ON PARAMETER ESTIMATION, CONFIDENCE INTERVALS AND OUTLIER DETECTION FOR SOME CIRCULAR MODELS

MOHD SYAZWAN BIN MOHAMAD ANUAR

DOCTOR OF PHILOSOPHY (STATISTICS)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

2018

ON PARAMETER ESTIMATION, CONFIDENCE INTERVALS AND OUTLIER DETECTION FOR SOME CIRCULAR MODELS

MOHD SYAZWAN BIN MOHAMAD ANUAR

Thesis Submitted to the Centre for Graduate Studies, Universiti Pertahanan Nasional Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Statistics)

JANUARY 2018

ABSTRACT

This study focuses on the parameter estimation, confidence interval estimation and outlier detection for several types of the circular model. The models consider in this study are Down and Mardia Circular Regression Model, Circular Functional Relationship Model (CFRM) and a new model known as the Simultaneous Circular Functional Relationship Model (CSFRM). Firstly, for the Down and Mardia (DM) Circular Regression model, we propose the roots of a polynomial function (polyroot) and minimizing the negative value of the log-likelihood function (ms) to obtain the parameter estimates. The confidence interval for this model are obtained based on the normal distribution where the parameter are estimated using polyroot or minimum sum (ms) and estimation using the bootstrap technique. Secondly, we consider the parameters estimation and confidence intervals for CFRM. We used the polyroot and ms to estimate the concentration, angular and slope parameters. The confidence intervals for the parameters are obtained based on the normal distribution where the parameters are estimated using polyroot or ms and estimation using the bootstrap technique. Thirdly, we consider the outlier problem in CFRM using two statistics namely the Functional Difference Mean Circular Error Using Cosine (FDMCEc) statistic and Functional Difference Mean Circular Error Using Sine (FDMCEs) statististic. This methods also used the row deletion method to calculate FDMCEc and FDMCEs values for the full and reduced data set. In our CFRM, we determine the cut-off points for several values of sample size and concentration parameter designed using simulation study. Lastly, a new simultaneous circular functional relationship model (CSFRM) which is an extended version of a CFRM is developed. We want to study the relationship for more than two circular variables. We derived the maximum likelihood function of the model and the estimate the variance-covariance of parameters based on Fisher information matrix. Then, the efficiency of the model is assessed using the biasness for the circular variables using circular mean, circular distance and mean resultant length and mean, absolute estimated bias and estimated root mean square error for continuous variables respectively. Model verification and assessment of all methods and model proposed in this study are examined using the simulation study. The wind and wave direction data set used for illustration and application to real data set.

ABSTRAK

Kajian ini memberi tumpuan kepada penganggaran parameter, selang keyakinan dan titik terpencil untuk beberapa jenis model bulatan. Pertama, kami mempertimbangkan penganggaran parameter dan selang keyakinan bagi model regresi membulat,. Kaedah penganggaran parameter yang digunakan adalah penganggaran menggunakan fungsi punca polynomial (polyroot) dan meminimumkan nilai negatif bagi fungsi log-kebolehjadian (ms). Keyakinan kaedah anggaran selang untuk model ini adalah teknik berdasarkan taburan normal di mana parameter dianggarkan menggunakan fungsi polyroot atau ms dan anggaran menggunakan teknik bootstrap. Kedua, kami mempertimbangkan penganggaran parameter dan selang keyakinan untuk model hubungan berfungsi bagi pembolehubah bulatan (CFRM). Kaedah-kaedah yang digunakan untuk menganggarkan parameter dan selang keyakinan adalah sama untuk kedua-dua model. Kami menggunakan polyroot dan ms untuk menganggarkan parameter-parameter menumpu, membulat dan kecerunan. Penganggaran selang keyakinan menggunakan teknik berdasarkan taburan normal di mana parameter dianggarkan menggunakan polyroot atau ms dan anggaran menggunakan teknik yang bootstrap. Ketiga, kami mempertimbangkan masalah titik terpencil dalam CFRM menggunakan dua statistik yang perbezaan. Statistik-statistik ini dikenali sebagai Perbezaan Ralat Bulatan Purata model Fungsian kosinus (FDMCEc) dan sinus (FDMCEs). Kaedah ini juga menggunakan kaedah penyingkiran baris untuk mengira nilai-nilai FDMCEc dan FDMCEs bagi set data penuh dan yang telah disingkirkan. Cerapan terpencil dikesan melalui perbezaan jarak terbesar antara set data penuh dan set data yang telah disingkirkan. Satu kajian simulasi dijalankan untuk mendapatkan nilai genting bagi beberapa nilai- nilai saiz sampel dan parameter menumpu. Akhir sekali, satu model hubungan berfungsi bagi pemboleh ubah bulatan serentak (CSFRM) yang merupakan lanjutan CFRM dibangunkan. Penganggar kebolehjadian maksimum bagi parameter dan penganggaran varians-kovarians bagi parameter dalam model serentak adalah berdasarkan Matriks Fisher Bermaklumat. Kemudian, hasil kajian dengan menggunakan simulasi menunjukkan bahawa parameter yang dianggarkan mempunyai bias yang kecil. Kaedah yang digunakan untuk menguji prestasi model bagi pemboleh ubah bulat dan pemboleh ubah bukan bulat adalah berbeza. Kesemua penilaian prestasi yang diuji di dalam kajian ini seperti kaedah-kaedah yang digunakan dan model yang dicadangkan dalam kajian ini diuji menggunakan kajian simulasi. Set data pergerakan arah angin dan ombak yang digunakan adalah sebagai ilustrasi kepada uji kaji menggunakan set data sebenar.

