Aust. N. Z. J. Stat. **43**(4), 2001, 461–471

THE SLIPPAGE PROBLEM FOR THE CIRCULAR NORMAL DISTRIBUTION

ASHIS SENGUPTA[1*] AND ARNAB KUMAR LAHA[1,2]

Indian Statistical Institute

Summary

The slippage problem occurs when an unspecified observation in a given random sample is from a distribution other than that for all the remaining observations. This paper studies the problem in terms of the 'slip' in the mean direction of a circular normal distribution. The slippage problem is first treated as a multiple decision problem with a prior which is invariant under the permutations of the hypotheses. The probabilities of accepting the various hypotheses for the Bayes rule with respect to this prior are explicitly obtained. The likelihood ratio tests for this slippage problem, for the cases when the mean directions are both known and unknown, are shown to be easily computable. The tests are illustrated through two well-known datasets. The performances of a range of tests are compared using extensive simulation.

Key words: Bayes rule; directional data; likelihood ratio test; locally most powerful type test; slippage problem.

1. Introduction

The slippage problem occurs if an unspecified observation in a given random sample comes from a distribution different from that for all the other remaining observations. In a given sample the corresponding observation may manifest itself as a 'surprising' observation lying away from the remaining set. However, such a manifestation may not always be apparent save the knowledge of the possibility of its occurrence. We must infer that slippage has occurred and also identify the 'slipped' observation. This can also be viewed as a problem of 'outlier' detection. Slippage assumes great importance in many practical situations, not only for linear data but also where directional data are encountered; for example, in applications to meteorological data, wind directions, movements of icebergs, propagation of cracks, biological and periodic phenomena, quality assurance and productivity measures, etc.

An outlier may occur in circular data due to a 'slip' in recording, an uninvited 'guest' in the host population, a poorly 'trained' or 'distracted' or highly 'brainwashed' individual in the test group, a physical 'distortion' in a segment of the otherwise homogeneous sampling site, a sudden 'shock' in the sampling environment, etc. Mardia (1972 p. 390) cites his experience of '... possible recording errors in axial data such as single observation being given as 321° while all others are reduced modulo 180°'. In studies on homing abilities with biological subjects, the sample may contain a few from a differing 'guest' species, as could possibly be the case with Jander's ant data; see Jander (1957) and SenGupta & Pal (2001). Such

Received December 1998; revised March 2000; accepted August 2000.

* Author to whom correspondence should be addressed.
[1] Applied Statistics Division, Indian Statistical Institute, Kolkata – 700 035, India.
e-mail: ashis@isical.ac.in
[2] Current address: National Institute of Management, Calcutta, Kolkata, India.

Acknowledgments. The authors thank the Associate Editor, the Technical Editor and the anonymous referee for helpful comments which resulted in this improved version of the paper.

situations can also arise with 'training' or 'brainwashing' experiments where only a few may be expected to wander away from or move towards the target direction — see, for example, the discussions by Fisher (1993 p.87) on the resultant directions of 22 sea stars after being displaced from their natural habitat. A segment of the sandstone layer may be 'disorientated' due to geological or environmental disruptions which may give rise to possible outliers in the sample of palæocurrent orientations from that layer. Such a dataset from Belford Anticline, New South Wales (Fisher & Lewis, 1983), is the first example analysed in Section 5. A wheel may experience a short mechanical instability or a sudden power surge resulting in a suspicious stopping position; for example, see the dataset on a roulette wheel given by Mardia (1972 p.2). We analyse this dataset in more detail in Section 5. This example also demonstrates that, unlike in linear data, an outlier in directional data may be any observation in the dataset and not necessarily only one of the largest or smallest. Unless the sample is first 'cleaned' of the outliers, or an extended model is justified incorporating the outliers, statistical inference for such data is problematical.

Though practical examples, such as those above on the problem of outliers in directional data, abound, not much seems to be known about the theoretical foundations of any general approach to solving the slippage problem under a parametric model; for example, a 'slip' in terms of the mean direction of the circular normal distribution (see, however, Collett, 1980; Bagchi & Guttman, 1988, 1990; Upton, 1993; Barnett & Lewis, 1994). Here, we deal with some formulations and derivations of statistical tests in this set-up. Compared to the contamination or mixture formulation which helps to determine only the presence of outliers, our approach both determines their presence and helps detect them.

Suppose $\Theta_1, \ldots, \Theta_n$ are independent random variables. Let $\mathrm{CN}(\mu, \kappa)$ denote the circular normal distribution. As is usual, we refer to μ as the mean direction and κ as the concentration parameter. We want to test H_0: Θ_j, $j = 1, \ldots, n$ are identically distributed as $\mathrm{CN}(\mu_0, \kappa)$ against H_i: $\Theta_1, \ldots, \Theta_{i-1}, \Theta_{i+1}, \ldots, \Theta_n$ are identically distributed as $\mathrm{CN}(\mu_0, \kappa)$ and Θ_i is distributed as $\mathrm{CN}(\mu_1, \kappa)$, $1 \le i \le n$, $\mu_1 > \mu_0$, μ_1, μ_0 and κ are all known. We take a decision theoretic route to solve this problem, and derive the test statistic along with its exact sampling distribution under H_0. These results are contained in Theorems 2.1–2.3.

