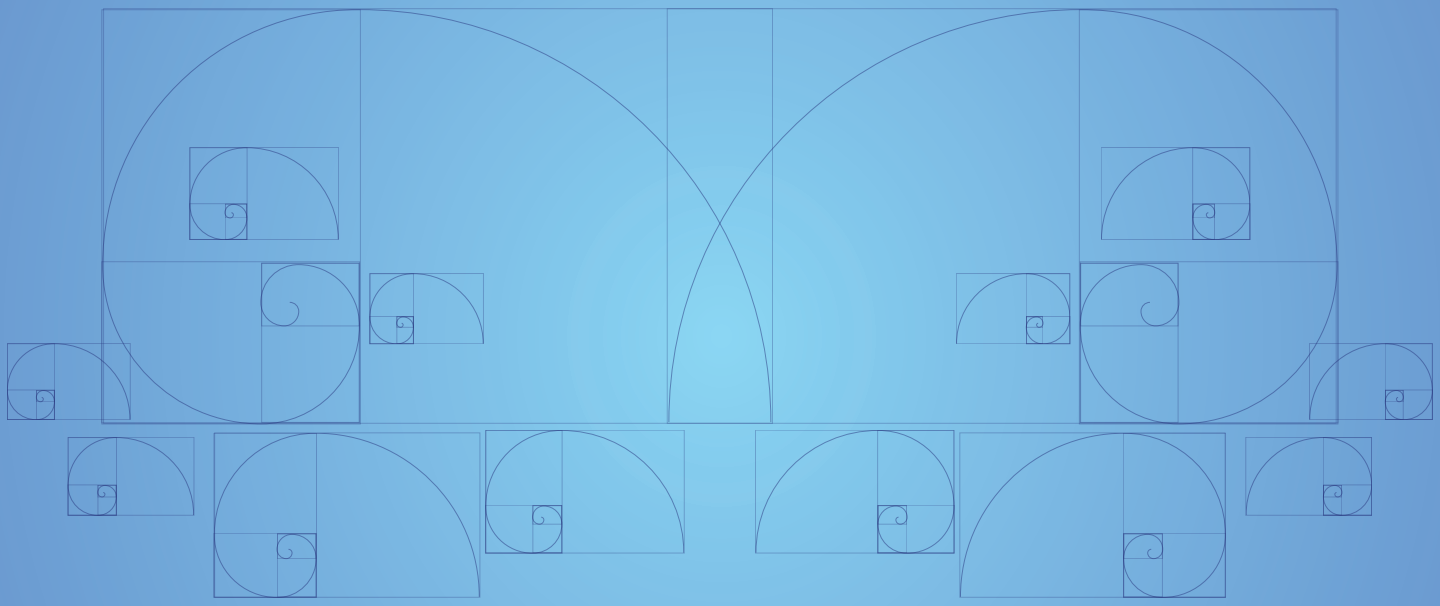
CONFERENCE PROCEEDINGS

INTERNATIONAL CONFERENCE ON MATHEMATICS, STATISTICS, AND FINANCIAL MATHEMATICS 2014 WITH IASC-ARS SESSIONS



ICMSFM 2014

18-19 November 2014, MALAYSIA

Organised by :

UTAR
UNIVERSITI TUNKU ABDUL RAHMAN

IASC-ARS
The Asian Regional Section Of
The International Association For Statistics Computing

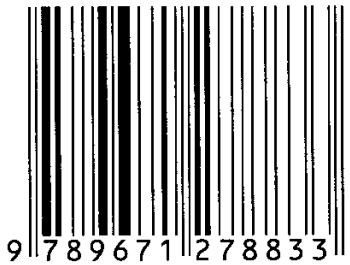
Conference Proceedings
International Conference on Mathematics, Statistics, and Financial Mathematics 2014 with IASC-ARS Sessions

---

Edited by:

Goh, Yong Kheng (Universiti Tunku Abdul Rahman)
Koh, Siew Khew (Universiti Tunku Abdul Rahman)

ISBN 978-967-12788-3-3



Published by:
Universiti Tunku Abdul Rahman
9, Jalan Bersatu 13/4,
46200 Petaling Jaya,
Selangor Darul Ehsan,
Malaysia
Tel: +603 – 4107 9802

ISBN 978-967-12788-3-3

Perpustakaan Negara Malaysia Data Pengkatalogan-dalam-Penerbitan

Printed by: Traxmedia Sdn. Bhd.

Contents

Biodata of Keynote Speakers

Professor HIGUCHI, Tomoyuki

Prof. Higuchi, Tomoyuki is currently the Director-General of the Institute of Statistical Mathematics, Japan. He obtained his B.E. (1984), M.E. (1986), and D.E. (1989) from the Department of Geophysics, University of Tokyo. He joined the Institute of Statistical Mathematics since 1989, and is appointed as a Professor in 2002, Director of Prediction and Knowledge Discovery Research Center in 2003, Director of Research and Development Center for Data Assimilation in 2011. His research interest is in the areas of spatial and time series modeling. He is also executive editor of Annals of the Institute of Statistical Mathematics (AISM) and an ordinary member of the Board of Directors of the Asian Regional Section of the International Association for Statistical Computing (IASC-ARS).

Professor NG, Michael Kwok-Po

Michael Ng is a Professor in the Department of Mathematics and Professor (Affiliate) of Department of Computer Science at the Hong Kong Baptist University. He obtained his B.Sc. degree in 1990 and M.Phil. degree in 1992 at the University of Hong Kong, and Ph.D. degree in 1995 at Chinese University of Hong Kong. He was a Research Fellow of Computer Sciences Laboratory at Australian National University (1995-1997), and an Assistant/Associate Professor (1997-2005) of the University of Hong Kong before joining Hong Kong Baptist University.

Professor YANG, Hailiang

Prof. Yang, Hailiang is from the Department of Statistics and Actuarial Science in the University of Hong Kong where he received his undergraduate and postgraduate degrees from Inner Mongolia University in 1982 and University of Alberta in 1993. He is an Associate of the Society of Actuaries. His research areas include insurance risk models, ruin theory, optimal dividends strategies, option pricing under regime switching models, optimal asset allocation and equity linked insurance products. Prof. Yang, Hailiang has published numerous papers related to his field.

Keynote Lecture: Stream Computing and Emulation in a World of the Edge Heavy Data

Professor HIGUCHI, Tomoyuki

The Institute of Statistical Mathematics
higuchi@ism.ac.jp

Abstract. Improvements to the temporal and spatial resolution of video imagery have led to data being generated at a further high rate. The rate of data generation continues to grow in places where information systems interfere with the real world. This phenomenon is also called the edge-heavy data problem. Stream computing is a computing technology that applies an information processing to massive sensor data on the spot, instead of transferring raw data to the cloud in its entirety. In this talk, we introduce a basic concept of stream computing, and explain computational algorithms for realizing it. These techniques essentially rely on sequential Bayesian filtering method. Its effectiveness is highly appreciated in a research community of data assimilation that is a fundamental tool to make a bridge between computational simulation and large-scale observation. Several examples of data assimilation which are carried out by our research group are demonstrated. A simple way of reducing computational burden necessary for data assimilation, which is called emulation, is addressed.

Keynote Lecture: Eigenvalues of Non-Negative Tensors: Algorithms and Applications

Professor NG, Michael Kwok-Po

Department of Mathematics, Hong Kong Baptist University
mng@math.hkbu.edu.hk

Abstract. In this talk, we discuss recent research development on the eigenvalues of non-negative tensors. We also discuss and develop some algorithms for applications in data mining, community discovery and information retrieval. Experimental results are reported to illustrate the usefulness of these algorithms.

Keynote Lecture: Valuing Equity-linked Death Benefits and Other Contingent Options

Professor YANG, Hailiang

Department of Statistics and Actuarial Science, The University of Hong Kong
hlyang@hku.hk

Abstract. Motivated by the Guaranteed Minimum Death Benefits in various deferred annuities, we investigate the calculation of the expected discounted value of a payment at the time of death. The payment depends on the price of a stock at that time and possibly also on the history of the stock price. If the payment turns out to be the payoff of an option, we call the contract for the payment a (life) contingent option. Because each time-until-death distribution can be approximated by a combination of exponential distributions, the analysis is made for the case where the time until death is exponentially distributed, i.e., under the assumption of a constant force of mortality. The time-until-death random variable is assumed to be independent of the stock price process which is a geometric Brownian motion or jump-diffusion. A substantial series of closed-form formulas is obtained, for the contingent call and put options, for lookback options, and for barrier options. (This talk is based on joint papers with Hans U. Gerber and Elias S. W. Shiu).

A Comparative Study of Outlier Detection Methods in Multiple Linear Regression Models

Yung-Seop Lee[1], Hee-Kyung Kim[2], and Chun Gun Park[3]

[1] Department of statistics, Dongguk University-Seoul, Seoul, Korea,
yung@dongguk.edu,
[2] Department of statistics, Dongguk University-Seoul, Seoul, Korea,
[3] Department of mathematics, Kyonggi University, Suwon, Korea.

Abstract. Outliers cause serious problems in statistical analyses such as estimation, inference, and model selection (Weisberg, 1985). In multiple linear regression model many outlier detection methods have been developed to overcome these problems. Some of them are focus on directly detecting outliers without estimating mean function (Hadi and Simono, 1993; Penna and Yohai, 1995). On the other hand, some other approaches are indirect methods which use residuals from a robust regression estimate to identify the outliers (Rousseeuw, 1984; Rousseeuw and Leroy, 1987; Simpson et al, 1992). As indirect methods, Least Median of Squares (LMS) (Rousseeum, 1984), Least Trimmed Squares (LTS) (Rousseeuw and Leroy, 1987), S estimators (Rousseeuw and Yohai, 1984), MM estimator (Yohai, 1987), GM estimator (Simpson et al, 1992), and S1S estimator (Coakley and Hettmansperger, 1993) were proposed. She et al (2011) proposed outlier detection methods using nonconvex Penalized regression and extended the technique to high-dimensional data. These existing indirect methods require the mean function estimated and then detect outliers. It is known that such raw residuals can fail to detect outliers at leverage points. Our main goal of this article is to include a newer outlier detection method (Park and Kim, 2014) to the comparative analysis done in earlier studies and study the performance of these methods under some situations which are designed with leverage points and outliers. This article is organized as follows. In Section 2, we briefly describe a newer outlier detection method which is developed with leave-one-out and a first difference based error variance estimator. Section 3 contains conclusions.

1 Introduction

Outliers cause serious problems in statistical analyses such as estimation, inference, and model selection (Weisberg, 1985). In multiple linear regression model many outlier detection methods have been developed to overcome these problems. Some of them are focus on directly detecting outliers without estimating mean function (Hadi and Simono, 1993; Penna and Yohai, 1995). On the other hand, some other approaches are indirect methods which use residuals from a robust regression estimate to identify the outliers (Rousseeuw, 1984; Rousseeuw and Leroy, 1987; Simpson et al, 1992). As indirect methods, Least Median of Squares (LMS) (Rousseeum, 1984), Least Trimmed Squares (LTS) (Rousseeuw and Leroy, 1987), S estimators (Rousseeuw and Yohai, 1984), MM estimator (Yohai, 1987), GM estimator (Simpson et al, 1992), and S1S estimator (Coakley and Hettmansperger, 1993) were proposed. She et al (2011) proposed outlier detection methods using nonconvex Penalized regression and extended the

technique to high-dimensional data. These existing indirect methods require the mean function estimated and then detect outliers. It is known that such raw residuals can fail to detect outliers at leverage points. Our main goal of this article is to include a newer outlier detection method (Park and Kim, 2014) to the comparative analysis done in earlier studies and study the performance of these methods under some situations which are designed with leverage points and outliers. This article is organized as follows. In Section 2, we briefly describe a newer outlier detection method which is developed with leave-one-out and a first difference based error variance estimator. Section 3 contains conclusions.

2 Methods

2.1 Multiple regression model

Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ be a response variable which is a $n \times 1$ vector, $\mathbf{x}_j = (x_{j1}, \ldots, x_{jn})^T$, $j = 1, \ldots, p$ be a explanatory variable, $X = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p)$ be a $n \times (p+1)$ matrix with $\text{rank}(X) = p+1$ and β be a regression coefficient parameter, $(p+1) \times 1$ column vector. Define $\mathbf{z} = X\beta$ and $\mathbf{z} = (z_1, \ldots, z_n)^T$. Without loss of generality, we assume that the elements of $\mathbf{z}$ are ordered as $z_1 < z_2 < \cdots < z_n$. Then multiple linear regression model without outliers is written as

$$\begin{aligned} \mathbf{y} &= X\beta + \sigma\epsilon; \\ &= \mathbf{z} + \sigma\epsilon, \end{aligned} \tag{1}$$

where β is the unknown regression coefficient parameter, σ^2 is the unknown variance parameter, and $\epsilon \sim N(0, I_n)$. Unknown parameter β can be estimated using ordinary least squares (OLS) and then $\hat{\mathbf{y}} = X\hat{\beta}$ can be obtained. Let $\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$ be a response variable with outliers. Define a column vector, $\mathbf{d}$, consists of d_i which are either zero or nonzero constant values, that is, $d_i = 0$ represents the ith observation is not outlier and $d_i \neq 0$ means the ith observation is outlier. We further define δ_i as follows; We then write the multiple regression model with outliers as following:

$$\begin{aligned} \mathbf{w} &= \mathbf{y} + \mathbf{d} \\ &= X\beta + \delta, \end{aligned} \tag{2}$$

where $\mathbf{d} = (d_1, \ldots, d_n)^T$ and $\delta = \mathbf{d} + \sigma\epsilon$. Unknown parameter β can be estimated using OLS and then $\hat{\mathbf{w}} = X\hat{\beta}$ can be obtained. Unknown variance parameter σ^2 can be estimated using the difference-based error variance estimator,

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (w_i - w_{i-1})^2.$$

The mean, variance, and mean squared error of the difference-based error variance estimator under (2), can be easily derived. See Park et al (2012) for more detail. In special cases where the errors are following a normal distribution with zero mean and variance σ^2, $\gamma_3 = 0$ and $\gamma_4 = 3$. Hence variance becomes $Var(\hat{\sigma}_R^2) = \{4\sigma^2 \mathbf{s}^T A^2 \mathbf{s} + 2\sigma^4 tr(A^2)\}/tr(A)^2$.

2.2 Leave-one-out difference-based error variance estimator

First we obtain the mean of leave-one-out difference-based error variance estimator when there are outliers. Consider the model (2),

$$\begin{aligned} \mathbf{w} &= X\beta + \delta \\ &= \mathbf{z} + \delta \end{aligned}$$

and assume $z_1 < z_2 < \ldots < z_n$ is ordered. Then the mean of leave-one-out difference-based error variance estimator, $\hat{\sigma}^2_{R(i)}$, obtained by deleting ith observation is

$$\begin{aligned} E(\hat{\sigma}^2_{R(i)}) = &\frac{1}{2(n-2)} \sum_{j=2}^{n} (z_j - z_{j-1})^2 + \frac{1}{n-2}(z_{i+1} - z_i)(z_i - z_{i-1}) \\ &+ \frac{1}{2(n-2)} \sum_{j=2}^{n} (d_j - d_{j-1})^2 + \frac{1}{(n-2)}(d_{i+1} - d_i)(d_i - d_{i-1}) \\ &+ \frac{1}{(n-2)} \sum_{j=2}^{n} (z_j - z_{j-1})(d_j - d_{j-1}) + \frac{1}{(n-2)}(z_{i+1} - z_i)(d_i - d_{i-1}) \\ &+ \frac{1}{(n-2)}(d_{i+1} - d_i)(z_i - z_{i-1}) + \sigma^2. \end{aligned}$$

Since the mean of leave-one-out difference-based error variance estimator depends on location where the outlier is located in consecutive order, we further derive the mean of leave-one-out difference-based error variance estimator under the following four cases: Case 1 is no outlier in three consecutive orders; Case 2 is one outlier in left side of the three consecutive orders; Case 3 is one outlier in the middle of the three consecutive orders; Case 4 is one outlier in the right of the three consecutive orders. These results are also summarized in Theorem 1.

Theorem 1

Assume the Model 2. The mean of leave-one-out difference-based error variance estimator is obtained under the following four cases:

- Case 1: no outlier in three consecutive order, $(i-1, i, i+1)$, $i = 2, \ldots, n-1$

$$E(\hat{\sigma}^2_{R(i)}) = \frac{1}{2(n-2)} \sum_{j=2}^{n} (z_j - z_{j-1})^2 + \frac{1}{2(n-2)} \sum_{j=2}^{n} (d_j - d_{j-1})^2$$

$$+\frac{1}{(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1}) + \sigma^2 + \frac{1}{(n-2)}(z_{i+1} - z_i)(z_i - z_{i-1});$$

– Case 2: one outlier in left side of the three consecutive orders, that is, one outlier at $(i-1)$th order in three consecutive order, $(i-1, i, i+1)$, $i = 2, \ldots, n-1$

$$\begin{aligned} E(\hat{\sigma}^2_{R(i)}) = & \frac{1}{2(n-2)}\sum_{j=2}^{n}(z_j - z_{j-1})^2 + \frac{1}{2(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1})^2 \\ & + \frac{1}{(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1}) + \sigma^2 + \frac{1}{(n-2)}(z_{i+1} - z_i)(z_i - z_{i-1}) \\ & - \frac{d_{i-1}}{(n-2)}(z_{i+1} - z_i); \end{aligned}$$

– Case 3: one outlier in the middle of the three consecutive order, that is, one outlier at (i)th order, $(i-1, i, i+1)$, $i = 2, \ldots, n-1$

$$\begin{aligned} E(\hat{\sigma}^2_{R(i)}) = & \frac{1}{2(n-2)}\sum_{j=2}^{n}(z_j - z_{j-1})^2 + \frac{1}{2(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1})^2 \\ & + \frac{1}{(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1}) + \sigma^2 + \frac{1}{(n-2)}(z_{i+1} - z_i)(z_i - z_{i-1}) \\ & + \frac{d_i}{(n-2)}\{(z_{i+1} - z_i) - (z_i - z_{i-1})\} - \frac{d_i^2}{(n-2)}; \end{aligned}$$

– Case 4: one outlier in the right of the three consecutive orders, that is, one outlier at $(i+1)$th order, $(i-1, i, i+1)$, $i = 2, \ldots, n-1$

$$\begin{aligned} E(\hat{\sigma}^2_{R(i)}) = & \frac{1}{2(n-2)}\sum_{j=2}^{n}(z_j - z_{j-1})^2 + \frac{1}{2(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1})^2 \\ & + \frac{1}{(n-2)}\sum_{j=2}^{n}(d_j - d_{j-1}) + \sigma^2 + \frac{1}{(n-2)}(z_{i+1} - z_i)(z_i - z_{i-1}) \\ & + \frac{d_{i+1}}{(n-2)}(z_i - z_{i-1}). \end{aligned}$$

2.3 Procedure of difference-based outlier detection

Our difference-based outlier detection approach consists of four parts; Part I is to obtain order of $\mathbf{w}$ which is summarized in Step 1-Step 2; Part II is to select clean data using both leave-one-out and leave-half-out difference-based error variance estimators, which is summarized in Step 3; and Part III is to detect outliers using clean data which is explained in Step 4. The procedure of our approach is described in the following:

– Step 1: Input data and fit the model

$$w_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \sigma\epsilon_i + d_i;$$

– Step 2: Determine the order of $\mathbf{w}$'s using $X\hat{\beta}$ and rearrange by the order of $\hat{\mathbf{w}}$, that is

$$w_{(1)} < \ldots < w_{(n)};$$

– Step 3: Select clean data (70%) using leave-one-out and first difference based error estimator
– Step 4: Detect outlier using 70% clean data.

3 Conclusion

In this paper, we propose a difference-based outlier detect approach which does not require to estimate mean model in multiple regression. Our approach is developed using a difference-based error variance estimator. Our approach is the leave-one-out type estimator based on difference-based error variance. The existing outlier detect approach using leave-one-out approach are highly affected by other outliers, while ours does not because our approach does not use the regression coefficient estimator. To the best of our knowledge, there is no such method for outlier detection purpose. We compare our approach with several existing methods using simulation study, suggesting outperformance of our approach under some situation such as leverage points and outliers.

References

1. Coakely, C. W., Hettmansperger, T. P.: A bounded influence, high breakdown, efficient regression estimator. Journal of the American Statistical Association. 88, 872–880 (1993)
2. Hadi, A. S., Simonoff, J. S.: Procedures for the Identification of Multiple Outliers in Linear Models. Journal of the American Statistical Association. 88, 1264-1272 (1993)
3. Park, C., Kim, I., Lee, Y.: Error variance estimation in nonparametric regression under lipschitz condition and small sample size. Journla of Statistical Planning and Inference. 142, 2369-2385 (2012)
4. Park, C., Kim, I.: Outlier Detection Using Difference Based Variance Estimators in Multiple Regression. manuscript. (2014)
5. Peña, D. and Yahai, V. J.: The detection of Influential subsets in linear regression by using an inflence matrix. Journal of the Royal Statistical Society: Series B. 57, 145-156, (1995)
6. Rousseeuw, P. J.:Least median of squares regression. Journal of the American Statistical Association. 79, 871-880, (1984)
7. Rousseeuw, P. J., Leroy, A. M.: Robust regression and outlier detection, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, New York: John Wiley & Sons Inc. (1987)
8. Rousseeuw, P., Yohai, V.: Robust regression by means of S-estimators, in Robust and nonlinear time series analysis (Heidelberg, 1983), Vol. 26 of Lecture Notes in Statist., New York: Springer, 256-272, (1984)

9. Simpson, D. G., Ruppert, D., Carroll, R., J.: On one-step GM estimates and stability of inferences in linear regression. Journal of the American Statistical Association. 87, 439-450, (1992)
10. She Y., Owen A. B.: Outlier detection using nonconvex penalized regression. Journal of the American Statistical Association. 106, 626-639, (2012)
11. Weisberg, S.: Applied Linear Regression 2nd ed., New York: Wiley, (1985)
12. Yohai, V. J.: High breakdown-point and high efficiency robust estimates for regression. Annals of Statistics. 15(2), 642-656, (1987)

Fuzzy GCO Based Web Page Retrieval

Gomathi.C[1], Dr.V.Rajamani[2]
[1] Department of information technology, AnnaUniversity, Barathidhasan Institute of Technology, Tiruchirappalli,Tamilnadu,India,
gomathi.m.e©gmail.com,
[2] Department of Electronics and communication, Veltech Multitech,Chennai, Tamilnadu,India
rajavmani@gmail.com

Abstract. In modern era - online user face the several problems like unwanted data over the unwanted web pages, security issues, retrieval time consumption, unwanted advertisement, etc., To overcome the web page retrieval problem, the novel approach 'Fuzzy GCO Based Web Page Retrieval' has been proposed to retrieve the web pages through the fuzzy value of relevance semantic relationship between user queries, keywords- web page weight calculation and the distance value between web pages. Leveraging of Semantic web technology, the fuzzy values are calculated through Geometric Interpretation (G), Cross web page (C), Optimal Properties of Proposed Distance (O). Finally, the user satisfactory value has been calculated to rank the web pages through semantic fuzzy relationship value.

Keywords: Geometric Interpretation (G), Cross web page (C), Optimal Properties of Proposed Distance (O), fuzzy relationship values.

1. Introduction

The user can retrieve the information from the web browser through the user query. Some web pages delivered to the user based on content and browser count. The browser count and content have been calculated by view time of the web page. The most hit web content opened as first link in the web browser. By applying this method, the newly uploaded useful content cannot open within first few links. The users spend more time to get the relevant document. To conquer this issue, an Annotation based retrieval system was introduced; it satisfies the user query with partial manner. But it retains the same problem. To overcome this problem, the fuzzy relationship values have been calculated for retrieved web pages and arrange it in decreasing order.

Most of the researcher proposed the page ranking algorithm, retrieval algorithm, personalization algorithm and analyzing tool, for an example Bettina Berendt, Myra Spiliopoulou [1]survived the web Usage Miner (WUM) tool specifications, to analyze the web pages. WUM Tool is used to retrieve the web pages through the web browser from different search engines. To justify that advanced web pages have been developed based on the Form based Web pages. The web pages retrieved based on the page ranking scheme. The rank allocated to the web pages based on the number of times the user visited.

Analyze the cloning method in retrieval of web pages. The web pages opened in web browser with dynamic clustering approach through state transition probabilities [2].

Page ranking system allocates the value to all the web pages. The large data have stored in the web database and indexing allocates for all data. The text document characterization is stored in the web database [3]. The different types of Markov models for prediction process in the web search can understand the frequency has been calculated for the web pages through this approach. The user keyword has been predicted based on the user query [4].

The individual user accesses the web pages as web personalization. The function of the new concept Conceptual Log (C-log), then it is interacting with web server logs. The concept based semantic data has been retrieved from the web page [5]. The web personalized information by applying the Markov model in calculating with maximum entropy value. Some general drawbacks are occurred to capture the web personalization in the web page [6].

The Web pages linked with some other web pages through the hub concept. The user query is classified in to different categories. The solution of the user query is different one for the proposed method. The web pages are categorized such that authoritative web pages and the normal web pages [7]. The different types of models present the documents with probabilistic approach. Markov chain stochastic process is used to calculate the probability value for all documents [8] [9].

The web pages are implemented automatically. The different types of user queries are classified the web page in differently [10]. Basically two algorithms are used in the block level link analysis algorithm such that hit algorithm and the Block Level hit algorithm. These algorithms make the hyperlinks and convert the web pages into blocks.

Based on the survey, on-line user require the optimal solution, this paper proposed the novel methodology to retrieve the web page, to acquire appropriate results the fuzzy relationship value has been calculated. To find fuzzy relationship values, the web page Geometric Interpretation weight, cross web page, optimal properties of proposed distance has been calculated based on the user query. Finally add these three values to get an optimal relational value of the web page, which one have more relational value that will open as first web page to the user, before that apply the stop word removal algorithm on user query to remove the unnecessary words.

2. Proposed system

Fig.1 Fuzzy GCO Based Web Page Retrieval System Architecture Model notifies that the user post the query in search engine. The search engines gather

the related results from the web database then it sends the web pages as results to the user with decreasing ordered rank.

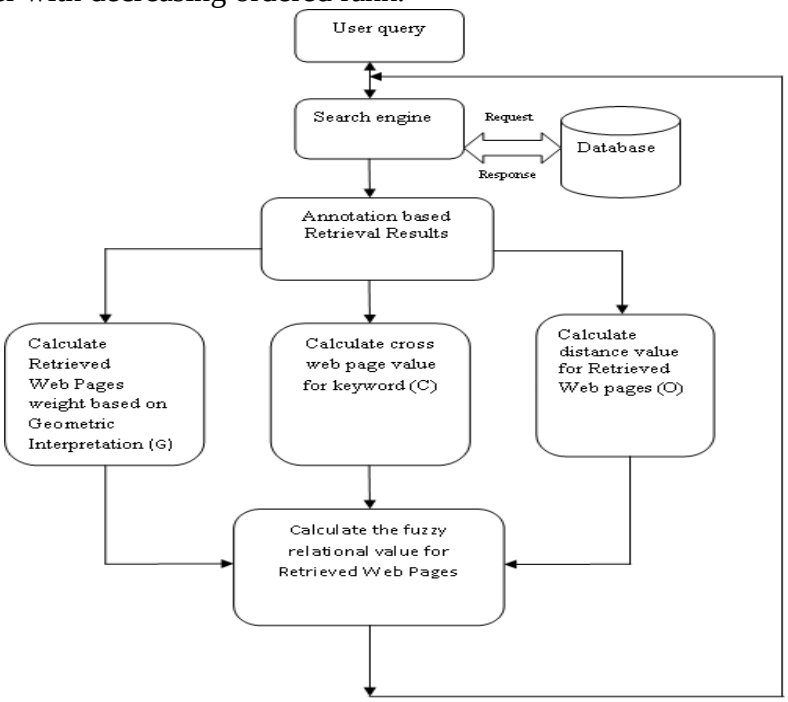


Fig. 1 Fuzzy GCO Based Web Page Retrieval System

An objective of this work is, to modify the search engine, which will provide the relevant document to the user request with more accuracy, it should also support the prediction mechanism with captured the user's requesting key words. To modify the working principle of search engine the following methods are adapted.

2.1 Probability Value Calculation for Web pages

The programmer develops the search engine which is based on the page ranking system, cluster based Markov models. The user calculates the probability value for all web pages using the Markov Semantic Indexing (MSI). At first, the user retrieves the web pages through the annotation based retrieval and makes the relationship with user keywords and the annotation based web pages. Calculate the relationship value between the user keywords and the web pages and calculate the distance value between the web pages. These two values are applied to the Markov chain property and make the calculation of finding the probability value for all web pages. Finally sort the probability value with descending order and allocates the rank with ascending order, that

web pages have more probability value. Further, based on the rank system, the web pages will be retrieved on the web browser. This calculation operation should be done in the design of the web search engine.

2.2 Annotation based retrieval

Annotation based retrieval specifies that automatically web pages retrieved based on the user typed keywords. The search engine analyze the user keyword and it compare with the stored documents in the web data base. Here, the web pages specify that web documents. The web documents has been retrieved dynamically from the web data base, based on the user typed query. The search engine checks the query with the data base using keywords. The keywords have been generated in the user typed query. Annotation based retrieval does not contain any algorithm or formula to retrieve the web documents itself. By having the keywords only, the search engine search the document in the entire data base and retrieve the results to the user.

2.2.1Geometric Interpretation (G)

Geometric interpretation specifies the web page weight value to retrieve the web documents which is based on the user query. A single user query may have more than one keyword in itself. The keyword considered as the component. A single user query may have multiple components. The search engines calculate the web page weight value separately to each component and apply the formula to calculate the web pages

$$\mathrm{Wp} = \frac{\mathrm{Wr}}{\mathrm{Wt}} \qquad (1)$$

In equation (1), W_P specifies web page weight value; W_r specifies relevant web pages based on the user query. W_t specifies total number of web pages on the particular search engine. The above equation may suit if the user query have single component. If the user queries have two or more components, the above equation does not suited directly.

$$\mathrm{P}(\mathrm{A},\mathrm{B}) = \frac{\mathrm{P}(\mathrm{A}\cap\mathrm{B})}{\mathrm{P}(\mathrm{A}\cup\mathrm{B})} \qquad (2)$$

Equation (2) describes, if the user query have two components itself then apply the equation (2) to find web page weight value between the components. In Equation (2), A and B are two different components $\mathrm{P}(\mathrm{A}\cap\mathrm{B})$, It specifies that intermediary relationship between the two components. $\mathrm{P}(\mathrm{A}\cup\mathrm{B})$,It

specifies that entire relationship between the two components. To find the $P(A \cup B)$ value, the search engine can use the formula

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (3)$$

Apply the equation (3) in equation (2), and then write the equation

$$P(A, B) = \frac{P(A \cap B)}{P(A) + P(B) - P(A \cap B)} \qquad (4)$$

Equation (4) used to find the web page weight value between the two components then may use the same formula to find web page weight value between multiple components.

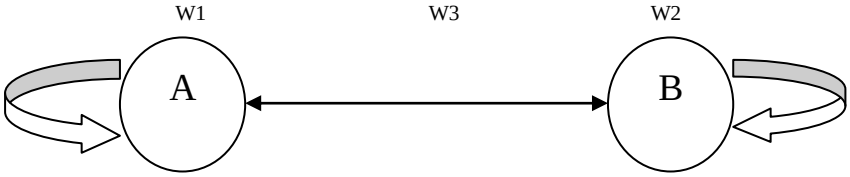


Fig.2, Simple Diagram for Calculating Geometric Interpretation

In Fig.2, The users have two components A & B. Two components specify two different keywords. The two different key words are occurred in the single statement as user typed query. The Search engine should calculate the web page weight value for these two different queries. At first, the search engines find the web page weight value W1 for A component and find the web page weight value W2 for B component in separately. Next, find the web page weight value W3 for between the two components. Finally the search engine finds the entire web page weight value for both components by applying the Equation (4).

The user has proved the results in Geometric Interpretation and the optimality properties of proposed distance. For example, if the user enter the query “How to cook briyani” in search engine and retrieve some results through the annotation based retrieval. At first, apply the stop word removal techniques to remove unwanted character. In the query, `to’ is the unwanted character and remove the character from this word.

Based on the user query, the user has three different keywords such that how, cook briyani. These three keywords consider as the three different components. First, the search engine calculates the web page weight value separately for each component. Then, it combined the keywords like How-

briyani, cook-briyani, how-cook to find the web page weight value. Finally, the search engine calculates the web page weight value for the full user query.

Table1. Web page weight value for retrieved content

S.No.	Retrieved Content	lower bound of W_p Value	W_p Value	Upper bound of W_p Value
1.	How to cook biryani	0.483	0.484	0.485
2.	How to make a chicken biryani	0.301	0.302	0.303
3.	How to cook vegetarian biryani	0.165	0.166	0.167
4.	How to prepare biryani	0.0569	0.0579	0.0589
5.	How to make muslim biryani	0.0469	0.0479	0.0489
6.	How to cook briyani in tamil	0.0211	0.0221	0.0231
7.	How to cook mutton biryani	0.307	0.308	0.309
8.	How to cook briyani rice	0.2769	0.2779	0.2789
9.	Chicken briyani recipe	0.323	0.324	0.325
10.	How to cook veg biryani	0.142	0.143	0.144
11.	How to cook egg biryani	0.124	0.125	0.126
12.	How to cook briyani in hindi	0.024	0.025	0.026
13.	Easy chicken biryani	0.294	0.295	0.296
14.	How to cook basmathi rice for briyani	0.0172	0.0182	0.0192

The above Table specifies the values that web page weight for relevant web document. In this paper the specified values are taken as fuzzy numbers for example, a user query is "How to cook briyani". The user gets 30 relevant documents and calculates the web page weight for 30 documents, for analytic purpose we taken only 14 fuzzy values. The user has assumed the maximum documents in the search engine 1, 00,000 lakhs documents.

Fuzzy numbers are uncertainty numbers so we need to find the crisp values through defuzzification [10] process. Defuzzification represent the average value of the trapezoidal fuzzy numbers, a universally accepted Robust's Ranking Technique is used for GCO.

$$\mu(x) = \begin{cases} \frac{x-0.483}{0.001}, 0.483 \le x \le 0.484 \\ 1, x = 0.484 \\ \frac{0.484-x}{0.001}, 0.484 \le x \le 0.485 \\ 0, otherwise \end{cases}$$

$$R(\tilde{a}_1) = R(0.483, 0.484, 0.485) = \int_0^1 0.5(a_\alpha^l, a_\alpha^u) d\alpha = 0.484$$ Where, (a^l, a^u) and a_i is the level cut of the fuzzy number ã. similarly convert other (W_p,W_d, W_c) fuzzy values in to crisp values, these values are used to find the relationship value for every retrieved content.

2.2.2 Optimal Properties of Proposed Distance (O)

Optimal properties derived from the Markov chain properties. These properties are used to calculate the web page distance value. Optimality properties measure distance value for retrieved web documents. Here, a web page notifies the web documents.

1. *Removal of Stop Word:* Removal of Stop Word removes the unwanted characters from the user query. Unwanted characters like `space', `:', `;', `!', `-', `_' These characters may be occurred in the user query. These characters can be removed by the stop word removal techniques and calculate the distance value after eliminating the unwanted characters.

2. *Distance Calculation:* Measure the distance value for retrieved web documents. The user considered the user query length and the retrieved web document length and used the formula to calculate the distance value

$$Wd = \frac{D(A) + D(B)}{Td} \qquad (5)$$

In Equation (5), W_d notifies retrieved web document's distance total value. D (A) notifies user query total length, D (B) notifies retrieved web content total length. T_d notifies that maximum query length in the search engine. D (B) value will be changed based on the retrieved web document.

Table 2.Web page distance value for retrieved content

Retrieved Content	lower bound of W_d Value	W_d Value	Upper bound of W_d Value
How to cook biryani	0.0008	0.0078	0.0148
How to make a chicken biryani	0.0047	0.0117	0.0187
How to cook vegetarian biryani	0.0057	0.0127	0.0197
How to prepare biryani	0.0022	0.0092	0.0162
How to make muslim biryani	0.0037	0.0107	0.0177
How to cook briyani in tamil	0.0042	0.0112	0.0182
How to cook mutton biryani	0.0037	0.0107	0.0177
How to cook briyani rice	0.0027	0.0097	0.0167
Chicken briyani recipe	0.0027	0.0097	0.0167
How to cook veg biryani	0.0022	0.0092	0.0162
How to cook egg biryani	0.0022	0.0092	0.0162
How to cook briyani in hindi	0.0042	0.0112	0.0182
Easy chicken biryani	0.0017	0.0087	0.0157
How to cook basmathi rice for biryani	0.0081	0.0151	0.0221

Table 2 describes that, the results for web document distance value. The user has given 30 document results for the query “How to cook briyani”. The user removes some unwanted characters and space in the document name to calculate the web page distance. Equation (5) is applied to measure the web page distance.

2.2.3 Cross Web Page Calculation(C)

The user enters the query in the search engine. The search engines have given the results to the user based on the user query. Lot of keywords is available in the user query. The search engine searches the web contents using the keywords. Basically keywords are classified into different category. Some of keywords have same meaning but names. The search engine should consider this problem. Because, the search engine search the content based on keywords. The user have found some solution to overcome this problem and find some words which is having two names in same meaning.

$$Wc = \left(\frac{Ks}{T} + 1\right) - (1 - \mu)(1 - \frac{Kd}{T}) \qquad (6)$$

in equation (6), W_c- specifies the total cross web page weight and K_s- specifies the total number of similar keywords weight, K_d- specifies the total number of dissimilar keywords weight, here dissimilar keywords means another alternate word for same keyword, T- specifies total number of keywords in the search engine and the symbol μ represents mean value of sample.

The similar and dissimilar keywords weight value by applying the equation (7). First, the user should analyze that the keywords have same meaning and different name. The user have identified some keywords have different name in same meaning.

Table 3. Different Keywords Weight Value

Keywords	(K_s /T) Value	Another Word for Keyword	(K_d/T) Value
Buy	0.1250	Purchase	0.4220
Big	0.1350	Large	0.0127
Quickly	0.2230	Speedily	0.0020
Middle	0.4260	Between	0.0085
Look	0.1080	See	0.0247
Mountain	0.2860	Hills	0.1880
Mobile	0.2080	Cell phone	0.0002
Happy	0.7140	Pleasure	0.1160
Rich	0.2260	Spicy	0.0004
At present	0.3370	Current time	0.0318
Glass	0.3310	Mirror	0.0014
Lady	0.3060	Women	0.0072
Silent	0.0886	Pease	0.0018
Pause	0.0771	Break	0.0032
Cook	0.1010	Make	0.5210

Table 3 shows that different keywords and each keyword have one alternate and that keyword weight value. In an above referred table the fourth column notify that dissimilar keywords weight value. The search engine applies these two values in Equation (6) then it will get the cross web page value.

3. Calculating the fuzzy Relationship value

By having the above three values, the search engine may find the relationship value for every retrieved content by using the formula

$$Vr = (\alpha * Wp) + (\beta * Wd) + (\gamma * Wc) \qquad (7)$$

In Equation (7) V_r notify the relationship value for retrieved web content.

4 Conclusion

This work hybrid the features of three different methods like Geometric interpretation, Cross web page calculation, Optimality properties of proposed distance. This Geometric Interpretation (G), Cross Web Page calculation(C), optimal properties of proposed distance (O) used for web page retrieval through the fuzzy relational value of web pages. Analytical results provide the detail of web page weight, cross web page calculation for keywords and web page distance value for optimal retrieval of web page. In future this work has extended to image and on-line video retrievals.

References

1. Bettina Berendt, Myra Spiliopoulou, "Analysis of Navigation Behavior in Web Pages Integrating Multiple Information Systems", Vol.9, Issue.1, The VLDB Journal, 2009, pp.56-75,
2. Jose Borges, Mark Levene, " A Dynamic Clustering Based Markov Model for Web Usage Mining", May 26, Google Scholar, 2004, pp.1-21.
3. Sergey Brin, Lawrence Page, "The Anatomy of a Large Scale Hyper Textual Web Search Engine", ACM, Vol.30, Issue.1-7, 1998, pp.107-117.
4. Mukund Despande, George Karypis, "Selective Markov Models for Predicting Web Page Access", ACM Transaction and internet technology, Vol.4, Issue.2,May 2004, pp.163-184.
5. M. Erinaki, M. Vazirgiannis, I. Varlamis, "Sewep: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process", ACM, ISBN 1-58113-737-0, pp.99-108.
6. Eren Manavoglu, Dmitry Pavlov, C. Lee Giles, "Probabilistic User Behavior Model", IEEE 3rd ICDM, ISBN 0-7695-1978-4, 2003, pp.203-210.
7. Jon M Kleinberg, "Authoritative Sources in Hyper Linked Environment", ACM – SIAM symposium, Vol.46, Issue.5, 1998, pp.604-632.
8. Mark steyvers, Padhraic smyth, Thomas Griffiths, "Probabilistic Author – Topic Models for Information Discovery" 10th ACM & Sig KDD and Data Mining,ISBN 1-58113-888-1, 2004, pp.306-315.
9. Rituparna Sen and Mark H Hansen, " Predicting Web User's Next Access Based on Log Date", Journal of Computational and Graphical Statistics, Volume.12, 2003, pp.101-108.
10. Mike Perkowitz, Oren Etzioni, "Towards Adaptive Web Sites: Conceptual Framework and Case Study", Science Direct, Vol.118, Issue.1-2, pp.245-275.
11. Yager, R.R.,"Multiple Objective Decision-Making Using Fuzzy sets", International journal of Man-Machine Studies 9, pp.375-382, 1977.

Shrinkage-based Modification of Fixed Rank Kriging for Fast Spatial Prediction

ShengLi Tzeng[1] and Hsin-Cheng Huang[2]

[1] Department of Public Health, China Medical University, Taichung 40402, Taiwan
slt.cmu@gmail.com
[2] Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan
hchuang@stat.sinica.edu.tw

Abstract Optimal spatial prediction is challenging when the number of observations is large. A large data set also comes with a non-stationary structure more often than not. Fixed rank kriging (FRK) is a popular method that handles both massive observations and non-stationarity. A family of non-stationary covariance functions is constructed based on a set of basis functions, and it allows for quite a few computational simplifications in spatial prediction. Its number of parameters, however, is too large to obtain stable estimation using least-squares methods. Even worse, a least-squares method leads to overfitting, i.e., it prefers larger-basis-number models which make estimation more unstable. We propose a penalized least-squares estimation, resulting in a simple shrinkage-based modification of FRK by cutting down the eigenvalues of a spatial covariance matrix. Our method can effectively control estimation variability via tuning the shrinkage parameter, and it also alleviates the overfitting problem. Original FRK is then a special setting of our method with a zero shrinkage parameter. Moreover, a fast estimation and prediction method still exists, which avoids taking a high-dimensional matrix inversion. Some numerical examples are given to demonstrate the effectiveness of the proposed method.

1 Introduction

Geostatistical models are becoming increasingly required in many environmental, atmospheric, and geophysical sciences. Optimal spatial prediction is challenging when the number of observations is large. A large data set also comes with a non-stationary structure more often than not. It's of great interest to develop a geostatistical model that is flexible to account for spatial covariances among observations. Consider a sequence of independent spatial processes, $\{y(\mathbf{s},t):\mathbf{s}\in D\}$; $t=1,\ldots,T$, defined on a d-dimensional spatial domain $D\subset\mathbb{R}^d$, where $T>1$. The processes are assumed to have a common spatial covariance function, $C(\mathbf{s},\mathbf{s}^*)=\mathrm{cov}(y(\mathbf{s},t),\,y(\mathbf{s}^*,t))$. Suppose that we observe data $\mathbf{z}_t\equiv(z(\mathbf{s}_1,t),\ldots,z(\mathbf{s}_n,t))'$; $t=1,\ldots,T$, at $\mathbf{s}_1,\ldots,\mathbf{s}_n\in D$ with additive noise $\boldsymbol{\varepsilon}_t$ according to

$$\mathbf{z}_t=\mathbf{y}_t+\boldsymbol{\varepsilon}_t;\quad t=1,\ldots,T, \tag{1}$$

where $\mathbf{y}_t\equiv(y(\mathbf{s}_1,t),\ldots,y(\mathbf{s}_n,t))'$, $\boldsymbol{\varepsilon}_t\sim(\mathbf{0},\sigma^2\mathbf{I}_n)$ is uncorrelated with $\mathbf{y}_t$, and $\boldsymbol{\varepsilon}_t$'s are mutually uncorrelated. We assume that σ^2 is known. The goal is to estimate $C(\cdot,\cdot)$ based on the data $\mathbf{z}_1,\ldots,\mathbf{z}_T$, from which the best linear unbiased prediction

of $\{y(\mathbf{s},t) : \mathbf{s} \in D\}$ for $t = 1,\ldots,T$, can be obtained. Note that neither the stationary assumption is made nor a parametric structure is assumed for $C(\cdot,\cdot)$. Throughout the paper, the mean of $y(\cdot,\cdot)$ is assumed known and, without loss of generality, to be 0.

The model (1) involves an $n \times n$ matrix $\text{var}(\mathbf{y}_t)$ to be estimated, whose inverse should be computed for prediction, and this make the analysis infeasible when n is large. [2] proposed a framework based on spatial random effect models, which handles both massive observations and non-stationarity. A family of non-stationary covariance functions is constructed based on a set of basis functions, and it allows for quite a few computational simplifications in spatial prediction. Kernel convolution is another widely used approach ([7]; [8]; [5]). These methods require no matrix inversion, but their performance depends on locations of knots and the form of the kernel, whose choices are not completely clear (see e.g., [3]).

Although the fixed rank kriging (FRK) approach in [2] can be very fast, its number of parameters is too large to obtain stable estimation using least-squares methods. Even worse, a least-squares method leads to overfitting, i.e., it prefers larger-basis-number models which make estimation more unstable. We propose a penalized least-squares estimation, resulting in a simple shrinkage-based modification of FRK by cutting down the eigenvalues of a spatial covariance matrix. Original FRK is then a special setting of our method with a zero shrinkage parameter.

[6] developed a regularization method mainly for multivariate geostatistics. In this paper, we shall emphasize its use in univariate problems, and get more understanding about when the shrinkage-based estimation works well. The rest of the paper is organized as follows. Section 2 starts by introducing our model and the proposed shrinkage method. A simulation example is demonstrated in Section 3. Finally Section 4 ends with some discussion.

2 Method

2.1 Spatial Random Effect Models

We consider a spatial random effect model for the latent process $y(\cdot,\cdot)$ (e.g.,[2]; [9]; [4]), which relies on a given set of basis functions $f_1(\cdot),\ldots,f_K(\cdot)$. The model assumes that $y(\cdot,\cdot)$ can be well represented as

$$y(\mathbf{s},t) = \mu(t) + \mathbf{w}_t'\mathbf{f}(\mathbf{s}) + \xi(\mathbf{s},t) = \mu(t) + \sum_{k=1}^{K} w_k(t) f_k(\mathbf{s}) + \xi(\mathbf{s},t), \qquad (2)$$

where $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}),\ldots,f_K(\mathbf{s}))'$ consists of $K \leq n$ known basis functions, $f_1(\cdot),\ldots,f_K(\cdot)$ such that $\mathbf{F} \equiv (\mathbf{f}(\mathbf{s}_1),\ldots,\mathbf{f}(\mathbf{s}_n))'$ is an $n \times K$ matrix of rank K, $\mathbf{w}_t = (w_1(t),\ldots,w_K(t))'$;

$t = 1, \ldots, T$, are random effects obeying $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{M})$, and $\xi(\mathbf{s}, t) \sim N(0, v^2)$ is a fine-scale white-noise process. Both $\mathbf{w}_t$ and $\xi(\mathbf{s}, t)$ are assumed to be uncorrelated between t's. For simplicity, $\boldsymbol{\Sigma}$ is set to be $\mathbf{I}_n$, $\mu(t)$ and σ^2 are assumed known and, without loss of generality we set $\mu(t) = 0$ throughout the paper. Given the bases $\{f_k(\cdot) : k = 1, \ldots, K\}$, the parameters that need to be estimated are $\mathbf{M}$ and v^2, where $\mathbf{M}$ is required to be non-negative definite and is denoted by $\mathbf{M} \succeq \mathbf{0}$.

2.2 Parameter Estimation

Let $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_T)'$. Then the sample covariance matrix $\mathbf{S} = \mathbf{Z}\mathbf{Z}'/T$ is an unbiased estimate of $\mathbf{V}$. However, $\mathbf{S}$ is a poor estimate of $\mathbf{V}$ when n is large or T is small, because of high estimation variability. Although the proposed model helps to reduce estimation variability, the unknown parameters in $\mathbf{M}$ still have high complexity unless K is very small. To reduce the estimation variability further while suitably controlling the bias, we consider a shrinkage approach. Specifically, we propose to estimate $\mathbf{M}$ and v^2 by shrinking the eigenvalues of $\mathbf{FMF}'$ in terms of the following objective function:

$$\phi(\mathbf{M}, v^2) = \frac{1}{2} \left\| \mathbf{FMF}' + v^2\mathbf{I}_n + \sigma^2\mathbf{I}_n - \mathbf{S} \right\|_F^2 + \tau \left\| \mathbf{FMF}' \right\|_*, \tag{3}$$

over all $\mathbf{M} \succeq \mathbf{0}$ and $v^2 \geq 0$, where $\|\mathbf{X}\|_F^2 = \mathrm{trace}(\mathbf{X}'\mathbf{X})$ is the squared Frobenius norm of $\mathbf{X}$, and $\|\mathbf{X}\|_* = \mathrm{trace}\big((\mathbf{X}'\mathbf{X})^{1/2}\big)$ is the nuclear norm of $\mathbf{X}$. The penalty $\|\mathbf{FMF}'\|_*$ is simply the trace of $\mathbf{FMF}'$ or the sum of the eigenvalues of $\mathbf{FMF}'$. Therefore, this penalty shrinks eigenvalues and force some small eigenvalues to become zeros exactly. The minimizers of (3) are denoted as $\big(\hat{\mathbf{M}}_\tau, \hat{v}_\tau^2\big)$. Clearly, original FRK is then a special setting of our method with a zero shrinkage parameter.

The following proposition shows that the minimizer of (3) has a simple closed form, whose proof is a direct result from Appendix in [6].

Proposition 1 Let $\mathbf{L} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1/2}$, and $\mathbf{PD}_K\mathbf{P}'$ be the eigen-decomposition of $\mathbf{L}'(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{L}$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ and $|d_1| \geq \cdots \geq |d_{K^*}| > 0 = |d_{K^*+1}| \cdots = |d_n|$. Then for any $\tau \geq 0$, the minimizers of (3) are given by

$$\hat{v}_\tau^2 = \underset{v^2 \geq 0}{\arg\min} \left\{ v^2(nv^2 - 2\,\mathrm{trace}(\mathbf{S} - \sigma^2\mathbf{I}_n)) - \sum_{k=1}^{K^*} (d_k - \tau - v^2)_+^2 \right\}, \tag{4}$$

$$\hat{\mathbf{M}}_\tau = (\mathbf{F}'\mathbf{F})^{-1/2}\mathbf{P}_{K^*}\,\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{P}'_{K^*}(\mathbf{F}'\mathbf{F})^{-1/2}, \tag{5}$$

where $x_+ \equiv \max(x, 0)$.

Note that $\hat{v}_\tau^2$ in (4) can be efficiently computed by a golden section search in $[0, d_1]$ when $d_1, \ldots, d_{K^*}$ are available. Since $(\mathbf{F}'\mathbf{F})^{-1/2}$ and $\mathbf{L}'(\mathbf{S} - \sigma^2\mathbf{I}_n)\mathbf{L}$ are $K \times K$ matrices, the computational complexity for $\hat{\mathbf{M}}_\tau$ is largely reduced when $K \ll n$.

2.3 Spatial Prediction

Suppose that $\mathbf{M}$ is estimated by $\hat{\mathbf{M}}_\tau$ and v^2 is estimated by $\hat{v}_\tau^2$ for some $\tau \geq 0$. Then for $\mathbf{s} \in D$ and $t = 1, \ldots, T$, the estimated best linear unbiased predictor (EBLUP) of $y(\mathbf{s}, t)$ and the corresponding estimated mean squared prediction error are given by

$$\hat{y}(\mathbf{s}, t) = \big(\mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\boldsymbol{\delta}'\big)\hat{\mathbf{V}}_\tau^{-}\mathbf{z}_t,$$
$$\widehat{\mathrm{var}}\left(\hat{y}(\mathbf{s}, t) - y(\mathbf{s}, t)\right) = \mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{f}(\mathbf{s}) + \hat{v}_\tau^2 - \big(\mathbf{f}(\mathbf{s})'\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\boldsymbol{\delta}'\big)\hat{\mathbf{V}}_\tau^{-}\big(\mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{f}(\mathbf{s}) + \hat{v}_\tau^2\boldsymbol{\delta}\big),$$

where $\boldsymbol{\delta} \equiv (I(\mathbf{s} = \mathbf{s}_1), \ldots, I(\mathbf{s} = \mathbf{s}_n))'$ and $\hat{\mathbf{V}}_\tau = \mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{F}' + \hat{v}_\tau^2\mathbf{I}_n + \sigma^2\mathbf{I}_n$. From (5), we have

$$\mathbf{F}\hat{\mathbf{M}}_\tau\mathbf{F}' = \mathbf{L}\mathbf{P}_{K^*}\,\mathrm{diag}\left((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\right)\mathbf{P}'_{K^*}\mathbf{L}', \qquad (6)$$

which implies

$$\hat{\mathbf{V}}_\tau = \mathbf{L}\mathbf{P}_{K^*}\,\mathrm{diag}\big((d_1 - \tau - \hat{v}_\tau^2)_+, \ldots, (d_{K^*} - \tau - \hat{v}_\tau^2)_+\big)\mathbf{P}'_{K^*}\mathbf{L}' + \hat{v}_\tau^2\mathbf{I}_n + \sigma^2\mathbf{I}_n.$$

Applying the Sherman-Morrison-Woodbury formula ([10]), the Moore-Penrose inverse of $\hat{\mathbf{V}}_\tau$ is

$$\begin{cases} \dfrac{1}{\hat{v}_\tau^2 + \sigma^2}\mathbf{I}_n - \dfrac{1}{\hat{v}_\tau^2 + \sigma^2}\mathbf{L}\mathbf{P}_{K^*}\mathrm{diag}\left(\dfrac{\hat{d}_1}{\hat{d}_1^2 + \hat{v}_\tau^2 + \sigma^2}, \ldots, \dfrac{\hat{d}_{K^*}}{\hat{d}_{K^*}^2 + \hat{v}_\tau^2 + \sigma^2}\right)\mathbf{P}'_{K^*}\mathbf{L}'; & \text{if } \hat{v}_\tau^2 + \sigma^2 > 0, \\ \mathbf{L}\mathbf{P}_{K^*}\left(\mathrm{diag}\big(\hat{d}_1, \ldots, \hat{d}_{K^*}\big)\right)^{-}\mathbf{P}'_{K^*}\mathbf{L}'; & \text{if } \hat{v}_\tau^2 + \sigma^2 = 0, \end{cases}$$

where $\hat{d}_k \equiv (d_k - \tau - \hat{v}_\tau^2)_+$; $k = 1, \ldots K^*$. Since $\mathbf{L}$ and $\mathbf{P}'_{K^*}$ are much smaller than $\hat{\mathbf{V}}_\tau$ if $K \ll n$, we can compute $\hat{\mathbf{V}}_\tau^{-}\mathbf{z}_t$ and $\hat{y}(\mathbf{s}, t)$ in $O(K^2 n)$, and gain considerable computation efficiency.

When σ^2 is unknown, we propose to estimate it by minimizing $\|\mathbf{F}\mathbf{M}\mathbf{F}' + \sigma^2\mathbf{I}_n - \mathbf{S}\|_F^2$ over all $\mathbf{M} \succeq \mathbf{0}$. Based on similar procedure of Proposition 1, we obtain

$$\hat{\sigma}^2 = \underset{\sigma^2 \geq 0}{\arg\min}\,\frac{1}{2}\left\{\sigma^2(n\sigma^2 - 2\,\mathrm{trace}(\mathbf{S})) - \sum_{k=1}^{K}(\gamma_k - \sigma^2)_+^2\right\}, \qquad (7)$$

where γ_k's are the eigenvalues of $(\mathbf{F}'\mathbf{F})^{-1/2}\mathbf{F}'\mathbf{S}\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1/2}$ with $\gamma_1 \geq \cdots \geq \gamma_K$.

2.4 Choice of the Shrinkage Parameter

An L-fold cross-validation is applied to choose the shrinkage parameter τ. We first randomly decompose $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ into L disjoint subsets $\mathcal{D}_1, \ldots, \mathcal{D}_L$ with $|\mathcal{D}_1|, \ldots, |\mathcal{D}_L|$ being equal or as close as possible. Let $\hat{y}^{(\ell)}(\mathbf{s}, t; \tau)$ be the EBLUP of

$y(\mathbf{s},t)$ based on only the data observed at $\{\mathbf{s}_1,\ldots,\mathbf{s}_n\}\setminus\mathcal{D}_\ell$ for $\mathbf{s}\in D$, $t=1,\ldots,T$ and $\ell=1,\ldots,L$. Then the proposed L-fold CV criterion is

$$\mathrm{CV}(\tau)=\sum_{\ell=1}^{L}\sum_{t=1}^{T}\sum_{\mathbf{s}\in\mathcal{D}_\ell}\left\{z(\mathbf{s},t)-\hat{y}^{(\ell)}(\mathbf{s},t;\tau)\right\}^2.$$

The final shrinkage parameter is $\hat{\tau}$, which minimizes $\mathrm{CV}(\tau)$ over $\tau\geq 0$.

3 Example

We considered a class of Fourier basis functions in $\mathbb{R}$ with $f_1(s)=1$ and

$$f_k(s)=\begin{cases}\cos(\pi k s); & \text{if } k\in\{2,4,\ldots\},\\ \sin(\pi(k-1)s); & \text{if } k\in\{3,5,\ldots\}.\end{cases}$$

Similar to [1], we generated the process $y(s,t)$ on $D=[0,1]$ according to (2) with $K=10$, $n=256$ and $v^2=0.041$, where $\mathbf{M}$ was chosen such that $\|\mathbf{FMF}'-\mathbf{V}_0\|_F$ is minimized, and $\mathbf{V}_0$ is an $n\times n$ matrix with its (i,j)-th component $\exp(-|i-j|/25)$. Figure 1 shows the resulting spatial covariance function $C(s,s^*)=\mathrm{cov}(y(s,t),y(s^*,t))$.

The data $\{\mathbf{z}_t : t=1,\ldots,T\}$ were simulated at $s_i=i/256$; $i=1,\ldots,n$, according to (1) with $\sigma^2=0.082$. We considered three values of $T=3,5,10$. For each T, we applied the proposed method developed in Sections 2.2 and 2.3 with the shrinkage parameter $\hat{\tau}$ selected by CV, where different values of $K=8,10,12,14,16$ were considered and σ^2 was assumed known. The proposed method is compared with a simple kriging method that estimates $C(s,s^*)$ using maximum likelihood (ML) based on the stationary exponential covariance model. The results in terms of the mean squared prediction error (MSPE),

$$\mathrm{MSPE}=\sum_{t=1}^{T}\int_{s\in D}\left|y(s,t)-\hat{y}(s,t)\right|_2^2, \tag{8}$$

are summarized in Table 1 based on 200 simulation replications, where $\hat{y}(s,t)$ is a generic estimate of $y(s,t)$. The MSPEs based on the proposed method with $\tau=0$ and those based the true covariance function of $C(s,s^*)$ are also shown for comparison.

Basically, our method based on either $\hat{\tau}$ or $\tau=0$ performs better than the exponential model. When T is small and K is large, the proposed method based on $\hat{\tau}$ performs better than that based on $\tau=0$, indicating the advantage of shrinkage. On the other hand, when T is large or K is small, CV tends to select $\hat{\tau}=0$, making the shrinkage less useful.

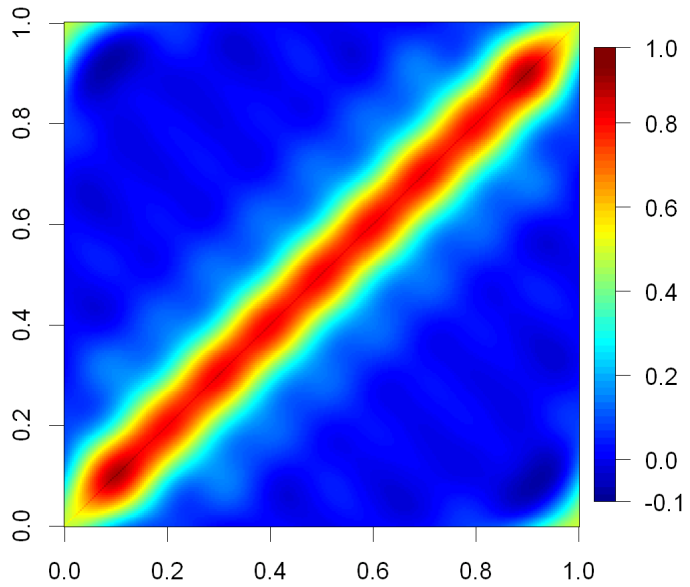


Figure 1. One-dimensional spatial covariance function of $y(s,t)$.

Table 1. MSPE performance of various methods based on different values of T and K, where the values in parentheses are the corresponding standard errors.

Methods		$T=3$	$T=5$	$T=10$
True		0.0457 (0.0001)	0.0457 (0.0001)	0.0457 (0.0001)
Exponential		0.0550 (0.0002)	0.0550 (0.0002)	0.0546 (0.0001)
CV	$K=10$	0.0461 (0.0002)	0.0461 (0.0001)	0.0460 (0.0001)
	$K=12$	0.0469 (0.0001)	0.0468 (0.0001)	0.0469 (0.0001)
	$K=14$	0.0475 (0.0002)	0.0475 (0.0001)	0.0476 (0.0001)
	$K=16$	0.0484 (0.0002)	0.0484 (0.0002)	0.0483 (0.0001)
$\tau=0$	$K=10$	0.0461 (0.0002)	0.0461 (0.0001)	0.0460 (0.0001)
	$K=12$	0.0469 (0.0001)	0.0469 (0.0001)	0.0469 (0.0001)
	$K=14$	0.0477 (0.0002)	0.0476 (0.0001)	0.0476 (0.0001)
	$K=16$	0.0486 (0.0002)	0.0485 (0.0002)	0.0483 (0.0001)

4 Discussion

Our penalized least-squares estimation proposal results in a simple shrinkage-based modification of FRK by cutting down the eigenvalues of a spatial covariance matrix. As shown in the example in Section 3, our method can effectively control estimation variability via tuning the shrinkage parameter, especially when T is small and K is large. Moreover, a fast estimation and prediction method still exists, which avoids taking a high-dimensional matrix inversion.

We did not really deal with the basis choice issue in this work, though various number of basis numbers are compared in the example. This issue can be a very challenging problem and deserves further research.

References

1. Bradley, J., Cressie, N., and Shi, T. (2011). Selection of rank and basis functions in the spatial random effects model. In *2011 Proceedings of the Joint Statistical Meetings*, pages 3393–3406. American Statistical Association Alexandria, VA.
2. Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:209–226.
3. Fanshawe, T. R. and Diggle, P. J. (2012). Bivariate geostatistical modelling: a review and an application to spatial variation in radon concentrations. *Environmental and Ecological Statistics*, 19:139–160.
4. Lemos, R. T. and Sansó, B. (2012). Conditionally linear models for non-homogeneous spatial random fields. *Statistical Methodology*, 9:275–284.
5. Majumdar, A. and Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology*, 39:225–245.
6. Tzeng, S. and Huang, H.-C. (2015). Non-stationary multivariate spatial covariance estimation via low-rank regularization. *Statistica Sinica*, to appear.
7. Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:275–294.
8. Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004). Flexible spatial models for kriging and cokriging using moving averages and the fast fourier transform (fft). *Journal of Computational and Graphical Statistics*, 13:265–282.
9. Wikle, C. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics*, pages 107–118. CRC Press, Boca Raton, Florida, USA.
10. Woodbury, M. A. (1950). Inverting modified matrices. Technical report, Memorandum report 42, Statistical Research Group, Princeton University, Princeton.

New dissimilarity measure for aggregated symbolic data with real and categorical variables

Nobuo Shimizu[1], Junji Nakano[1], and Yoshikazu Yamamoto[2]

[1] The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan
nobuo@ism.ac.jp, nakanoj@ism.ac.jp
[2] Tokushima Bunri University, 1314-1 Shido, Sanuki, Kagawa 769-2193, Japan
yamamoto@fe.bunri-u.ac.jp

Abstract. In recent "Big data" era, huge amount of individual data sets are available and should be analyzed for getting useful information. In many cases, they can be divided into some naturally defined groups, and the information of each group is expressed by several statistics calculated from the original data in the group. We consider that individual data consists of real and categorical variables and use up to second moments of data to express their characteristics. We call such set of descriptive statistics as "aggregated symbolic data", and define a dissimilarity measure between two of them based on log-likelihood ratio test statistics. We illustrate the usefulness of our dissimilarity by clustering real data examples.

1 Introduction

Nowadays we have often huge amount of complex data sets called "Big data" from various fields. We are often required to handle them and get useful information from them. As the number of data is too huge and the structure of variables are complicated, we need to develop new statistical methods for handling them. When we have too many individual data, it is almost impossible to examine each individual data. It often happens that individual data can be divided into some naturally defined groups, for example, species of living thing.

