THE SLIPPAGE PROBLEM FOR THE CIRCULAR NORMAL DISTRIBUTION

ASHIS SENGUPTA1∗ AND ARNAB KUMAR LAHA1,2

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
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 'disoriented' due to geological or environmental disruptions which may give rise to possible outliers in the sample of palaeocurrent 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 $\text{CN}(\mu, \kappa)$ denote the circular normal distribution. As is usual, we refer to $\mu$ as the mean direction and $\kappa$ as the concentration parameter. We want to test $H_0$: $\Theta_j, j = 1, \ldots, n$ are identically distributed as $\text{CN}(\mu_0, \kappa)$ against $H_i$: $\Theta_1, \ldots, \Theta_{i-1}, \Theta_i, \Theta_{i+1}, \ldots, \Theta_n$ are identically distributed as $\text{CN}(\mu_0, \kappa)$ and $\Theta_i$ is distributed as $\text{CN}(\mu_1, \kappa)$, $1 \leq i \leq n$, $\mu_1 > \mu_0$, $\mu_1$, $\mu_0$ and $\kappa$ 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 $H_i^*$; namely there exists $i$, $i$ unknown, such that $\Theta_i$ is distributed as $\text{CN}(\mu_1, \kappa)$ and $\Theta_1, \ldots, \Theta_{i-1}, \Theta_{i+1}, \ldots, \Theta_n$ are distributed as $\text{CN}(\mu_0, \kappa)$ and all are independent. In the case that $\mu_1 > \mu_0$, $\mu_1$, $\mu_0$ and $\kappa$ 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. $H_i^*$ when the parameters $\mu_0$ and $\mu_1$ are both unknown but $\kappa$ 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 $\kappa$ (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 $\Theta_1, \ldots, \Theta_n$ are independent $\text{CN}(\mu_i, \kappa)$ random variables with density

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

We assume that $\mu_0 < \mu_1$ and $n \geq 3$. We are interested in finding the Bayes rule for the multiple decision problem of accepting one of the $n + 1$ hypotheses $H_0, H_1, \ldots, H_n$ with respect to the prior distributions invariant under permutations of $H_1, \ldots, H_n$. 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 $H_1, \ldots, H_n$ give equal weight to $H_1, \ldots, H_n$ and hence they are of the form $\tau_p$, where $\tau_p(H_0) = 1 - np, 0 \leq p \leq (1/n), \tau_p(H_i) = p, 1 \leq i \leq n$. Let $\Theta = (\Theta_1, \ldots, \Theta_n)$ and let $\Phi(\Theta) = (\phi(1 | \Theta), \ldots, \phi(n | \Theta))$ denote a generalized critical function or a multiple decision rule (Ferguson, 1967 p. 299) with $\phi(i | \Theta)$, $i = 1, \ldots, n$ taking values 0 or 1, and $\sum \phi(i | \Theta) = 1$. Thus $\Phi$ chooses $H_j$ when $\Theta = \theta$ is observed if $\phi(i | \Theta) = 1$. Let $R_j$ denote the likelihood ratio:

$$R_j = \frac{f(\Theta_j; \mu_j)}{f(\Theta_i; \mu_0)} = \exp \left( \kappa \left( \cos(\Theta_j - \mu_i) - \cos(\Theta_j - \mu_0) \right) \right) \quad (j = 1, 2, \ldots, n).$$

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

**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,

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

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

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

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

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

(a) $\Pr(\phi(i; \Theta) = 0 | H_0 \text{ is true}) = 1 - \frac{1}{n} + \frac{G(u)^n}{n},$