Our problem in general is to test H_0 against H_1^*; namely there exists i, i unknown, such that Θ_i is distributed as $\mathrm{CN}(\mu_1, \kappa)$ and $\Theta_1, \Theta_2, \ldots, \Theta_{i-1}, \Theta_{i+1}, \ldots, \Theta_n$ are distributed as $\mathrm{CN}(\mu_0, \kappa)$ and all are independent. In the case that $\mu_1 > \mu_0$, μ_1, μ_0 and κ are all known, we derive the likelihood ratio test (LRT) and show that here also we are able to specify the exact null distribution of the LRT statistic. That is the content of our Theorem 3.1. We next consider the case of testing H_0 vs. H_1^* when the parameters μ_0 and μ_1 are both unknown but κ is known. We derive the form of the LRT that is given in Theorem 3.2.

In Section 4 we use extensive simulation to find the null distribution of the LRT statistic. We provide a table of cut-off points for various κ (Table 1) and also the P-values for $\kappa = 1$ (Table 2). We also provide a numerical description of the performance of the LRT in terms of its power (Table 3). In Section 5, we illustrate these results by analysing two well-known datasets of practical interest. In Section 6, we provide a locally most powerful type test (LMPTT) for detecting outliers. In Section 7, we provide a simulation-based comparison of the performance of the various statistics used for outlier detection: the L statistic, M statistic, LRT statistic, the LMPTT statistic and the Bayes test with different priors. It is seen that the Bayes test performs best in detecting outliers of large magnitude whereas the LRT performs best in detecting outliers of moderate magnitude and the LMPTT performs best in detecting outliers of small magnitude. Since no single test performs uniformly better than the others, we suggest that the Bayes test, LRT or LMPTT be used depending on the severity of the outlier to be detected in a dataset.

2. The decision theoretic approach

Suppose $\Theta_1, \dots, \Theta_n$ are independent $\mathrm{CN}(\mu_i, \kappa)$ random variables with density

$$f(\theta; \mu_i) = \frac{\exp\big(\kappa \cos(\theta - \mu_i)\big)}{2\pi I_0(\kappa)} \qquad (0 \le \theta < 2\pi,\ 0 \le \mu_i < 2\pi, \kappa > 0),\ i = 0, 1.$$

We assume that $\mu_0 < \mu_1$ and $n \ge 3$. We are interested in finding the Bayes rule for the multiple decision problem of accepting one of the $n+1$ hypotheses $H_0, H_1, \dots, H_n$ with respect to the prior distributions invariant under permutations of $H_1, \dots, H_n$. We use the loss function which assigns loss $= 0$ if the correct hypothesis is accepted and loss $= 1$ otherwise. The prior distributions invariant under permutations of $H_1, \dots, H_n$ give equal weight to $H_1, \dots, H_n$ and hence they are of the form τ_p, where $\tau_p(H_0) = 1-np$, $0 \le p \le (1/n)$, $\tau_p(H_i) = p$, $1 \le i \le n$. Let $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_n)$ and let $\Phi(\boldsymbol{\Theta}) = (\phi(1 \mid \boldsymbol{\Theta}), \dots, \phi(n \mid \boldsymbol{\Theta}))$ denote a generalized critical function or a multiple decision rule (Ferguson, 1967 p.299) with $\phi(i \mid \boldsymbol{\Theta})$, $i = 1, \dots, n$ taking values 0 or 1, and $\sum \phi(i \mid \boldsymbol{\Theta}) = 1$. Thus Φ chooses H_i when $\boldsymbol{\Theta} = \boldsymbol{\theta}$ is observed if $\phi(i \mid \boldsymbol{\Theta}) = 1$. Let R_j denote the likelihood ratio:

$$R_j = \frac{f(\Theta_j; \mu_1)}{f(\Theta_j; \mu_0)} = \exp\Big(\kappa\big(\cos(\Theta_j - \mu_1) - \cos(\Theta_j - \mu_0)\big)\Big) \qquad (j = 1, 2, \dots, n).$$

Theorem 2.1. *The Bayes test with respect to τ_p for H_0 against H_i, $1 \le i \le n$, is given by $\phi(0; \boldsymbol{\Theta}) = 0$ whenever $(1-np)/p < \max_j R_j$ and $\phi(i; \boldsymbol{\Theta}) = 0$ whenever either $R_i < \max_j R_j$ or $(1-np)/p > \max_j R_j$, $1 \le i \le n$.*

Proof. The result follows from the general theory given in Ferguson (1967 p.299) after some simplifications.