Symbolic data analysis [1–3] is one of techniques to handle such data for expressing and analyzing groups. In the traditional symbolic data analysis, several statistics calculated by using information about marginal distributions are mainly utilized. Among traditional symbolic data interval data was extensively studied. Various techniques for analyzing interval data such as principle component analysis, regression analysis, clustering analysis were developed.

In this paper, we consider the case where individual data has both real variables and categorical variables. We consider up to second order moments of data in a group as symbolic data, and call them "aggregated symbolic data" (ASD). By using the definition of ASD we propose a new dissimilarity measure between two groups because we hope to discriminate ASD.

2 Individual data and aggregated symbolic data

Let we have n individual data and they are divided into G meaningful groups. Let group 1 has $n^{(1)}$ individuals, ... , group G has $n^{(G)}$ individuals, so $n^{(1)} + n^{(2)} + \cdots + n^{(G)} = n$. We assume that each individual data is expressed by p real variables and q categorical variables. We denote the value of variable j of individual i in group g be $x_{ij}^{(g)}$. For $j = 1, \ldots, p$, $x_{ij}^{(g)}$ takes a real value, and for $j = 1, \ldots, q$, $x_{i,p+j}^{(g)}$ takes one of m_j categorical values, for example, "1",..., "m_j". Since it is difficult to deal with original categorical variables as it is, we use dummy variable to express them. We define a dummy variable $x_{ij}^{(g,k)}$ whose value is 1 when the value of categorical variable k of individual i in group g is "j", otherwise 0. Therefore, individual data of group g is expressed by an $n^{(g)} \times (p + m_1 + \cdots + m_q)$ matrix

$$X^{(g)} = \begin{bmatrix} x_{11}^{(g)} & \cdots & x_{1p}^{(g)} & x_{11}^{(g,1)} & \cdots & x_{1m_1}^{(g,1)} & \cdots & x_{11}^{(g,q)} & \cdots & x_{1m_q}^{(g,q)} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ x_{n^{(g)}1}^{(g)} & \cdots & x_{n^{(g)}p}^{(g)} & x_{n^{(g)}1}^{(g,1)} & \cdots & x_{n^{(g)}m_1}^{(g,1)} & \cdots & x_{n^{(g)}1}^{(g,q)} & \cdots & x_{n^{(g)}m_q}^{(g,q)} \end{bmatrix}.$$

Clearly an $n^{(g)} \times m_k$ matrix

$$X_2^{(g,k)} = \begin{bmatrix} x_{11}^{(g,k)} & \cdots & x_{1m_k}^{(g,k)} \\ \vdots & & \vdots \\ x_{n^{(g)}1}^{(g,k)} & \cdots & x_{n^{(g)}m_k}^{(g,k)} \end{bmatrix}$$

expresses k-th categorical variable of the group, $X_2^{(g)} = [X_2^{(g,1)} \cdots X_2^{(g,q)}]$ expresses q categorical variables and $X_1^{(g)}$expresses p real variables where $X^{(g)} = [X_1^{(g)} X_2^{(g)}]$.

When $n^{(g)}$ is large, we hope to describe the characteristics of $X^{(g)}$ concisely without losing too much information. For real variables, simple descriptive statistics are means, variances and covariances, that is, first and second moments of data. We use the same statistics for our data.

First order moments of $X^{(g)}$ is calculated by $\mathbf{1}'_{n^{(g)}} X^{(g)}$ where $\mathbf{1}_{n^{(g)}}$ is an $n^{(g)}$ dimensional vector whose elements are all 1. If we define $\bar{x}_l^{(g)}$is a mean of real variable l and $\hat{p}_j^{(g,k)} = \frac{1}{n^{(g)}} \sum_{i=1}^{n^{(g)}} x_{ij}^{(g,k)}$ is an empirical marginal probability of j-th value of categorical variable k, it is clear

$$\mathbf{1}'_{n^{(g)}} X^{(g)} = n^{(g)} [\bar{x}_1^{(g)} \cdots \bar{x}_p^{(g)} \hat{p}_1^{(g,1)} \cdots \hat{p}_{m_1}^{(g,1)} \cdots \hat{p}_1^{(g,q)} \cdots \hat{p}_{m_q}^{(g,q)}].$$

Second order moments of $X^{(g)}$ is given by$X^{(g)'} X^{(g)} = S^{(g)}$. $S^{(g)}$ is divided into four submatrix: $S_{ij}^{(g)} = X_i^{(g)'} X_j^{(g)}$. $S_{11}^{(g)}$ is a product sum matrix of p real

variables, $S_{21}^{(g)}$is a matrix expressing sum of real variables for each categorical value, and

$$S_{22}^{(g)} = \begin{bmatrix} S^{(g,11)} \cdots S^{(g,1q)} \\ \vdots \qquad\quad \vdots \\ S^{(g,q1)} \cdots S^{(g,qq)} \end{bmatrix}$$

is a Burt matrix where

$$S^{(g,k_1k_2)} = X_2^{(g,k_1)\prime} X_2^{(g,k_2)} = \begin{bmatrix} s_{11}^{(g,k_1k_2)} \cdots s_{1m_l}^{(g,k_1k_2)} \\ \vdots \qquad\quad \vdots \\ s_{m_{k_1}1}^{(g,k_1k_2)} \cdots s_{m_{k_1}m_{k_2}}^{(g,k_1k_2)} \end{bmatrix}$$

is a contingency table for categorical variables k_1 and k_2 [4].

We use $n^{(g)}$,$\mathbf{1}'_{n^{(g)}}X^{(g)}$,$S^{(g)}$ as aggregated symbolic data to describe the group g.

3 Dissimilarity measure between two groups

When we have groups of individual data, we often hope to define a dissimilarity measure among them. We define it by using the statistics to test the same stochastic structure between two groups. We have to note that the statistics should include just ASD, not individual data in our situation.

If two groups of data are similar, their data generating systems have same stochastic structures, that is, parameters of two stochastic generating processes are same. We call this assumption is a "same parameter model". If two groups have different stochastic structures, parameters can take different values for two groups. This assumption is called a "different parameter model". A dissimilarity between two groups can be expressed by a statistic for testing the same parameter model against the different parameter model. We use log-likelihood ratio statistic for this purpose.

As we have up to second moments, we can consider relationships between just two variables. We decide to ignore relationships among more than two variables. As we have real variables and categorical variables, we have three kinds of combinations: two categorical variables, one categorical variable and one real variable, and two real variables. Consider
$[x_{i1}^{(g)} \cdots x_{ip}^{(g)} x_{i1}^{(g,1)} \cdots x_{im_1}^{(g,1)} \cdots x_{i1}^{(g,q)} \cdots x_{im_q}^{(g,q)}]$ be an i-th realization of the random variable $[X_1^{(g)} \cdots X_p^{(g)} X_1^{(g,1)} \cdots X_{m_1}^{(g,1)} \cdots X_1^{(g,q)} \cdots X_{m_q}^{(g,q)}]$ whose stochastic structure describes the characteristics of group g.

3.1 Likelihood ratio test statistic for two categorical variables

First, we consider two categorical variables k_1 and k_2. This means that we consider two random variables $[X_1^{(g,k_1)} \cdots X_{m_{k_1}}^{(g,k_1)}]'$, $[X_1^{(g,k_2)} \cdots X_{m_{k_2}}^{(g,k_2)}]'$. We can

describe the stochastic structure of these random variables by parameters $Pr(X_{j_1}^{(g,k_1)} = 1, X_{j_2}^{(g,k_2)} = 1) = p_{j_1 j_2}^{(g,k_1 k_2)}$. This also means that a cell occurrence $s_{j_1 j_2}^{(g,k_1 k_2)}$ follows a multinomial distribution with probability $p_{j_1 j_2}^{(g,k_1 k_2)}$.

If we assume a different parameter model, the maximum likelihood estimators of $p_{j_1 j_2}^{(g,k_1 k_2)}$ is $s_{j_1 j_2}^{(g,k_1 k_2)}/n^{(g)}$. Thus the maximum log-likelihood is

$$\hat{l}_1^{(g_1 g_2, k_1 k_2)} = \sum_{a=1}^{2} \left\{ \log \left(\frac{n^{(g_a)}!}{\prod_{j_1=1}^{m_{k_1}} \prod_{j_2=1}^{m_{k_2}} s_{j_1 j_2}^{(g_a,k_1 k_2)}!} \right) + \sum_{j_1=1}^{m_{k_1}} \sum_{j_2=1}^{m_{k_2}} s_{j_1 j_2}^{(g_a,k_1 k_2)} \log \frac{s_{j_1 j_2}^{(g_a,k_1 k_2)}}{n^{(g_a)}} \right\}.$$

If parameters of distributions of g_1 and g_2 take the same set of values $\left\{p_{j_1 j_2}^{(g_1 g_2, k_1 k_2)}\right\}$, maximum likelihood estimators of $p_{j_1 j_2}^{(g_1 g_2, k_1 k_2)}$ is $\hat{p}_{j_1 j_2}^{(g_1 g_2, k_1 k_2)} = \frac{s_{j_1 j_2}^{(g_1,k_1 k_2)} + s_{j_1 j_2}^{(g_2,k_1 k_2)}}{n^{(g_1)} + n^{(g_2)}}$. Therefore the maximum log-likelihood for the same parameter model is

$$\begin{aligned} \hat{l}_0^{(g_1 g_2, k_1 k_2)} = & \sum_{a=1}^{2} \log \left\{ \frac{n^{(g_a)}!}{\prod_{j_1=1}^{m_{k_1}} \prod_{j_2=1}^{m_{k_2}} s_{j_1 j_2}^{(g_a,k_1 k_2)}!} \right\} \\ & + \sum_{j_1=1}^{m_{k_1}} \sum_{j_2=1}^{m_{k_2}} (s_{j_1 j_2}^{(g_1,k_1 k_2)} + s_{j_1 j_2}^{(g_2,k_1 k_2)}) \log \frac{s_{j_1 j_2}^{(g_1,k_1 k_2)} + s_{j_1 j_2}^{(g_2,k_1 k_2)}}{n^{(g_1)} + n^{(g_2)}}. \end{aligned}$$

Thus, log-likelihood ratio tests statistic for two categorical variables is defined by

$$\begin{aligned} d_{(cc)}^{(g_1,g_2)} = & -2 \sum_{k_1=1}^{q} \sum_{k_2=1}^{q} \sum_{j_1=1}^{m_{k_1}} \sum_{j_2=1}^{m_{k_2}} \left\{ (s_{j_1 j_2}^{(g_1,k_1 k_2)} + s_{j_1 j_2}^{(g_2,k_1 k_2)}) \log \frac{s_{j_1 j_2}^{(g_1,k_1 k_2)} + s_{j_1 j_2}^{(g_2,k_1 k_2)}}{n^{(g_1)} + n^{(g_2)}} \right\} \\ & + 2 \sum_{k_1=1}^{q} \sum_{k_2=1}^{q} \sum_{j_1=1}^{m_{k_1}} \sum_{j_2=1}^{m_{k_2}} \left\{ s_{j_1 j_2}^{(g_1,k_1 k_2)} \log \frac{s_{j_1 j_2}^{(g_1,k_1 k_2)}}{n^{(g_1)}} + s_{j_1 j_2}^{(g_2,k_1 k_2)} \log \frac{s_{j_1 j_2}^{(g_2,k_1 k_2)}}{n^{(g_2)}} \right\} \end{aligned}$$

3.2 Likelihood ratio test statistic for a real variable and a categorical variable

Here we consider the interaction between a real variable l and a categorical variable k. Under the condition that the categorical variable takes value "j", we assume that the real variable follows a normal distribution whose mean is $\mu_{(j)l}^{(g,k)}$ and the variance $\sigma_{ll}^{(g)}$. If we assume a different parameter model, maximum likelihood estimators of parameter $\mu_{(j)l}^{(g,k)}$ is $\hat{\mu}_{(j)l}^{(g,k)} = \frac{1}{n_j^{(g,k)}} \mathbf{x}_j^{(g,k)'} \mathbf{x}_l^{(g)}$ where $\mathbf{x}_j^{(g,k)'} = [x_{1j}^{(g,k)} \dots x_{n^{(g)}j}^{(g,k)}]$ and $\mathbf{x}_l^{(g)} = [x_{1l}^{(g)} \dots x_{n^{(g)}l}^{(g)}]'$. Note that these values are all given in $S_{12}^{(g)}$ and $S_{22}^{(g)}$. We should note that $\sigma_{ll}^{(g)}$ is a parameter which is not related with a categorical variable k. This value is given in $S_{11}^{(g)}$ described in the

next subsection. The marginal distribution of $X_l^{(g)}$ is clearly given by a normal mixture distribution $\sum_{j=1}^{m_k} \varphi(x_l^{(g)}|\mu_{(j)l}^{(g,k)}, \sigma_{ll}^{(g)}) p_j^{(g,k)}$ where $p_j^{(g,k)} = \sum_{j_2=1}^{m_{k_2}} p_{jj_2}^{(g,kk_2)}$.

If we assume a different parameter model, the maximum log-likelihood is

$$\hat{l}_1^{(g_1g_2,lk)} = \sum_{a=1}^{2} \left\{ -\frac{n^{(g_a)}}{2} \log(2\pi\sigma_{ll}^{(g_a)}) - \frac{1}{2\sigma_{ll}^{(g_a)}} \sum_{i|x_{ij}^{(g_a,k)}=1} (x_{il}^{(g_a)} - \hat{\mu}_{(j)l}^{(g_a,k)})^2 \right\}$$

where $\sum_{i|x_{ij}^{(g_a,k)}=1}$denotes the summation for i when $x_{ij}^{(g_a,k)} = 1$.

In a same parameter model, we assume $\mu_{(j)l}^{(g_1,k)} = \mu_{(j)l}^{(g_2,k)} = \mu_{(j)l}^{(g_1g_2,k)}$ and use different $\sigma_{ll}^{(g_1)}$ and $\sigma_{ll}^{(g_2)}$. Then the total maximum log-likelihood can be calculated as

$$d_{(rc)}^{(g_1,g_2)} = \sum_{l=1}^{p} \sum_{k=1}^{q} \sum_{j=1}^{m_k} \frac{n_j^{(g_1,k)} n_j^{(g_2,k)}}{\left(n_j^{(g_1,k)} + n_j^{(g_2,k)}\right)^2} \left(\frac{n_j^{(g_2,k)}}{\sigma_{ll}^{(g_1)}} + \frac{n_j^{(g_1,k)}}{\sigma_{ll}^{(g_2)}} \right) (\hat{\mu}_{(j)l}^{(g_1,k)} - \hat{\mu}_{(j)l}^{(g_2,k)})^2$$

3.3 Likelihood ratio test statistic for two real variables and overall dissimilarity

At last, we consider two real variables. If we have up to second order moments, it is usual to assume a bivariate normal distribution. However, as was shown in the previous subsection, each real variable follows a normal mixture distribution under our models. However, we still use bivariate normal distribution as an approximation of the normal mixture distribution. This is verified because we decide to consider just two variables at a time and use up to second moments of individual data.

As we have p real variables, we assume they are distributed by a multivariate normal distribution whose mean is $\boldsymbol{\mu}^{(g)}$ and covariance matrix $\Sigma^{(g)} = [\sigma_{ij}^{(g)}]$. So the log-likelihood function of them are written by

$$\begin{aligned} l(\boldsymbol{\mu}^{(g)}, \Sigma^{(g)}) = &-\frac{n^{(g)}p}{2} \log(2\pi) - \frac{n^{(g)}}{2} \log|\Sigma^{(g)}| \\ &-\frac{1}{2} \sum_{i=1}^{n^{(g)}} (\mathbf{x}_i^{(g)} - \boldsymbol{\mu}^{(g)})' \left(\Sigma^{(g)}\right)^{-1} (\mathbf{x}_i^{(g)} - \boldsymbol{\mu}^{(g)}) \end{aligned}$$

where $X_1^{(g)} = [\mathbf{x}_1^{(g)} \cdots \mathbf{x}_{n^{(g)}}^{(g)}]'$. Maximum likelihood estimators of $\boldsymbol{\mu}^{(g)}$ and $\Sigma^{(g)}$ are $\hat{\boldsymbol{\mu}}^{(g)} = \frac{1}{n^{(g)}} \sum_{i=1}^{n^{(g)}} \mathbf{x}_i^{(g)}$ and $\hat{\Sigma}^{(g)} = \frac{1}{n^{(g)}} \sum_{i=1}^{n^{(g)}} (\mathbf{x}_i^{(g)} - \hat{\boldsymbol{\mu}}^{(g)})(\mathbf{x}_i^{(g)} - \hat{\boldsymbol{\mu}}^{(g)})'$ for $g = g_1, g_2$ if we adopt a different parameter model.

If we assume a same parameter model, groups g_1 and g_2 have the same set of parameter values $\left\{\boldsymbol{\mu}^{(g_1g_2)}, \Sigma^{(g_1g_2)}\right\}$. In this case, the maximum likelihood

estimators of $\boldsymbol{\mu}^{(g_1g_2)}$ and $\Sigma^{(g_1g_2)}$ are $\hat{\boldsymbol{\mu}}^{(g_1g_2)} = \frac{n^{(g_1)}\hat{\boldsymbol{\mu}}^{(g_1)}+n^{(g_2)}\hat{\boldsymbol{\mu}}^{(g_2)}}{n^{(g_1)}+n^{(g_2)}}$ and $\hat{\Sigma}^{(g_1g_2)} = \frac{n^{(g_1)}\hat{\Sigma}^{(g_1)}+n^{(g_2)}\hat{\Sigma}^{(g_2)}}{n^{(g_1)}+n^{(g_2)}} + \boldsymbol{\delta}^{(g_1,g_2)}\boldsymbol{\delta}^{(g_1,g_2)'}$ where $\boldsymbol{\delta}^{(g_1,g_2)} = \frac{\sqrt{n^{(g_1)}n^{(g_2)}}}{n^{(g_1)}+n^{(g_2)}}(\hat{\boldsymbol{\mu}}^{(g_1)} - \hat{\boldsymbol{\mu}}^{(g_2)})$.

Therefore log-likelihood ratio test statistic of real variables is

$$d_{(rr)}^{(g_1,g_2)} = (n^{(g_1)} + n^{(g_2)}) \log|\hat{\Sigma}^{(g_1g_2)}| - n^{(g_1)} \log|\hat{\Sigma}^{(g_1)}| - n^{(g_2)} \log|\hat{\Sigma}^{(g_2)}|.$$

By combining three above log-likelihood ratio test statistics, we obtain overall dissimilarity between two groups:

$$d^{(g_1,g_2)} = d_{(rr)}^{(g_1,g_2)} + d_{(cc)}^{(g_1,g_2)} + d_{(rc)}^{(g_1,g_2)}.$$

4 Examples

Table 1 is a 2004 car data given by Unwin [5], which contains about 400 records with 10 real and 4 categorical variables. A categorical variable "Country" is added by us. We divide the real estate data into 6 groups by the values of "Country", and calculate dissimilarity of ASD proposed above and used them for hierarchical clustering.

Table 1. Part of car data in all over the world in 2004

Vehicle Name	Country	Price(USD)	···	Length(cm)	Type	···	Drive
Chevrolet Aveo 4dr	US	11690	···	167	Sedan	···	front
Hyundai Santa Fe GLS	Korea	21589	···	177	Sedan	···	front
Saab 9-5 Aero	Sweden	40845	···	190	Wagon	···	AWD
Honda Odyssey LX	Japan	24950	···	201	Mini Van	···	front
Nissan Murano SL	Japan	28739	···	188	Wagon	···	rear
Jaguar XKR coupe 2dr	UK	81995	···	187	Sports Car	···	rear
BMW X3 3.0i	Germany	37000	···	180	SUV	···	AWD
⋮	⋮	⋮	⋱	⋮	⋮	⋱	⋮

As a result of the hierarchical clustering of ASD using our dissimilarity measure and complete linkage technique, the dendrogram (Figure 1) is given. Germany and UK produced rather luxurious cars, US and Japan produced rather different kind of cars.

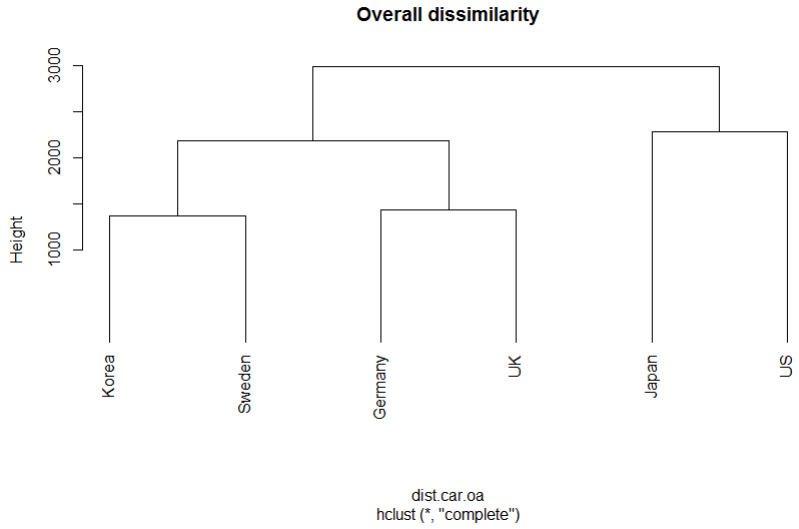


Fig. 1. Hierarchical clustering result of car data by using overall dissimilarity among 6 ASD

Next example is a real estate data in the special wards of Tokyo, Japan, which contains about 790,000 records with 5 real and 76 categorical variables. We divide the data into 23 groups by the categorical variable "Ward", and calculate dissimilarity of ASD and performed hierarchical clustering.

Table 2. Part of real estate data in special wards of Tokyo

No.	Ward	Rent(JPY)	Area(m^2)	$\cdots$	Deposit months	Bed rooms	Floor	$\cdots$	Immediately
1	Arakawa	82500	26.83	$\cdots$	0.00	1	3	$\cdots$	Yes
2	Meguro	51000	9.96	$\cdots$	1.00	0	2	$\cdots$	Yes
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
4594	Minato	372000	114.93	$\cdots$	2.00	3	5	$\cdots$	No
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
17277	Chuo	270000	89.72	$\cdots$	2.00	3	14	$\cdots$	Yes
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
494893	Itabashi	55000	20.16	$\cdots$	0.50	1	5	$\cdots$	Yes
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$

The resulted dendrogram was given in Figure 2. Chuo and Minato wards are both located in the center of Tokyo and the dissimilarity between these two wards and other wards are far, because there are very high-grade properties in these two wards.

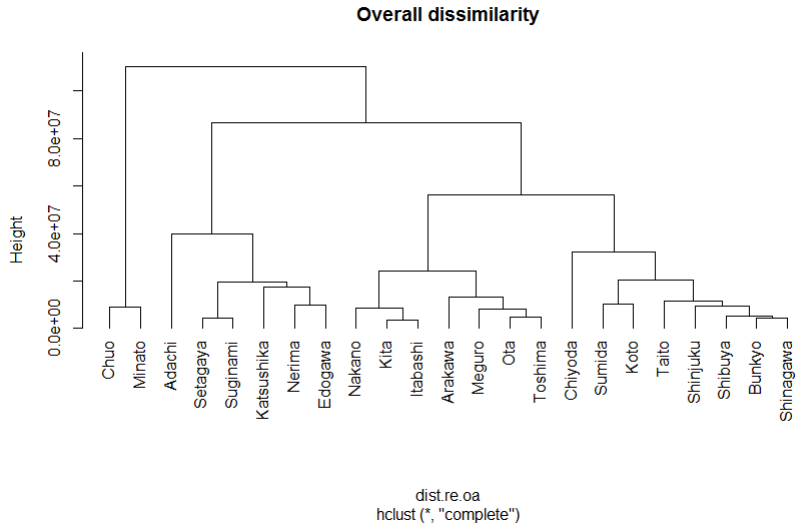


Fig. 2. Hierarchical clustering result of real estate data by using overall dissimilarity among 23 ASD

5 Concluding remark

As a kind of symbolic data, we propose aggregated symbolic data for groups of individual data which include real and categorical variables by using up to second moments. We define a dissimilarity between two aggregated symbolic data by using several log-likelihood ratio test statistics. We use this dissimilarity to two real data sets and obtain appropriate results.

References

1. Billard, L. and Diday, E.: *Symbolic data analysis: Conceptual statistics and data mining*, John Wiley & Sons Ltd, Chichester, UK, (2006).
2. Bock, H.-H., Diday, E.: *Analysis of symbolic data: exploratory methods for extracting statistical information from complex data*, Springer-Verlag, Berlin, (2000).
3. Diday, E., Noirhomme-Fraiture, M.: *Symbolic Data Analysis and the SODAS Software*, John Wiley & Sons Ltd, Chichester, UK, (2008).
4. Greenacre, M. and Blasius, J.: *Multiple correspondence analysis and related methods*, Chapman & Hall/CRC, London, (2006).
5. Unwin, A., Theus, M. and Hofmann, H.: *Graphics of large datasets: Visualizing a million*, Springer, New York, (2006). (Car data is available at http://stats.math.uni-augsburg.de/GOLD/data/2004Cars.txt)

Clusterwise linear regression model for modal multi-valued data

Hiroyasu Abe[1], Kensuke Tanioka, and Hiroshi Yadohisa

Doshisha University, Kyoto 610-0394, Japan
eio1001@mail4.doshisha.ac.jp

Abstract. In this paper, we propose a clusterwise linear regression model for modal multi-valued data that extends the linear regression model proposed by Billard and Diday. In both the original model and our proposed model, modal multi-valued data are described as proportions of categories. Furthermore, the original model is considered to be a multivariate multiple regression model for modal multi-valued data. Data are generated by calculating percentages of individuals described by categorical values, where individuals are aggregated by each concept. One of the advantages for such aggregation is that we can treat each concept equally, even if the number of individuals in each concept is different. Note that in our proposed clusterwise linear regression model, more than one linear regression structure is assumed.

1 Introduction

In recent years, data tables have become extremely large. Symbolic data analysis is one approach to analyzing such large datasets. In symbolic data analysis, a "concept" is defined as the set of individuals targeted for analysis. Concepts can be expressed using symbolic variables whose values represent the description of structures, including intervals, histograms, multi-value sets, and distributions. One of the advantages of symbolic data analysis is that data interpretation can be simplified by transforming targets from large numbers of trivial individual data points to a smaller number of concepts. For example, individual transactions in access log data comprise very large data files and are typically used only for sampling; however, if the target of analysis (i.e., the concept) is changed to age groups, the number of data points becomes substantially smaller and interpretation becomes easier because each group has a specific concept of age.

In this paper, we focus on modal multi-valued variables that describe a concept; more specifically, we focus on linear regression models for this type of variable. The values of a modal multi-valued variable together form a set of discrete values that can be assigned to the variable and corresponding weight values, for example, relative frequency. Because a dependent modal variable is considered to be a group of dependent variables, linear regression models for symbolic variables are the same as multiple linear regression models.

Linear regression models, however, have the following problem. If data have two or more regression structures, one regression line cannot be fitted to the

data. To solve this problem, we propose a clusterwise regression model for modal multi-valued data. In our analysis, some multiple linear regression models for the dependent weight matrix are estimated and some clusters of the concepts are identified in terms of the residual. Furthermore, our clusterwise linear regression model typically handles one dependent vector; however, our proposed model treats some dependent vectors, i.e., via a dependent matrix.

2 Modal multi-valued variables

Modal multi-valued variables are represented by weights that are attached to each discrete value that can be assigned to the variable. Here the weight signifies the relative frequency for each discrete value. Let p be the number of modal multi-valued variables for each concept c ($c = 1, \ldots, N$), and let q_j be the number of possible values for the jth variable. The symbolic value of the cth concept for the jth modal multi-valued variable ($c = 1, \ldots, N;\ j = 1, \ldots, p$) is represented as

$$\xi_{cj} = \{\{\gamma_{j1},\ w_{cj1}\},\ \ldots,\ \{\gamma_{jq_j},\ w_{cjq_j}\}\},$$

where γ_{jk} ($j = 1, \ldots, p;\ k = 1, \ldots, q_j$) is the kth possible value for the jth symbolic variable and w_{cjk} ($c = 1, \ldots, N;\ j = 1, \ldots, p;\ k = 1, \ldots, q_j$) is the relative frequency of the cth concept for γ_{jk}. Thus, we have $\sum_{k=1}^{q_j} w_{cjk} = 1$ ($c = 1, \ldots, N;\ j = 1, \ldots, p$).

In this paper, we rewrite the modal multi-valued variable using a matrix representation as $\boldsymbol{W}_j = (w_{cjk})$, which is an $N \times q_j$ matrix whose elements are the weights for the jth modal multi-valued variable. We represent all datasets as $\boldsymbol{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_p)$. Table.1. provides example modal multi-valued data.

Table 1. Example modal multi-valued data

concept	variable 1		variable 2		
	γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{23}
1	0.2	0.8	0.3	0.3	0.4
2	0.4	0.6	0.7	0.1	0.2
3	0.9	0.1	0.5	0.4	0.1

3 Linear regression model for modal multi-valued data

As described above, a modal multi-valued variable is represented as multiple variables; therefore, a linear regression model for a modal multi-valued variable is the set of regression models for weights of each corresponding value of the given variable[1].

Let $\boldsymbol{Y}$ be the $N \times r$ weight value matrix that represents one dependent modal multi-valued variable, and let $\boldsymbol{X}_j = (\boldsymbol{x}_{j1}, \ \ldots, \ \boldsymbol{x}_{jq_j})$ be the jth $N \times q_j$ weight value matrix that represents the jth independent modal multi-valued variable. Because the sum of the elements for each row in $\boldsymbol{X}_j$ $(j = 1, \ \ldots, \ p)$ is 1, which indicates that $\boldsymbol{X}_j$ is rank deficient, we use $\boldsymbol{X}_j^\dagger = (\boldsymbol{x}_{j1}, \ \ldots, \ \boldsymbol{x}_{jq_j-1})$ $(j = 1, \ \ldots, \ p)$ in the model. Let $\boldsymbol{X}^\dagger = (\boldsymbol{X}_1^\dagger, \ \ldots, \ \boldsymbol{X}_p^\dagger)$ be the $N \times (q-p)$ $(q = \sum_{j=1}^p q_j)$ weight value matrix that represents p independent modal multi-valued variables, and let $\boldsymbol{X}^\star = \boldsymbol{J}_N \boldsymbol{X}^\dagger$ be the columnwise mean-centered matrix. Here $\boldsymbol{J}_N = \boldsymbol{I}_N - 1/N \mathbf{1}_N \mathbf{1}_N'$ is the $N \times N$ centering matrix, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{1}_N$ is the vector of N ones. The linear regression model for modal multi-valued data is then

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X}^\dagger \boldsymbol{B} + \mathbf{1}_N \boldsymbol{\beta}_0' + \boldsymbol{E} \\
&= \boldsymbol{X}^\star \boldsymbol{B} + \mathbf{1}_N \boldsymbol{\beta}_0' + 1/N \mathbf{1}_N \mathbf{1}_N' \boldsymbol{X}^\dagger \boldsymbol{B} + \boldsymbol{E} \\
&= \boldsymbol{X}^\star \boldsymbol{B} + \mathbf{1}_N \boldsymbol{\alpha}_0' + \boldsymbol{E} \\
&= \boldsymbol{X}^* \boldsymbol{B}^* + \boldsymbol{E},
\end{aligned}$$

where $\boldsymbol{B}$ is the $(q-p) \times r$ coefficient matrix, $\boldsymbol{\beta}_0$ is the r intercept vector, $\boldsymbol{X}^* = (\mathbf{1}_N \ \boldsymbol{X}^\star)$, $\boldsymbol{\alpha}_0 = \boldsymbol{\beta}_0 + 1/N \boldsymbol{B}' \boldsymbol{X}^{\dagger\prime} \mathbf{1}_N$, $\boldsymbol{B}^* = (\boldsymbol{\alpha}^0 \ \boldsymbol{B}')'$, and $\boldsymbol{E}$ is the $N \times r$ error matrix. This form of the model is the same as that of the multiple linear regression model[3]. The solution of $\boldsymbol{B}^*$ is derived as the objective function

$$Q = \|\boldsymbol{E}\|^2 = \|\boldsymbol{Y} - \boldsymbol{X}^* \boldsymbol{B}^*\|^2$$

is minimized. Here $\|\cdot\|$ denotes the Frobenius norm. Consequently, the solution is

$$\boldsymbol{B}^* = (\boldsymbol{X}^{*\prime} \boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*\prime} \boldsymbol{Y}.$$

4 Clusterwise linear regression model

Our clusterwise linear regression model assumes that there are groups of individuals and each of the groups has a different linear regression model. The groups and coefficients for each of the groups are derived via an iterative method. In this section, we describe the model for classical data.

Let p be the number of classical variables for each individual i $(i = 1, \ \ldots, \ n)$. Given vector $\boldsymbol{y}$ for a dependent variable with length n and an $n \times p$ matrix $\boldsymbol{X}$ for independent variables, the model is

$$\begin{aligned}
\boldsymbol{y}_g &= \boldsymbol{X}_g \boldsymbol{\beta}_g + \mathbf{1}_{n_g} \beta_{g0} + \boldsymbol{e}_g \\
&= \boldsymbol{X}_g^\star \boldsymbol{\beta}_g + \mathbf{1}_{n_g} \beta_{g0} + 1/n_g \mathbf{1}_{n_g} \mathbf{1}_{n_g}' \boldsymbol{X}_g \boldsymbol{\beta}_g + \boldsymbol{e}_g \\
&= \boldsymbol{X}_g^\star \boldsymbol{\beta}_g + \mathbf{1}_{n_g} \alpha_{g0} + \boldsymbol{e}_g \\
&= \boldsymbol{X}_g^* \boldsymbol{\beta}_g^* + \boldsymbol{e}_g \quad (g = 1, \ \ldots, \ G),
\end{aligned}$$

where G is the number associated with the group and $\boldsymbol{y}_g$, $\boldsymbol{X}_g^*$, $\boldsymbol{\beta}_g^*$, and $\boldsymbol{e}_g$ are the dependent vector with length n_g, the $n_g \times p$ independent matrix, the coefficient vector with length $p+1$ and the error vector with length n_g of the gth group, respectively. Here n_g is the number of individuals in the gth group, therefore $\sum_{g=1}^{G} n_g = n$. Furthermore, $\boldsymbol{X}_g^*$ is the columnwise centered matrix with vector of n_g one, that is, $\boldsymbol{X}_g^* = (\mathbf{1}_{n_g}\ \boldsymbol{X}_g^\star) = (\mathbf{1}_{n_g}\ \boldsymbol{J}_{n_g}\boldsymbol{X}_g)$, and $\boldsymbol{\beta}_g^*$ is the coefficient vector corresponding to $\boldsymbol{X}_g^*$, that is, $\boldsymbol{\beta}_g^* = (\alpha_{g0}\ \beta_{g1}\ \cdots\ \beta_{gp}) = (\alpha_{g0}\ \boldsymbol{\beta}_g')$ and $\alpha_{g0} = \beta_{g0} + 1/n_g \mathbf{1}_{n_g}' \boldsymbol{X}_g \boldsymbol{\beta}_g = \beta_{g0} + \sum_{j=1}^{p} \beta_{gj}\bar{x}_{gj}$. The separation of the G groups and $\boldsymbol{\beta}_g$ $(g = 1, \ldots, G)$ are derived as the objective function

$$
\begin{aligned}
Q &= \sum_{g=1}^{G} \|\boldsymbol{e}_g\|^2 \\
&= \sum_{g=1}^{G} \|\boldsymbol{y}_g - \boldsymbol{X}_g^* \boldsymbol{\beta}_g^*\|^2 \\
&= \sum_{g=1}^{G} \sum_{i \in S_g} \left\{ y_{gi} - \left(\sum_{j=1}^{p} \beta_{gj} x_{gij} + \beta_{g0} \right) \right\}^2
\end{aligned}
$$

is minimized.

Our estimation algorithm is as follows.

Step1: Default separation S_g $(g = 1, \ldots, G)$ is selected at random.
Step2: $\boldsymbol{\beta}_g^*$ $(g = 1, \ldots, G)$ is estimated.
Step3: For the ith $(i = 1, \ldots, n)$ individual, residual errors $r_{gi} = y_i - \sum_{j=1}^{p} \beta_{gj} x_{ij} - \beta_{g0}$ $(g = 1, \ldots, G)$ are calculated.
Step4: The ith individual is assigned to the $\arg\min_g r_{gi}$th group.
Step5: Repeat Step2 through Step5 until convergence.

5 Clusterwise linear regression model for modal multi-valued data

Our clusterwise linear regression model for modal multi-valued data is the combination of models described in Section 3 and 4 above. In this section, we use the same notation as that in Section 3; i.e., we assume that the dependent and independent modal multi-valued variables $\boldsymbol{Y}$ and $\boldsymbol{X}^\dagger$ are given. Then, the model is

$$
\boldsymbol{Y}_g = \boldsymbol{X}_g^* \boldsymbol{B}_g^* + \boldsymbol{E}_g \quad (g = 1, \ldots, G), \tag{1}
$$

where $\boldsymbol{Y}_g$, $\boldsymbol{X}_g^*$, $\boldsymbol{B}_g^*$, and $\boldsymbol{E}_g$ are the $N_g \times r$ dependent matrix, the $N_g \times (q-p+1)$ independent matrix, the $(q-p+1) \times r$ coefficient matrix, and the $N_g \times r$ error

matrix of the gth group, respectively. Furthermore, N_g is the number of concepts in the gth group.

As seen in equation (1), the regression models for all r dependent weight variables are homogeneous within the same group of individuals. One cluster structure and $\boldsymbol{B}_g$ $(g = 1, \ldots, G)$ are derived via an iterative method as the objective function

$$\begin{aligned} Q &= \sum_{g=1}^{G} \|\boldsymbol{E}_g\|^2 \\ &= \sum_{g=1}^{G} \|\boldsymbol{Y}_g - \boldsymbol{X}_g^* \boldsymbol{B}_g^*\|^2 \\ &= \sum_{g=1}^{G} \sum_{c \in S_g} \sum_{\ell=1}^{r} \left\{ y_{gc\ell} - \left(\sum_{j=1}^{p} \sum_{k=1}^{q_j - 1} \beta_{g\ell jk} x_{gcjk} + \beta_{g\ell 0} \right) \right\}^2 \end{aligned}$$

is minimized.

The estimation algorithm is as follows.

Step1: Default separation S_g $(g = 1, \ldots, G)$ is selected at random.
Step2: $\boldsymbol{B}_g^*$ $(g = 1, \ldots, G)$ is estimated.
Step3: For the cth $(c = 1, \ldots, N)$ concept, residual errors $r_{gc} = \sum_{\ell=1}^{r} r_{gc\ell} = \sum_{\ell=1}^{r} \left(y_{c\ell} - \sum_{j=1}^{p} \sum_{k=1}^{q_j - 1} \beta_{g\ell jk} x_{cjk} - \beta_{g\ell 0} \right)$ $(g = 1, \ldots, G)$ are calculated.
Step4: The cth individual is assigned to the $\arg\min_g r_{gc}$th group.
Step5: Repeat Step 2 through Step 5 until convergence.

6 Numerical example

To verify the features of our proposed method, we show the estimation results of our model using example data. The data contain one dependent modal variable and one independent modal variable. We assumed that the dependent variable could be assigned one of three possible values, $\gamma_{\boldsymbol{Y}\ell}$ $(\ell = 1, 2, 3)$, and the independent variable could be assigned one of two possible values, $\gamma_{\boldsymbol{X}k}$ $(k = 1, 2)$. Data for the estimation are dependent on weight matrix $\boldsymbol{Y} = (\boldsymbol{y}_1\ \boldsymbol{y}_2\ \boldsymbol{y}_3)$ in which each column corresponds to the relative frequency of $\gamma_{\boldsymbol{Y}\ell}$ $(\ell = 1, 2, 3)$, and independent matrix $\boldsymbol{X} = (\boldsymbol{x}_1\ \boldsymbol{x}_2)$ in which each column corresponds to the relative frequency of $\gamma_{\boldsymbol{X}k}$ $(k = 1, 2)$. The number of concepts is $N = 60$, and data are generated from three true multiple linear regression models (i.e. the black lines shown in Fig.1) with 20 concepts. Because we need only one of the dependent weight vectors, $\boldsymbol{x}_1$ is used for the estimation.

Results of the estimation process are shown in Fig.1; the vertical axis represents the weight value of each $\gamma_{\boldsymbol{Y}1}$, $\gamma_{\boldsymbol{Y}2}$, and $\gamma_{\boldsymbol{Y}3}$, and the horizontal axis represents that of $\gamma_{\boldsymbol{X}1}$. The colors represent estimated clusters, whereas the tick marks represent true clusters.

From Fig.1, the estimation almost properly identifies the cluster structure and linear regression lines. We observe in the left plot that the red and blue

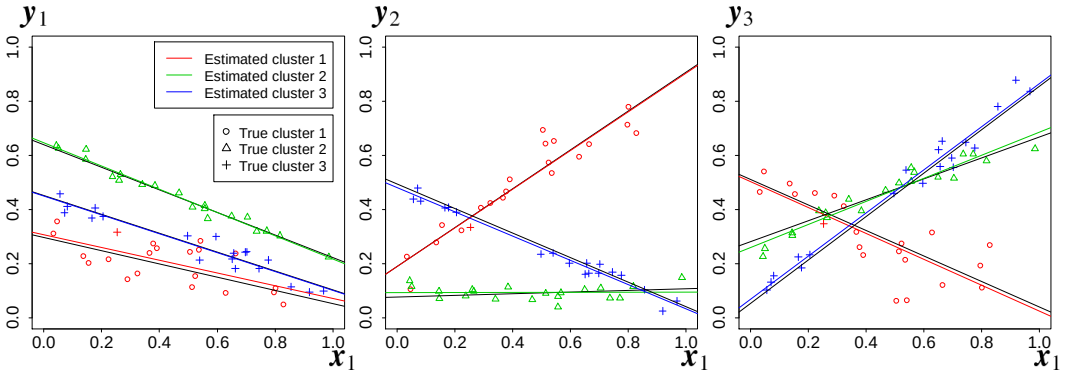


Fig. 1. Regression lines for example data

lines are similar; however, because of the structure of the other two dependent variables, estimation is completed for vector $\boldsymbol{y}_1$. If vector $\boldsymbol{y}_1$ is only used for estimating the cluster structure, it will fail to identify the true cluster structure. In other words, using our proposed method, multiple variables are treated as a single symbolic variable.

7 Conclusion

In this paper, we proposed a clusterwise linear regression model for modal multi-valued data. Applying our model to example numerical data, we showed that, in our model, the dependent symbolic variables are treated as a single variable although the symbolic variable is represented as multiple variables. Future work includes empirical analysis using our proposed model.

References

1. Billard, L., Diday, E.: Symbolic Data Analysis: Conceptual Statistics and Data Mining. Wiley series in computational statistics. Wiley, chichester (2006)
2. Helmuth, Späth.: Algorithm 39 Clusterwise linear regression. Computing. 22, 4, 367–373 (1979)
3. Johnson, R. A., Wichern, D. W.: Applied Multivariate Statistical Analysis. 6th ed. Pearson Education. 360–428 (2007)

Bayesian generalized lienar mixed models with general random effects covariance matrix

Keunbaik Lee[1] and Jae Keun Yoo[2]

[1] Sungkyunkwan University, Seoul, 110-745, Korea
keunbaik@skku.edu
[2] Ewha Womans University, Seoul, 120-750, Republic

Abstract. Random effects in generalized linear mixed models (GLMM) are used to explain the serial correlation of the longitudinal categorical data. The random effects covariance matrix is assumed to be constant over subjects and to be restricted such as AR(1) structure because of the positive definiteness of the matrix. However, these assumptions are too strong and can result in biased estimates of the fixed effects. In this paper we propose a Bayesian GLMM for long series of longitudinal binary data with regression models for parameters of the random effects covariance matrix. The covariance matrix is factored into generalized moving average parameters and innovation variances using the moving average Cholesky decomposition.

1 Bayesian Moving Average Regression Model for Random Effects Covariance Matrix

We now present Bayesian GLMMs with the random effects covariance matrix using the moving average Cholesky decomposition.

1.1 Moving Average Cholesky Decomposition of Random Effects Covariance Matrix

Let Y_{it} be the binary response for subject i $(i = 1, \cdots, N)$ at time t $(t = 1, \cdots, T)$ and let x_{it} be the corresponding vector of covariates. We assume that each Y_{it} is conditionally independent given random effects b_{it}, the responses for different subjects are independent, and the GLMM is given by

$$\text{logit} p_{it}(b_{it}) = x_{it}^T \beta + b_{it} \tag{1}$$

where $p_{it}(b_{it}) = P(Y_{it} = 1|b_{it})$ and β is the $p \times 1$ vector of regression coefficients. We assume that

$$b_i = (b_{i1}, \cdots, b_{in_i})^T \sim N(0, \Sigma_i) \tag{2}$$

where Σ_i is the random effects covariance matrix and b_i is a vector of random effects values for subject i.

To model Σ_i in (2), we assume that, for $t = 2, \cdots, n_i$,

$$b_{it} = \sum_{j=1}^{t-1} l_{i,tj} e_{ij} + e_{it}, \tag{3}$$

where $l_{i,tj}$ are generalized moving average (GMA) coefficients, and $e_i \sim N(0, D_i)$ for $e_i = (e_{i1}, \cdots, e_{in_i})^T$. Note that $b_{i1} = e_{i1}$. Then we rewrite (3) in matrix form as

$$b_i = L_i e_i$$

where L_i is a unique lower triangular having ones on its diagonal and $l_{i,tj}$ at its (t, j)th position for $j < t$. Then the covariance matrix for b_i is

$$\Sigma_i = L_i D_i L_i^T.$$

The constraint on theses parameters for Σ_i to be positive definite is that the IV need to be positive.

The parameters, GMA and IV, can be modeled using time and/or subject-specific covariate vectors $z_{i,tj}$ and $h_{i,t}$ by setting

$$l_{i,tj} = z_{i,tj}^T \eta, \quad \log(\sigma_{i,t}^2) = h_{i,t}^T \lambda, \tag{4}$$

where η and λ are $a \times 1$ and $b \times 1$ vectors of unknown dependence and variance parameters, respectively. Thus, this approach reduces the number of parameters for the covariance matrix using regression models for the IV and GMA parameters.

1.2 Prior Distributions

Using the MA Cholesky decomposition, we form priors that are diffused. The diffused prior distributions for β, η, and λ are given by

$$\beta \sim N(0, \sigma_\beta^2 I), \tag{5}$$

$$\eta \sim N(0, \sigma_\eta^2 I), \tag{6}$$

$$\lambda \sim N(0, \sigma_\lambda^2 I), \tag{7}$$

where σ_β^2, σ_γ^2, and σ_λ^2 are large values (for example, 100).

1.3 Bayesian Analysis via MCMC

We derive the likelihood function for the GLMM expressions (1) and (2). We let $\omega = (\beta^T, \eta^T, \lambda^T)$. The joint distribution of sample and the random effects in (1) and (2) is given by

$$L(\omega; y, b) = \prod_{i=1}^{N} \prod_{t=1}^{n_i} p_{it}(b_{it}, \beta)^{y_{it}} \left(1 - p_{it}(b_{it}, \beta)\right)^{1-y_{it}} \phi(b_i | \eta, \lambda) \tag{8}$$

where $p_{it}(b_{it}, \beta) = P(Y_{it} = 1|b_{it}, \beta, x_{it})$ and $\phi(b_i; \eta, \lambda)$ is a multivariate normal density with mean vector 0 and covariance matrix Σ_i which is simplified as

$$\phi(b_i|\eta, \lambda) = (2\pi)^{-\frac{n_i}{2}} \prod_{t=1}^{n_i} \left(\sigma_{i,t}^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} b_i^T L_i^{-T} D_i^{-1} L_i^{-1} b_i\right).$$

From the distribution (8) and prior distributions (5)-(7), the joint distribution is given by

$$P(y, b, \beta, \eta, \lambda) \propto \phi(\beta; \sigma_\beta^2)\phi(\eta; \sigma_\eta^2)\phi(\lambda; \sigma_\lambda^2) \left[\prod_{i=1}^{N} \left\{\prod_{t=1}^{n_i} (p_{it}^c(b_{it}, \beta))^{y_{it}} (1 - p_{it}^c(b_{it}, \beta))^{1-y_{it}}\right\} \phi(b_i|\eta, \lambda)\right].$$

Full conditional posterior distributions are required to implement the MCMC algorithm (Gelman et al., 2004). Since all full conditionals are intractable analytically and not easily generated from, we have to construct suitable proposals for a Metropolis-Hastings step (Hastings, 1970; Gamerman, 1997). In practice, Gibbs sampling is implemented using WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml). The MCMC algorithm simulates direct draws from the above full conditionals iteratively until convergence is achieved.

In Bayesian modeling, there are several model selection criteria such as posterior predictive loss, Bayes factor and deviance infromation criterion (DIC) (Daniels and Hogan, 2008). In this paper, we use DIC for Bayesian model comparison (Spiegelhalter et al., 2002). Let θ be a vector of all parameters and let $L(\theta|y)$ be the likelihood of $y = (y_1, \cdots, y_N)^T$. Then the DIC is given as

$$DIC = D(\theta) + p_D$$

where $D(\theta) = -2 \log L(\theta|y)$ measures the deviance evaluated at the posterior mean of the parameters (goodness-of-fit) and $p_D = \overline{Dev(\theta)} - D(\widetilde{\theta})$ measures the effective dimension (penalty of complexity) of model. The DIC is then defined analogously to AIC and models with smaller DIC should be preferred to models with larger DIC. The advantages of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the MCMC samples. In contrast, AIC and BIC require calculating the likelihood at its maximum values, which are not easily available from the MCMC simulation.

2 Acknowledgements

We would like to thank Dr. Myung-Ju Ahn of Samsung Medical Center in South Korea for providing the data and for their help in data collection and clarifying some issues with data. For Keunbaik Lee, this work was supported by Basic Science Research Program through the National Research Foundation

of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A1004002). For Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (KRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A1040077).

References

1. elman, A., Carlin, J., Stern, H., and Rubin, D.B.: *Bayesian Data Analysis.* 2nd Edition, Chapman & Hall, London, UK (2004)
2. amerman, D.: *Markov Chanin Monte Carlo: Stochastic Simulation for Bayesian Inference.* Chapman & Hall: London (1997)
3. astings, W.K.: Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* **57**, 97–109 (1970)
4. aniels, M.J. and Hogan, J.W.: *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* Chapman & Hall/CRC (2008)
5. piegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A.: Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B: Statistical Methodology,* 64, 583–616 (2002)

Nonlinear surface regression with sufficient dimension reduction

Takuma Yoshida[1]

Kagoshima University, Kagoshima, Kagoshima, Korimoto 1-21-24, Japan,
yoshida@sci.kagoshima-u.ac.jp ,

Abstract. We consider the nonlinear regression analysis with a scalar response and multiple predictors. An unknown regression function is approximated by radial basis function models. It is known that ordinal nonlinear estimation leads to overfitting. The purpose is to construct a smooth estimator for general robust regression. The proposed method is conducted by a two step procedure. First, the sufficient dimension reduction methods are applied to the response and radial basis functions for transforming the large number of radial bases to a small number of linear combinations of the radial bases without loss of information. In the second step, a multiple linear regression model between a response and the transformed radial bases is assumed and the ordinal is applied. Thus the final estimator is also obtained as a linear combination of radial bases. As the result, the proposed method leads to that the estimator has the curve with goodness and smoothness. Our method can be widely used in several regression problems such as quantile regression, robust regression and binary regression.

1 Introduction

1.1 Nonlinear regression model

Nonlinear regression has been developed to understand the relationship between an response Y and a d-dimensional predictor $\boldsymbol{x} = (x_1, \cdots, x_d)^T \in \mathbb{R}^d$. The general multiple regression model is defined as

$$Y = f(\boldsymbol{x}) + \varepsilon,$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is an unknown regression function and, in many cases, ε has symmetrical density. The function f is usually defined as the minimizer of $E[\rho(Y - a)|\boldsymbol{X} = \boldsymbol{x}]$, where $\rho(u)$ is convex and has a unique minimum at $u = 0$. When $\rho(u) = u^2$ is used, we obtain $f(\boldsymbol{x}) = E[Y|\boldsymbol{x}]$, and when $\rho(u) = |u|$ is used, $f(\boldsymbol{x})$ becomes the conditional median of Y given $\boldsymbol{x}$. These are often called the mean regression and the median regression, respectively The purpose of regression is often to estimate f using the sample $\{(Y_i, \boldsymbol{x}_i) : i = 1, \cdots, n\}$ of $(Y, \boldsymbol{x})$. In nonlinear regression, one of most popular approaches, the regression function f is assumed as radial basis function models

$$f(\boldsymbol{x}) = \sum_{k=1}^{K} c_k \phi_k(\boldsymbol{x}) = \boldsymbol{c}^T \boldsymbol{\phi}(\boldsymbol{x}),$$

where $\phi_1, \cdots, \phi_K$ are known radial basis functions; $c_1, \cdots, c_K$ are unknown parameters, where $\boldsymbol{c} = (c_1 \cdots c_K)^T$ and $\boldsymbol{\phi}(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}) \ \cdots \ \phi_K(\boldsymbol{x}))^T$. We estimate c_k instead of f directly. A radial basis function is defined in $\mathbb{R}^d$ and has a value that depends only on the distance between the predictor $\boldsymbol{x}$ and the knot $\boldsymbol{\kappa}_k$ and hence ϕ_k can be written in the form $\phi_k(\boldsymbol{x}) = \phi(||\boldsymbol{x} - \boldsymbol{\kappa}_k||)$. In generally, the coefficient $\boldsymbol{c}$ can be estimate as the minimizer of $\sum_{i=1}^n \rho(Y_i - f(\boldsymbol{x}_i))$. Although it is known that the above estimation may lead to wiggly estimator when K is large. Hence we should estimate $\boldsymbol{c}$ in order to avoid overfitting. As one of majority methods, the penalized method is useful. However in quantile and robust regression, te computation becomes difficult and the smoothing parameter selection method has not yet been developed. Therefore we suggest the alternative method to obtain smooth curve. In this conference, we use the method of sufficient dimension reductions (SDR) to constrain $\boldsymbol{c}$ but not penalization.

1.2 Sufficient dimension reduction

We briefly describe the usage of the sufficient dimension reduction methods. We consider a continuous one dimensional response Y and a continuous d-dimensional predictor $\boldsymbol{X} = (X_1, \cdots, X_d)$. If there exists a $J \times d (J \leq d)$ matrix C such that

$$Y \perp\!\!\!\perp \boldsymbol{X} | C\boldsymbol{X},$$

then C is called an SDR subspace. This means that the conditional distribution of $Y|\boldsymbol{X}$ equals that of $Y|C\boldsymbol{X}$. If C exists, then the dimension of the predictor can be reduced from d to J. In other words, it is sufficient to consider the univariate regression when $J = 1$. The SDR method estimates C or subspaces of C. For typical SDR methods, Li [6] proposed the sliced inverse regression (SIR), Cook and Weisberg [4] suggested the sliced average variance estimation (SAVE) and Li [7] studied the principal Hessian directions (PHD). SDR methods have been developed in nonlinear regression. The nonlinear SDR methods consider

$$Y \perp\!\!\!\perp \boldsymbol{\phi}(\boldsymbol{X}) | B\boldsymbol{\phi}(\boldsymbol{X}), \tag{1}$$

where $\boldsymbol{\phi} : \mathbb{R}^d \to \mathbb{R}^K$ is the nonlinear K-dimensional function and B is the $J \times K (K \leq J)$ matrix. The purpose is to estimate B.

The sufficient dimension reduction methods to estimate CDR space can be formulated as the eigenvalue problem:

$$M\boldsymbol{\beta}_j = \lambda_j G\boldsymbol{\beta}_j, \quad j = 1, \cdots, J, \tag{2}$$

where the K square matrix M is symmetric and nonnegative definite, the K square matrix G is positive definite, vectors $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_K$ are eigenvectors satisfying $\boldsymbol{\beta}_i^T G\boldsymbol{\beta}_j = 1$ for $i = j$, and 0 for otherwise, and $\lambda_1 \geq \cdots \geq \lambda_K \geq 0$ are corresponding eigenvalues. Then the matrix B is obtained as $B = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J]$.

Specifying M and G, several dimension reduction methods such as SIR, SAVE, PHD and others can be obtained. Let $\Sigma = E[\{\boldsymbol{\phi}(X) - E[\phi(X)]\}^2]$ and $\boldsymbol{Z} = \Sigma^{-1}(\boldsymbol{\phi}(X) - E[\boldsymbol{\phi}(X)])$ be a covariance matrix of $\boldsymbol{\phi}(X)$ and standardized version of $\boldsymbol{\phi}(X)$. We then obtain

SIR: $M = Cov(E[\boldsymbol{\phi}(X) - E[\boldsymbol{\phi}(X)]|y])$ and $G = \Sigma$.
SAVE: $M = \Sigma^{1/2}E[\{I - Cov(\boldsymbol{Z}|y)\}^2]\Sigma^{1/2}$ and $G = \Sigma$.
PHD: $M = \Sigma^{1/2}\Sigma_{yz}^2\Sigma^{1/2}$ and $G = \Sigma$, where $\Sigma_{yz} = E[\{Y - E[Y]\}ZZ^T]$.

Actually, since $\boldsymbol{\beta}_j$ is unknown, we should estimate $\boldsymbol{\beta}_j$ using sample version of above M and G. The estimation algorithm have been clarified in Li [6], Cook and Wiesberg [4], Li [7] and Cook [2]. Many useful SDR methods other than the above have been proposed by many authors. For example, the iterative PHD (Cook and Li, [3]), contour regression (Li et al., [8]), directional regression (Li and Wang, [**?**]), sparse SDR (Li, [9]), sparse SIR (Zhu et al., [11]) and coordinate independent sparse SDR (Chen et al., [1]) and so on. However we focus only on SIR, SAVE and PHD in this conference since the selection of SDR method is very difficult (although the heuristic search can be done).

1.3 Organization

We estimate the regression function f using nonlinear SDR methods. If B satisfying (1) exists, then the conditional distribution of $Y|\boldsymbol{\phi}(\boldsymbol{X})$ equals that of $Y|B\boldsymbol{\phi}(\boldsymbol{X})$. In other words, the regression function $f(\boldsymbol{x})$ can be expressed as $f(\boldsymbol{x}) = \boldsymbol{w}^T B\boldsymbol{\phi}(\boldsymbol{x})$, where $\boldsymbol{w} \in \mathbb{R}^J$ is an unknown vector. First we estimate B using the nonlinear SDR method. We next estimate $\boldsymbol{w}$ in the models of multiple linear regression. Thus the final estimator of f is established by a two step procedure. The SDR method constrains the behavior of the coefficients of the radial bases similarly to penalization. As such, the proposed two step estimation can avoid overfitting the estimate even if K is large. The efficiency of the proposed estimator is confirmed by numerical study in this paper.

This paper is organized as follows. Section 2 describes the nonlinear regression setting and Section 3 provides the two step estimation using the SDR method. In Section 4, we discuss the proposed method for binary response. Section 5 address a numerical simulation and an application are addressed. Conclusion and further study are provided in Section 6.

2 Nonlinear surface regression

For the dataset $\{(Y_i, \boldsymbol{X}_i) : i = 1, \cdots, n\}$ with an one dimensional response Y_i and a d-dimensional predictor $\boldsymbol{X}_i$, consider the nonlinear radial basis regression models:

$$Y_i = \sum_{k=1}^{K} c_k\phi_k(\boldsymbol{x}) + \varepsilon_i,$$

where $c_1, \cdots, c_K$ are unknown parameters, $\phi_1, \cdots, \phi_K$ are known radial basis functions and ε_i's are *i.i.d.* random error having symmetry density. In this paper, we use the thin plate spline basis for $\phi_k(\boldsymbol{x})$. For $k = 1, \cdots, K$, the thin plate spline is of the form

$$\phi_k(\boldsymbol{x}) = \phi(||\boldsymbol{x} - \boldsymbol{\kappa}_k||) = \begin{cases} ||\boldsymbol{x} - \boldsymbol{\kappa}_k||^{2m-d}, & d \text{ odd}, \\ ||\boldsymbol{x} - \boldsymbol{\kappa}_k||^{2m-d} \log ||\boldsymbol{x} - \boldsymbol{\kappa}_k||, & d \text{ even}, \end{cases}$$

where $\boldsymbol{\kappa}_1, \cdots, \boldsymbol{\kappa}_K$ are the locations of fixed d-dimensional knots and m is an integer satisfying $2m - d > 0$ that controls the smoothness of ϕ. Then we estimate $\boldsymbol{c} = (c_1, \cdots, c_K)^T$ and construct the estimator of f as for $\boldsymbol{x} \in \mathbb{R}^d$,

$$\hat{f}(\boldsymbol{x}) = \sum_{k=1}^{K} \hat{c}_k \phi_k(\boldsymbol{x}) = \hat{\boldsymbol{c}}^T \boldsymbol{\phi}(\boldsymbol{x}),$$

where $\hat{\boldsymbol{c}} = (\hat{c}_1, \cdots, \hat{c}_K)^T$ is the estimator of $\boldsymbol{c}$ and $\boldsymbol{\phi}(\boldsymbol{X}) = (\phi_1(\boldsymbol{X}), \cdots, \phi_K(\boldsymbol{X}))^T$. In general, the estimator $\hat{\boldsymbol{c}} = (\hat{c}_1, \cdots, \hat{c}_K)^T$ of $\boldsymbol{c}$ is obtained as the minimizer of

$$\sum_{i=1}^{n} \rho(Y_i - \boldsymbol{c}^T \boldsymbol{\phi}(\boldsymbol{x}_i)), \tag{3}$$

where ρ is the convex function on $\mathbb{R}$ with the unique minimizer at the origin. When $\rho(u) = u^2$, $f(x)$ is the conditional mean of Y given $\boldsymbol{X} = \boldsymbol{x}$ and $\hat{f}(x)$ is the its estimator. When $\rho(u) = |u|$, then $\hat{f}(x)$ is the estimator of the conditional median function. Thus specifying ρ, we capture several types of regression function and we can obtain the robust estimator of f.

In general, the estimator $\hat{f}(x)$ will have a wiggly curve if K is large. So we should improve the estimation method in order to obtain a smooth estimator. Many authors used the penalization method to avoid the overfitting curve in mean regression. But in median or other robust regression, the algorithm of penalization is complicated and the smoothing parameter selection methods have been developed. Therefore we suggest the new estimation method without penalization in next section.

3 Estimation via sufficient dimension reduction method

To avoid such overfitting, we estimate $\boldsymbol{c}$ by the two step procedure using the SDR method. We define the function $F : \mathbb{R}^{J+1} \to \mathbb{R}$ for $1 \leq J \leq K$, $z_1, \cdots, z_n, \varepsilon \in \mathbb{R}$ as

$$F(z1, z2, \cdots, z_J, \varepsilon) = \sum_{j=1}^{J} w_j z_j + \varepsilon,$$

where $w_1, \cdots, w_J \in \mathbb{R}$ are unknown parameters. Then the following nonlinear model

$$Y_i = F(\boldsymbol{\beta}_1^T \boldsymbol{\phi}(\boldsymbol{x}_i), \cdots, \boldsymbol{\beta}_J^T \boldsymbol{\phi}(\boldsymbol{x}_i), \varepsilon_i) \tag{4}$$

can be considered, where $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$ are unknown vectors of dimension K. This model implies that Y depends on X only through J linear combinations of the radial basis vector. For $j = 1, \cdots, J$, the estimator $\hat{\boldsymbol{\beta}}_j$ of $\boldsymbol{\beta}_j$ is obtained using an SDR method such as SIR, SAVE, PHD and so on. Next we substitute $\hat{\boldsymbol{\beta}}_j (j = 1, \cdots, J)$ into (4) and solve

$$\min_{w_1, \cdots, w_J} \sum_{i=1}^{n} \rho \left(Y_i - \sum_{j=1}^{J} w_j \hat{\boldsymbol{\beta}}_j^T \boldsymbol{\phi}(X_i) \right) \tag{5}$$

with respect to $\boldsymbol{w} = (w_1, \cdots, w_J)^T$. Finally we estimate f by

$$\hat{f}(\boldsymbol{x}) = \sum_{j=1}^{J} \hat{w}_j \hat{\boldsymbol{\beta}}_j^T \boldsymbol{\phi}(\boldsymbol{x}) = \hat{\boldsymbol{w}}^T \hat{B} \boldsymbol{\phi}(\boldsymbol{x})$$

where $\hat{\boldsymbol{w}} = (\hat{w}_1, \cdots, \hat{w}_J)^T$ is the minimizer of (5) and $\hat{B} = [\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_J]^T$. In other words, the estimator of $\boldsymbol{c}$ is constructed by $\hat{\boldsymbol{c}} = \hat{B}^T \hat{\boldsymbol{w}}$.

