-7 for large concentrations. WW and LRT being unusable due to their large sizes under small concentrations, their power function is not included in Figures 4 and 5 while Figures 6 and 7 include all tests.

Both the normalized and non-normalized power functions showed similar pattern in excess power (gain) attained by ILRT. A careful observation of the Figures 2-7 strongly supports the following points:

small concentrations:

i) As targeted, a notable gain for ILRT was observed over its competitors AW and HKG under small concentrations, namely $\kappa < 1$, and more prominently for $\kappa < .5$ (see Figures 2-5).

ii) For very small κ, the gain was increasing with the number of groups (p) for fixed values of other parameters, (see Figure 3).

iii) For two groups and/or $\kappa < .5$ the gain over AW was uniformly more than that over HKG even under large group sizes (Figure 3 and first row of Figure 2). However, this behaviour reversed for large number of groups and $.5 < \kappa < 1$. Under large group sizes and κ in the neighbourhood of 1, the three tests performed almost equally (last two rows of Figure-2).

Large concentrations:

iv) Under medium κ $(1 < \kappa < 2)$ and very small group sizes, likelihood based tests surpassed WW, the gain increasing with p (first row of Figures 6-7).

v) Under large concentrations and large group sizes, all tests including the regular LRT performed almost equally well (Figure 7). However, for two groups the power of AW declined in the farthest region from the null hypothesis, i.e. at $h = \pi$, more prominently under large concentrations and small group sizes (Figure 6).

Remarks: (i) As the unadjusted version of ILRT and WW are functionally related, namely $ILRT = (n-1)\log(WW-1)$, the normalized power functions of the two were virtually the same, and so WW is not included in Figures 2 -3. However their distributions and hence cut-off points are different. As noted above, under small concentrations WW acquires unduly large sizes making it practically unusable.

(ii) In a nutshell, the major benefit of ILRT was felt under small concentrations and/ or small group sizes as desired. ILRT not only improved over LRT and AW but also was superior to all other tests in this scenario. It compared well in all other cases to the best performers and hence can be uniformly used under all situations irrespective of the magnitude of the observed values of κ and group sizes.

The next section discusses an extension of ILRT under heterogeneous groups. An

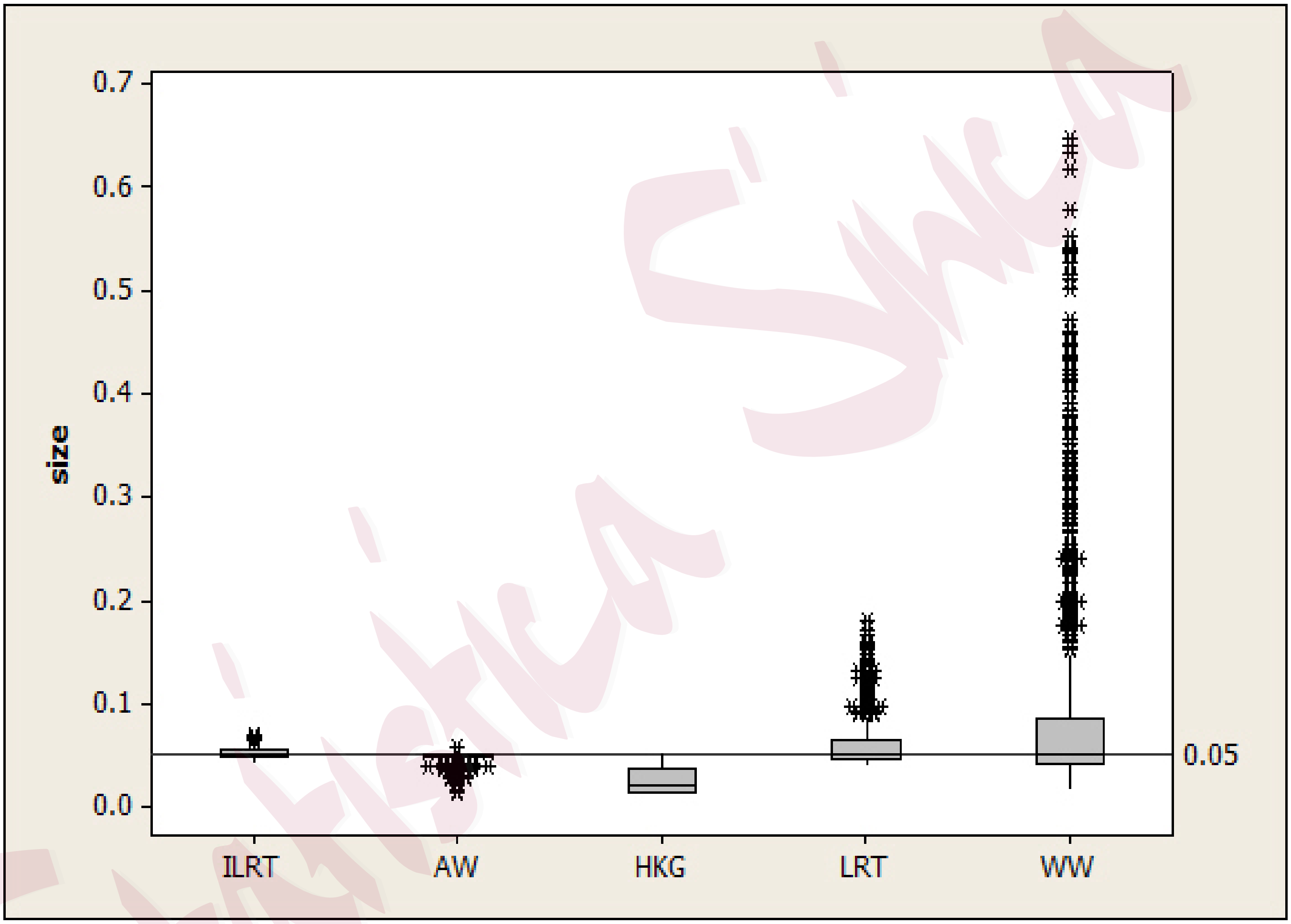

Figure 1: Box plots of simulated sizes of all tests for the parametric combinations reported in section 3.3.