The following theorem gives the performance of the Bayes rule given in Theorem 2.1 under the null hypothesis. Let,

$$K(t) = \exp\big(\kappa \cos(\delta + \pi - \sin^{-1} t)\big) + \exp\big(\kappa \cos(\delta + \sin^{-1} t)\big) + \exp\big(\kappa \cos(\delta + 3\pi - \sin^{-1} t)\big),$$

$$g(t) = \frac{K(t)}{2\pi I_0(\kappa)\sqrt{1-t^2}} \qquad (-1 < t < 1) \qquad \text{and} \qquad G(s) = \int_{-1}^{s} g(t)\, dt.$$

Further let $\mu_0 = \mu$, $\mu_1 = \mu + 2\delta$ and define

$$u = \frac{1}{2\kappa \sin \delta} \ln\Big(\frac{1-np}{p}\Big).$$

Theorem 2.2. *In the framework of Theorem 2.1,*

(a) $\Pr\big(\phi(i; \boldsymbol{\Theta}) = 0 \mid H_0 \textit{ is true}\big) = 1 - \dfrac{1}{n} + \dfrac{G(u)^n}{n}$,

(b) $\Pr\big(\phi(0; \boldsymbol{\Theta}) = 0 \mid H_0 \textit{ is true}\big) = 1 - G(u)^n$.

Proof. Observe that, $R_j = \exp\big(2\kappa \sin\delta \sin(\Theta_j - \mu - \delta)\big)$. Thus, as a consequence of our assumption $0 \le \mu_0 < \mu_1 < 2\pi$, $\max_j R_j = \exp\big((2\kappa \sin\delta) \max_j \sin(\Theta_j - \mu - \delta)\big)$ because $\exp(x)$ is an increasing function of x, and $2\kappa \sin\delta > 0$ because $0 < \delta < \pi$. Let $\eta_j = \sin(\Theta_j - \mu - \delta)$ for $j = 1, 2, \ldots, n$. Since $\Theta_1, \ldots, \Theta_n$ are independent and identically distributed (iid) it follows that $\eta_1, \eta_2, \ldots, \eta_n$ are iid. We first derive the distribution of η. If there is a branch of $\sin^{-1}\theta$ which is monotone on $[\pi/2, 3\pi/2]$, then the inverse tranformation defined uniquely through the domain of θ as partitioned below (where θ greater than or equal to 2π is to be interpreted as being reduced modulo 2π), is

$$\psi(\eta) = \begin{cases} \mu + \delta + (\pi - \sin^{-1}\eta) & \text{if } \mu + \delta \le \theta \le \frac{1}{2}\pi + \mu + \delta, \\ \mu + \delta + \sin^{-1}\eta & \text{if } \frac{1}{2}\pi + \mu + \delta \le \theta \le \frac{3}{2}\pi + \mu + \delta, \\ \mu + \delta + 3\pi - \sin^{-1}\eta & \text{if } \frac{3}{2}\pi + \mu + \delta \le \theta \le 2\pi + \mu + \delta. \end{cases}$$

After some calculation we find the density of η is $g(\eta)$.

Note that the density does not exist for the points –1 and 1, but the set $\{-1, 1\}$ has Lebesgue measure zero and hence any value can be put at these points without changing the distribution. We put 0 at these points. Define $W = \max_\ell \eta_\ell$. Then W has cumulative distribution function (cdf) H where $H(w) = G(w)^n$. Let $W_i^* = \max_{\ell \ne i} \eta_i$. Then the cdf of W_i^* is $H_i^*(w) = G(w)^{n-1}$. Now

$$\Pr\Big(\max_j R_j < \frac{1 - np}{p}\Big) = \Pr\Big(W < \frac{\ln\big(\frac{1-np}{p}\big)}{2\kappa \sin\delta}\Big) = H(u)\,.$$

Note that the event $R_i < \max_j R_j$ is equivalent to the event $\eta_i < W$. Further, we have $\Pr(\eta_i \ge W) = \Pr(\eta_i > \eta_s, s \ne i)$ because the distribution of the η_i is continuous. Thus, $\Pr(\eta_i > \max_s \eta_s) = \int_{-1}^{1} (1 - G(\eta))\, dH_i^*(\eta) = (n-1)\int_0^1 (1-y) y^{n-2} dy = 1/n$ where we put $y = G(\eta)$ and $dH_i^*(\eta) = (n-1) G(\eta)^{n-2} dG(\eta)$. Now

$$\Pr(\phi(i; \mathbf{\Theta}) = 0) = \Pr(\eta_i < W) + \Pr(W < u) - \Pr(\eta_i < W < u)\,.$$

Thus, the problem of finding $\Pr(\phi(i; \mathbf{\Theta}) = 0 \mid H_0 \text{ is true})$ is solved if $\Pr(\eta_i < W < u)$ is obtained. Now,

$$\begin{aligned} \Pr(\eta_i < W < u) = \Pr(\eta_i < W_i^* < u) &= \int_{-1}^{u} G(\eta)\, dH_i^*(\eta) \\ &= (n-1)\int_{-1}^{u} G(\eta)^{n-1}\, dG(\eta) = (n-1)\int_0^{G(u)} y^{n-1}\, dy = \frac{n-1}{n} G(u)^n. \end{aligned}$$