4 Extension to binary response

The proposed methods can be applied to the binary response data. We consider the data $\{(Y_i, \boldsymbol{X}_i) : i = 1, \cdots, n\}$, where $Y_i \in \{0, 1\}$ and $\boldsymbol{X}_i$ is similar to that given in Section 2. Then we model the relationship between response and predictor as

$$P(Y_i = 1 | \boldsymbol{X}_i = \boldsymbol{x}_i) = f(\boldsymbol{x}_i), \quad P(Y_i = 0 | \boldsymbol{X}_i = \boldsymbol{x}_i) = 1 - f(\boldsymbol{x}_i).$$

Since $Y_1, \cdots, Y_n$ are independently distributed as the Bernoulli distribution, the density g of y_i can be written as

$$g(y_i | \boldsymbol{x}_i) = f(\boldsymbol{x})^{y_i} \{1 - f(\boldsymbol{x})\}^{1 - y_i},$$

Here we assume that

$$f(\boldsymbol{x}_i) = \frac{\exp[\boldsymbol{c}^T \boldsymbol{\phi}(\boldsymbol{x}_i)]}{1 + \exp[\boldsymbol{c}^T \boldsymbol{\phi}(\boldsymbol{x}_i)]},$$

that is, the logistic regression is considered. To estimate $\boldsymbol{c}$, the maximum likelihood using Fisher-scoring is useful in the logistic regression. The Fisher-scoring iteration algorithm is detailed in many books of statistical analysis and hence we omit it in this paper.

Cook and Lee (1999) have proposed the SDR method for binary response. So we estimate $\boldsymbol{c}$ using SDR method for binary data and Fisher-scoring. We briefly introduce SIR, SAVE and PHD for binary data. Let $s_1 \geq s_J$ denote the eigenvalues of a matrix M and let $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_J$ denote the corresponding eigenvectors ($J \leq K$). By specifying M, SIR, SAVE and PHD can be obtained. The estimator $\hat{s}_j$ and $\hat{\boldsymbol{\beta}}_j$ of s_j and $\boldsymbol{\beta}_j$ are obtained by sample version of M. Let $\boldsymbol{\mu}_j = E[\boldsymbol{Z}|y = j]$ and $\Sigma_j = V[\boldsymbol{Z}|y = j]$ for $j = 0, 1$. Further, $\pi = P(Y = 1)$, $\boldsymbol{\mu}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ and $\Delta = \Sigma_1 - \Sigma_0$. When $M = \boldsymbol{\mu}\boldsymbol{\mu}^T$ is used, SIR is obtained. We note that since $\{\text{rank}\}(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}}) = 1$, the reduced dimension is $J = 1$. When $M = (1-\pi)(I - \Sigma_0)(I - \Sigma_0)^T + \pi(I - \Sigma_1)(I - \Sigma_1)^T$, where I is the identity matrix, we obtain SAVE. The PHD for binary data is defined as $M = \pi(1 - \pi)\{\boldsymbol{\Delta} + (1 - 2\pi)\boldsymbol{\mu}\boldsymbol{\mu}^T\}$.

In the first step estimation, using SDR method, the dimension of the radial basis functions $\{\phi_1(\boldsymbol{x}), \cdots, \phi_K(\boldsymbol{x})\}$ is reduced to row dimensional linear combination $\{\hat{\boldsymbol{\beta}}_1^T\boldsymbol{\phi}(\boldsymbol{x}), \cdots, \hat{\boldsymbol{\beta}}_J^T\boldsymbol{\phi}(\boldsymbol{x})\}$ for $J \leq K$. The second step, we consider the logistic regression using

$$f(\boldsymbol{x}_i) = \frac{\exp[\boldsymbol{w}^T\hat{B}\boldsymbol{\phi}(\boldsymbol{x}_i)]}{1 + \exp[\boldsymbol{w}^T\hat{B}\boldsymbol{\phi}(\boldsymbol{x}_i)]},$$

where $\hat{B}$ is same that given in Section 3 and $\boldsymbol{w} = (w_1, \cdots, w_J)^T$ and estimate $\boldsymbol{w}$ by maximum likelihood method. As the final estimator of $f(\boldsymbol{x})$, we obtain

$$\hat{f}(\boldsymbol{x}) = \frac{\exp[\hat{\boldsymbol{w}}^T\hat{B}\boldsymbol{\phi}(\boldsymbol{x})]}{1 + \exp[\hat{\boldsymbol{w}}^T\hat{B}\boldsymbol{\phi}(\boldsymbol{x})]},$$

where $\hat{w}$ is the maximum likelihood estimator of $\boldsymbol{w}$.

5 Numerical study

We will report the application to the real data in conference. But we omit in this paper.

6 Conclusion

This paper suggested nonlinear multiple regression using the sufficient dimension reduction method in the context of M-estimation. The proposed method was conducted via a two step algorithm. First we applied the sufficient dimension reduction method to the response and the radial basis function. We next estimate the regression function under multiple linear regression models. The first dimension reduction method controls the relationship between a response

and a radial basis function to avoid overfitting. In the second step, the model is regarded as a linear combination of the radial bases transformed in the first step. As the further study, it would be beneficial to extend the works in this paper to the partial linear model with the indicator predictor. We may also consider various extensions of the model concerned with extreme value response and so on.

Acknowledgements

The research of the author was partially supported by KAKENHI 80707141.

References

1. Chen,X., Zhou,C. and Cook,R.D.: Coordinate-Independent sparse sufficient dimension reduction and variable selection. Ann. Statist. 38, 3696-3723 (2010).
2. Cook,R.D.: Principle hessian directions revisited. J. Amer. Statist. Assoc. 93, 84-100 (1998).
3. Cook,R.D. and Li,B.: Dimension reduction for the conditional mean in regression. Ann. Statist. 30, 455-474 (2002).
4. Cook,R.D. and Weisberg,S.: Discussion of 'Sliced inverse regression for dimension reduction" by K.C.Li. J. Amer. Statist. Assoc. 86, 328-332 (1991).
5. Cook,D. and Lee,H.: Dimension reduction in binary response regression. J. Amer. Statist. Assoc. 94, 1187-1200 (1999).
6. Li,K.C.: Sliced inverse regression for dimension reduction. J. Amer. Statist. Assoc. 102, 997-1008 (1991).
7. Li,K.C.: On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. J. Amer. Statist. Assoc. 87, 1025-1039 (1992).
8. Li,B., Zha,H. and Chiaromonte,F.: Contour regression: A general approach to dimension reduction. Ann. Statist. 33, 1580-1616 (2005).
9. Li,L.: Sparse sufficient dimension reduction. Biometrika. 94, 603-611 (2007).
10. Li,Y. and Zhu,L.X.: Asymptotics for sliced average variance estimation. Ann. Statist. 35, 41-69 (2007).
11. Zhu,L., Miao,B. and Peng,H.: On sliced inverse regression with high-dimensional covariates. J. Amer. Statist. Assoc. 101, 630-643 (2006).

Efficient solution for nonlinear dynamic estimation problem with model-reality differences

Sie Long Kek[1], Kim Gaik Tay[2] and Kuan Chin Chua[3]

[1] Department of Mathematics and Statistics, Universiti Tun Hussein Onn Malaysia,
86400 Parit Raja, MALAYSIA
slkek@uthm.edu.my
[2] Department of Communication Engineering, Universiti Tun Hussein Onn Malaysia
86400 Parit Raja, MALAYSIA
[3] Department of Mathematical and Actuarial Sciences, Universiti Tunku Abdul Rahman
50744 Kuala Lumpur, MALAYSIA

Abstract

In this paper, a computational approach for solving the nonlinear dynamic estimation problem is proposed. Our aim is to estimate the nonlinear state dynamics. In our approach, the linear expectation model, which is added with the adjusted parameters, is introduced. On this basis, the differences between the original system and the model used can be measured repeatedly. Since the output is measureable from the original problem, it is fed back into the model, in turn, updates the estimation solution of the model used iteratively. As the convergence achieved, the model solution converges to the true solution of the original problem, in spite of model-reality differences. For illustration, an example is studied and the solution shows the efficiency of the approach proposed.

Keywords: nonlinear dynamic estimation, iterative solution, model-reality differences, adjusted parameters, output measurement

1 Introduction

Estimating the state dynamics accurately from a nonlinear dynamical system that is disturbed by Gaussian white noise sequences is a challenging task. This estimation is due to the fluctuation behavior appeared in the dynamic system that gives an unpredictable response, and makes the dynamic system even more complex. In this point of view, the Kalman filtering theory, which consists of the measurement and time updates, is proposed to give the optimal state estimate for the linear stochastic dynamic systems [1, 2, 3].

The idea of the Kalman filtering theory is then extended to nonlinear dynamical systems since most of engineering problems are nonlinear in nature, see for examples [4, 5, 6]. In implementation of the extended Kalman filter (EKF), the Jacobian matrices derived on the state and the measurement output equations are evaluated with the current predicted states. This linearization

would not give the optimal state estimate and the divergence could be happened towards the wrong estimated solution [7, 8].

To improve the EKF, the unscented Kalman filter (UKF) is investigated [9]. In such study, the probability density is approximated by a deterministic sampling of points using the unscented transformation [10, 11]. The UKF is more robust and more accurate than the EKF for the estimation errors. However, the UKF does not perform well for the bad initial state and its robustness is less than the optimization based state estimators, for instance, the moving horizon estimator [12]. Practically, state estimation with the Kalman filtering theory has been widely applied in engineering and sciences, which covers target tracking [13], robotic manipulators [14], reservoir modeling [15], biomedical applications [16], sensor data [17], and control systems with model-reality differences [18, 19].

In this paper, we propose an efficient computation approach, which is based on the association of the Kalman filtering theory and the principle of model-reality differences, for solving nonlinear dynamic estimation problem of stochastic system. In our approach, the adjusted parameters are introduced into the linear dynamic system, both for state and output equations. During the computation procedure, the output, which is measured from the real plant, is fed back into the model used, in turn, updates the model trajectory iteratively. In this way, the differences between the real plant and the model used are calculated at each iteration step. Consequently, the optimal solution of the model used approaches to the true optimal solution of the original estimation problem in spite of model-reality differences. On this basis, an iterative algorithm is then established for the estimation problem of nonlinear stochastic dynamical systems.

The rest of the paper is organized as follows. In Section 2, the estimation problem of a nonlinear stochastic dynamical system is described. For simplicity, a linear model-based estimation problem, which is added with the adjustable parameters, is formulated. In Section 3, an expanded estimation problem, which takes into account the differences between the real system and the model used, is introduced. The resulting iterative algorithm that is based on the Kalman filtering theory and the principle of model-reality differences is then derived for solving the nonlinear dynamics estimation problem. In Section 4, an illustrative example is studied for the efficiency. Finally, some concluding remarks are made.

2 Problem Statement

Consider a general class of difference equations given below:

$$x(k+1) = f(x(k),k) + G\omega(k) \tag{1a}$$
$$y(k) = h(x(k),k) + \eta(k) \tag{1b}$$

where $x(k) \in \Re^n$, $k = 0,1,...,N$, and $y(k) \in \Re^p$, $k = 0,1,...,N$, are, respectively, the state sequence and the output sequence. $\omega(k) \in \Re^q$, $k = 0,1,...,N-1$, and $\eta(k) \in \Re^p$, $k = 0,1,...,N$, are stationary Gaussian white noise sequences with zero mean and their covariance matrices are, respectively, given by $Q_\omega \in \Re^{q\times q}$ and $R_\eta \in \Re^{p\times p}$, which are positive definite matrices. $G \in \Re^{n\times q}$ is the process coefficient matrix, $f : \Re^n \times \Re \to \Re^n$ represents the plant dynamics and $h : \Re^n \times \Re \to \Re^p$ is the output measurement channel.

The initial state is

$$x(0) = x_0$$

where $x_0 \in \Re^n$ is a random vector with mean and covariance are, respectively, given by

$$E[x_0] = \bar{x}_0 \text{ and } E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^{\mathrm{T}}] = M_0 .$$

Here, $M_0 \in \Re^{n\times n}$ is a positive definite matrix. It is assumed that initial state, process noise and measurement noise are statistically independent.

Suppose the state mean propagation is given by

$$\bar{x}(k+1) = f(\bar{x}(k),k), \qquad \bar{x}(0) = \bar{x}_0 \tag{2a}$$
$$\bar{y}(k) = h(\bar{x}(k),k) \tag{2b}$$

where $\bar{x}(k) \in \Re^n$, $k = 0,1,...,N$, and $\bar{y}(k) \in \Re^p$, $k = 0,1,...,N$, are, respectively, the expected state sequence and the expected output sequence. Then, the aim is to find a sequence of the optimal state estimate $\hat{x}(k) \in \Re^n$, $k = 0,1,...,N$, such that the following weighted least squares error (WLSE) is minimized,

$$J_{mse}(x) = \frac{1}{2}\sum_{k=0}^{N}((x(k) - \bar{x}(k))^{\mathrm{T}}(M_x(k))^{-1}(x(k) - \bar{x}(k)) + (y(k) - \bar{y}(k))^{\mathrm{T}}(M_y(k))^{-1}(y(k) - \bar{y}(k))) \tag{3}$$

where $M_x(k) \in \Re^{n\times n}$, $k = 0,1,...,N$, and $M_y(k) \in \Re^{p\times p}$, $k = 0,1,...,N$, are, respectively, the state error covariance matrix and the output error covariance matrix. It is assumed that all functions in (1), (2) and (3) are continuously differentiable with respect to their respective arguments.

This problem is regarded as the nonlinear dynamic estimation problem, and is referred to as Problem (P). Since the exact state trajectory of Problem (P) is impossible to be obtained, and solving Problem (P) by using the nonlinear filtering theory is computationally demanding. In view of these, a linear model, which is referred to as Problem (M), is simplified from Problem (P) as follows:

$$\min_{x(k)} J_{mse}(x) = \frac{1}{2}\sum_{k=0}^{N}((x(k)-\bar{x}(k))^{\mathrm{T}}(M_x(k))^{-1}(x(k)-\bar{x}(k)) + (y(k)-\bar{y}(k))^{\mathrm{T}}(M_y(k))^{-1}(y(k)-\bar{y}(k)))$$

subject to (4)

$$\bar{x}(k+1) = A\bar{x}(k) + \alpha_1(k), \qquad \bar{x}(0) = \bar{x}_0$$
$$\bar{y}(k) = C\bar{x}(k) + \alpha_2(k)$$

where $\alpha_1(k) \in \Re^n$, $k = 0,1,...,N-1$, and $\alpha_2(k) \in \Re^p$, $k = 0,1,...,N$, are the adjusted parameters, $A \in \Re^{n\times n}$ and $C \in \Re^{p\times n}$ are, respectively, the state transition matrix and the output coefficient matrix. Note that both of these matrices can be obtained from the linearization of the plant dynamics and the measurement channel, respectively, at the known initial state.

Because of the different structure between these problems, only solving Problem (M) will not give the optimal solution of Problem (P). However, with adding the adjusted parameters into the model used, the differences between the original system and the model used can be calculated repeatedly once the solution of model used is obtained at each iteration step. On the other hand, the output, which is measurable from the real plant, is fed back into the model used in constructing the matching scheme, in turn, updates the model solution iteratively. In such a way, the repetitive solution converges to the true optimal solution of the original dynamic estimation problem, in spite of model-reality differences.

3 A Model-Reality Differences Approach

Now, let us define an expanded dynamic estimation problem, which is referred to as Problem (E), given by

$$\min_{x(k)} J_{mse}(x) = \frac{1}{2}\sum_{k=0}^{N}((x(k)-\bar{x}(k))^{\mathrm{T}}(M_x(k))^{-1}(x(k)-\bar{x}(k)) + (y(k)-\bar{y}(k))^{\mathrm{T}}(M_y(k))^{-1}(y(k)-\bar{y}(k))) + \tfrac{1}{2}r_1 \| \bar{x}(k)-z(k) \|^2$$

subject to (5)

$$\bar{x}(k+1) = A\bar{x}(k)+\alpha_1(k), \qquad \bar{x}(0)=\bar{x}_0$$
$$\bar{y}(k) = C\bar{x}(k)+\alpha_2(k)$$
$$Az(k)+\alpha_1(k) = f(z(k),k)$$
$$Cz(k)+\alpha_2(k) = h(z(k),k)$$
$$z(k) = \bar{x}(k)$$

where $z(k)\in\Re^n$, $k=0,1,...,N$, is introduced to separate the expected state estimate in the state estimation from the respective signal in the parameter estimation, and $\|\cdot\|$ denotes a usual Euclidean norm. The term of $\frac{1}{2}r_1 \| \bar{x}(k)-z(k) \|^2$ with $r_1\in\Re$ is introduced to improve the convexity and to facilitate the convergence of the resulting iterative algorithm. It is important to note that the algorithm is designed such that $z(k)=\bar{x}(k)$ is satisfied upon termination of the iterations, assuming that the convergence is achieved. The state estimate $z(k)$ is used for the computation of parameter estimation and the matching scheme, while the corresponding state estimate $\bar{x}(k)$ will give the optimal state sequence for state estimation. Thus, the optimal state estimation and parameter estimation are mutually interactive.

Then, we write the augmented cost function as

$$\begin{aligned} J'_{mse}(x) = {} & \frac{1}{2}\sum_{k=0}^{N-1}((x(k)-\bar{x}(k))^{\mathrm{T}}(M_x(k))^{-1}(x(k)-\bar{x}(k)) \\ & +(y(k)-\bar{y}(k))^{\mathrm{T}}(M_y(k))^{-1}(y(k)-\bar{y}(k))) \\ & +\tfrac{1}{2}r_1 \| \bar{x}(k)-z(k) \|^2 \\ & +p(k+1)^{\mathrm{T}}(A\bar{x}(k)+Bu(k)+\alpha_1(k)-\bar{x}(k+1)) \\ & +q(k)^{\mathrm{T}}(C\bar{x}(k)+\alpha_2(k)-\bar{y}(k)) \\ & +\mu(k)^{\mathrm{T}}(f(z(k),v(k),k)-Az(k)-Bv(k)-\alpha_1(k)) \\ & +\pi(k)^{\mathrm{T}}(h(z(k),k)-Cz(k)-\alpha_2(k)) \\ & +\beta(k)^{\mathrm{T}}(z(k)-\bar{x}(k)) \end{aligned} \quad (6)$$

where $p(k) \in \Re^n$, $q(k) \in \Re^p$, $\mu(k) \in \Re^n$, $\pi(k) \in \Re^p$, and $\beta(k) \in \Re^n$ are the appropriate multipliers to be determined later.

3.1 Optimal state estimate

By taking the first-order necessary condition $dJ'_{mse}(x) = 0$ for arbitrary $dx(k)$, the coefficients of $dx(k)$ must vanish. After carrying out some algebraic manipulations, the optimal state estimate, which is based on the measurement update, is yielded by

$$\hat{x}(k) = \overline{x}(k) + K_f(k)(y(k) - \overline{y}(k)) \tag{7}$$

and the optimal state estimate, which is based on the time update, is presented by

$$\overline{x}(k+1) = A\hat{x}(k) + \alpha_1(k), \qquad \overline{x}(0) = \overline{x}_0 \tag{8}$$

with the current output measurement, that is,

$$\overline{y}(k) = C\overline{x}(k) + \alpha_2(k) \tag{9}$$

where

$$K_f(k) = M_x(k)C^{\mathrm{T}}M_y(k)^{-1} \tag{10}$$

$$P(k) = M_x(k) - M_x(k)C^T M_y(k)^{-1} C M_x(k) \tag{11}$$

$$M_x(k+1) = AP(k)A^{\mathrm{T}} + GQ_\omega G^{\mathrm{T}}, \quad M_x(0) = M_0 \tag{12}$$

$$M_y(k) = CM_x(k)C^{\mathrm{T}} + R_\eta \tag{13}$$

Here, $K_f(k) \in \Re^{n \times p}$ is the filter gain matrix, $M_y(k) \in \Re^{p \times p}$, $P(k) \in \Re^{n \times n}$ and $M_x(k) \in \Re^{n \times n}$ are positive definite matrices [7, 8, 20, 21]. Notice that by adding the adjusted parameters, the state information (7) gives the minimum output error. It also improves the trajectory of the expected state sequence (8) and the corresponding measured output sequence (9) in the estimation of the original state dynamics.

In addition, the deterministic dynamic system, which is the combination of (7) and (8), is propagated to generate the following optimal state sequence and the corresponding measured output sequence,

$$\overline{x}(k+1) = A\overline{x}(k) + K_p(k)(y(k) - \overline{y}(k)) + \alpha_1(k), \qquad \overline{x}(0) = \overline{x}_0 \tag{14a}$$

$$\overline{y}(k) = C\overline{x}(k) + \alpha_2(k) \tag{14b}$$

where

$$K_p(k) = AK_f(k), \qquad k = 0,1,...,N-1 \tag{15}$$

is the predictor gain.

As a result, the modified model-based dynamic estimation problem, which is referred to as Problem (MM) and satisfies the conditions (7), (8) and (9), is defined as follows:

$$\begin{aligned} \min_{x(k)} J_{mse}(x) = \frac{1}{2}\sum_{k=0}^{N} & ((x(k)-\bar{x}(k))^{\mathrm{T}}(M_x(k))^{-1}(x(k)-\bar{x}(k)) \\ & +(y(k)-\bar{y}(k))^{\mathrm{T}}(M_y(k))^{-1}(y(k)-\bar{y}(k))) \\ & +\tfrac{1}{2}r_1 \| \bar{x}(k)-z(k) \|^2 \end{aligned}$$

subject to (16)

$$\begin{aligned} \bar{x}(k+1) &= A\bar{x}(k)+\alpha_1(k), \qquad \bar{x}(0)=\bar{x}_0 \\ \bar{y}(k) &= C\bar{x}(k)+\alpha_2(k) \end{aligned}$$

3.2 Parameter estimation

Furthermore, the coefficients of $d\mu(k)$ and $d\pi(k)$ are vanished as the first-order necessary condition $dJ'_{mse}(x)=0$ for arbitrary $d\mu(k)$ and $d\pi(k)$. That is, the adjusted parameters are computed from

$$\alpha_1(k) = f(z(k),k) - Az(k) \tag{17a}$$

$$\alpha_2(k) = h(z(k),k) - Cz(k) \tag{17b}$$

Hence, the differences between the real system and the model used are calculated.

The matching scheme is then established based on the separable variable

$$z(k) = \bar{x}(k)$$

where the optimal state estimate is employed to calculation of the adjusted parameters afterward. Notice that the following multipliers satisfy the first-order necessary condition $dJ'_{mse}(x)=0$,

$$\mu(k) = p(k+1) = 0,\ \pi(k) = q(k) = 0,\ \beta(k) = r_1(\bar{x}(k)-z(k))$$

in which the calculus of variation is applied [20, 22, 23].

3.3 Iterative algorithm

From the discussion above, the result can be summarized as an iterative algorithm, which takes into account the model-reality differences during the computation procedure. Therefore, the following calculation steps in the iterative algorithm are presented:

Iterative algorithm

Step 0 Compute a nominal solution. Assume $\alpha_1(k) = 0$, $\alpha_2(k) = 0$, and $r_1 = 0$, calculate $K_f(k)$, $P(k)$, $M_x(k)$, $M_y(k)$ and $K_p(k)$ from (10), (11), (12), (13) and (15), respectively. Then, solve Problem (M) defined by (4) to obtain $\bar{x}(k)$ and $\bar{y}(k)$. Set $i = 0$, $z(k)^0 = \bar{x}(k)^0$ and $\hat{y}(k)^0 = \bar{y}(k)^0$.

Step 1 Compute the adjusted parameters $\alpha_1(k)^i$, $k = 0,1,...,N-1$, and $\alpha_2(k)^i$, $k = 0,1,...,N$, from (17). This step is called the *parameter estimation* step.

Step 2 With the specific $\alpha_1(k)^i$, $\alpha_2(k)^i$, $z(k)^i$ and $\hat{y}(k)^i$, solve Problem (MM) defined by (16). This step is called the *state estimation* step.

3.1 Use (14a) to obtain the new optimal state estimate $\bar{x}(k)^i$, $k = 0,1,...,N$..

3.2 Use (14b) to obtain the new optimal output estimate $\bar{y}(k)^i$, $k = 0,1,...,N$.

Step 3 Test the convergence and update the optimal state estimate of Problem (P). In order to provide a mechanism for regulating convergence, a simple relaxation method is employed:

$$z(k)^{i+1} = z(k)^i + k_z(\bar{x}(k)^i - z(k)^i) \quad (18a)$$

$$\hat{y}(k)^{i+1} = \hat{y}(k)^i + k_y(\bar{y}(k)^i - \hat{y}(k)^i) \quad (18b)$$

where $k_z, k_y \in (0,\ 1]$ are scalar gains. If $z(k)^{i+1} = z(k)^i$, $k = 0,1,...,N$, and $\hat{y}(k)^{i+1} = \hat{y}(k)^i$, $k = 0,1,...,N$, within a given tolerance, stop; else set $i = i+1$ and repeat from Step 1.

Remarks:

(a) The off-line calculation is done, as stated in Step 0, to calculate $K_f(k)$, $P(k)$, $M_x(k)$, $M_y(k)$ and $K_p(k)$. Then, these parameters are

used for solving Problem (M) in Step 0 and for solving Problem (MM) in Step 2, respectively.

(b) The variables $\alpha_1(k)^i$ and $\alpha_2(k)^i$ are zero in Step 0. Their calculated values, as stated in Step 1, change from iteration to iteration.

(c) Problem (P) is not necessary to be linear, and the WLSE is the quadratic cost function for both Problem (P) and Problem (M).

(d) The default value of the scalar gains (k_z, k_y) is 0.9, and this value can be chosen from 0.1 to 0.9 for an optimal number of iteration.

(e) The convergence of $z(k)$ and $\hat{y}(k)$ in Step 3 is verified by comparing the following 2-norm with the given tolerance

$$\| z^{i+1} - z^i \|_2 = \sqrt{\sum_{k=0}^{N} \| z(k)^{i+1} - z(k)^i \|^2} \tag{19a}$$

$$\| \hat{y}^{i+1} - \hat{y}^i \|_2 = \sqrt{\sum_{k=0}^{N} \| \hat{y}(k)^{i+1} - \hat{y}(k)^i \|^2} \tag{19b}$$

4 Illustration

Consider a nonlinear dynamical system [24] in Problem (P) given below:

$$\begin{aligned} x_1(k+1) &= 0.99x_1(k) + 0.2x_2(k) \\ x_2(k+1) &= -0.1x_1(k) + \frac{0.5x_2(k)}{1+(x_2(k))^2} + \omega(k) \\ y(k) &= x_1(k) - 3x_2(k) + \eta(k) \end{aligned}$$

where the initial condition $x(0) = x_0$ is a random vector with mean and covariance are, respectively, given by

$$\overline{x}_0 = \begin{bmatrix} 1.0 \\ 0.8 \end{bmatrix} \text{ and } M = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The stationary Gaussian white noise sequences are $\omega(k)$ and $\eta(k)$ with zero mean and their respective covariance matrices are given by

$$Q_\omega = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad R_\eta = 1.$$

The simplified model in Problem (M) is given below:

$$\bar{x}_1(k+1) = 0.99\bar{x}_1(k) + 0.2\bar{x}_2(k) + \alpha_{11}(k)$$
$$\bar{x}_2(k+1) = -0.1\bar{x}_1(k) + 0.95\bar{x}_2(k) + \alpha_{12}(k)$$
$$\bar{y}(k) = \bar{x}_1(k) - 3\bar{x}_2(k) + \alpha_2(k)$$

with the initial condition

$$\bar{x}_1(0) = 1.0\,, \quad \bar{x}_2(0) = 0.8\,, \quad k = 0,1,...,20$$

and the adjusted parameters $\alpha_1(k) = (\alpha_{11}(k) \quad \alpha_{12}(k))^{\mathrm{T}}$ and $\alpha_2(k)$.

Table 1 shows the simulation result, where there is a 46 percent of the error reduction done by the algorithm proposed. Here, the final WLSE, which is 0.2863, is preferred since this value is smaller than the mean square error (MSE) of the EKF that is 0.4468. In Figure 1, the dynamics of the plant and state estimate are shown, where the state estimate tracks the plant dynamics slightly. In Figure 2, the behavior for the model output is similar equivalently to the original output, which shows the effectiveness of the algorithm proposed.

Table 1. Simulation Result

Number of iteration	Elapsed time	Initial WLSE	Final WLSE
37	0.032799	0.5293	0.2863

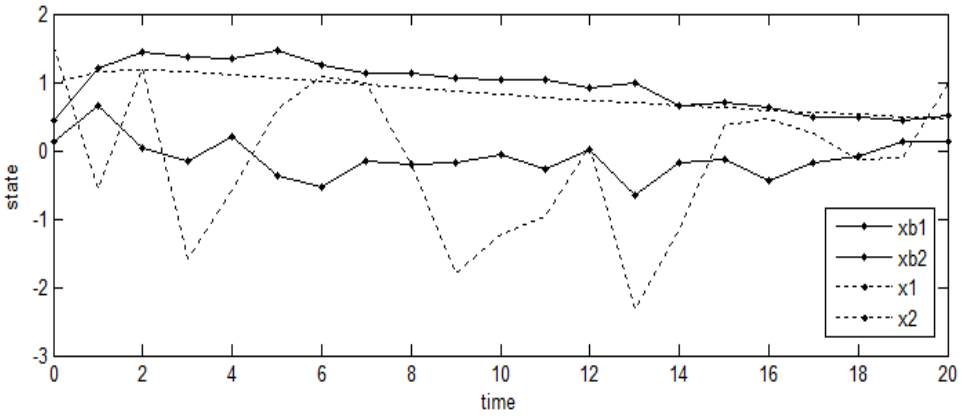


Fig. 1. State Trajectories

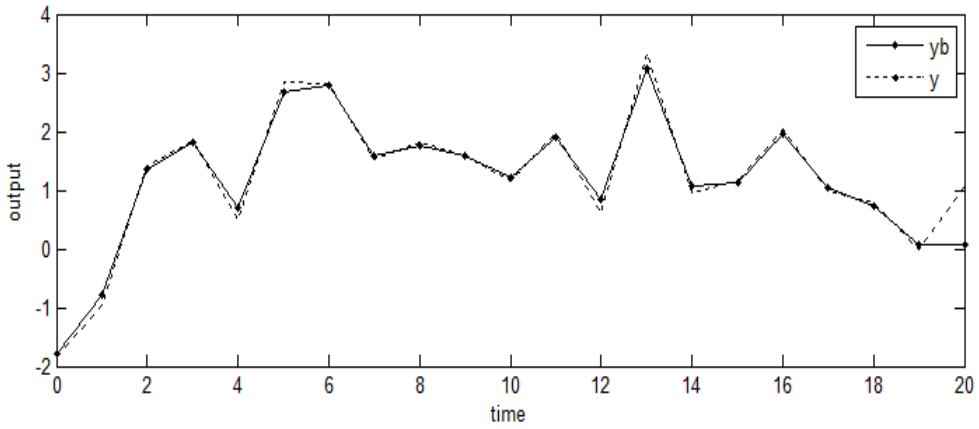


Fig. 2. Output Trajectories

5 Concluding Remarks

The efficient computation approach for solving the nonlinear dynamic estimation problem was discussed in this paper. To solve this problem, the simplified linear model-based estimation problem with adding the adjusted parameters is introduced. During the computation procedure, the differences between the real system and the model used could be taken into account. The real output, which is measured from the real plant, is fed back into the model used in order to update the optimal solution of the model. This is done iteratively. As a result, the iterative solution converges to the true optimal solution of the original estimation problem despite model-reality differences when the convergence is achieved. For illustration, an example was studied and the results showed the efficiency of the algorithm proposed. In conclusion, the applicable of the algorithm proposed to nonlinear dynamic estimation problem is highly recommended.

References

1. Kalman, R. E.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering. 35-45 (1960).
2. Kalman, R. E.: Contributions to the theory of optimal control. Bol. Soc. Mat. Mexicana. 102-119 (1960).
3. Kalman, R. E. and Bucy, R.S.: New results in linear filtering and prediction theory. Journal of Basic Engineering. 95-108 (1961).
4. McElhoe, B. A.: An assessment of the navigation and course corrections for a manned flyby of Mars or Venus, Aerospace and Electronic Systems, IEEE Transactions on. AES-2, 613-623 (1966).

5. Smith, G. L., Schmidt, S. F. and McGee, L. A.: Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle. United States: National Aeronautics and Space Administration. (1962).
6. Ahmed, N. U.: Linear and Nonlinear Filtering for Scientists and Engineers. World Scientific Publishers, Singapore, New Jersey, London, Hong Kong. (1999).
7. Anderson, B. D. O. and More, J. B.: Optimal filtering. Englewood Cliffs NJ: Prentice-Hall. (1979).
8. Bagchi, A.: Optimal control of stochastic systems. New York: Prentice-Hall. (1993).
9. Julier S. and Uhlmann, J.: Unscented filtering and nonlinear estimation. Proceedings of the IEEE. 92, 401–422 (2004).
10. Uhlmann, J.: Dynamic map building and localization: new theoretical foundations. Ph.D. thesis, University of Oxford. (1995).
11. Julier, S. and Uhlmann, J.: A new method for the nonlinear transformation of means and covariances in nonlinear filters. IEEE Trans. on Automatic Control. 45, 477-482 (2000).
12. Rajamani, M. R.: Data-based techniques to improve state estimation in model predictive control. Ph.D. thesis, University of Wisconsin-Madison. (2007).
13. Olfati, S. R.: Collaborative target tracking using distributed Kalman filtering on mobile sensor networks. American Control Conference (ACC). 29 June-1 July, Dartmouth Coll., Hanover, NH, USA. 1100-1105 (2011).
14. Rigatos, G. G.: Derivative-free nonlinear Kalman filtering for MIMO dynamical systems: application to multi-DOF robotic manipulators. International Journal Advanced Robotic Systems. 8 (6), 47-61 (2011).
15. Krymskaya, M. V., Hanea, R. G. and Verlaan, M.: An iterative ensemble Kalman filter for reservoir engineering applications. Computational Geosciences. 13 (2), 235-244 (2009).
16. Andrea, F., Giovanni, S. and Claudio, C.: Enhanced accuracy of continuous glucose monitoring by online extended Kalman filtering. Diabetes Technology & Therapeutics. 12 (5), 353-363 (2010).
17. Feng, Z. G., Teo, K. L., Ahmed, N. U., Zhao, Y. and Yan, W. Y.: Optimal fusion of sensor data for Kalman filtering. Discrete and Continuous Dynamical Systems–Series A. 14 (3), 483-503 (2006).
18. Kek, S. L., Teo K. L. and Mohd Ismail, A. A.: An integrated optimal control algorithm for discrete-time nonlinear stochastic system. International Journal of Control. 83 (12), 2536-2545 (2010).
19. Kek, S. L., Teo, K. L. and Mohd Ismail, A. A.: Filtering solution of nonlinear stochastic optimal control problem in discrete-time with model-reality differences. Numerical Algebra, Control and Optimization (NACO). 2 (1), 207-222 (2012).
20. Bryson, A. E. and Ho,Y. C.: Applied optimal control, Washington, DC: Hemisphere (1975).
21. Bar-Shalom, Y., Li, X. R and Kirubarajan, T.: Estimation with applications to tracking and navigation. New York: John Wiley & Sons, Inc (2001).
22. Lewis, F. L. and Syrmos, V. L.: Optimal Control, 2nd Ed., John Wiley & Sons (1995).
23. Simon, D.: Optimal state estimation: Kalman, H-infinity and nonlinear Approaches. John Wiley & Sons, Inc., Hoboken, New Jersey (2006).
24. Wu, Y., Hu, D., Wu, M. and Hu, X.: Unscented Kalman filtering for additive noise case: augmented vs. non-augmented. American Control Conference, June 8-10, Portland, OR, USA, 4051-4055 (2005).

Algorithm for constructing partition of difference sets

Editha Rivera Jorda
Technological University of the Philippines
Email add: reyjorda@yahoo.com

Abstract The Difference System of Sets (DSS) are combinatorial structures that are a generalization of cyclic difference sets that arise in connection with code synchronization.

The paper discussed some recent combinatorial constructions of the difference systems of sets (DSS) from the partitioning of the multiplicative group $\boldsymbol{GF(n)^*}$ for any prime number $\boldsymbol{n}$ to difference sets of quadratic residue (Paley type) for $\boldsymbol{n \equiv 3(mod\ 4)}$ and the set of quadratic residues for $\boldsymbol{n \equiv 1\ (mod\ 4)}$. The DSS obtained are perfect and regular which gives optimal codes with respect to the Levenshtein bound.

The paper developed a program for constructing DSS from partitioning the multiplicative group $GF(n)$ of any prime n to difference sets of quadratic residues (Paley type) for $n \equiv 3(mod\ 4)$.

Key words: difference system of sets, cyclic difference sets, code synchronization

1 Introduction

When we transmit data, we are concerned about sending a message over a channel that could be affected by the noise. We wish to be able to encode and decode the information in a manner that will allow the detection, and possibly the correction, of errors caused by noise. .

Below is a simple model of a communications system for transmitting and receiving coded messages (Figure 1). Uncoded message may be composed of letters or characters, but usually they consist of binary m – tuples. These messages are encoded into codewords, consisting of binary n - tuples by a device called an encoder. The messages is transmitted and then decoded. We will consider the occurrence of errors during transmission. An error occurs if there is a change in one or more bits in the

codeword. A decoding scheme is a method that either converts an arbitrarily received n - tuples to a meaningful decoded message or gives an error message for that n - tuple. If the received message is a codeword (one of the special n - tuples allowed to be transmitted), then the decoded message must be the unique message that will encoded into the codeword. For received non coded words, the decoding scheme will give an error indication and try to correct the error and reconstruct the original message. The goal is to transmit error-free messages as cheaply and quickly as possible.

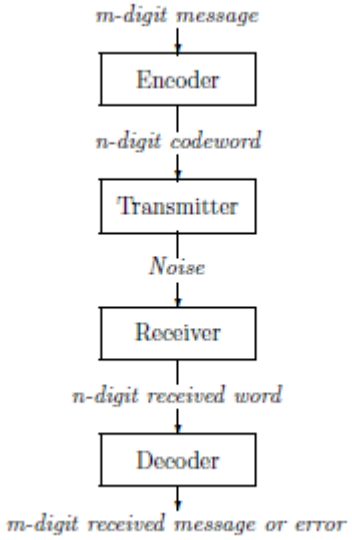


Figure 1. Communication system for transmitting and receiving coded message

Tonchev [8] considered the process of transmitting data over a channel, where the data being sent can be thought of as a stream of symbols from a finite alphabet $F_q = \{0, 1, \dots, q-1\}$. The data stream consists of consecutive messages, each message being a sequence of n consecutive symbols:

$$\dots [x_1 \dots x_n][y_1 \dots y_n] \dots$$

where $\mathbf{x} = x_1 \dots . x_n$ and $y = y_1 \dots y_n$ are two consecutive messages.

Therefore in order that the message decoded by the receiver resembles, as close as possible, the original message, the code should be synchronized. The synchronization problem that arises at the receiving end is the task of partitioning correctly the data stream into messages of length of n, as opposed to conceiving incorrectly a sequence of n symbols being the

concatenation of the end of one message with the beginning of another message as a single message.

$$\dots [x_{i+1} \dots x_n y_1 \dots y_i] \dots$$

The comma-free codes, first introduced in [Crick et al], are designed to solve the code synchronization problem. The comma-free index $\rho(C)$ allows one to distinguish a code word from a joint of two code words (and hence provides for synchronization of code words) provided that at most $\lfloor \rho(C)/2 \rfloor$ errors have occurred in the given code word.

Levenstein [4],[5],and[6] introduced difference systems of sets for the construction of codes with prescribed comma-free index that allow for the synchronization in the presence of errors, that is, the redundancy (r) of a code counts the number of positions added for error detection and corrections. Hence, it is required that the redundancy is as small as possible. Thus degree of resemblance will depend on how good the code is in relation to the channel.

A difference systems of sets (DSS) with parameters $\left(n, \{\tau_0, \tau_1, \dots \tau_{q-1}\}, \rho\right)$ *is a collection of* q *disjoint set* $Q \subseteq \{1, \dots, n\}, |Q| = \tau_i,\ 0 \le i \le q-1,$ *such that the multi-set*

$$M = \{a - b(mod\ n) | a \in Q_i, b \in Q_j, i \neq j\} \qquad (1)$$

contains every number $k, 1 \le k \le n-1,$ at least ρ times. A DSS is perfect if every number $i,\ (1 \le i \le n-1)$ appears exactly ρ times in the multi-set of differences (1). A DSS is regular if all subsets Q_i are of the same size that is, $\tau_0 = \tau_1 = \dots = \tau_{q-1\tau} = m.$ The notation used for the regular DSS on n points with q subsets of size m is (n, m, q, p).

Since DSS are a generalization of cyclic difference sets, Tonchev [8] direct constructions of perfect and regular DSS is obtained by using general partitions of cyclic difference sets, will be discussed in Sect 2.

In Sect.3, a general algorithm for constructing DSS obtained from cyclic difference sets with given parameters n, m, q, ρ will be described. The out put of the program and tables of DSS found by using this algorithm are given in Sect.4

2 DSS as partitions of difference sets

In this section, we consider partition of the multiplicative group $GF(n)^*$ of a finite field of prime order n defined by a subgroup of $GF(n)^*$ and its cosets, and the difference sets of Paley type, or equivalently, by partitioning the set of quadratic residues Q modulo a prime number $n \equiv 3(mod\ 4)$.

Let $D = \{x_1, x_2, \dots, x_k\}$ be a (v, k, λ) difference sets([1],[2],[7]}, that is a subset of k residues modulo v such that every positive residue modulo v occurs exactly λ times is the multi-set of differences.

$$\{x_i - x_j (mod\ v) | x_i, x_j \in D, x_i \neq x_j\} \quad (2)$$

Then the collection of singletons $Q_1 = \{x_1\}, \dots, Q_{k-1} = \{x_k\}$ is a perfect regular DSS with parameters $(n = v, m = 1`, q = k, \rho = \lambda)$. Thus, DSS are a generalization of cyclic difference sets.

The following lemma generalizes this simple construction by using more general partitions of difference sets

Lemma 2.1 (Tonchev [8]). *Let $D \subseteq \{1,2, \dots, n\}, |D| = k$, be a cyclic (n, k, λ) difference set. Assume that D is partitioned into q disjoint subsets $Q_0, \dots, Q_{q-1}$ that are the base blocks of a cyclic design B with block sizes $\tau_i = |Q_i|, i = 0, \dots, q-1$ such that every two points are contained in at most λ_i blocks. Then the sets $Q_0, \dots, Q_{q-1}$ forms a DSS with parameters $(n, \tau_0, \dots \tau_{q-1}, \rho = \lambda - \lambda_1)$. The DSS $\{Q_i\}_{i=0}^{q-1}$ is perfect if and only if B is a pairwise balanced design with every two points occurring together in exactly λ_1 blocks.*

We consider partition of the multiplicative group $GF(n)^*$ of a finite field of prime order n defined by a subgroup of $GF(n)^*$ and its cosets as the first application of Lemma 2.1 in the following theorem.

Theorem 2.2 (Tonchev [8]) *Let $n = mq + 1$ for some $m, q \in \mathbb{Z}_n$ and let α be a primitive element of the finite field order $n, GF(n)$. The collections of sets:*

$$Q_0 = \{\alpha^q, \alpha^{2q}, \dots, \alpha^{mq} = 1\}, \quad Q_1 = \alpha Q_0, \dots, \ Q_{q-1} = \alpha^{q-1} Q_0$$

is a perfect regular $(n, m, q, \rho = n - m - 1)$ DSS.

Example 2.3 *Let $n = 19 = 3 \cdot 6 + 1, q = 3, m = 6$ and $\alpha = 2$ as its primitive element. The DSS from Theorem 2.2 has $\rho = 12$ and the three sets Q_i of size 6 are*

$$Q_0 = \{8,7,18,11,12,1\},\ Q_1 = \{16,14,17,3,5,2\},\ Q_2 = \{13,9,15,6,10,4\}$$

Hence, the collection $Q_0, Q_1, and Q_2$ is a perfect regular $(19, 6, 3, 12)$ DSS .

The next theorem gives perfect regular DSS's obtained as partitions of

differences of quadratic- residues (QR), or Paley type. It uses partition of a subgroup of the multiplicative group of a finite field of prime order $n \equiv 3(mod\ 4)$.

Theorem 2.4 *(*Tonchev[8]). *For every prime* $n = 2q + 1 \equiv 3(mod\ 4)$ *there exists a perfect regular DSS with parameters* $(n, m, q, \rho = (n - 2m - 1)/4)$. *The set of quadratic residues is defined as*

$$Q = \left\{\alpha^{2i} \middle| 1 \le i \le (n-1)/2\right\} \qquad (3)$$

and let $D_m = \left\{\left(\alpha^{2iq}\right) \middle| 1 \le i \le m\right\}$ *be the cyclic subgroup of* Q *of order m. We defined the collection of sets* $Q_1, Q_2, \dots, Q_{q-1}$ *to be cosets of* D_m *in* Q:

$$Q_0 = D_m, \quad Q_1 = D_m\alpha^2, \dots, Q_{q-1} = D_m\alpha^{2(q-1)}$$

Example 2.5 *Let* $n = 2 \cdot 3 \cdot 3 + 1,\ q = 3, m = 3,\ \alpha = 2$ *as its primitive element. The sets* Q *and* D_3 *defined as in Theorem 2.4 are:*

$$Q = \{4,16,7,9,178,11,6,5,1\} \text{ and } D_3 = \{7,11,1\}$$

and the three sets of Q_i *of size* 3 *are*

$$Q_0 = D_3 = \{7,11,1\}, Q_1 = \{9,6,4\}, Q_2 = \{17,5,16\}$$

Hence, the collection $Q_0, Q_1, and\ Q_2$ *is a perfect regular* $(19,3,3,15)$ *DSS.*

3 An algorithm for constructing DSS

In this section, we describe an algorithm in constructing DSS obtained as partition of the multiplicative group $GF(n)^*$ of a finite field of prime order n and differences of quadratic residues or Paley type.

The following are the minimum system requirements for a computer to run the C++ language,and how the program operates algorithm.

3.1 Minimum system requirements

The minimum systems requirements for a computer to run the C++ language is at least it has Windows 98 operating system, the executable file (DSSv6.exe) may be run in DOS mode and the program(DSSv6.cpp) was developed using Bloodshed C++.

3.2 How the Algorithm/Program Works?

Generally, the algorithm follows the concept of modular programming. There are five modules or procedures (main() was excluded): accept(), testrange(), testprime(), primitive(), and findalpha() and findalpha2().

3.2.1 Accept() procedure

This procedure signals the user to input an integer.

3.2.2 Testrange() procedure

After the user inputted an integer, this procedure checks whether the Input integer is accepted or not. Theoretically, the program can accept any positive integer greater than or equal to 3. However, since C++ data structures have limitations (the input integer is declared as **integer** data type– having a value range of -32,768 to 32767), the program cannot accept any integer greater than 32,767. Moreover, even though the program can accept such integer, the output of the program will clutter the screen. Having said this, it is advisable we choose arbitrarily 3,571 as the highest allowable input.

3.2.3 Testprime() procedure

Once an accepted input was found (that is, within the range), the previous procedure calls **testprime()** to determine whether an input is composite or prime. For example, if the user inputs 8. 8 is within the range yet a composite number. If this is the case, **testprime()** will call **accept()** for another valid input. On one hand, if the user inputs 7, which is within the range and a prime number, the **primitive()** produce will be called.

3.2.4 Primitive() procedure

Primitive() procedure allows the user to input the number of sets (q) of the DSS. The user will be asked whether if he wants to compute for partition of any prime n (multiplicative group of a finite field) or Partition of quadratic residues. The procedure will keep on running until the user input a correct q. This is shown in the given code below:

```
ret: gotoxy(0,2); cout<<"\t--------------------------------------------------------------------------------";
   cout<<"\n\tWhich one do you want to perform?\n\n";
cout<<"\t[1] Partition of any prime n (multiplicative group of a finite field)\n";
cout<<"\t[2] Partition of quadratic residues\n";  cout<<"\t[3] Exit";
cout<<"\n\n\tChoice:  "; cin>>ch;
if(ch>3 || ch<1) {  cout<<"\tInvalid choice! Try again.\n"; Sleep(1000);
   gotoxy(18,8);  cout<<"  "; gotoxy(8,10);  cout<<"\t\t\t\t"; goto ret;  } else {
if(ch==1){  cout<<"\t----------------------------------------------------------------------";
here:  gotoxy(0,10);  cout<<"\n\tEnter the value of q:  "; cin>>q;
```

```
        while(((n-1)%q)!=0) {
          cout<<"\t"<<q<<" does not divide "<<n-1<<". Press enter to try again.";
          cin.sync(); cin.get(); gotoxy(8,12);  cout<<"\t\t\t\t\t\t\t"; goto here; }
        m=(n-1)/q; findalpha(); }
 else if(ch==2) { if((n%4)==3){ cout<<"\t-------------------------------------------------";
    enterq:  gotoxy(0,10); cout<<"\n\tEnter the value of q:  "; cin>>q;
      while((n-1)%(2*q)!=0) {
        cout<<"\t"<<q*2<<" does not divide "<<n-1<<". Press enter to try again.";
        cin.sync(); cin.get(); gotoxy(8,12);  cout<<"\t\t\t\t\t\t\t"; goto enterq;  }
      m=(n-1)/(2*q); findalpha2(); }
 else { cout<<"\tInvalid input! Press enter to try again.\t\n"; cin.sync(); cin.get();
 main(); } } else { exit(0);  } getche();
```

3.2.5 Findalpha() procedure

This procedure finds the smallest primitive element of the integer *n*. The variable *alpha* is initialized to 2 as the smallest possible primitive element of the input integer *n*. Special case is given if n=3 since the 3 mod 2 = 1. For any other integers, the program will find *alpha* from 2 up to n-1. The program will halt if a primitive element is found; otherwise, alpha will be increased by 1. *alpha* is found if *flag* is less than 2. *flag* determines whether there is duplication of elements when *alpha raised to (1 up to n-1) mod (n-1)*. If there is a duplication of elements, therefore, alpha is not the "true" alpha. As mentioned earlier, alpha will be increased by 1 until it reaches (n-1) or until flag is less than 2.

The code is given below.

```
alpha=2;
do{ element[0]=alpha;  result=alpha*alpha;          //trivial base
  if(result >= n)  element[1]=result % n;
  else{  if(n==3){ element[1]= result % n; } else{ element[1]=result; } }
  for(i=2;i<=(n-1)&&(flag<2);i++){
    result=(result*alpha)%n; element[i]=result;
    if(result==1)  flag++; } if(flag<2)  {found=1;}   else { alpha++; flag=0; }
} while(found!=1); cout<<"\nThe primitive element is "<<alpha<<".";
```

The code below generates the sets $Q_0, Q_1, \ldots, Q_{q-1}$.

```
for(int i=(q-1);i<(n-1);i=i+q) {      //find the initial set Q0
  Q0[a]=element[i];  cout<<Q0[a];
  if((i+q)<(n-1)){  cout<<","; } a++; }
for(int q2=0,t=q;q2<=(q-2);q2++,t++) {
  cout<<"Q"<<q2+1<<"={";
  for(int m1=0,k=0;m1<m;m1++,k++)  {   //find the sets  Q1,Q2,Q3,...,Qq-1
    if(t>=(n-1)){  t=t%(n-1); } dss[q2][m1]=element[t];  t=t+q;
    cout<<dss[q2][m1];  if((k+1)<m)  cout<<","; } l++; cout<<"} \n "; }
```

If the user choose the partition of quadratic residues, findalpha2() will be called. The code is given below.

```
char ans;  int result,flag=0; int element[n-1];  alpha=2;
do { element[0]=alpha;  result=alpha*alpha;          //trivial base
  if (result >= n)  {element[1]=result % n;} else{ element[1]=result;}
 for(i=2;i<=(n-1)&&(flag<2);i++) { result=(result*alpha)%n;
    element[i]=result;  if(result==1)  flag++; }
  if(flag<2) { found=1;}  else { alpha++;  flag=0;  }
 } while(found!=1);  cout<<"\nThe primitive element is "<<alpha<<".";
cout<<"("<<n<<", "<<m<<", "<<q<<", "<<(n-2*m-1)/4<<")\n\n";
int b=0; int Q[m]; int a=0; cout<<"\n";  cout<<"Q={";
for(int i=1;i<(n-1);i=i+2)  { Q[a]=element[i]; cout<<Q[a];
if(i<n-2) cout<<",";   a++; } cout<<"}\n"; double dss[q][m];
 int k,t,l=0,counter=0;  for(int q2=0,t=(q-1);q2<=(q-1);q2++,t=q-1+l){
 cout<<"Q"<<q2<<"={"; for(int m1=0,k=0;m1<m;m1++,k++) {
  if(t>=m*q)  t=t%q;  dss[q2][m1]=Q[t]; t=t+q;  cout<<dss[q2][m1];
 if((k+1)<m)  cout<<","; } l++; cout<<"}\n"; }
 cout<<"Press B to go back to main menu: \t\t\t\t\t"; ans=getche();
if(ans=='b' || ans=='B') { first=1; main(); } else { first=1; ver=0; proc1(); }
```

4 Tables of DSS

In this section, Tables 1 and 2 of DSS partitioned from any prime n and set of quadratic residues generated by the algorithm described in Sec. 3. Here only partitions that satisfy the necessary conditions of Theorem 2.2 and Theorem 2.4 are shown.

Table 1. DSS partitioned from any prime n

n	GF(n)*	q	m	alpha	Sets	Parameters
3	{1,2}	2	1	2	Q0={1} Q1={2}	(3,1,2,1)
5	{1,2,3,4}	2	2	2	Q0={4,1} Q1={3,2}	(5,2,2,2)
7	{1,2,3,4,5,6}	2	3	3	Q0={2,4,1} Q1={6,5,3}	(7,3,2,3)
7	{1,2,3,4,5,6}	3	2	3	Q0={6,1} Q1={4,3} Q2={5,2}	(7,2,3,4)
11	{1,2,3,4,5,6,7,8,9,10}	2	5	2	Q0={4,5,9,3,1} Q1={8,10,7,6,2}	(11,5,2,5)
11	{1,2,3,4,5,6,7,8,9,10}	5	2	2	Q0={10,1} Q1={9,2} Q2={7,4} Q3={3,8} Q4={6,5}	(11,2,5,8)
13	{1,2,3,4,5,6,7,8,9,10,11,12}	2	6	2	Q0={4,3,12,9,10,1} Q1={8,6,11,5,7,2}	(13,6,2,6)

n	GF(n)*	q	m	alpha	Sets	Parameters
13	{1,2,3,4,5,6,7,8,9,10,11,12}	6	2	2	Q0=(12,1) Q1=(11,2) Q2=(9,4) Q3=(5,8) Q4=(10,3) Q5=(7,6)	(13,2,6,10)
13	{1,2,3,4,5,6,7,8,9,10,11,12}	3	4	2	Q0=(8,12,5,1) Q1=(3,11,10,2) Q2=(6,9,7,4)	(13,4,3,8)
13	{1,2,3,4,5,6,7,8,9,10,11,12}	4	3	2	Q0=(3,9,1) Q1=(6,5,2) Q2=(12,10,4) Q3=(11,7,8)	(13,3,4,9)
17	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16}	8	2	3	Q0=(16,1) Q1=(14,3) Q2=(8,9) Q3=(7,10) Q4=(4,13) Q5=(12,5) Q6=(2,15) Q7=(6,11)	(17,2,8,14)
17	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16}	2	8	3	Q0=(9,13,15,16,8,4,2,1) Q1=(10,5,11,14,7,12,6,3)	(17,8,2,8)
17	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16}	8	2	3	Q0=(16,1) Q1=(14,3) Q2=(8,9) Q3=(7,10) Q4=(4,13) Q5=(12,5) Q6=(2,15) Q7=(6,11)	(17,2,8,14)
17	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16}	4	4	3	Q0=(13,16,4,1) Q1=(5,14,12,3) Q2=(15,8,2,9) Q3=(11,7,6,10)	(17,4,4,12)
19	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}	2	9	2	Q0=(4,16,7,9,17,11,6,5,1) Q1=(8,13,14,18,15,3,12,10,2)	(19,9,2,9)
19	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}	9	2	2	Q0=(18,1) Q1=(17,2) Q2=(15,4) Q3=(11,8) Q4=(3,16) Q5=(6,13) Q6=(12,7) Q7=(5,14) Q8=(10,9)	(19,2,9,16)
19	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18}	6	3	2	Q0=(7,11,1) Q1=(14,3,2) Q2=(9,6,4) Q3=(18,12,8) Q4=(17,5,16) Q5=(15,10,13)	(19,3,6,15)

n	GF(n)*	q	m	alpha	Sets	Parameters
23	{1,2,3,4,5, 6,7,8,9,10, 11,12,13,14, 15,16,17,18, 19,20,21,22}	2	11	5	Q0={2.4,8,16,9,18, 13,3,6,12,1} Q1={10,20,17,11,22, 21,19.15,7,14,5}	(23,11,2,11)
23	{1,2,3,4,5, 6,7,8,9,10, 11,12,13,14, 15,16,17,18, 19,20,21,22}	11	2	5	Q0={22,1} Q1={18,5} Q2={21,2} Q3={13,10} Q4={19,4} Q5={3,20} Q6={15,8} Q7={6,17} Q8={7,16} Q9={12,11} Q10={14,9}	(23,2,11,20)

Table 2. DSS partitioned from set of quadratic residues

n	QR(n)*	q	m	alpha	Sets	Parameters
3	{1}	1	1	2	Q0={1}	(3,1,1,1)
7	{2,4,1}	1	3	3	Q0={2,4,1}	(7,3,1,3)
7	{2,4,1}	3	1	3	Q0={1} Q1={2} Q2={4}	(7,1,3,5)
11	{4,5,9,3,	1	5	2	Q0={4,5,9,3,1}	(11,5,1,5)
11	{4,5,9,3, 1}	5	1	2	Q0={1} Q1={4} Q2={5} Q3={9} Q4={3}	(11,1,5,9)
19	{4,16,7,9, 17,11,6,5,	1	9	2	Q0={4,16,7,9,17,11, 6,5,1}	(19,9,1,9)
19	{4,16,7,9, 17,11,6,5, 1}	3	3	2	Q0={7,11,1} Q1={9,6,4} Q2={17,5,16}	(19,3,3,15)
23	{2,4,8,16, 9,18,13,3, 6,12,1}	1	11	5	Q0={2,4,8,16,9, 18,13,3,6,12, 1}	(23,11,1,11)
23	{2,4,8,16, 9,18,13,3, 6,12,1}	11	1	5	Q0={1} Q1={2} Q2={4} Q3={8} Q4={16} Q5={9} Q6={18} Q7={13} Q8={3} Q9={6} Q10={12}	(23,1,11,21)

References

1. Beth,T., D. Jungnickel, H. Lenz: Design Theory, 2nd ed, Cambridge University Press, Cambridge, (1999)
2. Colbourn, C.J., and Dinitz, J.F.: The CRC Handbook of Combinatorial Designs. CRC Press, Boca Raton,(1996)
3. Golomb, S.W., Gordon B., Welch, L.R.: Comma-free codes. Canad. J. Math vol.10, no.2, 202-209, (1958)
4. Levenshtein, V.I.: One method of constructing quasilinear codes providing synchronization in the presence of errors. Problem of Information Transmission, vol.7, no.3, 215-222 (1971)
5. Levenshtein, V.I.: Combinatorail problems motivated by comma-free codes. J. Combin. Designs,12, 184-196 (2004)
6. Levenshtein, V. I. And Tonchev, V. D.: Constructions of difference systems of sets. Algebraic and Combinatorial Coding Theory, Eight International Workshop Proc., St Peterburg, Russia, 194-197 (2002)
7. Tonchev, V.D.: Combinatorial Configurations. Wiley, New York (1988)
8. Tonchev, V.D.: Partition of difference sets and code synchronization. Finite fields and their application, 601-621, (2005)

The Fekete-Szegö Problem for Class of p-Valent Functions with Respect to Symmetric Points

Aini Janteng[1], and Part Leam Loh[1]

[1]Faculty of Science and Natural Resources, Universiti Malaysia Sabah,
Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
aini_jg@ums.edu.my, travis_loh@hotmail.com

Abstract. Let A_p be the class of $p-$valent functions $f(z)$ of the form $f(z) = z^p + \sum_{n=1}^{\infty} a_{p+n} z^{p+n}$ which are analytic in the open unit disc $D = \{z: |z| < 1\}$ and $p \in \mathbb{N} = \{1,2,3, \dots\}$. In this paper, two classes $M_s(p, A, B)$ and $N_s(p, A, B)$ are considered which are consisting of functions $f(z) \in A_p$ and satisfying $\frac{2zf'(z)}{p[f(z)-f(-z)]} \prec \frac{1+A(z)}{1+B(z)}$ and $\frac{2[zf'(z)]'}{p[f(z)-f(-z)]'} \prec \frac{1+A(z)}{1+B(z)}$ respectively with $-1 \leq B < A \leq 1$and $z \in D$. The Fekete-Szegö inequality for functions $f(z)$ belongs to these classes are obtained.

1 Introduction

Let U be the class of functions which are analytic in the open unit disc $D = \{z: |z| < 1\}$ given by

$$w(z) = \sum_{n=1}^{\infty} b_n z^n \tag{1}$$

and satisfying the conditions

$$w(0) = 0, \qquad |w(z)| < 1, \qquad z \in D.$$

Let A_p denote the class of functions f which are analytic and $p-$valent in D of the form

$$f(z) = z^p + \sum_{n=1}^{\infty} a_{p+n} z^{p+n}, \quad p \in \mathbb{N} = \{1,2,3, \dots\}, z \in D. \tag{2}$$

Also, let $M_s(p)$ be the subclass of A_p consisting of functions given by (2) satisfying

$$Re\left[\frac{zf^{'}(z)}{p[f(z)-(f-z)]}\right]>0, \quad z\in D$$

and $N_s(p)$ be the subclass of A_p consisting of functions given by (2) satisfying

$$Re\left[\frac{[zf^{'}(z)]'}{p[f(z)-f(-z)]'}\right]>0, \quad z\in D.$$

Further, let $f,g\in U$. Then we say that f is subordinate to g, and we write $f\prec g$, if there exixts a function $w\in U$ such that $f(z)=g\big(w(z)\big)$ for all $z\in D$. Specially if g is univalent in D, then $f\prec g$ if and only if $f(0)=g(0)$ and $f(D)\subseteq g(D)$.

In term of subordination, let $M_s(p,A,B)$ denote the class of functions of the form (2) and satisfying the condition

$$\frac{2zf'(z)}{p[f(z)-f(-z)]}\prec\frac{1+Az}{1+Bz}, \qquad -1\le A<B\le 1, \quad z\in D$$

and $N_s(p,A,B)$ denote the class functions of the form (2) and satisfying the condition

$$\frac{2[zf^{'}(z)]'}{p[f(z)-f(-z)]'}\prec\frac{1+Az}{1+Bz}, \qquad -1\le A<B\le 1, \quad z\in D.$$

By definition of subordination, it follows that $f\in M_s(p,A,B)$ if and only if

$$\frac{2zf'(z)}{p[f(z)-f(-z)]}=\frac{1+A[w(z)]}{1+B[w(z)]}, \quad w\in U \tag{3}$$

and $f\in N_s(p,A,B)$ if and only if

$$\frac{2[zf^{'}(z)]^{'}}{p[f(z)-f(-z)]'}=\frac{1+A[w(z)]}{1+B[w(z)]}, \quad w\in U. \tag{4}$$

2 Preliminary Result

The following lemma is required to prove our main result.

Lemma 1 ([1]) If $w(z)$ is given by (1), then

$$|b_1| \leq 1 \ , \ \ |b_2| \leq 1 - |b_1|^2$$

3 Main Result

Theorem 1 If $f(z) \in M_s(p, A, B)$ and μ is a complex number, then

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \begin{cases} \dfrac{p(A-B)}{2}, & if\ |\lambda - \mu| \leq v \\ \dfrac{p^2(A-B)^2}{(p+1)^2}|\lambda - \mu|, & if\ |\lambda - \mu| \geq v \end{cases}$$

where

$$\lambda = \frac{B(p+1)^2}{2p(B-A)} \text{ and} \tag{5}$$

$$v = \frac{(p+1)^2}{2p(A-B)} \tag{6}$$

Proof: Since $f(z) \in M_s(p, A, B)$, from (3) we have

$$\begin{aligned} &2\left[pz^p + (p+1)a_{p+1}z^{p+1} + (p+2)a_{p+2}z^{p+2} + \cdots\right] \\ &= p \cdot \left[(1-(-1)^p)z^p + (1+(-1)^p)a_{p+1}z^{p+1} + (1-(-1)^p)a_{p+2}z^{p+2}\right. \\ &\qquad \left. + \cdots\right] \cdot [1 + (A-B)b_1 z + (A-B)(b_2 - Bb_1^2)z^2 + \cdots] \end{aligned}$$

Identifying the term, we have

$$a_{p+1} = \frac{p(A-B)b_1}{(p+1)} \tag{7}$$

$$a_{p+2} = \frac{p(A-B)}{2}b_2 - \frac{pB(A-B)}{2}b_1^2 \tag{8}$$

From (7) and (8), we get

$$a_{p+2} = \frac{p(A-B)}{2} b_2 + \frac{p^2(A-B)^2}{(p+1)^2} \lambda b_1^2$$

where λ is defined by (5). Therefore

$$a_{p+2} - \mu a_{p+1}^2 = \frac{p(A-B)}{2} b_2 + \frac{p^2(A-B)^2}{(p+1)^2} (\lambda - \mu) b_1^2$$

Applying triangle inequality

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \frac{p(A-B)}{2} |b_2| + \frac{p^2(A-B)^2}{(p+1)^2} |\lambda - \mu| |b_1|^2 \tag{9}$$

and using Lemma 1 with $|b_2| \leq 1 - |b_1|^2$, (9) gives

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \frac{p(A-B)}{2} + \frac{p^2(A-B)^2}{(p+1)^2} [|\lambda - \mu| - v] |b_1|^2$$

where v is defined by (6).

