IWMS-2015 Souvenir Booklet

PREFACE

During February 2015 the International Organising Committee of IWMS-2015 was encouraged to consider the publication of a Pre-Workshop Proceedings. At that stage we had not made any final decisions regarding the publishingof a formal refereed proceedings as we had not been able to reach any commitments with a publisher or a journal. However, we wanted to give participants the opportunity to present either an extended abstract or a formal paper. With the pressure of time we were conscious that a formal refereeing process prior to the beginning of the Workshop was not going to be possible.

The International Organising Committee was very grateful for the offer of financial support from Professor Kai-Tai Fang to assist us in our quest. He also arranged with Associate Professor Yong-Dao Zhou to provide technical support to undertake such an exercise. As a consequence it was decided to proceed with the publication of a Pre-Workshop non-refereed Proceedings. Following the offer from the journal *Special Matrices* to publish a formal refereed Proceedings for the International Workshop, we decided to name this publication as the *IWMS-2015 Souvenir Booklet*.

Participants of the Workshop were invited to submit a paper or an extended abstract on their intended presentation. We did not make it mandatory to make a submission. The papers in this booklet have been reproduced exactly as received from the authors (other than to ensure that they conformed with a standard presentation as per a formal template).
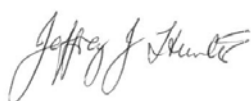
The presentations are presumed to be essentially that which will be given in the Workshop. Although the papers were scrutinised prior to publication for suitability, the papers have not been formally refereed and no claim is made by the Editors or the International Organising Committee as to the accuracy or originality of the contents of the papers.

This *Souvenir Booklet* has not been copyrighted by the International Organizing Committee of the IWMS-2015although some authors have retained the copyright of their material. If no copyright notice is indicated it is presumed that the author(s) have not copyrighted their material and the contents may be freely copied provided the source is cited. Publication in this volume of any paper does not preclude the authors from submitting papers to other publications elsewhere based on their submissions, including the formal Workshop Proceedings.

A recommended citation of articles in this publication is of the form:
Puntanen, S. & Styan, G.P.H. (2015). "Twenty-six years of the International Workshop on Matrices and Statistics (IWMS): 1990-2015". In: Souvenir Booklet of the 24th International Workshop on Matrices and Statistics, (25-28 May 2015), Haikou, Hainan, China. Pages x-y. Ed. Jeffrey J. Hunter.

I would like to express my thanks to those participants who have taken this opportunity to present their work in this publication. The support of Associate Professor Yong-Dao Zhou who has single-handedly compiled the booklet from the submissions is gratefully appreciated. I would also like to thank Dr Simo Puntanen has worked with me in screening the submissions and also to Professor Kai-Tai Fang not only for pushing us to put together this publication but for his generous financial support without which this would not have been possible. I also appreciate the assistance of Professor Chuanzhong Chen in arranging for the design of a cover page for the booklet.

Jeffrey J Hunter
Editor,*IWMS-2015 Souvenir Booklet*

IWMS-2015 Souvenir Booklet

PREFACE

The first workshop in the "International Workshop on Matrices and Statistics" (IWMS) series took place at the University of Tampere in Tampere, Finland, 6 – 8 August 1990, and the IWMS has been held annually since. The IWMS has made great contributions to the promotion of the development of the Matrices and Statistics discipline. I am very excited that the 24th International workshop (2015) on Matrices and Statistics will be held at the Hainan Normal University School of Mathematics and Statistics. The workshop will have a catalytic role in promoting the development of Metrics and Statistics discipline in China, especially it will drive a rapid development of Hainan's related scientific fields. It also provides us an opportunity to make our contributions to the promotion of the development of the Matrices and Statistics discipline.

Since received the invitation to host the 2015 24th IWMS in October, 2013, we have been making every effort for more than one year and finally completed all of the preparatory work. I hereby express my gratitude to Professors Kai-Tai Fang, Jeffrey J. Hunter and Simo Puntanen who had done a great deal of preparatory work for the workshop. I would also like to thank members of the organizing team, in particular Ms. Wang Li, who had put their hard work for the various preparatory work. Hainan Province is a very beautiful area. It has plenty of sunlight, the sky-blue ocean and beautiful beach. School of Mathematics and Statistics, Hainan Normal University has a long history of 66 years. It is our honor to host IWMS-2015 here. We hope that the workshop will be of a great success; the participants will be enjoying the workshop and have a great time. The beautiful season in Hainan is between December and March. It is regretful that the workshop is not able to be held during the best season. The month of May is getting hot in Hainan.

Finally, I would like to say that, after experiencing several twists and turns, the IWMS-2015 Souvenir Booklet becomes quite a bonus for all of us. Professors Jeffrey J. Hunter, Kai-Tai Fang and Simo Puntanen were considering the publication of a Pre-Workshop Proceedings during February 2015. The Pre-Workshop non-refereed Proceedings will not preclude authors from submitting their papers to other journals for publication. Professor Jeffrey J. Hunter contacted De Gruyter Open Ltd. in March, 2015 which agreed to publish the Proceedings of the 24th International Workshop on Matrices and Statistics free of charge. However the Proceedings will not be formally published until after the workshop, possibly one or two year later. It was decided that the IWMS-2015 Souvenir Booklet was printed before the workshop according to the original plan.

Again, I would like to thank Professors Kai-Tai Fang, Jeffrey J. Hunter, Simo Puntanen and Yong-Dao Zhou for their efforts for the production of the Souvenir Booklet, especially Professor Kai-Tai Fang for his financial support.

Chuanzhong Chen

Chair, Local Organizing Committee

Organizing Committees of the 24th IWMS

International Organizing Committee:

Jeffrey J. Hunter (New Zealand, Chair), Simo Puntanen (Finland, Vice-chair), George P.H. Styan (Canada, Honorary Chair), Hans Joachim Werner (Germany), Dietrich von Rosen (Sweden), Julia Volaufova (U.S.A), S. Ejaz Ahmed (Canada), Augustyn Markiewicz (Poland), Goetz Trenkler (Germany)

Scientific Program Committee:

Jeffrey J. Hunter (Chair), Simo Puntanen (Vice-chair), George P.H. Styan, Hans Joachim Werner, Dietrich von Rosen, Julia Volaufova, S. Ejaz Ahmed, Augustyn Markiewicz, Kai-Tai Fang (China), Yonghui Liu (China), Chuanzhong Chen (China), Jianxin Pan (United Kingdom), Shuangzhe Liu (Australia), Kimmo Vehkalahti (Finland), Tsung-I Lin (Taiwan)

Local Organizing Committee:

Chuanzhong Chen (Chair), Li Wang (China)

Contents

3. The Journal *Special Matrices*

A Conversation with Kai-Tai Fang

Agnes W. L. Loie[1]　　　Lucinda Li[2,*]　　　Simo Puntanen[3,†]
George P. H. Styan[4,‡]

[1]Information and Public Relation Office, Hong Kong Baptist University, Hong Kong
[2]BNU-HKBU United International College, Zhuhai, 519085, China
[3]School of Information Sciences, FI-33014 University of Tampere, Finland
[4]Department of Mathematics and Statistics, McGill University, Montréal, Canada H3A 2K6

Abstract　Kai-Tai Fang was born in 1940 in Taizhou city, in Jiangsu Province in eastern China. He received his secondary education at the renowned Yangzhou High School in Jiangsu. In 1957, he entered Peking University to read mathematics and in 1963 he undertook graduate studies at the Institute of Mathematics, Academia Sinica, Beijing. After graduation from the Institute of Mathematics, Academia Sinica, in 1967, Kai-Tai Fang was appointed Assistant Researcher in the Institute of Mathematics, Academia Sinica, a position he held until 1978 when he was promoted to Assistant Professor. In the following year, he was transferred to the Institute of Applied Mathematics, Academia Sinica, shortly after which he was promoted to Associate Professor in October 1980. In 1984, Kai-Tai Fang was appointed Associate Director of the Institute and in July 1986, he became a Professor.