It follows that

$$\begin{aligned} &\Pr(\phi(i; \mathbf{\Theta}) = 0 \mid H_0 \text{ is true}) = 1 - \frac{1}{n} + \frac{G(u)^n}{n}\,; \\ &\Pr(\phi(i; \mathbf{\Theta}) = 1 \mid H_0 \text{ is true}) = 1 - \Pr(\phi(i; \mathbf{\Theta}) = 0 \mid H_0 \text{ is true}) = \frac{1}{n} - \frac{G(u)^n}{n}\,; \\ &\Pr(\phi(0; \mathbf{\Theta}) = 0 \mid H_0 \text{ is true}) = 1 - G(u)^n\,; \\ &\Pr(\phi(0; \mathbf{\Theta}) = 1 \mid H_0 \text{ is true}) = \Pr(W \le u) = G(u)^n. \end{aligned}$$

Theorem 2.3 gives the performance of the Bayes rule when H_j is true. Let η_i be as in the proof of Theorem 2.2 and let G be its cdf when $\Theta_i \overset{\mathrm{d}}{=} \mathrm{CN}(\mu_0, \kappa)$ and let G^* denote its cdf when $\Theta_i \overset{\mathrm{d}}{=} \mathrm{CN}(\mu_1, \kappa)$.

Theorem 2.3. *In the framework of Theorem 2.1, let* $1 \le i, j \le n$,

(a) $\Pr(\phi(i;\boldsymbol{\Theta}) = 0 \mid H_j$ *is true*)

$$= 1 + \int_{-1}^{u} G(w)^{k-2} G^*(w)\, dG(w) - (k-2)\int_{-1}^{1} G(w)^{k-3}\big(1 - G(w)\big) G^*(w)\, dG(w) - \int_{-1}^{1} G(w)^{k-2}\big(1 - G(w)\big)\, dG^*(w) \quad \textit{where} \quad i \ne j,$$

(b) $\Pr(\phi(0;\boldsymbol{\Theta}) = 0 \mid H_j$ *is true*)

$$= 1 - (k-1)\int_{-1}^{u} G(w)^{k-2} G^*(w)\, dG(w) - \int_{-1}^{u} G(w)^{k-1}\, dG^*(w)\,,$$

(c) $\Pr(\phi(j;\boldsymbol{\Theta}) = 0 \mid H_j$ *is true*)

$$= \int_{-1}^{u} G(w)^{k-1}\, dG^*(w) + (k-1)\int_{-1}^{1} G(w)^{k-2} G^*(w)\, dG(w).$$

Proof. The proof follows the lines of the proof of Theorem 2.2 but with more tedious computation after noting that under $H_j, \eta_1, \eta_2, \ldots, \eta_n$ are no longer iid. In fact the distribution of η_j differs from the rest.

The two probabilities given in Theorem 2.2 can be readily computed for the circular normal distribution. Note, however, that the choice of $\tau_i = p$ reflects the case of equal ignorance. In the case that we have other prior information, this may be incorporated and the corresponding probabilities computed as above, but the resulting expressions would be quite complicated.

3. The likelihood ratio test

Case I. We first consider the case of testing H_0 against H_1^* when $0 \le \mu_0 < \mu_1 < 2\pi$, μ_1, μ_0 and κ are all known. As in the previous section, let $\mu_0 = \mu$ and $\mu_1 = \mu + 2\delta$, $\delta > 0$. In this case we prove the following theorem. Let G be the distribution function of $\sin(\Theta - \mu - \delta)$ under H_0.

Theorem 3.1. *In testing* H_0 *against* H_1^* *the LRT statistic is equivalent to the statistic* $V = \max_j \sin(\Theta_j - \mu - \delta)$. *The exact sampling distribution function of* V *under* H_0 *is given by* $M(\theta) = G(\theta)^n$.

Proof . The likelihood under H_0 is

$$L_0(\theta_1, \theta_2, \ldots, \theta_n) = \frac{\exp\big(\kappa \sum_i \cos(\theta_i - \mu)\big)}{\big(2\pi I_0(\kappa)\big)^n} \qquad (0 \le \theta_i < 2\pi,\ \kappa > 0,\ 0 \le \mu < 2\pi),$$

and that under H_1^* is,

$$L_1^*(\theta_1, \theta_2, \ldots, \theta_n) = \max_j \left\{ \big(2\pi I_0(\kappa)\big)^{-n} \exp\left(\kappa \sum_{i \ne j} \big(\cos(\theta_i - \mu) + \cos(\theta_j - \mu - 2\delta)\big)\right)\right\}$$
$$(0 \le \theta_i < 2\pi,\ \kappa > 0,\ 0 \le \mu < 2\pi,\ \delta > 0).$$