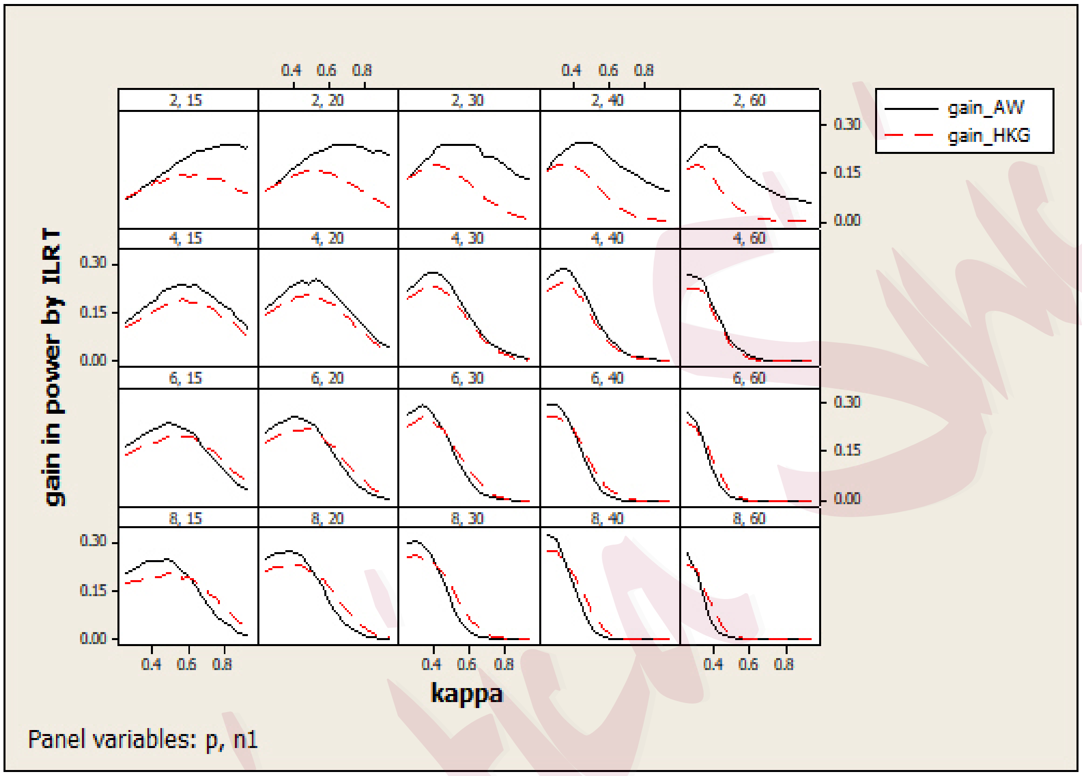

Figure 2: Gain in normalized power by ILRT over AW and HKG vs $\kappa(< 1)$ for various group sizes (n_1) and number of groups (p). Panel headings are values of the pairs p, n_1.

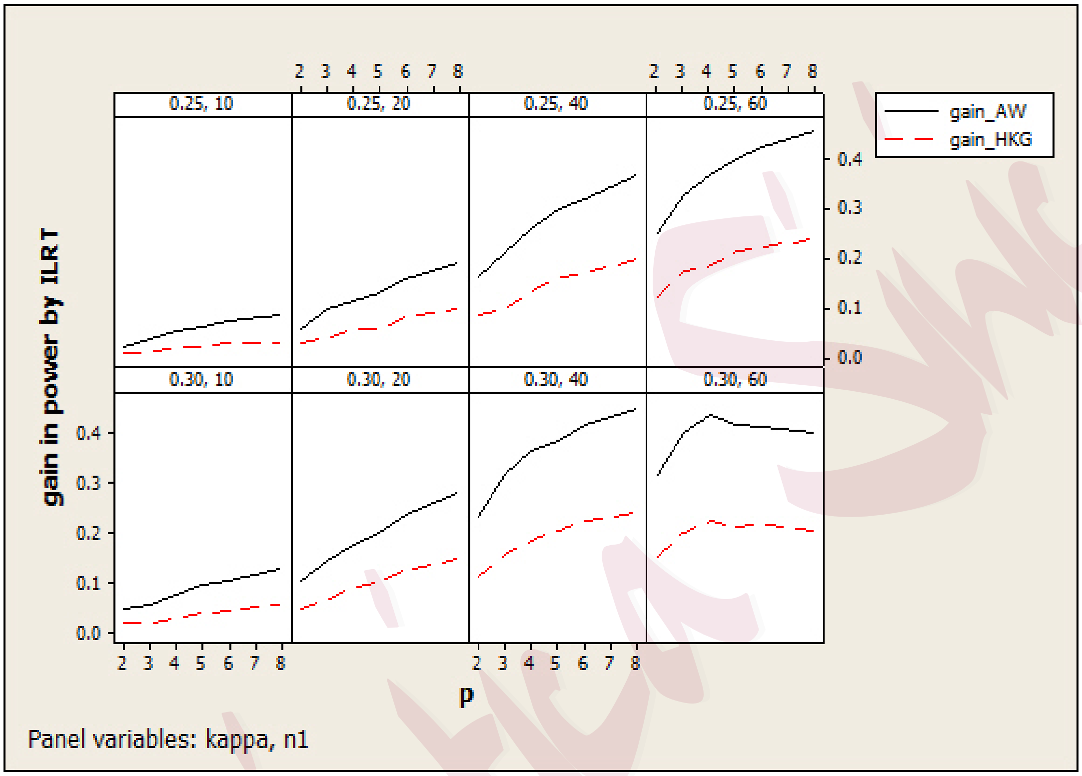

Figure 3: Gain in normalized power by ILRT over AW and HKG vs p for various group sizes (n_1) and concentrations $\kappa = .25$ and $.3$. Panel headings are values of the pairs κ, n_1.

