Abstract—The directional tests consider the sample on \([0, 2\pi]\) although the angular data of line segments are on \([0, \pi]\). We will derive the LM test statistic for the conditional von Mises distribution for testing non-uniformity of angles of line segments spread on the two dimensional plane, which has relatively high performance even for small samples as the extension of V-test to half circle. Recently, the source of disasters should be investigated statistically for the view of anti-disaster. We illuminate the performance of this test statistic by conducting simulation studies and also apply to angular data of active faults in Japan. Finally, we proposed a new non-hierarchical clustering method based on angular dispersions.

Index Terms—angular data, Rayleigh test, LM test, von Mises distribution, \(k\)-means.

I. INTRODUCTION

In Japan, there are so many earthquakes including smaller ones. On occasion of the 2011 Tohoku earthquake and the Great Hanshin earthquake, we had the great loss of human life in these disaster. The precise prediction of occurrence of the earthquakes has been expected for many years, however it is very difficult to include the estimations of locations and magnitudes. In the recent studies on seismology, active faults are useful for estimating locations and magnitudes of earthquakes. Active faults are the discontinuity of strata which has the distortion of each rock plane. The active faults may raise up the risk to bring about earthquakes.

The research on active faults and earthquakes has been studied for a long time. Especially the investigation into active faults and outbreak probabilistic model is also studied and discussed in Japan and other countries. We handle the active fault database collected by National Institute of Advanced Industrial Science and Technology[1]. The table I shows the specifications of the active faults in Japan. We should explore the property of active faults for the stand of anti-disaster from the statistical view-point.

In this article we investigate the distribution of active faults focusing on angles of faults devising the new test statistic with high performance even for small samples.

<table>
<thead>
<tr>
<th>TABLE I SPECIFICATIONS OF THE ACTIVE FAULTS IN JAPAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of faults named</td>
</tr>
<tr>
<td>Number of the sum of elements by segment</td>
</tr>
<tr>
<td>Total length [km]</td>
</tr>
</tbody>
</table>

In the first place, we shall be aware of dangerousness of the active faults in the same directions, and we are interested in testing uniformity of the directions of faults against the uniform directions[2]. We study the uniformity test of angular data of line segments.

Mardia[3] described the Rayleigh test when mean direction is given based on von Mises distribution ([4], [5]), whose density function is expressed by the following:

\[
f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)),
\]

where \(\kappa\) and \(\mu\) discribe the parameter of concentration and mean direction respectively. \(I_0(\kappa)\) is the modified Bessel function of order zero,

\[
I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos x)dx.
\]

The score statistic is obtained by using the convenient parameter transformation,

\[
\omega = (\kappa \cos \mu, \kappa \sin \mu)^T.
\]

Uniformity is corresponding to \(\kappa = 0\), and the transformation gives rise to \(\omega = 0\), which can give us very simple score statistic free from any evaluation of parameter estimation. The Rayleigh test statistic \(S_1\) is follows: then the null distribution is asymptotically \(\chi^2_2\).

\[
S_1 = 2n \bar{R}^2,
\]

where \(\bar{R}\) is the resultant vector of sample illustrated in appendix B. Another test is a just variant of the above test statistic, which is also known as Rayleigh test or V-test ([6], [7], [8], [9], [3], [10], [11], [12], [13], [14]):

\[
S_2 = \frac{2}{n} \left\{ \sum_{i=1}^{n} \cos(\mu - \theta_i) \right\}^2.
\]

The asymptotic distribution of \(S_2\) under uniformity is \(\chi^2_1\). The parameter \(\mu\) included in the statistic \(S_2\) is replaced by the specified direction \(\theta_0\). In our study we replaced \(\theta\) parameter by theoretical mean value of uniform distribution on \([0, 2\pi]\), and \(E(\theta_i) = \pi\). In our experiences \(S_1\) statistic does not have good performance in that it does not assure that \(\alpha\)-critical levels and high powers against alternatives (small \(\kappa\)) especially for small samples. In this article we will propose the new test statistic based on the LM test and investigate the behavior of this statistic and apply this test to active faults data in Japan. The normal von Mises distribution is defined on \([0, 2\pi]\), although the angles of the line segments such active faults must be the distribution defined on \([0, \pi]\). We
consider the following conditional distribution,
\[
g(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu))
\]
\[
= \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu))d\theta
\]
\[
(0 \leq \theta, \mu < \pi, \ k \geq 0)
\]

Tedious calculations of the likelihood give us the following new test statistic:
\[
S_3 = \left\{ \frac{2n}{\pi} \sin \mu + \sum_{i=1}^{n} \cos(\mu - \theta_i) \right\}^2
\]
\[
= \left\{ \frac{2n}{\pi} \sum_{i=1}^{n} \cos(\mu - \theta_i) \right\}^2
\]
Here we note that the theoretical mean value \(\pi/2\) is preferable to \(\mu\) to test uniformity on \([0, \pi]\). \(S_3\) is distributed \(\chi^2\) asymptotically.