(b) $\Pr(\phi(0; \Theta) = 0 | H_0 \text{ is true}) = 1 - G(u)^n.$
Proof. Observe that, \( R_j = \exp \{2x \sin \delta \sin(\Theta_j - \mu - \delta)\} \). Thus, as a consequence of our assumption \( 0 \leq \mu_0 < \mu_1 < 2\pi \), \( R_j = \exp \{2x \sin \delta \sin(\Theta_j - \mu - \delta)\} \) because \( \exp(x) \) is an increasing function of \( x \), and \( 2x \sin \delta > 0 \) 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 \( \eta \). If there is a branch of \( \sin^{-1} \theta \) which is monotone on \([\pi/2, 3\pi/2]\), then the inverse transformation defined uniquely through the domain of \( \theta \) as partitioned below (where \( \theta \) greater than or equal to \( 2\pi \) is to be interpreted as being reduced modulo \( 2\pi \)), is

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

After some calculation we find the density of \( \eta \) 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 \( W = \max\eta_i \). Then \( W \) has cumulative distribution function (cdf) \( H \) where \( H(u) = G(u)^n \). Let \( W^*_i = \max_{\ell \neq i} \eta_i \). Then the cdf of \( W^*_i \) is \( H^*_i(u) = G(u)^n \). Now

\[
\Pr\left( \max R_j < \frac{1 - np}{p} \right) = \Pr\left( W < \frac{\ln\left(1 - \frac{np}{n}\right)}{2x \sin \delta} \right) = H(u).
\]

Note that the event \( R_j < \max R_j \) is equivalent to the event \( \eta_j < W \). Further, we have \( \Pr(\eta_j \geq W) = \Pr(\eta_j \geq \eta_s, s \neq i) \) because the distribution of the \( \eta_i \) is continuous. Thus, \( \Pr(\eta_j \geq \max \eta_s) = \int_1^1 (1 - G(\eta)) dH^*_i(\eta) = (n - 1) \int_0^1 (1 - y)^{n-2} dy = 1/n \) where we put \( y = G(\eta) \) and \( dH^*_i(\eta) = (n - 1)G(\eta)^{n-2}dG(\eta) \). Now \( \Pr(\phi(i; \Theta) = 0) = \Pr(\eta_j < W) + \Pr(W < u) - \Pr(\eta_j < W < u) \).

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

\[
\Pr(\eta_j < W < u) = \Pr(\eta_j < W^*_j < u) = \int_1^u G(\eta) dH^*_i(\eta) = (n - 1) \int_0^u G(\eta)^{n-1} dG(\eta) = (n - 1) \int_0^G(u) y^{n-1} dy = \frac{n - 1}{n} G(u)^n.
\]

It follows that

\[
\Pr(\phi(i; \Theta) = 0 \mid H_0 \text{ is true}) = 1 - \frac{1}{n} + \frac{G(u)^n}{n};
\]

\[
\Pr(\phi(i; \Theta) = 1 \mid H_0 \text{ is true}) = 1 - \Pr(\phi(i; \Theta) = 0 \mid H_0 \text{ is true}) = 1 - \frac{G(u)^n}{n};
\]

\[
\Pr(\phi(0; \Theta) = 0 \mid H_0 \text{ is true}) = 1 - G(u)^n;
\]

\[
\Pr(\phi(0; \Theta) = 1 \mid H_0 \text{ is true}) = \Pr(W \leq u) = G(u)^n.
\]

Theorem 2.3 gives the performance of the Bayes rule when \( H_j \) is true. Let \( \eta_j \) be as in the proof of Theorem 2.2 and let \( G \) be its cdf when \( \Theta_j \equiv \Xi \) and let \( G^* \) denote its cdf when \( \Theta_i \equiv \Xi \).
Theorem 2.3. In the framework of Theorem 2.1, let $1 \leq i, j \leq n$,

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

$$= 1 + \int_{-1}^{u} G(w)^{k-2}G^*(w) dG(w) - (k - 2) \int_{-1}^{1} G(w)^{k-3}(1 - G(w))G^*(w) dG(w) - \int_{-1}^{1} G(w)^{k-2}(1 - G(w))G^*(w) dG(w)$$

where $i \neq j$,

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

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

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

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

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 $\eta_j$ 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 $0 \leq \mu_0 < \mu_1 < 2\pi, \mu_1, \mu_0$ and $\kappa$ 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

