Session 2(c): Adoption of Data Science and Analytics by the Academia III

Modelling Wind Direction using Simultaneous Linear Functional Relationship for von Mises Distribution

N. A. MOKHTAR[a], Y. Z. ZUBAIRI[b] & A. G. HUSSIN[c]
[a,c] Faculty of Defence Sciences and Technology,
National Defence University of Malaysia,
Kem Sungai Besi, 57000 Kuala Lumpur, Malaysia.
[b]Centre for Foundation Studies in Science,
University of Malaya, 50603 Kuala Lumpur, Malaysia.

PRESENTATION OUTLINE

- Introduction
- Literature Review
- Simultaneous Linear Functional Relationship Model
- Parameter Estimation
- Derivation of Covariance Matrix
- Simulation Study
- Simulation Result
- Application on Real Wind Direction Data
- Reference

INTRODUCTION

- Wind direction data is importantly used in meteorology as the knowledge of wind direction on a specific location contributes to accurate estimation of real power transmission capacity.
- The data is angular and to be more specific, it is circular.
- In this study, we consider the simultaneous linear functional relationship model for multivariate circular wind direction data.
- The three variables of the wind direction data in Malaysia coastal stations in Jan-Feb 2013 are considered with the von Mises distribution.
- x=wind direction data in Kota Kinabalu
- y1=wind direction in Kuala Terengganu
- y2= wind direction in Alor Setar

INTRODUCTION

Circular statistics

- Data are measured in the range $[0°, 360°)$ or $[0,2\pi)$ radians.

- Formal analysis cannot be done to circular data with usual statistical technique due to the wrap-around nature of a circle.

- The von Mises distribution is said to be the most useful distribution for circular data ***(Mardia and Jupp, 2000).***

- The p.d.f. of von Mises distribution: $$g(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta-\mu)}$$ I

where $I_0(\kappa)=\frac{1}{2\pi}\int_0^{2\pi}e^{\kappa\cos\theta}d\theta$; $0 \le x < 2\pi,\ 0 \le \mu < 2\pi,\ \kappa > 0.$

LITERATURE REVIEW

- Examples of circular data:

  - Kendall (1974) studied the bird navigation ***(Best and Fisher, 1979)***.
  - geologists studying the direction of the earth's magnetic pole ***(Batschelet, 1981)***.
  - directions of the stars ***(Fisher,1993)***
  - In medical application, Jammaladaka et al. (1986) discussed about the angle of knee flexion to assess the recovery of ortheopaedic patients ***(Jammalamadaka and Sengupta, 2001)***.

Simultaneous Linear Functional Relationship Model

- $Y_j = \alpha_j + X \ (mod\ 2\pi)$ 2
- $x_i = X_i + \delta_i$ → error: $\delta_i \sim VM(0, \kappa)$
- $y_{ji} = Y_{ji} + \varepsilon_{ji}$ → error: $\varepsilon_i \sim VM(0, v_j)$
- for $i = 1,2, \dots, n$ and $j = 1, \dots, q$
- P.d.f of von Mises distribution: $g(x; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)}$; 3

$$0 \le x < 2\pi,\ 0 \le \mu < 2\pi,\ \kappa > 0$$

- Assume Ratio of concentration parameter $\lambda = \frac{\nu}{\kappa} = 1$, thus $\nu = \kappa$
- $\log L\left(\alpha_j, \kappa, X_1, \dots, X_n; x_1, \dots x_n, y_{11}, \dots, y_{qn}\right) =$

$$-2n\log(2\pi) - (1+q)n\log I_0(\kappa) + \kappa\sum_{i=1}^{n}\cos(x_i - X_i) + \kappa\sum_{j=1}^{q}\sum_{i=1}^{n}\cos(y_{ji} - \alpha_j - X_i)$$

PARAMETER ESTIMATION

➢ Maximum Likelihood Estimator of $\boldsymbol{\alpha_j}, (\boldsymbol{j = 1, \ldots, q})$