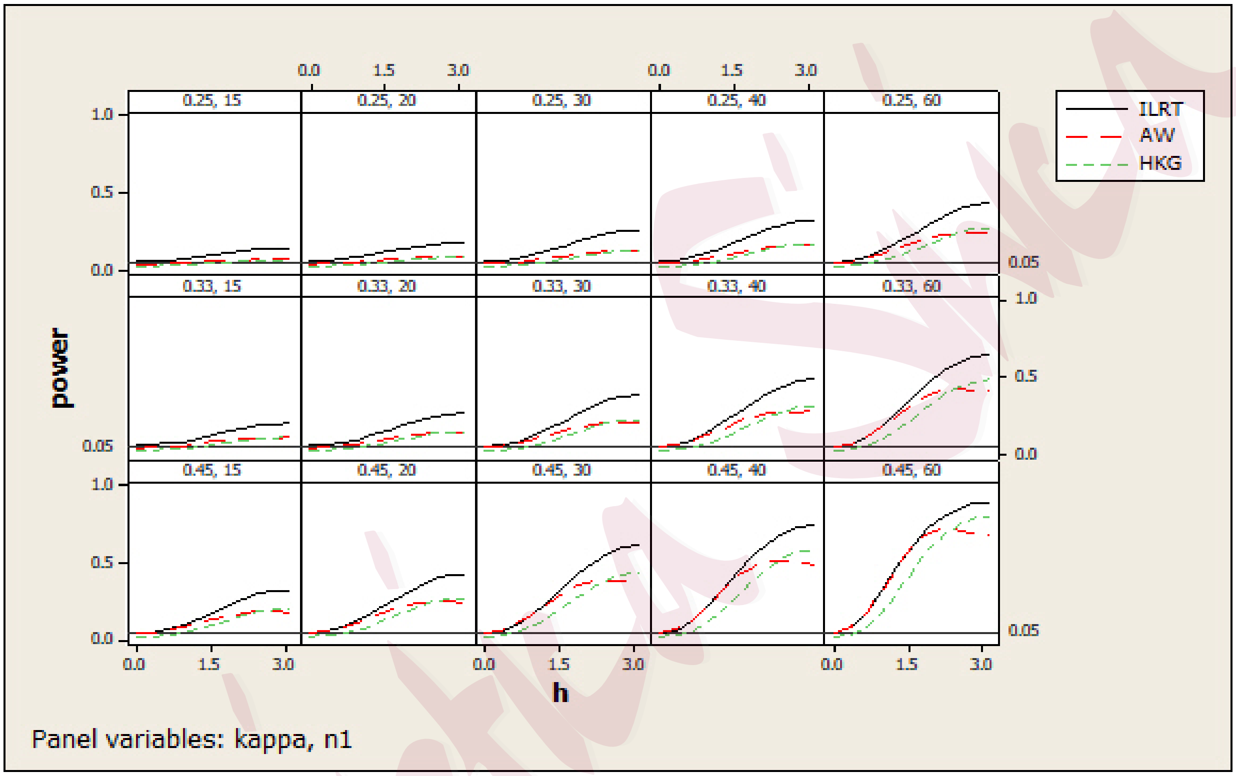

Figure 4: Simulated Power functions of ILRT, AW, HKG; under 2 groups and small κ. Panel headings are values of the pairs κ, n_1.

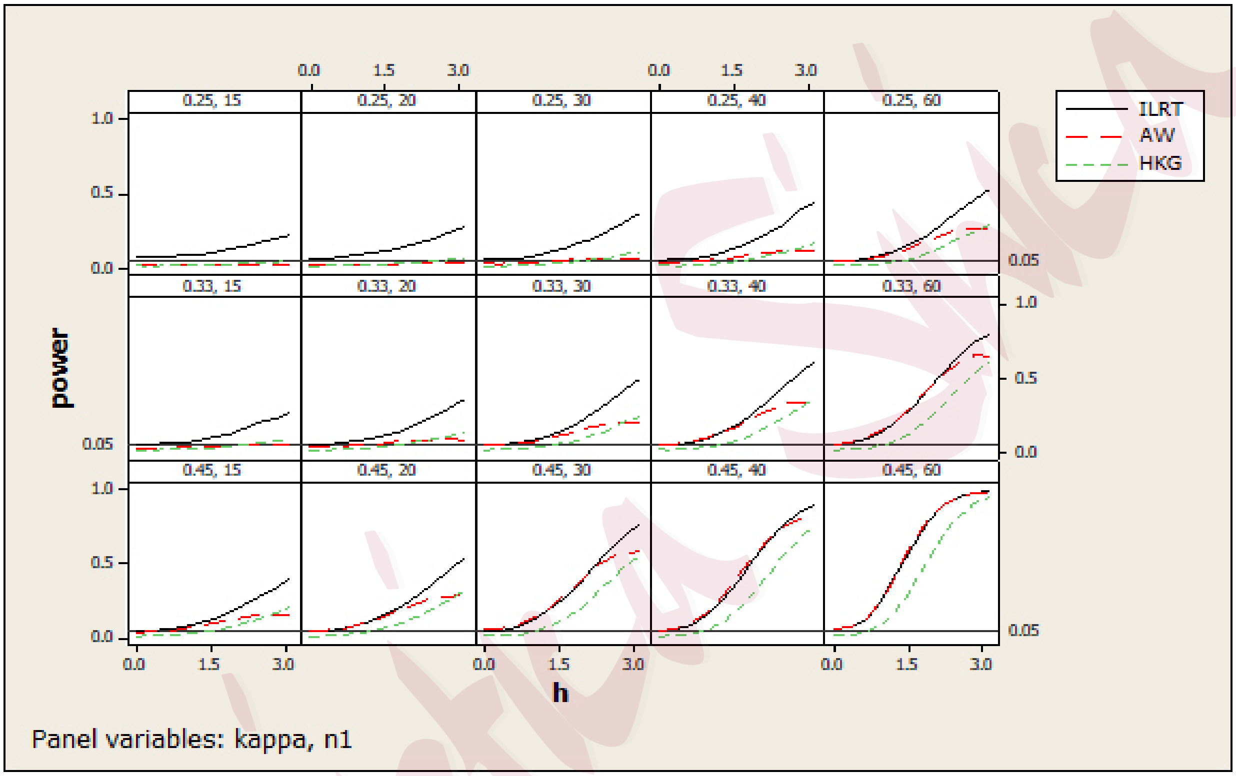

Figure 5: Simulated Power functions of ILRT, AW, HKG; 8 groups, small κ. Panel headings are values of the pairs κ, n_1.