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

and that under $H_1^*$ is,

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

© Australian Statistical Publishing Association Inc. 2001
Thus the LRT statistic \( \Lambda \) is given by
\[
- \ln \Lambda = \max_j \left\{ \kappa \left( \sum_{i \neq j} \cos(\theta_i - \mu) + \cos(\theta_j - \mu - 2\delta) - \sum_i \cos(\theta_i - \mu) \right) \right\}
\]
\[
= \max_j \left\{ \kappa \left( \cos(\theta_j - \mu - 2\delta) - \cos(\theta_j - \mu) \right) \right\}
\]
\[
= \max_j \left\{ \kappa \left( 2\sin(\theta_j - \mu - \delta) \sin \delta \right) \right\}.
\]

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 \( \kappa \) is known but \( \mu \) and \( \delta \) 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 \( \mu \) 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 \( j^\ast \) 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 \( \Lambda \) is given by
\[
- \ln \Lambda = \kappa \left( \left( \sum_{i \neq j} \cos \theta_i \right) (\cos \hat{\mu}_1^* - \cos \hat{\mu}_0) + \left( \sum_{i \neq j} \sin \theta_i \right) (\sin \hat{\mu}_1^* - \sin \hat{\mu}_0) + 1 - \cos(\theta_j - \hat{\mu}_0) \right).
\]

**Proof.** The log-likelihood under \( H_0 \) is
\[
\ln L_0(\theta_1, \ldots, \theta_n) = -n \ln \left( 2\pi I_0(\kappa) \right) + \kappa \sum_i \cos(\theta_i - \mu)
\]
\[
= -n \ln \left( 2\pi I_0(\kappa) \right) + \kappa \left( \cos \mu \sum_i \cos \theta_i + \sin \mu \sum_i \sin \theta_i \right).
\]

Putting \( \partial \ln L_0 / \partial \mu = 0 \) and solving for \( \mu \) gives \( \hat{\mu}_0 = \tan^{-1} \left( \sum_i \sin \theta_i / \sum_i \cos \theta_i \right) \) with \( \tan^{-1} \) so defined as to be unique. Under \( H_1^* \) the log-likelihood is
\[
\ln L_1^*(\theta_1, \ldots, \theta_n) = \max_j \left\{ -n \ln 2\pi I_0(\kappa) + \kappa \left( \cos \mu \sum_{i \neq j} \cos \theta_i + \sin \mu \sum_{i \neq j} \sin \theta_i + \cos(\theta_j - \dot{\mu}_0 - 2\delta) \right) \right\}.
\]

Thus, \( L_1^* = \max_j L_j \). Fix \( 1 \leq j \leq n \). We now compute \( \hat{\mu}_j \) and \( \hat{\delta}_j \) which are the maximum likelihood estimates (MLEs) of \( \mu \) and \( \delta \) under \( H_j \). Setting
\[
\frac{\partial \ln L_j}{\partial \mu} = 0 \quad \text{and} \quad \frac{\partial \ln L_j}{\partial \delta} = 0
\]
and solving for \( \mu \) gives
\[
\hat{\mu}_j = \tan^{-1} \left( \frac{\sum_{i \neq j} \sin \theta_i}{\sum_{i \neq j} \cos \theta_i} \right) \quad \text{and} \quad \hat{\delta}_j = \frac{\theta_j - \hat{\mu}_j}{2}.
\]

Let \( j^* \) be that \( j \) for which \( L_j \) attains its maximum value after substituting \( \hat{\mu}_j \) and \( \hat{\delta}_j \). Thus
under $H_1^*$ the estimate of $\mu$ is $\hat{\mu}_j (= \bar{\mu}_1^*)$ and that of $\delta$ is $\hat{\delta}_j (= \bar{\delta})$. Therefore the LRT statistic $\Lambda$ is given by

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

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 $\Lambda$ 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, $CN(\Delta, 1)$ with $\Delta = 20(20)180$ (in degrees), and repeating it 5000 times.