❑ $$\hat{\alpha}_j = \begin{cases} tan^{-1}\left(\frac{S}{C}\right) & when\ S > 0, C > 0 \\ tan^{-1}\left(\frac{S}{C}\right) + \pi & when\ C < 0 \\ tan^{-1}\left(\frac{S}{C}\right) + 2\pi & when\ S < 0, C > 0 \end{cases} \quad (4)$$

➢ Maximum Likelihood Estimator of X_i

❑ $$\hat{X}_{i1} \approx \hat{X}_{i0} + \frac{sin(x_i - \widehat{X_{i0}}) + \sum_{j=1}^{q} sin(y_{ji} - \widehat{\alpha_j} - \widehat{X_{i0}})}{cos(x_i - \widehat{X_{i0}}) + \sum_{j=1}^{q} cos(y_{ji} - \widehat{\alpha_j} - \widehat{X_{i0}})} \quad (5)$$

PARAMETER ESTIMATION

➢ Maximum Likelihood Estimator of $\boldsymbol{\kappa}$

- $\hat{\kappa}$ can be obtained by using the approximation given by Fisher (1993)

$$A^{-1}(w) = \begin{cases} 2w + w^3 + \frac{5}{6}w^5 & when\ w < 0.53 \\ -0.4 + 1.39w + \frac{0.43}{(1-w)} & when\ 0.53 \le w < 0.85 \\ \frac{1}{w^3 - 4w^2 + 3w} & when\ w \ge 0.85 \end{cases} \quad (6)$$

- Hence, $\hat{\kappa} = A^{-1}(w)$ where

$$w = \left(\frac{1}{n(1+q)}\right)\left\{\sum_{i=1}^{n} cos\left(x_i - \hat{X}_i\right) + \sum_{j=1}^{q}\sum_{i=1}^{n} cos\left(y_{ji} - \hat{\alpha}_j - \hat{X}_i\right)\right\} \quad (7)$$

DERIVATION OF COVARIANCE MATRIX

- **The covariance matrix is derived based on Fisher Information matrix**

- $$F = \begin{bmatrix} R & 0 & W \\ 0 & S & 0 \\ W^T & 0 & U \end{bmatrix}$$

8

❑ $$R = \begin{bmatrix} (1+q)\kappa A(\kappa) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (1+q)\kappa A(\kappa) \end{bmatrix} \longrightarrow (n \times n)\ matrix$$

DERIVATION OF COVARIANCE MATRIX

- $W = \begin{bmatrix} \kappa A(\kappa) \\ \vdots \\ \kappa A(\kappa) \end{bmatrix} \longrightarrow (n \times 1) matrix$

  $and \quad W^T = [\kappa A(\kappa) \quad \dots \quad \kappa A(\kappa)] \longrightarrow (1 \times n)\ matrix$

- $S = [n(1+q)A'(\kappa)] \longrightarrow (1 \times 1)\ matrix$
- $U = [n\kappa A(\kappa)\ ] \longrightarrow (1 \times 1)\ matrix$
- $W^T R^{-1} W = [\kappa A(\kappa) \quad \dots \quad \kappa A(\kappa)] \begin{bmatrix} \frac{1}{(1+q)\kappa A(\kappa)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{(1+q)\kappa A(\kappa)} \end{bmatrix} \begin{bmatrix} \kappa A(\kappa) \\ \vdots \\ \kappa A(\kappa) \end{bmatrix}$

  $= \left[\frac{n\kappa A(\kappa)}{(1+q)}\right] \longrightarrow (1 \times 1) matrix$

DERIVATION OF COVARIANCE MATRIX

- $U - W^T R^{-1} W = [\, n\kappa A(\kappa)\,] - \left[\frac{n\kappa A(\kappa)}{(1+q)}\right] = n\kappa A(\kappa)\left[1 - \frac{1}{(1+q)}\right] = n\kappa A(\kappa)\left[\frac{1+q-1}{1+q}\right] = n\kappa A(\kappa)\left[\frac{q}{1+q}\right]$