We can say that the performance of the test statistics are as follows:
\[
S_1 \sim S_3 \approx S_2.
\]

However, as we conventionally use the test statistics \(S_1\) and \(S_2\) defined on full circle, and we must use the \(S_3\) from the conditioned distribution.

### TABLE II

<table>
<thead>
<tr>
<th>(n)</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.005</td>
<td>0.027</td>
<td>0.067</td>
<td>0.142</td>
</tr>
<tr>
<td>5</td>
<td>0.017</td>
<td>0.017</td>
<td>0.035</td>
<td>0.097</td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.106</td>
</tr>
<tr>
<td>20</td>
<td>0.000</td>
<td>0.065</td>
<td>0.146</td>
<td>0.235</td>
</tr>
<tr>
<td>50</td>
<td>0.001</td>
<td>0.015</td>
<td>0.043</td>
<td>0.095</td>
</tr>
<tr>
<td>100</td>
<td>0.022</td>
<td>0.054</td>
<td>0.097</td>
<td>0.168</td>
</tr>
</tbody>
</table>

\(A\) \(S_3\) (the width of sample range for \(\mu\)).

\(B\) \(S_3\) (the sample mean for \(\mu\)).

\(C\) \(S_3\) (the theoretical mean for \(\mu\)).

**IV. THE APPLICATION TO ACTIVE FAULTS DATA**

As we mentioned in the previous section, the active faults in the same angle may be the source of earthquakes. We test the uniformity of angles of line segments approximating the faults by picking up both ends of observed faults. The Figure 2, Table IV and Table V show the \(q = 1 - p\)-value calculated by prefecture of Japan. Thick color is corresponding to higher value of \(q\). The high \(q\)-values indicates possibilities of non-uniformity of angles of faults.

**A. generalized test**

In addition, we consider that angular data exist only in the restricted range \([c_1, c_2]\). We derived the former conditional

Fig. 1. The active faults in Japan

III. SMALL SAMPLE BEHAVIOR OF THE TEST STATISTICS

We study the small sample behavior of the test statistics based on the random segment spread on the two-dimensional space. The test statistic \(S_3\) is basically restricted to the half angular space \([0, \pi]\). For comparison of these statistics we generate the uniform random numbers on \([0, \pi]\), and we use the 2\(\theta_i\) for \(S_1\) and \(S_2\), \(\theta_i\) for \(S_3\), because the former two statistics are ones designed for testing for the uniformity of \([0, 2\pi]\). In this article we need just test statistic defined on half circle \([0, \pi]\). The Table II and Table III are the results of the simulation for 1,000,000 times which show that our proposed statistics conditioned on the half circle based on the conditional distribution has moderately good performance.
### Table III