Since the power function is symmetric about $180^\circ$ 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 $\kappa$ increases.

Since the assumptions involved in the usual large sample approximation $(-2 \ln \Lambda \approx \chi^2)$ 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 $\kappa$ even as small as 4. Also, the convergence of the power to 1 increases rapidly with $\kappa$. 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 (κ = 1)

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1353</td>
<td>0.1353</td>
<td>0.1353</td>
</tr>
<tr>
<td>2</td>
<td>0.1354</td>
<td>0.1353</td>
<td>0.1353</td>
</tr>
<tr>
<td>5</td>
<td>0.1357</td>
<td>0.1354</td>
<td>0.1354</td>
</tr>
<tr>
<td>10</td>
<td>0.1367</td>
<td>0.1356</td>
<td>0.1354</td>
</tr>
<tr>
<td>25</td>
<td>0.1441</td>
<td>0.1371</td>
<td>0.1362</td>
</tr>
<tr>
<td>50</td>
<td>0.1786</td>
<td>0.1448</td>
<td>0.1396</td>
</tr>
<tr>
<td>75</td>
<td>0.2638</td>
<td>0.1707</td>
<td>0.1527</td>
</tr>
<tr>
<td>90</td>
<td>0.3884</td>
<td>0.2224</td>
<td>0.1774</td>
</tr>
</tbody>
</table>

TABLE 3
Power of the LRT (α = 0.05, κ = 1)

<table>
<thead>
<tr>
<th>Δ (°)</th>
<th>n = 10</th>
<th>n = 20</th>
<th>n = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.056</td>
<td>0.055</td>
<td>0.055</td>
</tr>
<tr>
<td>40</td>
<td>0.061</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td>60</td>
<td>0.062</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>80</td>
<td>0.063</td>
<td>0.057</td>
<td>0.059</td>
</tr>
<tr>
<td>100</td>
<td>0.065</td>
<td>0.081</td>
<td>0.055</td>
</tr>
<tr>
<td>120</td>
<td>0.074</td>
<td>0.086</td>
<td>0.058</td>
</tr>
<tr>
<td>140</td>
<td>0.087</td>
<td>0.090</td>
<td>0.059</td>
</tr>
<tr>
<td>160</td>
<td>0.085</td>
<td>0.089</td>
<td>0.057</td>
</tr>
<tr>
<td>180</td>
<td>0.088</td>
<td>0.095</td>
<td>0.065</td>
</tr>
</tbody>
</table>

TABLE 4
Variation in the power of the LRT with κ (α = 0.05, n = 10)

<table>
<thead>
<tr>
<th>Δ (°)</th>
<th>κ = 1</th>
<th>κ = 2</th>
<th>κ = 4</th>
<th>κ = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.058</td>
<td>0.050</td>
<td>0.057</td>
<td>0.070</td>
</tr>
<tr>
<td>30</td>
<td>0.056</td>
<td>0.049</td>
<td>0.062</td>
<td>0.141</td>
</tr>
<tr>
<td>60</td>
<td>0.060</td>
<td>0.062</td>
<td>0.165</td>
<td>0.619</td>
</tr>
<tr>
<td>90</td>
<td>0.063</td>
<td>0.098</td>
<td>0.437</td>
<td>0.964</td>
</tr>
<tr>
<td>120</td>
<td>0.074</td>
<td>0.146</td>
<td>0.776</td>
<td>1.000</td>
</tr>
<tr>
<td>150</td>
<td>0.077</td>
<td>0.240</td>
<td>0.947</td>
<td>1.000</td>
</tr>
<tr>
<td>180</td>
<td>0.082</td>
<td>0.293</td>
<td>0.986</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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 ˆκ, 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) disoriented 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 $\mu, \kappa$ to be known and $\delta$ to be unknown. Fix $1 \leq j \leq n$. 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

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

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:

reject $H_0$ if $\max_{1 \leq j \leq n} \sin(\theta_j - \mu) > c$,

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 $\theta_j$ as the outlier if $H_0$ is rejected and $\sin(\theta_j - \mu) = \max_{1 \leq j \leq n} \sin(\theta_j - \mu)$.