- $[U - W^T R^{-1} W]^{-1} = n\kappa A(\kappa)\left[\frac{q+1}{q}\right]$

- $S^{-1} = \left[\frac{1}{n(1+q)A'(\kappa)}\right]$

- $cov\begin{bmatrix}\hat{\kappa} \\ \hat{\alpha}_j\end{bmatrix} = \begin{bmatrix}\frac{1}{n(1+q)A'(\kappa)} & 0 \\ 0 & n\kappa A(\kappa)\left[\frac{q+1}{q}\right]\end{bmatrix}$ (9)

- $var[\hat{\kappa}] = \frac{1}{n(1+q)A'(\kappa)}$ (10)

- $var[\hat{\alpha}_j] = n\kappa A(\kappa)\left[\frac{q+1}{q}\right]$ (11)

SIMULATION STUDY

- $Y_j = \alpha_j + X \ (mod\ 2\pi) \quad for \ \ j = 1, \dots, q$
- $x_i = X_i + \delta_i \ and \ \ y_{ji} = Y_{ji} + \varepsilon_{ji}$ for $i = 1,2, \dots, n$
- Programming was run where the simulation, *s*=5000
- Testing the biasness of parameters in the simultaneous model.
- $\kappa = 5, 8, 10, 15$
- *n*=30, 70, 100, 150 and 300

$\lambda = \frac{\nu}{\kappa} = 1, \ \text{thus}\ \nu = \kappa.$

SIMULATION STUDY

- Biasness of $\hat{\alpha}_j$
  - Mean of circular parameter $\hat{\alpha}_j, \bar{\hat{\alpha}}_j$:

$$\bar{\hat{\alpha}}_j = \begin{cases} tan^{-1}\left(\frac{S}{C}\right) & when\ S > 0, C > 0 \\ tan^{-1}\left(\frac{S}{C}\right) + \pi & when\ C < 0 \\ tan^{-1}\left(\frac{S}{C}\right) + 2\pi & when\ S < 0, C > 0 \end{cases}$$

$$where\ c = \sum_{j=1}^{s} \cos(\hat{\alpha}_j) \quad and \quad S = \sum_{j=1}^{s} \sin(\hat{\alpha}_j)$$

  - Circular distance, $d = \pi - |\pi - |\bar{\hat{\alpha}} - \alpha||$
  - Mean resultant length, $R = \frac{1}{s}\sqrt{\left(\sum_{j=1}^{s} cos(\hat{\alpha}_j)\right)^2 + \left(\sum_{j=1}^{s} sin(\hat{\alpha}_j)\right)^2}$

SIMULATION STUDY

➢ Biasness of $\tilde{\kappa}$

- Mean of $\tilde{\kappa}$, $\bar{\tilde{\kappa}} = \frac{1}{s}\sum_{j=1}^{s} \tilde{\boldsymbol{\kappa}}_i$
- Estimated bias, $EB = \bar{\tilde{\kappa}} - \kappa$
- Estimated Root Mean Square Errors, ERMSE=

$$\sqrt{\frac{1}{s}\sum_{i=1}^{s}(\tilde{\kappa}_i - \kappa)^2}$$

Maximum Likelihood Estimator of κ, which is $\hat{\kappa}$ has been corrected by the division by 2/3. Hence, a better approximation is then become $\tilde{\kappa} = \frac{2\hat{\kappa}}{3}$.

SIMULATION RESULT-Biasness of $\hat{\alpha}_1$

- Mean of circular parameter $\hat{\alpha}_j, \bar{\hat{\alpha}}$:

- $$\bar{\hat{\alpha}}_j = \begin{cases} tan^{-1}\left(\frac{S}{C}\right) & when\ S > 0, C > 0 \\ tan^{-1}\left(\frac{S}{C}\right) + \pi & when\ C < 0 \\ tan^{-1}\left(\frac{S}{C}\right) + 2\pi & when\ S < 0, C > 0 \end{cases}$$

$$where\ c = \sum_{i=1}^{s} \cos(\hat{\alpha}_{ji})\quad and\quad S = \sum_{i=1}^{s} \sin(\hat{\alpha}_{ji})$$

n	κ=5	κ=8	κ=10	κ=15
30	0.7843	0.7870	0.7850	0.7858
70	0.7871	0.7838	0.7847	0.7845
100	0.7852	0.7859	0.7855	0.7851
150	0.7856	0.7847	0.7856	0.7845
300	0.7854	0.7852	0.7852	0.7857

- Table 1 : Circular mean of $\hat{\alpha}_1$
- Circular mean of $\hat{\alpha}_1$ becomes closer to the real value of $\alpha_1 = \frac{\pi}{4} = 0.7854$, for large sample size n and κ.