If $|\lambda - \mu| \leq v$, then

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \frac{p(A-B)}{2}$$

which the bound is sharp for $w(z) = z^2$.

If $|\lambda - \mu| \geq v$, then by Lemma 1 with $|b_1| \leq 1$

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \frac{p^2(A-B)^2}{(p+1)^2} |\lambda - \mu|$$

This bound is sharp for $w(z) = z$.

Theorem 2 If $f(z) \in N_s(p, A, B)$ and μ is a complex number, then

$$\left|a_{p+2} - \mu a_{p+1}^2\right| \leq \begin{cases} \dfrac{p(A-B)}{2(p+2)}, & if\ |\lambda - \mu| \leq v \\ \dfrac{p^4(A-B)^2}{(p+1)^4}|\lambda - \mu|, & if\ |\lambda - \mu| \geq v \end{cases}$$

where

$$\lambda = \frac{B(p+1)^4}{2p^2(p+2)(B-A)} \text{ and} \tag{10}$$

$$v = \frac{(p+1)^4}{2p^2(p+2)(A-B)} \tag{11}$$

Proof: Since $f(z) \in N_s(p, A, B)$, from (4) we have

$$\begin{aligned} &2\left[p^2 z^{p-1} + (p+1)^2 a_{p+1} z^p + (p+2)^2 a_{p+2} z^{p+1} + \cdots\right] \\ &= p \cdot \left[p(1-(-1)^p) z^{p-1} + (p+1)(1+(-1)^p) a_{p+1} z^p + (p\right. \\ &\qquad \left. + 2)(1-(-1)^p) a_{p+2} z^{p+1} + \cdots\right] \cdot [1 + (A-B) b_1 z \\ &\qquad + (A-B)(b_2 - B b_1^2) z^2 + \cdots] \end{aligned}$$

Identifying the term, we have

$$a_{p+1} = \frac{p^2(A-B)}{(p+1)^2} b_1 \tag{12}$$

$$a_{p+2} = \frac{p^2(A-B)}{2(p+2)} b_2 - \frac{p^2 B(A-B)}{2(p+2)} b_1^2 \tag{13}$$

From (12) and (13), we get

$$a_{p+2} = \frac{p^2(A-B)}{2(p+2)} b_2 + \frac{p^4(A-B)^2}{(p+1)^4} \lambda b_1^2$$

where λ is defined by (10). Therefore

$$a_{p+2} - \mu a_{p+1}^2 = \frac{p^2(A-B)}{2(p+2)} b_2 + \frac{p^4(A-B)^2}{(p+1)^4} [\lambda - \mu] b_1^2$$

Applying triangle inequality

$$\left|a_{p+2}-\mu a_{p+1}^2\right| \leq \frac{p^2(A-B)}{2(p+2)}|b_2| + \frac{p^4(A-B)^2}{(p+1)^4}|\lambda-\mu||b_1|^2 \qquad (14)$$

and using Lemma 1with $|b_2| \leq 1-|b_1|^2$, (14) gives

$$\left|a_{p+2}-\mu a_{p+1}^2\right| \leq \frac{p^2(A-B)}{2(p+2)} + \frac{p^4(A-B)^2}{(p+1)^4}[|\lambda-\mu|-v]|b_1|^2$$

where v is defined by (11).

If $|\lambda-\mu| \leq v$, then

$$\left|a_{p+2}-\mu a_{p+1}^2\right| \leq \frac{p^2(A-B)}{2(p+2)}$$

which the bound is sharp for $w(z)=z^2$.

If $|\lambda-\mu| \geq v$, then by Lemma 1 with $|b_1| \leq 1$, (14) gives

$$\left|a_{p+2}-\mu a_{p+1}^2\right| \leq \frac{p^4(A-B)^2}{(p+1)^4}|\lambda-\mu|$$

This bound is sharp for $w(z)=z$.

By taking $p=1$ in Theorem 1 and Theorem 2, we obtain the following corollaries respectively.

Corollary 1 ([2]) If $f(z) \in S_s^*(A,B)$ and μ is a complex number, then

$$|a_3-\mu a_2^2| \leq \begin{cases} \frac{B-A}{2}\left[B+\frac{\mu}{2}(A-B)\right] & if\ \ \mu \leq -2\left[\frac{1+B}{A-B}\right] \\ \frac{A-B}{2} & if\ \ -2\left[\frac{1+B}{A-B}\right] \leq \mu \leq 2\left[\frac{1-B}{A-B}\right] \\ \frac{A-B}{2}\left[B+\frac{\mu}{2}(A-B)\right] & if\ \ \mu \geq 2\left[\frac{1-B}{A-B}\right] \end{cases}$$

Corollary 2 ([2]) If $f(z) \in C_s(A,B)$ and μ is a complex number, then

$$|a_3 - \mu a_2^2| \leq \begin{cases} \frac{B-A}{6}\left[B + \frac{3}{8}\mu(A-B)\right] & if \ \ \mu \leq -\frac{8}{3}\left[\frac{1+B}{A-B}\right] \\ \frac{A-B}{6} & if \ \ -\frac{8}{3}\left[\frac{1+B}{A-B}\right] \leq \mu \leq \frac{8}{3}\left[\frac{1-B}{A-B}\right] \\ \frac{A-B}{6}\left[B + \frac{3}{8}\mu(A-B)\right] & if \ \ \mu \geq \frac{8}{3}\left[\frac{1-B}{A-B}\right] \end{cases}$$

References

1. Mehrok, B. S., Singh, H.: A coefficient inequality for a certain class of analytic functions. Int. Journal of Math. Analysis 5(7), 311-318 (2011)
2. Shamugam, T. N., Ramachandran, C., Ravichandran, V.: Fekete-Szegö problem for subclasses of starlike functions with respect to symmetric points. Bull. Korean Math. Soc. 43(3), 589-598 (2006)

The Fekete Szegö Problem for a Subclass of Quasi-Convex Functions with Respect to Symmetric Points

Aini Janteng[1], and Puoi Choo Chuah[1]

[1]Faculty of Science and Natural Resources, Universiti Malaysia Sabah,
Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.
aini_jg@ums.edu.my, colorwind.cpc@gmail.com

Abstract. The purpose of the present paper is to introduce the class $K_s^*(\alpha,\beta)$, subclass of quasi-convex functions with respect to (w.r.t) symmetric points. Sharp upper bounds for $|a_n|, n=2,3,4$ and the Fekete-Szegö inequalities are considered for functions belonging to the class $K_s^*(\alpha,\beta)$.

1 Introduction

Let S be the class of functions f which are analytic and univalent in the open unit disc $D=\{z:|z|<1\}$ given by

$$f(z)=z+\sum_{n=2}^{\infty}a_n z^n \tag{1}$$

where a_n is a complex number.

For $0\le\alpha<1$, Janteng in [1] was introduced a new subclass of quasi-convex functions denoted by $K_s^*(\alpha)$.

Definition 1 Let f be given by (1) and $0\le\alpha<1$. Then $f\in K_s^*(\alpha)$, if there exists a $g\in C_s$ such that for $z\in D$,

$$\mathrm{Re}\left\{\frac{2\alpha\left(z^2f''(z)\right)'}{\left(g(z)-g(-z)\right)'}+\frac{2\left(zf'(z)\right)'}{\left(g(z)-g(-z)\right)'}\right\}>0.$$

Note: The definition above is also equivalent to the following:
$f\in K_s^*(\alpha)$, if there exists a $h=zg'\in S_s^*$ such that

$$\operatorname{Re}\left\{\frac{2\alpha z\left(z^2 f^{'/}(z)\right)'}{h(z)-h(-z)}+\frac{2z\left(zf^{'}(z)\right)'}{h(z)-h(-z)}\right\}>0.$$

We now consider a class which covers some well-known classes of univalent functions as follows:

Definition 2 Let f be given by (1). Then $f\in K_s^*(\alpha,\beta), 0\le\alpha<1, 0\le\beta<1,$ if there exists a $g\in C_s(\beta)$ such that for $z\in D$,

$$\operatorname{Re}\left\{\frac{2\alpha\left(z^2 f^{''}(z)\right)'}{\left(g(z)-g(-z)\right)'}+\frac{2\left(zf^{'}(z)\right)'}{\left(g(z)-g(-z)\right)'}\right\}>0.$$

Note: The definition above is also equivalent to the following:
$f\in K_s^*(\alpha,\beta)$, if there exists a $h=zg^{'}\in S_s^*(\beta)$ such that

$$\operatorname{Re}\left\{\frac{2\alpha z\left(z^2 f^{''}(z)\right)'}{h(z)-h(-z)}+\frac{2z\left(zf^{'}(z)\right)'}{h(z)-h(-z)}\right\}>0. \tag{2}$$

We note that the class $K_s^*(\alpha,0)=K_s^*(\alpha)$ in Janteng [1] and $K_s^*(0,0)=K_s^*$ which was considered by Janteng, Abdul Halim and Darus in [2].

2 Preliminary Result

There is a preliminary lemma required for proving our results.

Lemma 1 ([3]) Let k be analytic in D with $\operatorname{Re}\{k(z)\}>0$ and be given by $k(z)=1+c_1z+c_2z^2+\cdots$ for $z\in D$, then

$$|c_n|\le 2 \qquad (n\ge 1)$$

and

$$\left|c_2-\frac{c_1^2}{2}\right|\le 2-\frac{|c_1|^2}{2}.$$

3 Main Result

Theorem 1 Let $f \in K_s^*(\alpha,\beta)$, $0 \le \alpha < 1$, $0 \le \beta < 1$, and be given by (1) then

$$|a_n| \le \frac{n-(n-2)\beta}{n^2[(n-1)\alpha+1]}$$

for $n = 2,3,4.$

Proof. Since $h \in S_s^*(\beta)$, it follows that

$$2zh'(z) = [\beta + (1-\beta)H(z)](h(z) - h(-z))$$

for $z \in D$,with $\operatorname{Re} H(z) > 0$ where $H(z) = 1 + p_1 z + p_2 z^2 + p_3 z^3 + \cdots$. Upon equating coefficients, we obtain

$$2b_2 = (1-\beta)p_1, \qquad 2b_3 = (1-\beta)p_2 \tag{3}$$

It follows from (2) that

$$2\alpha z(z^2 f''(z))' + 2z(zf'(z))' = (h(z) - h(-z))k(z) \tag{4}$$

where $\operatorname{Re}\{k(z)\} > 0$. Writing $k(z) = 1 + c_1 z + c_2 z^2 + c_3 z^3 + \cdots$ and equating coefficients in (4) gives

$$4(\alpha+1)a_2 = c_1, \qquad 9(2\alpha+1)a_3 = c_2 + b_3 \tag{5}$$

and

$$16(3\alpha+1)a_4 = c_3 + b_3 c_1$$

The result now follows on using classical inequalities $|p_n| \le 2, |c_n| \le 2, n \ge 2$, and the inequality $|b_3| \le 1-\beta$ which follow from (3).

Now we consider the functional $|a_3 - \mu a_2^2|$ for a complex μ.

Theorem 2 For $f \in K_s^*(\alpha,\beta)$, $0 \le \alpha < 1, 0 \le \beta < 1$ and μ complex,

$$\left|a_3 - \mu a_2^2\right| \le \frac{3-\beta}{9(2\alpha+1)} \max\left(1, \frac{4(1-\beta)(\alpha+1)^2 + \left|8(\alpha+1)^2 - 9(2\alpha+1)\mu\right|}{4(3-\beta)(\alpha+1)^2}\right)$$

The result obtained is sharp.

Proof. From (5), we write

$$a_3 - \mu a_2^2 = \frac{1}{9(2\alpha+1)}\left(c_2 - \frac{c_1^2}{2}\right) + \frac{\left\{8(\alpha+1)^2 + 9(2\alpha+1)\mu\right\}c_1^2}{144(2\alpha+1)(\alpha+1)^2} + \frac{(1-\beta)p_2}{18(2\alpha+1)} \tag{6}$$

It follows from (6) and Lemma 1 that

$$\left|a_3 - \mu a_2^2\right| \le \frac{1}{9(2\alpha+1)}\left|c_2 - \frac{c_1^2}{2}\right| + \frac{\left|8(\alpha+1)^2 + 9(2\alpha+1)\mu\right||c_1|^2}{144(2\alpha+1)(\alpha+1)^2} + \frac{(1-\beta)|p_2|}{18(2\alpha+1)} \tag{7}$$

$$\le \frac{1}{9(2\alpha+1)}\left\{2 - \frac{|c_1|^2}{2}\right\} + \frac{\left|8(\alpha+1)^2 + 9(2\alpha+1)\mu\right||c_1|^2}{144(2\alpha+1)(\alpha+1)^2} + \frac{(1-\beta)|p_2|}{18(2\alpha+1)}$$

$$= \frac{1}{9(2\alpha+1)} + \frac{\left\{\left|8(\alpha+1)^2 + 9(2\alpha+1)\mu\right| - 8(\alpha+1)^2\right\}|c_1|^2}{144(2\alpha+1)(\alpha+1)^2} + \frac{(1-\beta)|p_2|}{18(2\alpha+1)} \tag{8}$$

which, on using $|c_1| \le 2$ and $|p_2| \le 2$ gives

$$\left|a_3 - \mu a_2^2\right| \le \begin{cases} \dfrac{3-\beta}{9(2\alpha+1)}, & \text{if } \kappa(\alpha) \le 8(\alpha+1)^2 \\ \dfrac{1-\beta}{9(2\alpha+1)} + \dfrac{\left|8(\alpha+1)^2 - 9(2\alpha+1)\mu\right|}{36(2\alpha+1)(\alpha+1)^2} & \\ \text{if } \kappa(\alpha) \ge 8(\alpha+1) & \end{cases}$$

where $\kappa(\alpha) = \left|8(\alpha+1)^2 - 9(2\alpha+1)\mu\right|$.

Letting $c_1 = 0, c_2 = p_2 = 2$ and $c_1 = c_2 = p_2 = 2$ respectively in (6) shows that the results is sharp.

Next, we consider the real number μ as follows.

Theorem 3 For $f \in K_s^*(\alpha, \beta)$, $0 \le \alpha < 1$, $0 \le \beta < 1$ and μ real,

$$\left|a_3 - \mu a_2^2\right| \le \begin{cases} \dfrac{3-\beta}{9(2\alpha+1)} - \dfrac{\mu}{4(\alpha+1)^2}, & \text{if} \quad \mu \le 0 \\ \dfrac{3-\beta}{9(2\alpha+1)}, & \text{if} \quad 0 \le \mu \le \dfrac{16(\alpha+1)^2}{9(2\alpha+1)} \\ \dfrac{\mu}{4(\alpha+1)^2} - \dfrac{1+\beta}{9(2\alpha+1)}, & \text{if} \quad \mu \ge \dfrac{16(\alpha+1)^2}{9(2\alpha+1)} \end{cases}$$

The result obtained is sharp.

Proof. We consider two cases. At first, we suppose that $\mu \le \dfrac{8(\alpha+1)^2}{9(2\alpha+1)}$. From (8) and using the fact that $|c_1| \le 2$ and $|p_2| \le 2$, we obtain

$$\left|a_3 - \mu a_2^2\right| \le \begin{cases} \dfrac{3-\beta}{9(2\alpha+1)} - \dfrac{\mu}{4(\alpha+1)^2}, & \text{if} \quad \mu \le 0 \\ \dfrac{3-\beta}{9(2\alpha+1)}, & \text{if} \quad 0 \le \mu \le \dfrac{8(\alpha+1)^2}{9(2\alpha+1)} \end{cases}$$

Letting $c_1 = c_2 = p_2 = 2$ and $c_1 = 0, c_2 = p_2 = 2$ respectively in (6) shows that the results is sharp.

Next, we suppose that $\mu \ge \dfrac{8(\alpha+1)^2}{9(2\alpha+1)}$. In this case, it follows from (8) and Lemma 1 that

$$\left|a_3 - \mu a_2^2\right| \le \begin{cases} \dfrac{3-\beta}{9(2\alpha+1)}, & \text{if} \quad \dfrac{8(\alpha+1)^2}{9(2\alpha+1)} \le \mu \le \dfrac{16(\alpha+1)^2}{9(2\alpha+1)} \\ \dfrac{\mu}{4(\alpha+1)^2} - \dfrac{1+\beta}{9(2\alpha+1)}, & \text{if} \quad \mu \ge \dfrac{16(\alpha+1)^2}{9(2\alpha+1)} \end{cases}$$

Letting $c_1 = 0, c_2 = p_2 = 2$ and $c_1 = 2i, c_2 = -2, p_2 = 2$ respectively in (6) shows that the results is sharp.

References

1. Janteng, A.: Assortment of problem for certain classes of analytic functions. PhD Thesis. Kuala Lumpur, Universiti Malaya (2006)
2. Janteng, A., Halim, S. A., Darus, M.: Functions close-to-convex and quasi-convex with respect to other points. International Journal of Pure and Applied Mathematics 30, 225 – 236 (2006)
3. Pommerenke, Ch.: Univalent functions. Göttingen, Vandenhoeck and Ruprecht (1975)

L_1-consistent adaptive multivariate histograms from a randomized queue prioritized for statistically equivalent blocks

Gloria Teng[1], Jennifer Harlow[2], and Raazesh Sainudiin[2]

[1] Universiti Tunku Abdul Rahman, Kuala Lumpur 53300, Malaysia
gloriateng@utar.edu.my
[2] University of Canterbury, Christchurch 8041, New Zealand
jenny.harlow@canterbury.ac.nz
raazesh.sainudiin@gmail.com

Abstract. An L_1-consistent data-adaptive histogram estimator driven by a randomized queue prioritized by a statistically equivalent blocks rule is obtained. Such data-dependent histograms are formalized as real mapped regular pavings ($\mathbb{R}$-MRP). A regular paving (RP) is a binary tree obtained by selectively bisecting boxes along their first widest side. A statistical regular paving (SRP) augments an RP by mutably caching the recursively computable sufficient statistics of the data. Mapping a real value to each element of the partition gives an $\mathbb{R}$-MRP that can be used to represent a piecewise-constant function density estimate on a multidimensional domain. $\mathbb{R}$-MRPs are closed under addition and allow for efficient averaging of histograms with different partitions in any dimension. A partitioning strategy driven by a randomized priority queue of the current leaf nodes of an SRP is formalized as a Markov chain over the space of SRPs and the conditions for its L_1-consistency are obtained.

1 Introduction

Suppose our random variable X has an unknown density f on $\mathbb{R}^d$, then for all Borel sets $A \subseteq \mathbb{R}^d$,

$$\mu(A) := \Pr\{X \in A\} = \int_A f(x)dx \ .$$

Any density estimate $f_n(x) = f_n(x; X_1, X_2, \ldots, X_n) : \mathbb{R}^d \times \left(\mathbb{R}^d\right)^n \to \mathbb{R}$, is simply a map from $\left(\mathbb{R}^d\right)^{n+1}$ to $\mathbb{R}$. The objective in density estimation is to estimate the unknown f from an independent and identically distributed sample $X_1, X_2, \ldots, X_n$ drawn from f. This density estimate f_n of the unknown f gives us a means of computing the probabilities of any Borel set $A \in \mathcal{B}^d$ or of computing the density at any point $x \in \mathbb{R}^d$. Density estimation is often the first step in many learning tasks, including, classification, regression and clustering.

There are two general approaches to density estimation: parametric density estimation and nonparametric density estimation. Here we are concerned with nonparametric density estimation. Histograms and kernel density estimates are

the two most common forms of nonparametric density estimate for data assumed to be drawn from a continuous distribution. Both can be used for univariate and multivariate data. Other density estimators include orthogonal series estimators and nearest neighbour estimators [16, chap. 2]. Adaptations and specializations of these density estimation methods may be used for particular types of data, high-dimensional data, and very large data sets [15, 8]. However it is formed, the density estimate is some smoothed representation of the observed data [18]. The density estimation method determines how this smoothing is performed. Data-adaptive density estimation methods adapt the amount of smoothing to the local density of the data [16, chap. 2].

For a given prior distribution over SRPs, the posterior mean can be thought of as an L_2-loss minimizing Bayesian nonparametric density estimate. Such a Bayesian smoothing based on the sample mean of a sequence of histogram states visited by an MCMC algorithm with stationary samples from the posterior distribution was given in [12]. The crucial strategy to initialize the MCMC chain from states with high posterior probability, in order to minimize the chance that the chain gets stuck in low posterior states, was done in a an *ad hoc* manner in [12]. This paper proposes the use of an L_1-consistent and data-dependent tree-based histogram, that is built using a data structure known as statistical regular paving (SRP), as an initializing strategy for the MCMC in [12]. SRP is an extension of a regular paving (RP) [13, 6, 5], a class of space-partitioning trees that can facilitate efficient arithmetical operations. A real mapped regular paving ($\mathbb{R}$-MRP) is an extension of an RP designed to represent a piecewise-constant function and allow efficient arithmetical operations with them, including the averaging of $\mathbb{R}$-MRP histogram states with *different* partitions that are visited by an MCMC algorithm as in [12]. An SRP augments an RP by mutably caching recursively computable sufficient statistics of the data. A histogram density estimate represented as an $\mathbb{R}$-MRP can then be created from an SRP. Moreover, such histogram density estimates allow for a wide range of subsequent statistical operations, such as, creating marginal and conditional density estimates or evaluating the density estimate at a large set of query points, to be performed efficiently [5].

The paper is laid out as follows. Section 2 reviews various tree-based histogram estimators in the literature. Section 3 introduces the arithmetic and algebra for RPs, $\mathbb{R}$-MRPs, and SRPs, and explains how a histogram can be built using these structures. Section 4 illustrates the use of a randomized priority queue to partition the histogram and a proof of the L_1-consistency of this adaptive partitioning scheme. Section 5 concludes.

2 Tree-based histograms

A histogram is based on a partition of the data space; the elements of the partition are commonly known as bins. The choice of bin width(s) is the smoothing problem: wider bins give more smoothing, narrower bins less smoothing. The bins of a *regular* histogram are all equally-sized; the bins of an *irregular* histogram can vary in size.

Regular partitioning with small enough bins to suit the modes of the density will give too many bins in low or flat density areas [11]. Regular partitioning with a bin width more suited to the overall variability of the data may compromise the potential of the histogram to show important local features in the highest density areas. Multivariate histograms with a single bin width are not able to adapt to spatially varying smoothing requirements [7, chap. 17]. A data-dependent partition allows the bin width to vary in a way that is determined by the data. Data-dependent partitions can provide estimates which are theoretically superior to those using partitions based simply on the number of data points in the data set [17], and under certain conditions, a histogram density estimate with a data-dependent partition can be strongly L_1-consistent [9].

A tree structure can be used in algorithms for creating data-adaptive histograms. This is especially suitable where the algorithm uses some form of recursive partitioning strategy, often in association with a penalty function to control complexity. A *greedy* partitioning algorithm makes locally optimal decisions (with respect to the chosen optimality criterion) based on the immediately available information in each step but is not guaranteed to find a globally optimal solution. Several greedy data-adaptive tree-structured histogram algorithms have been developed, including methods that grow the tree (partition) step-by-step or that grow the tree to represent the most complex allowable partition and then use a greedy algorithm to prune to reduce the tree (reduce the number of elements in the partition. Partitioning trees can also be used in non-greedy complexity-penalized optimization algorithms that perform an exhaustive comparison of a limited set of possible partitions. For a discussion of such tree-based approaches see [7, chap. 17 & 18] and the references therein.

In general, computational efficiency of the methods described above suffer in two basic ways. First, they are not well-suited to very high-dimensional data because the computational complexity of most density estimation algorithms grows exponentially with the dimensions [14, chap. 7], irrespective of the complexity of the underlying density. Second, these methods cannot cope with large volumes of data, say with sample size n around 10^5 or 10^6, even in small dimensions, say dimension d up to 4 or 5. A Metropolis-Hastings Markov chain method was developed in [12] with stationary distribution given by a posterior distribution over regular paving histograms and used to estimate the posterior expectation by exploiting the arithmetic properties of regular pavings when averaging

histogram samples from the chain. The averaged regular paving histogram density estimate was tested with uniform data in up to $d = 1,000$ dimensions and found that the method coped well with this type of high-dimensional unstructured data. Results using data simulated from uniform mixture approximations to non-uniform structured densities such as multivariate Gaussian and Rosenbrock densities showed that the number of dimensions in which the method is computationally feasible with reasonably smaller mean integrated absolute errors is much lower, about $d = 5$ or $d = 6$. However, the method can computationally cope with large volumes of data, even n as high as 10^7, in stark contrast with other available methods.

Many conventional kernel density estimation methods are only effective with data in less than five or six dimensions [3, 19] with sample sizes around few thousands and generally reach computational bottle-necks when the sample size reaches 10^4. The posterior histogram estimate of [12] therefore has some attractions as a density estimation method in up to around five dimensions especially in situations where there is a large amount of sample data available and where it is advantageous to be able to carry out subsequent statistical operations efficiently, directly on the density estimate itself. Such statistical operations include (i) evaluating of the density over a large set of query points for cross-validation, (ii) obtaining the highest density or coverage regions, (iii) getting marginal densities as $\mathbb{R}$-MRPs by tree-based integration over a subset of the coordinates, or (iv) producing conditional densities as $\mathbb{R}$-MRPs for subsequent regression, according to the algorithms in [5].

3 Regular pavings and histograms

This section introduces the notions of RPs, SRPs, and $\mathbb{R}$-MRPs, and explains how a histogram density estimate can be built using these data structures.

3.1 Regular pavings (RPs)

Let $\boldsymbol{x} := [\underline{x}, \overline{x}]$ be a compact real interval with lower bound $\underline{x}$ and upper bound $\overline{x}$, where $\underline{x} \leq \overline{x}$. Let the space of such intervals be $\mathbb{IR}$. The width of an interval $\boldsymbol{x}$ is $\mathrm{wid}\,(\boldsymbol{x}) := \overline{x} - \underline{x}$. The midpoint is $\mathrm{mid}\,(\boldsymbol{x}) := (\underline{x} + \overline{x})/2$. A box of dimension d with coordinates in $\Delta := \{1, 2, \ldots, d\}$ is an interval vector:

$$\boldsymbol{x} := [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_d, \overline{x}_d] =: \bigboxtimes_{j \in \Delta} [\underline{x}_j, \overline{x}_j] \ .$$

The set of all such boxes is $\mathbb{IR}^d$, i.e., the set of all interval real vectors in dimension d. Consider a box $\boldsymbol{x}$ in $\mathbb{IR}^d$. Define the index ι to be the first coordinate of maximum width:

$$\iota := \min \left(\operatorname*{argmax}_i (\mathrm{wid}\,(\boldsymbol{x}_i)) \right) \ .$$

A *bisection* or *split* of $\boldsymbol{x}$ perpendicularly at the mid-point along this first widest coordinate ι gives the left and right child boxes of $\boldsymbol{x}$

$$\boldsymbol{x}_{\mathsf{L}} := [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_\iota, \operatorname{mid}(\boldsymbol{x}_\iota)) \times [\underline{x}_{\iota+1}, \overline{x}_{\iota+1}] \times \ldots \times [\underline{x}_d, \overline{x}_d] \ ,$$

$$\boldsymbol{x}_{\mathsf{R}} := [\underline{x}_1, \overline{x}_1] \times \ldots \times [\operatorname{mid}(\boldsymbol{x}_\iota), \overline{x}_\iota] \times [\underline{x}_{\iota+1}, \overline{x}_{\iota+1}] \times \ldots \times [\underline{x}_d, \overline{x}_d] \ .$$

Such a bisection is said to be *regular*. Note that this bisection gives the left child box a half-open interval $[\underline{x}_\iota, \operatorname{mid}(\boldsymbol{x}_\iota))$ on coordinate ι so that the intersection of the left and right child boxes is empty.

A recursive sequence of selective regular bisections of boxes, with possibly open boundaries, along the first widest coordinate, starting from the root box $\boldsymbol{x}$ in $\mathbb{IR}^d$ is known as a *regular paving* [6] or n-tree [13] of $\boldsymbol{x}$. A regular paving of $\boldsymbol{x}$ can also be seen as a binary tree formed by recursively bisecting the box $\boldsymbol{x}$ at the root node. Each node in the binary tree has either no children or two children. When the root box $\boldsymbol{x}$ is clear from the context we refer to an RP of $\boldsymbol{x}$ as merely an RP. Each node of an RP is associated with a sub-box of the root box that can be attained by a sequence of selective regular bisections. Each node in an RP can be distinctly labelled by the sequence of child node selections from the root node. We label these nodes and the associated boxes with strings composed of L and R for left and right, respectively. For example, in Figure 1, the root node associated with root box $\boldsymbol{x}_\rho$ is labeled ρ.

The relationship of trees, labels and partitions is illustrated in Figure 1 using a simple one-dimensional example. The root node associated with root interval $\boldsymbol{x}_\rho \in \mathbb{IR}$ is labelled ρ. First, ρ is split into two child nodes, and the left child and right child nodes are labelled $\rho\mathsf{L}$ and $\rho\mathsf{R}$, respectively. The left half of $\boldsymbol{x}_\rho$ that is now associated with node $\rho\mathsf{L}$ is labelled $\boldsymbol{x}_{\rho\mathsf{L}}$. Similarly, the right half of $\boldsymbol{x}_\rho$ that is associated with the right child node $\rho\mathsf{R}$ is labelled $\boldsymbol{x}_{\rho\mathsf{R}}$. $\rho\mathsf{L}$ and $\rho\mathsf{R}$ are a pair of *sibling nodes* since they share the same parent node ρ. A node with no child nodes is called a *leaf node*. A *cherry node* is a sub-terminal node with a pair of child nodes that are both leaves. This pair of sibling nodes can be *reunited* or *merged* to its parent cherry node ρ, thereby turning the cherry node into a leaf node.

Returning to Figure 1, the left node $\rho\mathsf{L}$ is split to get its left and right child nodes $\rho\mathsf{LL}$ and $\rho\mathsf{LR}$ with associated sub-intervals $\boldsymbol{x}_{\rho\mathsf{LL}}$ and $\boldsymbol{x}_{\rho\mathsf{LR}}$ respectively, formed by the bisection of interval $\boldsymbol{x}_{\rho\mathsf{L}}$ (because the root interval $\boldsymbol{x}_\rho$ is one-dimensional, each bisection is always on that single coordinate).

Let the j-th interval of a box $\boldsymbol{x}_{\rho\mathsf{v}}$ be $[\underline{x}_{\rho\mathsf{v},j}, \overline{x}_{\rho\mathsf{v},j}]$. The volume of a d-dimensional box $\boldsymbol{x}_{\rho\mathsf{v}}$ associated with the node $\rho\mathsf{v}$ of an RP of $\boldsymbol{x}_\rho$ is the product of the side-lengths of the box, that is, $\operatorname{vol}(\boldsymbol{x}_{\rho\mathsf{v}}) = \prod_{j=1}^{d} (\overline{x}_{\rho\mathsf{v},j} - \underline{x}_{\rho\mathsf{v},j})$.

The volume is associated with the depth of a node. The depth of a node $\rho\mathsf{v}$ in an RP is denoted by $\mathsf{d}_{\rho\mathsf{v}}$. A node has depth $\mathsf{d}_{\rho\mathsf{v}} = k$ in the tree if it can be reached by k splits from the root node. If an RP has root box $\boldsymbol{x}_\rho$ and a node $\rho\mathsf{v}$ in

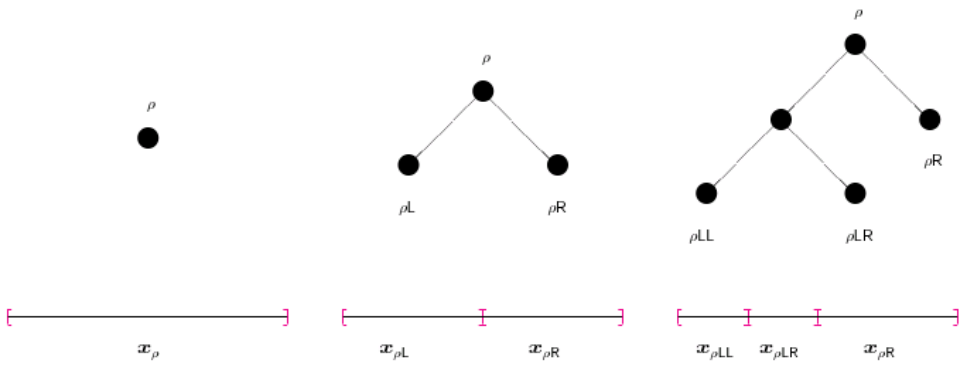


Fig. 1. A sequence of selective bisections, starting from the root, produces an RP.

the regular paving has depth k, then the volume of the box $\boldsymbol{x}_{\rho\mathsf{v}}$ associated with that node is $\mathrm{vol}\,(\boldsymbol{x}_{\rho\mathsf{v}}) = 2^{-k}\mathrm{vol}\,(\boldsymbol{x}_\rho)$. This is because any split always results in the child node's box having half the volume of the parent node's box.

In general, an RP is denoted by s. The set of all nodes of an RP is denoted by $\mathbb{V} := \rho \cup \{\rho\{\mathsf{L},\mathsf{R}\}^j : j \in \mathbb{N}\}$. The set of all leaf nodes of an RP is denoted by $\mathbb{L}$. The boxes associated with the leaf nodes of an RP are the partition of the root box $\boldsymbol{x}_\rho$. The set of leaf boxes of a regular paving s with root box $\boldsymbol{x}_\rho$ is denoted by $\boldsymbol{x}_{\mathbb{L}(s)}$. Let $\mathbb{S}_k$ be the set of all regular pavings with root box $\boldsymbol{x}_\rho$ made of k splits. Note that the number of leaf nodes $m = |\mathbb{L}(s)| = k+1$ if $s \in \mathbb{S}_k$.

The number of distinct binary trees with k splits is equal to the Catalan number C_k.

$$C_k = \frac{1}{k+1}\binom{2k}{k} = \frac{(2k)!}{(k+1)!(k!)} \quad . \tag{1}$$

For $i, j \in \mathbb{Z}_+$, where $\mathbb{Z}_+ := \{0, 1, 2, \ldots\}$ and $i \leq j$, let $\mathbb{S}_{i:j}$ be the set of regular pavings with k splits where $k \in \{i, i+1, \ldots, j\}$. The space of all regular pavings is then $\mathbb{S}_{0:\infty} := \lim_{j\to\infty} \mathbb{S}_{0:j}$. The size of the space of all regular pavings with between i and j splits, $|\mathbb{S}_{i:j}|$, is given by the sum of Catalan numbers:

$$|\mathbb{S}_{i:j}| = \sum_{k=i}^{j} C_k \quad . \tag{2}$$

The size of the space of all regular pavings with up to k splits is $|\mathbb{S}_{0:k}|$.

3.2 Real mapped regular pavings ($\mathbb{R}$-MRPs)

A real mapped regular paving ($\mathbb{R}$-MRP) is an extension of an RP. Let $s \in \mathbb{S}_{0:\infty}$ be an RP with root node ρ and root box $\boldsymbol{x}_\rho \in \mathbb{IR}^d$. Let ${}^{\square}f : \mathbb{V}(s) \to \mathbb{R}$ map each node of s to an element in $\mathbb{R}$ as follows:

$$\{\rho\mathsf{v} \mapsto f_{\rho\mathsf{v}} : \rho\mathsf{v} \in \mathbb{V}(s), f_{\rho\mathsf{v}} \in \mathbb{R}\} \quad .$$

Such a map ${}^{\square}f$ is called an $\mathbb{R}$-*mapped regular paving* ($\mathbb{R}$-MRP). Thus, an $\mathbb{R}$-MRP ${}^{\square}f$ is obtained by augmenting each node $\rho\mathsf{v}$ of the RP tree s with an additional data member $f_{\rho\mathsf{v}} \in \mathbb{R}$.

The sets of all nodes and leaf nodes of an $\mathbb{R}$-MRP ${}^{\square}f$ are denoted by $\mathbb{V}({}^{\square}f)$ and $\mathbb{L}({}^{\square}f)$, respectively. The set of all leaf node boxes is denoted by $\boldsymbol{x}_{\mathbb{L}({}^{\square}f)}$. The class of $\mathbb{R}$-MRPs over the leaf boxes of regular pavings of a root box $\boldsymbol{x}_\rho \in \mathbb{IR}^d$ is then

$$ {}^{\square}\mathcal{F} := \{\{\rho\mathsf{v} \mapsto f_{\rho\mathsf{v}} : \rho\mathsf{v} \in \mathbb{V}(s), f_{\rho\mathsf{v}} \in \mathbb{R}\} : s \in \mathbb{S}_{0:\infty}\} $$

Arithmetic operations in $\mathbb{R}$ can be extended to $\mathbb{R}$-MRPs [5]. For example, given any two $\mathbb{R}$-MRPs ${}^{\square}f^{(1)}$ and ${}^{\square}f^{(2)}$ with the same root box $\boldsymbol{x}_\rho$ and a binary operation $\star \in \{+, -, \cdot, /\}$, the $\mathbb{R}$-MRP ${}^{\square}f = {}^{\square}f^{(1)} \star {}^{\square}f^{(2)}$ can be obtained. An $\mathbb{R}$-MRP ${}^{\square}f$ can also be transformed using any standard function $\tau \in \mathfrak{S} := \{\exp, \sin, \cos, \tan, \ldots\}$ to obtain the $\mathbb{R}$-MRP $\tau({}^{\square}f)$. Finally, a binary operation of the form ${}^{\square}f \star x$ for an $\mathbb{R}$-MRP ${}^{\square}f$ and $x \in \mathbb{R}$ can also be carried out, and again the result ${}^{\square}g = {}^{\square}f \star x$ is an $\mathbb{R}$-MRP. All these properties are used to show that ${}^{\square}\mathcal{F}$ satisfies the conditions of a Stone-Weierstrass theorem and therefore dense in $\mathcal{C}(\boldsymbol{x}_\rho, \mathbb{R})$, the algebra of real-valued continuous functions over $\boldsymbol{x}_\rho$ [5]. This ensures that we can uniformly approximate any continuous density $f : \boldsymbol{x}_\rho \to \mathbb{R}$ using $\mathbb{R}$-MRPs in ${}^{\square}\mathcal{F}$.

$\mathbb{R}$-MRPs are important structures in this paper because an $\mathbb{R}$-MRP can be used to represent a piecewise-constant function. [5] describes function approximation using $\mathbb{R}$-MRPs in general. The advantage of an $\mathbb{R}$-MRP representation is that all the arithmetic operations between real-valued simple functions in ${}^{\square}\mathcal{F}$ described above can be carried out efficiently and recursively using trees. A box in any type of RP has real volume ($\mathrm{vol}\,(\boldsymbol{x}_{\rho\mathsf{v}}) \in \mathbb{R}$). This allows operations using both node volume and node value, such as integrating, normalizing and marginalizing, to be carried out on $\mathbb{R}$-MRPs. A *non-negative* $\mathbb{R}$-MRP ${}^{\square}f$ can be used to represent a (possibly non-normalized) probability density function. An $\mathbb{R}$-MRP ${}^{\square}f$ is non-negative if $f_{\rho\mathsf{v}} \geq 0 \ \forall \rho\mathsf{v} \in \mathbb{L}({}^{\square}f)$. Figure 2 shows an $\mathbb{R}$-MRP density estimate of an example density which is a mixture of two bivariate Normal densities for $x \in \mathbb{R}^2$.

3.3 Statistical regular pavings (SRPs)

A *statistical regular paving* (SRP) is an extension of the RP structure that is able to act as a partitioned 'container' and responsive summarizer for multivariate data. An SRP can be used to create a histogram of a data set. An SRP is effectively an association of a collection of data (the *data sample* or *data set*) with an RP-based structure where the nodes have additional properties:

– A node of an SRP tree can be associated with a subset of the sample data;

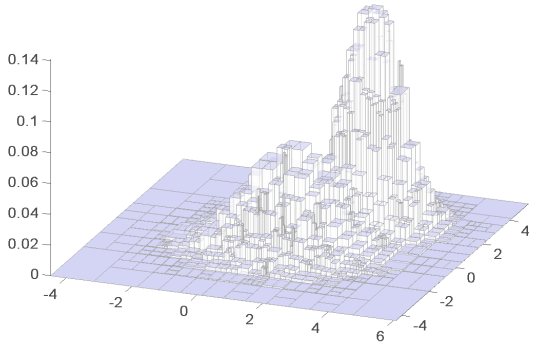


Fig. 2. $\mathbb{R}$-MRP density estimate of a bivariate Gaussian mixture.

– A node of an SRP tree records recursively computable statistics relating to this sample subset.

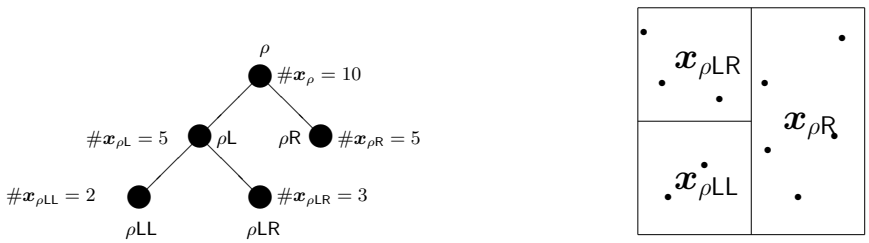
An SRP is denoted by s. Denote $\mathbb{S}_{i:j}$ as the set of all statistical regular pavings with a given root box and k splits where $k \in \{i, i+1, \ldots, j\}$, where $i, j \in \mathbb{Z}_+$ and $i \leq j$. The space of all statistical regular pavings with a given root box is then $\mathbb{S}_{0:\infty} := \lim_{i \to \infty} \mathbb{S}_{0:i}$

Take a data sample of size n, $X_1, X_2, \ldots, X_n$ and an SRP s. For convenience the sample will be referred to as ${}^{n}X$. Let ${}^{\subset n}X$ be a subset of ${}^{n}X$ and let ${}^{\subset n}X_{\rho\mathsf{v}}$ be the subset of ${}^{n}X$ contained in the box $\boldsymbol{x}_{\rho\mathsf{v}}$ associated with a node $\rho\mathsf{v}$ in s.

A recursively computable statistic of some data is a statistic whose value can be updated from the addition of new data using only the current value of the statistic and the new data (i.e., it is not necessary to know the individual data values from which the current value of the statistic is calculated). Formally, if $T({}^{\subset n}X)$ is some statistic of ${}^{\subset n}X$ and a new data point x is added to ${}^{\subset n}X$ so that $n' = n + 1$ and ${}^{\subset n'}X = {}^{\subset n}X \cup x$, then $T({}^{\subset n'}X)$ can be calculated using $u(T({}^{\subset n}X), x)$ where u is some updating function.

For the purpose of this paper, the only statistic that an SRP node $\rho\mathsf{v}$ is required to keep is the count of the number of data points in ${}^{\subset n}X_{\rho\mathsf{v}}$. This count is denoted by $\#\boldsymbol{x}_{\rho\mathsf{v}} = |{}^{\subset n}X_{\rho\mathsf{v}}|$. A leaf node $\rho\mathsf{v}$ with $\#\boldsymbol{x}_{\rho\mathsf{v}} > 0$ is a non-empty leaf node. The set of non-empty leaves of an SRP s is $\mathbb{L}^+(s) := \{\rho\mathsf{v} \in \mathbb{L}(s) : \#\boldsymbol{x}_{\rho\mathsf{v}} > 0\} \subseteq \mathbb{L}(s)$.

Figure 3 depicts a small SRP s with root box $\boldsymbol{x}_\rho \in \mathbb{IR}^2$. The number of sample data points in the root box $\boldsymbol{x}_\rho$ is 10. Figure 3(a) shows the tree, including the count associated with each node in the tree. Figure 3(b) shows the partition of the root box represented by this tree, with the sample data points superimposed on the box.



(a) SRP tree. (b) SRP paving.

Fig. 3. A small SRP.

3.4 Statistical regular paving (SRP) histograms as $\mathbb{R}$-MRPs

Given the count data recorded by each node, an SRP associated with data ${}^nX$ can be used to form a histogram. The bins are the elements in the partition, i.e., the boxes associated with the leaf nodes $\boldsymbol{x}_{\mathbb{L}(s)}$. If the total number of data points associated with the whole of an SRP s with root node ρ and root box $\boldsymbol{x}_\rho$ is $n = \#\boldsymbol{x}_\rho = \sum_{\rho\mathsf{v}\in\mathbb{L}(s)} \#\boldsymbol{x}_{\rho\mathsf{v}}$, then the corresponding histogram is:

$$\hat{f}_n(x) = \sum_{\rho\mathsf{v}\in\mathbb{L}(s)} \frac{\mathbb{1}_{\boldsymbol{x}_{\rho\mathsf{v}}}(x)}{n} \left(\frac{\#\boldsymbol{x}_{\rho\mathsf{v}}}{\mathrm{vol}\,(\boldsymbol{x}_{\rho\mathsf{v}})} \right) \quad . \tag{3}$$

A histogram obtained using Equation (3) is referred to as an SRP histogram. SRP histograms have some similarities to dyadic histograms [7, chap. 18]. Both are binary tree-based and partition so that a box may only be bisected at the mid-point of one of its coordinates, but the RP structure restricts partitioning further by only bisecting a box on its first widest coordinate in order to make ${}^\square\mathcal{F}$ closed under addition and scalar multiplication and thereby allow for computationally efficient averaging of histograms with different partitions.

This SRP histogram is a piecewise-constant function that can be represented as an $\mathbb{R}$-MRP in ${}^\square\mathcal{F}$. Thus all the $\mathbb{R}$-MRP operations described above can be carried out with the $\mathbb{R}$-MRP histogram density estimate formed from the SRP. Section 4 discusses how the partitioning of the root box of an SRP can be carried out in a data-adaptive and asymptotically L_1-consistent manner.

4 Randomized priority queue for adaptive partitioning

A randomized priority queue (RPQ) partitioning method orders the elements of $\mathbb{L}^\triangledown(s)$, the splittable leaf nodes of an SRP s, according to some priority function $\psi : \mathbb{L}^\triangledown(s) \to \mathbb{R}$, in order to select the next node to be split from

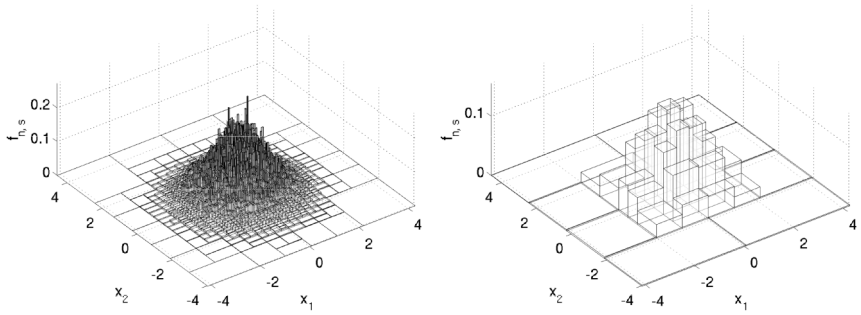


Fig. 4. Two histogram density estimates for the standard bivariate Gaussian density. The left figure shows a histogram with 1485 leaf nodes where $\overline{\#} = 50$ and the histogram on the right has $\overline{\#} = 1500$ resulting in 104 leaf nodes.

$\text{argmax}_{\rho\mathsf{v}\in\mathbb{L}^{\triangledown}(s)}\,\psi(\rho\mathsf{v})$, the set of splittable leaf nodes of s which are equally 'large' when measured using ψ. If there is more than one such 'largest' node the choice is made uniformly at random from this set; this is the 'randomized' aspect of the process. Two criteria can be specified to stop the RPQ partitioning. A straightforward stopping condition is to stop partitioning when the number of leaves in the SRP reaches a specified maximum $\overline{m}$. The other stopping condition relates to the priority function so that partitioning stops when the value of the largest node under the priority function ψ is less than or equal to a specified value $\overline{\psi}$. An RPQ will also stop partitioning if there are no splittable leaf nodes in the SRP.

The RPQ process generates a sequence of states $\{S(t)\}_{t\in\mathbb{Z}_+}$ on $\mathbb{S}_{1:\overline{m}-1}$. If the initial state $S(t=0)$ is the root $s \in \mathbb{S}_0$ then this can be seen as a sequence $\{S(k)\}_{k\in\mathbb{Z}_+}$ on $\mathbb{S}_{0:\overline{m}-1}$ such that $S(k) \in \mathbb{S}_k$, i.e. the $(k+1)^{\text{th}}$ state has $k+1$ leaves or k splits.

A statistically equivalent block (SEB)-based SRP partitioning scheme driven by an RPQ can be used to create a final SRP where each leaf node has at most $\overline{\#}$ of the sample data points associated with it and the total number of leaves is at most $\overline{m}$. The SEB priority function is given by $\psi(\rho\mathsf{v}) = \#\boldsymbol{x}_{\rho\mathsf{v}}$.

Figure 4 shows two different SRP histograms constructed using two different values of $\overline{\#}$ for the same dataset of $n = 10^5$ points simulated under the standard bivariate Gaussian density. A small $\overline{\#}$ produces a histogram that is under-smoothed with unnecessary spikes (left) while the other histogram with a larger $\overline{\#}$ used as the SEB stopping criterion is over-smoothed (right). An approach to solve the smoothing problem by a Bayesian MCMC that exploits the efficient averaging of $\mathbb{R}$-MRP histograms in ${}^{\square}\mathcal{F}$ was proposed in [12] (as discussed in Section 2). The crucial initialization strategy for the MCMC was not justified in [12]. It is proved here to be L_1-consistent.

We now show that an RMRP density estimate based on an SRP created using the SEB RPQ partitioning scheme is asymptotically L_1-consistent provided that $\overline{\#}$ and $\overline{m}$ grow with the sample size n at appropriate rates. This is done by proving the three conditions in Theorem 1 of [9]. We will need to show that as the number of sample points increases linearly, the following conditions are met:

1. the number of leaf boxes grows sub-linearly;
2. the partition grows sub-exponentially in terms of a combinatorial complexity measure;
3. and the volume of the leaf boxes in the partition are shrinking.

Let $\{S_n(i)\}_{i=0}^{\dot{I}}$ on $\mathbb{S}_{0:\infty}$ be the Markov chain formed using SEB RPQ. The Markov chain terminates at some state $\dot{s}$ with partition $\ell(\dot{s})$. Associated with the Markov chain is a fixed collection of partitions

$$\mathcal{L}_n := \Big\{\ell(\dot{s}) : \dot{s} \in \mathbb{S}_{0:\infty}, Pr\{S(\dot{I}) = \dot{s}\} > 0\Big\}$$

and the size of the largest partition $\ell(\dot{s})$ in $\mathcal{L}_n$ is given by

$$m(\mathcal{L}_n) := \sup_{\ell(\dot{s})\in\mathcal{L}_n} |\ell(\dot{s})| \leq \overline{m}$$

such that $\mathcal{L}_n \subseteq \{\ell(s) : s \in \mathbb{S}_{0:\overline{m}-1}\}$.

Given n fixed points $\{X_1, \ldots, X_n\} \in \left(\mathbb{R}^d\right)^n$. Let $\mathit{\Pi}\left(\mathcal{L}_n, \{X_1, \ldots, X_n\}\right)$ be the number of distinct partitions of the finite set $\{X_1, \ldots, X_n\}$ that are induced by partitions $\ell(\dot{s}) \in \mathcal{L}_n$:

$$\mathit{\Pi}(\mathcal{L}_n, \{X_1, \ldots, X_n\}) := |\{\{\boldsymbol{x}_{\rho v} \cap \{X_1, \ldots, X_n\} : \boldsymbol{x}_{\rho v} \in \ell(\dot{s})\} : \ell(\dot{s}) \in \mathcal{L}_n\}| \ .$$

For any fixed set of n points, the growth function of $\mathcal{L}_n$ is then

$$\mathit{\Pi}^*(\mathcal{L}_n, \{X_1, \ldots, X_n\}) = \max_{\{X_1,\ldots,X_n\}\in\left(\mathbb{R}^d\right)^n} \mathit{\Pi}(\mathcal{L}_n, \{X_1, \ldots, X_n\}) \ .$$

Let $A \subseteq \mathbb{R}^d$. Then the diameter of A is the maximum Euclidean distance between any two points of A, i.e., $\mathrm{diam}(A) := \sup_{x,y\in A} \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$. Thus, for a box $\boldsymbol{x} = [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_d, \overline{x}_d]$, $\mathrm{diam}(\boldsymbol{x}) = \sqrt{\sum_{i=1}^d (\overline{x}_i - \underline{x}_i)^2}$.

We now check the three conditions for L_1-consistency of the histogram estimate constructed using SEB RPQ.

Theorem 1 (L_1-Consistency). *Let $X_1, X_2, \ldots$ be independent and identical random vectors in $\mathbb{R}^d$ whose common distribution μ has a non-atomic density f, i.e., $\mu \ll \lambda$. Let $\{S_n(i)\}_{i=0}^{\dot{I}}$ on $\mathbb{S}_{0:\infty}$ be the Markov chain formed using SEB*

RPQ with terminal state $\dot{s}$ and histogram estimate $f_{n,\dot{s}}$ over the collection of partitions $\mathcal{L}_n$. As $n \to \infty$, if $\overline{\#} \to \infty$, $\overline{\#}/n \to 0$, $\overline{m} \geq n/\overline{\#}$, and $\overline{m}/n \to 0$ then the density estimate $f_{n,\dot{s}}$ is strongly consistent in L_1, i.e.

$$\int |f(x) - f_{n,s}(x)|dx \to 0 \text{ with probability } 1.$$

Proof. We will assume that $\overline{\#} \to \infty$, $\overline{\#}/n \to 0$, $\overline{m} \geq n/\overline{\#}_n$, and $\overline{m}/n \to 0$, as $n \to \infty$, and show that the three conditions:

(a) $n^{-1}m(\mathcal{L}_n) \to 0$,
(b) $n^{-1} \log \Pi^*_n(\mathcal{L}_n) \to 0$, and
(c) $\mu(x : \text{diam}(\boldsymbol{x}(x)) > \gamma) \to 0$ with probability 1 for every $\gamma > 0$,

are satisfied. Then by Theorem 1 of Lugosi and Nobel (1996) our density estimate $f_{n,\dot{s}}$ is strongly consistent in L_1.

Condition (a) is satisfied by the assumption that $\overline{m}/n \to 0$ since $m(\mathcal{L}_n) \leq \overline{m}$ (see Remark 1).

The largest number of distinct partitions of any n point subset of $\mathbb{R}^d$ that are induced by the partitions in $\mathcal{L}_n$ is upper bounded by the size of the collection of partitions $\mathcal{L}_n \subseteq \mathbb{S}_{0:\overline{m}-1}$, i.e.

$$\Pi^*_n(\mathcal{L}_n) \leq |\mathcal{L}_n| \leq \sum_{i=0}^{\overline{m}_n - 1} C_i$$

where i is the number of splits.

The growth function is thus bounded by the total number of partitions with 0 to $\overline{m}_n - 1$ splits, i.e. the $(\overline{m}_n - 1)$-th partial sum of the Catalan numbers. The partial sum can be asymptotically approximated as ([10]):

$$\sum_{k=0}^{\overline{m}-1} C_k \to \frac{4^{\overline{m}}}{\left(3(\overline{m}-1)\sqrt{\pi(\overline{m}-1)}\right)} \quad \text{as } \overline{m} \to \infty \ .$$

Taking logs and dividing by n on both sides we get

$$\begin{aligned} \log \Delta^*_n(L_n)/n &\leq \log \left(\frac{4^{(m(\mathcal{L}_n)+1)}}{3m(\mathcal{L}_n)\sqrt{\pi m(\mathcal{L}_n))}} \right) /n \\ &\leq \tfrac{1}{n}(m(\mathcal{L}_n)+1)\log 4 - \tfrac{1}{n}\log 3\sqrt{(\pi)} - \tfrac{3}{2n}\log m(\mathcal{L}_n). \end{aligned}$$

The first and third term goes to 0 by an application of condition (a). The second term which is just a constant divided by n also vanishes as $n \to \infty$. Therefore, condition (b) is satisfied.

We now prove the final condition. Fix $\gamma, \xi > 0$. There exists a box $\hat{\boldsymbol{x}} = [-M, M]^d$ for a large enough M, such that, $\mu(\hat{\boldsymbol{x}}^c) < \xi$. Consequently,

$$\mu(\{x : \text{diam}(\boldsymbol{x}(x)) > \gamma\}) \leq \xi + \mu(\{x : \text{diam}(\boldsymbol{x}(x)) > \gamma\} \cap \hat{\boldsymbol{x}}).$$

Using 2^{di} hypercubes of equal volume $(2M)^d/2^{di}, i = \left\lceil \log_2 \left(2M\sqrt{d}/\gamma \right) \right\rceil$ with side length $2M/2^i$ and diameter $\sqrt{d(\frac{2M}{2^i})^2}$, we can have at most 2^{di} boxes in the interior of $\hat{\mathbb{X}}$ and δ boxes at the lower dimensional boundaries of $\hat{\mathbb{X}}$, i.e. there are at most m_γ disjoint boxes in $\hat{\boldsymbol{x}}$ that have diameter greater than γ, where

$$m_\gamma < 2^{di} + \delta, \quad \delta = \left(2^d + \sum_{j=1}^{d-1} 2^{d-j} \binom{d}{j} 2^{ij} \right) \ . \tag{4}$$

By choosing i large enough we can upper bound m_γ by $(2M\sqrt{d}/\gamma)^d + 2^d + \sum_{j=1}^{d-1} 2^{d-j} \binom{d}{j} (2M\sqrt{d}/\gamma)^j$, a quantity that is independent of n, such that

$$\begin{aligned}
\mu(x : \text{diam}(\boldsymbol{x}(x)) > \gamma) &\leq \xi + \mu\left(\{x : \text{diam}(\boldsymbol{x}(x)) > \gamma\} \cap \hat{\boldsymbol{x}}\right) \\
&\leq \xi + m_\gamma \left(\max_{\boldsymbol{x} \in \ell(\dot{s})} \mu(\boldsymbol{x}) \right) \\
&\leq \xi + m_\gamma \left(\max_{\boldsymbol{x} \in \ell(\dot{s})} \mu_n(\boldsymbol{x}) + \max_{\boldsymbol{x} \in \ell(\dot{s})} |\mu(\boldsymbol{x}) - \mu_n(\boldsymbol{x})| \right) \\
&\leq \xi + m_\gamma \left(\frac{\overline{\#}}{n} + \sup_{\boldsymbol{x} \in \mathbb{IR}^d} |\mu(\boldsymbol{x}) - \mu_n(\boldsymbol{x})| \right).
\end{aligned}$$

The first term in the parenthesis converges to zero since $\overline{\#}/n \to 0$ by assumption. For $\epsilon > 0$, the second term goes to zero by applying the Vapnik-Chervonenkis theorem to boxes in $\mathbb{IR}^d$ with shatter coefficient $s(\mathbb{IR}^d, n) = 2^{2d}$ [1, p. 220], i.e.

$$Pr\left\{ \sup_{\boldsymbol{x} \in \mathbb{IR}^d} |\mu_n(\boldsymbol{x}) - \mu(\boldsymbol{x})| > \epsilon \right\} \leq 8 \cdot 2^{2d} \cdot e^{-n\epsilon^2/32} \ .$$

By the Borel-Cantelli lemma,

$$\lim_{n \to \infty} \sup_{\boldsymbol{x} \in \mathbb{IR}^d} |\mu_n(x) - \mu(x)| = 0 \quad \text{w.p. } 1 \ .$$

Thus for any $\gamma, \xi > 0$,

$$\limsup_{n \to \infty} \mu(\{x : \text{diam}(\boldsymbol{x}(x)) > \gamma\}) \leq \xi.$$

Therefore, condition (c) is satisfied and this completes the proof.

Remark 1. We can choose $\overline{\#}$ to be some sub-linear function of n, say n^α. Then $\alpha > 0$ so that $\overline{\#} \to \infty$ and $\alpha < 1$ so that $\overline{\#}/n \to 0$. Now let $\overline{m} = n^\beta$, then $\beta > 0$ so that $\overline{m} \geq n/\overline{k}_n$. The above constraints imply that $\alpha + \beta \geq 1$. Finally $\beta < 1$ such that $\overline{m}/n \to 0$.

5 Conclusions

In this paper we formalized the RP data structure and its extensions, SRP and $\mathbb{R}$-MRP, and showed that by using an SEB RPQ partitioning scheme, an L_1-consistent adaptive histogram can be obtained. This can be used to initialize MCMC as in [12] to obtain Bayesian smoothed density estimates.

Note that the regular paving structure places some restrictions on the density estimate due to the way the bisections are selected, but has the advantage of allowing a wide range of statistical operations to be carried out efficiently on piecewise-constant density estimate from the dense class of ${}^{\square}\mathcal{F}$, real mapped regular pavings [5]. In addition, a collection of histogram density estimates from ${}^{\square}\mathcal{F}$ with different partitions in any dimension can be efficiently averaged even with large sample sizes. Up to a given prior distribution, the resulting nonparametric Bayesian density estimate is a smoothed $\mathbb{R}$-MRP representation of the posterior sample mean with lower mean integrated absolute error than any particular $\mathbb{R}$-MRP histogram state visited by the MCMC algorithm [12]. A major advantage of the SEB-based RPQ algorithm is that, run for a limited number of states, it can be an effective way to create an over-smoothed histogram that reflects the major features of the density of the sample data. This gives an SRP that provides a much better starting point for further data-adaptive partitioning than the original, unpartitioned, SRP. By partitioning deeper into the state space with small $\overline{\#}$ we can find histograms with higher posterior densities along the asymptotically consistent path taken by the SEB-based RPQ Markov chain. Such high posterior states can be used to initialize the MCMC algorithm of [12] and thereby minimize the mixing time as done in [4, chap. 6].

However, a drawback of the SRP RPQ algorithm as a density estimator is that, without some idea of the characteristics of the density to be estimated, it is extremely hard to determine suitable values for the parameters controlling the partitioning process. As with the other greedy algorithms discussed in Section 2, the locally optimal choices made by an RPQ algorithm may be globally suboptimal. An SEB-based RPQ has nevertheless been shown to be able to produce an asymptotically consistent density estimate. Cross-validation or minimum distance estimation [2, chap. 6], or other smoothing techniques, could potentially be used with SRP RPQs to produce $\mathbb{R}$-MRP density estimates. These possibilities are currently being explored.

Acknowledgements

This research was partly supported by RS's external consulting revenues from the New Zealand Ministry of Tourism, University of Canterbury (UC) College of Engineering Sabbatical Grant in 2014 and UC MSc Scholarship to JH.

References

1. Luc Devroye, László Györfi, and G'aabor Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, New York, 1996.
2. Luc Devroye and G'aabor Lugosi. *Combinatorial Methods in Density Estimation.* Springer-Verlag, New York, 2001.
3. Alexander G Gray and Andrew W Moore. Nonparametric Density Estimation: Towards Computational Tractability. In *SIAM International Conference on Data Mining*, pages 203–211. SIAM, 2003.
4. J. Harlow. Data-adaptive multivariate density estimation using regular pavings, with applications to simulation-intensive inference. Master's thesis, University of Canterbury, 2013.
5. J. Harlow, R. Sainudiin, and W. Tucker. Mapped regular pavings. *Reliable Computing*, 16:252–282, 2012.
6. M. Kieffer, L. Jaulin, I. Braems, and E. Walter. Guaranteed set computation with subpavings. In W. Kraemer and J.W. Gudenberg, editors, *Scientific Computing, Validated Numerics, Interval Methods, Proceedings of SCAN 2000*, pages 167–178. Kluwer Academic Publishers, New York, 2001.
7. Jussi Klemelä. *Smoothing of Multivariate Data: Density Estimation and Visualization.* Wiley, Chichester, United Kingdom, 2009.
8. Dongryeol Lee and Alexander Gray. Fast High-Dimensional Kernel Summations Using the Monte Carlo Multipole Method. In *Advances in Neural Information Processing Systems (NIPS), 21 (2008)*, pages 929–936. MIT Press, 2009.
9. Gábor Lugosi and Andrew Nobel. Consistency of Data-Driven Histogram Methods for Density Estimation and Classification. *The Annals of Statistics*, 24(2):687–706, 1996.
10. S. Mattarei. Asymptotics of partial sums of central binomial coefficients and Catalan numbers. arXiv.0906.4290v3, January 2010.
11. Jorma Rissanen, TP Speed, and Bin Yu. Density Estimation by Stochastic Complexity. *IEEE Transactions on Information Theory*, 38(2):315–323, 1992.
12. R. Sainudiin, G. Teng, J. Harlow, and D. S. Lee. Posterior expectation of regularly paved random histograms. *ACM Transactions on Modeling and Computer Simulation*, 23(26), 2013.
13. H. Samet. *The Design and Analysis of Spatial Data Structures.* Addison-Wesley Longman, Boston, 1990.
14. David W. Scott. *Multivariate Density Estimation.* Wiley, New York, 1992.
15. David W Scott and Stephan R Sain. Multidimensional Density Estimation. In C. R. Rao, E. J. Wegman, and J. L. Solka, editors, *Handbook of Statistics*, volume 24, chapter 9, pages 229–262. Elsevier, Amsterdam, The Netherlands, 2005.
16. B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.
17. Charles J Stone. An Asymptotically Optimal Histogram Selection Rule. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II*, pages 513–520, Belmont, CA, 1985. Wadsworth.
18. P. Whittle. On the Smoothing of Probability Density Functions. *Journal of the Royal Statistical Society . Series B (Methodological)*, 20(2):334–343, 1958.
19. Xibin Zhang, Maxwell L. King, and Rob J. Hyndman. A Bayesian Approach to Bandwidth Selection for Multivariate Kernel Density Estimation. *Computational Statistics & Data Analysis*, 50(11):3009–3031, July 2006.