© Australian Statistical Publishing Association Inc. 2001
TABLE 5

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

\[
\begin{array}{cccccccccc}
\mu_1 & H_0 \text{ not rejected} & H_1 \text{ accepted} & & & & & & & \\
\hline
 & p = 0.05 & p = 0.06 & p = 0.07 & p = 0.05 & p = 0.06 & p = 0.07 & p = 0.05 & p = 0.06 & p = 0.07 \\
15 & 1000 & 1000 & 1000 & 0 & 0 & 0 & 0 & 0 & 0 \\
30 & 1000 & 1000 & 1000 & 0 & 0 & 0 & 0 & 0 & 0 \\
45 & 1000 & 1000 & 731 & 0 & 121 & 294 & 0 & 144 & 175 \\
60 & 1000 & 735 & 373 & 0 & 121 & 294 & 0 & 144 & 333 \\
75 & 754 & 453 & 271 & 146 & 295 & 395 & 100 & 252 & 334 \\
90 & 527 & 343 & 204 & 315 & 420 & 506 & 158 & 237 & 290 \\
120 & 343 & 213 & 136 & 533 & 625 & 662 & 124 & 162 & 202 \\
150 & 240 & 157 & 97 & 666 & 722 & 765 & 94 & 121 & 138 \\
180 & 214 & 125 & 84 & 712 & 763 & 785 & 84 & 112 & 131 \\
\end{array}
\]

TABLE 6

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

\[
\begin{array}{cccccccccccc}
\mu_1 & H_0 \text{ not rejected} & H_1 \text{ accepted} & & & & & & & & & \\
\hline
 & LRT & LMP & L & M & LRT & LMP & L & M & LRT & LMP & L & M \\
15 & 955 & 948 & 948 & 965 & 6 & 14 & 9 & 4 & 39 & 38 & 43 & 31 \\
30 & 910 & 943 & 938 & 965 & 24 & 21 & 11 & 3 & 66 & 36 & 51 & 32 \\
45 & 935 & 940 & 930 & 970 & 18 & 24 & 12 & 4 & 47 & 36 & 58 & 26 \\
75 & 906 & 924 & 902 & 964 & 18 & 24 & 12 & 4 & 47 & 36 & 58 & 26 \\
90 & 877 & 922 & 893 & 966 & 18 & 24 & 12 & 4 & 47 & 36 & 58 & 26 \\
120 & 766 & 936 & 823 & 951 & 18 & 24 & 12 & 4 & 47 & 36 & 58 & 26 \\
150 & 698 & 938 & 732 & 934 & 18 & 24 & 12 & 4 & 47 & 36 & 58 & 26 \\
180 & 681 & 945 & 743 & 912 & 271 & 5 & 170 & 84 & 48 & 50 & 87 & 4 \\
\end{array}
\]

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 \(CN(\mu_1, 2)\). We vary \(\mu_1\) 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 \(P_2\) (the probability that the presence of outlier is signalled and the outlier is correctly identified). A good test has high \(P_1\) and low \(P_1 - P_2\) (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 \(\mu_1\), the most difficult ones to detect, as expected. The LRT performs best for moderate values of \(\mu_1\) and the Bayes rule performs best for large values of \(\mu_1\).
values of $\mu_1$ 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. $\mu_1$ 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 $\kappa$ is known. The case where $\kappa$ is unknown is quite challenging and no work on it seems to be available. The difficulties stem from the fact that $\kappa$ 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 $\kappa$. In case an exact or small sample solution is required, one may try to use a conditional likelihood approach by taking $L_c(\mu) = L(\mu; S | C = c)$, where $S = \sum_i \sin \theta_i$ and $C = \sum_i \cos \theta_i$, whose distribution is free of $\kappa$ 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 $\kappa$, 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