During 1990–2005, Kai-Tai Fang was working in various leading positions in the Hong Kong Baptist University (HKBU), at the Department of Mathematics and at the Statistics Research and Consultancy Centre. From 2006 onwards he has been the Director of the Institute of Statistics and Computational Intelligence, BNU-HKBU United International College, Zhuhai Campus of Beijing Normal University (BNU).

Sections 1–7 of this conversation appeared partly in Loie (2005) and that part was extended by Lucinda Li in 2014. The article of Loie (2005) was published in Fan & Li (2005), a monograph being dedicated to Professor Fang on his 65th birthday in June 2005. Most of the articles in that book were presented at the International Conference on Statistics in Honor of Professor Kai-Tai Fang's 65th Birthday, 20–24 June 2005, Hong Kong. The permit for reproduction of Loie (2005) by the World Scientific Publishing Company is gratefully acknowledged. Section 8 is based on communication between Kai-Tai Fang, Lucinda Li, Simo Puntanen and George P. H. Styan which took place in autumn 2014 and spring 2015.

---

[*]Email: lucindali@uic.edu.hk
[†]Corresponding author. Email: simo.puntanen@uta.fi
[‡]Email: geostyan@gmail.com

PHOTOGRAPH 1: Professor Kai-Tai Fang in Toronto, August 2013.

1 Some Background

Professor Kai-Tai Fang has received several awards for his statistical works, which have had a profound effect on developments in a wide range of fields. In 1982 he was awarded the special prize of the Ministry of Light Industry for the standardization of adult Chinese dress. In 1984, his unified approach to the distribution of restricted occupancy problems won him the second-class prize for Science & Technology of Academia Sinica. Kai-Tai Fang's precision test methodology and determination gained him second-class prize for National Standardization in 1988. His solutions to the distribution of some random military coverage problems also won him another Science & Technology prize from the Academia Sinica the following year.

In 1992, Kai-Tai Fang received a first-class award for Most Excellent Textbook from the State Statistical Bureau of the PRC for his authorship of *Statistical Distributions* (Fang & Xu, 1987). In the same year, his book *Generalized Multivariate Analysis* (Fang & Zhang, 1990) won him a special nationwide award for Most Excellent Book in China by the Government Information and Publication Administration, Beijing. In 1998, the number of citations of Kai-Tai's works reached the ninth highest in the country, according to the Chinese Science Citation Database.

In Hong Kong, Kai-Tai Fang was presented with the President's Award for Outstanding Performance in Scholarly Work by Hong Kong Baptist University in 2001. He was made an honorary member of the Hong Kong Statistical Society in 2002. Two years later in 2004, he was named the Outstanding Author by the Science Press in Beijing. In 2008, together with Professor Yuan Wang, Kai-Tai Fang was presented the highly honorable State Natural Science Award at the Second Level, the highest level award given in that year.

Later in 2012, he was awarded Guangdong Excellent Teacher and Zhuhai Advanced Teacher by the Guangdong Government and Zhuhai Government, respectively, for his teaching excellence. That was further enhanced by the National Statistical Research Award at the First Level conferred to him for his book *Design and Modeling of Experiments*, co-written with Min-Qian Liu and Yong-Dao Zhou in 2013; see Fang, Liu & Zhou (2011). The most recent honor was the 2014 ICSA Distinguished Achievement Award given by the International Chinese Statistical Association (ICSA), for his outstanding achievements in statistical research and teaching, and his contribution to the successful founding of the association.

Kai-Tai Fang has authored and co-authored 22 textbooks and monographs, published more than 250 research papers, 9 papers in mathematical/statistical culture and 45 statistics popularization papers. He has served on numerous editorial boards, including *Acta Mathematicae Applicatae Sinica*, *Statistics & Probability Letters*, *Statistica Sinica*, *Journal of Multivariate Analysis* and *International Statistical Review*. He was the Editor-in-Chief of the book series "Modern Applied Mathematics Methods" in China from 1990 to 2003 and has been the Editor-in-Chief for the book series "Statistics Textbook Series for Higher Education" since 2010.

Kai-Tai Fang has been instrumental in organizing several influential conferences and workshops, both internationally and nationally. He has supervised the research of many graduate students and provided useful advice, encouragement and collaboration for students and their peers around the world. In honor of his scholarly contributions, Kai-Tai Fang has been elected a Fellow of the Institute of Mathematical Statistics in 1992 and a Fellow of the American Statistical Association in 2001. The latter honored him "for (his) outstanding contributions to multivariate analysis, quasi-Monte Carlo methods and design

of experiments; for (his) leadership in statistical education, consultation and administration; and for editorial service". He has also been elected a Member of the International Statistical Institute in 1985.

2   Early Life, Interest in Mathematics: 1951–57

Loie: *When did you start to develop an interest in mathematics and statistics? Did it have anything to do with your family background?*

Fang: Not really. Shortly after I was born, World War II broke out and it was an era of complete chaos. My parents took the seven of us to find refuge in rural villages. After the war, things were still chaotic and a formal education system was not in place. The teachers were not serious about teaching and that gave us a perfect excuse to follow suit with regard to learning. I remember that because of a shortage of space and manpower, two classes of different levels shared the same classroom and the same teachers. That meant that the teacher could only devote half his time to teaching us and the other half to teaching the higher level. We lacked interest and the desire to study hard. We aimed only at a mere pass. It was not until headmaster Konghou Wang stepped into our classroom that I took a positive twist in my learning attitude. It was also then that I began to develop an interest in mathematics.

Loie: *In what way did he inspire you?*

Fang: Every day Mr. Wang would give an extra 30-minutes' tuition for our grade six class and he would come up with a list of questions for us to work on. Those who finished first and got all the answers right would be allowed to leave the classroom to play. The questions he set were far from routine and were in fact pretty interesting. There was one that I remember in particular. The question was a farmer who was selling a basket of eggs. He approached the first family who bought half the basket of eggs plus half an egg. The second family bought half of the remaining total plus half an egg. The third bought the remaining half of the total plus half an egg, then the whole basket of eggs was sold out. How many eggs were there in the basket? It did not take me long to come up with the answer: seven. I was the first to hand in the answer and was instantly allowed to go out and play. Even my elder sister, who was in the same class and ranked first in class, could not get the answer right.

That was the first time in my life that I discovered my strength and competitiveness; it was also the first time I realised I had an edge over my classmates. After that, I was almost always the first to leave the classroom. This self-discovery, coupled with Mr. Wang's recognition, worked miracles in building up my confidence and had a far-reaching impact on my self-esteem.

Loie: *Would you regard Mr. Wang your first mentor?*

Fang: Indeed he was. He was instrumental in stimulating my interest in mathematics, an area that I undertook as my lifelong career.