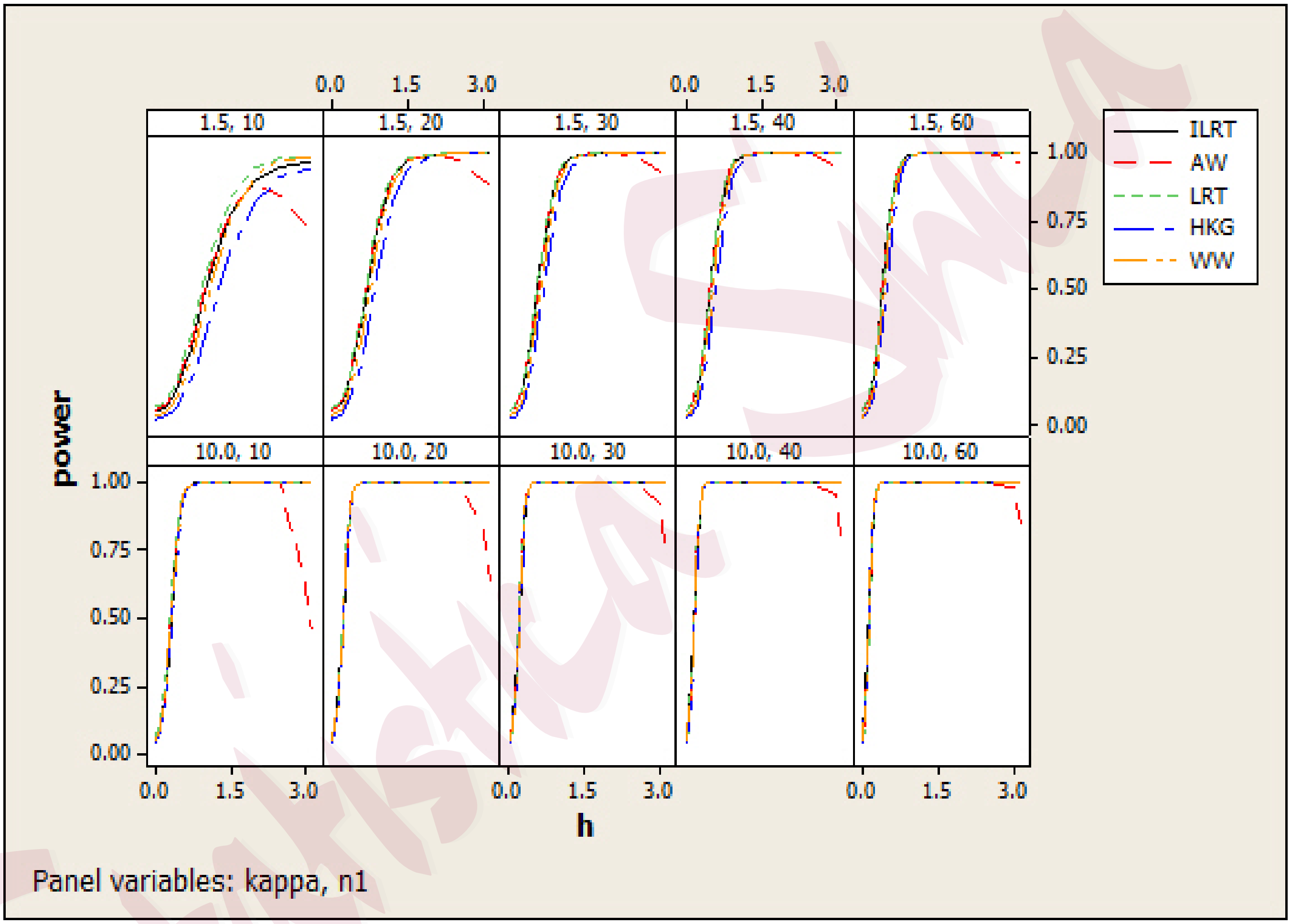

Figure 6: Simulated Power functions of ILRT,AW,LRT, HKG, WW;

2 groups, large κ. Panel headings are values of the pairs κ, n_1.

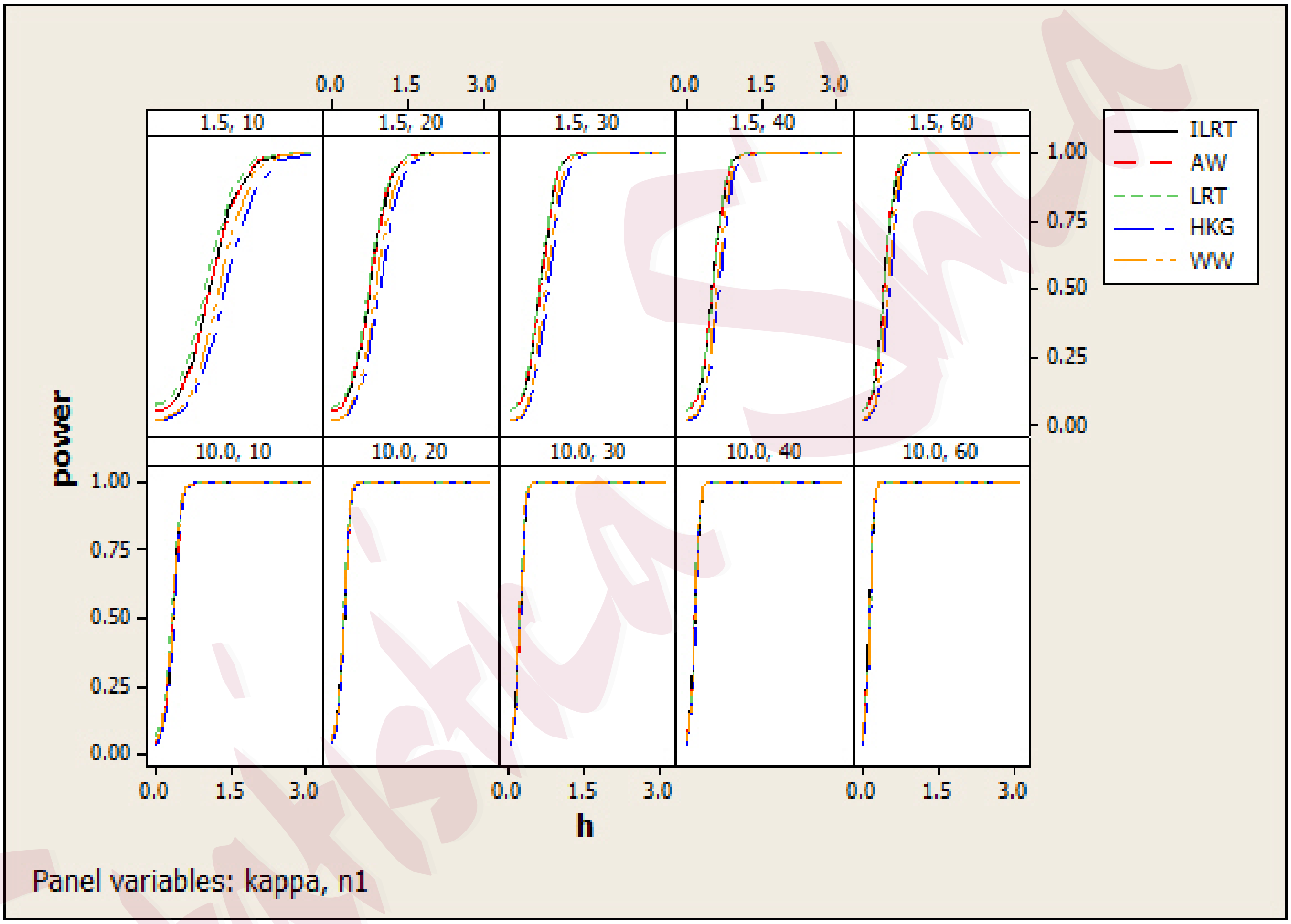

Figure 7: Simulated Power functions of ILRT,AW,LRT, HKG, WW;

8 groups, large kappa.Panel headings are values of the pairs κ, n_1.

extension to GvM and the Batschelet distributions is also outlined. It is to be noted that our approach gets adapted easily and elegantly for generalization of ANOMED to distributions on hyper-spheres.