Semiparametric inference based on weighted estimating equations for additive hazards model with covariates missing at random

Xiaolin Chen and Yunquan Song

China University of Petroleum, Qingdao 266580, P.R. China
xlchen@amss.ac.cn and syqfly1980@163.com

Abstract. In this paper, the statistical inference for additive hazards model with some covariates missing at random is considered. The nonparametric kernel smoothing techniques based weighted estimating equation method (Qi, et al., 2005) is applied to the ad hoc estimating equations (Lin and Ying, 1994). It is shown that the simple inverse probability weighted estimator with nonparametrically estimated nonmissingness probability and augmented inverse probability weighted estimator with the nonparametrically estimated or true nonmissingness probability have the same efficiency. And they are more efficient than the simple inverse probability weighted estimator with true nonmissingness probability. The finite sample performance of the methods established in this article is evaluated through extensive simulation studies, which also verified our findings. The Mouse Leukemia data is used further to illustrate our proposed methods.

1 Introduction

In the analysis of survival data, it is common to encounter the situation, in which some covariates are missing due to various reasons. For example, in the case-cohort study, covariates are assembled only for the cases and a subcohort to reduce the cost. To analyze the survival data with missing covariates, the Cox's proportional hazards model (Cox, 1972) is most frequently employed, in which the conditional hazard function is specified by a product of an unknown baseline hazard function and an exponential regression. Statistical inference for survival data with missing covariates under the Cox's proportional hazards model has been investigated by Wang and Chen (2001), Qi et al. (2005), Xu et al. (2009), Luo et al. (2009), Cook et al. (2011) etc. Among these methods, Qi et al. (2005) proposed a weighted estimating equation method based on the inverse probability weighting idea (Horvitz and Thompson, 1952) and the nonparametric kernel smoothing techniques. This method does not require to specify the model for the nonmissingness probability and any conditional distribution. So it is rather robust in real application.

However, the popularity of Cox's proportional hazards model is not only from its utility and wide applications, but also from convention and the availability of software. Sometimes the Cox's proportional hazards model may not fit the data well. Let T and Z be the survival time and covariates respectively.

An important alternative model is the additive hazards model (Lin and Ying, 1994), which assumes the conditional hazard function of T given Z as

$$\lambda(t|Z) = \lambda_0(t) + \beta_0^T Z, \tag{1}$$

where $\lambda_0(t)$ is the true baseline hazard function and β_0 is the true regression coefficients.

For survival data with fully observed covariates, Lin and Ying (1994) proposed a pseudoscore estimating equation method in an ad hoc fashion by mimicking the partial likelihood score equation of the Cox's proportional hazards model. Some literature, such as Kulich and Lin (2000a, 2000b), Jiang and Zhou (2007) and so on, proposed methods to deal with the situation when some covairates are missing with design. Recently, Lin (2011) applied the simple and augmented inverse probability methods to the pseudoscore estimating equation of Lin and Ying (1994) to handle the situation whether some covariates are missing by design or chance. But it is noted that the methods of Lin (2011) need to specify a parametric model of nonmissingness probability and some conditional distributions which are hard to be validated. In this paper, we will overcome this drawback by applying the nonparametric kernel smoothing techniques.

Denote the censoring variable to be C and assume that T and C are conditional independent given Z. Let $X = \min(T, C)$ and $\delta = I(T \leq C)$. Because some covariates are missing sometimes, we decompose Z as (Z_o, Z_m), where Z_o represents covariates that are always observed and Z_m represents covariates that are missing sometimes. Define $\xi = 1$ if Z_m is observed and $\xi = 0$ otherwise. Throughout this paper, we assume that Z_m is missing at random (MAR), i.e., $Pr(\xi = 1|X, \delta, Z) = Pr(\xi = 1|X, \delta, Z_o) \equiv \pi(W)$, where $W = (X, \delta, Z_o)$.

The remainder of this article is organized as follows. Simple and augmented inverse probability weighted estimating equations methods with true and nonparametrically estimated nonmissingness probability are established in Sections 2 and 3 respectively. Section 4 reports the simulation studies, while analysis of the mouse leukemia study is given in Section 5. The regularity conditions and proofs of the theorems are given in the Appendix.

2 Simple inverse probability weighted estimators

Suppose that the observed data are $\{(X_i, \delta_i, Z_{i,o}, \xi_i, \xi_i Z_{i,m}), i = 1, \ldots, n\}$. Define $M_i(t) = N_i(t) - \int_0^t Y_i(u)\{d\Lambda_0(u) + \beta_0^T Z_i du\}$ for $0 \leq t \leq \tau$, where $\Lambda_0(t)$ is the true cumulative hazard function and τ is the maximum follow-up time. By the fact that $M_i(t), i = 1, \ldots, n$ are zero-mean martingales with respect to the σ-filtration $\sigma\{N_i(u), Y_i(u+), Z_i : 0 \leq u \leq t, i = 1, \ldots, n\}$ (Fleming and Harrington, 1991), when the covariates are fully observed, the estimating

function for β can be arrived as

$$U_n(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \{Z_i - \bar{Z}(t)\}\{dN_i(t) - Y_i(t)\beta^T Z_i dt\}, \tag{2}$$

where $\bar{Z}(t) = \frac{S^{(1)}(t)}{S^{(0)}(t)}$ with $S^{(k)}(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i(t)Z_i^{\otimes k}$ for $k = 0$ and 1. For a vector a, $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^T$. In fact, (2) is the pseudoscore estimating function advised by Lin and Ying (1994).

When some covariates are missing, the estimating function (2) is not applicable. In this section, we consider the simple inverse probability weighted estimating equation methods with true and estimated nonmissingness probability. Firstly, we assume that the nonmissingness probability is known, which is the case for the case-cohort design.

Applying the idea of inverse probability weighting (Horvitz and Thompson, 1952) to (2), the following estimating function for β can be obtained,

$$U_{n,sw}(\beta, \pi) = \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^{\tau} \{Z_i - \bar{Z}_{sw}(t, \pi)\}\{dN_i(t) - Y_i(t)\beta^T Z_i dt\}, \tag{3}$$

where $\bar{Z}_{sw}(t, \pi) = \frac{S_{sw}^{(1)}(t,\pi)}{S_{sw}^{(0)}(t,\pi)}$ with $S_{sw}^{(k)}(t, \pi) = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} Y_i(t)Z_i^{\otimes k}$ for $k = 0$ and 1. From (3), the simple inverse probability weighted estimator for β is

$$\begin{aligned}\hat{\beta}_{sw}(\pi) = &[\sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^{\tau} \{Z_i - \bar{Z}_{sw}(t, \pi)\}^{\otimes 2} Y_i(t)dt]^{-1} \\ &\times [\sum_{j=1}^{n} \frac{\xi_j}{\pi(W_j)} \int_0^{\tau} \{Z_j - \bar{Z}_{sw}(t, \pi)\}dN_j(t)].\end{aligned}$$

Then we can obtain the estimator for $\Lambda_0(t)$,

$$\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\pi), \pi) = \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^{t} \frac{\{dN_i(u) - Y_i(u)\hat{\beta}_{sw}^T(\pi)Z_i du\}}{\sum_{j=1}^{n} \frac{\xi_j}{\pi(W_j)} Y_j(u)}.$$

In fact, the estimator $\hat{\beta}_{sw}(\pi)$ and $\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\pi), \pi)$ are already obtained by Lin (2011). We provide the procedure here just for the ease of the following presentation. The asymptotic properties of $\hat{\beta}_{sw}(\pi)$ and $\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\pi), \pi)$ are obtained by Lin (2011), which are stated in the following Theorem 1.

Before presenting the asymptotic results, we introduce the following notations. Let $s^{(k)}(t) = E\{Y(t)Z^{\otimes k}\}, k = 0, 1, 2$, $\bar{z}(t) = \frac{s^{(1)}(t)}{s^{(0)}(t)}$, $A = E[\int_0^{\tau}\{Z_i - \bar{z}(t)\}^{\otimes 2} Y_i(t)dt]$, $M_{Z_i} = \int_0^{\tau}\{Z_i - \bar{z}(t)\}dM_i(t)$ and $\Sigma = E[M_{Z_i}^{\otimes 2}] = E[\int_0^{\tau}\{Z_i - \bar{z}(t)\}^{\otimes 2} dN_i(t)]$.

THEOREM 1 (LIN, 2011). *Under regular conditions* $(C1)$ *to* $(C4)$ *in the Appendix,*

1. $n^{\frac{1}{2}}(\hat{\beta}_{sw}(\pi) - \beta_0)$ *is asymptotically normal with mean zero and covariance matrix* $A^{-1}\Sigma_{sw}(\pi)A^{-1}$*, where* $\Sigma_{sw}(\pi) = \Sigma + E\{(1-\pi(W_i))\pi(W_i)^{-1}M_{Z_i}^{\otimes 2}\}$.

2. $n^{\frac{1}{2}}(\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\pi), \pi) - \Lambda_0(t))$ *converges weakly to a zero-mean Gaussian process with the covariance function being* $\Gamma_{sw}(s,t,\pi) = E(\psi_{i,sw}(s,\pi)\psi_{i,sw}(t,\pi))$*, where* $\psi_{i,sw}(t,\pi) = \frac{\xi_i}{\pi(W_i)}\int_0^t \frac{1}{s^{(0)}(u)}dM_i(u) - \int_0^t \bar{z}^T(u)duA^{-1}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t)$.

The asymptotic covariance matrix in Theorem 1 can be consistently estimated by the usual plug-in method.

In practice, the nonmissingness probability is frequently unknown. Furthermore, the estimating function (3) only use the information from the subjects with completely observed covariates. So in order to employ more information in observed data, the estimated nonmissingness probability can be applied. Lin (2011) proposed that the nonmissingness probability can be modelled by a parametric model and established the according asymptotic properties. But it is noted that the parametric method is not robust and suffers from the model misspecification. So in order to overcome this drawback, similar to Qi et al. (2005), we propose to estimate the nonmissingness probability π by the nonparametric kernel methods in this paper. Specifically, let $W = (W^{(1)}, W^{(2)})$, where $W^{(1)}$ consists of continuous components of W and $W^{(2)}$ consists of discrete components of W. Denote $K(\cdot)$ to be the kernel function and $K_h(\cdot) = K(\cdot/h)$, where h is the bandwidth which is tending to zero as n goes to infinity. Then the probability of nonmissingness can be estimated by the following Nadaraya-Watson estimator, $\hat{\pi}(w) = \hat{\pi}(w^{(1)}, w^{(2)}) = \frac{\sum_{i=1}^n \xi_i I(W_i^{(2)}=w^{(2)})K_h(w^{(1)}-W_i^{(1)})}{\sum_{j=1}^n I(W_j^{(2)}=w^{(2)})K_h(w^{(1)}-W_j^{(1)})}$.

Replacing π by $\hat{\pi}$ in (3), the simple inverse probability weighted estimators with nonparametrically estimated nonmissingness probability for β_0 and $\Lambda_0(t)$ can be obtained,

$$\hat{\beta}_{sw}(\hat{\pi}) = [\sum_{i=1}^n \frac{\xi_i}{\hat{\pi}(W_i)}\int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}^{\otimes 2}Y_i(t)dt]^{-1} \times [\sum_{j=1}^n \frac{\xi_j}{\hat{\pi}(W_j)}\int_0^\tau \{Z_j - \bar{Z}_{sw}(t,\hat{\pi})\}dN_j(t)]$$

and

$$\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\hat{\pi}), \hat{\pi}) = \sum_{i=1}^n \frac{\xi_i}{\hat{\pi}(W_i)}\int_0^t \frac{\{dN_i(u) - Y_i(u)\hat{\beta}_{sw}^T(\hat{\pi})Z_i du\}}{\sum_{j=1}^n \frac{\xi_j}{\hat{\pi}(W_j)}Y_j(u)}.$$

The asymptotic properties of $\hat{\beta}_{sw}(\hat{\pi})$ and $\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\hat{\pi}), \hat{\pi})$ are stated in the following theorem.

THEOREM 2. *Under regular conditions* $(C1)$ *to* $(C8)$ *in the Appendix, we have*

1. $n^{\frac{1}{2}}(\hat{\beta}_{sw}(\hat{\pi}) - \beta_0)$ *is asymptotically normal with mean zero and covariance matrix* $A^{-1}\Sigma^*_{sw}(\pi)A^{-1}$*, where*

$$\begin{aligned}\Sigma^*_{sw}(\pi) &= \Sigma_{sw}(\pi) - E[(\pi(W_i)^{-1} - 1)\{E(M_{Z_i}|W_i)\}^{\otimes 2}] \\ &= \Sigma + E\{(1 - \pi(W_i))\pi(W_i)^{-1}Var(M_{Z_i}|W_i)\}.\end{aligned}$$

2. $n^{\frac{1}{2}}(\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\hat{\pi}), \hat{\pi}) - \Lambda_0(t))$ *converges weakly to a zero-mean Gaussian process with the covariance function being* $\Gamma^*_{sw}(s,t,\pi) = E(\psi^*_{i,sw}(s,\pi)\psi^*_{i,sw}(t,\pi))$*, where*

$$\begin{aligned}\psi^*_{i,sw}(t,\pi) = &\frac{\xi_i}{\pi(W_i)}\int_0^t \frac{1}{s^{(0)}(u)}dM_i(u) + (1 - \frac{\xi_i}{\pi(W_i)})E[\int_0^t \frac{1}{s^{(0)}(u)}dM_i(u)|W_i] \\ &- \int_0^t \bar{z}^T(u)duA^{-1}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t) \\ &- \int_0^t \bar{z}^T(u)duA^{-1}(1 - \frac{\xi_i}{\pi(W_i)})E[\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t)|W_i].\end{aligned}$$

From Theorem 2, compared with Theorem 1, we can see that the simple inverse probability weighted estimator with nonparametrically estimated nonmissingness probability has smaller asymptotic covariance matrix than that with true nonmissingness probability.

Let $M^0_{Z_i} = E\{M_{Z_i}|W_i\}$. Denote $\hat{M}^0_{Z_i}$ to be the Nadaraya-Waston estimator of $M^0_{Z_i}$ based on $\hat{M}_{Z_i}, i = 1, \ldots, n$, which are defined below. The asymptotic covariance matrix $A^{-1}\Sigma^*_{sw}(\pi)A^{-1}$ can be consistently estimated by

$$\hat{A}^{-1}(\hat{\Sigma} + \hat{E}\{(1 - \pi(W_i))\pi(W_i)^{-1}Var(M_{Z_i}|W_i)\})\hat{A}^{-1},$$

where

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\hat{\pi}(W_i)}\int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}^{\otimes 2}Y_i(t)dt,$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i}{\hat{\pi}(W_i)}\int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}^{\otimes 2}dN_i(t)$$

and

$$\hat{E}\{(1 - \pi(W_i))\pi(W_i)^{-1}Var(M_{Z_i}|W_i)\} = \frac{1}{n}\sum_{i=1}^n \frac{\xi_i(1 - \hat{\pi}(W_i))}{\hat{\pi}(W_i)^2}(\hat{M}_{Z_i} - \hat{M}^0_{Z_i})^{\otimes 2}$$

with

$$\hat{M}_{Z_i} = \int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}\{dN_i(t) - Y_i(t)\hat{\beta}_{sw}(\hat{\pi})^T Z_i dt - Y_i(t)d\hat{\Lambda}_{sw}(t, \hat{\beta}_{sw}(\hat{\pi}), \hat{\pi})\}.$$

3 Augmented inverse probability weighted estimators

It is known that augmented inverse probability weighted estimators are doubly robust and more efficient than the simple inverse probability weighted estimators when the nonmissingness probability and some conditional distributions are known or prespecified by parametric models (Tsiatis, 2006). Lin (2011) also studied the augmented inverse probability weighted method, in which both the nonmissingness probability and the conditional distributions of the missing covariates given the observed data are modelled by parametric models. But in this situation, the conditional distributions of the missing covariates given the observed data is hard to be specified. So we propose that both the nonmissingness probability and the conditional distributions of the missing covariates given the observed data are estimated by the nonparametric kernel method. In this paper, we find that the augmented inverse probability weighted estimators with known or nonparametricly estimated nonmissingness probability have the same asymptotic distribution as that of the simple inverse probability weighted estimators with nonparametrically estimated nonmissingness probability. This phenomenon is also found by Qi et al. (2005) for the survival data with missing covariates under the Cox's proportional hazards model.

Firstly, the nonmissingness probability is assumed to be known in advance, such as in the case-cohort design. But because the conditional distributions of the missing covariates given the observed data is always unknown, the nonparametric kernel smoothing technique is applied to obtain the estimator of the conditional expectations of the missing covariates given the observed data. By adding augmentation terms to the estimating function (3), we arrive at the following augmented inverse probability weighted estimating function,

$$\begin{aligned}
&U_{n,aw}(\beta,\pi)\\
&=\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau\{Z_i-\bar{Z}_{aw}(t,\pi)\}[dN_i(t)-Y_i(t)\beta^TZ_idt]\\
&+\sum_{i=1}^{n}(1-\frac{\xi_i}{\pi(W_i)})\int_0^\tau[\hat{E}(Z_i|W_i)dN_i(t)-Y_i(t)\hat{E}[Z_i^{\otimes 2}|W_i]\beta dt\\
&-\bar{Z}_{aw}(t,\pi)\{dN_i(t)-Y_i(t)\hat{E}[Z_i^T|W_i]\beta dt\}],
\end{aligned}\tag{4}$$

where $\bar{Z}_{aw}(t,\pi)=\frac{S_{aw}^{(1)}(t,\pi)}{S_{aw}^{(0)}(t,\pi)}$ with $S_{aw}^{(0)}(t,\pi)=\frac{1}{n}\sum_{i=1}^{n}Y_i(t)$ and

$$S_{aw}^{(1)}(t,\pi)=\frac{1}{n}\sum_{i=1}^{n}[\frac{\xi_i}{\pi(W_i)}Y_i(t)Z_i+(1-\frac{\xi_i}{\pi(W_i)})Y_i(t)\hat{E}(Z_i|W_i)],$$

$\hat{E}[Z_i|W_i]$, $\hat{E}(Z_i^{\otimes 2}|W_i)$ are the Nadaraya-Watson estimators for $E[Z_i|W_i]$, $E[Z_i^{\otimes 2}|W_i]$ respectively. Denote the solution of $U_{n,aw}(\beta,\pi)=0$ to be $\hat{\beta}_{aw}(\pi)$. Then we can

get

$$\begin{aligned}\hat{\beta}_{aw}(\pi) = [&\sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^\tau \{Z_i - \bar{Z}_{aw}(t,\pi)\} Y_i(t) \beta^T Z_i dt \\ &+ \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^\tau Y_i(t) \{\hat{E}[Z_i^{\otimes 2}|W_i] dt - \bar{Z}_{aw}(t,\pi) \hat{E}[Z_i^T|W_i]\}]^{-1} \\ &\times [\sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^\tau \{Z_i - \bar{Z}_{aw}(t,\pi)\} dN_i(t) \\ &+ \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^\tau \{\hat{E}[Z_i|W_i] - \bar{Z}_{aw}(t,\pi)\} dN_i(t)].\end{aligned}$$

Then, we can get the augmented inverse probability weighted estimator $\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\pi), \pi)$ for $\Lambda_0(t)$. The asymptotic properties of the $\hat{\beta}_{aw}(\pi)$ and $\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\pi), \pi)$ are stated in the following theorem.

THEOREM 3. *Under regularity conditions* $(C1)$ *to* $(C9)$ *in the Appendix, we have*

1. $n^{\frac{1}{2}}(\hat{\beta}_{aw}(\pi) - \beta_0)$ *is asymptotically normal with mean zero and covariance matrix* $A^{-1}\Sigma^*_{sw}(\pi)A^{-1}$.

2. $n^{\frac{1}{2}}(\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\pi), \pi) - \Lambda_0(t))$ *converges weakly to a zero-mean Gaussian process with the covariance function being* $\Gamma^*_{sw}(s,t,\pi)$.

In the situation where nonmissingness probability is unknown, we estimate it by the method as that in Section 3 and plug it in (4). Denote the estimating function (4) with estimated nonmissingness probability to be $U_{n,aw}(\beta, \hat{\pi})$ and the according solution to be $\hat{\beta}_{aw}(\hat{\pi})$. Accordingly the estimator of $\Lambda_0(t)$ is $\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\hat{\pi}), \hat{\pi})$. The asymptotic properties of the $\hat{\beta}_{aw}(\hat{\pi})$ and $\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\hat{\pi}), \hat{\pi})$ are stated in the following theorem.

THEOREM 4. *Under regularity conditions* $(C1)$ *to* $(C9)$ *in the Appendix, we have*

1. $n^{\frac{1}{2}}(\hat{\beta}_{aw}(\hat{\pi}) - \beta_0)$ *is asymptotically normal with mean zero and covariance matrix* $A^{-1}\Sigma^*_{sw}(\pi)A^{-1}$.

2. $n^{\frac{1}{2}}(\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\hat{\pi}), \hat{\pi}) - \Lambda_0(t))$ *converges weakly to a zero-mean Gaussian process with the covariance function being* $\Gamma^*_{sw}(s,t,\pi)$.

The asymptotic covariance matrixes in Theorems 3 and 4 can be consistently estimated by the method similar to that used in Theorem 2. So we omit them here for simplicity.

4 Simulation Studies

In this section, extensive simulation studies were conducted to verified the performance of the proposed methods in Sections 2 and 3.

Table 1. Simulation results for the case with $(\beta_{10}, \beta_{20}) = (0, 0.7)$.

n	Method	β_{10}				β_{20}			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
300	Full	-0.000	0.084	0.087	0.949	0.009	0.105	0.106	0.946
	CC	0.044	0.102	0.101	0.935	-0.091	0.137	0.136	0.865
	SW(π)	0.005	0.134	0.137	0.949	0.006	0.165	0.167	0.948
	SW($\hat{\pi}$)	0.011	0.090	0.093	0.951	-0.045	0.141	0.144	0.932
	AW(π)	0.000	0.099	0.097	0.963	0.012	0.157	0.159	0.947
	AW($\hat{\pi}$)	0.001	0.099	0.097	0.964	0.013	0.155	0.162	0.944
500	Full	0.000	0.064	0.066	0.944	0.003	0.080	0.082	0.951
	CC	0.039	0.077	0.080	0.907	-0.101	0.103	0.101	0.800
	SW(π)	-0.003	0.102	0.107	0.941	0.004	0.127	0.125	0.951
	SW($\hat{\pi}$)	0.009	0.070	0.072	0.936	-0.040	0.110	0.101	0.922
	AW(π)	0.000	0.075	0.074	0.955	0.008	0.120	0.121	0.945
	AW($\hat{\pi}$)	-0.001	0.075	0.075	0.952	0.006	0.118	0.121	0.941

The survival time T was generated from the distribution with hazard function $\lambda(t|Z) = \lambda_0 + \beta_{10}Z_o + \beta_{20}Z_m$. Let Z_o and Z_m be generated from the Bernoulli distribution with success probability 0.5 independently. The regression coefficients were set to be $(\beta_{10}, \beta_{20}) = (0, 0.7)$ or $(\beta_{10}, \beta_{20}) = (1, -1)$. The censoring variable C follows the uniform distribution on $(0, c_0)$ independently of Z_o and Z_m. Different $c_0's$ and $\lambda_0's$ were chosen to get approximatively 40% censoring rate (CR) under different parameter values. The covariate Z_m is observed with the probability $\pi(W) = (1 + \exp(-W^T\theta))^{-1}$, where $W = (1, \delta, X, Z_o)$. Different θ's were set to produce the approximatively 60% missing rate (MR) under various cases. The univariate Gaussian kernel function was applied with bandwidth $h = 5\sigma_X n^{-\frac{1}{3}}$ with σ_X being the standard error of X, because there is only one continuous component X in W. More detailed discussion about the choice of kernel function and bandwidth can be found in Qi et al. (2005). The sample sizes were set to be $n = 300$ and $n = 500$, while the replications was 1000.

The simulation results were summarized in Tables 1 and 2, in which Full represents full data method; CC, complete-case method; SW(π), simple inverse probability weighted method with true nonmissingness probability; SW($\hat{\pi}$), simple inverse probability weighted method with estimated nonmissingness probability; AW(π), augmented inverse probability weighted method with true nonmissingness probability; AW($\hat{\pi}$), augmented inverse probability weighted method with estimated nonmissingness probability. From the results, we can see that the complete-cases analysis produces the inconsistent estimators of the regression coefficients for the cases we considered and the increase of sample size can not improve the performance of the complete-cases analysis. The simple inverse probability weighted estimators with $\hat{\pi}$ have the largest biases among the methods except for the complete-cases analysis, while the rest of methods have the similar performance with regard to bias. For all the methods, the av-

Table 2. Simulation results for the case with $(\beta_{10}, \beta_{20}) = (1, -1)$.

n	Method	β_{10}				β_{20}			
		Bias	ESE	SSE	CP	Bias	ESE	SSE	CP
300	Full	0.008	0.135	0.137	0.948	-0.005	0.134	0.137	0.944
	CC	-0.074	0.167	0.163	0.899	0.148	0.185	0.184	0.838
	SW(π)	0.012	0.196	0.197	0.945	-0.010	0.204	0.218	0.947
	SW($\hat{\pi}$)	-0.064	0.143	0.143	0.912	0.059	0.166	0.176	0.918
	AW(π)	0.007	0.160	0.151	0.964	-0.017	0.184	0.197	0.937
	AW($\hat{\pi}$)	0.021	0.163	0.153	0.964	0.039	0.191	0.204	0.944
500	Full	-0.001	0.104	0.104	0.946	0.001	0.103	0.103	0.951
	CC	-0.085	0.127	0.126	0.877	0.153	0.142	0.135	0.781
	SW(π)	-0.004	0.151	0.154	0.941	-0.004	0.159	0.161	0.951
	SW($\hat{\pi}$)	-0.065	0.109	0.110	0.910	0.055	0.130	0.131	0.917
	AW(π)	-0.005	0.120	0.114	0.959	-0.008	0.142	0.143	0.955
	AW($\hat{\pi}$)	0.004	0.122	0.115	0.958	-0.021	0.145	0.146	0.946

erage estimated standard errors (ESE) agree with the sample standard errors (SSE) of the estimators. The simple inverse probability weighted estimator with $\hat{\pi}$, augmented inverse probability weighted estimator with π and $\hat{\pi}$ have smaller asymptotic variance than the simple inverse probability weighted estimator with π. It also can be seen that frequently the coverage probability (CP) of complete-cases analysis deviates from the nominal level 95% largely.

5 The Mouse Leukemia Study

In this section, we illustrate the proposed methods by analyzing the the mouse leukemia data given by Kalbfleisch and Prentice (2002). This study was conducted by Nowinski et al. (1979) to examine the influence of genetic and viral factor on the development of spontaneous leukemia and collected data about 204 mice dying of thymic leukemia, nonthymic leukemia, or other natural causes during 2 years. In our analysis, we mainly investigate the influence of level of endogenous murine leukemia virus and the Gpd-1 phenotype on the survival time.

At first, following Qi et al. (2005), Huang and Wang (2010) and so on, we categorize the virus level into a binary variable with 0 representing values below 10^4 PFU/ml and 1 otherwise. The Gpd-1 phenotype was subject to substantial missingness, while the virus level was recorded almost for all the mice. For simplicity, following Wang and Chen (2001), Qi et al. (2005) and so on, the mice with missing endogenous murine leukemia was excluded from the analysis. Then the data about 175 mice was used in our analysis, among which 67 mice died of leukemia. Furthermore, 56 mice died of thymic leukemia and 11 mice died of nonthymic leukemia. The survival time was scaled by year. As noted by many authors, the missing at random assumption is reasonable in this analysis.

Table 3. Estimators for the regression coefficients of virus level and Gpd-1 phenotype.

Method	Case 1		Case 2	
	Virus	Gpd-1	Virus	Gpd-1
CC	0.1113(0.0549)	-0.1931(0.0818)	0.1098(0.0557)	-0.2119(0.0841)
SW($\hat{\pi}$)	0.2068(0.0705)	-0.2285(0.1024)	0.1967(0.0767)	-0.2703(0.1089)
AW($\hat{\pi}$)	0.2426(0.0673)	-0.2371(0.0947)	0.2250(0.0723)	-0.2911(0.0996)

The kernel function and the bandwidth were chosen as those in Section 4. Two kinds of analysis were conducted: the first assumed that time to mortality due to the thymic leukemia as the failure time, while the second assumed that time to mortality due to the thymic or nonthymic leukemia as the failure time. Case 1 is corresponding to the analysis based on the time to mortality due to thymic leukemia as the failure time, while Case 2 is corresponding to the analysis based on the time to mortality due to thymic or nonthymic leukemia as the failure time. The numbers in the parentheses represent the standard error. Table 3 summarizes the analysis results, from which we can conclude that the virus level has positive association with the hazard function of mice with leukemia, while the Gpd-1 has negative association. These results coincides with the early research about this study, such as Qi et al. (2005) and Lin (2011). But it is noted that estimators from simple inverse probability weighted estimating equations are close to estimators from augmented inverse probability weighted estimating equations, while estimators from the complete-case method have smaller estimators.

6 Acknowledgements

Chen's research was supported by the National Natural Science Foundation of China (11201484 and 11326184) and the Fundamental Research Funds for the Central Universities (14CX02009A).

7 Appendix

We give the outlined proof of Theorems 2 and 3. Theorem 4 can be proven by the same manner as those of Theorems 2 and 3. So we omit it here. In the following, we list the necessary regularity conditions.

($C1$): $\Lambda_0(\tau) < \infty$ and $\Pr\{Y(\tau) = 1\} > 0$;

($C2$): Z is bounded with probability 1 and time-independent;

($C3$): The matrix A, which is defined in Section 2, is positive definitive.

($C4$): The nonmissingness probability π is bounded away from 0 and has r continuous bounded partial derivatives with respect to the continuous components of W a.e.;

($C5$): The probability density function $f(w)$ of W and the conditional probability density function $f_{W|\xi}(w)$ of W given ξ are bounded away from 0 and has r continuous bounded partial derivatives with respect to the continuous components of W a.e.; The conditional probability density functions $f_{W|\xi=0}(w)$ and $f_{W|\xi=1}(w)$ have the same support, and $c(w) = f_{W|\xi=1}(w)/f_{W|\xi=0}(w)$ is bounded over the support.

($C6$): The conditional expectations $E(Z|W)$ and $E[Z^{\otimes 2}|W]$ have r continuous bounded partial derivatives with respect to the continuous components of W a.e.;

($C7$): $r > d$, $nh^{2d} \to \infty$ and $nh^{2r} \to 0$, as $n \to \infty$;

($C8$): $\bar{M}_{1,n}(t)$ and $\bar{M}_{2,n}(t)$ converge to zero-mean Gaussian processes with continuous sample path, where $\bar{M}_{1,n}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)^2}(\hat{\pi}(W_i) - \pi(W_i))M_i(t)$ and $\bar{M}_{2,n}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)^2}(\hat{\pi}(W_i) - \pi(W_i))E[M_i(t)|W_i]$.

($C9$): $\sup_{0\le t\le\tau}|\bar{M}_{3,n}(t)| = o_p(1)$ and $\sup_{0\le t\le\tau}|\bar{M}_{4,n}(t)| = o_p(1)$, where $\bar{M}_{3,n}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n}(1-\frac{\xi_i}{\pi(W_i)}) \int_0^t(\hat{E}[dM_i(u)|W_i] - E[dM_i(u)|W_i])$ and $\bar{M}_{4,n}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n}(1-\frac{\xi_i}{\hat{\pi}(W_i)}) \int_0^t(\hat{E}[dM_i(u)|W_i] - E[dM_i(u)|W_i])$.

It is noted that ($C8$) and ($C9$) are not standard for the proof of the nonparametric kernel technique. But our problem complicated by nonparametric technique, censored and missing data is very challenging. The investigation of the asymptotic properties is very hard. Also, the similar conditions were used in Qi et al. (2005) for the Cox's model with missing covariates.

Proof of Theorem 2. (1) It is easy to verify that

$$n^{\frac{1}{2}}\{\hat{\beta}_{sw}(\hat{\pi}) - \beta_0\} = V_{n,sw}(\hat{\pi})^{-1} n^{-\frac{1}{2}} U_{n,sw}(\beta_0, \hat{\pi}),$$

where $V_{n,sw}(\hat{\pi}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_i}{\hat{\pi}(W_i)} \int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}^{\otimes 2} Y_i(t)dt$.

Under the Conditions ($C4$), ($C5$) and ($C7$), it can be shown that

$$\sup_{t\in[0,\tau]} \|S_{sw}^{(k)}(t,\hat{\pi}) - s^{(k)}(t)\| = o_p(1), \tag{5}$$

for $k = 0, 1, 2$. So we arrive at

$$\begin{aligned} & V_{n,sw}(\hat{\pi}) - A \\ &= \int_0^\tau \{S_{sw}^{(2)}(t,\hat{\pi}) - s^{(2)}(t)\}dt - \int_0^\tau \{\frac{S_{sw}^{(1)}(t,\hat{\pi})^{\otimes 2}}{S_{sw}^{(0)}(t,\hat{\pi})} - \frac{s^{(1)}(t)^{\otimes 2}}{s^{(0)}(t)}\}dt = o_p(1). \end{aligned} \tag{6}$$

In the following, we prove that $n^{-\frac{1}{2}} U_{n,sw}(\beta_0, \hat{\pi})$ is asymptotically normal when n goes to infinity. It is easy to see that

$$n^{-\frac{1}{2}} U_{n,sw}(\beta_0, \hat{\pi}) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\hat{\pi}(W_i)} \int_0^\tau \{Z_i - \bar{Z}_{sw}(t,\hat{\pi})\}dM_i(t).$$

By the Taylor expansion of $\hat{\pi}(W_i)^{-1}$ about $\pi(W_i)$, we obtain

$$n^{-\frac{1}{2}}U_{n,sw}(\beta_0,\hat{\pi}) = T_1 - T_2 + T_3 - T_4 + o_p(1),$$

where

$$T_1 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t),$$

$$T_2 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi^2(W_i)}(\hat{\pi}(W_i) - \pi(W_i))\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t),$$

$$T_3 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{\bar{z}(t) - \bar{Z}_{sw}(t,\hat{\pi})\}dM_i(t)$$

and

$$T_4 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi^2(W_i)}(\hat{\pi}(W_i) - \pi(W_i))\int_0^\tau \{\bar{z}(t) - \bar{Z}_{sw}(t,\hat{\pi})\}dM_i(t).$$

Under the Conditions $(C1)$, $(C2)$, $(C4)$, $(C5)$ and $(C7)$, similar to Step B2 of Qi et al. (2005), it can be gotten that

$$T_2 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i - \pi(W_i)}{\pi(W_i)}E\left[\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t)\Big|W_i\right] + o_p(1).$$

By Condition $(C4)$ and (5), we can reach $\sup_{t\in[0,\tau]}\|Z_{sw}(t,\hat{\pi}) - \bar{z}(t)\| = o_p(1)$. Then by $(C8)$, similar to step A.2 of Qi et al. (2005), it can be proven that $T_3 = o_p(1)$. Similarly, $T_4 = o_p(1)$. Finally, we arrive at

$$\begin{aligned} & n^{-\frac{1}{2}}U_{n,sw}(\beta_0,\hat{\pi}) \\ = & n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t) \\ & + n^{-\frac{1}{2}}\sum_{i=1}^{n}(1 - \frac{\xi_i}{\pi(W_i)})E\left[\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t)\Big|W_i\right] + o_p(1). \end{aligned} \tag{7}$$

Based on (6) and (7), we get

$$\begin{aligned} & n^{\frac{1}{2}}\{\hat{\beta}_{sw}(\hat{\pi}) - \beta_0\} \\ = & A^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t) \\ & + A^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}(1 - \frac{\xi_i}{\pi(W_i)})E\left[\int_0^\tau \{Z_i - \bar{z}(t)\}dM_i(t)\Big|W_i\right] + o_p(1). \end{aligned} \tag{8}$$

By (8) and the central limit theorem, the proof of first part is finished.

(2) By some algebraic manipulations, it can be seen that $n^{\frac{1}{2}}\{\hat{\Lambda}_{sw}(t,\hat{\beta}_{sw}(\hat{\pi}),\hat{\pi})-\Lambda_0(t)\} = T_5 + T_6$, where $T_5 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\hat{\pi}(W_i)}\int_0^t \frac{1}{S^{(0)}_{aw}(u,\hat{\pi})}dM_i(u)$ and $T_6 = \int_0^t Z_{sw}(u,\hat{\pi})^T dun^{\frac{1}{2}}\{\hat{\beta}_{sw}(\hat{\pi})-\beta_0\}$. Similar to the proof of first part, we can prove that

$$T_5 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^t \frac{1}{s^{(0)}(u)}dM_i(u) - n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i-\pi(W_i)}{\pi(W_i)}E\left[\int_0^t \frac{1}{s^{(0)}(u)}dM_i(u)\Big|W_i\right] + o_p(1)$$

and $T_6 = \int_0^t \bar{z}(u)dun^{\frac{1}{2}}\{\hat{\beta}_{sw}(\hat{\pi})-\beta_0\} + o_p(1)$. So we have $n^{\frac{1}{2}}\{\hat{\Lambda}_{sw}(t,\hat{\beta}_{sw}(\hat{\pi}),\hat{\pi}) - \Lambda_0(t)\} = n^{\frac{1}{2}}\psi^*_{i,sw}(t,\pi) + o_p(1)$, where $\psi^*_{i,sw}(t,\pi)$ is defined in Theorem 2. The finite dimensional convergence of $n^{\frac{1}{2}}\{\hat{\Lambda}_{sw}(t,\hat{\beta}_{sw}(\hat{\pi}),\hat{\pi}) - \Lambda_0(t)\}$ can be proven by (A.6). The tightness is obtained by the empirical process theory (Lin et al., 2000). Finally the weak convergence is proven.

Proof of Theorem 3.

(1) It is easy to verify that $n^{\frac{1}{2}}\{\hat{\beta}_{aw}(\pi)-\beta_0\} = V_{n,aw}(\pi)^{-1}n^{-\frac{1}{2}}U_{n,aw}(\beta_0,\pi)$, where

$$\begin{aligned}V_{n,aw}(\pi) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau\{Z_i-\bar{Z}_{aw}(t,\pi)\}Y_i(t)Z_i^T dt \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_i}{\pi(W_i)})\int_0^\tau\{\hat{E}(Z_i^{\otimes 2}|W_i) - \bar{Z}_{aw}(t,\pi)\hat{E}(Z_i^T|W_i)\}Y_i(t)dt \\ &= T_7 + T_8.\end{aligned}$$

By the law of large numbers and similar to Step B1 in Qi et al. (2005), it can be proven that $T_7 = A + o_p(1)$ and $T_8 = o_p(1)$. So we get $V_{n,aw}(\pi) = A + o_p(1)$.

In the following, we prove that $n^{-\frac{1}{2}}U_{n,aw}(\beta_0,\pi)$ is asymptotically normal when n goes to infinity. Through some algebraic calculation, it can be verified that $n^{-\frac{1}{2}}U_{n,aw}(\beta_0,\pi) = T_9 + T_{10}$, where $T_9 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau\{Z_i - \bar{Z}_{aw}(t,\pi)\}dM_i(t)$ and

$$T_{10} = n^{-\frac{1}{2}}\sum_{i=1}^{n}(1-\frac{\xi_i}{\pi(W_i)})\int_0^\tau\{\hat{E}(Z_i dM_i(t)|W_i) - \bar{Z}_{aw}(t,\pi)\hat{E}(dM_i(t)|W_i)\},$$

with $\hat{E}(Z_i dM_i(t)|W_i) = \hat{E}(Z_i|W_i)dN_i(t) - Y_i(t)\{d\Lambda_0(t) - \hat{E}(Z_i^{\otimes 2}|W_i)\beta_0\}dt$ and $\hat{E}(dM_i(t)|W_i) = dN_i(t) - Y_i(t)\{d\Lambda_0(t) - \hat{E}(Z_i^T|W_i)\beta_0\}dt$. Similar to Step A.2 of Qi et al. (2005), it can be seen that $T_9 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{\xi_i}{\pi(W_i)}\int_0^\tau\{Z_i - \bar{z}(t)\}dM_i(t) +$

$o_p(1)$. By a simple decomposition, it is easy to see that

$$T_{10} = n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^\tau \{E[Z_i - \bar{z}(t)] dM_i(t) | W_i\} + o_p(1).$$

Then we arrive at

$$\begin{aligned} & n^{-\frac{1}{2}} U_{n,aw}(\beta_0, \pi) \\ = & n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) \\ & + n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^\tau \{E[Z_i - \bar{z}(t)] dM_i(t) | W_i\} + o_p(1). \end{aligned}$$

Finally we have

$$\begin{aligned} & n^{\frac{1}{2}} \{\hat{\beta}_{aw}(\pi) - \beta_0\} \\ = & A^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) \\ & + A^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) E \left[\int_0^\tau \{Z_i - \bar{z}(t)\} dM_i(t) \Big| W_i \right] + o_p(1). \end{aligned} \quad (9)$$

By (9) and the central limit theorem, the proof of first part is finished.

(2) By some algebraic manipulations, it can be seen that $n^{\frac{1}{2}} \{\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\pi), \pi) - \Lambda_0(t)\} = T_{11} + T_{12}$, where

$$T_{11} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^t \frac{1}{\bar{Y}(u)} dM_i(u) - \int_0^t \frac{S_{sw}^{(1)}(u, \pi)^T}{\bar{Y}(u)} du n^{\frac{1}{2}} \{\hat{\beta}_{aw}(\pi) - \beta_0\}$$

and

$$\begin{aligned} T_{12} = & n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^t \frac{1}{\bar{Y}(u)} E[dM_i(u) | W_i] \\ & - \int_0^t \frac{\frac{1}{n} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) Y_i(u) E(Z_i) | W_i}{\bar{Y}(u)} du n^{\frac{1}{2}} \{\hat{\beta}_{aw}(\pi) - \beta_0\} \\ & - n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^t \frac{1}{\bar{Y}(u)} \{E(dM_i(t) | W_i) - \hat{E}(dM_i(t) | W_i)\} du \hat{\beta}_{aw}(\pi) \end{aligned} \quad (10)$$

with $\bar{Y}(t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(t)$. For T_{11}, we can see that

$$T_{11} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\xi_i}{\pi(W_i)} \int_0^t \frac{1}{s^{(0)}(u)} dM_i(u) - \int_0^t \bar{z}(u) du n^{\frac{1}{2}} \{\hat{\beta}_{aw}(\pi) - \beta_0\} + o_p(1).$$

By the law of large numbers, the fact $n^{\frac{1}{2}}\{\hat{\beta}_{aw}(\pi) - \beta_0\} = O_p(1)$ and Condition $(C9)$, the last two terms of (12) converges to 0 in probability. So we can get

$$T_{12} = n^{-\frac{1}{2}} \sum_{i=1}^{n} (1 - \frac{\xi_i}{\pi(W_i)}) \int_0^t \frac{1}{s^0(t)} E[dM_i(u)|W_i] + o_p(1).$$

Finally we arrive at $n^{\frac{1}{2}}\{\hat{\Lambda}_{aw}(t, \hat{\beta}_{aw}(\pi), \pi) - \Lambda_0(t)\} = n^{\frac{1}{2}}\psi^*_{i,sw}(t, \pi) + o_p(1)$. The finite dimensional convergence of is easy to seen. The tightness is obtained by the empirical process theory (Lin et al., 2000). Finally the weak convergence is proven.

References

1. Cook, V.J., Hu, X.J. and Swartz, T.B.: Cox regression with covariates missing not at random. Statistics in Biosciences, 3, 208-222 (2011)
2. Cox, D.R.: Regression models and life-table (with discussion). Journal of the Royal Statistical Society: Series B, 34, 187-220 (1972)
3. Fleming, T.R. and Harrington, D.P.: (1991) Counting processes and survival analysis. New York: Wiley. (1991)
4. Huang, B. and Wang, Q.H.: Semiparametric analysis based on weighted estimating equations for transformation models with missing covariates. Journal of Multivariate Analysis, 101, 2078-2090 (2010)
5. Horvitz, D.G. and Thompson, D.J.: A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663-685 (1952)
6. Kalbfleisch, J.D. and Prentice, R.L.: The statistical analysis of failure time data. New York: Wiley. (2002)
7. Jiang, J. and Zhou, H.: Additive hazard regression with auxiliary covariates. Biometrika, 94, 359-369 (2007)
8. Kulich, M. and Lin, D.Y.: Additive hazards regression for case-cohort studies. Biometrika, 87, 73-87 (2000a)
9. Kulich, M. and Lin, D.Y.: Additive hazards regression with covariate measurement error. Journal of the American Statistical Association, 95, 238-248 (2000b)
10. Lin, D. and Ying, Z.: Semiparametric analysis of the additive risk model. Biometrika, 81, 61-71 (1994)
11. Lin, D., Wei, L., and Ying, Z.: Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B, 62, 711-730 (2000)
12. Lin, W.: Missing covariates and high-dimensional variable selection in additive hazards regression. PhD thesis of University of Southern California. (2011)
13. Luo, X.D., Tsai, W.Y. and Xu, Q.: Pseudo-partial likelihood estimators for the Cox regression model with missing covariates. Biometrika, 96, 617-633 (2009)
14. Nowinski, R.C., Brown, M., Doyle, T., and Prentice, R.L.: Genetic and viral factors influencing the development of spontaneous tumors in AKR mice. Virology 96, 186-204 (1979)
15. Qi, L.H., Wang C.Y. and Prentice, R.L.: Weighted estimators for proportional hazards regression with missing covariates. Journal of the American Statistical Association, 100, 1250-1263 (2005)
16. Tsiatis, A.A.: Semiparametric theory and missing data. New York: Springer. (2006)
17. Wang, C.Y. and Chen, H.Y.: Augmented inverse probability weighted estimator for Cox missing covariate regression. Biometrics, 57, 414-419 (2001)
18. Xu, Q., Paik, M.C., Luo, X.D. and Tsai, W.Y.: Reweighting estimators for Cox regression with missing covariates. Journal of the American Statistical Association, 104, 1155-1167 (2009)

Five-Band Toeplitz Universal Portfolios

Choon Peng Tan[1] and Sheong Wei Phoon[2]

[1,2]Department of Mathematical and Actuarial Sciences
UniversitiTunku Abdul Rahman
JalanGenting Kelang, 53300 Setapak
Kuala Lumpur, Malaysia
tancp@utar.edu.my, dinoleon@live.com

Abstract. Mahalanobis universal portfolios generated by five-bandToeplitaz matrices are studied in this paper. The structure of the companion matrix of the generating matrix is determined. The empirical performance of the portfolios on selected stock-price data sets from the local stock exchange is analysed. These data sets comprise of company stocks traded over periods of 1500 days. It is shown that investment wealth can be increased significantly by using the five-band Toeplitz universal portfolio.

Keywords: universal portfolio, five-band Toeplitz matrix, empirical performance

INTRODUCTION

Universal portfolios have been studied by a number of authors (Cover(1991), Cover and Ordentlich (1996)). They can be used in investment settings where no assumption is made on the stochastic model of the stock prices. A study of two-stock portfolios of stocks listed on the New York Stock Exchange was done by Cover(1991) indicating a tremendous increase of investment wealth over a 22-year period. To overcome the substantial memory and time requirements of the Cover uniform portfolio, Helmbold et al. (1998) introduced a multiplicative-update universal portfolio which can achieve a comparable performance. A generalization of the Helmbold universal portfolio to an additive-update universal portfolio was proposed by Tan and Lim (2012) using the Mahalanobis squared divergence.

An empirical study was made on the performance of diagonal –matrix generated universal portfolios on selected Malaysian Stocks demonstrating good investment returns. In this study, the empirical performance of five-band Toeplitz universal portfolios will be analyzed.

Some Preliminaries and Theory

Consider a market of m stocks, where the price relative of a stock on any day is defined as the ratio of its closing price to its opening price. Let $x_n = (x_{ni})$

denote the price-relative vector on the n^{th} trading day, where x_{ni} is the price relative of thei^{th} stock for $i = 1,2,3, \dots, m$. A portfolio vector is an investment strategy used on a trading day, where each component is a proportion of current wealth invested on a particular stock. Suppose $\boldsymbol{b}_n = (b_{ni})$ is the portfolio strategy used on day n, where b_{ni} is the proportion of current wealth invested on stock *i,*for $i = 1,2,3, \dots, m$. An initial wealth of one unit is assumed. The wealth S_n at the end of the n^{th} trading day is given by:

$$S_n = \prod_{j=1}^{n} \boldsymbol{b_j^t x_j} \quad (1)$$

where $\boldsymbol{b}_j^t \boldsymbol{x}_j = \sum_{i=1}^{m} b_{ji} x_{ji}$.

Let A be anmxm matrix and suppose the initial portfolio $\boldsymbol{b_1}$is given. Then the *Mahalanobis universal portfolio* $(\boldsymbol{A}, \boldsymbol{b_1}, \xi)$ generated by $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{b_1}$ is given by:

$$\boldsymbol{b_{n+1}} = \boldsymbol{b_n} + \frac{\xi}{(b^t x_n)} [Ax_n - \left(\frac{\boldsymbol{1^t} Ax_n}{\boldsymbol{1^t} A\boldsymbol{1}}\right) A\boldsymbol{1}] \quad \textbf{(2)}$$

for $n = 0,1,2, \dots,$ where $\boldsymbol{1} = (1,1,1, \dots, 1)$ and ξ is any real number such that $\boldsymbol{b}_{n+1} \geq 0$. For **A** and $\boldsymbol{b_1}$fixed, there exists an interval of ξ containing $\xi = 0$ such that $\boldsymbol{b}_{n+1} \geq 0$. For ξ in the valid interval, the family of universal portfolios $\left(\boldsymbol{A}, \boldsymbol{b_1}, \xi\right)$ form a parametric family. The focus of this study is on the wealth achieved by this parametric family where **A** is a $5x5$ Toeplitaz matrix. A simplification of (2) is given by Tan and Lim (2012), namely,

$$\boldsymbol{b_{n+1}} = \boldsymbol{b_n} + \frac{\xi}{(b^t x_n)} \boldsymbol{C} x_n \quad (3)$$

for$n = 0,1,2, \dots$,where **C** is known as the *companion matrix* of **A** defined as:

$$c_{ij} = a_{ij} - \frac{a_{i.} a_{.j}}{a_{..}} for i, j = 1,2, \dots, m \quad (4)$$

where $a_{i.} = \sum_{i=1}^{m} a_{ij}$, $a_{.j} = \sum_{j=1}^{m} a_{ij}$, $a_{..} = \sum_{i,j=1}^{m} a_{ij}$.

If the generating matrix **A** is positive definite, then $\boldsymbol{A^{-1}}$is positive definite and the Mahalanobis squared divergence generated by $\boldsymbol{A^{-1}}$is well-defined. The portfolio $\boldsymbol{b_{n+1}}$ defined by (2) maximizes a linear sum of the rate of wealth increase and the Mahalanobis squared divergence of the portfolio vectors $\boldsymbol{b_{n+1}}$ and $\boldsymbol{b_n}$(see Tan and Lim (2012)). If **A** is not positive definite, the portfolio $\boldsymbol{b_{n+1}}$defined by (2) is said to be pseudo-Mahalanobis. No emphasis is made on whether **A** is positive definite or not in this empirical study. It is observed that if **C** is the companion matrix of **A,** then $\gamma \boldsymbol{C}$is the companion matrix of $\gamma \boldsymbol{A}$for any scalar γ**,** from (3) and (4). It is clear from (3) that the parametric class of portfolios $(\boldsymbol{A}, \boldsymbol{b_1}, \xi)$ is the same as the parametric class $(\gamma \boldsymbol{A}, \boldsymbol{b_1}, \xi)$ for any scalar γ.

The mxm matrix $\boldsymbol{A} = (a_{ij})$ is a *symmetric Toeplitz matrix* of bandwidth $2k + 1$ for a positive integer k if $a_{ij} = a_{|i-j|}$ for $|i - j| \leq k$ and $a_{|i-j|} = 0$ for $|i - j| > k$. In this paper, $k = 2$ and the following 5-band Toeplitz matrix is studied:

$$A = \begin{pmatrix} 1 & r & s & 0 & 0 \\ r & 1 & r & s & 0 \\ s & r & 1 & r & s \\ 0 & s & r & 1 & r \\ 0 & 0 & s & r & 1 \end{pmatrix} \quad (5)$$

The companion matrix of **A**is $(5 + 8r + 6s)^{-1}\boldsymbol{C}$ where **C** is given by:

$$C = \begin{pmatrix} \varphi_1 & \varphi_4 & \varphi_5 & \varphi_6 & \varphi_7 \\ \varphi_4 & \varphi_2 & \varphi_8 & \varphi_9 & \varphi_6 \\ \varphi_5 & \varphi_8 & \varphi_3 & \varphi_8 & \varphi_5 \\ \varphi_6 & \varphi_9 & \varphi_8 & \varphi_2 & \varphi_4 \\ \varphi_7 & \varphi_6 & \varphi_5 & \varphi_4 & \varphi_1 \end{pmatrix} (6)$$

Where

$$\phi_1 = 4 + 6r + 4s - 2rs - r^2 - s^2 ,$$

$$\phi_2 = 4 + 4r + 4s - s^2 - 4rs - 4r^2 ,$$

$$\phi_3 = 4 + 4r + 2s - 4r^2 - 4s^2 - 8rs ,$$

$$\phi_4 = 6r^2 + 2r + 3rs - s^2 - 2s - 1 ,$$

$$\phi_5 = 4s^2 + 2s + 4rs - 2r^2 - 3r - 1\,,$$

$$\phi_6 = -(2r^2 + s^2 + 3r + 2s + 3rs + 1),$$

$$\phi_7 = -(r + s + 1)^2\,,$$

$$\phi_8 = 4r^2 + r - 2s^2 - 3s - 1\,,$$

$$\phi_9 = 5s^2 + 3s + 4rs - 4r^2 - 4r - 1$$

Since the parametric classes of $(\boldsymbol{A}, \boldsymbol{b_1}, \xi)$ and $((5 + 8r + 6s)\boldsymbol{A}, \boldsymbol{b_1}, \xi)$ are the same, it suffices to study the class $((5 + 8r + 6s)\boldsymbol{A}, \boldsymbol{b_1}, \xi)$ with companion matrix $\mathbf{C}$ given by (6). The portfolio components of $\boldsymbol{b_{n+1}}$ for this class is given by:

$$b_{n+1,1} = b_{n,1} + \frac{\xi}{b_n^t x_n}(\phi_1 x_{n1} + \phi_4 x_{n2} + \phi_5 x_{n3} + \phi_6 x_{n4} + \phi_7 x_{n5})$$

$$b_{n+1,2} = b_{n,2} + \frac{\xi}{b_n^t x_n}(\phi_4 x_{n1} + \phi_2 x_{n2} + \phi_8 x_{n3} + \phi_9 x_{n4} + \phi_6 x_{n5})$$

$$b_{n+1,3} = b_{n,3} + \frac{\xi}{b_n^t x_n}(\phi_5 x_{n1} + \phi_8 x_{n2} + \phi_3 x_{n3} + \phi_8 x_{n4} + \phi_5 x_{n5}) \qquad (7)$$

$$b_{n+1,4} = b_{n,4} + \frac{\xi}{b_n^t x_n}(\phi_6 x_{n1} + \phi_9 x_{n2} + \phi_8 x_{n3} + \phi_2 x_{n4} + \phi_4 x_{n5})$$

$$b_{n+1,5} = b_{n,5} + \frac{\xi}{b_n^t x_n}(\phi_7 x_{n1} + \phi_6 x_{n2} + \phi_5 x_{n3} + \phi_4 x_{n4} + \phi_1 x_{n5})$$

for $n = 1,2, \ldots$.

EMPPIRICAL PERFORMANCE

Five-stock portfolios traded on the Kuala Lumpur Stock Exchange are selected for the empirical study. The Malaysian companies in the five-stock portfolios are listed in Table 1. The period of trading is from 1^{st} March 2006 until

2^{nd} August 2012, consisting of a total of 1500 trading days. The Toeplitz universal portfolio (7) is run on the five data sets.

The values of the best wealth S_{1500} achieved after 1500 trading days for data sets D, E, F, G, H are listed in Tables 2, 3, 4, 5, 6. It is observed that the best wealth in Tables 2-6 for data sets D – H are 2.7021, 9.3827, 2.0973, 5.1826 and 5.1100 respectively. The highest wealth of 9.3827 is achieved for data set E and the lowest wealth of 2.0973 is achieved for set F. The variation of wealth in Tables2-6 for data sets D-H are about 0.07, 0.68, 0.10, 0.18 and 0.32 respectively. The largest variation of wealth is observed for set E and the smallest variation for set D. Thus, changing the parameters *r* and *s* may lead to an increase or decrease of wealth in the range of 0.07-0.68 units.

A study of performance may also be made by fixing one parameter, say s and varying the other parameter *r*. The best wealth S_{1500} achieved after 1500 trading days for data set D-H are displayed in Tables7-11. The best wealth in the tables are 2.6633, 9.3199, 2.0973, 5.1397 and 4.9653 for data sets D-H. The highest wealth 9.3199 achieved is for set E and the lowest wealth 2.0973 is for set F. It is clear that the best wealth achieved depends on the data set, that is, the component stocks of the portfolio. Here the stocks in portfolio E perform well whereas the stocks in portfolio F perform below expectations. The variation of wealth in Tables 7-11 for data sets D-H are about 018, 0.46, 0.21, 0.36 and 0.36 respectively. It is observed that the variation of wealth in Tables 7-11 is larger than the corresponding variation in Tables 2-6 except for Table 8 which has a smaller variation than Table 3. By changing the parameters *r* and *s* in these tables, it is possible to increase or decrease the wealth by 0.18-0.46 units.

The Toeplitz universal portfolio is a useful alternative investment strategy to other universal portfolios. The empirical study here indicates that it is a viable strategy available to investors who wish to obtain better investment returns.

TABLE 1. The five Malaysian companies in the five-stock portfolios D, E, F, G and H.

Data Set	Malaysian Companies in Each Portfolio
D	IOI Corporation, Carlsberg Brewery Malaysia, British American Tobacco (M), Nestle (M), Digi.Com
E	Public Bank, Kulim (M), KLCC Property Holdings, Aeon Co. (M), Kuala Lumpur Kepong (KLK)
F	AMMB Holdings, Berjaya Sports Toto, Air Asia, Gamuda, Genting
G	Aeon Co.(M), British American Tobacco (M), Kulim (M), Nestle (M), Digi.Com
H	Digi.Com, Public Bank, KLCC Property Holdings, Carlsberg Brewery Malaysia, KLK

TABLE 2: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501} given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for the selected values of r which give the highest return of S_{1500} among the same r-value for thesvalue from -0.5 to 0.5 for data Set D.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
D	-0.5	-0.5	0.0864	2.6312	0.1591	0.2515	0.2764	0.1950	0.1180
	-0.4	-0.2	-0.2238	2.6394	0.1898	0.2696	0.2612	0.1727	0.1067
	-0.3	-0.3	-0.2929	2.6593	0.1925	0.2641	0.2452	0.1832	0.1150
	-0.2	-0.3	-0.1584	2.6476	0.1897	0.2586	0.2436	0.1917	0.1165
	-0.1	-0.5	-0.7026	2.7021	0.2215	0.2413	0.1971	0.1914	0.1487
	0.1	-0.5	-0.1391	2.6633	0.1899	0.2447	0.2085	0.2235	0.1334
	0.2	-0.5	-0.0979	2.6527	0.1904	0.2431	0.2005	0.2306	0.1354
	0.3	-0.5	-0.0756	2.6457	0.1919	0.2417	0.1917	0.2372	0.1375
	0.4	-0.5	-0.0616	2.6401	0.1938	0.2406	0.1825	0.2436	0.1396
	0.5	-0.5	-0.0520	2.6354	0.1959	0.2396	0.1730	0.2499	0.1416

TABLE3: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for the selected values of r which give the highest return of S_{1500} among the same r-value for thesvalue from -0.5 to 0.5 for data Set E.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
E	-0.5	-0.4	0.1501	8.7025	0.3496	0.1276	0.0688	0.1858	0.2682
	-0.4	-0.3	-1.3009	9.0232	0.2728	0.1199	0.2556	0.1996	0.1521
	-0.3	-0.3	-0.3302	9.0440	0.3514	0.1072	0.1276	0.1820	0.2318
	-0.2	-0.4	-0.4726	9.2239	0.3417	0.0977	0.1655	0.1877	0.2075
	-0.1	-0.5	-0.7207	9.3648	0.3012	0.1104	0.2185	0.2108	0.1591
	0.1	-0.5	-0.2011	9.3827	0.3765	0.0539	0.1933	0.1568	0.2195
	0.2	-0.5	-0.1440	9.3192	0.3731	0.0500	0.2113	0.1458	0.2198
	0.3	-0.1	-0.0647	9.2578	0.3951	0.0663	0.1962	0.1113	0.2311
	0.4	0.2	-0.0501	9.1584	0.3945	0.0972	0.1946	0.0884	0.2253
	0.5	0.4	-0.0447	9.0555	0.3870	0.1227	0.2072	0.0689	0.2142

TABLE4: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for the selected values of r which give the highest return of S_{1500} among the same r-value for thesvalue from -0.5 to 0.5 for data Set F.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
F	-0.5	0.5	0.0171	2.0764	0.2219	0.1392	0.0455	0.2254	0.3680
	-0.4	0.5	0.0169	2.0773	0.2163	0.1354	0.0464	0.2273	0.3746
	-0.3	0.5	0.0169	2.0777	0.2096	0.1310	0.0483	0.2293	0.3818
	-0.2	0.5	0.0172	2.0787	0.2018	0.1254	0.0504	0.2317	0.3907
	-0.1	0.5	0.0179	2.0808	0.1922	0.1181	0.0523	0.2346	0.4027
	0.1	0.5	0.0207	2.0870	0.1647	0.0961	0.0583	0.2426	0.4382
	0.2	0.5	0.0232	2.0911	0.1441	0.0789	0.0635	0.2481	0.4654
	0.3	0.5	0.0273	2.0973	0.1145	0.0528	0.0705	0.2556	0.5067
	0.4	0.5	0.0241	2.0492	0.1014	0.0621	0.1148	0.2516	0.4701
	0.5	0.5	0.0203	1.9997	0.0942	0.0771	0.1571	0.2462	0.4254

TABLE5: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for the selected values of r which give the highest return of S_{1500} among the same r-value for the$s$value from -0.5 to 0.5 for data Set G.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
G	-0.5	0.5	-0.0381	5.1826	0.1954	0.1051	0.1874	0.1237	0.3885
	-0.4	0.5	-0.0388	5.1361	0.1777	0.1134	0.1747	0.1408	0.3935
	-0.3	0.5	-0.0405	5.0813	0.1551	0.1236	0.1594	0.1620	0.4000
	-0.2	0.5	-0.0432	5.0083	0.1255	0.1371	0.1404	0.1898	0.4073
	-0.1	-0.5	0.3451	5.0879	0.3213	0.0029	0.3996	0.1094	0.1668
	0.1	-0.5	0.0746	5.0509	0.3294	0.0090	0.3704	0.0955	0.1956
	0.2	-0.1	0.0337	5.0707	0.3705	0.0082	0.3744	0.0628	0.1841
	0.3	-0.1	0.0282	5.1068	0.3691	0.0105	0.3674	0.0572	0.1958
	0.4	-0.1	0.0240	5.1397	0.3683	0.0123	0.3613	0.0521	0.2061
	0.5	-0.1	0.0206	5.1659	0.3671	0.0145	0.3551	0.0481	0.2152

TABLE6: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for the selected values of r which give the highest return of S_{1500} among the same r-value for the$s$value from -0.5 to 0.5 for data Set H.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
H	-0.5	-0.5	-0.0465	4.8304	0.1380	0.2248	0.3246	0.0721	0.2405
	-0.4	-0.3	0.5126	4.8418	0.1824	0.2198	0.2456	0.0739	0.2782
	-0.3	-0.4	2.1126	5.1100	0.3059	0.1647	0.0588	0.1647	0.3059
	-0.2	-0.4	0.1657	4.8914	0.1643	0.1973	0.3066	0.0663	0.2655
	-0.1	-0.5	0.2404	4.9672	0.1681	0.1839	0.3146	0.0591	0.2743
	0.1	-0.5	0.0600	4.9326	0.1758	0.1673	0.3159	0.0595	0.2816
	0.2	-0.5	0.0437	4.9384	0.1814	0.1580	0.3148	0.0584	0.2873
	0.3	-0.5	0.0344	4.9464	0.1872	0.1486	0.3136	0.0573	0.2933
	0.4	-0.5	0.0284	4.9556	0.1931	0.1391	0.3123	0.0560	0.2995
	0.5	-0.5	0.0242	4.9653	0.1991	0.1295	0.3109	0.0548	0.3057

TABLE7: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for ten selected values of r with the difference interval of 0.1 and given that $s = 0.1$ for data Set D.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
D	0.1	-0.5	-0.1391	2.6633	0.1899	0.2447	0.2085	0.2235	0.1334
	0.1	-0.4	-0.0895	2.6440	0.1914	0.2477	0.2142	0.2182	0.1285
	0.1	-0.3	-0.0671	2.6324	0.1952	0.2507	0.2187	0.2118	0.1236
	0.1	-0.2	-0.0545	2.6239	0.2000	0.2542	0.2228	0.2048	0.1182
	0.1	-0.1	-0.0329	2.5742	0.2056	0.2581	0.2268	0.1971	0.1124
	0.1	0.1	-0.0622	2.5513	0.2169	0.2597	0.2312	0.1816	0.1107
	0.1	0.2	-0.0275	2.5480	0.2212	0.2586	0.2316	0.1754	0.1133
	0.1	0.3	-0.0233	2.5247	0.2249	0.2577	0.2316	0.1701	0.1157
	0.1	0.4	-0.0200	2.5042	0.2280	0.2571	0.2315	0.1656	0.1178
	0.1	0.5	-0.0174	2.4863	0.2308	0.2568	0.2313	0.1615	0.1195

TABLE8: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for ten selected values of r with the difference interval of 0.1 and given that $s = 0.2$for data Set E.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
E	0.2	-0.5	-0.1440	9.3192	0.3731	0.0500	0.2113	0.1458	0.2198
	0.2	-0.4	-0.1113	9.3163	0.3845	0.0503	0.2006	0.1368	0.2278
	0.2	-0.3	-0.0930	9.3199	0.3937	0.0535	0.1898	0.1299	0.2330
	0.2	-0.2	-0.0781	9.2597	0.3930	0.0652	0.1791	0.1270	0.2357
	0.2	-0.1	-0.0678	9.2029	0.3918	0.0777	0.1682	0.1249	0.2374
	0.2	0.1	-0.0547	9.0953	0.3885	0.1044	0.1454	0.1222	0.2395
	0.2	0.2	-0.0503	9.0416	0.1188	0.1332	0.1215	0.2401	0.1188
	0.2	0.3	-0.0468	8.9873	0.3841	0.1338	0.1203	0.1211	0.2406
	0.2	0.4	-0.0439	8.9304	0.3813	0.1496	0.1070	0.1212	0.2409
	0.2	0.5	-0.0411	8.8614	0.3765	0.1665	0.0940	0.1225	0.2406

TABLE9: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for ten selected values of r with the difference interval of 0.1 and given that $s = 0.3$for data Set F.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
F	0.3	-0.5	-0.0582	1.8925	0.2710	0.1455	0.2922	0.2539	0.0374
	0.3	-0.4	0.0000	1.8792	0.2000	0.2000	0.2000	0.2000	0.2000
	0.3	-0.3	0.0406	1.8889	0.1450	0.2027	0.1262	0.1448	0.3813
	0.3	-0.2	0.0362	1.9067	0.1429	0.1891	0.1213	0.1532	0.3935
	0.3	-0.1	0.0332	1.9261	0.1401	0.1747	0.1158	0.1628	0.4066
	0.3	0.1	0.0296	1.9710	0.1328	0.1418	0.1034	0.1865	0.4356
	0.3	0.2	0.0287	1.9979	0.1282	0.1225	0.0958	0.2007	0.4527
	0.3	0.3	0.0282	2.0283	0.1233	0.1008	0.0873	0.2170	0.4717
	0.3	0.4	-0.0278	2.0619	0.1185	0.0773	0.0786	0.2354	0.4903
	0.3	0.5	0.0273	2.0973	0.1145	0.0528	0.0705	0.2556	0.5067

TABLE10: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for ten selected values of r with the difference interval of 0.1 and given that $s = 0.4$ for data Set G.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
G	0.4	-0.5	0.0304	5.1070	0.3336	0.0138	0.3540	0.0810	0.2176
	0.4	-0.4	0.0268	5.1110	0.3399	0.0137	0.3543	0.0757	0.2163
	0.4	-0.3	0.0249	5.1191	0.3476	0.0130	0.3560	0.0692	0.2141
	0.4	-0.2	0.0240	5.1275	0.3566	0.0128	0.3581	0.0617	0.2107
	0.4	-0.1	0.0240	5.1397	0.3683	0.0123	0.3613	0.0521	0.2061
	0.4	0.1	0.0251	5.1317	0.3922	0.0222	0.3618	0.0325	0.1912
	0.4	0.2	0.0245	5.0621	0.3919	0.0445	0.3487	0.0333	0.1817
	0.4	0.3	0.0242	4.9814	0.3910	0.0707	0.3332	0.0345	0.1705
	0.4	0.4	0.0243	4.8901	0.3907	0.1012	0.3155	0.0356	0.1570
	0.4	0.5	0.0248	4.7836	0.3908	0.1378	0.2945	0.0366	0.1403

TABLE11: Values of S_{1500} (using the best ξ) and portfolio vector b_{1501}given that $b_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ by using the universal portfolio generated from the 5-bandToeplitz matrix A for ten selected values of r with the difference interval of 0.1 and given that $s = 0.5$ for data Set H.