Loie: *What happened after that?*

Fang: In 1951, I was admitted to Yangzhou High School in Jiangsu Province, one of the most reputable, well-established secondary schools in the whole nation. The school adopted a serious and professional manner and they (the school board) even employed university professors to teach us. The deputy headmaster, for example, was a famous

PHOTOGRAPH 2: Kai-Tai Fang as three years old.

PHOTOGRAPH 3: Kai-Tai Fang (2nd row from top, 5th from left) and other senior high graduates of Yangzhou High School in 1957.

English Professor. The quality of the teachers was exceptional and they attached great emphasis to independent thinking. The school had produced many famous graduates such as Zemin Jiang, Jiaomu Hu who was Mao Zedong's secretary; and more than 10 members of the Chinese Academy of Sciences.

Loie: *Was your interest in mathematics further enhanced here?*

Fang: Yes. I met my second teacher who deepened my passion for mathematics. He was Guangzhao Fang. He adopted an enlightened approach by first asking questions before giving a lecture. This was to inspire us to think. I was always among the first two to answer the questions. His lectures were stimulating and I was captivated by what he had to say. When I was invited by the school to give a talk on how to learn mathematics in the late 1980s, I was so pleased to meet Mr. Fang again and I highlighted his teaching approach in my talk.

Loie: *I heard that Yangzhou High School was renowned for its whole-person education and its emphasis on encouraging students to develop a wide variety of interest as well as nurturing their psychological quality. How did your secondary education benefit your development as a whole?*

Fang: When I entered senior high, I read many books on self-development to boost my psychological quality and I set a number of targets for myself. For example, to increase my perseverance level, I planned a series of target studies for every weekend, a practice that I maintained even when I entered university. Another example is that our school then had a scheme to encourage students to exercise. Those who succeeded in running a certain number of kilometers would be awarded a souvenir. I challenged myself to run every day, even in the severe cold winter climate. All these self-training exercises helped equip me with the determination to overcome future problems, both academic and otherwise; they also gave me the will to succeed. I never give up easily, regardless of the scale of any problem. In addition, our teachers also wanted us to help the poor people.

In the 1950s, I was deeply distressed to see so many people, particularly women, living a hard life because of their low education. In those days, most women were still illiterate. They could not read, write, or take a job. All they could do was to bear children for their husbands and depend on the male members of their families for their whole lives, financially, socially, and emotionally. Many did not even have a name and were only identified by the surnames of their own clan and their husbands'. Hoping to help them make a change, I decided despite my young age to take part in the voluntary service of combating illiteracy among poor people in the rural areas. Being able to write their own names was already something made them proud of and in themselves.

Seeing that people were battered hard by the tough lives, our school was calling for efforts to raise the spirit of the community. I took the initiative to organize the street carnival during the Spring Festivals. I mobilized students and also other supporters in the community to decorate the places, put up stalls, and stage street shows to lighten up the community as well as to broaden the minds and experience of the young.

3 University Studies, Peking University and Academia Sinica: 1957–67

Loie: *Can you tell me something about your university studies?*

Fang: Professors at Peking University had high expectations about their students. Peking University, famous as it was, wanted to do just as well as Moscow University, which then ranked first in a number of areas. Their education strategy was to let the best professors teach first-year students so that the latter would have a solid foundation for their studies. This, I think, was a wise strategy because despite my 10-year stoppage in my studies due to the Cultural Revolution, I still had a firm grasp of mathematical techniques. Because of the keen competition between Peking and Moscow Universities and also among students, all of us were under tremendous pressure. Many of my classmates were filled with a sense of negativism even though they performed exceptionally well in their secondary school days.

Loie: *How did you surpass all the difficulties and keen competition you faced in learning advanced mathematics?*

Fang: While at Peking University, I came across a book *How to Solve It: A New Aspect of Mathematical Method*, authored by George Pólya (1957), who was then a professor at the Swiss Federal Institute (where Albert Einstein graduated), and later at Stanford University. This book embodies a wealth of wisdom on thinking skills. In a nutshell, the book establishes a close-knit link between the specific and general, and advises readers to be general in order to be specific and vice-versa. It also warns readers that it is better not to have a book at all than to believe all that is written in the book. Professor Pólya's book also challenges readers to do something positive to exceed the teachings of books they read. I was deeply moved by Pólya's teachings and I put them into practice. I set high expectations of myself and required myself to look for solutions rather than seeking help from my teachers and classmates. Pólya's book has had a life-long impact on me and I have applied his teachings to my academic studies and research ever since. It never occurred to me that I would have a chance to thank Professor Pólya in person. That chance came when I visited Stanford University in 1982. Professor Kai Lai Chung took me to see him and I told him that to me, he was first and foremost a great educationalist and I trusted that his readers would agree with what I said.
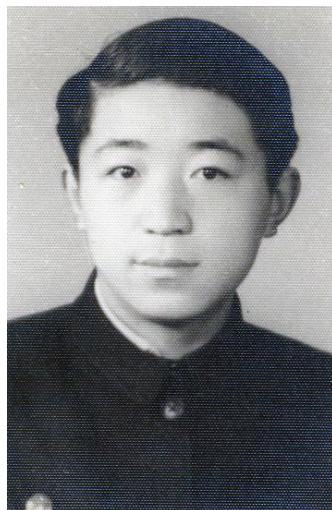
Photograph 4: Kai-Tai Fang during his university days.

Loie: *I understand that you were a student of the renowned Pao-Lu Hsu[1] and became greatly influenced by his supervision.*

Fang: Yes, the next person that impacted me was my supervisor, Professor Pao-Lu Hsu, a UK-educated scholar who laid a solid foundation for multivariate statistical analysis and who had four papers acknowledging his contributions published in the same issue of the prestigious international journal *The Annals of Statistics* in 1980. This was an exceptional treatment by the journal in recognition of his outstanding contributions. Hsu was severely ill in 1962 and was instructed to take full rest by his doctor. Despite his illness, he continued to work full time and take up both research and supervision duties. He required us to study a 50-page book written by a Russian mathematician, a Stalin Award winner, and asked us to try to improve his results and make a report in class. After our presentation, Hsu told students that if they followed the Russian approach, they could only come up with a single dimension. He then showed us how to solve the problem using different approaches and came up with a more powerful answer that catered not only to one-dimensional statistics but also to high-dimensional statistics. This was an eye-opening experience for me.

Hsu's insistence in fulfilling his teaching obligations, despite his weak physical condition, and his dedication to research exerted great influence on my future academic career. In fact, my first paper "The Limiting distribution of linear permutation statistics and its applications" was completed under his supervision. Hsu said to me that there was a gap in a paper originally published in *The Annals of Statistics* and should I be able to identify and fill in that gap, I would be qualified to graduate. Very soon, I was able to identify the gap and fill it in and even discovered that the paper could well be extended. I put in many new angles which produced some interesting results. After two weeks, I handed in my paper and, after reading it, Hsu told me that I could now graduate. He even recommended that *Acta Peking University* should publish my paper, which it willingly accepted. It was however most unfortunate that before my paper was published, a political movement took place and all publications and newsletters came to a halt. Fortunately, the paper was published 19 years later in *Acta Mathematicae Applicatae Sinica*, see Fang (1981).