4. Extensions of ILRT to other cases.

4.1 ILRT under unequal concentration parameters.

Often the p groups may follow von-Mises distributions with differing concentration parameters. The set-up is similar to the one in section 3.2 except that now $\theta_{ij} \sim$ von-Mises $(\mu_i, \kappa_i), i = 1, ..., p$. The likelihood function is given by

$$L^*(\boldsymbol{\mu}, \boldsymbol{\kappa}|\boldsymbol{\theta}) = \prod_{i=1}^{p} L_i, \quad L_i = \frac{1}{I_0(\kappa_i)^{n_i}} \exp\left[\kappa_i\{\sum_{j=1}^{n_i} \cos(\theta_{ij} - \mu_i)\}\right].$$

The only existing test for unknown and unequal concentrations for this problem is the likelihood ratio test suggested by Watson (1983) (WW^* in the sequel) given by

$$T_{WW^*} = 2\left(\sum_{i=1}^{p} \hat{\kappa}_i R_i - R_W\right) \sim \chi^2_{p-1},$$

where

$$R_W = \left\{\left(\sum_{i=1}^{p} \hat{\kappa}_i R_i \cos\theta_{i.}\right)^2 + \left(\sum_{i=1}^{p} \hat{\kappa}_i R_i \sin\theta_{i.}\right)^2\right\}^{1/2},$$

$\hat{\kappa}_i$ being the MLE of κ_i under i^{th} group.

To identify the specific parametric region where improvement over WW^* is essential, power values were simulated with details as that for equal κ case. Additionally, increments in the concentration parameters by .25 and .5 for successive groups were introduced. Though the sizes of actual LRT were unduly large, the size-normalized power function was reasonably good even for small group sizes. Size corrective multiplicative adjustments however may depend on the pattern of concentrations across the groups in a complicated way and can not be easily derived.

ILRT for this situation is developed next. Adopting a parallel approach as in section 3.2 with the product prior

$$\Pi^*(\boldsymbol{\kappa}) = \prod_{i=1}^{p} \Pi_i, \quad \Pi_i = (I_0(\kappa_i))^{n_i} \kappa_i^{(n_i-1)/2-1} \exp(-n_i\kappa_i), \quad \kappa_i > 0 \; \forall \; i, \tag{10}$$

the resulting integrated likelihood function is

$$\bar{L}^*(\boldsymbol{\mu}|\boldsymbol{\theta}) = \prod_{i=1}^{p} \int L_i \Pi_i d(\kappa_i), \tag{11}$$

leading to

$$\bar{L}^*(\boldsymbol{\mu}|\boldsymbol{\theta}) \propto \prod_{i=1}^{p} [n_i - \sum_{j=1}^{n_i} \cos(\theta_{ij} - \mu_i)]^{-(n_i-1)/2}.$$

The maximizer of $\bar{L}^*$ with respect to $\boldsymbol{\mu}$ under H_1 is still $\bar{\mu}_i^* = \hat{\mu}_i = \theta_{i.}, \; i = 1, ..., p$. However, under H_0, $\bar{L}^*$ is maximized at $\bar{\mu}_0^*$, a solution to the equation

$$\sum_{i=1}^{p} (S_{i.} \cos(\bar{\mu}_0^*) - C_i \sin(\bar{\mu}_0^*))/(n_i - C_{i.} \cos(\bar{\mu}_0^*) - S_{i.} \sin(\bar{\mu}_0^*)) = 0, \tag{12}$$

where S_i and C_i are defined in section 2.1. This leads to the integrated likelihood ratio

$$\bar{\lambda}^* = \prod_{i=1}^{p} \left[\frac{n_i - R_i}{n_i - C_{i.} \cos(\bar{\mu}_0^*) - S_{i.} \sin(\bar{\mu}_0^*)} \right]^{(n_i-1)/2}.$$

The log likelihood ratio statistic is

$$T_{ILRT^*} = -2log\bar{\lambda}^* = \sum_{i=1}^{p} (n_i - 1). \log\left(\frac{n_i - R_i}{n_i - C_{i.} \cos(\bar{\mu}_0^*) - S_{i.} \sin(\bar{\mu}_0^*)} \right).$$

As before, we have the following result:

Theorem 2: The asymptotic distribution of T_{ILRT^*} is χ^2_{p-1}.

Proof of Theorem 2. Note from equation (11) that

$$\begin{aligned}\bar{L}^*(\boldsymbol{\psi}|\boldsymbol{\theta}) &= \prod_{i=1}^{p}\int L_i\Pi_i d(\kappa_i)\\ &= \prod_{i=1}^{p} f_i\end{aligned}$$

where $f_i = \int L_i\Pi_i d(\kappa_i)$. Treating each group separately and employing parallel arguments as in the proof of Theorem 1 on each L_i separately, analogues of equations (4) through (8) hold for each f_i, so that

$$\begin{aligned}\bar{l}_i(\boldsymbol{\psi}_i) &= \log(f_i)\\ &= l_{A_i}(\boldsymbol{\psi}_i) + log(\Pi_i(\hat{\kappa}_{i_{\boldsymbol{\psi}_i}})) + log(g(\hat{\kappa}_{i_{\boldsymbol{\psi}_i}})) + log(\{1 + O(n^{-1.5})\})\end{aligned}$$

where $l_{A_i}(\boldsymbol{\psi}_i)$ is the Cox-Reid adjusted profile likelihood corresponding to L_i. Taking logarithms, This gives

$$\begin{aligned}\bar{l}^*(\boldsymbol{\psi}) &= \sum_{i=1}^{p}\log(f_i)\\ &= \sum_{i=1}^{p} l_{A_i}(\boldsymbol{\psi}_i) + log(\Pi_i(\hat{\kappa}_{i_{\boldsymbol{\psi}_i}})) + log(g(\hat{\kappa}_{i_{\boldsymbol{\psi}_i}})) + log(\{1 + O(n^{-1.5})\}).\end{aligned}$$

the MLE('s) of $\boldsymbol{\mu}$ under H_1 are $\boldsymbol{\theta}_{i.}$ as given in section (2.1) while under H_0 is $\bar{\mu}_0^*$ given in equation (12). The MLE's $\hat{\kappa}_i$ of κ_i (solutions to the equations $A(\hat{\kappa}_i) = R_i/n_i,\ i = 1, ..., p$, where R_i are defined in section 2.1) are same under both H_0 and