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.008</td>
<td>0.123</td>
<td>0.234</td>
<td>0.365</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.004</td>
<td>0.053</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td>0.012</td>
<td>0.029</td>
<td>0.054</td>
<td>0.114</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.134</td>
<td>0.272</td>
<td>0.396</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.144</td>
<td>0.244</td>
<td>0.338</td>
<td>0.451</td>
</tr>
<tr>
<td></td>
<td>0.008</td>
<td>0.069</td>
<td>0.208</td>
<td>0.381</td>
</tr>
<tr>
<td></td>
<td>0.015</td>
<td>0.034</td>
<td>0.065</td>
<td>0.123</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.007</td>
<td>0.067</td>
<td>0.146</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td>0.188</td>
<td>0.273</td>
<td>0.374</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.248</td>
<td>0.343</td>
<td>0.432</td>
<td>0.537</td>
</tr>
<tr>
<td></td>
<td>0.113</td>
<td>0.233</td>
<td>0.353</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>0.021</td>
<td>0.045</td>
<td>0.079</td>
<td>0.141</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.007</td>
<td>0.067</td>
<td>0.146</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>0.168</td>
<td>0.268</td>
<td>0.362</td>
<td>0.473</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.305</td>
<td>0.400</td>
<td>0.488</td>
<td>0.591</td>
</tr>
<tr>
<td></td>
<td>0.310</td>
<td>0.453</td>
<td>0.571</td>
<td>0.689</td>
</tr>
<tr>
<td></td>
<td>0.032</td>
<td>0.063</td>
<td>0.106</td>
<td>0.177</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.325</td>
<td>0.470</td>
<td>0.594</td>
<td>0.720</td>
</tr>
<tr>
<td></td>
<td>0.312</td>
<td>0.425</td>
<td>0.523</td>
<td>0.628</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.343</td>
<td>0.446</td>
<td>0.539</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>0.738</td>
<td>0.831</td>
<td>0.888</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>0.067</td>
<td>0.121</td>
<td>0.185</td>
<td>0.280</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.858</td>
<td>0.923</td>
<td>0.957</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td>0.651</td>
<td>0.748</td>
<td>0.816</td>
<td>0.877</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$1%$</th>
<th>$2.5%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.400</td>
<td>0.515</td>
<td>0.613</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>0.966</td>
<td>0.983</td>
<td>0.991</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>0.141</td>
<td>0.227</td>
<td>0.319</td>
<td>0.436</td>
</tr>
<tr>
<td>Rayleigh test</td>
<td>0.997</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.914</td>
<td>0.950</td>
<td>0.970</td>
<td>0.983</td>
</tr>
</tbody>
</table>

A $S_4$ (the width of sample range for $\mu$).
B $S_3$ (the sample mean for $\mu$).
C $S_3$ (the theoretical mean for $\mu$).

**Algorithm 1** angular data clustering

**Input:**

$K$ (number of clusters)

$\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ (angular data of line segment)

**Output:**

$A = \{a_1, a_2, \ldots, a_k\}$

$l(i) | i = 1, \ldots, n$ (cluster labels of $\theta_i$)

$A \leftarrow$ choose $K \alpha_j$‘s from angular data sample $\Theta$ as the nodes randomly

**repeat**

$A_{previous} \leftarrow A$

**for** $i = 1$ to $n$ do

$l(i) \leftarrow \arg \min_j [1 - \cos (2\theta_i - 2\alpha_j)]$

**end for**

**for** $j = 1$ to $k$ do

$A \leftarrow \text{mean.direction}(l(i) | l(i) = j)$ from (13)

**end for**

until $A_{previous} \neq A$

For comparing our proposed method and general $k$-means, we plotted the histogram of angular data colored by the
clusters. As illustration, look at the lineament data appeared the circular test book [15]. Our proposed method can handle the connectivity of the 0 and π in the case of the angles of line segments data. Here we note that usual circular data have property the connection of 0 and 2π. Then the histogram Figure 3 is seemed to be separated dark colored one cluster considering the 0-π connectivity. However the standard k-means method results in detecting unusual two clusters owning to disconnection of 0 and π (see Figure 4).

The result of clustering Japanese active faults with the number of clusters k = 2. Then we had the LM test the each clustered data and the p-values are 0.00075, 0.00015.

VI. CONCLUSION AND FUTURE

The purpose of this article was to derive the test statistic of angular data of line segments for describing the property of data. Then we proposed the LM test statistic.

APPENDIX A
DERIVATION OF LM TEST STATISTIC

Likelihood function and Log likelihood function of the conditional von Mises distribution (6) is

\[
L(\theta; \mu, \kappa) = \prod_{i=1}^{n} f(\theta_i; \mu, \kappa),
\]

\[
l = \sum_{i=1}^{n} \log f(\theta_i; \mu, \kappa) = \sum_{i=1}^{n} [\kappa \cos(\theta_i - \mu) - \log I_0(\kappa) - \log(2\pi) - \log Z(\mu, \kappa)],
\]

where \(I_0\) denotes the modified Bessel function of the first kind and order 0, and \(Z(\mu, \kappa)\) can be defined by

\[
Z(\mu, \kappa) = \int_{0}^{\pi} \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)) d\theta.
\]

The score statistic with \(\kappa = 0\) is calculated using first and second derivatives. Then, \(Z(\mu, 0) = 1/2\). Take partial differentiation of \(Z(\mu, 0)\) with respect to each parameters \(\mu\) and \(\kappa\).