Loie: *That was an indelibly dark era for mainland residents. How did you survive this period?*

Fang: China then was really shrouded in an intense political climate and people became distant from one another, fearing that any outpouring of genuine feelings would be betrayed, especially if they were about government and policies. Because of this, my years at the university were unhappy—a sharp contrast from my high school days. My dislike for the chaotic political movement in Peking University prompted my decision to pursue further studies at the Institute of Mathematics, Academia Sinica (later changed to Chinese Academy of Sciences) and became the first postgraduate student of Professor Minyi Yue.

Loie: *What did you do in Academia Sinica?*

Fang: My first two years at the Institute of Mathematics were fruitful under a favorable academic ambience. In 1965 I was assigned to An Shan Steel and Iron Co and was forced to turn to application instead of just theory. The engineers there treated us nicely and

---

[1]Pao-Lu Hsu (1910–1970) graduated from Tsinghua University in 1933, majoring in mathematics and then worked at Peking University as a teacher. In 1936–1940 he studied mathematical statistics at University College London, earning his Ph.D. in 1938 and Sc.D. in 1940, with Jerzy Neyman and Egon S. Pearson as advisors. In 1945, he went to the United States, visiting the University of California at Berkeley, Columbia University, and the University of North Carolina at Chapel Hill. In 1947, he returned back to Peking University. Further reference: Chen & Olkin (2012).

PHOTOGRAPH 5: George Pólya sent this photo to Kai-Tai Fang by person in 1982.
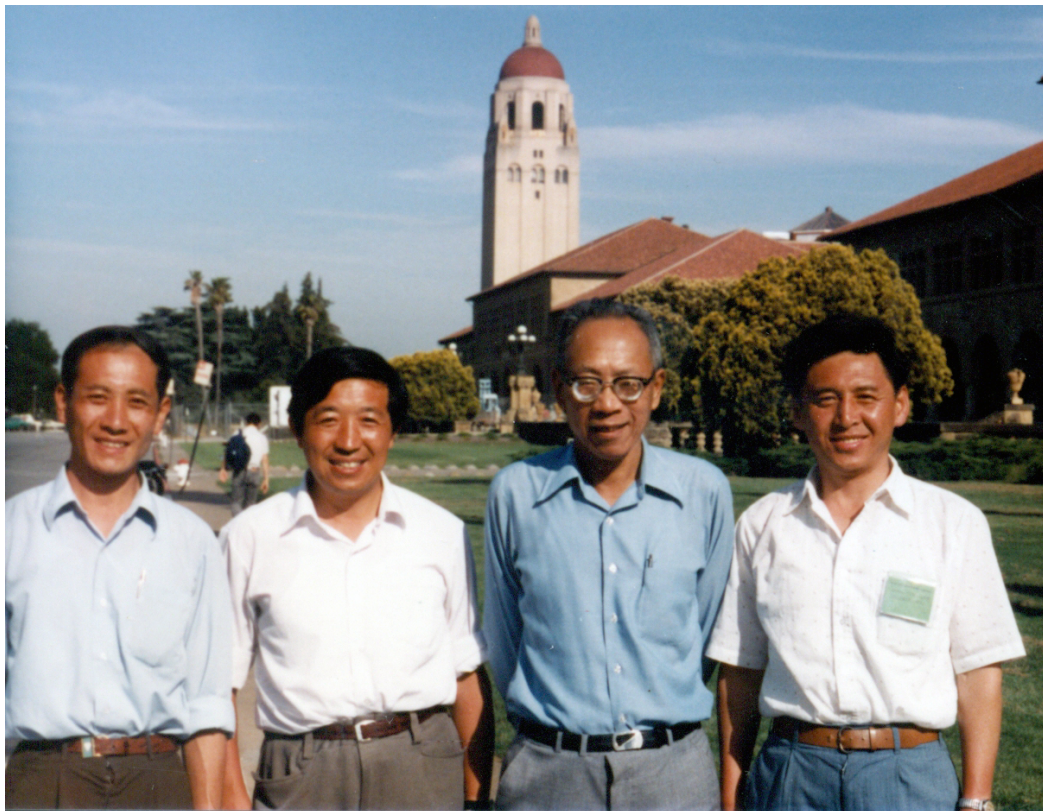
PHOTOGRAPH 6: Kai-Tai Fang and Kai Lai Chung (those in the middle), in Stanford University, 1981.

PHOTOGRAPH 7: Pao-Lu Hsu (front row, 2nd from left) with Kai-Tai Fang (back, 2nd from left) and his other graduate students at the Peking University, 1963.

PHOTOGRAPH 8: Celebration of Minyi Yue's 90th birthday and Yuan Wang's 80th birthday in 2010. Professor Wang (4th from left) and Professor Yue (5th from left) are in the middle.

had a high expectation of us. At that time, I used non-linear regression analysis to analyse the data collected. This period signaled a positive change for me in that I could apply my knowledge to meet the high expectations people bestowed on us. What I learned in Peking University focused merely on theory and did not touch on any applications, which was the weakness of the Russian education system. While I was at An Shan, I was asked to give lectures to the engineers. I covered eight topics in statistics, all of which were published for staff reference, an indication of the high regard in which they held me. My first two years at the Institute were fruitful under a favorable academic ambience. However, this favorable situation did not last long as the political movement took shape the following year.

Loie: *That was a prelude to a political storm, with the Cultural Revolution just round the corner?*

Fang: Indeed. The subsequent years were a complete waste of time. All of us were deprived of the right and opportunity to pursue our studies and research. In 1965 and 1966, I was sent to the villages as a laborer. The following two years came the Cultural Revolution and the political movement lasted until 1976. We were all under tremendous pressure and stress and were uncertain about tomorrow. What was important then was that we could survive today and we did not even dare to think about tomorrow.

As much as I hated this period, it did strengthen the psychological side of me, enabling me to face each and every bitter challenge with an unyielding manner. Like Professor Grace Yang of the University of Maryland once said me: "You have recovered your 10 lost years." She gave me a lot of encouragement on a number of issues.

4   Orthogonal Design, Uniform Design

Loie: *When did you learn the orthogonal design and start conducting experiments with this method?*

Fang: During the early 1970s, staff from Peking University and the Institute of Mathematics, Academia Sinica, attempted to promote and apply orthogonal design to the industrial sector. In 1972, I had the opportunity to go to the Tsingtao Beer Factory and other factories. I supervised the engineers there to apply orthogonal design to industrial experiments. It was a precious experience for me to witness the substantial potential of applying orthogonal design to practical industrial use. However, I also detected the considerable difficulties faced by the engineers in understanding statistical methods, especially in calculating the ANOVA Table without the help of computers or calculators. I came to realize the need for statisticians to simplify the complicated statistical theories and methods, and later created "Visualization Analysis" for analytical use on experiment data. Very soon this method was commonly used on the mainland, triggering a great sense of encouragement and inspiration on my part.

Loie: *There were quite a number of contributions that you made to the orthogonal design. What were they?*

Fang: During my process of promoting the common use of orthogonal design, I encountered quite a number of complicated multi-factor and non-linear issues. The engineers were unable to identify a satisfactory parameter value combination for a long time. An example was a porcelain insulator factory in Nanjing.

The factory had a team of staff assigned to conduct experiments continually to identify a satisfactory parameter value combination. Although they had achieved much in their

PHOTOGRAPH 9: Kai-Tai Fang (front, 2nd from right), other postgraduate students and new staff of Academia Sinica were sent to work in rural villages during the Cultural Revolution.