H_1. Consequently, the terms containing the estimates of κ_i in the log integrated likelihood ratio get cancelled out leaving the resulting ILRT statistic :

$$-2\log\bar{\lambda}^* = -2\sum_{i=1}^{p}\left[(l_{A_i}(\hat{\boldsymbol{\mu}}) - l_{A_i}(\hat{\mu}_0^*)) + O_P(n_i^{-1.5})\right].$$

Ignoring the $O_P(n_i^{-1.5})$ terms, noting the asymptotic χ_1^2 distribution of the adjusted Cox-Reid log likelihood ratio for each group, additive property of χ^2 under independence across the groups and a common estimate of mean under the null establish the asymptotic χ^2_{p-1} distribution of T_{ILRT^*}. $\square$

Use of the overall sample mean $\theta_{..}$ in place of $\bar{\mu}_0^*$ gave good approximation. Also minor fine-tuning with the multipliers 1.085 for $(.7 < k_0 < 1)$, 1.05 for $(2 < k_0 < 5)$ and 1.15 for $(1 < k_0 < 2)$ further enhanced the size performance. Here k_0 is the smallest one among the estimates of concentration parameters for the p groups. Equally good performance was also exhibited by size-adjusted .88 T_{WW*} with multiplier .88 for $\kappa_0 > .7$. However for very small concentrations, namely $k_0 < .7$ none of the test gave satisfactory results and this case needs further investigation.

The next section briefs the development of ILRT for the GvM in the circular case.

4.2.ANOMED for GvM (A case of two nuisance parameters)

Note that the ILRT-based treatment for the nuisance parameter is likely to be effective under orthogonality between the nuisance parameters and the parameters of interest more effectively when the prior does not depend on the parameters of interest.

This fact can be used to construct tests for ANOMED for some more distributions, preferably where the normalizing constant does not depend on the parameter of interest. In this case, its influence can be circumvented by including its reciprocal in the prior while keeping the prior free of the parameter of interest. These conditions are, for example, satisfied for a three parameter GvM and Batschelet(1981) distributions, as discussed below. Suppose $\boldsymbol{\theta} = \{\theta_{ij},\ \ i = 1, ..., p,\ \ j = 1, ...n_i\}$ are i.i.d. observations from the generalized von Mises distribution with pdf

$$f(\theta_{ij}) = [2\pi G_0(k_1, k_2)]^{-1}\ \exp[k_1 \cos(\theta_{ij} - \mu) + k_2 \cos 2\ (\theta_{ij} - \mu)], k_1 > 0,\ k_2 > 0,$$

where $\mu \in [0, 2\pi)$ is a location parameter and $G_0(k_1, k_2)$ is the normalizing constant. The prior $\Pi(\kappa_1, \kappa_2) = [G_0(\kappa_1, \kappa_2)]^n \exp[-n\kappa_1 - n\kappa_2] k_1^{a_n-1} k_2^{a_n-1}$, is the most appropriate and yields

$$\bar{L}(\boldsymbol{\mu}|\boldsymbol{\theta}) \propto [n - \sum_i R_i^1]^{a_n} [n - \sum_i R_i^2]^{a_n} \tag{13}$$

where, $R_i^l(x) = \sum_{j=1}^{n_i} \cos l(\theta_{ij} - x),\ l = 1, 2$, leading to the ILR statistics

$$T_{GvM\_ILRT} \equiv 2a_n \log\Big[\frac{[n - \sum_i R_i^1(\hat{\mu}_0)][n - \sum_i R_i^2(\hat{\mu}_0)]}{[n - \sum_i R_i^1(\hat{\mu}_{1i})][n - \sum_i R_i^2(\hat{\mu}_{1i})]}\Big]. \tag{14}$$

Here, $\hat{\mu}_{1i}, i = 1, ..., p$ and $\hat{\mu}_0$ are maximizers of $\bar{L}(\boldsymbol{\mu})$ under H_1 and H_0 respectively and can be obtained using numerical methods. Since the domain of maximization is bounded, this should not pose much difficulty. The choice of a_n can be based on the Satterthwaite-Welch type technique in line with the arguments in section 3.2.

A parallel approach holds for the Batschelet distribution(1981)with density function

$$f(\theta) = C^{-1}\exp[\kappa\cos(\theta-\mu)+\nu\sin(\theta-\mu)];\ -\pi\leq\theta,\ \mu<\pi;\ \kappa\geq 0;\ -\infty<\nu<\infty,$$

with $R_i^2(x)$ replaced by $\sum_{j=1}^{n_i}\sin(\theta_{ij}-x)$ in equations (13) and (14). However fine tuning adjustments as in section A.1 of Appendix-A may need to be developed under small concentrations.

5. Examples

This section illustrates some applications of the ILRT with real life examples representing situations where ANOMED is most appropriate. The computational details are summarized in Table 1. For WW, HKG, LRT and AW the computational formulae given in section 2.2 are used. ILRT is computed using equation (3) together with the multiplicative correction factor suggested in section A.1 of Appendix A, and replacing κ_0 by its estimate $\hat{\kappa}_0$, as reported in Table 1. For the data sets, (except the data set D_3 where the raw data was not available), the assumptions of von-Mises distribution and equal concentration parameter for the groups were validated (cf MJ (2000), Fisher (1993)). These examples are also informative of the proper usage of appropriate tests.

5.1 Epidemic onset data:D1,D2

In certain epidemic diseases, like acute primary angle closure glaucoma (APACG), the exact date of attack can be reliably determined. As suggested by Gao *et al.*

(2006) (GAO henceforth), each date of onset within a year can be represented as an angle by treating the 365 days of a year equivalent to 360^o ($2.\pi$ radians). So, one day is equivalent to $360/365 = 0.986^o$. Then a well-fitted von-Mises distribution with a single peak (mode) (indicating a prevalent date of onset), would indicate the seasonal influence on such data. Furthermore we note here that a significant difference between the peak dates of onset for the groups corresponding to the different levels of an attribute like age-group, gender etc., can be considered to be indicative of interaction between the seasonal effect and the attribute under consideration.

Gender, adverse environmental conditions and amount of sun light are known to be influential factors in causing APACG (cf Ivanisevic *et al.* (2002), Sharpec *et al.* (2010), Hillman *et al.* (1977)). As the latter two factors vary with season, a seasonal impact on the onset of APACG is expected. This may vary across the gender and age groups, perhaps due to differing capabilities of sustaining the adverse conditions.