- Circular distance, $d = \pi - |\pi - |\bar{\hat{\alpha}} - \alpha||$

n	κ=5	κ=8	κ=10	κ=15
30	0.0011	0.0016	0.0004	0.0004
70	0.0017	0.0016	0.0007	0.0009
100	0.0002	0.0005	0.0001	0.0003
150	0.0002	0.0007	0.0002	0.0009
300	0.0000	0.0002	0.0002	0.0003

- Table 2 :The circular distance of $\hat{\alpha}_1$

- There is a small circular distance or bias which suggests a good estimation of $\boldsymbol{\alpha_1}$

SIMULATION RESULT- Biasness of $\hat{\alpha}_1$

- Mean resultant length, $R = \frac{1}{s}\sqrt{\left(\sum_{j=1}^{s} cos\left(\hat{\alpha}_{ji}\right)\right)^2 + \left(\sum_{j=1}^{s} sin\left(\hat{\alpha}_{ji}\right)\right)^2}$

n	κ=5	κ=8	κ=10	κ=15
30	0.9919	0.9953	0.9965	0.9976
70	0.9965	0.9981	0.9985	0.9990
100	0.9976	0.9986	0.9989	0.9993
150	0.9984	0.9991	0.9993	0.9995
300	0.9992	0.9995	0.9996	0.9998

- Table 3 :The mean resultant length of $\hat{\alpha}_1$
- The mean resultant length becomes closer to 1 with increasing n and increasing κ, where this suggests a better accuracy.

SIMULATION RESULT- Biasness of $\hat{\alpha}_2$

n	κ=5	κ=8	κ=10	κ=15
30	0.7861	0.7856	0.7854	0.7859
70	0.7886	0.7840	0.7842	0.7837
100	0.7862	0.7861	0.7850	0.7844
150	0.7852	0.7854	0.7851	0.7845
300	0.7853	0.7852	0.7853	0.7857

- Table 4 : Circular mean of $\hat{\alpha}_2$
- Circular mean of $\hat{\alpha}_2$ becomes closer to the real value of $\alpha_2 = \frac{\pi}{4} = 0.7854$, for large sample size n and κ.

n	κ=5	κ=8	κ=10	κ=15
30	0.0007	0.0002	0.0000	0.0005
70	0.0032	0.0014	0.0012	0.0017
100	0.0008	0.0007	0.0004	0.0010
150	0.0002	0.0000	0.0003	0.0008
300	0.0001	0.0002	0.0001	0.0003

- Table 5 :The circular distance of $\hat{\alpha}_2$
- There is a small circular distance or bias which suggests a good estimation of $\boldsymbol{\alpha}_2$.

SIMULATION RESULT-Biasness of $\hat{\alpha}_2$

n	κ=5	κ=8	κ=10	κ=15
30	0.9920	0.9953	0.9965	0.9977
70	0.9966	0.9980	0.9984	0.9990
100	0.9976	0.9987	0.9990	0.9993
150	0.9984	0.9991	0.9993	0.9996
300	0.9992	0.9995	0.9997	0.9998

- Table 6 :The mean resultant length of $\hat{\alpha}_2$
- The mean resultant length becomes closer to 1 with increasing *n* and increasing κ, where this suggests a better accuracy.