PHOTOGRAPH 10: Kai-Tai Fang, S. R. Wang, Grace Yang and Lucien LeCam with his wife, in the Summer Palace, Beijing, 1987.

experiments, they still failed to get one of the responses to meet the requirement, thus failing to deliver the glass insulator products. (At that time, the factory received a large number of orders for glass insulators but was unable to deliver the products.) In view of the complexity of the issue, I adhered to the principle of "big net catching big fish". I

conducted 25 experiments and arranged six 5-level factors by an orthogonal design.

From a statistical point of view, the experiment model was non-estimable and was therefore incorrect. However, in those 25 experiments, one had all the responses fulfilling the requirements. That was great news to the factory in-charge. Should one liken the outcome to winning the US lottery or was it significant? In fact, using orthogonal design to conduct 25 experiments actually represented 15,625 experiments, thus greatly increasing the likelihood of attaining an ideal technical/manufacturing condition. In my opinion, the power of fractional factorial design was that the experimental points have a good representation. Since then, I have used the same strategy to solve many of the "lasting, major and difficult" problems of the factories. This success has also injected in me the necessary courage to initiate the Uniform Design theory and method.

Loie: *Can you tell me how you came up with uniform design, an approach so well known in the statistical field?*

Fang: Using the "big net" approach to get the best level-combination of the factors, I mustered the courage to create another new method—the uniform design approach. In 1970, I came across several occasions which called for the application of a more powerful statistical approach. For example, in 1975, a factory manufacturing steel for automobiles wanted to come up with a nationwide standard that needed the numerical calculation of many five-dimension integrals. At a time when computers were much less powerful, it was almost impossible to do so. Luogeng Hua and Yuan Wang came up with a method to solve high dimension integral problems. Professor Wang taught me how to use their method and I realized five years later that the method might be applicable to experimental design.

In 1978, there were three major missile-related projects covering land, sea and aerospace. A problem-solving approach was needed to tackle all the projects. Again, it was highly challenging. I had to come up with a new method, one that could approximate a complicated system by a simple method with required accuracy. The great challenge was a motivating force to me.

I collaborated with Yuan Wang and we worked out the uniform design. This method made possible the calculation of an accurate answer in .00001 seconds with the required accuracy. It was both time- and cost-saving and provided a valuable alternative since it could also be used in computer experiments as well as laboratory experiments. Several years after the uniform design theory was proposed, I discovered that it was being used extensively in the mainland. Not only was it used for military purposes, it was also adopted by and for civilians.

Loie: *There was some opposition to the uniform design in the mainland, why?*

Fang: In the cases mentioned above, we promoted the use of uniform design in computer experiments which was a completely new concept to many at that time. While most traditional experimental designs used ANOVA as the main tool for data analysis, we proposed the use of regression analysis for modelling. During the 1970s, especially just after the Cultural Revolution, many scholars in China were still adhering to the modeling of the traditional experimental designs. But our uniform design approach, although not quite supported by the academics, was greatly welcomed by the engineers.

5 Overseas: 1980–82

Loie: *The 1980s marked a significant chapter in your life as you started to play a key role in the global scene. Would you consider that as an epoch-making era for you?*

PHOTOGRAPH 11: Yuan Wang and Kai-Tai Fang. This photo was taken by CCTV in 1994.

Fang: In a way, yes. In 1980 when I had the opportunity for an overseas visit, I did not plan to go initially in view of my wife's illness. Tingmui suffered from asthma and, of course, my two daughters were quite young. My standard of English then was low and basically I could not communicate in the language. Besides, we were blocked from the outside world during the 10 years of political movement and we were ignorant of what was happening around the world.

After much deliberation, I finally visited Yale University with one focus in mind—to learn as much as I could within a limited period of time. The eight months I spent at Yale was, on one hand, tremendously difficult as I was learning from scratch and yet, on the other hand, it was fulfilling as I succeeded in coming up with several papers, one of which was published in *Encyclopedia of Statistical Sciences*. While at Yale, Professor Samuel Kotz sent me one of his books which inspired me to come up with a paper— my first paper, Fang (1982), to be published in a Western journal (*Journal of Applied Probability*)—and therefore a series of collaborations started between us.

Then in 1981 and 1982, I went to Stanford University for a visit and there I met Professor T. W. Anderson, a Princeton graduate who later taught at Columbia and Stanford Universities. He asked me to read two papers and then we had some idea about generalized multivariate analysis. Many statisticians wanted to generalize multivariate analysis to non-normal populations, but they failed. The combination of the essence of Pólya's and Hsu's teaching—that you need to seek different approaches and you need a powerful tool in order to exceed the work of others—prompted me to undertake a search for such a powerful tool: the $\underline{\underline{d}}$ operator. I systematically developed this tool and wrote two papers on the topic. After publication, Professor Anderson identified a vast potential for further development and subsequently selected a topic for extended research. He engaged more of his students in this research and on my return to the mainland, I also brought many students into the activity. This collaborative research of ours continued for eight years, during which more than 50 articles, two monographs and a collection of papers were published. This was a big leap forward compared to the situation prior to my overseas visit when it was a norm for one project to generate only one or two papers. I realized that for

PHOTOGRAPH 12: Visiting Samuel Kotz and his wife, 1988.

PHOTOGRAPH 13: T. W. Anderson and Kai Tai Fang in Hong Kong, 1997.

people of high standards, the choice of topics was of paramount importance.

During my visit to Stanford University, Professor Ingram Olkin organized a small seminar group on "multivariate multiple comparisons" which met every week. The participants included T. W. Anderson, Mary Ellen Bock, Zhongguo Cheng and me.

Our collaboration brought me high international reputation. I was invited by *Encyclopedia* and other journals to act as referee, author or Associate Editor. I was deeply appreciative of Professor Anderson for his identification of the potential and productivity of the generalized multivariate analysis as a research topic.

Loie: *Did you have the chance to visit other universities?*

Fang: Yes, in October 1982, I was invited to give talks at several universities, including Princeton, Yale and Columbia, and the universities of Pennsylvania, Maryland, Rutgers and George Washington. These visits were useful to my work, especially after I was promoted to Associate Director of the Institute of Applied Mathematics, Academia Sinica. Then in 1985–86, upon Professor Ingram Olkin's recommendation, I taught two subjects in the Swiss Federal Institute as a Guest Professor. It posed another challenge for me as it was the first time I had to teach in English. During my stay there, I seized the opportunity to visit various prestigious institutions including Oxford, Cambridge, University College London and Imperial College London, etc. The visits helped me to network which has proved useful throughout my career. In fact, it was during these overseas trips that I met Professors Colin White, C. Radhakrishna Rao, Norman Lloyd Johnson, D. R. Cox, A. P.

PHOTOGRAPH 14: Ingram Olkin and Kai-Tai Fang in Beijing, 1984.

Dawid, Yung Liang Tong, Y. S. Chow, George Tiao, Michael Stephens, Dennis K. J. Lin, and Agus Sudjianto. I am glad that our paths crossed as they were all inspirational to me in one way or another.