ACKNOWLEDGEMENT

I am grateful to Allah the most merciful. I also would like to express my sincere gratitude to my supervisors, Prof. Dr. Abdul Ghapor Hussin and Assoc. Prof. Dr. Yong Zulina Zubairi for their guidance, continuous encouragement and constant support in making this Doctor of Philosophy (Statistics) study possible. I greatly appreciate their guidance from the initial to the final level that enabled me to develop an understanding of this research thoroughly. Without their advice and assistance, it would be a lot tougher to completion.

Secondly, I acknowledge my sincere indebtedness and gratitude to my beloved parents Mohamad Anuar Harithuddin and Azizah Shaari, my mother-in-law Munariza Abd Halim, my beloved wife Nur Hidayu Ab Rahman and my beloved children Muhammad Daniyal Hakimi and Dhia Nur Qaisara for their love, dream and sacrifice throughout my life. I am really thankful for their support, patience, and understanding that makes this work possible.

Thirdly, I would like to express very special thanks to my seniors Zanariah and Iqbal for guiding me until this research has been successfully completed. Last but not least, I want to thank some of my friends especially Khairany for helping me during this study and always supporting me till the end of the study period.

Last but not least, I am grateful to UPNM for granting me the study leave and KPT for the financial assistant thus enabling me to pursue my PhD studies.

APPROVAL

I certify that an Examination Committee has met on **18 January 2018** to conduct the final examination of Mohd Syazwan Bin Mohamad Anuar on his degree thesis entitled — On Parameter Estimation, Confidence Intervals and Outlier Detection for Some Circular Models. The committee recommends that the student be awarded Doctor of Philosophy (Statistics). Members of the Examination Committee were as follows.

Muhd Zu Azhan Yahya, PhD
Professor
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

Wan Rozita Wan Din, PhD
Associate Professor
Center for Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

Azami Zaharim , PhD
Professor
Faculty of Engineering and Built Environment
University Kebangsaan Malaysia
(External Examiner)

Muhammad Qaiser Shahbaz, PhD
Professor
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia
(External Examiner)

APPROVAL

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of Doctor of Philosophy (Statistics). The members of the Supervisory Committee were as follows.

Abdul Ghapor Hussin, PhD
Professor
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Assoc. Prof. Yong Zulina Zubairi, PhD
Associate Professor
Center for Foundation Studies in Science
Universiti Malaya
(Co-Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Author's full name : Mohd Syazwan Bin Mohamad Anuar

Date of birth : 23 April 1986

Title : On Parameter Estimation, Confidence Intervals and Outlier Detection for Some Circular Models.

Academic session : 2013/2014

I declare this thesis is classified as:

☐ CONFIDENTIAL (Contains confidential information under the Official Secret Act 1972)*

☐ RESTRICTED (Contains restricted information as specified by the organization where research was done)*

☐ OPEN ACCESS I agree that my thesis to be published as onlines open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

SIGNATURE SIGNATURE OF MAIN SUPERVISOR

I/C NUMBER NAME OF MAIN SUPERVISOR

Date: Date:

Note: *If the thesis is CONFIDENTIAL or RESTRICTED, please attach the letter from the organisation stating the period and reasons for confidentiality and restriction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

κ	Concentration parameter
$\hat{\kappa}$	Estimate value of concentration parameter
$\mathbf{x}$	A unit vector in the plane
θ	An angle on the circle (A circular data)
z	A unit complex number on a circle
R	Resultant length
$\bar{R}$	Sample Mean resultant length
$\bar{\theta}$	Sample Mean direction
μ	Mean direction
$\hat{\mu}$	Estimate value of Mean direction
ρ	Mean resultant length (precision parameter)
$A(\kappa)$	Mean resultant length for von Mises distribution
n	Sample sizes
$\tilde{\theta}$	Sample Median direction
V	Sample circular variance
υ	Sample circular standard deviation
$\hat{\delta}$	Sample circular dispersion

$d_0(\alpha)$	Circular mean deviation
$\bar{D}_0$	Circular mean difference
B	Bootstrap sample sizes
y	Dependent (response) variable
x	Independent (predictor) variable
e	Angular error
$\hat{e}$	Fitted error
ε	Circular random error for dependent variable
δ	Circular random error for independent variable
λ	Ratio of error concentration parameters in a functional relationship model
d	Circular distance
d_{ij}	Distance between observation i and j
$\bar{\hat{\kappa}}$	Means of concentration parameter
$I_p(\kappa)$	Modified Bessel Function of the First Kind and Order p
$I_0(\kappa)$	Modified Bessel Function of the First Kind and Order zero
v	Dependent (response) variable for Down and Mardia model
u	Independent (predictor) variable for Down and Mardia model