Thus the LRT statistic Λ is given by

$$\begin{aligned}-\ln\Lambda &= \max_j\left\{\kappa\left(\sum_{i\neq j}\cos(\theta_i-\mu)+\cos(\theta_j-\mu-2\delta)-\sum_i\cos(\theta_i-\mu)\right)\right\}\\ &= \max_j\left\{\kappa\big(\cos(\theta_j-\mu-2\delta)-\cos(\theta_j-\mu)\big)\right\}\\ &= \max_j\left\{\kappa\big(2\sin(\theta_j-\mu-\delta)\sin\delta\big)\right\}.\end{aligned}$$

Since $0<\delta<\pi$, we have $\sin\delta>0$. Thus $-\ln\Lambda$ is equivalent to the statistic $\max_j\sin(\Theta_j-\mu-\delta)=V$. Now note that under H_0, $\sin(\Theta_j-\mu-\delta)$ are iid. Hence the distribution function of V is $M(\theta)=G(\theta)^n$.

Observe that $G(\theta)$ can be evaluated numerically and hence the cut-off points for the LRT are readily available.

Case II. We wish to test H_0 against H_1^* when κ is known but μ and δ are unknown. The form of the LRT is given by Theorem 3.2. Let $\hat\mu_0$ and $\hat\mu_1^*$ denote the estimates of μ under H_0 and H_1^* respectively. Further, let L_j denote the likelihood when there is a slip at j ($j=1,2,\ldots,n$). Let $\hat\jmath$ be the value of j for which L_j attains its maximum.

Theorem 3.2. *In testing H_0 against H_1^* the LRT statistic Λ is given by*

$$\begin{aligned}-\ln\ \Lambda = \kappa\Big(&\big(\textstyle\sum_{i\neq\hat\jmath}\cos\theta_i\big)(\cos\hat\mu_1^*-\cos\hat\mu_0)\\ &+\big(\textstyle\sum_{i\neq\hat\jmath}\sin\theta_i\big)(\sin\hat\mu_1^*-\sin\hat\mu_0)+1-\cos(\theta_{\hat\jmath}-\hat\mu_0)\Big).\end{aligned}$$

Proof. The log-likelihood under H_0 is

$$\begin{aligned}\ln L_0(\theta_1,\ldots,\theta_n) &= -n\ln\big(2\pi\ I_0(\kappa)\big)+\kappa\textstyle\sum_i\cos(\theta_i-\mu)\\ &= -n\ln\big(2\pi\ I_0(\kappa)\big)+\kappa\big(\cos\mu\textstyle\sum_i\cos\theta_i+\sin\mu\sum_i\sin\theta_i\big).\end{aligned}$$

Putting $\partial\ln L_0/\partial\mu=0$ and solving for μ gives $\hat\mu_0=\tan^{-1}\big(\sum_i\sin\theta_i/\sum_i\cos\theta_i\big)$ with $\tan^{-1}$ so defined as to be unique. Under H_1^* the log-likelihood is

$$\begin{aligned}\ln L_1^*(\theta_1,\ldots,\theta_n) = \max_j\Big\{&-n\ln 2\pi I_0(\kappa)\\ &+\kappa\big(\cos\mu\textstyle\sum_{i\neq j}\cos\theta_i+\sin\mu\displaystyle\sum_{i\neq j}\sin\theta_i+\cos(\theta_j-\mu-2\delta)\big)\Big\}.\end{aligned}$$

Thus, $L_1^*=\max_j L_j$. Fix $1\le j\le n$. We now compute $\hat\mu_j$ and $\hat\delta_j$ which are the maximum likelihood estimates (MLEs) of μ and δ under H_j. Putting

$$\frac{\partial\ln L_j}{\partial\mu}=0\qquad\text{and}\qquad\frac{\partial\ln L_j}{\partial\delta}=0$$

and solving for μ gives

$$\hat\mu_j=\tan^{-1}\left(\frac{\sum_{i\neq j}\sin\theta_i}{\sum_{i\neq j}\cos\theta_i}\right)\qquad\text{and}\qquad\hat\delta_j=\frac{\theta_j-\hat\mu_j}{2}.$$

Let $\hat\jmath$ be that j for which L_j attains its maximum value after substituting $\hat\mu_j$ and $\hat\delta_j$. Thus

TABLE 1

Cut-off points of the LRT ($\alpha = 0.05$, *various* κ)

κ	n =10	n = 20	n = 30
0.5	0.3680	0.3679	0.3679
1.0	0.1357	0.1354	0.1354
1.5	0.0504	0.0499	0.0498
2.0	0.0199	0.0187	0.0185
4.0	0.0139	0.0062	0.0034
10.0	0.0210	0.0102	0.0064

under H_1^* the estimate of μ is $\hat{\mu}_{\hat{j}}(=\hat{\mu}_1^*)$ and that of δ is $\hat{\delta}_{\hat{j}}(=\hat{\delta})$. Therefore the LRT statistic Λ is given by

$$-\ln\Lambda = \kappa\big(1 + \textstyle\sum_{i\neq\hat{j}}\cos(\theta_i - \hat{\mu}_1^*) - \sum_i \cos(\theta_i - \hat{\mu}_0)\big)\,,$$

which after some calculations gives the claimed expression.