\[
\frac{\partial Z(\mu, 0)}{\partial \mu} = 0,
\]

\[
\frac{\partial Z(\mu, 0)}{\partial \kappa} = \int_{0}^{\pi} \left[ \frac{1}{2\pi} \cos(\theta - \mu) \right] d\theta.
\]

\[
\frac{\partial^2 Z(\mu, 0)}{\partial \mu^2} = 0.
\]

\[
\frac{\partial^2 Z(\mu, 0)}{\partial \mu \partial \kappa} = \int_{0}^{\pi} \left[ \frac{1}{4\pi} \sin(\theta - \mu) \right] d\theta.
\]

\[
\frac{\partial^2 Z(\mu, 0)}{\partial \kappa^2} = \int_{0}^{\pi} \left[ \frac{1}{2\pi} + \frac{1}{2\pi} \cos(\theta - \mu) \right]^2 d\theta.
\]

Thus, substitute each partial differentiation for this result where \(A(\kappa) = I_1(\kappa)/I_0(\kappa)\).

\[
\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} \left[ 0 \cdot \sin(\theta_i - \mu) - \frac{1}{Z(\mu, 0)} \frac{\partial Z(\mu, 0)}{\partial \mu} \right] = 0.
\]

\[
\frac{\partial l}{\partial \kappa} = \sum_{i=1}^{n} \left[ \cos(\theta_i - \mu) - A(0) - \frac{1}{Z(\mu, 0)} \frac{\partial Z(\mu, 0)}{\partial \kappa} \right] = \sum_{i=1}^{n} [\cos(\theta_i - \mu)] - \frac{2n}{\pi} \sin(\mu).
\]
Fig. 5. The result of clustering Japan active faults

\[
\frac{\partial^2 l}{\partial \mu^2} = \sum_{i=1}^{n} \left[ -2 \cdot \cos(\theta_i - \mu) \right]
\]

\[
= \left( \frac{-1}{Z^2(\mu, 0)} \frac{\partial Z(\mu, 0)}{\partial \mu} \right)^2
\]

\[
+ \frac{1}{Z(\mu, 0)} \frac{\partial^2 Z(\mu, 0)}{\partial \mu^2} \right]
\]

\[
= 0.
\]

\[
\frac{\partial^2 l}{\partial \mu \partial \kappa} = \sum_{i=1}^{n} \left[ \sin(\theta_i - \mu) \right]
\]

\[
= \left( \frac{-1}{Z^2(\mu, 0)} \frac{\partial Z(\mu, 0)}{\partial \kappa} \right)^2
\]

\[
+ \frac{1}{Z(\mu, 0)} \frac{\partial^2 Z(\mu, 0)}{\partial \kappa^2} \right]
\]

\[
= \sum_{i=1}^{n} \left[ \sin(\theta_i - \mu) \right] - \frac{2n}{\pi} \cos(\mu).
\]

\[
\frac{\partial^2 l}{\partial \kappa^2} = \sum_{i=1}^{n} \left[ -A'(0) \right]
\]

\[
= \left( \frac{-1}{Z^2(\mu, 0)} \frac{\partial Z(\mu, 0)}{\partial \kappa} \right)^2
\]

\[
+ \frac{1}{Z(\mu, 0)} \frac{\partial^2 Z(\mu, 0)}{\partial \kappa^2} \right]
\]

\[
= \frac{4n}{\pi^2} \sin^2(\mu) - \frac{n}{2}.
\]

Finally, calculate the score statistic by Lagrange multiplier using Hessian matrix. we call this statistic LM test statistic.

\[
H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \kappa} \\ \frac{\partial^2 l}{\partial \mu \partial \kappa} & \frac{\partial^2 l}{\partial \kappa^2} \end{bmatrix}.
\]

\[
J = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial l}{\partial \mu} \\ \frac{\partial l}{\partial \kappa} \end{bmatrix}.
\]

\[
LM = d_2^T (H_{22} - H_{12}^T H_{11}^{-1} H_{21})^{-1} d_2.
\]

\[
S_4 = \frac{-2n}{\pi} \sin \hat{\mu} + \sum_{i=1}^{n} \cos (\hat{\mu} - \theta_i)^2.
\]

\[ (12) \]

\[
\bar{\theta} = \arg \left\{ \sum_{j=0}^{n} \cos \theta_j + i \sum_{j=0}^{n} \sin \theta_j \right\}.
\]

\[ (13) \]

Also, \( \hat{R} \) that is the length of resultant vector \( R \) is used for a measure of concentration of a data set.

\[
R = \left( \sum_{i=1}^{n} \cos \theta_i, \sum_{i=1}^{n} \sin \theta_i \right).
\]

\[ (14) \]

\[
\hat{R} = \frac{\| R \|}{n}.
\]

\[ (15) \]
REFERENCES