SIMULATION RESULT-Biasness of $\tilde{\kappa}$

- Mean of $\tilde{\kappa}$, $\bar{\tilde{\kappa}} = \frac{1}{s}\sum_{j=1}^{s} \tilde{\boldsymbol{\kappa}}_i$

n	κ=5	κ=8	κ=10	κ=15
30	4.9868	8.2941	10.4789	15.9079
70	4.7099	7.9817	10.0778	15.2226
100	4.6642	7.9346	10.0058	15.1279
150	4.6181	7.8564	9.9340	14.9858
300	4.5685	7.8179	9.8498	14.9296

- Table 7 : The mean of $\tilde{\kappa}$
- The mean of $\tilde{\kappa}$ becomes closer to the true value of κ as the sample size n and the concentration parameter κ increase.

- Estimated bias, $EB = \bar{\tilde{\kappa}} - \kappa$

n	κ=5	κ=8	κ=10	κ=15
30	-0.0132	0.2941	0.4789	0.9079
70	-0.2901	-0.0183	0.0778	0.2226
100	-0.3358	-0.0654	0.0058	0.1279
150	-0.3819	-0.1436	-0.0660	-0.0142
300	-0.4315	-0.1821	-0.1502	-0.0704

- Table 8 : The estimated bias of $\tilde{\kappa}$
- There is also a good estimation of κ in particular for large sample size and large κ.

SIMULATION RESULT-Biasness of $\tilde{\kappa}$

- Estimated Root Mean Square Errors, ERMSE= $\sqrt{\frac{1}{s}\sum_{i=1}^{s}(\tilde{\kappa}_i-\kappa)^2}$

n	κ=5	κ=8	κ=10	κ=15
30	1.0972	1.6272	2.0103	3.1910
70	0.7986	1.0262	1.2523	1.8476
100	0.7217	0.8455	1.0326	1.5468
150	0.6477	0.7050	0.8218	1.2180
300	0.5656	0.5225	0.5864	0.8576

- Table 9 : The estimate root mean square error (ERMSE) of $\tilde{\kappa}$
- The estimate root mean square error (ERMSE) has decreased with the increasing value of n and κ . This suggests that $\tilde{\kappa}$ is a good estimator of κ.

APPLICATION TO REAL WIND DIRECTION DATA

- x=wind direction data in Kota Kinabalu
- y1=wind direction in Kuala Terengganu
- y2= wind direction in Alor Setar
  - $\hat{\alpha}_1 = 0.43609$
  - $\hat{\alpha}_2 = 0.82331$
  - $var\ (\hat{\alpha}_j) = 0.02628$
  - $\tilde{\kappa} = 1.67425$
  - $var\ (\tilde{\kappa}) = 0.02386$
  - Therefore, the estimated relationship for the variables are
    - $Y_1 = 0.43609 + X(mod\ 2\pi)$
    - $Y_2 = 0.82331 + X(mod\ 2\pi)$

REFERENCE

- Mardia K.V. and Jupp P. E. (2000) Directional Statistics, John Wiley & Sons.
- Batschelet E 1981. *Circular Statistic in Biology*. Academic Press.
- Best D. J. and Fisher N. I. 1979. Efficient Simulation of the von Mises Distribution. *Applied Statistics*: 28(2), 152-157.
- Jammalamadaka S. R. and Sengupta A. (2001). *Topics in Circular Statistics*, World Scientific Publishing.
- Hussin A.G, Hassan S. F. and Zubairi Y.Z. (2010) A Statistical Method to Describe the Relationship of Circular Variables Simultaneously, Pakistan Journal of Statistics: 26 (4), 593-607.
- Hussin A. G., Hassan S. F. , Zubairi Y. Z. , Zaharim A. and Sopian K. (2009) Approximation of Error Concentration Parameter for Simultaneous Circular Functional Model. European Journal of Scientific Research: 27(2), 258-263.
- Mardia K.V. (1972)Statistics of Directional Data, Academic Press London and New York.
- N. A. Mokhtar, Y. Z. Zubairi & A. G. Hussin. (2015). Parameter Estimation of Simultaneous Linear Functional Relationship Model for Circular Variables Assuming Equal Error Variances. Pakistan Journal of Statistics. Vol. 31(2): 251-265.