Loie: *I heard that you had a particular connection with the University of North Carolina (UNC) at Chapel Hill. How is that so?*

Fang: Well, in 1988 I first went to the University of North Carolina at Chapel Hill—where my former supervisor Pao-Lu Hsu once taught as a visiting professor between 1945 and 1947—and taught generalized multivariate analysis. In that year, Professor Norman Lloyd Johnson was in his last semester before his retirement, and he left me his teaching notes on multivariate statistical analysis. It is a unique gift to me. Since then, I have been linked to the UNC in different ways. In addition to Pao-Lu Hsu, my former student Jianqing Fan, a recipient of the COPSS Award and now professor at Princeton, was also a faculty member of the UNC. Later, my daughter Yan Fang completed her Master's degree in biostatistics at UNC and my son-in-law, Runze Li obtained his doctoral degree in statistics at the same university. Runze was my former student in Beijing. Both of Jianqing and Runze have been Editors-in-Chiefs of *The Annals of Statistics*.

PHOTOGRAPH 15: Runze Li and Yan Fang, graduate ceremony in University of North Carolina (UNC) at Chapel Hill, 2000.

PHOTOGRAPH 16: Kai-Tai Fang, Norman Lloyd Johnson and Gordon Simons in UNC, 1988.

6 Back to China: 1982

Loie: *What happened when you returned to the Academia Sinica in 1982?*

Fang: I returned to Beijing in October 1982 after 2 years visiting in USA (Yale and Stanford Universities). In the spring of 1984, I was appointed the Deputy Director of Institute of Applied Mathematics, Academia Sinica. In 1985 I was approved to become Doctoral Supervisor in Applied Mathematics including Statistics by the National Committee. There were few professors who could be doctoral supervisors in China then. There were about 6 doctoral supervisors in statistics around the country in 1985. Many young scholars wanted to be my postgraduate students or to attend our seminars. Most of those young scholars eventually went overseas after my supervision. Jianqing Fan (to Princeton), Gang Li (to UCLA), and Jianxin Pan (to the University of Manchester) were among this group.

Loie: *What were your main contributions to the development of statistics in China in the 1980s?*

Fang: During that period of time, I was devoted to the promotion of research on and the use of multivariate statistical analysis. I organized several national conferences on the topic. After a lot of hard work with Professor George Tiao and others, I proposed and managed to have the first Sino-American statistical meeting held in Beijing in 1987. This was the first time that so many statisticians from USA came to visit China.

Loie: *As the Deputy Director of the Institute of Applied Mathematics, Academic Sinica, what were your special contributions to the Institute?*

Fang: I was very active in encouraging research staff to expand their research porfolio by seeking external projects. I suggested, and was accepted by the Director of the Institute to change the research funding system from center-based to project-based. That succeeded in raising the initiatives of the staff, as well as the income of the Institute and the researchers. In addition, I established the first technical report series in the Institute of

Applied Mathematics. To help save printing cost and enhance presentation of the reports, I bought an HP laser printer which was not available in China then, and carried it all the way from California to Beijing for the Institute. I even undertook to train the Institute staff on how to use LaTeX.

Loie: *Did you still have time for your own research?*

Fang: Meanwhile, I was also productive in my own work. I produced two monographs, *Symmetric Multivariate and Related Distributions* (Fang, Kotz & Ng, 1990) and *Generalized Multivariate Analysis* (Fang & Zhang, 1990), and two reference books, *Statistical Distributions* (Fang & Xu, 1987) and *Applied Regression Analysis* (Fang, Quan & Chen, 1988).

7 Hong Kong Baptist University: 1990–2005

Loie: *When did you move to Hong Kong and why?*

Fang: I always have had more interest in research than administration. When I knew that the Academia Sinica had a plan to promote me from the Deputy Director to Director of the Institute of Applied Mathematics, I decided to take leave from the mainland for a few years. I joined the Hong Kong Baptist College which happened to be in the process of planning to become a university. She needed scholars who excelled in research. Due to the need of the College and its continuing development after successful upgrading, I decided to stay there permanently.

Loie: *How would you describe your academic and research pursuits at the Hong Kong Baptist University (HKBU)?*

Fang: With the encouragement of Kai-Wang Ng, I moved to the Hong Kong Baptist University (then College) in 1990. My years at the HKBU were the happiest and smoothest of my academic life. Many of my important papers were all published during this period in international journals, gaining me global exposure and reputation. The academic ambience here was stimulating and the congeniality among colleagues was notable. The support I gained from the top administrators was keen, facilitating my collaboration with overseas and mainland academics. The number of PhD and MPhil students supervised by me was comparatively large. All these factors created an environment conducive to both academic and research developments. I must thank our former Presidents Dr. Daniel C. W. Tse and Professor C. F. Ng, as well as Academic Vice-President Professor Herbert H. Tsang for their encouragement and support. It was at the HKBU that I reaped the most fruitful harvest in terms of academic and research pursuits. It was also here that I received various honors and awards.

Loie: *You have developed or further developed quite a number of methods during your academic career at the HKBU. Can you tell me more about them?*

Fang: The Quasi-Monte Carlo method was one that was expanded and further developed here in Hong Kong. We started applying the method in Beijing to develop the uniform design. In Hong Kong we continued to apply the Quasi-Monte Carlo method to experimental design, and also to a variety of statistical problems, including simulation and statistical inference. In 1994, I co-authored a book, *Number-Theoretic Methods in Statistics*, with Yuan Wang to further promote the method and its applications; see Fang & Wang (1994).

我所部分师生与来访外宾合影。

PHOTOGRAPH 17: Samuel Kotz (left 5) and Yung Liang Tong (left 6) visited the Institute of Applied Mathematics, Beijing, 1987.

PHOTOGRAPH 18: Kai-Tai Fang (fifth from left) with C. F. Ng (centre), then Dean of Science and now President and Vice-Chancellor, and departmental colleagues at the Hong Kong Baptist University, 1993.

It was also in the same year that I became President of the newly-established Uniform Design Association of China, a post which I held for 10 years until 2003.

There was a hiccup in the development of uniform design as the uniformity was categorized as a geometric criterion instead of a statistical one. This criticism provided an excuse for people to reject our papers. In view of this, I decided to spend more time on the uniform design theory. In 1992, a participant from North Carolina State University attended my conference in Hong Kong and told me that it was a pity that the Western community did not know about the uniform design. This remark was of great encouragement to me and inspired me to work on more solid basic theory for the uniform design method.

Loie: *How did you overcome the technical difficulties of promoting the uniform design theory and method?*

Fang: I indeed encountered a number of problems in tackling the issue. First, I was not familiar with the typical tools employed by the Quasi-Monte Carlo method as they were invented by mathematicians such as Luogeng Hua and Yuan Wang. Besides, I am a statistician and not a pure mathematician. One way to solve the problem was for me to learn to use the tools but it would not be effective in light of my age and time.

Second, the uniform design theory in itself was difficult. I therefore spent the first four years, i.e., from 1992–96, working on it. It was like an exploration for me and I made slow progress. It was necessary for me to identify the tools that suited me—on which I spent an enormous amount of time. As the Chinese saying goes: "*It is of little use for peonies to blossom only by themselves. They need green leaves to bloom with them.*" I was stimulated to focus more of my time on the uniform design. In fact, 90 per cent of my academic pursuits has focused on uniform design since then. My collaboration with several scholars led to the discovery of a breakthrough that suited me. I came up with the conjecture that most orthogonal designs were uniform in a certain sense. If that was the case, we could link up orthogonal design with uniform design and obtain a vast development potential for uniform design.