Set	s	r	Best ξ	S_{1500}	b_{1501}				
H	0.5	-0.5	0.0242	4.9653	0.1991	0.1295	0.3109	0.0548	0.3057
	0.5	-0.4	0.0213	4.9449	0.2002	0.1298	0.3085	0.0575	0.3041
	0.5	-0.3	0.0193	4.9216	0.2010	0.1305	0.3057	0.0611	0.3017
	0.5	-0.2	0.0180	4.8972	0.2018	0.1311	0.3031	0.0647	0.2993
	0.5	-0.1	0.0171	4.8693	0.2024	0.1320	0.2999	0.0693	0.2964
	0.5	0.1	0.0163	4.8020	0.2036	0.1340	0.2923	0.0806	0.2894
	0.5	0.2	0.0163	4.7603	0.2040	0.1352	0.2876	0.0880	0.2851
	0.5	0.3	0.0166	4.7113	0.2044	0.1364	0.2821	0.0969	0.2803
	0.5	0.4	0.0170	4.6514	0.2045	0.1383	0.2745	0.1090	0.2736
	0.5	0.5	-0.0184	4.5978	0.1796	0.2601	0.1288	0.2917	0.1398

REFERENCES

1. T.M.Cover, *Mathematical Finance***1**, 1-29 (1991).
2. T.M.Cover and E. Ordentlich, *IEEE Transactions on Information Theory***42**, 348-363 (1996).
3. D.P. Helmbold et al., *Mathematical Finance***8**, 325-347 (1998).
4. C.P. Tan and W.X. Lim, "Universal Portfolios Generated by the Mahalanobis Squared Divergence" in *Contributions in Mathematics and Application IV*, edited by Y. Lenbury and Nguyen V.S., East-West Journal of Mathematics, Bangkok, Thailand, 2012, pp. 225-235.

Performance of Universal Portfolios Generated by Dominant-Diagonal Matrices and Probability Distributions

Choon Peng Tan[1] and Kee Seng Kuang[2]

[1,2]Department of Mathematical and Actuarial Sciences
UniversitiTunku Abdul Rahman
JalanGenting Kelang, 53300 Setapak
Kuala Lumpur, Malaysia
tancp@utar.edu.my, kuangks@utar.edu.my

Abstract. Dominant-diagonal matrices are well-known to be positive definite and hence they generate a class of Mahalanobis universal portfolios. The objective of this paper is to study the empirical performance of some dominant-diagonal-matrix-generated universal portfolios with two parameters. Low-order universal portfolio generated by some common distributions like the lognormal and inverse Gaussian distributions are also studied with regard to their performance on selected stock-price data sets. It is possible to achieve higher wealth by using these portfolios in investment.

Keywords: universal portfolio, dominant-diagonal matrix, common probability distribution, empiricalperformance
PACS: 02.50.-r, 89.65.Gh

INTRODUCTION

In non-parametric investment, an investment strategy not depending on the stochastic model of the stock prices is needed. Such a strategy is known as a *universal portfolio*. Earlier work on universal portfolios is discussed in Cover (1991). A generalization of Cover (1991) using the Dirichlet distribution to generate a universal portfolio is given in Cover and Ordentlich(1996). The Cover-Ordentlich universal portfolio consumes a lot of computer memory and time in its implementation, where the memory and time grow exponentially with the number of stocks in the portfolio. To reduce the amount of memory and time to practical implementation, Tan (2013)proposes a finite-order universal portfolio generated by a probability distribution. It is the objective of this paper to study the empirical performance of such a portfolio.

A multiplicative-update universal portfolio is introduced by Helmbold et al. (1998).An extension to an additive-update universal portfolio is studied by Tan

and Lim (2012).This type of additive-update universal portfolio is generated by the Mahalanobis squared divergence associated with a symmetric, positive definite matrix. The focus of this paper is to study the empirical performance of the Mahalanobis universal portfolio generated by a dominant-diagonal matrix.

Some Basic Theory

A universal portfolio of*m* stocks is considered, where a portfolio $\boldsymbol{b} = (b_i)$ is a probability vector consisting of the proportions of the current wealth invested in each stock. Let $\boldsymbol{b}_n = (b_{ni})$ be the portfolio strategy used on the n^{th} trading day, where b_{ni} is the proportion of the current wealth at the beginning of the n^{th} trading day invested on stock *i,* for $i = 1, 2, \cdots, m$. Suppose $\boldsymbol{x}_n = (x_{ni})$ is the price-relative vector on the n^{th} trading day, where x_{ni} is the price relative of stock *i*which is defined as the ratio of the closing of stock *i*to its opening price on day *n*, for $i = 1, 2, \cdots, m$. The sequence of *n* price-relative vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ is denoted by $\boldsymbol{x}_1^n$. The wealth $S_n(\boldsymbol{x}_1^n)$ at the end of the n^{th} trading day is computed as $S_n(\boldsymbol{x}_1^n) = \prod_{j=1}^{n} \boldsymbol{b}_j^t \boldsymbol{x}_j$ where $\boldsymbol{b}_j^t \boldsymbol{x}_j = \sum_{i=1}^{m} b_{ji} x_{ji}$ and the initial wealth S_0 is assumed to be 1 unit.

Given $\boldsymbol{b}_1$, the Mahalanobis universal portfolio $(A, \boldsymbol{b}_1, \xi)$ generated by an $m \times m$ symmetric, positive definite matrix $A = (a_{ij})$ is the sequence of universal portfolios $\{\boldsymbol{b}_{n+1}\}$ given by:

$$\boldsymbol{b}_{n+1} = \boldsymbol{b}_n + \frac{\xi}{(\boldsymbol{b}_n^t \boldsymbol{x}_n)} C \boldsymbol{x}_n \tag{1}$$

whereξ is any real number such that $\boldsymbol{b}_{n+1} \geq \boldsymbol{0}$for $n = 1, 2, \cdots; C = (c_{ij})$ is the companion matrix of *A* defined as:

$$c_{ij} = a_{ij} - \frac{a_{i\cdot} a_{\cdot j}}{a_{\cdot\cdot}} \tag{2}$$

where $a_{i\cdot} = \sum_{j=1}^{m} a_{ij}$, $a_{\cdot j} = \sum_{i=1}^{m} a_{ij}$ and $a_{\cdot\cdot} = \sum_{i,j=1}^{m} a_{ij}$ denote the i^{th} row sum, j^{th} column sum and total sum of entries of *A* respectively (refer to Tan and Lim (2012)). Given $\boldsymbol{b}_1$ and*A*, there exists an interval of ξ such that $\boldsymbol{b}_{n+1} \geq \boldsymbol{0}$for all $n = 1, 2, \cdots$. Then (1) is the parametric family of Mahalanobis universal portfolios $(A, \boldsymbol{b}_1, \xi)$ generated by $\boldsymbol{b}_1$ and *A*.

A matrix *A* is *row/column dominant* if the modulus of each diagonal element is greater than the sum of the moduli of the corresponding off-diagonal row/column elements. A matrix *A* is *dominant-diagonal*if it is either row or column dominant. A special type of 5×5 dominant-diagonal matrix *A* given by (3) is studied in this paper:

$$A = \begin{pmatrix} r & \frac{s}{4} & \frac{s}{5} & \frac{s}{6} & \frac{s}{7} \\ \frac{s}{4} & r & \frac{s}{6} & \frac{s}{7} & \frac{s}{8} \\ \frac{s}{5} & \frac{s}{6} & r & \frac{s}{8} & \frac{s}{9} \\ \frac{s}{6} & \frac{s}{7} & \frac{s}{8} & r & \frac{s}{10} \\ \frac{s}{7} & \frac{s}{8} & \frac{s}{9} & \frac{s}{10} & r \end{pmatrix} \tag{3}$$

where $r > \frac{319}{420}s$, $r > 0$ and $s > 0$ are sufficient for A to be dominant-diagonal and hence positive definite. Let the companion matrix C be:

$$C = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 \\ \theta_2 & \theta_6 & \theta_7 & \theta_8 & \theta_9 \\ \theta_3 & \theta_7 & \theta_{10} & \theta_{11} & \theta_{12} \\ \theta_4 & \theta_8 & \theta_{11} & \theta_{13} & \theta_{14} \\ \theta_5 & \theta_9 & \theta_{12} & \theta_{14} & \theta_{15} \end{pmatrix} \tag{4}$$

and from (2),

$$\theta_1 = r - \frac{1}{d}\left[r + \frac{319}{420}s\right]^2 ,$$
$$\theta_2 = \frac{s}{4} - \frac{1}{d}\left[\left(r + \frac{319}{420}s\right)\left(r + \frac{115}{168}s\right)\right] ,$$
$$\theta_3 = \frac{s}{5} - \frac{1}{d}\left[\left(r + \frac{319}{420}s\right)\left(r + \frac{217}{360}s\right)\right] ,$$
$$\theta_4 = \frac{s}{6} - \frac{1}{d}\left[\left(r + \frac{319}{420}s\right)\left(r + \frac{449}{840}s\right)\right] ,$$
$$\theta_5 = \frac{s}{7} - \frac{1}{d}\left[\left(r + \frac{319}{420}s\right)\left(r + \frac{2414}{5040}s\right)\right] ,$$
$$\theta_6 = r - \frac{1}{d}\left[r + \frac{115}{168}s\right]^2 ,$$
$$\theta_7 = \frac{s}{6} - \frac{1}{d}\left[\left(r + \frac{115}{168}s\right)\left(r + \frac{217}{360}s\right)\right] ,$$
$$\theta_8 = \frac{s}{7} - \frac{1}{d}\left[\left(r + \frac{115}{168}s\right)\left(r + \frac{449}{840}s\right)\right] ,$$
$$\theta_9 = \frac{s}{8} - \frac{1}{d}\left[\left(r + \frac{115}{168}s\right)\left(r + \frac{2414}{5040}s\right)\right] ,$$
$$\theta_{10} = r - \frac{1}{d}\left[r + \frac{217}{360}s\right]^2 ,$$
$$\theta_{11} = \frac{s}{8} - \frac{1}{d}\left[\left(r + \frac{217}{360}s\right)\left(r + \frac{449}{840}s\right)\right] ,$$

$$\theta_{12} = \frac{s}{9} - \frac{1}{d}\left[\left(r + \frac{217}{360}s\right)\left(r + \frac{2414}{5040}s\right)\right] \quad ,$$

$$\theta_{13} = r - \frac{1}{d}\left[r + \frac{449}{840}s\right] \quad ,$$

$$\theta_{14} = \frac{s}{10} - \frac{1}{d}\left[\left(r + \frac{449}{840}s\right)\left(r + \frac{2414}{5040}s\right)\right] \quad ,$$

$$\theta_{15} = r - \frac{1}{d}\left[r + \frac{2414}{5040}s\right]^2$$

where $d = 5r + \frac{964}{315}s$. The empirical performance of the portfolio will be presented in the next section.

Let $Y_1, Y_2, \ldots, Y_m$ be m mutually independent random variables with positive moments. From Tan (2013), the orders 1, 2, 3 universal portfolios generated by $Y_1, Y_2, \ldots, Y_m$ are defined by (5), (6) and (7) respectively, namely,

$$\hat{b}_{n+1,k} = \frac{\sum_{i=1}^{m} x_{ni} E[Y_k Y_i]}{\sum_{k=1}^{m}\left(\sum_{i=1}^{m} x_{ni} E[Y_k Y_i]\right)} \quad \text{for} k = 1, 2, \cdots, m; n = 1, 2, \cdots; \tag{5}$$

$$\hat{b}_{n+1,k} = \frac{\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} x_{ni_1} x_{n-1,i_2} E\left[Y_k Y_{i_1} Y_{i_2}\right]}{\sum_{k=1}^{m}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m} x_{ni_1} x_{n-1,i_2} E\left[Y_k Y_{i_1} Y_{i_2}\right]} \quad \text{for} k = 1, 2, \cdots, m; n = 2, 3, \cdots; \tag{6}$$

$$\hat{b}_{n+1,k} = \frac{\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\sum_{i_3=1}^{m} x_{ni_1} x_{n-1,i_2} x_{n-2,i_3} E\left[Y_k Y_{i_1} Y_{i_2} Y_{i_3}\right]}{\sum_{k=1}^{m}\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\sum_{i_3=1}^{m} x_{ni_1} x_{n-1,i_2} x_{n-2,i_3} E\left[Y_k Y_{i_1} Y_{i_2} Y_{i_3}\right]} \quad \text{for} k = 1, 2, \cdots, m; n = 3, 4, \cdots; \tag{7}$$

For values of n not defined by (5), (6) and (7), the portfolio component $\hat{b}_{n+1,k}$ can be arbitrary. The empirical study is on Y_1, Y_2 and Y_3 which are independent and coming from the lognormal or inverse Gaussian distributions. The random variable Y has a lognormal (μ, σ) distribution if the probability density function is

$$f(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(lny-\mu)^2}{2\sigma^2}}, y > 0 \tag{8}$$

and

$$E[X^n] = e^{n\mu} + \frac{n^2\sigma^2}{2}. \tag{9}$$

The random variable Y has an inverse Gaussian (μ, λ) distribution if the probability density function is

$$f(y; \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi y^3}} e^{-\frac{\lambda(y-\mu)^2}{2y\mu^2}}, y > 0 \tag{10}$$

and its moment generating function is:

$$m(t) = e^{\frac{\lambda}{\mu}\left[1-\sqrt{1-\frac{2\mu^2 t}{\lambda}}\right]}. \quad (11)$$

EMPIRICAL RESULTS

Five data sets A, B, C, D and E are obtained for this study. They comprise of the daily stock prices of selected Malaysian companies traded on the Kuala Lumpur Stock Exchange and listed in Table 1.There are a total of 500 trading days for the three-stock portfolios A, B and C and the trading period is from January 1, 2003 until November 30, 2004. The trading period of the five-stock portfolios D and E is from March 1, 2006 under August 2, 2012 consisting of a total of 1500 trading days.

The empirical performance of the dominant-diagonal universal portfolio generated by A in (3) is studied by running the portfolio on data sets D and E. The initial wealth S_0 is assumed to be 1 unit. The best wealth S_{1500} achieved after 1500 trading days is displayed in Tables 2 and 3 for valid values of the parameter ξ in the interval containing $\xi = 0$and ten selected pairs of (r, s) given the uniform initial portfolio $\boldsymbol{b}_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$. The parameter ξ achieving the best wealth S_{1500} and the final portfolio $\boldsymbol{b}_{1501}$ are also displayed in the two tables. It is observed that for data set D, the wealths achieved are in the range 2.579-2.604 units. Larger proportions of wealth in the portfolio D tend to move to stocks 2 and 3 in the portfolio for better performance. For data set E, the wealth achieved is in the range 9.168-9.229 units. The largest proportion of wealth in portfolio E tends to move to the first stock for better performance. An increase of wealth of about 0.1 unit may be obtained by changing the values of r and s in the parameter vector (r, s) for portfolio D or E. It is clear that the stocks in D do not perform well compared with the outstanding performance of the stocks in E with a wealth different of 3.5 times.

The orders 1, 2 and 3 universal portfolios (5) – (7) generated by the lognormal $\big((\mu_1, \sigma), (\mu_2, \sigma), (\mu_3, \sigma)\big)$ and inverse Gaussian $\big((\mu_1, \lambda), (\mu_2, \lambda), (\mu_3, \lambda)\big)$ distributions are run on data sets A, B and C for nine selected parametric vectors. The wealth S_{500} achieved after 500 trading days is displayed in Tables 4 and 5 where the initial vector $\boldsymbol{b}_1 = (0.3333, 0.3333, 0.3334)$. The wealths achieved by the portfolios may be compared with the best-constant-rebalanced-portfolio (BCRP) wealth. The BCRP wealth for a fixed data set is the highest wealth achieved among all the constant rebalanced portfolios which are run on the data

set. The BCRP wealths for A, B and C are 1.8534, 4.2970 and 4.2970 units respectively.

The abbreviation $(\boldsymbol{\mu}, \sigma)$ is used to represent the 3 parametric vectors (μ_1, σ), (μ_2, σ) and (μ_3, σ). In Table 4, the wealths achieved by the order 1 lognormal $(\boldsymbol{\mu}, \sigma)$ universal portfolios are in the ranges of 1.1637 – 1.8534, 1.3677 – 4.2969 and 0.9596 – 4.2968 units for A, B and C respectively. The highest wealths achieved are close to the BCRP wealths for A, B and C indicating a good performance. However the variation of wealths achieved for the lognormal $(\boldsymbol{\mu}, \sigma)$ universal portfolios on B and C is large, namely, about 3 units. For instance the order 1 lognormal $\big((5,5),(10,5),(1,5)\big)$ universal portfolio achieves a low wealth of 1.5570 units compared with the highest wealth of 4.2968 achieved by the order 1 lognormal $\big((5,5),(1,5),(10,5)\big)$ universal portfolio for set B. For the order 2 lognormal $(\boldsymbol{\mu}, \sigma)$ universal portfolios, the wealth achieved are in the ranges of 1.1637 – 1.8783, 1.3738 – 4.2970 and 1.3728 – 4.2970 for A, B and C respectively. The highest wealth of 1.8783 units achieved by the order 2 lognormal $\big((10,5),(5,5),(1,5)\big)$portfolio for A exceeds the BCRP wealth of 1.8534 units. The variation of wealths of about 3 units is large for B and C, indicating that certain parametric vectors lead to poor performance. The highest wealths for orders 2 and 3 portfolios match the BCRP wealth of 4.2970 units, indicating a good performance for the parametric vectors $\big((5,5),(1,5),(10,5)\big)$. For the order 3 lognormal $(\boldsymbol{\mu}, \sigma)$ universal portfolios, the wealths achieved are in the ranges of 1.1637 – 1.8783, 1.3498 – 4.2970 and 0.9860 – 4.2970 units for A, B and C respectively. The highest wealths for B and C match the BCRP wealth of 4.2970 units whereas the highest order 3 wealth for A exceeds the BCRP wealth of 1.8534 units. A large variation of order 3 wealths is observed for B and C.

The abbreviation $(\boldsymbol{\mu}, \lambda)$ is used to represent the 3 parametric vectors (μ_1, λ), (μ_2, λ) and (μ_3, λ). In Table 5, the wealth S_{500} achieved by orders 1, 2 and 3 inverse Gaussian $(\boldsymbol{\mu}, \lambda)$ universal portfolios are displayed. Higher variation of wealths is observed for orders 2 and 3 universal portfolios for sets B and C. The highest orders 2 and 3 wealths for set A exceed the BCRP wealth of 1.8534 units. The highest wealths of 4.0579 and 4.1782 units achieved by the respective orders 2 and 3 inverse Gaussian $\big((10,10),(20,10),(50,10)\big)$ universal portfolios are close to the BCRP wealth of 4.2970 units. A similar situation is observed for the highest wealths of 3.9974 and 4.1466 units achieved by the respective orders 2 and 3 inverse Gaussian $\big((20,10),(10,10),(50,10)\big)$ universal portfolios which are close to the BCRP wealth of 4.2970 units.

Among the orders 1, 2 and 3 universal portfolios studied, there is no clear superiority in performance of one order over another order. Thus the choice of order can be arbitrary. It is obvious that the least-order universal portfolio requires the least time and computer memory in its implementation. The selection of the parametric vectors of the generating probability distribution is crucial in the ultimate performance of the universal portfolio.

TABLE 1. The three Malaysian companies in the three-stock portfolios A, B, C and the five Malaysian companies in the five-stock portfolios D and E

Data Set	Malaysian Companies in Each Portfolio
A	Malayan Banking, Genting, Amway (M) Holdings
B	Public Bank, Sunrise, YTL Corporation
C	Hong Leong Bank, RHB Capital, YTL Corporation
D	IOI Corporation, Carlsberg Brewery Malaysia, British America Tobacco (M), Nestle (M), Digi.Com
E	Public Bank, Kulim (M), KLCC Property Holdings, Aeon (M), Kuala Lumpur Kepong

TABLE 2. Wealth, S_{1500}achieved by dominant-diagonal universal portfolio generated by A in (3) for ten selected pairs of (r, s) with the best ξin the valid interval containing $\xi = 0$ and portfolio vector $\boldsymbol{b}_{1501}$ given that $\boldsymbol{b}_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ for the data set D

Set	(r, s)	Best ξ	S_{1500}	$\boldsymbol{b}_{1501}$
D	(0.1, 0.1)	-2.4507	2.57904	[0.2139 0.2489 0.2394 0.1832 0.1049]
	(0.2, 0.1)	-1.1669	2.59309	[0.2112 0.2611 0.2398 0.1828 0.1051]
	(0.3, 0.1)	-0.7673	2.59798	[0.2103 0.2620 0.2400 0.1827 0.1050]
	(0.4, 0.1)	-0.5707	2.59996	[0.2098 0.2623 0.2400 0.1827 0.1052]
	(0.5, 0.1)	-0.4547	2.60137	[0.2095 0.2626 0.2401 0.1826 0.1052]
	(0.6, 0.1)	-0.3773	2.60194	[0.2093 0.2627 0.2401 0.1826 0.1053]
	(0.7, 0.1)	-0.3227	2.60249	[0.2092 0.2628 0.2401 0.1826 0.1053]
	(0.8, 0.1)	-0.2827	2.60367	[0.2091 0.2631 0.2402 0.1826 0.1050]
	(0.9, 0.1)	-0.2507	2.60374	[0.2091 0.2630 0.2402 0.1825 0.1052]
	(1.0, 0.1)	-0.2253	2.60397	[0.2090 0.2631 0.2402 0.1825 0.1052]

TABLE 3. Wealth, S_{1500}achieved by dominant-diagonal universal portfolio generated by A in (3) for ten selected pairs of (r, s) with the best ξin the valid interval containing $\xi = 0$ and portfolio vector $\boldsymbol{b}_{1501}$ given that $\boldsymbol{b}_1 = (0.2, 0.2, 0.2, 0.2, 0.2)$ for the data set E

Set	(r, s)	Best ξ	S_{1500}	$\boldsymbol{b}_{1501}$
E	(0.1, 0.1)	-4.0333	9.17264	[0.3659 0.1214 0.1112 0.1461 0.2554]
	(0.2, 0.2)	-2.0133	9.17030	[0.3656 0.1215 0.1114 0.1462 0.2553]
	(0.3, 0.3)	-1.34	9.16797	[0.3654 0.1217 0.1115 0.1452 0.2552]
	(0.4, 0.4)	-1.0067	9.17040	[0.3656 0.1215 0.1114 0.1461 0.2554]
	(0.5, 0.5)	-0.8068	9.17282	[0.3659 0.1214 0.1112 0.1461 0.2554]
	(0.6, 0.7)	-0.6989	9.20268	[0.3626 0.1236 0.1111 0.1457 0.2570]
	(0.7, 0.9)	-0.6173	9.22761	[0.3603 0.1252 0.1110 0.1453 0.2582]
	(0.8, 1.0)	-0.5358	9.22134	[0.3612 0.1246 0.1109 0.1453 0.2579]
	(0.9, 1.1)	-0.4726	9.21462	[0.3616 0.1243 0.1111 0.1455 0.2565]
	(1.0, 1.3)	-0.4332	9.22877	[0.3598 0.1255 0.1111 0.1454 0.2582]

TABLE 4. The wealth S_{500}achieved by the orders 1, 2, 3 universal portfolios generated from the lognormal $(\boldsymbol{\mu}, \sigma)$ distribution for data sets A, B, C and nine selected vectors of $(\boldsymbol{\mu}, \sigma)$

$\sigma = 5$			Set A Order			Set B Order			Set C Order		
μ_1	μ_2	μ_3	1	2	3	1	2	3	1	2	3

1	1	1	1.56822	1.56541	1.55418	2.17065	2.19625	2.20528	1.83489	1.85355	1.86753
5	5	5	1.56822	1.56541	1.55418	2.17065	2.19625	2.20528	1.83489	1.85355	1.86753
10	10	10	1.56822	1.56541	1.55418	2.17065	2.19625	2.20528	1.83489	1.85355	1.86753
1	5	10	1.56392	1.63944	1.62518	4.29686	4.29702	4.29702	4.29675	4.29702	4.29702
1	10	5	1.16369	1.16366	1.16366	1.55707	1.59869	1.65630	0.95960	0.97994	0.98689
5	1	10	1.65396	1.63944	1.62518	4.29681	4.29702	4.29702	4.29681	4.29702	4.29702
5	10	1	1.16369	1.16366	1.16366	1.55699	1.59869	1.65630	0.95955	0.97994	0.98689
10	1	5	1.85338	1.87827	1.87827	1.36779	1.37379	1.34979	1.36651	1.37282	1.38571
10	5	1	1.85336	1.87827	1.87827	1.36773	1.37379	1.34979	1.36642	1.37282	1.38571

TABLE 5. The wealth S_{500}achieved by the orders 1, 2, 3 universal portfolios generated from the inverse Gaussian $(\boldsymbol{\mu}, \lambda)$ distribution for data sets A, B, C and nine selected vectors of $(\boldsymbol{\mu}, \lambda)$

$\lambda = 10$			Set A Order			Set B Order			Set C Order		
μ_1	μ_2	μ_3	1	2	3	1	2	3	1	2	3
10	10	10	1.56579	1.56218	1.56369	2.16256	2.20005	2.20269	1.83429	1.85813	1.87247
20	20	20	1.56627	1.56383	1.56113	2.16418	2.19649	2.20360	1.83441	1.85563	1.87144
50	50	50	1.56700	1.56517	1.55785	2.16661	2.19441	2.20443	1.83459	1.85369	1.86964
10	20	50	1.61155	1.62393	1.61872	3.66830	4.05787	4.17817	3.44854	3.96084	4.12944
10	50	20	1.25595	1.19739	1.18018	1.78280	1.68365	1.69501	1.18687	1.06274	1.02608
20	10	50	1.66514	1.64612	1.62853	3.61517	4.02966	4.16302	3.53413	3.99736	4.14655
20	50	10	1.26624	1.20229	1.18245	1.63134	1.62437	1.66798	1.08763	1.02586	1.01036
50	10	20	1.80533	1.85795	1.86772	1.59837	1.45947	1.39067	1.56272	1.44734	1.42046
50	20	10	1.76152	1.84035	1.86071	1.48412	1.41797	1.37312	1.39741	1.38436	1.39324

REFERENCES

1. T. M. Cover, *Mathematical Finance***1**, 1-29 (1991).
2. T. M. Cover and E. Ordentlich, *IEEE Transactions on Information Theory***42**, 348-363 (1996).
3. D. P. Helmbold et al., *Mathematical Finance***8**, 325-347 (1998).
4. C. P. Tan, "Performance Bounds for the Distribution-Generated Universal Portfolio" in *Proceedings of the 59th ISI World Statistic Congress*, International Statistical Institute, Netherlands, 2013, pp. 5327-5332.
5. C. P. Tan and W. X. Lim, "Universal Portfolios generated by the Mahalanobis Squared Divergence" in *Contributions in Mathematics and Applications IV* edited by Y. Lenbury and Nguyen V.S., East-West Journal of Mathematics, Bangkok, Thailand, 2012, pp. 225-235.

Designing 3^n Conjoint Choice Experiments Using Partially Confounded Factorial Designs

Chin Khian Yong[1] and Joyce Wong Kah Kei[1]

Universiti Tunku Abdul Rahman, Kuala Lumpur, 53300, Malaysia,
yongck@utar.edu.my

Abstract. Conjoint choice experiment has been use widely in marketing to assist researchers understand how people make complex judgments such as purchase decision and product valuation by posing a series of choices about product or services. Data generated from conjoint choice experiments are used to estimate marketplace behavior. Over years, different type of conjoint choice experiment had been constructed and developed such as full factorial conjoint choice experiment, fractional factorial conjoint choice experiment, completely confounded factorial conjoint choice experiment and partially confounded factorial conjoint choice experiment (PCFCCE). PCFCCE overwhelmed other designs since it allows estimation of interaction effects as well as able to estimate all effects including effects confounded with block.
Three level PCFCCE with three to nine number of factors (attributes) with two replicate and three replicate had been constructed. A laptop preference study which conducted as 3^5 PCFCCE was used to illustrate the used of three level PCFCCE. Log-Odds Transformation method was applied to estimate the corresponding coefficients and approach for analyzing a 3^n PCFCCE such as relative importance and willingness to pay was developed. The results indicated that all main effects are significant and seven out of ten first order interaction effects are significant.

1 Introduction

The modeling of consumer preferences among multi-attribute alternatives has been one of the major activities in consumer behavior. Experimental analysis has been the fundamental tool in modeling consumer preferences in marketing and economic. Traditional experimental analysis normally ask respondents which product features or attributes are important, when purchasing a product or services. Respondents are not required to make any trade-off between each feature. This is unable to reveal the true situation of the market. For example, all pro features are unable to find in one product as compared to other products such as cheaper, better in quality and nicer in outlook. Other than that, respondents normally will end up listing all the features are important. The shortcoming of traditional experimental analysis had been overcome by the development of conjoint analysis. Conjoint analysis is generally agreed to be started in 1964 with the seminal paper by Luce, a mathematical psychologist and Tukey, a statistician [4]. Conjoint analysis is then brought to marketing field by [1]. In conjoint analysis, products or services are described on a limited number of relevant attributes(characteristics) each with limited number of

levels (specific quantities). These products are called profiles. For example, in the consumer laptop preference study, some of the relevant attributes are price, battery hours, RAM, storage and screen size, while, price has three levels, RAM has two levels, battery hours has three levels, storage has one level and screen size has two levels as shown in Table 1 below.

Table 1. Laptops' Attributes and Corresponding Levels

Laptops' Attributes	Attributes' Levels
Price	RM 1800
	RM 2300
	RM 2800
RAM	2
	4
Battery Hours	6
	12
	18
Storage	640
LCD Screen Size	13
	14

Conjoint analysis shows respondents different sets of profile in order to measure consumer preferences for products' features, to forecast the likely acceptance of a product if brought to the market.

Thereafter, conjoint choice experiment (CCE) had developed as the extension of conjoint analysis. Conjoint choice experiment was introduced by [3]. In CCE two or more profiles are group in a choice sets and respondents are required to choose the most preferable profile. Hence, respondents are required to make trade-offs when choosing, which mimic the real market decision. This enable us to understand how human beings make judgment when presenting them with a series of choices about product and services. CCE reveals better information on consumer preferences when purchasing goods and services. An example of choice set for the laptop example is shown in Table 2. The choice set shown below is form by three profile and one neither options.

Table 2. Sample of Choice Sets for Laptop

1.1	**Option A**	**Option B**	**Option C**	**Option D**
Battery Life	6	12	18	
LCD Screen Size	13	14	15	
Price	1800	2300	2800	None of it
Storage	320	640	1000	
RAM	2	4	8	
I would choose :				

In CCE, respondents are required to evaluate a number of choice sets and each choice set are independent. Respondents need to choose the most preferable alternatives from each choice sets. From the data collected, the preferences of respondents are derived in terms of utilities. Utility is a numerical expression representing the satisfaction respondents have on an attribute. The higher the utility value, the bigger the satisfaction, vice versa.

An effective CCE is an experiment that able to reflect a better image of consumer preferences with the following characteristics [2]:

1. level balance (each level occurs equally often within each factor)
2. orthogonal (parameter estimates are uncorrelated)
3. minimal level overlaps (minimal repeated level in each choice set)
4. utility balance (swapping and re-labelling of levels ensure no alternative will dominate the other alternative)

CCE should provide as much information as possible on the relevant attributes. However, problem exists when the number of attributes increases; the numbers of choice sets will also increase proportionally. Hence, each respondent is required to answer more choice sets. This increase in quantity of information is balanced off with decrease in quality of information due to effects such as boredom [7]. Over decades, researchers had work onto to it in order to have an efficient design that can obtain maximum information from the minimum number of choice sets.

Designs such as fractional factorial designs and completely confounded factorial designs are view as a solution to handle large number of attributes. Fractional factorial design is an experimental design that only runs a fraction of complete factorial design whereas completely confounded factorial design is a design that arranges factorial experiment into blocks, where the block size is smaller than the number of treatment combinations in a full factorial design. Both designs are able to reduce the number of runs yet, completely confounded factorial designs are more efficient than fractional factorial designs because fractional factorial designs can't estimate interaction effects and have larger bias [8]. Yet, completely confounded factorial designs which have the same sets of interactions confounded in all replicates causes all the information on confounded interactions are lost, because it is impossible to separate confounded effects from block effects. For example, to construct a three replicates 3^5 design with three blocks, there are a total number of thirteen effects being confounded. Information on the thirteen effects is lost.

Partially confounded factorial designs are view as extension of completely confounded factorial design in CCE. Partially confounded means different sets of interactions are confounded in different replicates and the confounded interactions can be recovered from those replicates in which they are not confounded. Partially confounded factorial conjoint choice experiments (PCFCCE) are able

to handle large number attributes and all effects can be estimated as well as estimation of interaction effects. In the past, partially confounded factorial design in CCE has been focusing on two levels partially confounded factorial designs [6]. In this study, partially confounded factorial conjoint choice experiments are extend to three levels partially confounded factorial designs.

Thus, the following are the objectives of this study:

1. To develop a construction method for 3^N partially confounded factorial designs that can be applied to conjoint choice experiments.
2. To develop an approach for analyzing 3^N partially confounded factorial designs in conjoint choice experiment.

2 Construction of 3^N Partial Confounded Factorial Conjoint Choice Experiment

Partially Confounded factorial design is view as a solution to completely confounded factorial design. Completely confounded design is constructed by using the technique of confounding effects into blocks and all the effects confounded is unable to be estimated. On the other hand, partially confounded factorial design is able to estimate all the effects including those confounded effects by confounding different effects in each replicate. In other words, the confounded effects in one replicate can be estimated from other replicates. Hence, the prior criteria for a partial confounded factorial design is the number of replicate must be more than one(at least two) and blocks of each replicate is constructed by using p different effects. For example, let 2^2 design with two replicates and two blocks for each replicate, first replicate confounded by effect AB and second replicate confounded by effect B. Therefore, first replicate can use to estimate effects A and B; second replicate can use to estimate effect A and AB. All the effects in 2^2 design are estimated.

When applying partially confounded factorial design to conjoint choice experiments, two or more treatment combinations are positioned in a choice set. Each treatment combination represents one choice(alternative). It is necessary (1) to ensure that the treatment combination and its respective treatment combinations are in the same block or same choice sets and (2) all effects are able to be estimate when designing or none of the effects are confounded in every block in each replicate when designing CCE as partially confounded factorials [9].

In order to have the treatment combination and its respective treatment combinations in the same block or same choice sets for 3^N design or requirement (1), the word length, l must have a modulus of 3 equal to 0 or With the restriction of the the first nonzero α_i must be unity in 3^N design, there are $\sum_{i=0}^{N-1} 3^i$

effects that can be chosen to be the p independent effects to be confounded with blocks. In this study, only effects with word length of multiple three can be chosen to be confounded with blocks to fulfill requirement (1). In 3^N design, there are total of $\sum_{i=0}^{N-2} 3^i$ effects with word length of multiple of 3.

For example, in a 3^5 design, there are total of 283 effects with only 40 effects that can be chosen to be confounded with blocks. To construct an 3^p incomplete blocks, there are $p + (3^p - 2p - 1)/2$ effects to be confounded where all effects are of word length, In addition, each replicate must have the same number of blocks. Hence, firstly, it is important to determine the number of word length of multiple three effects and then determine the number of blocks that can be generated in each replicate with the condition that none of the effects are confounded in every block in each replicate. Table 3 shows the possible design of a two replicate 3^N partially confounded factorial CCE that fulfill all the requirement mentioned above. More replicates are needed to further reduce the number of choice sets in a fraction in order to fulfill requirement(2).

The choice sets constructed in this study have four options per choice sets including the treatment combination, two compliments of the particular treatment combination and none of the above alternative option. None of the above alternative option is added in each choice set is to mimic the real situation when sometimes consumers do not make a choice based on the alternatives given. Each block represents one questionnaire and every questionnaire is answer by different individuals.

Table 3. Defining contrast for two replicate 3^N partially confounded factorial CCE which ensure no same effects is confounded in the two replicates.

		Replicate 1		Replicate 2			
n	**p**	**Defining Contrast**	**Other Effects**	**Defining Contrast**	**Other Effects**	**B**	**CS**
3	1	ABC	-	AC^2	-	3	3
4	1	ABC	-	ABD	-	3	9
5	1	$ABCDE^2$	-	$ABCD^2E$	-	3	27
	2	ABC^2DE, AB^2DE^2	ACD , BCE	ABC^2D^2, $ABCDE^2$	ABE , CDE	9	9
6	1	ABCDEF	-	$ABCDE^2$	-	3	81
	2	ABCDEF, AB^2CD^2	$ACEF^2$, BDE^2F^2	ABC^2D^2 , $ABCDF^2$	ABF , CDF	9	27

Note:
n: Number of Factors
p: Number of independent effects used as defining contrast
B: Number of blocks in each replicate
CS: Number of choice sets in a block

In this study, a data set from a laptop preference study on respondent's preferences was conducted as a partial confounded factorial CCE. The main objective of this study is to evaluate the effects of the laptops attributes toward

respondent's decisions making. In this study, five main laptop's attributes was included since it is impossible to include all attributes in a CCE. The associated attributes are Price (Factor A), Random Access Memory (Factor B), Storage (Factor C), LCD Screen Size (Factor D) and Battery Life (Factor E), each with three levels (as shown in Table 4).

Table 4. Laptops' Attributes and Corresponding Levels in the Choice Sets

Laptops' Attributes	**Factor Levels**
Price	RM 1800
	RM 2300
	RM 2800
Random Access Memory (RAM)	2 GB
	4 GB
	8 GB
Storage	320 GB
	640 GB
	1000 GB
LCD Screen Size	13"
	14"
	15"
Battery Life	6 hours
	12 hours
	18 hours

The total number of laptops descriptions based on the five attributes are $3^5 = 243$. The principal of CCE is to obtain information on the attributes without asking respondents to answer more than a moderate number of choice sets. Other than that, information overload will leads to respondent fatigue and boredom, consequently affecting the precision of the results [7]. Hence, it is inappropriate for each respondent to evaluate all 243 alternatives. Thus, confounded factorial CCE is used and the 243 descriptions are distributed over nine blocks, left 27 descriptions in each fractions. In order to estimate all the effects, partially confounded CCE is used in this laptop preference study. A PCFCCE must have at least two replicates and each replicate must have sufficient number of respondents. Thus, only Two replicates are constructed in this study due to cost constraints.

The construction of the design can be found in Table 3 in $n = 5$ and $p = 2$ row. In each replicate, two different independent effects are chosen to be confounded with blocks forming a total of nine blocks in each replicate. The chosen confounded independent effects are ABC^2DE and AB^2DE^2 for replicate one while ABC^2D^2 and $ABCDE^2$ for replicate two. In addition, there are two other generalized interactions which are confounded with blocks which are ACD and BCE for replicate one whereas ABE and CDE for replicate two. Confounded effects in replicate one: ABC^2DE, AB^2DE^2, ACD and BCE are able to esti-

mate by using information from replicate two. On the other side, confounded effects in replicate two: ABC^2D^2, $ABCDE^2$, ABE and CDE can be estimated from replicate one. Hence, all the effects in this laptop preference study can be estimated.

Since the chosen confounded effects are of word length, $l \bmod 3 = 0$, the laptops descriptions are arranged in such a way that all compliment combinations are in the same fractions. Each fraction has a total of 27 treatment combinations or it can also known as 9 group of compliment combinations where each group consists of 3 treatment combinations. Other than that, each group of compliment combinations (total of three laptops descriptions) plus a null option or base alternative form a choice sets. This ensure that none of the levels in each choice sets are overlap and increase the efficiency of the design [8]. This form the nine choice sets and an extra choice set is added to each block for reliability check but it is not included in data analysis. While, each fraction represents a questionnaire and there are a total of 18 questionnaires. In other word, each respondent is responsible to respond to one questionnaire with a total of 10 choice sets and each choice sets with four alternatives. Besides, the orders of attributes in each choice sets are randomized to avoid response bias and obtain more precise results. Table 5 shows one of the choice sets from the questionnaire. With this design, all the main effects and first-order interactions are able to estimate.

Table 5. A fraction of the choice sets from questionnaire

1.1	**Option A**	**Option B**	**Option C**	**Option D**
Battery Life	6	12	18	None of it
LCD Screen Size	13	14	15	
Price	1800	2300	2800	
Storage	320	640	1000	
RAM	2	4	8	
I would choose :				

3 Analysis for 3^N Partial Confounded Factorial Choice Based Conjoint Experiments

Respondents of the laptop preference study were separated into two main groups which are students and employees who ranging between 19 and 25 years old and each of them only responded once to the questionnaire. Students were randomly chosen from Faculty of Engineering and Science (FES) of University Tunku Abdul Rahman, whereas employees were randomly chosen from Underwriter and Claim departments of AIA Company. Students were responded to 13 fractions of the questionnaire and each questionnaire was responded by 25 students. On

the other hand, employees from AIA had responded to 5 fractions of the questionnaire and each questionnaire was also responded by 25 employees. Hence, there were 325 and 125 respondents who were students from UTAR and employees from AIA respectively. This summed up to the total of 450 respondents for the laptop preference study. However, some of the respondents did not pass the reliability check and left a total of 369 questionnaires, with an average of 20 respondents per fraction that can be used for analysis. The analysis of the laptop preference study was done by using Log-Odds Transformation method.

Log-Odds Transformation method was developed by [5] used for data analyzing for data with single fraction. Log-odds transformation linearizes the multinomial logit model into an additive model of the systematic components:

$$\ln\left(\frac{\pi_{ispl}}{\pi_{isp0}}\right) = \ln\left(\frac{\frac{e^{\nu_{ispl}}}{\sum\limits_{k\epsilon C_{isp}} e^{\nu_{ispl}}}}{\frac{e^{\nu_{isp0}}}{\sum\limits_{k\epsilon C_{isp}} e^{\nu_{ispl}}}}\right) = \ln(e^{\nu_{ispl}} - e^{\nu_{isp0}}) = \nu_{ispl} - 0 = x'_{ispl}\beta \qquad (1)$$

where π_{ispl} is the true probability that alternative one is selected from the isp^{th} choice set and π_{isp0} is the true probability for the base alternative in choice set isp^{th} .

Covariance structure proposed by [5] for the multi-vector multinomial distribution is as follow:

$$Cov(P^{(n)}_{\sim l}, P^{(m)}_{\sim k}) = \begin{cases} \frac{1}{N}p^{(n)}_l(1-p^{(n)}_l), & n=m,\ l=k \\ -\frac{1}{N}p^{(n)}_l p^{(m)}_k, & n=m, l\neq k \\ \frac{1}{N}p^{(nm)}_{lk} - p^{(n)}_l p^{(m)}_k, & n\neq m \end{cases} \qquad (2)$$

where $p^{(n)}_{\sim l}$ and $p^{(m)}_{\sim k}$ are two multinomial vectors symbolize proportion,n and m are indexes symbolizing choice set, l and k are indexes symbolizing alternatives within choice set as well as N represents number of respondents.

Based on [8],for confounded factorial CCE, multiple fractions are present in the designs and separate questionnaires are developed for each fraction/block of subjects are assigned to each fraction. Thus, to apply the log-odds transformation method when blocks are assumed random, we should incorporate the random block effects, u into the model. The corresponding mixed model for confounded factorial CCE is

$$Y = X\beta + Zu + \varepsilon \qquad (3)$$

where X is matrix associated with attribute levels of main effects and first-order interaction effects and Z is matrix associated with random block effects u and u is assumed to be normally distributed with mean zero and covariance matrix

u, $G = Cov(u)$. ε is normally distributed with mean zero and covariance matrix $R = Cov(\varepsilon)$. Thus,

$$Cov(Y) = ZGZ' + R. \tag{4}$$

This approach can also be used in PCFCCE.

In the laptop preference study, each respondent had answered a total of ten choice sets, nevertheless only nine choice sets were used in data analysis; in view that one of the choice set was served as internal reliability check. Each choice sets consists of four alternatives. Thus, with each fraction of respondents responded to 36 (9 choice sets $\times$ 4 alternatives) responses, there were total of 13284 responses. On the other hand, each fraction consists of 36 observed proportions and hence there were a total of 648 (36×18 fractions) observed proportions. Yet, the fourth alternative (Option D) in the choice set was omitted and set as based alternatives, leaving 27 observed proportions in each fraction and a total of 486 observed proportions. Thus, for each choice set, A_n where $n = 1, ..., 9$, the vector of log-odds value are defined as:

$$Y_n(p) = \begin{cases} \ln\left(\frac{p_{n1}}{p_{n4}}\right) \\ \ln\left(\frac{p_{n2}}{p_{n4}}\right) \\ \ln\left(\frac{p_{n3}}{p_{n4}}\right) \end{cases}$$

where the proportion of each alternatives ,p_{n1} ,p_{n2} and p_{n3} (Option A, B and C) was divided by the proportion of the base alternative, p_{n4} (Option D) in a choice sets.

Based on [8],for confounded factorial CCE, multiple fractions are present in the designs and separate questionnaires are developed for each fraction/block of subjects are assigned to each fraction. Thus, to apply the log-odds transformation method when blocks are assumed random, we should incorporate the random block effects, u into the model. The corresponding mixed model for confounded factorial CCE is

$$Y = X\beta + Zu + \varepsilon \tag{5}$$

where X is known matrix associated with attribute levels of main effects and first-order interaction effects and Z is matrix associated with the random block effects u and u is assumed to be normally distributed with mean zero and covariance matrix u, $G = Cov(u)$. ε is normally distributed with mean zero and covariance matrix $R = Cov(\varepsilon)$. Thus,

$$Cov(Y) = ZGZ' + R \tag{6}$$

For the proportion of times that the alternatives are selected, the corresponding covariance matrix is a direct matrix calculation in the presence of

data:

$$\widehat{Cov}(p) = \frac{1}{N(N-1)}\left(D'D - \frac{1}{N}D'11'D\right) \tag{7}$$

where D is a matrix with each column represents alternatives in all choice sets while each row corresponds to a respondent in presented order . Elements in D are set to be either 1s or 0s: 1s for selected alternatives and 0s otherwise.

For each choice set, A_n where $n = 1, ..., N$, the vector of log-odds value are defined as:

$$Y_n(p) = \begin{cases} \ln\left(\frac{p_{n1}}{p_{nJ}}\right) \\ \ln\left(\frac{p_{n2}}{p_{nJ}}\right) \\ \ln\left(\frac{p_{n3}}{p_{nJ}}\right) \\ ... \\ ... \\ \ln\left(\frac{p_{n(J-1)}}{p_{nJ}}\right) \end{cases} \tag{8}$$

where J is the number of alternatives in each choice set and p_{nJ} is the base alternative which each proportion is divided by it.

When Y'=$[Y_1(p) \quad Y_2(p) \quad ... \quad Y_N(p)]$ where the elements are functions of p with continuous derivate, there exists a consistent estimator for the covariance matrix of Y, $Cov(Y)$ is

$$\widehat{Cov}(Y) = H[\widehat{Cov}(p)]H' \ , \ H = \left[\frac{\partial Y(\pi)}{\partial \pi}\right]_{\pi=p} \tag{9}$$

where $H_n(n = 1, 2, ..., N)$ is:

$$H_n = \begin{bmatrix} \frac{1}{p_{n1}} & 0 & 0 & ... & 0 & -\frac{1}{p_{nJ}} \\ 0 & \frac{1}{p_{n2}} & 0 & ... & 0 & -\frac{1}{p_{nJ}} \\ 0 & 0 & \frac{1}{p_{n3}} & 0 & ... & -\frac{1}{p_{nJ}} \\ . & & & & & \\ . & & & & & \\ 0 & 0 & ... & 0 & \frac{1}{p_{n(J-1)}} & -\frac{1}{p_{nJ}} \end{bmatrix} \tag{10}$$

and the H-matrix is:

$$\mathbf{H} = \begin{bmatrix} H_1 & 0 & 0 & 0 & ... & 0 \\ 0 & H_2 & 0 & 0 & ... & 0 \\ 0 & 0 & H_3 & 0 & ... & 0 \\ 0 & 0 & 0 & H_4 & ... & 0 \\ . & . & . & . & . & . \\ 0 & 0 & ... & 0 & 0 & H_N \end{bmatrix} \tag{11}$$

The estimates of the model coefficients and their covariance matrix is:

$$\hat{\beta} = |X'[\widehat{Cov}(Y)]^{-1}X|^{-1}X'[\widehat{Cov}(Y)]^{-1}Y \tag{12}$$

$$\widehat{Cov}(\hat{\beta}) = |X'[\widehat{Cov}(Y)]^{-1}X|^{-1} \tag{13}$$

4 Estimation of 3^N partial confounded Factorial Conjoint Choice Experiment

The result of the laptop preference study obtained from log-odd transformation method. From the result obtained, all main effects were statistically significant (p-value $< \alpha = 0.05$). On the other hand, seven out of the ten first order interaction effects were also statistically significant (p-value $< \alpha = 0.05$), first order interaction effects that were significant were Battery Life and LCD Screen Size, Battery Life and Price, Battery Life and RAM, LCD and RAM, Price and Storage, Price and RAM as well as Storage and RAM. Thus, it was important to estimate first order interactions effects to avoid large bias to the main effects.

Table 6. The respective odds and percentage of battery life with given level

Battery Life Level (hours)	Respective Odds	Percentage (%)
6	$e^{(-0.3193\times 6)} \approx 0.1472$	85.5172%
12	$e^{(-0.3193\times 12)} \approx 0.02167$	12.5943%
18	$e^{(-0.3193\times 18)} \approx 0.00319$	1.8540 %

The effect of odds of a unit increase in battery life was $e^{\beta_1} = e^{-0.3193} = 0.727$, holding other effects constant. It can also interpreted as the odds of purchasing the product were decreasing as battery life increases. The odds of respondents choosing laptop with one unit less of battery life (e.g. 1 hours) were 1.38 times the odds of respondents choosing laptop with one unit more of battery life (e.g. 2 hours). Respondents were more favorable to laptop with lower battery life. In the laptop preference study, 86% of the respondents were expected to be likely to battery life of 6 hours, 13% were favor to battery life of 12 hours and 2% were likely to battery life of 18 hours. The reason behind was the interaction effect between RAM and battery life. When RAM life increased by one unit, the rate of change of odds with respect to one unit increased in battery life, increased by a factor of 0.9966678, holding other effects at fixed value. The rate of change of odds with respect to one unit increased in battery life decreased as the RAM increased. Hence, to increase the rate of change of odds, battery life increased and RAM decreased or RAM increased and battery life decreased. In the laptop preference study, RAM was more important as compared to battery

Table 7. The respective odds and percentage of LCD screen size with given level

LCD Screen Size Level (inches)	Respective Odds	Percentage (%)
13	$e^{(0.07364\times 13)} \approx 2.6046$	30.9111%
14	$e^{(0.07364\times 14)} \approx 2.80362$	33.2732%
15	$e^{(0.07364\times 15)} \approx 3.01786$	35.8158 %

life (as discussed in relative importance section later). So, with high RAM, respondents were favorable to battery life in low unit .

A unit increase in LCD screen size affects the odds of a positive response multiplicatively by factor of $e^{0.07364} = 1.07642$, holding other effects at fixed value. LCD screen size increased by 1 inch, respondent's desire to purchase the product was affected by 1.0764155. It can also deduced as the odds of respondents choosing laptop with one inch more LCD screen size (e.g. 2 inch) were 1.08 times the odds of respondents choosing laptop with one inch (e.g. 1 inch) less LCD screen size. Respondents preferred laptops with larger LCD screen size. 31%, 33% and 35% of respondents in the laptop preference study were expected to favor to products with LCD screen size of 13 inches, 14 inches and 15 inches respectively.

Table 8. The respective odds and percentage of price with given level

Price Level (RM)	Respective Odds	Percentage (%)
1800	$e^{(-0.001511\times 1800)} \approx 0.06585$	59.1604%
2300	$e^{(-0.001511\times 2300)} \approx 0.03093$	28.7877%
2800	$e^{(-0.001511\times 2800)} \approx 0.014528$	13.0519 %

The coefficient of price, holding other effects constant showed that a unit increased in price the odds of respondents choosing the product was affected by multiplicative factor of 0.9984898. The odds of respondents choosing laptop with RM 1 lesser were 1.0015 times the odds of respondents choosing laptop with RM 1 extra. Respondents were more favorable to laptop with lower price. In the laptop preference study, 59% of the respondents were expected to be likely to price of RM 1800, 29% were favor to price of RM 2300 and 13% were likely to price of RM 2800. Besides, the odds of respondents choosing the cheapest laptop (RM 1800) were 0.06585/0.014528=4.5326 times the odds of respondents choosing the most expensive laptop (RM 2800).

A unit increased in storage affect the odds of respondents choosing the product by 1.0015899, holding other effects constant. The odds of respondents choosing laptop with an extra gigabyte of storage (e.g. 2 GB) was 1.002 times the odds of respondents choosing laptop with a gigabyte less in storage (e.g. 1 GB). In the laptop preference study, 18% of respondents were expected to choose laptop with 320 GB storage, 30% of respondents were favor to laptop with 640

Table 9. The respective odds and percentage of storage with given level

Storage Level (GB)	Respective Odds	Percentage (%)
320	$e^{(0.001588596\times 320)} \approx 1.662547$	17.8318%
640	$e^{(0.001588596\times 640)} \approx 2.764062$	29.6463%
1000	$e^{(0.001588596\times 1000)} \approx 4.896869$	52.5219 %

GB and 53% of respondents choose laptop with 1000 GB storage in the laptop preference study. Hence, respondents were favor to laptop with higher storage.

Table 10. The respective odds and percentage of random access memory (RAM) with given level

RAM Level (GB)	Respective Odds	Percentage (%)
2	$e^{(0.8181\times 2)} \approx 5.1360$	0.7061%
4	$e^{(0.8181\times 4)} \approx 26.3787$	3.6269%
8	$e^{(0.8181\times 8)} \approx 695.8349$	95.6671 %

The coefficient for RAM says that, holding other effects at a fixed value, showed that the odds of respondents purchasing the laptops increased by 2.267 for one-unit increase in RAM. In the laptop preference study, the odds of respondents choosing laptop with 8 GB were 135.48 ($= \frac{695.83}{5.136}$) times the odds of respondents choosing laptop with 2 GB. The table above also shows that 95% of the respondents in the laptop preference study were expected to choose laptop with 8 GB whereas only 0.71% and 3.63% were expected to choose laptop with 2 GB and 4 GB respectively. Hence, changed of unit in RAM were significant to respondents; small changed in the unit of RAM increased the odds of respondents choosing the product drastically.

When LCD screen size increased by one unit, the rate of change of odds with respect to one unit increased in battery life, $Y_2 - Y_1|X_1$increased by a factor of $e^{\beta_{12}} = e^{0.027116904} = 1.0275$, holding other effects at fixed value. It can also be interpreted as the rate of change of odds with respect to one unit increased in LCD screen size increased by a factor of 1.0274879 when the battery life increased by one unit. However, just presenting this result would be rather incomplete. Hence, it was wise to add estimates of the effect of LCD screen size at different inches, 13 and 15 inches.

The rate of change of odds with respect to one unit increased in battery life when the LCD screen size was 15 inches was 1.05573 ($= \frac{1.5019}{1.4227}$) when the LCD screen size was 13 inches. Thus, larger LCD screen size was more favorable to the rate of change of odds with respect to one unit increased in battery life.

The coefficient for interaction terms between battery life and price showed that, when price increased by RM 1, the rate of change of odds with respect to one unit increased in battery life, affected by a factor of $e^{\beta_1 3} = e^{0.00001644} = 1.000016$, holding other effects constant. It can also be deduced as the rate of

change of odds with respect to one unit increased in price was increased by a factor of 1.0000164 when battery life increased by one hour. The rate of change of odds with respect to one unit increased in battery life when the price was at RM 2800 was 1.01657($= \frac{1.030}{1.047}$) times when the price was at RM 1800.

When RAM increased by one unit, the rate of change of odds with respect to one unit increased in battery life, increased by a factor of $e^{\beta_{15}} = e^{-0.003338} = 0.997$, holding other effects at fixed value. The rate of change of odds with respect to one unit increased in battery life decreased as the RAM increased; the rate of change of odds with respect to one unit increment in RAM decreased as the battery life increased. The rate of change of odds with respect to one unit increment in RAM when the battery life was at 6 hours was 1.04 times when the battery life was at 18 hours. Hence, lesser battery life was more likely to the rate of change of odds with respect to one unit increment in RAM.

The coefficient for interaction terms between LCD screen size and RAM shows that, when LCD screen size increases by 1 inches, the rate of change of odds with respect to one unit increase in RAM, affected by a factor of $e^{\beta_{25}} = e^{-0.035} = 0.966$, holding other effects constant. The rate of change of odds with respect to one unit increased in RAM decreased as the LCD screen size increased; the rate of change of odds with respect to one unit increment in LCD screen size decreased as the RAM increased. The rate of change of odds with respect to one unit increased in RAM when the LCD screen size was at 13 inches was 1.07($= \frac{0.6344}{0.5915}$) times when the LCD screen size was at 15 inches.

5 Conclusion

In this study, a data set from laptop preference study on respondents' preference on laptop was conducted as PCFCCE. The constructed 3^5 designs was applied in the laptop preference study with five attributes which are: battery life, LCD screen size, price, storage and RAM, all with three levels. Data had been collected by asking respondents to answer a questionnaire with ten choice sets and four alternatives each. There are a total of 18 questionnaires (blocks) and answered by 369 respondents. Each respondent responded to the questionnaire once. The estimation of the coefficients was obtained based on Log-Odds Transformation method which is discussed in Chapter 4. Estimation was done on main effects and first order interactions, higher order interactions are neglected. From the result obtained, all the main effects and seven first order interactions are significant (p-value $< \alpha =$0.05).

In this study, all the attributes in the laptop preference study are quantitative attributes. The study can be extent by focusing on qualitative attributes such as type of graphic cards and brand of the laptops or mixture between qualitative and quantitative attributes. Design construction on qualitative attributes is same as quantitative attributes; yet, method for estimation and analysis of

quantitative attributes are yet to be develop. Besides, in this study, log-odds transformation method was incorporated for coefficients estimation. However, for log-odds transformation method, large samples size is required to increase estimation accuracy. The average respondents for each questionnaire can be increased to about 602 respondents for higher accuracy in the estimation. Analysis in this study only focuses on overall respondent's preferences due to time limitation. Hence, analysis can be further extent to compare respondents preferences of different group such as compare between different gender, different courses and different age-groups. In the meanwhile, other appropriate method can be developed to construct 3^N PCFCCE design such as Markov Chain Monte Caarlo to replace simpler and less reliable design procedures. Other than that, other efficient approaches on analyzing 3^N PCFCCE can also be explored such as Generalized Linear Model.

References

1. P. E. Green and V. R. Rao. Conjoint measurement for quantifying judgemental data. *Journal of Marketing Research*, 8:355–363, 1971.
2. J. Huber and K. Zwerina. The importance of utility balance in efficient choice design. *Journal of Marketing Research*, 33:307–317, 1996.
3. Jordan J. Louviere and G. Woodworth. Design and analysis of simulated consumer choice or allocation experiment : An approach based on aggregate data. *Journal of Marketing Research*, 20:350–367, 1983.
4. R. D. Luce and John W. Tukey. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1:1–27, 1964.
5. G. Montopoli and D.A. Anderson. The analysis of discrete choice experiments with correlated error structure. *Comunication Statistics, Theory and Method*, 30(4):615–626, 2001.
6. C. L. Ng. Designing conjoint choice experiment using partially confounded factorial. Master's thesis, Universiti Tunku Abdul Rahman, 2011.
7. Z. Sandor and M. Wedel. Designing conjoint choice experiment using managers' prior beliefs. *Journal of Marketing Research*, 38:430–444, 2001.
8. C. K. Yong. *Designing Conjoint Choice Experiments Using Confounded Factorial Designs*. PhD thesis, Department of Statistics University of Nebraska Lincoln, 2004.
9. C. K. Yong, K.M. Eskridge, and C. R. Calkin. Confounded factorial designs useful for conjoint choice experiments. In *Proceedings of The Joint Statistical Meetings,*, pages 2453–2460, Minneapolis, MN, August 2005. Joint Statistical Meetings.