The exact sampling distribution of the LRT statistic is formidable — no closed form seems to be possible nor even any analytic representation for it in small samples.

We note that Collett (1980) considers a problem very similar to the above. He tests for no slippage versus a slippage alternative and derives an LRT statistic for it, which he calls the L statistic. However, the difference from our approach is that he first uses a data-based measure to detect the outlier candidate. Subsequently, formal tests are conducted to statistically validate the hypothesis that it is in fact an outlier. The choice of the measure may not be reasonable for even symmetric datasets where clusters may appear far from the mean direction. In our procedure the detection and testing for the outlier is based on the LRT and is entirely probabilistic.

4. Simulation

We use simulation to obtain the null distribution of Λ as well as its power. The simulation results for the null distribution are based on 5000 repetitions with sample size n, $n = 10, 20, 30$. The random sample from a circular normal distribution is drawn using the IMSL library routine. For simulating the power a particular observation is drawn from a circular normal population, CN(Δ, 1) with $\Delta = 20(20)180$ (in degrees), and repeating it 5000 times. Since the power function is symmetric about 180° the above computation is sufficient. Table 1 gives the cut-off points for the LRT at 5% level of significance. By looking at the null distribution of the test statistic it can be seen to get increasingly concentrated with the increase in sample size (see Table 2). Further the null distribution is seen to be increasingly concentrated around 0 as the value of κ increases.

Since the assumptions involved in the usual large sample approximation ($-2\ln\Lambda \overset{\mathrm{d}}{\approx} \chi^2$) are violated here, it is not appropriate to use this approximation for the distribution of an LRT. We obtain the power of the LRT with $\kappa = 1$ through extensive simulations. The power of this LRT does not show a perceptible increase with increase in sample size (see Table 3).

The LRT shows (see Table 4) encouraging power performance for κ even as small as 4. Also, the convergence of the power to 1 increases rapidly with κ. This is expected because with the higher value of the concentration parameter the observations tend to be close together making it 'easier' for us to detect an outlier.

TABLE 2

Percentiles of the null distribution of the LRT ($\kappa = 1$)

Percentiles	$n = 10$	$n = 20$	$n = 30$
1	0.1353	0.1353	0.1353
2	0.1354	0.1353	0.1353
5	0.1357	0.1354	0.1354
10	0.1367	0.1356	0.1354
25	0.1441	0.1371	0.1362
50	0.1786	0.1448	0.1398
75	0.2638	0.1707	0.1527
90	0.3884	0.2224	0.1774

TABLE 3

Power of the LRT ($\alpha = 0.05, \kappa = 1$)

Δ (°)	$n = 10$	$n = 20$	$n = 30$
20	0.056	0.055	0.055
40	0.061	0.058	0.057
60	0.062	0.059	0.059
80	0.063	0.057	0.059
100	0.065	0.081	0.055
120	0.074	0.086	0.058
140	0.087	0.090	0.059
160	0.085	0.089	0.057
180	0.088	0.095	0.065

TABLE 4

Variation in the power of the LRT with κ ($\alpha = 0.05, n = 10$)

Δ (°)	$\kappa = 1$	$\kappa = 2$	$\kappa = 4$	$\kappa = 10$
15	0.058	0.050	0.057	0.070
30	0.056	0.049	0.062	0.141
60	0.060	0.062	0.165	0.619
90	0.063	0.098	0.437	0.964
120	0.074	0.146	0.776	1.000
150	0.077	0.240	0.947	1.000
180	0.082	0.293	0.986	1.000

5. Examples

We illustrate the above tests through two well-known examples on directional data. For both these examples, we assume that the 'known' value of κ is $\hat{\kappa}$, the MLE, as obtained from the data. The relevant computations as needed below were done through DDSTAP (SenGupta, 1998), a statistical package for the analysis of directional data. We tested both these datasets for circular uniformity using the Rayleigh test and found them to be compatible with the assumptions of the circular normal model.

Example 1. Fisher & Lewis (1983) give data from three samples of palæocurrent orientations from three bedded sandstone layers, measured on the Belford Anticline, New South Wales. We consider here the first sample. The dataset is 284°, 311°, 334°, 320°, 294°, 137°, 123°, 166°, 143°, 127°, 244°, 243°, 152°, 242°, 143°, 186°, 263°, 234°, 209°, 267°, 315°, 329°, 235°, 38°, 241°, 319°, 308°, 127°, 217°, 245°, 169°, 161°, 263°, 209°,

228°, 168°, 98°, 278°, 154°, 279°. Our LRT picked up the observation 38 as an outlier with an observed value of $\Lambda = 0.1675$ and $P = 0.01$. This outlier could be attributed to the segment of the sandstone layer corresponding to this observation being (inconsistently) disorientated by some 'external shocks'.