I spent one year with Peter Winker of Germany, a doctoral student then and a professor now, to prove with the computer that my conjecture was true. It was exciting to find that my conjecture was true in that many existing orthogonal designs were also uniform designs. Our result was based on the measure of uniformity proposed by my colleague, Fred J. Hickernell. This discovery was of mutual benefit to both Fred and myself. For him, his proposed measure of uniformity was initially not appreciated by many but his measure became necessary in uniform design. For me, his measure of uniformity helped prove that many existing orthogonal designs were uniform designs. With this, we still had one step to complete—to come up with a mathematical proof.

To achieve this, I invited Rahul Mukerjee, Professor of the Indian Institute of Management in Calcutta, to collaborate with me. Rahul is a worldwide expert in experimental design. After two weeks, he told me that my conjecture was not always true, even for a two-level factorial case. However, he came up with an excellent result—that we could link up uniformity with orthogonality. A criterion "aberration" was used to measure orthogonal design. For uniform design, the centered discrepancy was used to assess uniform design. With this, Rahul and I established an analytic relationship between centered discrepancy and aberration. This discovery was immediately published in a top journal, *Biometrika*, see Fang & Mukerjee (1999). It opened up an entirely new area that linked up uniform design and factorial design, an area in which I collaborated with Chang-Xing Ma and others, and which resulted in the publication of more than 20 papers since 1999–2000.

Then in 2000, I began collaboration with S. G. Ge from Suzhou University and Min-Qian Liu from Nankai University to link up combinatorial design and uniform design.

PHOTOGRAPH 19: Peter Winker and Kai-Tai Fang in Germany, 2004.

PHOTOGRAPH 20: Jianxing Yin, Min-Qian Liu, Rahul Mukerjee, Kai-Tai Fang, Hong Ma and Yizhen Liang, 2001.

Photograph 21: Chang-Xing Ma, Kai-Tai Fang and Dennis K. J. Lin, 2010.

Photograph 22: Kai-Tai Fang and Fred J. Hickernell, 2004.

Another new direction was established and this also led to the publication of many research papers.

The breakthrough we achieved in relation to uniform design won international recognition. The *Encyclopedia of Statistics Science (Second Edition)* has chosen uniform design as an entry, see Fang (2006), while the *Handbook of Statistics (Volume 22)* already included uniform design as a chapter; Fang & Lin (2003). Springer's *Handbook of Engineering Statistics* invited us to write a chapter on uniform design for engineers and this too will

soon be published; Fang & Chan (2006).

Uniform design also won national acclaim. The Uniform Design Association of China, for example, reflected the need to conduct national conferences, training courses, workshops and other activities to meet the calls to promote the applications of uniform design.

Loie: *Are there applications of uniform design in real case studies?*

Fang: Application-wise, there were numerous successful applications of uniform design in China. With the keyword "uniform design", you can call up (on the Internet) hundreds of published case studies. The application of uniform design by Ford Motor Co in the USA is exemplary of the applicability of this method. At Ford, Dr. Agus Sudjianto introduced to us that the technique had become a critical enabler for them to execute "Design for Six Sigma" to support new product development, in particular, automotive engine design. I was told that today, computer experiments using uniform design have become standard practices at Ford Motor Company to support early stage of product design before hardware is available. As a result, Runze Li, Agus Sudjianto and myself decided to publish a textbook/monograph *Design and Modeling for Computer Experiments*, Fang, Li & Runze (2005), where many case studies were from the real cases in Ford Motor Co. In 2001 the 50th Gordon Research Conference: the Statistics in Chemistry & Chemical Engineering invited the topic "Uniform design for simulation experiments"as one of the nine topics, and each topic was given 3.5 hours for introduction and discussion. Professors Dennis K. J. Lin, Yizhen Liang (a chemist) and myself formed a panel for this topic.

Loie: *Apart from research, you also spent much time on statistical education. Can you elaborate on that?*

Fang: To promote statistical education, I wrote international monographs, textbooks for undergraduate and postgraduate studies as well as textbooks for engineers in the mainland and for various targets on different occasions. I was also willing to take up guest professorships. Often, many of the participants who are now professors and industry and university leaders came to me and said they had listened to my lectures on various occasions or studied my textbooks when they were students. I found that quite rewarding.

I have been told that one of my textbooks, *An Introduction to Multivariate Analysis*, Zhang & Fang (1982), has been assigned as a compulsory textbook for Analytical Chemistry students. That was beyond my expectations. Although promoting statistical education has increased both my exposure and reputation as a by-product, what I found most gratifying and encouraging was the fact that I can make some contributions to my country.

I also understand that some of my textbooks and articles have been published in layman's terms for different professions so that the non-statistical sectors could also conduct research with statistics including uniform design. For example, the application of multivariate statistics to devise a common dress standard in 1976–78 was successful and the National Standards Bureau invited me to write a series of lectures. The published articles were collected as a book entitled Statistics and Standardization. Another example was the An Shan Steel and Iron Co which I mentioned earlier.

Loie: *I know that you have organized or co-organized quite a number of significant conferences, both international and national. Organizing conferences of this scale requires an enormous amount of time and attention. How did you find time to organize these activities amid your already hectic schedule?*

Fang: On a national basis, China had fallen behind for at least two decades because of the political turbulence. It was necessary to bring it to par with our counterparts overseas. Collaboration was useful in this regard. I took part in organizing seven nationwide multivariate analysis conferences since 1979, with one part of it theoretical and the other on applications, to provide a platform for establishing collaboration between the two. To attract international collaboration, I organized the Sino-American Statistical Meeting in 1987, which attracted more than 200 participants. In Hong Kong, I organized

- the International Symposium on Multivariate Analysis and Their Applications in 1992,
- the International Workshop on Quasi-Monte Carlo Methods and Their Applications in 1995,
- the International Symposium on Contemporary Multivariate Analysis and Its Applications in 1997,
- the Symposium on Theory of Uniform Design and Its Applications in 1999,
- the 4th Monte Carlo and Quasi-Monte Carlo Conference in Scientific Computing in 2000 and
- the Symposium on the Uniform Experimental Design in 2003.

Loie: *You have been accorded high international reputation as a result of your contributions to the global statistical field. The honors and awards bestowed on you include Fellow of Institute of Mathematical Statistics (1993), Fellow of American Statistical Association (2001) and numerous awards for your outstanding contributions to multivariate analysis, quasi Monte Carlo methods, design of experiments, and for your leadership in statistical education, consultation and administration as well as for your editorial service. Despite all these prestigious honors, you are still a modest man of high integrity, as reflected in the tributes dedicated to you by your peers. How did you manage to always conduct yourself in such a good manner and with such a positive attitude? Was there any advice you took to heart and which helped shape you the way you are today?*

Fang: My parents set a good example for me. To this day, I still remember vividly the advice my father gave me. He said if you extend your help to others, you should forget about it. On the contrary, if you receive assistance from other people, you should always keep that in mind and return the favor. Before I reached 40, the country had become poor. My wife Tingmui was weak and my daughters were still young. Our standard of living was basically minimal. Some of my friends helped me, but I was incapable of paying them back. When my economic situation improved, I paid back all the debts and whenever we came across a friend who needed financial assistance, we never hesitated to lend a helping hand.