Determination of Motor Insurance Rates

W. Y. Pan[1], H. C. Soo[2], and A. H. Pooi[3]

[1] UniversitiTunku Abdul Rahman, Malaysia
[2] Heriot-Watt University Malaysia
[3] Sunway University Business School, Malaysia

Abstract. Consider the third party Motor Insurance data fromSweden for 1977 described by Andrews and Herzberg in 1985. These data show the claim amount per insured (C) when the kilometers travelled per year (K), geographical zone (Z), no claims bonus (B) and make of the car (M) are given. As the variables Z and M are categorical and having respectively 7 and 9 categories, we may code them by using respectively the vectors $(Z_1, Z_2, \dots, Z_6)$ and $(M_1, M_2, \dots, M_8)$ of binary variables. The variable C is next modelled to be dependent on $\boldsymbol{X}^* = (K, Z_1, Z_2, \dots, Z_6, B, M_1, M_2, \dots, M_8)$ via a conditional distribution which is derived from a 17-dimensional power-normal distribution. A suitably chosen quantile of the conditional distribution may then be used to determine the motor insurance rate when the value of $\boldsymbol{X}^*$is given.

1. Introduction

Several authors have considered the problem of finding fair and accurate tariffs based on the car insurance data which include the numbers of claims and the total costs for the claims. Some authors studied the car insurance data using decision-trees and combination of regression techniques ([1], [2], [5] and [6]). Assuming that the number claims is Poisson distributed and the cost for individual claims is gamma distributed, Jørgensen and Souza [4] modelled the expected cost μ of claims per insured unit as a function of the explanatory variables. Smyth and Jørgensen [8] modelled the dispersion of the costs and the mean of the costs simultaneously using the double generalized linear models. They applied the double generalized linear models to the Swedish third party automobile portfolio of 1977 to estimate the tariffs. These authors used the main effects model as their final model because the model which includes interactions is too complex for practical use in setting insurance tariffs.

This paper uses yet another method to estimate the motor insurance rate based on the Swedish data for 1977. In using the method, initially we code the qualitative explanatory variables using vectors of binary variables. A multivariate power-normal distribution is next fitted to the $(k\text{x}1)$ vector $\boldsymbol{y}$ of which the initial components are given by the components in the vectors of binary variables, and the original quantitative variables, and the last component is given by the amount of claims per insured. Let $\boldsymbol{X}^*$ be the vector formed by

the $k-1$ initial values of $\boldsymbol{y}$. When the value of $\boldsymbol{X}^*$ is given, we may find a conditional distribution for the last component y_k which represents the amount of claims per insured.

The $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points of the conditional distribution may be regarded as the lower and upper limits of the nominally $100(1-\alpha)\%$ in-sample prediction interval for the amount of claims per insured.

From the Swedish data, we can form a total of 2183 values of $\boldsymbol{X}^*$. Thus we can obtain a total of 2183 prediction intervals. The proportion of prediction intervals which cover the observed amount of claims per insured may be used to estimate the coverage probability of the prediction interval. Meanwhile the average length of the 2183 prediction intervals may be used to estimate the expected length of the prediction interval.

When the estimated coverage probability is close to the target value $1-\alpha$ and the average length is small, the prediction interval is considered to be able to predict the future amount of claims per insured. It turns out that when we choose $\alpha = 0.05$, the estimated coverage probability is 0.937 which is close to the target value 0.95. However the average length is 1811.791 which is a fairly large value. An examination of the conditional distributions of the amount of claims per insured shows that these distributions are often skewed to the right and having a large variation in their variances. Thus it seems plausible to determine the car insurance rate using a suitably chosen quantile for each category of customers.

The layout of the paper is as follows. In Section 2, the method based on multivariate power-normal distribution is outlined. Section 3 presents an analysis of the Swedish third party motor insurance data for 1977. Section 4 concludes the paper.

2. Procedure based on multivariate power-normal distribution

To introduce the multivariate power-normal distribution, we may begin with the following power transformation introduced in Yeo and Johnson [9]:

$$\tilde{\varepsilon} = \psi(\lambda^+, \lambda^-, z) = \begin{cases} \left[(z+1)^{\lambda^+} - 1\right]/\lambda^+, & (z \geq 0, \lambda^+ \neq 0) \\ \log(z+1), & (z \geq 0, \lambda^+ = 0) \\ -\left[(-z+1)^{\lambda^-} - 1\right]/\lambda^-, & (z < 0, \lambda^- \neq 0) \\ -\log(-z+1), & (z < 0, \lambda^- = 0) \end{cases} \quad (2.1)$$

If z in Equation (2.1) has the standard normal distribution, then $\tilde{\varepsilon}$ is said to have a power-normal distribution.

Let $\boldsymbol{y}$ be a column vector consisting of k correlated random variables. The vector $\boldsymbol{y}$ is said to have a k-dimensional power-normal distribution with parameters $\boldsymbol{\mu}, \mathbf{H}, \lambda_i^+, \lambda_i^-, \sigma_i, 1 \leq i \leq k$ if

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{H}\boldsymbol{\varepsilon} \tag{2.2}$$

where $\boldsymbol{\mu} = E(\boldsymbol{y})$, $\mathbf{H}$ is an orthogonal matrix, and $\boldsymbol{\varepsilon}$ is a column vector comprising the uncorrelated variables $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ of which the i^{th} variable ε_i is given by $\varepsilon_i = \sigma_i[\tilde{\varepsilon}_i - E(\tilde{\varepsilon}_i)]/\{\text{var}(\tilde{\varepsilon}_i)\}^{\frac{1}{2}}$ with $\sigma_i > 0$ representing a constant, and $\tilde{\varepsilon}_i$ denoting a random variable having the power-normal distribution with parameters λ_i^+ and λ_i^-.

We may use the procedure given in Pooi [7] to fit a multivariate power-normal distribution to the observed values of $\boldsymbol{y}$. From the multivariate power-normal distribution, we may use the method in Pooi [7] to find the conditional probability density function (pdf) of y_k given the value of $y_1, y_2, \ldots, y_{k-1}$.

However these are cases in which some of the variables $y_1, y_2, \ldots, y_{k-1}$ may originally be discrete. Consider, for example, the case when one of the variables, y_i say, is originally binary with two possible values given by 0 and 1. Let $y_i^{(0)}$ be the original binary y_i, and $y_i^{(a)}$ the i^{th} component of $\boldsymbol{y}$ given by Equation (2.2). The event that "$y_i^{(0)} = 0$" may then be considered as equivalent to the event that "$y_i^{(a)} \in E_i^{(0)} = (-\infty, 0.5]$", while the event that "$y_i^{(0)} = 1$" may be taken to be equivalent to the event that "$y_i^{(a)} \in E_i^{(1)} = (0.5, \infty)$".

When the original y_i is continuous, we may choose a small positive value δ_i and use the event that "$y_i \in E_i = (y_i^* - \delta_i, y_i^* + \delta_i]$" to describe approximately the situation when $y_i = y_i^*$.

Thus when some of the variables $y_1, y_2, \ldots, y_{k-1}$ are discrete, we may need to find the conditional distribution of y_k when y_i is conditional to lie in a certain subset E_i for $1 \leq i \leq k-1$. A way to find this conditioned distribution is by means of simulation. Initially we generate a large number of the values of $\boldsymbol{y}$ using Equation (2.2).

Suppose $y_k^{[j]}$ is the value of y_k when the generated $\boldsymbol{y}$ satisfies "$y_i \in E_i$ for $1 \leq i \leq k-1$" for the j^{th} time. Let N be the total number of such $y_k^{[j]}$. We next fit the values of $y_k^{[1]}, y_k^{[2]}, \ldots, y_k^{[N]}$ by a power-normal distribution. The fitted power-normal distribution may now be taken to be an approximation for the conditional distribution of y_k when $y_i \in E_i$ for $1 \leq i \leq k-1$.

From the conditional pdf of y_k, we may obtain the $100(\alpha/2)\%$ point, L_α, and $100(1-\alpha/2)\%$ point, U_α, of the conditional pdf to form a nominally

$100(1-\alpha)\%$ in-sample prediction interval $[L_\alpha, U_\alpha]$ for the observed value of y_k. The important characteristics of the prediction interval $[L_\alpha, U_\alpha]$ are

(a) its coverage probability P_c which is defined as the probability that $[L_\alpha, U_\alpha]$ will cover the observed y_k, and
(b) its expected length E_l which is defined as the expected value of the length $U_\alpha - L_\alpha$.

3. Swedish third party motor insurance

Consider the Swedish Third Party Motor Insurance data for 1977 ([1] and [3]). From these data we can obtain the claims amount per insured (C) when the kilometers travelled per year (K), geographical zone (Z), no claims bonus (B) and make of the car (M) are given. Variables K and B are quantitative and having 5 and 7 levels respectively. On the other hand, Z and M are respectively 7-level and 9-level qualitative variables. A qualitative variable V of n_v levels may be coded by using a vector of $n_v - 1$ binary variables. For example, for $1 \le i \le n_v - 1$, the i^{th} category of V may be coded as $(0,0,\dots,1,0,\dots,0)$ in which the value 1 appears as the i^{th} entry, while the ${n_v}^{th}$ category may be coded as a vector of zeros. Thus variables Z and M may be coded by using the vectors $(Z_1, Z_2, \dots, Z_6)$ and $(M_1, M_2, \dots, M_8)$ of binary variables. From the Swedish data for 1977, we can obtain a total of 2183 observed values of the vector $\boldsymbol{y} = (K, Z_1, Z_2, \dots, Z_6, B, M_1, M_2, \dots, M_8, C)$.

After fitting a 17-dimensional power-normal distribution to the observed values of $\boldsymbol{y}$, we may next find a conditional distribution for C when the value of $\boldsymbol{X}^* = (K, Z_1, Z_2, \dots, Z_6, B, M_1, M_2, \dots, M_8)$ is given. The mean of the conditional distribution will provide an estimate of the claims amount per insured when the value of $\boldsymbol{X}^*$ is given. To examine the performance of the estimate, we may compare

$A_{ij}^{(o)} =$ the average value of the observed claims amount per insured at the j^{th} level of variable i, obtained by averaging over all the levels of the other variables, with

$A_{ij}^{(e)} =$ the average value of the mean of the conditional distribution at the j^{th} level of variable i, obtained by averaging over all the levels of the other variables.

Figure i displays the values of $A_{ij}^{(o)}$ and $A_{ij}^{(e)}$ for $1 \le i \le 4$. Figure 1 shows that $A_{1j}^{(o)}$ is fairly close to $A_{1j}^{(e)}$ for $1 \le j \le 5$. Figures 2-4 show that there are some differences between the values $A_{ij}^{(o)}$ and $A_{ij}^{(e)}$ for some i and j.

Nevertheless for $2 \leq i \leq 4$, the trend exhibited by the values of $A_{ij}^{(o)}$ when j varies is still fairly similar to that shown by the values of $A_{ij}^{(e)}$ when j varies.

A nominally $100(1-\alpha)\%$ prediction interval $[L_\alpha, U_\alpha]$ may also be obtained for each given value of $\boldsymbol{X}^*$. When α is chosen to be 0.05, the estimated coverage probability of the prediction interval for C is found to be equal to 0.937 which is close to the target value 0.95. On the other hand, the average length of the prediction interval for C is found to be equal to 1811.791.

The lower and upper limits of the first 100 prediction intervals are shown in Figure 5. The figure shows that the upper limits of the prediction intervals can be quite far from the mean of the conditional distribution. These upper limits also exhibit a fairly large variation. Thus it seems plausible to determine the car insurance rate using a q-quantile for some suitably chosen $q > 0.5$ for each category of customers.

4. Conclusion

The proposed method based on multivariate power-normal distribution for determining the car insurance rates is quite promising as it produces prediction intervals which have good coverage probability, and are capable of determining an individual rate for each category of customers. The proposed method is also fairly versatile as it can deal with the situation when the explanatory variables are a mixture of quantitative and qualitative variables.

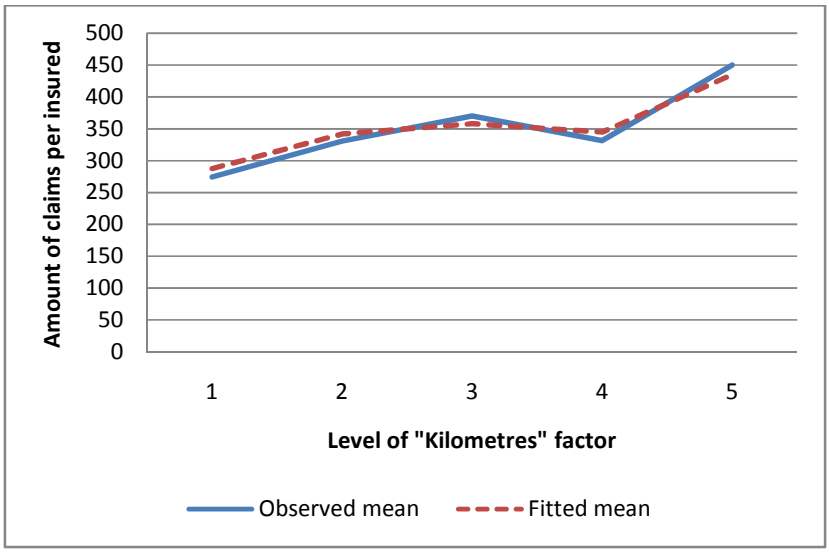


Fig. 1. Mean of each level of the "Kilometres" factor obtained by averaging over all the levels of the other factors

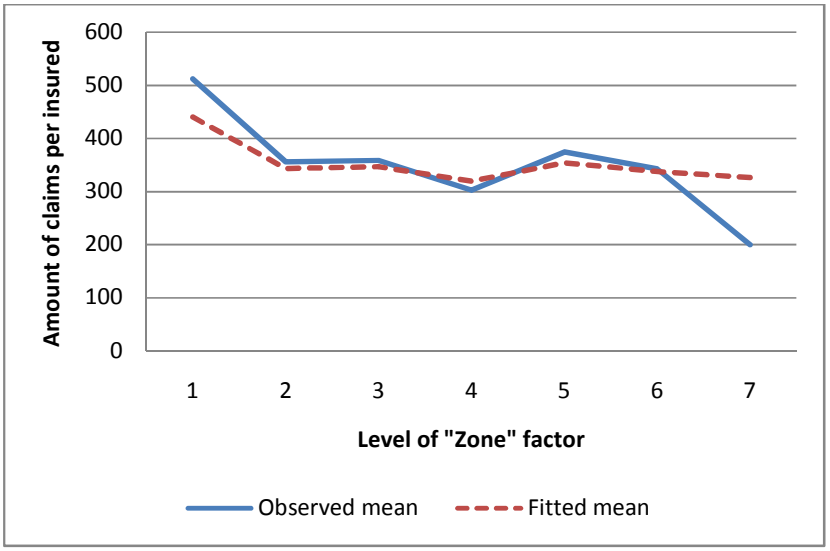


Fig. 2. Mean of each level of the "Zone" factor obtained by averaging over all the levels of the other factors

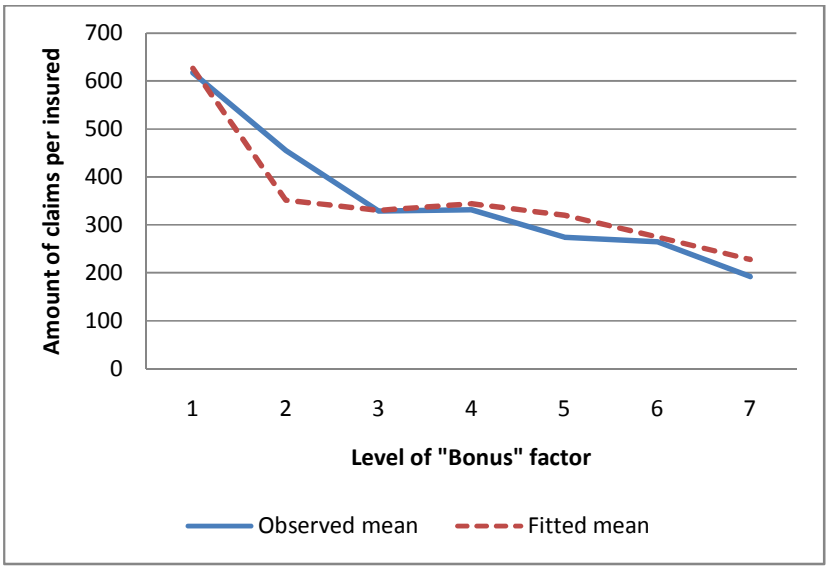


Fig. 3. Mean of each level of the "Bonus" factor obtained by averaging over all the levels of the other factors

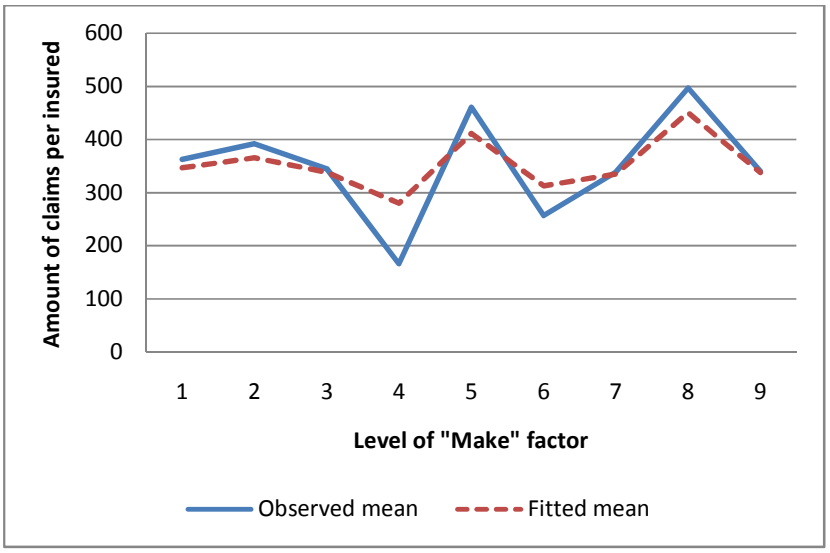


Fig. 4. Mean of each level of the "Make" factor obtained by averaging over all the levels of the other factors

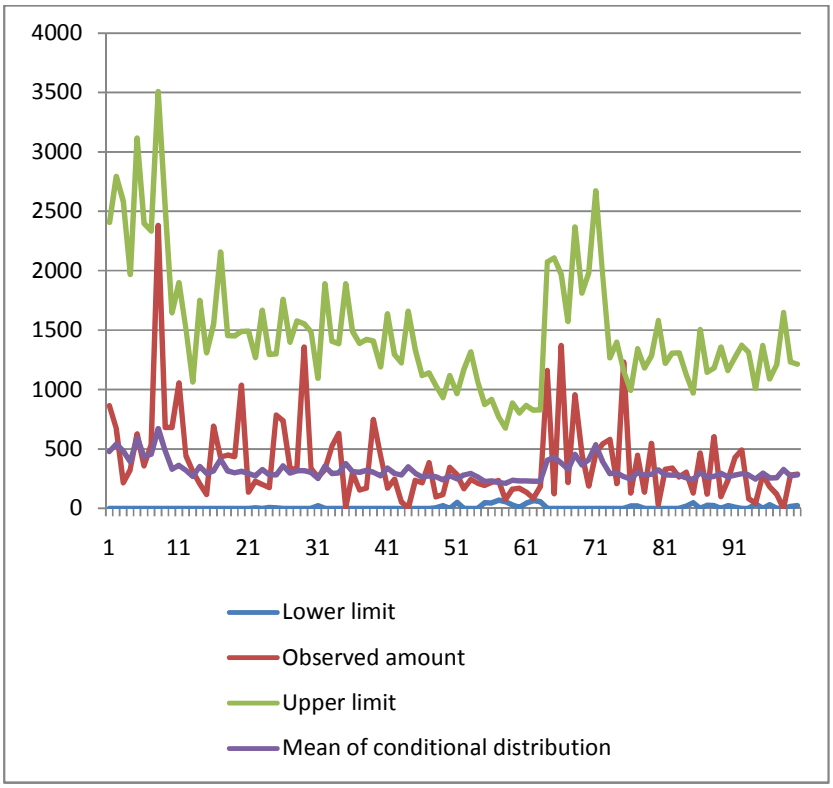


Fig. 5. Lower and upper limits of first 100 prediction intervals for the amount of claims per insured ($\alpha = 0.05$)

References

1. Andrews, D. and Herzberg, A.: Data: a collection of problems from many fields for the student and research worker. Spring Verlag. (1985)
2. Christmann, A.: Empirical risk minimization for car insurance data. (2006)
3. Hallin, M., Ingerbleek, J.F.: The Swedish automobile portfolio in 1977. A statistical study. Scandinavian Actuarial Journal. 49-64 (1983)
4. Jørgensen, B. and De Souza, M.C.P.: Fitting Tweedie's compound Poisson model to insurance claims data. Scandinavian Actuarial Journal. 69-93 (1994)
5. Loh, W.: Regression by parts: Fitting visually interpretable models with GUIDE. Handbook of Data Visualization. 447-469 (2008)
6. Marin-Galiano, M. and Chriatmann, A.: Insurance: an R-program to model insurance data. Technical Report 49, Department of Statistics, University of Dortmund, Germany (2004)

7. Pooi, A.H.: A model for time series analysis. Applied Mathematical Sciences. 6(115), 5735-5748 (2012)
8. Smyth, G.Y. and Jørgensen, B.: Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. Astin Bulletin. 32, 143-157 (2002)
9. Yeo, I.K. and Johnson, R.A.: A new family of power transformation to improve normality or symmetry. Biometrika. 87, 954-959 (2000)

Spreading Dynamic Model of a Contagious Disease in Heterogenic Population of Living Beings Using Multi Group Model Approach

Basuki Widodo[1] Nur Asiyah[2], Suhud Wahyudi[3], M. Setijo Winarko[4]
[1,2,3,4]Laboratory of Analysis and Algebra
Mathematics Department of ITS Surabaya-Indonesia
b_widodo@ matematika.its.ac.id

***Abstract*:** A heterogeneous ecosystem can be divided into many groups of homogenous. Therefore, if there is spreading of disease inside the ecosystem then the occurring of cross relation infection between those groups are able to be analyzed.

A mathematical model of a spreading contagious disease (epidemic) in a population of living beings, which are grouped as many of homogenous one, will be analyzed for the local and global stability from the system critical point with a multi group model approach. This paper provides a case of two groups where its infection rate is nonlinear. The method used here is to utilize Routh-Hurwitz criteria to analyze its local stability and the construction of Liapunov function for its global stability by implementing vector graph. By using graph theory approach, the system can be drawn as a network where each vertex represents a homogenous group and an edge (j, i), which is present if and only if the disease can be contagious from group i to group j.

After knowing the system critical point stability type of this nonlinear infection rate multi group epidemic model, then the number of reproduction R_O can be obtained to understand whether this epidemic disease will be extinguished later or becoming endemic instead.

***Keywords*:** Stability Analysis, Liapunov Function, Nonlinear infection rate, Network

1. Introduction

Real time problem in this life can be accommodated by using certain assumptions to build a mathematical model which help us to easily find solutions of that problem analytically or numerically. Coupled system from nonlinear differential function on a network has been used to model many things such as: inspecting coupled system of nonlinear oscillator, understanding the spreading of contagious disease, and analyzing the stability and complexity of coupled system in a complex ecosystem model [4].

Researches have been conducted to analyse the stability model of the spreading of contagious disease by neglecting the heterogeneity of a population [5, 6, 8], where this heterogeneity can be caused by many factors. A group of

individuals can be divided into groups of homogenous according to their different relational pattern on to something such as: group classification based on their age on the spreading of measles or goitre disease, or group classification based on their sex relation on a disease which the spreading media is through sex like HIV/AIDS [4].

Due to possibility of different infection rate within each group, therefore a network concept on the spreading of contagious disease is needed. By using approach from graph theory, the system can be drawn as a network where each vertex shows a homogenous group and an edge (j, i) will be present if and only if the disease is infectious from group i to group j. Some vertexes will be connected by a vector edge indicating connection between vertexes in the system. In this model, a vertex can present as an oscillator, a big ecology of community or a path, and can also present as homogenous groups for a common contagious disease, while the interaction between vertexes can be a physical connection between those oscillator, a spread between small groups in the path, or even a cross relation infection between homogenous groups within a heterogeneous group [4].

In this paper, local and global stability of the epidemic model of two groups with nonlinear infection rate will be analysed. A Routh-Hurwitz criterion is utilized for the local stability analysis, while graph theory is implemented for the global stability.

2. Epidemic Model of Two Groups and the Solution Area

Compartment is a flow that describes the spread of disease from individuals. There are phases in a compartment, they are:

S : *Susceptible*; healthy individual but is not immune to disease.

E : *Exposed;* infected individual but hasn't shown a symptom (incubation area).

I : *Infective*; Infected individual and also able to infect others.

R : *Removed*; Immune individual after being infected.

Epidemic model that consist of two groups are differential equation of epidemic model system with $i, j = 1,2$or can also be called as epidemic 2-Group. If given the equation of the infection rate $f_{ij}(S_i, I_j) = I_j^{p_j} S_i^{q_i}$, then system of differential equations become :

$$\dot{S}_1 = \Lambda_1 - d_1^S S_1 - \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} \qquad (1)$$

$$\dot{E}_1 = \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} - (d_1^E + \epsilon_1) E_1 \quad (2)$$

$$\dot{I}_1 = \epsilon_1 E_1 - (d_1^I + \gamma_1) I_1 \quad (3)$$

$$\dot{S}_2 = \Lambda_2 - d_2^S S_2 - \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} \quad (4)$$

$$\dot{E}_2 = \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} - (d_2^E + \epsilon_2) E_2 \quad (5)$$

$$\dot{I}_2 = \epsilon_2 E_2 - (d_2^I + \gamma_2) I_2 \quad (6)$$

The figure of compartment diagram from the model (2)-(6) is depicted as follow:

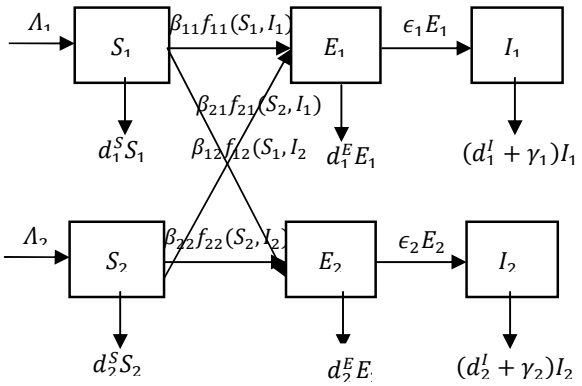


Figure 1: Compartment Diagram of Epidemic Two Group Model with Nonlinear Infection Rate

Assuming that $\epsilon_1, \epsilon_2 > 0$, and $d_i^* > 0$ where $d_1^* = \min\{d_1^S,\ d_1^E, d_1^I + \gamma_1\}$, $d_2^* = \{d_2^S,\ d_2^E, d_2^I + \gamma_2\}$, $f_{ij}(S_i, I_j) = I_j^{p_j} S_i^{q_i} > 0 \text{for} S_i > 0, I_j > 0,$ and p_j, q_i is a positive constant. By taking summation of equation 1-3 will produce

$$\dot{S}_1 + \dot{E}_1 + \dot{I}_1 = \Lambda_1 - d_1^S S_1 - \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} I_i + \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} - (d_1^E + \epsilon_1) E_1$$
$$+ \epsilon_1 E_1 - (d_1^I + \gamma_1) I_1$$
$$= \Lambda_1 - d_1^S S_1 - d_1^E E_1 - (d_1^I + \gamma_1) I_1$$
$$\leq \Lambda_1 - d_1^* (S_1 + E_1 + I_1)$$

Therefore, $\lim \sup_{t\to\infty}(S_1 + E_1 + I_1) \leq \frac{\Lambda_1}{d_1^*}$, and the equations (4) – (6) will give

$$\begin{aligned}\dot{S}_2 + \dot{E}_2 + \dot{I}_2 &= \Lambda_2 - d_2^S S_2 - \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} + \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} - (d_2^E + \epsilon_2)E_2 \\ &\quad + \epsilon_2 E_2 - (d_2^I + \gamma_2)I_2 \\ &= \Lambda_2 - d_2^S S_2 - d_2^E E_2 - (d_2^I + \gamma_2)I_2 \\ &\leq \Lambda_2 - d_2^*(S_2 + E_2 + I_2)\end{aligned}$$

and $\lim \sup_{t\to\infty}(S_2 + E_2 + I_2) \leq \frac{\Lambda_2}{d_2^*}$

Thus the solutions area of the epidemic two group model can be written as

$$\Gamma = \left\{(S_1, E_1, I_1, S_2, E_2, I_2)) \in \mathbb{R}_+^{3n} | S_i \leq \frac{\Lambda_i}{d_i^S}, (S_i + E_i + I_i) \leq \frac{\Lambda_i}{d_i^*}, i = 1,2\right\} \quad (7)$$

3. Model Equilibrium Point

Equilibrium point is a point where it's characteristic is invariant with time. Then the equilibrium points are taken from $\frac{dS_i}{dt} = 0, \frac{dE_i}{dt} = 0, \frac{dI_i}{dt} = 0$ and the equations 1 – 6 become:

$$\Lambda_1 - d_1^S S_1 - \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} = 0 \quad (8)$$

$$\sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} - (d_1^E + \epsilon_1)E_1 = 0 \quad (9)$$

$$\epsilon_1 E_1 - (d_1^I + \gamma_1)I_1 = 0 \quad (10)$$

$$\Lambda_2 - d_2^S S_2 - \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} = 0 \quad (11)$$

$$\sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} - (d_2^E + \epsilon_2)E_2 = 0 \quad (12)$$

$$\epsilon_2 E_2 - (d_2^I + \gamma_2)I_2 = 0 \quad (13)$$

There are four equilibrium points achieved. They are equilibrium point of disease free, namely endemic point, a point where first group are disease free while second group are endemic, and a point where first group are endemic while second group are disease free.

3.1 Disease Free Equilibrium Point

When, $I_1, I_2 = 0$, it will have free disease equilibrium point P_0, where all the individuals are *Susceptible* or will have no any spread of disease in both of groups. By substituting $I_1^0, I_2^0 = 0$ to equation (10) and (13) will give:

$$\epsilon_1 E_1 - (d_1^I + \gamma_1) I_1 = 0$$
$$\epsilon_1 E_1^0 - 0 = 0$$
$$E_1^0 = 0$$

And

$$\epsilon_2 E_2 - (d_2^I + \gamma_2) I_2 = 0$$
$$\epsilon_2 E_2^0 - 0 = 0$$
$$E_2^0 = 0$$

This result substituted to equation (8) and (11) resulting

$$\Lambda_1 - d_1^S S_1 - \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} = 0$$
$$\Lambda_1 - d_1^S S_1^0 - 0 = 0$$
$$S_1^0 = \frac{\Lambda_1}{d_1^S}$$

And it will give:

$$\Lambda_2 - d_2^S S_2 - \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} = 0$$
$$\Lambda_2 - d_2^S S_2^0 - 0 = 0$$
$$S_2^0 = \frac{\Lambda_2}{d_2^S}$$

Thus the disease free equilibrium point for this epidemic two group model can be written as $P_0 = (S_1^0, 0,0, S_2^0, 0,0)$ where $S_1^0 = \frac{\Lambda_1}{d_1^S}, S_2^0 = \frac{\Lambda_2}{d_2^S}$.

3.2 Endemic Equilibrium Point

Endemic equilibrium point is a condition where the disease will always be inside the population in both of group. Endemic equilibrium point $P^* = (S_1^*, E_1^*, I_1^*, S_2^*, E_2^*, I_2^*)$ depend on the infected populations from contagious disease where $I_1, I_2 \geq 0$. It is given from taking $\frac{dS_1}{dt} = 0, \frac{dE_1}{dt} = 0, \frac{dI_1}{dt} = 0, \frac{dS_2}{dt} = 0, \frac{dE_2}{dt} = 0, \frac{dI_2}{dt} = 0$.Thus the endemic equilibrium point from the this two group model is $P^* = (S_1^*, E_1^*, I_1^*, S_2^*, E_2^*, I_2^*)$ where S_1^*, E_1^*, I_1^* satisfying equations below:

$$\frac{dS_1}{dt} = 0$$

$$\Lambda_1 - d_1^S S_1^* - \sum_{j=1}^{n} \beta_{1j} I_j^{*p_j} S_1^{*q_1} = 0$$

$$\Lambda_1 = d_1^S S_1^* + \sum_{j=1}^{n} \beta_{1j} I_j^{*p_j} S_1^{*q_1} \quad (14)$$

$$\frac{dE_1}{dt} = 0$$

$$\sum_{j=1}^{n} \beta_{1j} I_j^{*p_j} S_1^{*q_1} - (d_1^E + \epsilon_1) E_1^* = 0$$

$$(d_1^E + \epsilon_1) E_1^* = \sum_{j=1}^{n} \beta_{1j} I_j^{*p_j} S_1^{*q_1} \quad (15)$$

$$\frac{dI_1}{dt} = 0$$

$$\epsilon_1 E_1^* - (d_1^I + \gamma_1) I_1^* = 0$$

$$\epsilon_1 E_1^* = (d_1^I + \gamma_1) I_1^* \quad (16)$$

$$\frac{dS_2}{dt} = 0$$

$$\Lambda_2 - d_2^S S_2^* - \sum_{j=1}^{n} \beta_{2j} I_j^{*p_j} S_2^{*q_2} = 0$$

$$\Lambda_2 = d_2^S S_2^* + \sum_{j=1}^{n} \beta_{2j} I_j^{*p_j} S_1^{*q_1} \quad (17)$$

$$\frac{dE_2}{dt} = 0$$

$$\sum_{j=1}^{n} \beta_{2j} I_j^{*p_j} S_1^{*q_1} - (d_2^E + \epsilon_2) E_2^* = 0 \quad (18)$$

$$(d_2^E + \epsilon_2) E_2^* = \sum_{j=1}^{n} \beta_{2j} I_j^{*p_j} S_1^{*q_1}$$

$$\frac{dI_2}{dt} = 0$$

$$\epsilon_2 E_2^* - (d_2^I + \gamma_2) I_2^* = 0$$

$$\epsilon_2 E_2^* = (d_2^I + \gamma_2) I_2^* \quad (19)$$

3.3 Equilibrium Point Where First Group is Disease Free While Second Group is Endemic

On this equilibrium point, the first group are free disease. While the second group the disease are always present. Equilibrium point of $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$ depends on the infected population in the first group $I_1 = 0$ and the infected population from contagious disease in which $I_2 \geq 0$. This point will be given by substituting $I_1^0 = 0$ to (10).

$\epsilon_1 E_1 - (d_1^I + \gamma_1) I_1 = 0$

$\epsilon_1 E_1^0 - 0 = 0$

$E_1^0 = 0$

This result is substituted to equation (8) and (8) giving

$$\Lambda_1 - d_1^S S_1 - \sum_{j=1}^{2} \beta_{1j} I_j^{p_j} S_1^{q_1} = 0$$

$$\Lambda_1 - d_1^S S_1^0 - 0 = 0$$

$$S_1^0 = \frac{\Lambda_1}{d_1^S}$$

Thus the equilibrium point for two group epidemic model is $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$ with $S_1^0 = \frac{\Lambda_1}{d_1^S}$ and S_2^*, E_2^*, I_2^* should satisfy equations (17) – (19).

3.4 EquilibriumPoint Where First Group is Endemic While Second Group is Disease Free

On this equilibrium point disease is always present in first group while on the second group all individuals are *susceptible* or in a case where the disease spread are not happening. Equilibrium point of $P_2^* = (S_1^*, E_1^*, I_1^*, S_2^0, 0,0)$ depends on the infected population in first group $I_1 \geq 0$ and the infected population in second group is $I_2 = 0$. To find this equilibrium point, $I_2^0 = 0$ is substituted to equation (13)

$\epsilon_2 E_2 - (d_2^I + \gamma_2) I_2 = 0$

$\epsilon_2 E_2^0 - 0 = 0$

$E_2^0 = 0$

The result is substituted once again into equation (12) and (11) giving

$$\Lambda_2 - d_2^S S_2 - \sum_{j=1}^{2} \beta_{2j} I_j^{p_j} S_2^{q_2} = 0$$

$$\Lambda_2 - d_2^S S_2^0 - 0 = 0$$

$$S_2^0 = \frac{\Lambda_2}{d_2^S}$$

Thus the equilibrium point of disease free for this two model epidemic group is $P_2^* = (S_1^*, E_1^*, I_1^*, S_2^0, 0,0)$ where S_1^*, E_1^*, I_1^* satisfy the equations (14) – (16) and $S_2^0 = \frac{\Lambda_2}{d_2^S}$.

4. Basic Reproduction Number

From previous examination, it is known that

$$R_0 = \rho\left(\left[\frac{\beta_{ij}\epsilon_i C_{ij}(S_i^0)}{(d_i^E + \epsilon_i)(d_i^I + \gamma_i)}\right]\right)$$

Assuming that if $R_0 = \rho(M_0)$ with

$$M_0 = M(S_1^0, S_2^0) = \left(\frac{\beta_{ij}\epsilon_i C_{ij}(S_i^0)}{(d_i^E + \epsilon_i)(d_i^I + \gamma_i)}\right)_{1\le i,j\le 2}$$

because of $f_{ij}(S_i, I_j) = I_j^{p_j} S_i^{q_i}$ then $C_{ij}(S_i^0) = p_j I_j^{p_j - 1} {S_i^0}^{q_i}$

Therefore

$$R_0 = \rho\begin{pmatrix} \frac{\beta_{11}\epsilon_1 p_1 I_1^{p_1-1} {S_1^0}^{q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} & \frac{\beta_{12}\epsilon_2 p_2 I_2^{p_2-1} {S_1^0}^{q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} \\ \frac{\beta_{21}\epsilon_2 p_1 I_1^{p_1-1} {S_2^0}^{q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)} & \frac{\beta_{22}\epsilon_2 p_2 I_2^{p_2-1} {S_2^0}^{q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)} \end{pmatrix}$$

To find R_0 it should be found the Eigen value of the matrix M_0.

$\det|M_0 - \lambda I| = 0$

$$\begin{vmatrix} \frac{\beta_{11}\epsilon_1 p_1 I_1^{p_1-1} {S_1^0}^{q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} - \lambda & \frac{\beta_{12}\epsilon_2 p_2 I_2^{p_2-1} {S_1^0}^{q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} \\ \frac{\beta_{21}\epsilon_2 p_1 I_1^{p_1-1} {S_2^0}^{q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)} & \frac{\beta_{22}\epsilon_2 p_2 I_2^{p_2-1} {S_2^0}^{q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)} - \lambda \end{vmatrix} = 0$$

And the polynomial equation can be converted into: $\lambda^2 - (p+q)\lambda + (pq - rs) = 0$ where:

$$p = \frac{\beta_{11}\epsilon_1 p_1 I_1^{p_1-1} {S_1^0}^{q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)}$$

$$q = \frac{\beta_{22}\epsilon_2 p_2 I_2^{p_2-1} {S_2^0}^{q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)}$$

$$r = \frac{\beta_{12}\epsilon_2 p_2 I_2^{p_2-1} S_1^{0\,q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)}$$

$$s = \frac{\beta_{21}\epsilon_2 p_1 I_1^{p_1-1} S_2^{0\,q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)}$$

Because $R_0 = \rho(M_0) = \max_i|\lambda_i|$, where $\rho$is a spectral radius, by using ABC equations resulting:

$$\lambda = \frac{(p+q) + \sqrt{(p+q)^2 - 4(pq - rs)}}{2}$$

$$\lambda = \frac{1}{2}\left[(p+q) + \sqrt{(p-q)^2 + 4rs}\right]$$

By substituting back variables p, q, r, s then it will give

$$R_0 = \frac{1}{2}\left[\left(\frac{\beta_{11}\epsilon_1 p_1 I_1^{p_1-1} S_1^{0\,q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} + \frac{\beta_{22}\epsilon_2 p_2 I_2^{p_2-1} S_2^{0\,q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)}\right) + \sqrt{\left(\frac{\beta_{11}\epsilon_1 p_1 I_1^{p_1-1} S_1^{0\,q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)} - \frac{\beta_{22}\epsilon_2 p_2 I_2^{p_2-1} S_2^{0\,q_2}}{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)}\right)^2 + 4\frac{\beta_{12}\epsilon_2 p_2 I_2^{p_2-1} S_1^{0\,q_1} \beta_{21}\epsilon_2 p_1 I_1^{p_1-1} S_2^{0\,q_2}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)(d_2^E + \epsilon_2)(d_2^I + \gamma_2)}}\right]$$

This R_0 represents basic reproduction number for the two group epidemic model with nonlinear infection rate.

5. Local Equilibrium Point Stability

5.1 Local Equilibrium Point Stability of Free Disease Condition

On the free disease equilibrium point $P_0 = (S_1^0, 0, 0, S_2^0, 0, 0)$with$S_1^0 = \frac{\Lambda_1}{d_1^S}, S_2^0 = \frac{\Lambda_2}{d_2^S}$, it is known that $\beta_{1j} = \beta_{2j} = \epsilon_1 = \epsilon_2 = \gamma_1 = \gamma_2 = 0$, so then its Jacobian matrix:

$$J(P_0) = \begin{bmatrix} -d_1^S & 0 & 0 & 0 & 0 & 0 \\ 0 & -d_1^E & 0 & 0 & 0 & 0 \\ 0 & 0 & -d_1^I & 0 & 0 & 0 \\ 0 & 0 & 0 & -d_2^S & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_2^E & 0 \\ 0 & 0 & 0 & 0 & 0 & -d_2^I \end{bmatrix}$$

The Eigen value can be taken from $\det|J(P_0) - \lambda I| = 0$, therefore resulting:

$\lambda_1 = -d_1^S$ $\lambda_2 = -d_1^E$
$\lambda_3 = -d_1^I$ $\lambda_4 = -d_2^S$
$\lambda_5 = d_2^E$ $\lambda_6 = -d_2^I$

Due to Eigen value of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ are negative on the real components, therefore the equilibrium point $P_0 = \left(\frac{\Lambda_1}{d_1^S}, 0,0, \frac{\Lambda_2}{d_2^S}, 0,0\right)$ is asymptotic stable.

5.2 Local Equilibrium Point of Both Groups Are Endemic

On the equilibrium point of $P^* = (S_1^*, E_1^*, I_1^*, S_2^*, E_2^*, I_2^*)$, both groups are on endemic condition. So that its Jacobian matrix is:

$$J = \begin{bmatrix} -a-b & 0 & -c & 0 & 0 & -d \\ b & -e & c & 0 & 0 & d \\ 0 & f & -g & 0 & 0 & 0 \\ 0 & 0 & -h & -i-j & 0 & -k \\ 0 & 0 & h & j & -l & k \\ 0 & 0 & 0 & 0 & m & -n \end{bmatrix} \quad (20)$$

With :

$a = d_1^S$ $h = \beta_{21} p_1 I_1^{*p_1-1} S_2^{*q_2}$
$b = \sum_{j=1}^{2} \beta_{1j} I_j^{*p_j} q_1 S_1^{q_1-1}$ $i = d_2^S$
$j = \sum_{j=1}^{2} \beta_{2j} I_j^{*p_j} q_2 S_2^{*q_2-1}$
$c = \beta_{11} p_1 I_1^{*p_1-1} S_1^{*q_1}$
$d = \beta_{12} p_2 I_2^{p_2-1} S_1^{*q_1}$ $k = \beta_{22} p_2 I_2^{*p_2-1} S_2^{*q_2}$
$e = (d_1^E + \epsilon_1)$ $l = (d_2^E + \epsilon_2)$
$f = \epsilon_1$ $m = \epsilon_2$
$g = (d_1^I + \gamma_1)$ $n = (d_2^I + \gamma_2)$

Then it should be found $\det|J - \lambda I| = 0$

Resulting characteristic equation with model of

$$a_0\lambda^6 + a_1\lambda^5 + a_2\lambda^4 + a_3\lambda^3 + a_4\lambda^2 + a_5\lambda + a_6 = 0$$

With a_0=1

$$a_1 = a + b + e + g + i + j + l + n$$

$$a_2 = ae + be + ag + bg - cf + ai + aj + bi + bj + eg + al + bl + ei + an + ej + bn + gi + el + gj + en + gl + gh + il + jl + in + jn - km + ln$$

$$a_3 = -acf + aeg + beg + aei + aej + bei + agi + bej + ael + agj + bgi - cfi + bel + bgj - cfj + aen + agl + ben + bgl - cfl + egi + agn + ail + egj + ajl + bgn + bil - cfn + ain + bjl + egl + ajn - akm + bin + bjn - bkm + egn + eil + aln + ejl + bln + ein + gil + ejn - ekm + gjl + gin + eln + gjn - gkm + gln - ikm + iln + jln$$

$$a_4 = beln + bgjn - bgkm - cfjn + cfkm + egil + agln - aikm + egjl + bgln - bikm - cfln + egin + ailn + egjn - egkm + ajln + biln + bjln + egln - eikm + eiln - gikm + ejln + giln + gjln - acfi - acfj - acfl + aegi + aegj + begi - acfn + begj + aegl + begl + aegn + aeil + aejl + begn + beil + aein + agil + bejl + aejn - aekm + +agjl + bein + bgil - cfil + agin + bejn - bekm + bgjl - cfjl - dfhm + aeln + agjn - agkm + bgin - cfin$$

$$a_5 = -acfil - acfjl - adfhm - acfin - acfjn + acfkm + aegil + aegjl + begil - acfln + aegin + begjl + aegjn - aegkm + begin + begjn - begkm + aegln - aeikm + begln - beikm - dfhim + aeiln - agikm + aejln + beiln - bgikm + cfikm + agiln + begjln + agjln + bgiln - cfiln + bgjln - cfjln - egikm + egiln + egjln$$

$$a_6 = -begikm - aegikm - acfjln - acfiln - adfhim + begjln + begiln + aegjln + aegiln + acfikm$$

To know the stability of that equilibrium point, a parameter will be chosen to refer to the solution area as below:

$\Lambda_1 = 0.2$	$\Lambda_2 = 0.1$	$S_1^* = 3$	$\beta_{11} = 0.005$
$d_1^S = 0.003$	$d_2^S = 0.002$	$E_1^* = 8$	$\beta_{12} = 0.002$
$d_1^E = 0.004$	$d_2^E = 0.002$	$I_1^* = 10$	$\beta_{21} = 0.002$
$d_1^I = 0.005$	$d_2^I = 0.002$		$\beta_{22} = 0.002$
$\epsilon_1 = 0.02$	$\epsilon_2 = 0.02$	$S_2^* = 3$	
$\gamma_1 = 0.011$	$\gamma_2 = 0.012$	$E_2^* = 4$	
		$I_2^* = 6$	

$p_1 = 1, q_1 = 1, p_2 = 1, q_2 = 1$

By inputting that value to characteristic equation, will result:

$a_0 = 1$

$a_1 = 0.175$

$a_2 = 0.0113$

$a_3 = 0.0003719$

$a_4 = 0.000006534$

$a_5 = 0.000000046932$

$a_6 = 0.0000000002461$

Then Routh-Hurwitz criterion stability is used to analyze the stability of the endemic equilibrium point.

λ^6	a_0	a_2	a_4	a_6
λ^5	a_1	a_3	a_5	0
λ^4	$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}$	$b_2 = \frac{a_1 a_4 - a_0 a_5}{a_1}$	a_6	0
λ^3	$c_1 = \frac{b_1 a_3 - a_1 b_2}{b_1}$	$c_2 = \frac{b_1 a_5 - a_1 a_6}{b_1}$	0	0
λ^2	$d_1 = \frac{c_1 b_2 - b_1 c_2}{c_1}$	a_6	0	0
λ^1	$e_1 = \frac{d_1 c_2 - c_1 a_6}{d_1}$	0	0	0
λ^0	a_6	0	0	0

Results from the first row are:

$a_0 = 1$
$a_1 = 0.175$
$b_1 = 0.0092$
$c_1 = 0{,}00025454$
$d_1 = 0.0000047421$
$e_1 = 0.000000029032$
$a_6 = 0.0000000002461$

It is shown that values on the first column have the same sign, positive numbers, so then the stability of the equilibrium point $P^*$is local asymptotic stable.

5.3 Local Point Stability of the First Group is Free Disease and the Second Group is Endemic.

Equilibrium point $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$ with $S_1^0 = \frac{\Lambda_1}{d_1^S}$ and S_2^*, E_2^*, I_2^* satisfy equation (17) – (19). Due to equilibrium point P_1^* happened if the first group is free disease and the second is endemic, then $\beta_{11} = \beta_{12} = \beta_{21} = \epsilon_1 = \gamma_1 = 0$. It will be shown that

$$R_1 = \frac{1}{2}\left[\left(\frac{\beta_{22}\epsilon_2 p_2 I_2^{*p_2-1} S_2^{*q_2}}{(d_2^E+\epsilon_2)(d_2^I+\gamma_2)}\right) + \sqrt{\left(\frac{\beta_{22}\epsilon_2 p_2 I_2^{*p_2-1} S_2^{*q_2}}{(d_2^E+\epsilon_2)(d_2^I+\gamma_2)}\right)^2}\right]$$

$$R_1 = \frac{\beta_{22}\epsilon_2 p_2 I_2^{*p_2-1} S_2^{*q_2}}{(d_2^E+\epsilon_2)(d_2^I+\gamma_2)}$$

And the Jacobian matrix of P_1^*. Then it should be found the Eigen value of $J(P_1^*)$, and it's taken from

$$\det|J(P_1^*) - \lambda I| = 0 \qquad (21)$$

The result shows that $-d_1^S, -d_1^E, -(d_1^I+\gamma_1)$is the Eigen value of that Jacobian Matrix on the equilibrium point $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$.

Thus giving three next Eigen values as

$$-\lambda^3 - \Big((d_2^I+\gamma_2) + (d_2^E+\epsilon_2) + d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1}\Big)\lambda^2$$
$$-\Big((d_2^E+\epsilon_2)(d_2^I+\gamma_2) + d_2^S(d_2^I+\gamma_2) + (d_2^I+\gamma_2)\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1} + d_2^S(d_2^E+\epsilon_2) + (d_2^E+\epsilon_2)\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1} - \epsilon_2\beta_{22} p_2 I_2^{*p_2-1} S_2^{*q_2}\Big)\lambda - \Big(d_2^S(d_2^E+\epsilon_2)(d_2^I+\gamma_2) + (d_2^E+\epsilon_2)(d_2^I+\gamma_2)\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1} - d_2^S\epsilon_2\beta_{22} p_2 I_2^{*p_2-1} S_2^{*q_2}\Big) = 0$$

(22)

Equation (22) can be written in third polynomial order becoming:

$$a_0\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0$$

With the coefficient $a_0, a_1, a_2 a_3$as below

$$a_0 = -1$$
$$a_1 = -\Big((d_2^I+\gamma_2) + (d_2^E+\epsilon_2) + d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1}\Big)$$
$$a_2 = -\Big((d_2^I+\gamma_2)\Big((d_2^E+\epsilon_2) + d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1}\Big) + (d_2^E+\epsilon_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1}\big) - \epsilon_2\beta_{22} p_2 I_2^{*p_2-1} S_2^{*q_2}\Big)$$
$$a_3 = -\Big((d_2^E+\epsilon_2)(d_2^I+\gamma_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2-1}\big) - d_2^S\epsilon_2\beta_{22} p_2 I_2^{*p_2-1} S_2^{*q_2}\Big)$$

To calculate at its stability by using root characteristic (Eigen value of λ) from third polynomial order Routh-Hurwitz criterion is used.

λ^3	a_0	a_2
λ^2	a_1	a_3
λ^1	$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1}$	0
λ^0	a_3	0

A system can be called stable when the root of characteristic equation from a matrix has a real negative Eigen value if only the element from the first column $(a_0,\ a_1,\ b_1,\ a_3)$has common sign. From the coefficient above it is known that the value of

$$a_0 = -1 < 0$$

$$a_1 = -\left((d_2^I + \gamma_2) + (d_2^E + \epsilon_2) + d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}\right) < 0$$

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1} = \frac{a_1 a_2}{a_1} - \frac{a_0 a_3}{a_1} = a_2 - \frac{a_0 a_3}{a_1}$$

$$a_3 = -\Big((d_2^E + \epsilon_2)(d_2^I + \gamma_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}\big) - d_2^S \epsilon_2 \beta_{22} p_2 I_2^{*p_2 - 1} S_2^{*q_2}\Big)$$

It can be ensured that $a_0 < 0$ dan $a_1 < 0$, while for b_1 dan $a_3$can't still be ensured the value whether it is negative or positive. In order for endemic equilibrium point $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$ stable then the value of b_1 and $a_3$should be a negative. There it needs two conditions and they are:

$$a_3 < 0$$

$$\Leftrightarrow -\Big((d_2^E + \epsilon_2)(d_2^I + \gamma_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}\big) - d_2^S \epsilon_2 \beta_{22} p_2 I_2^{*p_2 - 1} S_2^{*q_2}\Big) < 0$$

$$\Leftrightarrow \left((d_2^E + \epsilon_2)(d_2^I + \gamma_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}\big) > d_2^S \epsilon_2 c\right)$$

$$\Leftrightarrow \frac{(d_2^E + \epsilon_2)(d_2^I + \gamma_2)\big(d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}\big)}{d_2^S (d_2^E + \epsilon_2)(d_2^I + \gamma_2)} > \frac{d_2^S \epsilon_2 \beta_{22} p_2 I_2^{*p_2 - 1} S_2^{*q_2}}{d_2^S (d_2^E + \epsilon_2)(d_2^I + \gamma_2)}$$

$$\Leftrightarrow \frac{d_2^S + \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}}{d_2^S} > R_1$$

$$\Leftrightarrow 1 + \frac{\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}}{d_2^S} > R_1$$

$$\Leftrightarrow \frac{\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}}{d_2^S} > R_1 - 1$$

Because $\frac{\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}}{d_2^S} > 0$resulting $R_1 > 1$

$$b_1 = \frac{a_1 a_2 - a_0 a_3}{a_1} < 0$$

a_1 has been known as negative value and $a_0 = -1$, so it will give

$a_1 a_2 + a_3 > 0$

$a_1 a_2 > -a_3$

Because $a_3 < 0$ then $a_1 a_2 > 0$, where it has been known that $a_1 < 0$, in order for $b_1 < 0$then $a_2 < 0$, giving

$$(d_2^E + \epsilon_2)(d_2^I + \gamma_2) + d_2^S(d_2^I + \gamma_2) + (d_2^I + \gamma_2)\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1} + d_2^S(d_2^E + \epsilon_2) + (d_2^E + \epsilon_2)\beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1} - \epsilon_2 \beta_{22} p_2 I_2^{*p_2 - 1} S_2^{*q_2} < 0$$

To ease the writings of this equations, it should be assumed that $a = (d_2^E + \epsilon_2)$, $b = (d_2^I + \gamma_2)$,

$c = \beta_{22} I_2^{*p_2} q_2 S_2^{*q_2 - 1}$, $R = \epsilon_2 \beta_{22} p_2 I_2^{*p_2 - 1} S_2^{*q_2}$

and $R_1 = \frac{R}{ab}$ and it become

$\Leftrightarrow (b(a + d_2^S + c) + a(d_2^S + c) - R) < 0$

$\Leftrightarrow R > ab + d_2^S b + bc + d_2^S a + ac$

$\Leftrightarrow R > ab + (b + a)d_2^S + (a + b)c$

$$\Leftrightarrow \frac{R}{ab} > \frac{ab + (b + a)d_2^S + (a + b)c}{ab}$$

$$\Leftrightarrow R_1 > 1 + \frac{(b + a)d_2^S + (a + b)c}{ab}$$

Because $\frac{(b+a)d_2^S + (a+b)c}{ab} > 0$then $R_1 > 1$

Thus the equilibrium point of $P_1^* = (S_1^0, 0,0, S_2^*, E_2^*, I_2^*)$is locally stable for $R_1 > 1$.

5.4 Local Stability of First Group Endemic and Second Group Free Disease

The equilibrium point $P_2^* = (S_1^*, E_1^*, I_1^*, S_2^0, 0,0)$ happen if the first group is endemic and the second group is free disease, so it can be written $\beta_{12} = \beta_{22} = \beta_{21} = \epsilon_2 = 0$, dan $S_2^0 = \frac{\Lambda_2}{d_2^S}$. And it will give

$$R_2 = \frac{1}{2}\left[\left(\frac{\beta_{11}\epsilon_1 p_1 I_1^{*p_1 - 1} S_1^{*q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)}\right) + \sqrt{\left(\frac{\beta_{11}\epsilon_1 p_1 I_1^{*p_1 - 1} S_1^{*q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)}\right)^2}\right]$$

$$= \frac{\beta_{11}\epsilon_1 p_1 I_1^{*p_1 - 1} S_1^{*q_1}}{(d_1^E + \epsilon_1)(d_1^I + \gamma_1)}$$

And the Jacobian matrix of P_2^* is $J(P_2^*)$

Eigen value is taken from $\det|J(P_2^*) - \lambda I| = 0$, obtained

from the result it can be understood that $-d_2^S, -d_2^E, -(d_2^I - \gamma_2)$ are Eigen values of $J(P_2^*)$, and in order to obtain next three Eigen values

$$\Leftrightarrow \begin{vmatrix} -d_1^S - \beta_{11} I_1^{*p_1} q_1 S_1^{*q_1-1} - \lambda & 0 & -\beta_{11} p_1 I_1^{*p_1-1} S_1^{*q_1} \\ \beta_{11} I_1^{*p_1} q_1 S_1^{*q_1-1} & -(d_1^E + \epsilon_1) - \lambda & \beta_{11} p_1 I_1^{*p_1-1} S_1^{*q_1} \\ 0 & \epsilon_1 & -(d_2^I + \gamma_2) - \lambda \end{vmatrix} = 0$$

This Matrix is similar to (22), and it give the same result which equilibrium point of P_2^* stabil is locally stable for $R_2 > 1$.

6. Simulation

To interpret the analysis result of this two group epidemic model, a simulation is made by using Matlab software. And the analysis of simulation is done on four equilibrium point and they are free disease equilibrium point, where $R_0 \leq 1$, an endemic equilibrium point where $R_0 > 1$, and equilibrium point where both groups do not interact each another.

6.1 Free Disease Condition ($R_0 \leq 1$)

Parameters taken for simulation is based on the solution area of the model which has been explained on the previous section, so then it is chosen satisfying parameters as below:

First Group	Second Group	First Value	Infection Rate
$\Lambda_1 = 0.2$	$\Lambda_2 = 0.1$	$S_1(0) = 15$	$\beta_{11} = 0.001$
$d_1^S = 0.009$	$d_2^S = 0.007$	$E_1(0) = 1$	$\beta_{12} = 0.001$
$d_1^E = 0.02$	$d_2^E = 0.017$	$I_1(0) = 4$	$\beta_{21} = 0.001$
$d_1^I = 0.025$	$d_2^I = 0.02$	$S_2(0) = 10$	$\beta_{22} = 0.001$
$\epsilon_1 = 0.003$	$\epsilon_2 = 0.001$	$E_2(0) = 1$	
$\gamma_1 = 0.003$	$\gamma_2 = 0.002$	$I_2(0) = 3$	

And the coefficient $p_1 = 1, q_1 = 1, p_2 = 1, q_2 = 1$.

By using those parameters, a number $R_0 = 0.095 < 1$ is achieved, and the rate of change graph is shown as:

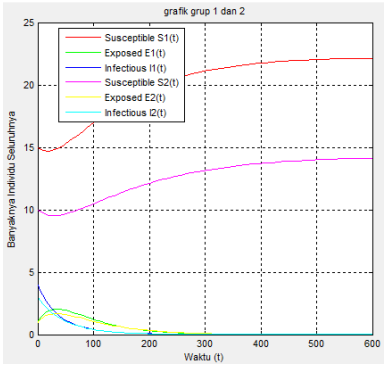


Figure 2: The Rate of Change for Disease Free Condition ($R_0 \leq 1$)

The Rate of Change on the *Susceptible* Population

In the beginning of the rate of change on the S*usceptible*, both groups experience a decrease in number due to the first infected individual and also slow infection rate, but subsequently the *Susceptible* population increase and become stable afterward, when the number of population is around 22 for the first group and around 14 for the second group. It shows that in free disease condition more individual will become susceptible on the contagious disease.

The Rate of Change on the *Exposed* Population

In the beginning of the graph on the *Exposed*, both groups experience an increase in number due to slow infection rate and first infected individual, therefore some individuals on the *Susceptible* population become E*xposed* population. After that it experience a decrease in number and become constant to zero, meaning that this system are free disease condition.

The Rate of Change on the *Infected* Population

The rate of change of infected population on both groups experiences a decrease in number and also become constant to zero, so that it can be concluded that the disease has disappeared from the population.

6.2 Endemic Condition($R_0 > 1$)

To have the endemic condition it can be done by increasing the infection rate which happen in free disease condition or by defining other parameters which also satisfy the model solution area. Parameters chosen for this condition are:

First Group	Second Group	First Value	Infection Rate
$\Lambda_1 = 0.2$ $d_1^S = 0.003$	$\Lambda_2 = 0.1$ $d_2^S = 0.002$	$S_1(0)$ $= 30$	$\beta_{11} = 0.005$ $\beta_{12} = 0.002$

$d_1^E = 0.004$ $d_1^I = 0.005$ $\epsilon_1 = 0.02$ $\gamma_1 = 0.011$	$d_2^E = 0.002$ $d_2^I = 0.002$ $\epsilon_2 = 0.02$ $\gamma_2 = 0.012$	$E_1(0) = 2$ $I_1(0) = 8$ $S_2(0)$ $= 20$ $E_2(0) = 2$ $I_2(0) = 6$	$\beta_{21} = 0.002$ $\beta_{22} = 0.002$

With the coefficient $p_1 = 1, q_1 = 1, p_2 = 1, q_2 = 1$.

By using these parameters, it will give $R_0 = 9.06 > 1$, and the rate of change can be shown as below:

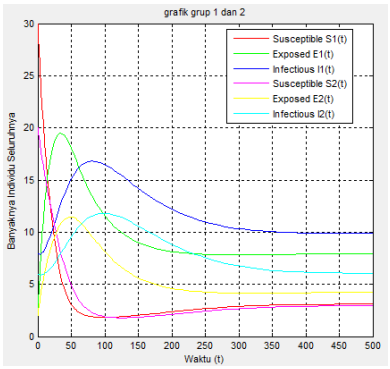


Figure 3: The rate of change of endemic condition ($R_0 > 1$)

The Rate of Change on the *Susceptible* Population

The rate of change of *Susceptible* population experience a decrease in both groups due to endemic condition of a disease, resulting more individuals from *Susceptible* population are infected and counted as in *Exposed* population. After that it increases a bit due to a big amount of recruitment rate and stable into 4.

The Rate of Change on the *Exposed* Population

The rate of change of *Exposed* population experiences an increase in number on both groups. But then it decrease due to the individuals in *Exposed* now showing number of *infections* and eventually they are sorted as *Infected* population and later on it stabilize on around 8 for the first group and around 4 for the second group due to no more addition from *Susceptible* that are infected.

The Rate of Change of the *Infected* Population

The rate of change of the *Infected* population experiences an increase in number on both groups because disease is endemic on the second group. The increasing rate of change on the *Infected* population depends on how many

individuals from the *Exposed* population have shown the symptoms of contagious disease. After that the rate of change on the *Infected* population decreases and become stable due to no more addition of the *Exposed* individuals showing the symptoms of the disease.

6.3 Condition where the First Group is Free Disease and the Second Group is Endemic ($R_1 > 1$)

Parameters chosen from the model solution area of this condition are:

First Group	Second Group	First Value	Infection Rate
$\Lambda_1 = 0.2$ $d_1^S = 0.009$ $d_1^E = 0.02$ $d_1^I = 0.025$ $\epsilon_1 = 0.003$ $\gamma_1 = 0.003$	$\Lambda_2 = 0.1$ $d_2^S = 0.002$ $d_2^E = 0.002$ $d_2^I = 0.002$ $\epsilon_2 = 0.02$ $\gamma_2 = 0.012$	$S_1(0) = 15$ $E_1(0) = 1$ $I_1(0) = 4$ $S_2(0) = 20$ $E_2(0) = 2$ $I_2(0) = 6$	$\beta_{11} = 0$ $\beta_{12} = 0$ $\beta_{21} = 0$ $\beta_{22} = 0.007$

With the coefficient $p_1 = 1, q_1 = 1, p_2 = 1, q_2 = 1$.

By using these parameters, it will give$R_1 = 9.09 > 1$, and the graph of the rate of change as below:

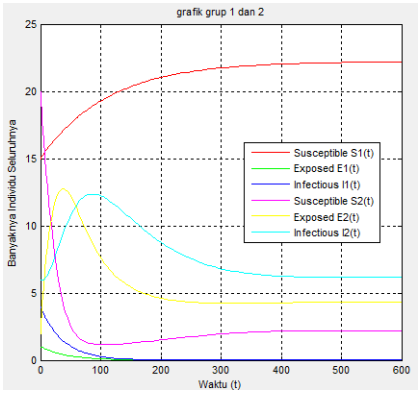


Figure 4: The Rate of Change Graph for First Group is Free Disease while Second Group is Endemic.

The Rate of Change on the *Susceptible* Population

The rate of change on the *Susceptible* population on the first group experience an increase than become constant due to no disease spreading. On

both groups the rate of change on the *Susceptible* population decrease because there is disease spread so that some of the Susceptible population on both groups are infected and sorted as *Exposed* population.

The Rate of Change on the *Exposed* Population

The rate of change on the *Exposed* population on first group decreases and become constant to zero due to no disease spreading. While on the second group the rate of change of the *Exposed* population increases and decrease becoming constant due to no disease spread.