Example 2. We next consider the famous roulette wheel data obtained from Mardia (1972). The dataset is 43°, 45°, 52°, 61°, 75°, 88°, 88°, 279°, 357°. This dataset has previously been analysed by Bagchi & Guttman (1990), who assumed a circular normal distribution for it. With this assumption, the LRT picked up the observation 279° as an outlier with $\Lambda = 0.0276$ and $P = 0.12$. This could be attributed to, for example, subject to verification, a brief mechanical failure or a sudden electrical surge during that spinning time of the wheel which resulted in the observation 279°.

The identification of the observation 279° in Example 2 illustrates the fact that, unlike the linear case, internal values can actually be outliers in the context of directional data. This incidentally also coincides with the analysis done by Bagchi & Guttman (1990).

Once an outlier has been detected as above, one can discard it and proceed with further statistical analyses as needed using the rest of the dataset. Alternatively, one may fit an extended model to the entire dataset, say a contaminated or a mixture model with a circular normal distribution, which should give a better fit than the original with only a circular normal distribution.

6. Locally most powerful type test for outliers

In this section we assume μ, κ to be known and δ to be unknown. Fix $1 \le j \le n$. We derive an LMP test (LMPT) of H_0 against H_j. Motivated by this we propose an LMPTT of H_0 against H_1^*.

Theorem 6.1. *In testing H_0 against H_j the* LMPT *is:*

Reject H_0 if $\sin(\Theta_j - \mu) > c$ for some constant c depending on the size of the test.

Proof. The log-likelihood under H_j is given by

$$\ln L_j(\delta; \theta_1, \dots, \theta_n) = K + \sum_{i \neq j} \cos(\theta_i - \mu) + \cos(\theta_j - \mu - 2\delta)\,,$$

where K is constant. Then, the score function is given by

$$S_j(\delta) = \frac{\partial \ln L_j}{\partial \delta} = 2\sin(\theta_j - \mu - 2\delta)\,.$$

Hence $S_j(0) = 2\sin(\theta_j - \mu)$. Thus the LMPT statistic for testing H_0 against H_j is $\sin(\Theta_j - \mu)$: we reject H_0 if $\sin(\Theta_j - \mu) > c$ for some constant c.

Motivated by Theorem 6.1, for testing H_0 against H_1^* we propose the test:

$$\text{Reject } H_0 \text{ if } \max_{1 \le j \le n} \sin(\Theta_j - \mu) > c,$$

where c is a constant to be determined from the size condition. The exact sampling null distribution of the test statistic can be obtained using standard techniques. Recall that we designate $\theta_{\hat{j}}$ as the outlier if H_0 is rejected and $\sin(\theta_{\hat{j}} - \mu) = \max_{1 \le j \le n} \sin(\theta_j - \mu)$.

TABLE 5

Performance of the Bayes test for outlier ($\kappa = 2,\ n = 10,\ p = 0.05, 0.06$ *and* 0.07)

μ_1	H_0 not rejected			H_1 accepted Correct observation			H_1 accepted Wrong observation		
	$p = 0.05$	$p = 0.06$	$p = 0.07$	$p = 0.05$	$p = 0.06$	$p = 0.07$	$p = 0.05$	$p = 0.06$	$p = 0.07$
15	1000	1000	1000	0	0	0	0	0	0
30	1000	1000	1000	0	0	0	0	0	0
45	1000	1000	731	0	0	94	0	0	175
60	1000	735	373	0	121	294	0	144	333
75	754	453	271	146	295	395	100	252	334
90	527	343	204	315	420	506	158	237	290
120	343	213	136	533	625	662	124	162	202
150	240	157	97	666	722	765	94	121	138
180	214	125	84	712	763	785	84	112	131

TABLE 6

Performance of the LRT, LMP, L and M test for outlier ($\kappa = 2, n = 10$)

μ_1	H_0 not rejected				H_1 accepted Correct observation				H_1 accepted Wrong observation			
	LRT	LMP	L	M	LRT	LMP	L	M	LRT	LMP	L	M
15	955	948	948	965	6	14	9	4	39	38	43	31
30	910	943	938	965	24	21	11	3	66	36	51	32
45	935	940	930	970	18	24	12	4	47	36	58	26
75	906	924	902	964	45	44	30	21	49	32	68	15
90	877	922	893	966	78	41	41	27	45	37	66	7
120	766	936	823	951	181	25	89	44	53	39	88	5
150	698	938	732	934	253	13	184	64	49	49	84	2
180	681	945	743	912	271	5	170	84	48	50	87	4

7. Comparison of the various procedures of identifying an outlier

We provide a simulation-based comparison of the various procedures for identifying an outlier. The statistics which are included in this comparison are the L statistic (Collett, 1980), the M statistic (Mardia, 1975), the Bayes statistic, the LRT statistic and the LMPTT statistic.