Gao give data on exact dates of onset converted to angles for 132 APACG patients from Singapore along with information on other attributes like age group, sex, etc. The data-set D1 is extracted from this data-base and displays the dates of onset of APACG for male patients partitioned into four age groups, namely, below 50, 50 to 59, 60 to 69 and above 70. Referring to Table-1 for D1, the estimated concentration under H_0 ($\hat{\kappa}_0 = .2563$) is very small as are the group sizes. Following recommendations of section 3.3, inference based on ILRT is most reliable. ILRT clearly rejected the hypothesis of no difference (p-value: 0.0151), as did the next favored HKG (cf: (B)-(i); section 3.3) (p-value: 0.0132). This indicates that the seasonal impact varies among the age groups (i.e. the mean dates of onset across the age groups are significantly different). The strength with which WW and LRT rejected the hypothesis (p-vale $< 10^{-3}$) is untrustworthy owing to their large type-I errors (cf - (A)-(i), section 3.3), though they are at parity with ILRT. The least powerful test AW accepted the null, perhaps a wrong decision. Note that in some situations, for example in case-control studies for assessing effectiveness of a treatment on gait pattern under Cerebral Palsy where even small angular differences with respect to gait pattern are of great clinical importance, such a decision could be risky.

Similar hypothesis for female patients (data set D2) extracted from the same data-base under same age groups was unanimously accepted by all the tests except WW (which rejected the hypothesis (p-value: 0.0391) conforming to its aforementioned tendency of false alarms under small concentrations). Such a decision can also be undesirable in some situations like drug testing where falsely declaring a drug to be superior over others could be harmful.

In conclusion, males are prone to age dependent seasonal impact while for females seasonal influence does not depend on age. This also indicates a kind of three way interaction among the gender, age-group and seasonal influence on the dates of onset of APACG. Gao et al observed such differences but were not able to establish the same statistically using circular regression, possibly because interaction effects were not accounted for in their regression model. A further clinical investigation and research is needed in this context, since the results observed here may give some important clues and insights.

5.2 Light pulse treatment on the pineal melatonin rhythm:D3

It is widely assumed that the circadian system adapts to local environmental cues such as light and temperature, which vary enormously across habitats. Moore & Menaker (2012) examined effect of light pulse treatment on the pineal melatonin rhythm of five Anolis lizards species. The data set D3, with small group sizes from a control group and a treatment group, was analysed in a similar manner for the *A. gundlachi* species, As reported in Table 1, the light pulse treatment caused a significant phase delay in the circadian rhythm. This indicates that the circadian system of the species under consideration gets itself adapted for the Light Pulse treatment. This is high concentrated data ($\hat{\kappa}_0 = 9.0176$) and with the analysis as per the recommendations of section 3.3 (cf (B)-(v), section 3.3), all tests are equally competent. This is reflected in the unanimous decision of rejecting the null hypothesis given by all the tests (all p-values $<$0.05). However here also ILRT rejects the hypothesis much strongly than other tests (p-value=0.0006), favoring the conjectured behavior.

6. Concluding remarks

Our motivation for this paper was to develop an efficient parametric test for testing

Table 1: Computational details for the three data sets.

Data	Group sizes	Resultant lengths	$\hat{\kappa}_0$	p-values
D_1	$n_1 = 5$	R_1=4.0986	0.2563	ILRT: 0.0151
p=4	$n_2 = 9$	R_2=3.9193		AW:0.0869
	$n_3 = 12$	R_3=6.146		HKG: 0.0132
	$n_4 = 9$	R_4=3.104		WW: 0.0007
	N=35	R_0=4.4494		LRT: 0.0007
D_2	$n_1 = 8$	R_1=4.7977	0.4116	ILRT:0.1815
p=4	$n_2 = 22$	R_2=3.7932		AW:0.1414
	$n_3 = 36$	R_3=6.8435		HKG: 0.1284
	$n_4 = 31$	R_4= 10.7446		WW: 0.0391
	N=97	R=19.5524		LRT:0.0923
D_3	$n_1 = 9$	R_1=8.73	9.0186	ILRT:0.0006
p=2	$n_2 = 7$	R_2=6.65		AW: 0.0211; HKG: 0.0154
	N=16	R=15.085		WW: 0.0196; LRT:0.0115

homogeneity of mean directions of several independent circular populations, which can be universally implemented in practice. The need of this test emerged from the fact that there is no such universal or omnibus test in the existing literature with acceptable performance as applicable to several diverse realistic situations, e.g. low concentrations and large number of small size groups. We have derived an universal, yet simple and elegant test statistic. It was demonstrated that our method can also be extended in a straight forward manner to a rich class of distributions: asymmetric, bimodal, sharply-peaked, flat-topped, etc., as modelled by e.g. generalized von Mises and Batschelet distributions. The hurdle of the non location-scale nuisance parameters $\boldsymbol{\kappa}$ was overcome by introducing a new approach through the integrated likelihood ratio test. It was also established by extensive simulations that our test outperforms in the usual parametric region and uniformly well-competes with the best among the available ones. Finally, our approach is amenable to elegant and almost straight forward generalizations to higher dimensions, i.e. to hyper-spherical, e.g., Langevin, populations. This last observation is currently being studied in further details.

APPENDIX-A

A.1 Corrective adjustments for small concentrations:

The corrective multiplicative adjustment cf_{Ilrt} given below for controlling the sizes of ILRT under small and equal concentrations was derived by regressing the ratio of theoretical 95^{th} quantile of the desired χ^2_{p-1} distribution to the simulated 95^{th} quantiles of T_{ILRT} under H_0 based on 200000 simulations. A large number of parametric combinations of input parameters n, p and κ were used and then κ was replaced by its estimate $\hat{\kappa}_0$. The densely clustered sizes around the target level of the multiplicatively adjusted ILRT as seen in the corresponding box-plot in Figure 1 are indicative of a closer conformation to the desired χ^2_{p-1} distributional assumption.

cf_{Ilrt}=

$$\begin{cases} 0.563 - 0.0029\, n_1 + 0.029\, p + 0.93\, \kappa_0 - 0.32\, \sqrt{p} - 0.12\, \log(N) & \\ +0.32\, \log(p) - 0.186\, \log(\kappa_0) + 0.019\, n1\, \kappa_0) & \text{if } \kappa_0 < .4, \\ (1.92 - 0.0186\, \sqrt{p} + 0.0544\, \log(N) - 0.985\, \sqrt{\kappa}_0 + \log(\kappa_0) & \\ -0.002\, \sqrt{N} + 0.001\, n1 - 0.01\, \sqrt{n}_1) & \text{if } .4 < \kappa_0 < 1. \end{cases}$$

Furthermore, as mentioned in the proof of Theorem 1 a little fine-tuning for moderate values of $\kappa \in (1, 9)$ namely, 1.11 for $\{1 < \kappa_0 < 1.25\} \cup \{3 < \kappa_0 < 4.25\}$; 1.17 for $1.25 < \kappa_0 < 3$; 1.04 for $4.25 < \kappa_0 < 9$ gave excellent results. Also for $\kappa_0 > 15, a_n = n - 1.5$ in place of $n - 1$ gave more accurate results.