The Rate of Change on the *Infected* Population

The rate of change on the *Infected* population on first group experience a decrease and become constant to zero. It indicates that the disease has been disappeared. While the rate of change of the *infected* population on the second group experience an increase in number because some of the *Exposed* population are sorted as *Infected* population. It decrease and become constant due to no more addition of the *Exposed* individuals that shows disease symptoms.

6.4 First Group is Endemic and Second Group is Free Disease ($R_2 > 1$)

Parameters chosen from the model solution area of this condition are:

First Group	Second Group	First Value	Infection Rate
$\Lambda_1 = 0.2$	$\Lambda_2 = 0.1$	$S_1(0) = 30$	$\beta_{11} = 0.005$
$d_1^S = 0.003$	$d_2^S = 0.007$	$E_1(0) = 2$	$\beta_{12} = 0$
$d_1^E = 0.004$	$d_2^E = 0.017$	$I_1(0) = 8$	$\beta_{21} = 0$
$d_1^I = 0.005$	$d_2^I = 0.02$	$S_2(0) = 10$	$\beta_{22} = 0$
$\epsilon_1 = 0.02$	$\epsilon_2 = 0.001$	$E_2(0) = 1$	
$\gamma_1 = 0.011$	$\gamma_2 = 0.003$	$I_2(0) = 3$	

With the coefficient $p_1 = 1, q_1 = 1, p_2 = 1, q_2 = 1$.

By using those parameters, it will give $R_2 = 7.81 > 1$, and the rate of change graph as below:

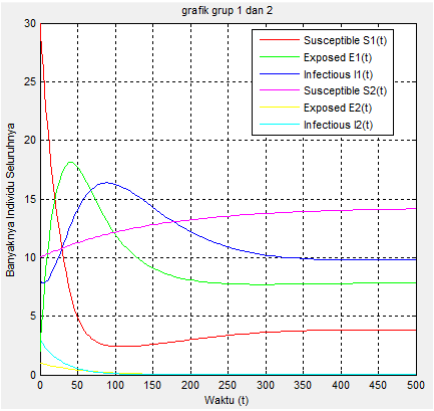


Figure 5: The Rate of Change Graph on First Group is Endemic and Second Group is Free Disease

The Rate of Change on the *Susceptible* Population

First group experiences a decrease in number while the second group increase due to no disease spread.

The Rate of Change on the *Exposed* Population

The rate of change on the Exposed population on first group increase, but then decrease and becoming constant.While the Exposed population on second group decrease and becoming constant as well to zero. The rate of change on the *Exposed* population is inversely related with a case where the first group is free disease and the second group is endemic.

The Rate of Change on the *Infected* Population

The first group increase due to the spread of the disease. But then decrease and become constant. It shows that on the first group the disease is endemic. While the second group decrease then becoming constant to zero.

7. Conclusion

According to the analysis and the explanation above we can conclude that:

1. Basic reproduction number $R_0 = \rho\left(\frac{\beta_{ij}\epsilon_i C_{ij}(S_i^0)}{(a_i^E+\epsilon_i)(a_i^I+\gamma_i)}\right)_{1\le i,j\le n}$, where ρrepresents spectral distance.
2. With the analysis of stability from the model provided previously, assuming $B = (\beta_{ij})$will not be reduced, and resulting:
   a. If $R_0 \le 1$ the equilibrium point of free disease P_0 is local stable asymptotic.

b. If $R_0 \leq 1$ the equilibrium point of free disease P_0 global stable asymptotic.
c. If $R_0 > 1$ the equilibrium point of free disease P^* local stable asymptotic.
d. If $R_0 > 1$ the equilibrium point of free disease P^* global stable asymptotic l.

3. The result of simulation using Matlab shows the basic reproduction number and also the relation between the stability analysis in free disease or endemic condition equilibrium point.

8. References

[1] Berman, A. and Plemmons, R. J. 1979. *Nonnegative Matrices in the Mathematical Science*. New York : Academic Press.

[2] Budayasa, K. 2007. *Teori Graph dan Aplikasinya*. Surabaya : Unesa University Press.

[3] Finizio N. and Ladas G. 1988. *Ordinary Differential Equations with Modern Applications*. California : Wadsworth Publishing Company Belmont.

[4] Li, M. Y., Shuai, Z. *Global-stability for Coupled Systems of Differential Equation on Network*. J. Differential Equation 248 (2010) 1-20.

[5] Rahmalia, D. 2010.*Pemodelan Matematika dan Analisis Stabilitas dari Penyebaran Penyakit Flu Burung*. Tugas Akhir S1 Jurusan Matematika ITS Surabaya.

[6] Sari, A.N. 2011. *Analisis Stabilitas dari Model Penyebaran Penyakit Menular Melalui Transportasi Antar Dua Wilayah (Kota).* Tugas Akhir S1 Jurusan Matematika ITS Surabaya.

[7] Wiggins, S. 1990. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. New York : Splinger-Verlag.

[8] Widodo,B. 2012. *Pemodelan Matematika*, ITS Press Surabaya – Indonesia.

FNPR: Using Priority Rules with Fuzzy Serious Queues

G. Geetharamani[1] , J. Arun Pandian[2]

[1,] Department Of Mathematics, Anna University, Barathidhasan Institute of Technology, Tiruchirappalli,Tamilnadu,India,
geeramdgl@rediffmail.com
[2,] Department Of Mobile and Pervasive Computing, Anna University, Barathidhasan Institute of Technology, Tiruchirappalli,Tamilnadu,India,
aparunpandian@gmail.com

Abstract. This paper determines the optimal results for scheduling and releasing packet transmission in wireless network using priority rules with fuzzy serious queues. In wireless network data transmission are not crisp and deterministic, also the data cannot be described precisely. In the proposed Flexible Network Packet Releasing (FNPR) model, the packet scheduling, controlling and releasing are performed using Approximate Reasoning (AR) technique. In this technique the priority rules are applied for packet scheduling and dispatching method is used for packet releasing. Slack time, waiting time and data criterion are considered as linguistic variable. By using these parameters membership values and degree of packet are identified then the values are forwarded to fuzzy inference system. The system has to decide the packet transmission based on the parameters and queue size. It will reduce the packet loss and enhance the throughput , transmission time.

Key words: Packet Releasing, packet scheduling, priority rules, fuzzy serious queues.

1. Introduction

Ubiquitous access to information anytime and anywhere will characterize whole new kinds of information systems in the future. This is being made possible by rapidly emerging communication systems that exploit wireless technologies. These systems have the potential to change how societies will evolve, as people are no longer constrained by information location or communication mechanisms. High speed networks technology, called also HSN, is a key component of these future wireless networks [2]. HSNs are likely to expand their presence in future applications [24].

Numerous challenges must be overcome to realize the practical benefits of HSN technology. These include power management, medium access, security, and of

principal interest here, Quality of Service (QoS) issue. It is expected that HSN technology will carry diverse kinds of multimedia services characterized by their high exigency level of quality delivery [3]. However, for various sensitive services like multimedia streaming or VoIP applications, additional Quality of Service (QoS) is required, especially in terms of bandwidth and end-to-end delay guarantees [4]. These services depend on the provision of adequate QoS support when deployed in the wireless networks.

One of the most crucial mechanisms of a model for providing QoS support is the traffic regulation [5][6]. This later has the task of measuring the state of the network's resources and thereby to decide which application data flows can be adapted [7]. In the past few years, some "conventional" QoS traffic adaptation models have been proposed [8]. These models, e.g. SWAN, are essentially based on traffic engineering mechanisms and QoS optimization techniques [9][10]. A number of works found in the literature have proposed techniques that build on a combination of well-established algorithms to provide distributed traffic control in wireless networks. For instance, AIMD (Additive Increase Multiplicative Decrease) and fair queuing have proven to be efficient components to achieve a distributed traffic control [11][12][13][14][15]. AIMD algorithm is implemented to control the sending rates of competing sources. While many of the proposals can provide some levelof QoS support they are based on a set of architectural assumptions where all nodes in the network implement a certain set of end-to-end control algorithms, or require the support of QoS-capable MAC at each node along the path. SWAN is the best example of protocols using AIMD algorithm [16]. However, one of the drawbacks of SWAN is how to calculate the threshold rate limiting any excessive delay that might be experienced [17] [18]. SWAN adopts engineering techniques that attempt to set the admission threshold rate at nodes to operate under the saturation level of the wireless channel.

In this paper, we attempt to exploit the efficiency of intelligent learning theories such as Fuzzy Logic theory and Fuzzy Approximate Reasoning in networks to achieve the traffic adaptation in HSNs. The use of fuzzy logic theory in our proposal is justified by the fact that this theory is well adapted to systems characterized by imprecise states, dynamic nature, and uncertain information, as in the case of HSNs. Furthermore, the use of fuzzy logic can add more flexibility and capability for managing both the random traffic state and the arbitrary buffer change occupancy of devices due to the dynamic of HSNs. This theory has been successfully applied in many industrial systems, such as embedded systems and telecommunications control.

The dynamic nature of traffic behavior in HSNs may generate uncertain information [24]. Then, it is useful to conceive new intelligent models that deal with the imprecision information caused by the dynamic events of a network. Our proposed model contributes to this issue by dealing essentially with two parameters: packets delay and buffer size in intermediate nodes of HSNs. By manipulating these parameters, our model is capable to generate the convenient traffic adaptation rate for multimedia sessions, in the aim to avoid congestion in network. The efficiency of using fuzzy logic and Petri nets theories to ensure the traffic adaptation is proved by the extensive simulations we achieved compared to the conventional AIMD algorithm and IEEE 802 standard. The information generated by these mathematical tools will be used to study the behavior the traffic states with the network and to introduce a new detection and recovery algorithm that deals with the possible happening of errors that may affect the reliability of the proposed model. Note that through this paper, we use the terms "Slack time", "Waiting time," and "Time criterion" interchangeably.

2. FuzzyHSN Model

The proposed model (Fig. 1) aims to study the traffic adaptation behavior in order to ensure a convenient QoS decision regarding the traffic regulation rate according to network conditions. Based on Slack time measurement and packets waiting time, the traffic adaptation decision policy is established as a set of production rules "IF THEN" that define the action to be performed in a response to the current network state. Each production rule includes a set of antecedent input conditions "IF-part" and consequent output propositions "THEN-part".

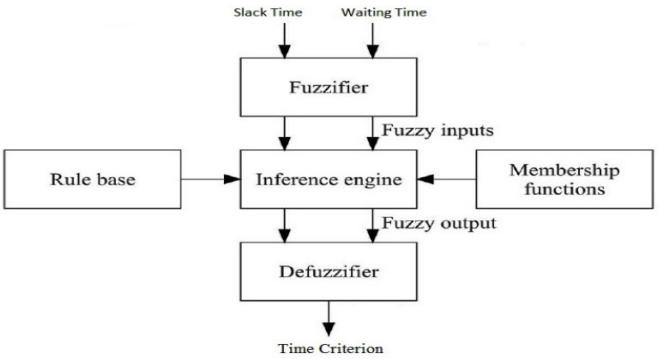


Fig 1: proposed fuzzyHSN model

The slack time and waiting time represents the time it took to send the packet between the transmitter and receiver including the total deferred time plus the time to fully acknowledge the packet. This is simply measured at the source node by subtracting the time that a packet is passed to the MAC layer from the time an ACK packet is received from the receiver.

It is important to note that the information about the status of a network in terms of congestion can be obtained tanks to the packet delay parameter. A big value of this parameter means that congestion may have appeared in the network; in this case, the process of traffic adaptation can start. The adaptation is performed in three steps:

2.1. Fuzzification: The fuzzy inputs in our system are the slack time and the waiting time obtained as feedback from MAC layer. The traffic adaptation rate represents the fuzzy output. These two parameters have to be converted into fuzzy sets. In fuzzy theory, a fuzzy set is different from an ordinary set, because a fuzzy set may contain elements that have different degree of membership in a set, whereas the elements of an ordinary set are considered members of a particular set if they have full membership in this set.

During the design of FuzzyHSN model, the choice of the rules base is performed depending on the manner how the system should behave to ensure the control of traffic rate. After classification of the appropriate rules, the membership functions associated to each parameter's rule are identified. For that aim, a variety of membership functions may be applied such as triangular, Gaussian, and trapezoidal functions.

Fuzzy logic is principally a multi-valued logic that allows transitional values to be defined between conventional evaluations like yes/no, true/false, and 0/1. Notations similar to warm cold or very cold can be formulated mathematically and developed by computers. Fuzzy numbers are generally used in a variety of fields, as the illustration uncertainty. In perform most of the things encountered in life are imprecise. Fuzzy logic is inherently robust since it does not necessitate precise, noise free inputs and humiliates gradually when system apparatus fail like if a feedback sensor quits or is destroyed. .The most frequently used shapes of fuzzy numbers are triangular and trapezoidal, which is appraised as follows.

Triangular Fuzzy Number

A triangular fuzzy number *A* can be parameterized by a triplet (a_1, a_2, a_3), where the membership function of the triangular fuzzy number *A* is distinct by

$$u_A(u) = \begin{cases} 0, & u < a_1 \\ \dfrac{u - a_1}{a_2 - a_1}, & a_1 < u < a_2 \\ \dfrac{a_3 - u}{a_3 - a_2}, & a_2 < u < a_3 \\ 0, & u > a_3 \end{cases} \quad (1)$$

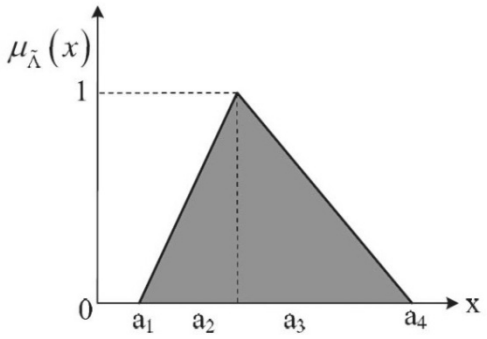


Fig 2: A triangular fuzzy number

Trapezoidal Fuzzy Number

A trapezoidal fuzzy number *A* can be parameterized by a four (a_1, a_2, a_3, a_4), where the membership function of the trapezoidal fuzzy number *A* is distinct by,

$$u_A(u) = \begin{cases} 0, & u < a_1 \\ \dfrac{u - a_1}{a_2 - a_1}, & a_1 < u < a_2 \\ 1, & a_2 < u < a_3 \\ \dfrac{a_3 - u}{a_3 - a_2}, & a_3 < u < a_4 \end{cases} \quad (2)$$

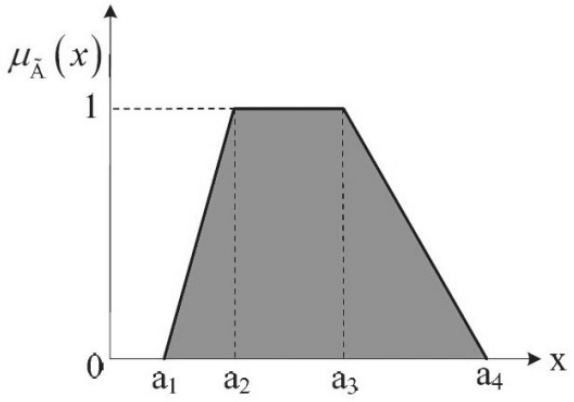


Fig 3: A trapezoidal fuzzy number

In our study, we have chosen triangular functions because of their simplicity in computation. A fuzzy set in the universe of discourse *U*, $U = (u_1, u_2 \ldots u_n)$ which can be described by a membership function u_A:$U \rightarrow$ [0,1] , represented by,

$$u_A = u_A(u_1)/u_1 + u_A(u_2)/u_2 + \cdots u_A(u_n)u_n \quad (3)$$

where $u_A(u_i)$, indicates the grade of membership of u_i in the fuzzy set *A* (e.g., to indicate the waiting time in the sets "*high*", "*medium*", or "*low*").
Let *A* and *B* be two triangular fuzzy numbers parameterized by the triplets, (a_1, a_2, a_3), (b_1, b_2, b_3), respectively. The addition and multiplication operations between the triangular fuzzy numbers *A* and *B* is defined, respectively, as follows:

$$A \oplus B = (a_1, a_2, a_3) \oplus (b_1, b_2, b_3)$$
$$= (a_1 + b_1,\ a_2 + b_2,\ a_3 + b_3)$$

$$A \otimes B = (a_1, a_2, a_3) \otimes (b_1, b_2, b_3)$$
$$= (a_1 * b_1,\ a_2 * b_2,\ a_3 * b_3)$$

The maximum operation Θ between *A* and *B* is defined as follows:

$$A \Theta B = (a_1, a_2, a_3) \Theta (b_1, b_2, b_3)$$

$$=(a_1 v b_1,\ a_2 v b_2,\ a_3 v b_3)$$

Where an Î [0, 1].

2.2. Rules Evaluation: in this step, a set of rules is applied using an inference engine to describe the different steps of the traffic adaptation process. The rule base includes a set of conditional statements as follows:

R1: IF waiting time is long **AND** slack time is Critically-short **THEN** time criterion is urgent (1.0)
R2: IF waiting time is medium **AND** slack time is Critically-short **THEN** time criterion is urgent (0.8)
R3: IF waiting time is short **AND** slack time is Critically-short **THEN** time criterion is urgent (0.6)
R4: IF waiting time is long **AND** slack time is short **THEN** time criterion is urgent (0.5)
R5: IF waiting time is medium **AND** slack time is short **THEN** time criterion is urgent (0.2)
R6: IF waiting time is medium **AND** slack time is short **THEN** time criterion is not-urgent (0.7)

2.3. Defuzzification: In this step, the decision sets concerning the traffic adaptation are converted into precise quantities. There are several heuristics methods that permit to perform the defuzzification: mean of maximum, center of area, and max criterion. In our model, we have used the mean of maxima (MoM) method as defuzzifier method because of its light computational complexity. The evaluation result is obtained as the average of the elements that reach the maximum grade in a fuzzy set.
In what follows, we will describe the modeling concept of traffic adaptation rules. FuzzyHSN considers both the traffic state and the network resources availability parameters. The input parameters are represented by both the Waiting time and the Slack time while the output parameter is represented by the Time criterion.

3. Experimental Result:

In this paper, we have made some assumptions regarding the routing protocol. We are assuming that a reactive multipath routing protocol AOMDV exists that makes node disjoint or link disjoint multiple paths from source to destination. The proposed scheme will assist the AOMDV in reducing the delay when route failure occurs.

3.1. Performance Metrics

Following performance metrics are used to measure the performance of the proposed technique.

3.1.1 Packet Loss Percentage

The percentage of data packets dropped in the network either at the source or at intermediate nodes. Performance will get better, when the packet loss will decrease with the proposed technique.

3.1.2. End-To-End Delay

It is the time interval between the event when the source has transmitted a data packet and the event when destination receives that data packet. In the case, no link breakage prediction algorithm and multipath routing protocol are present then this delay will be high because the time spent in building a route is included in the end-to-end delay.

3.1.3. Route Discovery Time

The aggregate number of route requests generated by all sources per second. With the proposed technique, this time will be decreased.

3.2. Simulation Setup

For the evaluation of the proposed idea, the model for simulation that was used here is same as discussed in [1]. The version of network simulator used was ns2.34. A total of 100 nodes network was used in an area of 1000×1000 m2. The nodes were deployed randomly in the simulation area. The mobility model that was used for the node movement was random waypoint mobility model. There was no pause time between node movements, so nodes were moving continuously without any movement pause. The average speed v of the nodes was kept varying so that the mobility rate can change. The real speed of nodes was selected from the defined range of 0.9v to 1.1v.

Table I. SIMULATION PARAMETER VALUES

Parameter	**Value**
Simulation time	100 sec
Number of nodes	90
Number of stationary nodes	10
MAC layer	WLAN
Traffic type	CBR
Packet size	512 bytes
No. of UDP connections	Many
Movement pause time	0 sec
Maximum speed	Varying

The traffic was created with many UDP sessions with CBR traffic type. Each UDP session was started randomly in the first 100 seconds of the simulation time and the traffic sources and destinations were also selected randomly. The traffic sessions last till the end of the simulation. Total simulation time was 1000 seconds. The parameters that were configured are shown in Table 1.The membership function of waiting time, slack time and Time criteria are defined (fig 4-6).

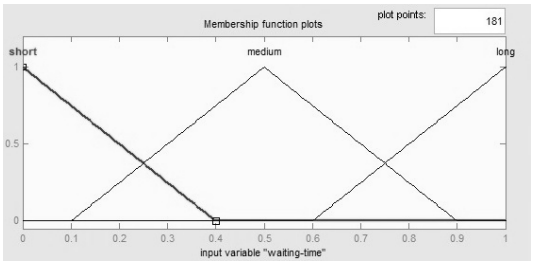


Fig 4. Input Membership function of packet waiting time

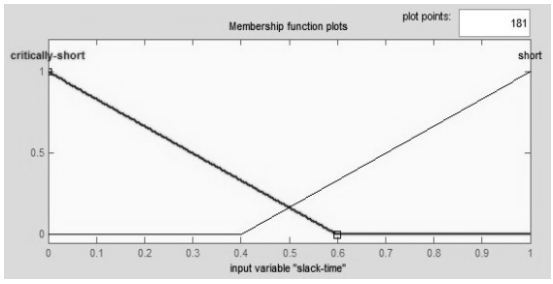


Fig 5. Input Membership values of packet Slack time

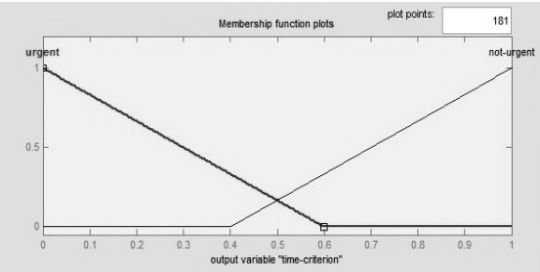


Fig 6. Output Membership values of time criterion

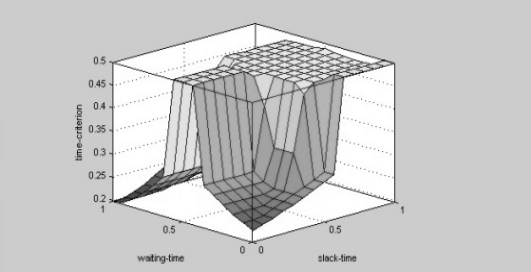


Fig 7. Surface value of the proposed model

Figure 8 shows the Surface of the proposed fuzzy controller, Figure 8 shows the percentage of packets lost with the increase in node velocity. The packet loss here is due to the mobility of intermediate nodes. When an intermediate data relaying node moves away, then the packets sent by the neighbor node on the transmission path are lost.

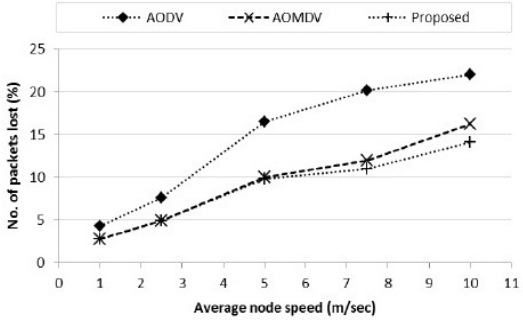


Fig 8. Packet loss comparison with varying speed.

Also, when an intermediate node finds no path towards a destination, it drops the packet. Packets are also dropped by the sender in case of buffer overflow or when it fails to get the route after making several route discovery requests. It is

also evident from the Figure 9 that the percentage of packet loss is increasing for all the three protocols with an increase in the average speed of the node. The simulations were carried out 5 times setting varying node speeds.

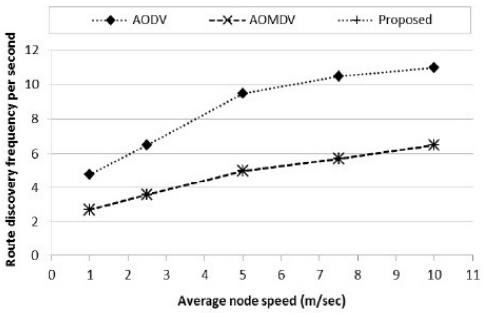


Fig 9. Route discovery frequency comparison with varying speed

The AOMDV and proposed protocol always drop less number of packets as compared to the standard AODV protocol. The reason for this smaller packet loss is the availability of alternate paths to forward the packets when the primary path fails. The packet loss with the proposed idea is still less than the AOMDV because the proposed idea predicts in advance the status of a link and may shift the traffic to alternate path if an alternate path exists at that node or it generates the route error message, so that a new path can be established.

Figure 9 shows the comparison of route discovery frequency of proposed protocol with the AODV and AOMDV. The route discovery frequency also increases with an increase in the average node speed. When the nodes are more mobile, then the path failures are frequent that resulted in the generation of more route discovery procedures.

The route discovery frequency of the proposed idea and AOMDV are same. It is because, the proposed idea uses AOMDV for routing and helps it in detecting the link status in advance, thus reducing the delay in route discovery process rather than reducing the route discovery frequency. So, no improvement in the route discovery frequency is observed as compared to the AOMDV.

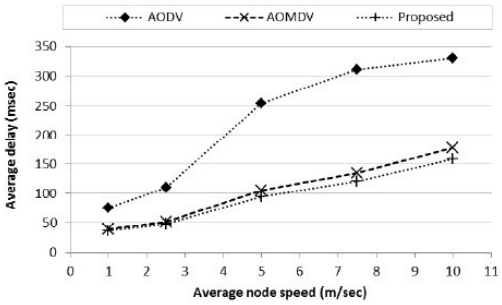


Fig 10. End-to-end delay comparison with varying speed

Figure 10 shows the comparison of average end-to-end delay with varying speed. Average end-to-end delay increases with the mobility increase. The reason is that, with an increase in the number of route failures, the route discovery operation is triggered that increases the end-to-end delays. The average end-to-end delay for the proposed idea is lowest as compared to the other two because the proposed idea predicts in advance the status of a link and in case the link status is about-to-break, then it can trigger the route discovery procedure in advance, which will minimize the average end-to-end delay.

4. Conclusion:

In this paper, we have introduced a new traffic adaptation model in HSNs based on fuzzy logic theory. The model named FuzzyHSN aims to enhance the traffic regulation rate decision of multimedia real time sessions, based on the slack time computed as the feedback delay measurement between the source and destinations nodes, and the packet waiting time in intermediate nodes. We have evaluated the performance of FuzzyHSN, compared to AIMD-SWAN and IEEE 802.11, under diverse traffic and network conditions. The experimental results have demonstrated that the QoS design based on fuzzy theory promises to be an efficient solution and may constitute a good alternative to some conventional methods based on complex optimization and overloaded traffic engineering mechanisms. The promising results indicate that the proposed model can be hopeful to deal with the dynamics of mesh networks when managing QoS delivery especially for supporting multimedia services.

References

1. M. K. Marina and S. R. Das, "Ad hoc on-demand multipath distance vector routing," *Wiley Wirel. Commun. Mob. Comput.*, vol. 6, pp. 969–988, (2006).
2. H. Jiang, P. Wang, H. V. Poor, and W. Zhuang, "A distributedMAC scheme supporting voice services in mobile ad hoc networks," Wireless Commun. Mob. Comput., vol. 10, no. 4, pp. 547–558, (2010).
3. R. Ben Ali, A. Hafid, and J. Rezgui, "An Enhanced reservation based medium access control with scheduling and admission control for voice over wireless mesh networks," IEEE Trans. Wireless Commun., vol. 11, no. 11, pp. 3540–3549, (2012).
4. P. Wang and W. Zhuang, "A collision-free mac scheme for multimedia wireless mesh backbone," IEEE Trans. Wireless Commun., vol. 8, no. 7, pp. 3577–3589, (2009).
5. G.-S. Ahn, A. T. Campbell, A. Veres, and L.-H. Sun, "Supporting service differentiation for real-time and best-effort traffic in stateless wireless ad hoc networks (swan)," IEEE Trans. Mobile. Comput., vol. 1, pp. 192–207, (2002).
6. A. El-Masri, L. Khoukhi, A. Sardouk, and D. Gaiti: "Wirs: resource reservation and traffic regulation for QoS support in wireless mesh networks," in Proc. IEEE Global Telecommun. Conf. (IEEE Globecom), Houston, USA, (2011).
7. S.-Y. Oh, G. Marfia, and M. Gerla, "Manet QoS support without reservations," Security Commun. Netw., vol. 4, no. 3, pp. 316–328, (2011).
8. L. Khoukhi and S. Cherkaoui, "Intelligent QoS management for multimedia services support in wireless mobile ad hoc networks," Comput. Netw. (COMNet), vol. 54, no. 10, pp. 1692–1706, (2010).
9. K. Chen, K. Nahrstedt, and N. Vaidya, "The utility of explicit ratebased flow control in mobile ad hoc networks," in IEEE Wireless Commun. Netw. Conf. (WCNC), pp. 1921–1926, (2004).
10. L. Khoukhi, A. El Masri, A. Sardouk, A. Hafid, and D. Gaiti, "Traffic adaptation in wireless mesh networks: Fuzzy-based model," in IEEE Wireless Commun. Mobile Comput. Conf. (IWCMC), (2011).
11. D. Bansal and H. Balakrishnan, "Tcp-friendly congestion control for real-time streaming applications," MIT Lab. Comp. Sci., Tech. Rep. MIT-LCS-TR-806, (2000).
12. K. K. Ramakrishnan and R. Jain, "A binary feedback scheme for congestion avoidance in computer networks," ACM Trans. Comput. Syst., vol. 8, no. 2, pp. 158–181, (1990).

13. Y. Xia, L. Subramanian, I. Stoica, and S. Kalyanaraman, "One more bit is enough," IEEE/ACM Trans. Netw., vol. 16, no. 6, pp. 1281–1294, Dec. (2008).
14. E. Altman, I. Bp, and F. D. Ingeniera, "Novel delayed ack techniques for improving tcp performance in multihop wireless networks," in Personal Wireless Commun. (PWC) , pp. 23–25, (2003)
15. D. Kliazovich and F. Granelli, "Cross-layer congestion control in ad hoc wireless networks," Ad Hoc Netw., vol. 4, (2006).
16. G. Marfia, P. Lutterotti, S. Eidenbenz, G. Pau, and M. Gerla, "Faircast: fair multi-media streaming in ad hoc networks through local congestion control," in Proc. 11th Int. Symp. Model. Anal. Simul. Wireless Mobile Syst. (MSWiM), (2008).
17. C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of service support in IEEE 802.16 Networks," IEEE Network, vol. 20, no. 2, pp. 50–55, (2006).
18. A. Sayenko, O. Alanen, and T. Hämäläinen, "Scheduling solution for the IEEE 802.16 base station," Comput. Netw., vol. 52, no. 1, 96–115, (2008).
19. J. Lakkakorpi and A. Sayenko, "Backhaul as a bottleneck in IEEE 802.16e networks," in Proc. IEEE Global Telecommun. Conf. (GLOBECOM), (2008).
20. E.V. Broekhoven and B.D. Baets, "Fast and accurate center of gravity defuzzification of fuzzy systemoutputs defined on trapezoidal fuzzy partitions," Fuzzy Sets Syst., vol. 157, no. 7, pp. 904–918, (2006).
21. W. V. Leekwijck and E. E. Kerre, "Defuzzification: Criteria and classification," Fuzzy Sets Syst., 108, 2, 159–178, (1999).
22. National Instruments Corporationl, Labview 2011 PID and Fuzzy Logic Toolkit Help, (2011).
23. M. B. Dwyer and L. A. Clarke, "A compact petri net representation and its implications fornanalysis," IEEE Trans. Softw. Eng., 22, 11, pp. 794–811, Nov. (1996).
24. Endicott, D and Frost, V. "Performance Experiences With Wide Area Highspeed Networks", Communications Magazine, IEEE, 35, 8, 106 – 108, (1997).
25. B.Palpandi, Dr. G.Geetharamani and J.Arun Pandian, "Performance Enhancement in OSI Network Model using Fuzzy Queue", Indian journal of applied research, 4, (2014).

An approximation solution of fuzzy differential equations using a new two step RK method

Ali Ahmadian[1,2], Mohamed Suleiman[2], Norazak Senu[1,2], and Soheil Salahshour[3]

[1] Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia.
ahmadian.hosseini@gmail.com,
[2] Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia.
[3] Department of Computer Engineering, College of Technical and Engineering, Mashhad Branch, Islamic Azad University,Mashhad, Iran.

Abstract. In this paper we study numerical methods for addressing fuzzy differential equations by an application of an accelerated Runge-Kutta method using a Bede's characterization theorem . We use an accelerated Runge-Kutta method in order to enhance the order of accuracy of solutions by reusing of function evaluations from the previous step which is reduced the number of function evaluations.We state a convergence result and give numerical example to illustrate the efficiency of the proposed method.
Keywords: Fuzzy ordinary differential equation; Bede's Characterization Theorem; Numerical methods; Accelerated Runge-Kutta method; Seikkala derivative.

1 Introduction

The application of fuzzy differential equations (FDEs) has been rapidly growing in recent years. Particularly (FDEs) appear in the diverse fields of sciences and engineering as particle systems [23], quantum optics and gravity [24], medicine [4], population models [13], synchronize hyperchaotic systems [39] and bioinformatics and computational biology [12].
The applicable definitions of fuzzy derivative and the fuzzy integral was studied by Dubois and Prade [14] and Puri and Ralescu [32]. Using Hukuhara derivative, Kaleva [21] and Seikkala [36] started to develop a theory of FDEs. Subsequently, several authors have been inspired to apply numerical and analytical methods for the solution of these equations [1, 3, 5–9, 25–27, 29, 30, 34, 35]. The most important contribution on these numerical methods is the Euler method provided by Ma et al. in [22]. Abbasbandy and Allahviranloo in [2] developed four-stage order Runge-Kutta (RK) methods for a Cauchy problem with a fuzzy initial value. Also Palligkinis and Papageorgiou [28] applied RK methods for a more general category of problems, and they proved convergence for s-stage RK methods. Finally, Effati and Pakdaman [15] solved FDEs by artificial neural network method.

In [11], Bede proved a Characterization Theorem which states that under certain conditions a fuzzy differential equation under Hukuhara differentiability is equivalent to a system of ordinary differential equations (ODEs). Bede also remarked that this Characterization Theorem can help to solve FDEs numerically by converting them to a system of ODEs, which can then be solved by any numerical method suitable for ODEs. More specifically, in [10], Bede wrote, "in order to obtain numerical solutions of FDEs under Hukuhara differentiability, it is not necessary to rewrite the whole literature on numerical solutions of ODEs in the fuzzy setting, but instead we can use any numerical method for the ODEs directly". The contribution of this paper is to use Bede's Characterization Theorem to solve FDEs numerically by the fourth-order accelerated Runge-Kutta (ARK) method.
The rest of the paper is organized as follows. In Section 2, we will give some necessary notations and definitions of fuzzy set theory, fuzzy differential equations, characterization theorem and a family of accelerated RungeKutta methods. Fourth order ARK method for solving FDEs is introduced in Section 3. Also Convergence and stability of the mentioned method are proved in this section. The proposed algorithm is illustrated by solving three examples in Section 4. At the end of the paper we present some conclusions and further research topics.

2 Preliminaries

We give some definitions and introduce the necessary notation which will be used throughout the paper. See for example [16, 18].

2.1 Definitions and notations

We consider R, the set of all real number. A fuzzy number is mapping $u : R \to [0, 1]$ with the following properties:
(a) u is upper semi-continuous,
(b) u is fuzzy convex, i.e., $u(\lambda x + (1 - \lambda)y \geq min\{u(x), u(y)\}$ for all $x, y \in R, \lambda \in [0, 1]$,
(c) u is normal, i.e., $\exists x_0 \in R$ for which $u(x_0) = 1$,
(d) $supp\ u = \{x \in R | u(x) > 0\}$ is the support of the u, and its closure cl(supp u) is compact.

Let E be the set of all fuzzy number on R. The r-level set of a fuzzy number $u \in E, 0 \leq r \leq 1$, denoted by $[u]_r$, is defined as

$$[u]_r = \begin{cases} \{x \in R | u(x) \geq r\} & if\ 0 < r \leq r \\ cl(supp\ u) & if\ r = 0 \end{cases}$$

It is clear that r-level set of a fuzzy number is a closed and bounded interval $[\underline{u}(r), \overline{u}(r)]$, where $\underline{u}(r)$ denotes the left-hand endpoint of $[u]_r$ and $\overline{u}(r)$ denotes the right-hand endpoint of $[u]_r$.

Definition 1. *Let $f : R \to E$ be a fuzzy function. We say f is differentiable at $t_0 \in R$, if there exists an element $f^{'}(t_0) \in E$ such that limits*

$\lim_{h\to 0^+} \frac{f(t_0+h)\ominus f(t_0)}{h}$ *and* $\lim_{h\to 0^+} \frac{f(t_0)\ominus f(t_0-h)}{h}$
exist and are equal to $f^{'}(t_0)$. Here the limits are taken in the metric space (E, D).

Next we review one of the main results from Bede [11](let $\|.\|$ denote the usual Euclidean norm).

Theorem 1. *(Characterization Theorem) [11] Let us consider the fuzzy initial value problem (FIVP)*

$$\begin{cases} x' = f(t, x), \\ x(t_0) = x_0, \end{cases} \tag{1}$$

where $f : [t_0, t_0 + a] \times E \to E$ is such that
(i) $[f(t, x)]^r = [\underline{f}^r(t, \underline{x}, \overline{x}), \overline{f}^r(t, \underline{x}, \overline{x})]$,
(ii) $\underline{f}^r$ and $\overline{f}^r$ are equicontinuous (that is, for any $\epsilon > 0$ there is a $\delta > 0$ such that $|\underline{f}^r(t, x, y) - \underline{f}^r(t, x, y)| < \epsilon$ and $|\overline{f}^r(t, x, y) - \overline{f}^r(t, x, y)| < \epsilon$ for all $r \in [0, 1]$, whenever $(t, x, y), (t_1, x_1, y_1) \in [t_0, t_0 + a] \times R^2$ and $\|(t, x, y) - (t_1, x_1, y_1)\| < \delta$ and uniformly bounded on any bounded set,
(iii) there exists an $L > 0$ such that
$|\underline{f}^r(t, x, y) - \underline{f}^r(t, x, y)| \le L \max\{|x_2 - x_1|, |y_2 - y_1|\}$ *for all $r \in [0, 1]$,*
$|\overline{f}^r(t, x, y) - \overline{f}^r(t, x, y)| \le L \max\{|x_2 - x_1|, |y_2 - y_1|\}$ *for all $r \in [0, 1]$.*

Then the FIVP (1) and system of ODEs

$$\begin{cases} (\underline{x}^r(t))' = \underline{f}^r(t, \underline{x}^r, \overline{x}^r) \\ (\overline{x}^r(t))' = \overline{f}^r(t, \underline{x}^r, \overline{x}^r) \\ \underline{x}^r(t_0) = (x_0^r) \\ \overline{x}(t_0) = (\overline{x_0^r}) \end{cases} \tag{2}$$

are equivalent.

2.2 Fourth order accelerated Runge-Kutta

The general form of the ARK methods presented in this paper is

$$y_{n+1} = c_0 y_n - c_{-0} y_{n-1} + c_1 k_1 - c_{-1} k_{-1} + \sum_{i=2}^{v} c_i (k_i - k_{-i}) \quad for\ 1 \le n \le N-1, \tag{3}$$

Table 1. Optimized ARK4 parameters set .

Parameters	constant set
c_0	1.0
c_{-0}	0.0
c_1	1.022831928839203211581411
c_{-1}	0.022831928839203211581410 16
c_2	- 0.045158301883180231641969 73
c_3	- 0.086187006135813174734622 00
c_4	0.608513379179790194795185 5
a_1	0.246418984804535202766398 8
a_2	0.379427607085112010701626 9
a_3	0.756756177970740702853666 9

where

$$\begin{cases} k_1 = hf(y_n), & k_{-1} = hf(y_{n-1}), \\ k_2 = hf(y_n + a_1k_1), & k_{-2} = hf(y_{n-1} + a_1k_{-1}), \\ k_3 = hf(y_n + a_2k_2), & k_{-3} = hf(y_{n-1} + a_2k_{-2}), \\ k_4 = hf(y_n + a_3k_3), & k_{-4} = hf(y_{n-1} + a_3k_{-3}). \\ k_5 = hf(y_n + a_4k_4), & k_{-5} = hf(y_{n-1} + a_4k_{-4}). \end{cases} \tag{4}$$

where v denotes the number of function evaluations performed at each time step and increases with the order of local accuracy of the ARK method. The fourth-order ARK with using of (3),(4) and $v = 4$ is of the following form:

$$y_{n+1} = y_n + c_1k_1 - c_{-1}k_{-1} + c_2(k_2 - k_{-2}) + c_3(k_3 - k_{-3}) + c_4(k_4 - k_{-4}), \tag{5}$$

where

$$\begin{aligned} &k_1 = f(y_n), k_{-1} = f(y_{n-1}), k_2 = f(y_n + a_1k_1), k_{-2} = f(y_{n-1} + a_1k_1), \\ &k_3 = f(y_n + a_2k_2), k_{-3} = f(y_{n-1} + a_2k_{-2}), k_4 = f(y_n + a_3k_3), k_{-4} = f(y_{n-1} + a_2k_{-4}). \end{aligned} \tag{6}$$

Specific nonzero constants for the fourth order ARK which we applied in this paper, are in table 1 . The local truncation error is [31]

$$T(t, h) = \frac{31h^5}{360}y^{(5)}(\zeta_n) + O(h^6).$$

Table 1 is a set of approximate values and naturally contains some round-off error. The effects of this round-off error (in parameter values) can be significant if the error produced by the ARK methods is small enough to be of the same order. Therefore, to measure the accuracy of the ARK methods correctly (see [31]) ,we have considered the parameter values using 25 digits of computational precision in MATLAB. Note, however, that such high precision is not required for the ARK methods to be effective.
It is important to note that at the first step, there is no previous step. Therefore,

ARK methods cannot be self-starting. A one-step method must supply the approximate solution y_1 at the end of the first step t_1. The one-step method must be of sufficient order to ensure that the difference $y_1 - y(t_1)$ is of order p or higher. For example, the ARK4 method can be started by the RK4 method.

3 Fourth order fuzzy accelerated Runge-Kutta method

Consider the autonomous fuzzy initial value problem (FIVP)

$$\begin{aligned} &x'(t) = f(x(t)),\\ &x(0) = x_0, \quad t \in I = [0,T], \end{aligned} \tag{7}$$

where $f : R \to E$ be a fuzzy valued function and $x_0 \in E$.
Theorem 1 shows us a way to translate the FIVP (7) into a system of ODEs which is equivalent with the FIVP. Let $[x(t)]^r = [\underline{x}_r(t), \overline{x}_r(t)]$. If $x(t)$ is Seikkala' s differentiable then $[x'(t)]^r = [\underline{x}'_r(t), \overline{x}'_r(t)]$, and (7) translates into the following system of ODEs:

$$\begin{cases} \underline{x}'(t) = \underline{f}_r(\underline{x}_r, \overline{x}_r), \underline{x}(0) = \underline{x_0},\\ \overline{x}'(t) = \overline{f}_r(\underline{x}_r, \overline{x}_r), \overline{x}(0) = \overline{x_0}, \end{cases} \tag{8}$$

where $[f(x)] = [\underline{f}_r(\underline{x}_r, \overline{x}_r), \overline{f}_r(\underline{x}_r, \overline{x}_r)]$. Then, it can state that if we ensure that the solution $[\underline{x}_r, \overline{x}_r]$ of the system (8) are valid level sets of a fuzzy number valued function and $[\underline{x}'_r, \overline{x}'_r]$ are valid level sets of a fuzzy valued function, then by the Stacking Theorem [21], it is possible to construct the solution of FIVP (7).
The Characterization Theorem 1 [11] proved that FIVP (7) is equivalent to the system of ordinary differential equations (8) under ceratin conditions.
Now, we consider fuzzy initial value problem (7) with step-size $h = \frac{T}{N}$ and grid points

$$t_j = j.h, \qquad j = 0, 1, ..., N.$$

at which the exact solution $[Y(t)]^r = [\underline{Y}_r(t), \overline{Y}_r(t)]$ is approximated by $[y(t)]^r = [\underline{y}_r(t), \overline{y}_r(t)]$.The generalized fourth-order ARK method based on the fourth-order approximation of $\underline{Y}'_r(t)$ and $\overline{Y}'_r(t)$ and Eqs. (5),(6) and (8) is obtained as follows:

$$\begin{cases} \underline{y}_{n+1}(r) = \underline{y}_n(r) + F(\underline{y}_n(r), \overline{y}_n(r)),\\ \overline{y}_{n+1}(r) = \overline{y}_n(r) + G(\underline{y}_n(r), \overline{y}_n(r)),\\ y(0;r) = \underline{y}_0(r),\\ y(0;r) = \overline{y}_0(r), \end{cases} \tag{9}$$

where

$$\begin{cases} F(\underline{y}_n(r), \overline{y}_n(r)) = c_1\underline{k_1} - c_{-1}\underline{k_{-1}} + c_2(\underline{k_2} - \underline{k_{-2}}) + c_3(\underline{k_3} - \underline{k_{-3}}) + c_4(k_4 - \underline{k_{-4}}),\\ G(\underline{y}_n(r), \overline{y}_n(r)) = c_1\overline{\overline{k_1}} - c_{-1}\overline{\overline{k_{-1}}} + c_2(\overline{k_2} - \overline{\overline{k_{-2}}}) + c_3(\overline{k_3} - \overline{\overline{k_{-3}}}) + c_4(\overline{k_4} - \overline{\overline{k_{-4}}}), \end{cases}$$

at which

$$\begin{cases} \underline{k}_1 = h\underline{f}(y_n(r)),\ \overline{k}_1 = h\overline{f}(y_n(r)), \\ \underline{k}_2 = h\underline{f}(y_n(r) + a_1\underline{k}_1),\ \overline{k_2} = h\overline{f}(y_n(r) + a_1\overline{k_1}), \\ \underline{k}_3 = h\underline{f}(y_n(r) + a_2\underline{k}_2),\ \overline{k_3} = h\overline{f}(y_n(r) + a_2\overline{k_2}), \\ \underline{k}_4 = h\underline{f}(y_n(r) + a_3\underline{k}_3),\ \overline{k_4} = h\overline{f}(y_n(r) + a_3\overline{k_3}), \end{cases}$$

and

$$\begin{cases} \underline{k_{-1}} = h\underline{f}(y_{n-1}(r)),\ \overline{k_{-1}} = h\overline{f}(y_{n-1}(r)), \\ \underline{k_{-2}} = h\underline{f}(y_{n-1}(r) + a_1\underline{k_{-1}}),\ \overline{k_{-2}} = h\overline{f}(y_{n-1}(r) + a_1\overline{k_{-1}}), \\ \underline{k_{-3}} = h\underline{f}(y_{n-1}(r) + a_2\underline{k_{-2}}),\ \overline{k_{-3}} = h\overline{f}(y_{n-1}(r) + a_2\overline{k_{-2}}), \\ \underline{k_{-4}} = h\underline{f}(y_{n-1}(r) + a_3\underline{k_{-3}}),\ \overline{k_{-4}} = h\overline{f}(y_{n-1}(r) + a_3\overline{k_{-3}}). \end{cases}$$

where y_0 is an initial value.

4 Examples

Example 1. [17] Consider the fuzzy initial value problem

$$y'(t) = v_1 y^2(t) + v_2, \quad y(0) = y_0, \qquad (10)$$

at which $\underline{y}_0 = -0.0012 + 0.0012r, \overline{y}_0 = 0.0012 - 0.0012r$ and $v_i > 0$ for $i = 1,$ are fuzzy numbers.

Regarding to a Note published by Karimi et al. [33], the exact solution is given by

$$Y_1(t,r) = \frac{A\tan(\frac{At}{4} + \arctan(\frac{-3+3r^2}{1250A}))}{2+2r},$$

$$Y_2(t,r) = \frac{-\sqrt{2}B\tan(\frac{\sqrt{2}}{4}[Bt + 2\sqrt{2}\arctan(\frac{3\sqrt{2}(-3+r)(-1+r)}{2500B})]}{-6+2r},$$

where

$$A = \sqrt{6 + 8r + 2r^2},$$
$$B = \sqrt{(-5+r)(-3+r)}.$$

The result of ARK4 method, RK4 formula and Euler method with $h = 0.1$ at $t = 1$ are shown in Table 2.
The exact and approximate solutions by Euler, ARK4 and RK4 methods are compared and plotted at $t = 1$ in Fig. 1.

5 Conclusion

A family of accelerated Runge-Kutta formulae has been applied in this paper. The ARK methods are more computationally efficient than RK methods at

Table 2. The result of the accelerated ARK method for Example **??** with $h = 0.1$.

r	$\underline{y}_n^r$	$\underline{yrk4}_n^r$	$\underline{Eu.}$	$\underline{Y}$	$\lvert\underline{y}_n^r - \underline{Y}\rvert$	$\lvert\underline{yrk4}_n^r - \underline{Y}\rvert$	$\lvert\underline{Eu.} - \underline{Y}\rvert$
0	8.585372e-1	8.585647e-1	8.381748e-1	8.585385e-1	1.320206e-6	2.612539e-5	2.036377e-2
0.1	9.060921e-1	9.061242e-1	8.807003e-1	9.060941e-1	2.059972e-6	3.002945e-5	2.539387e-2
0.2	9.568802e-1	9.569156e-1	9.254290e-1	9.568834e-1	3.182790e-6	3.215571e-5	3.145444e-2
0.3	1.011350e0	1.011397e0	9.725813e-1	1.011354e0	4.885827e-6	4.20363e-5	3.877366e-2
0.4	1.070031e0	1.070088e0	1.022411e0	1.070039e0	7.472675e-6	4.852846e-5	4.762847e-2
0.5	1.133558e0	1.133625e0	1.075207e0	1.133570e0	1.141538e-5	5.497288e-5	5.836302e-2
0.6	1.202690e0	1.202776e0	1.131286e0	1.202708e0	1.745615e-5	6.765738e-5	7.142234e-2
0.7	1.278352e0	1.278456e0	1.191015e0	1.278379e0	2.677684e-5	7.633849e-5	8.736466e-2
0.8	1.361678e0	1.361809e0	1.254799e0	1.361719e0	4.128642e-5	8.940021e-5	1.069205e-1
0.9	1.454074e0	1.454242e0	1.323101e0	1.454138e0	6.411828e-5	1.039685e-4	1.310370e-1
1	1.557307e0	1.557406e0	1.396394e0	1.557407e0	1.005098e-4	1.724654e-5	1.610137e-1

r	$\overline{y}_n^r$	$\overline{yrk4}_n^r$	$\overline{Eu.}$	$\overline{Y}$	$\lvert\overline{y}_n^r - \overline{Y}\rvert$	$\lvert\overline{yrk4}_n^r - \overline{Y}\rvert$	$\lvert\overline{Eu.} - \overline{Y}\rvert$
0	4.467569e0	4.490504e0	2.605837e0	4.499280e0	3.171124e-2	8.776448e-3	1.893443e0
0.1	3.792954e0	3.803834e0	2.425309e0	3.807681e0	1.472720e-2	3.847755e-3	1.382372e0
0.2	3.291519e0	3.296986e0	2.263145e0	3.298843e0	7.323670e-3	1.857307e-3	1.035698e0
0.3	2.904280e0	2.907066e0	2.116890e0	2.908122e0	3.842863e-3	1.05686e-3	7.912328e-1
0.4	2.596040e0	2.597496e0	1.984427e0	2.598145e0	2.105239e-3	6.490281e-4	6.137180e-1
0.5	2.344600e0	2.345375e0	1.863970e0	2.345795e0	1.194614e-3	4.200984e-4	4.818250e-1
0.6	2.135311e0	2.135693e0	1.754030e0	2.136009e0	6.978694e-4	3.162470e-4	3.819792e-1
0.7	1.958132e0	1.958317e0	1.653308e0	1.958549e0	4.176625e-4	2.325667e-4	3.052415e-1
0.8	1.805960e0	1.806030e0	1.560724e0	1.806214e0	2.550668e-4	1.843243e-4	2.454903e-1
0.9	1.673631e0	1.673640e0	1.475339e0	1.673789e0	1.584210e-4	1.496930e-4	1.984506e-1
1	1.557308e0	1.557406e0	1.396394e0	1.557407e0	9.978339e-5	1.724654e-6	1.610137e-1

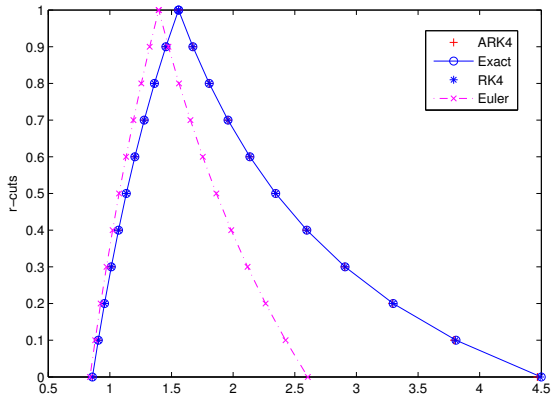


Fig. 1. $h = 0.1$

providing the same order of local accuracy because they reuse function evaluations from previous steps of integration. We introduced the fuzzy ARK method of order four for the approximation solution of a fuzzy initial value problem and illustrated it by a numerical example. Numerical example show that the accuracy-optimized fuzzy ARK4 presented here are superior to fuzzy RK methods that computationally cost the same in terms of the number of function evaluations. Also, the fuzzy ARK methods are comparable to fuzzy RK methods that have the same order of local accuracy.

References

1. A. Ahmadian, M. Suleiman, S. Salahshour, D. Baleanu, A Jacobi operational matrix for solving fuzzy linear fractional differential equation, Adv. Difference Equ. 2013 (2013):104.
2. S. Abbasbandy, T. Allahviranloo, Numerical solution of fuzzy differential equation by Runge-Kutta method, Nonlinear Studies 11 (1) (2004) 117-129.
3. S. Abbasbandy, T.Allahviranloo, O. Lopez-Pouso, J. J. Nieto, Numerical methods for fuzzy differential inclusions. Computer and Mathematics With Applications 48/10-11(2004)1633-1641.
4. M.F. Abbod, D.G.Von Keyserlingk, D.A. Linkens, M. Mahfouf, Survey of utilisation of fuzzy technology in medicine and healthcare. Fuzzy Set Syst. 120, (2001) 331-349.
5. T. Allahviranloo, N. Ahmady, E. Ahmady, Numerical solution of fuzzy differential equations by predictor-corrector method. Information Sciences, 177(2007) 1633-1647.
6. T. Allahviranloo, S. Salahshour, Euler method for solving hybrid fuzzy differential equation, Soft Comput. J,15 (2011) 1247-1253.
7. T. Allahviranloo, S. Salahshour, M. Khezerloo, Maximal and minimal symmetric solutions of fully fuzzy linear systems, J. Comput. Appl. Math. 235 (2011) 4652-4662.
8. T. Allahviranloo, Z. Gouyandeh, A. Armand, Fuzzy fractional differential equations under generalized fuzzy Caputo derivative, J. of Intelligent and Fuzzy Systems, DOI 10.3233/IFS-130831, (2013), In press.
9. M. Amirfakhrian, An iterative Gauss-Newton method to solve an algebraic fuzzy equation with crisp coefficients, J. of Intelligent and Fuzzy Systems, 22 (2011) 207-216.
10. B. Bede, I.J. Rudas, A.L. Bencsik, First order linear fuzzy differential equations under generalized differentiability, Information Sciences 177 (2007)1648-1662.
11. B. Bede, Note on "Numerical solutions of fuzzy differential equations by predictor corrector method", Information Sciences 178 (2008) 1917-1922.
12. Casasnovas, J., Rossell, F., 2005. Averaging fuzzy biopolymers. Fuzzy Set Syst. 152, 139-158.
13. M. Gue, R. Li, Implusive functional differential inclusions and fuzzy population models. Fuzzy Sets and Systems, 138 (2003) 601-615.
14. D. Dubios, H.Prade, Towards fuzzy differential calculus-part3. Fuzzy Sets and Systems 8 (1982) 225-234.
15. S. Effati, M. Pakdaman, Artifficial neural network approach for solving fuzzy differential equations, Inf. Sci. 180 (2010) 1434-1457.
16. M. Friedman, M.Ming, A. Kandel, Numerical solution of fuzzy differential and integral equations. Fuzzy Sets and System 106 (1999) 35-48.
17. B. Ghazanfari, A. Shakerami,Numerical solutions of fuzzy differential equations by extended RungeKutta-like formulae of order 4, Fuzzy Sets and Systems,189 (2011) 74-91.
18. R. Goetschel, W. Voxman, Elementery calculus. Fuzzy Sets and Systems 18 (1986) 31-43.
19. D. Goeken, O. Johnson, Runge-Kutta with higher derivative approximations, Appl. Numer. Math. 39 (2000) 249-257.
20. Z. Jackiewicz and S. Tracogna, A general class of two-step Runge-Kutta methods for ordinary differential equations, SIAM Journal on Numerical Analysis, 32(1995) 1390-1427.
21. O. Kaleva, Fuzzy differential equations. Fuzzy Sets and Systems 24 (1987) 301-317.

22. M. Ma, M. Friedman, A. Kandel, Numerical solution of fuzzy differential equations, Fuzzy Sets Syst. 105 (1999) 133-138.
23. MS. El Naschie , A review of E-infinite theory and the mass spectrum of high energy particle physics. Chaos, Solitons & Fractals 19 (2004)209-36.
24. MS. El Naschie From experimental quantum optics to quantum gravity via a fuzzy Khler manifold. Chaos, Solitons & Fractals 25 (2005) 969-977.
25. M. Mazandarani , A. Vahidian Kamyad , Modified fractional Euler method for solving Fuzzy Fractional Initial Value Problem, Commun. Nonlinear Sci. Numer. Simulat. 18 (2013) 12-21.
26. M. Matinfar, M. Ghanbari, R. Nuraei, Numerical solution of linear fuzzy Volterra integro-differential equations by variational iteration method, J. of Intelligent and Fuzzy Systems, 24 (2013) 575-586.
27. J.J. Nieto, A. Khastan, K. Ivaz, Numerical solution of fuzzy differential equations under generalized differentiability, Nonlinear Analysis: Hybrid Systems vol.3, (2009), 700-707.
28. S.Ch. Palligkinis, G. Papageorgiou, I.Th. Famelis, Runge-Kutta methods for fuzzy differential equations, Appl. Math. Comput. 209, (2009) 97-105.
29. N. Parandin, Numerical solution of fuzzy differential equations of nth-order by RungeKutta method, Neural Comput & Applic 21 (2012) 347-355.
30. S. Pederson, M. Sambandham, The Runge-Kutta method for hybrid fuzzy differential equations, Nonlinear Anal. Hybrid Syst. 2 (2008) 626-634.
31. P. Phohomsiri, F. E. Udwadia, Acceleration of Runge-Kutta integration schemes, Discrete Dynamics in Nature and Society, vol. 2004, no. 2 (2004) 307-314.
32. M. L. Puri, D. Ralescu, Differential for fuzzy function. J. Math. Anal. Appl. 91 (1983) 552-558.
33. A. Karimi Dizicheha, S. Salahshour, F. Ismail, A note on "Numerical solutions of fuzzy differential equations by extended RungeKutta-like formulae of order 4", Fuzzy Sets Syst. (2013), In press.
34. S. Salahshour, T. Allahviranloo, S. Abbasbandy, D. Baleanu , Existence and uniqueness results for fractional differential equations with uncertainty, Advances in Difference Equations (2012), 2012:112.
35. S.Salahshour, T. Allahviranloo, S. Abbasbandy, Solving fuzzy fractional differential equations by fuzzy Laplace transforms,Commun. Nonlinear. Sci. Numer. Simulat. 17 (2012) 1372-1381.
36. S. Seikkala, On the fuzzy initial value problem. Fuzzy Sets and Systems 24 (1987) 319-330.
37. F. E. Udwadia, A. Farahani, Accelerated Runge-Kutta Methods,Discrete Dynamics in Nature and Society, Vol. 2008,Article ID 790619,(2008).
38. X. Wu, J. Xia, Extended Runge-Kutta-like formulae, Appl. Numer. Math. 56 (2006) 1584-1605.
39. H. Zhang, X. Liao, J. Yu, Fuzzy modeling and synchronization of hyperchaotic systems, Chaos, Solitons & Fractals, 26 (2005), 835-843.

List of IASC-ARS and Contributed Talks

Nonlinear Surface Regression with Sufficient Dimension Reduction, *Takuma Yoshida*

Efficient Solution for Nonlinear Dynamic Estimation Problem with Model-reality Differences, *Sie Long Kek, Kim Gaik Tay and Kuan Chin Chua*

FNPR: Using Priority Rules with Fuzzy Serious Queues, *G. Geetharamani and J. Arun Pandian*

An Approximation Solution of Fuzzy Differential Equations Using a New Two Step RK Method, *Ali Ahmadian, Mohamed Suleiman, Norazak Senu and Soheil Salahshour*

A Study of Customer Classification Model Using Time Series Pattern, *Jung Jin Lee and Ji-Hye Baek*

A Comparative Study of Outlier Detection Methods in Multiple Linear Regression Models, *Yung-Seop Lee, Hee-Kyung Kim and ChunGun Park*

Feature Selection Using Conditional Mutual Information, *Donguk Kim and Chi Kyung Ahn*

Fuzzy GCO Based Web Page Retrieval, *Gomathi. C and V.Rajamani*

***L*1-Consistent Adaptive Multivariate Histograms from a Randomized Queue Prioritized for Statistically Equivalent Blocks,** *Gloria Teng, Jennifer Harlow and Raazesh Sainudiin*

Semiparametric Inference Based on Weighted Estimating Equations for Additive Hazards Model with Covariates Missing at Random, *Xiaolin Chen and Yunquan Song*

ICMSFM2014 Committees

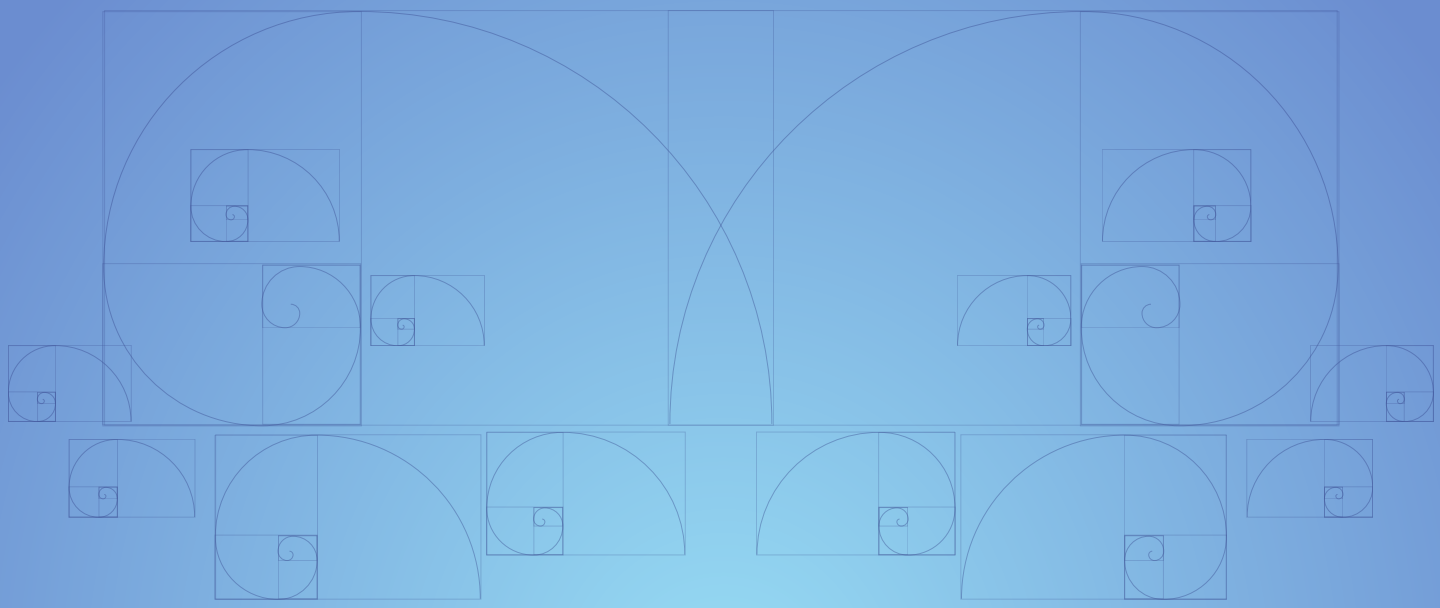
International Advisory Committee

Prof. Dr. Hailiang YANG (University of Hong Kong)
Prof. Dr. Michael Kwok-Po NG (Hong Kong Baptist University)
Prof. Dr. Chun-Houh CHEN (Academia Sinica, Taiwan)
Prof. Dr. Junji NAKANO (The Institute of Statistical Mathematics, Japan)

Local Organising Committee

Chair: Dr. Goh Yong Kheng
Secretary: Dr. Koh Siew Khew
Treasurer: Dr. Chang Yun Fah
Members:

Ms. Amirtha Sangeetha Ganesan
Dr. Chen Huey Voon
Mr. Chew Chun Yong
Dr. Chua Kuan Chin
Mr. Chuah Hock Lung
Dr. Lee Min Cherng
Dr. Liew Kian Wah
Ms. Ng Wei Shean
Dr. Tan Choon Peng
Dr. Yeo Heng Giap Ivan
Dr. Wong Voon Hee



ICMSFM 2014

ISBN 978-967-12788-3-3