For comparison, 1000 samples of size 10 each containing one outlier are generated such that nine observations are drawn from CN(0, 2) while the outlying observation is drawn from CN(μ_1, 2). We vary μ_1 to measure the effectiveness of the procedures in detecting outliers of different severity. For L, M, LRT and LMPTT statistics we record the frequency of acceptance of the null hypothesis, the alternative hypothesis and also the number of times the correct observation is identified as the outlier. For the Bayes rule we also need to specify the prior probability p of any of the observations being an outlier. Since the performance of the Bayes rule is seen to depend on the value of p we examine the performance of the test for several values of p. The results of these investigations are given in Tables 5–6.

There are several well-known criteria for comparing tests of outliers (Barnett & Lewis, 1994). We compare the above test statistics based on P_1 (the power of the test) and P_3 (the probability that the presence of outlier is signalled and the outlier is correctly identified). A good test has high P_1 and low $P_1 - P_3$ (which is the probability that the test wrongly identifies a good observation as an outlier). Based on these criteria, we see from Tables 5 and 6 that the LMPTT performs best for small values of μ_1, the most difficult ones to detect, as expected. The LRT performs best for moderate values of μ_1 and the Bayes rule performs best for large values of μ_1 which incidentally may attract the most penalties in practical situations. We also note that the Bayes rule gives increasingly better results with larger values of p, which is expected since the dataset actually contains an outlier.

The Bayes rule with $p = p_0$, where p_0 is such that $1 - np_0 = 0.5$, can be used in cases when there is no prior information about an outlier being present in the dataset. Since the efficacy of the Bayes rule in detecting the presence of an outlier is seen to increase with p, a higher value of p should be specified in cases where the presence of outlier is suspected. The LMPTT appears to perform slightly better than the LRT when outliers of lesser severity are to be detected (i.e. μ_1 is small) which is expected due to the nature of the LMPTT. However, if we are interested in detecting outliers of moderate to large severity the LRT performs better than the LMPTT. Moreover, we note that the Bayes procedure, the LRT and the LMPTT all perform better than the tests based on the L and M statistics.

In this paper we have considered only the case where κ is known. The case where κ is unknown is quite challenging and no work on it seems to be available. The difficulties stem from the fact that κ is neither a location nor a scale parameter, so that the usual techniques of reduction by invariance or similarity cannot be applied to get unconditional tests free of the nuisance parameter κ. In case an exact or small sample solution is required, one may try to use a conditional likelihood approach by taking $L_c(\mu) = L(\mu; S \mid C = c)$, where $S = \sum_i \sin\theta_i$ and $C = \sum_i \cos\theta_i$, whose distribution is free of κ under H_0. Another approach could be to try with a Bayesian idea by using the integrated likelihood (Aitkin & Rubin, 1985). We set a proper prior density (with respect to Lebesgue measure on $\mathbb{R}^+$) for κ, say $\pi(\kappa)$ and obtain the integrated likelihood $L^*(\mu)$. $L_c(\mu)$ or $L^*(\mu)$ may then replace the likelihood function used in our approaches above. If a large sample is available, one can try to use techniques available to construct such tests as to be free of a general nuisance parameter. The test statistics do not however seem to come out in elegant forms. We are currently investigating several such approaches to this problem.

References

AITKIN, M. & RUBIN, D.B. (1985). Estimation and hypothesis testing in finite mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **47**, 67–75.

BAGCHI, P. & GUTTMAN, I. (1988). Theoretical considerations of the multivariate von Mises–Fisher distribution. *J. Appl. Statist.* **15**, 149–169.

BAGCHI, P. & GUTTMAN, I. (1990). Spuriosity and outliers in directional data. *J. Appl. Statist.* **17**, 341–350.

BARNETT, V. & LEWIS, T. (1994). *Outliers in Statistical Data*, 3rd edn. New York: John Wiley and Sons.

COLLETT, D. (1980). Outliers in circular data. *Appl. Statist.* **29**, 50–57.

FERGUSON, T.S. (1967). *Mathematical Statistics — A Decision Theoretic Approach.* London: Academic Press.

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

FISHER, N.I. & LEWIS, T. (1983). Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika* **70**, 333–341. Errata (1984) **71**, 655.

JANDER, C. (1957). Die optiche Richtungsorientierung der roten Waldameise (*Formica rufa* L.). *Z. vergl. Physiologie* **40**, 162–238.

MARDIA, K.V. (1972). *Statistics of Directional Data.* London: Academic Press.

MARDIA, K.V. (1975). Statistics of directional data (with Discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **37**, 349–393.

SENGUPTA, A. (1998). *DDSTAP — A Software Package for Statistical Analysis of Directional Data.* Indian Statistical Institute.

SENGUPTA, A. & PAL, C. (2001). On optimal tests for isotropy against the symmetric wrapped stable-circular uniform mixture family. *J. Appl. Statist.* **28**, 129–143.

UPTON, G. (1993). Outliers in circular data. *J. Appl. Statist.* **20**, 229–235.