A.2 Piece-wise approximation of $A(\kappa)$:

Note that for $\kappa \in [1, \infty)$, $\omega = 1/\kappa \in (0, 1]$. By computing $A(\omega)$ on a very fine mesh of $(0, 1]$ and regressing $A(\omega)$ verses ω piece-wise on the partition given in Table 2, (chosen selectively) the approximation of $A(\kappa)$ with error less than $10^{(-3)}$ reported in Table 2 was obtained.

Acknowledgements

Part of this work was done by Prof. Ashis SenGupta as Distinguished Professor at

Table 2: Details of the piece-wise approximation $c + b\omega$, $\omega = 1/\kappa$ for $A(\kappa)$

Domain for κ	c	b
[1,1.45)	0.391	1.84
[1.45,3)	.235	1.95
[3, 4.25)	.149	1.98
[4.25,10)	.0805	1.99
[10, 15)	.046	2
[15, 50)	.0181	2
$>$50	.007	2

Middle East Technical University, Ankara, Turkey. A major part of this work was completed during the visits of Dr. H. V. Kulkarni at ASU, Indian Statistical Institute, Kolkata, India (ISI). She was partially supported by ISI, The National Board for Higher Mathematics(NBHM), Department of Atomic Energy (DAE), India, under the grant no. 2/48(34)/2016/NBHM(RP)/R&D II/4531 and Department of science and technology (DST) through the grant SR/FST/MSI-103 dated 18/11/2015 awarded to the department of statistcs, Shivaji University, Kolhapur; and is thankful to them for financial support for this research work.

We are deeply thankful to an anonymous referee, whose careful reading and valuable suggestions greatly improved the manuscript.

References

Amaral, G.J.A., Dryden, I.L. and Wood, D. A. (2007). Pivotal Bootstrap Methods for k-Sample Problems in Directional Statistics and Shape Analysis. *J. Am. Statist. Assoc.* **102**, 695-707.

Anderson, C.M. and Wu, C.F. (1995). Measuring Location Effects from Factorial Experiments with a Directional Response. *Int. Statist. Review* **63**, 345-363.

Batschelet, E. (1981).Circular statistics in biology **Vol. 111**. *London: Academic press.*

Beran, R. and Fisher, N.I. (1998). Nonparametric Comparison of Mean Directions or Mean Axes. *The Ann. Statist.* **26**, 472-493.

Berger, J.O., Liseo, B.and Wolpert, R. (1999). Integrated likelihood functions for eliminating nuisance parameters (with discussion). *Statist. Sci.* **14**, 1-28.

Chamberlain, G. (2007).Decision theory applied to an instrumental variables model. *Econometrica* **75**, 609-652.

Cox, D.R. and Reid, N. (1987).Parameter Orthogonality and Approximate Conditional Inference. *J. R. Statist. Soc.* B **49**, 1-39.

Evans, M.and Swartz, T.(2000). Approximating integrals via Monte Carlo and deterministic methods. USA: Oxford University Press.

Fisher, N.I.(1993). *Statistical Analysis of Circular Data*. Cambridge, MA: Cambridge University Press.

Gao, F.,Chia, K.S., Krantz, I., Nordin, P. and Machin, D. (2006).On the application of the von Mises distribution and angular regression methods to investigate the seasonality of disease onset. *Statist. in Medicine* **25**, 1593-1618.

Ghosh, M., Datta, G.S., Kim, D.and Sweeting, T.J. (2006). Likelihood-based inference for ratios of regression coefficients in linear models. *Ann. of the Institute of Statist. Mathematics* **58**, 457-473.

Harrison, D., Kanji, G.K. and Gadsden, R.J. (1986). Analysis of variance for circular data. *J. of Applied Statist.* **13**, 123-138. *J. of Applied Statist.* **15**, 197-224.

Hillman, J.S., and Turner, D.C. (1977).Association between acute glaucoma and the weather and sunspot activity. *Brit. J. of Ophthalmology* **249**, 223-233.

Ivanisevic, M., Erceg, M., Smoljanovic, A. and Trosic, Z. (2002). The Incidence and Seasonal Variations of Acute Primary Angle-Closure Glaucoma. *Collegium Antropologicum* **1**, 41-45.

Jammalamadaka, S.R. and SenGupta, A. (2001). *Topics in Circular Statistics*. USA: World Scientific.

Kalbfleisch, J.D. and Sprott, D.A. (1970).Application of likelihood methods to models involving large numbers of parameters (with discussion). *J. R. Statist. Soc.* B **32**, 175-208.

Larsen, P.V., Blaesild, P. and Sorensen, M.K. (2002).Improved likelihood Ratio Tests on the von Mises-Fisher Distribution. *Biometrika* **89**, 947-951.

Liseo, B. (1993).Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295-304.

Lozano, Y. R. (2016). *Landmark processing by head direction cells.* Diss. UCL (University College London).

Malley, J.D., Redner, R.A., Severini, T.A., Badner, J.A., Pajevic, S. and Bailey-Wilson, J.E. Estimation of linkage and association from allele transformation data. *Biometrical Journal* **45**, 349-366.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics.* Willy Series in Probability and Statistics.

Moore, A.F. and Menaker, M. (2012). Photic resetting of the circadian clock is correlated with photic habitat in Anolis lizards. *J. of Comparative Physiology* A **198**, 375-387.

SenGupta, A. and Roy, R. (2011). Testing Homogeneity of Mean Directions for Circular Dispersion Models. *Calcutta Statistical Association Bulletin(Special 7-th Triennial Proceedings Volume* **63**, 249-263.

Severini, T.A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika* **94**, 529-542.

Severini, T.A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika* **97**, 481-496.

Severini, T.A. (2011). Frequency Properties of inferences Based on an integrated Likelihood function. textitStatistica Sinica **21**, 433-447.

Stephens, M.A. (1972). Multisample Tests for the von Mises Distribution. *J. Am. Stat. Asso.* **67**, 456-461.

Subak-Sharpe I.S., Low, S., Nolan, W. and Foster, P.J. (2010). Pharmacological and environmental factors in primary angle-closure glaucoma. *Br. Med. Bull.* **93**, 125-143.

Watson, G.S. and Williams, E.J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika* **43**, 344-352.

Indian Statistical Institute, Kolkata, WB 700108, India, and Augusta University, Augusta, GA 30912, USA.

E-mail: (amsseng@gmail.com)

Department of Statistics, Shivaji University, Kolhapur,India, 416004.

E-mail: (kulkarni.hemangi@gmail.com)