My mother was exemplary of how one should conduct one's self, even in an adverse environment. She came from a village background and had no educational opportunity or cultural heritage. It was not until after her marriage that she had the chance to learn how to read and write. My father, on the other hand, came from the upper class. My mother kept a low profile but she learned exceptionally fast. And she always presented herself well, regardless of the situation. I learned from my mother that if you want to adapt yourself to a new environment you have to learn to be aware of your surroundings and should not be self-centered. I adhered strictly to this philosophy when I first visited the United States, a country so vastly different from my own in almost every aspect— in cultural, logic, systems, terms, etc. I thought of my mother and I began to watch attentively other people's behaviour, their culture, their logic, their way of thinking, their

PHOTOGRAPH 23: A. P. Dawid, Fred J. Hickernell, Kai-Tai Fang and T. W. Anderson at a conference in Hong Kong, 1997. Sitting behind Kai-Tai is C. Radhakrishna Rao.

PHOTOGRAPH 24: Kai-Tai Fang poses with his former students; from left: J. J. Liang, Jianxin Pan, Jianqing Fan, Kai-Tai, H. B. Fang and M. Y. Xie, during a conference break, Hong Kong, 1997.

strengths and their weaknesses. I became happy when I began to appreciate differences in my environment. I learned the things that were desirable and brought them back to the mainland. My positive attitude allowed me to keep an open mind in my management style and, because of this, I was promoted to Deputy Director of the Institute of Applied Mathematics, Academia Sinica, in a mere one-and-a-half years after my return to the mainland.

Another person who influenced my personal development was Professor Kai Lai Chung[2], the first Ph.D. student of Pao-Lu Hsu. While I was in the USA, he told me that many people hid themselves in the office or laboratories to do research and declined to mix with Americans. He asked me why I went to the States and advised me to go out and mix with the people. Following his advice, I joined an activity every two-and-a-half days, be it a seminar or party or social gathering. I benefited greatly by joining these activities and I became aware of a significant improvement in my English communication skills and in my understanding of the Western culture.

I am a firm believer that great achievements involve great risks. I encountered a dilemma in 1980 when I was offered an opportunity to go abroad as a visiting scholar. My wife then was sick and my two kids still small. This, coupled with my unfavorable financial situation, somehow deterred me from making a positive move. One of my friends said to me that if I did not go abroad, my career development would be limited. He said every person must face at least one difficult period in his lifespan and that one must face it with bravery and courage in order to overcome it. This remark was inspirational to me. I therefore took the risk of traveling abroad and was psychologically prepared to come back anytime should I receive a telegram with bad news. My wife, who was told by some of the neighbours that she might not be able to see me again should I go abroad, supported my move. I deeply appreciate her much-needed understanding and unfailing support.

Loie: *Do you have any motto and if yes, could you share it with the younger generation?*

Fang: I don't have one in particular but I think that to me, the most important thing is to be a person of integrity. Good character precedes good academic achievements. Also, don't be afraid of difficulties. Face the problems head-on and find a way to solve them. Remember that there is always a way out for those who look for one. I went through the 10-year Cultural Revolution without even knowing whether there would be a tomorrow and I survived. It is important to build a strong psychological shield to shelter yourself from external attack. I also encourage youngsters to work hard. If you decide to go for something, do it with all your might and give the best you can.

---

[2]Kai Lai Chung (1917–2009) graduated from the Department of Mathematics of Tsinghua University, Beijing. In December 1945 he went to Princeton University and obtained his PhD in 1947, with Harald Cramér and John W. Tukey as his advisors. In 1950s, Chung taught at the University of Chicago, Columbia University, Berkeley, Cornell University and Syracuse University. In 1961 he transferred to Stanford University, providing fundamental contributions to modern probability theory.

PHOTOGRAPH 25: Kai-Tai Fang with his parents, sisters and brother.

PHOTOGRAPH 26: A family picture taken at the Summer Palace in Beijing, 1981.

PHOTOGRAPH 27: Kai-Tai and wife Tingmui in 1985.

PHOTOGRAPH 28: Kai-Tai Fang's family with C. Radhakrishna Rao's family, 2003.
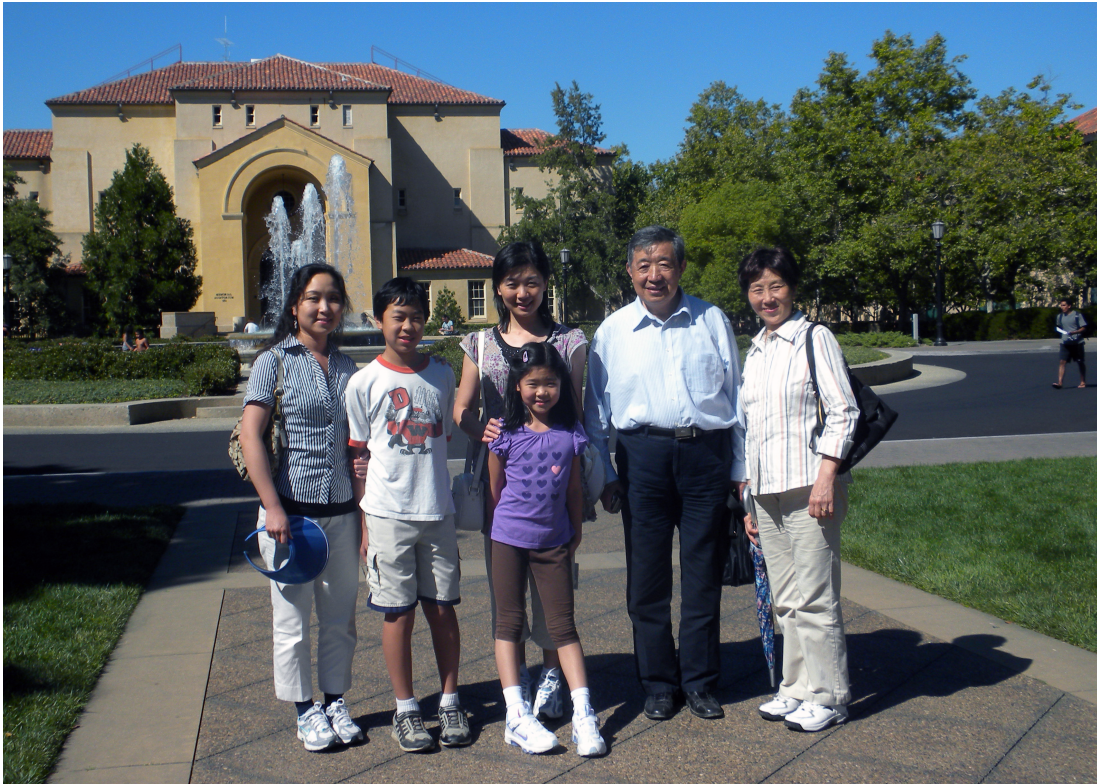
PHOTOGRAPH 29: A happy family reunion in Stanford University, 2010.

8 HBNU-HKBU United International College, Zhuhai Campus: 2006 onwards

Li, Puntanen & Styan = LPS.

LPS: *When did you retire from the HKBU?*

Fang: The usual retirement age for the public sector in the last century of Hong Kong was 60. Thus I should have retired in the year 2000. But due to the development need of the HKBU, the senior management of the University made a special exception in my case to retire in January 2006.

LPS: *You were working in Hong Kong since you emigrated there in 1990, how come you are teaching in a college back in China?*

Fang: Everyone nearing retirement has to make a decision as to what to do, where to stay, and how to contribute to the host society after retirement. Several universities of the US and New Zealand invited me to teach there. At the same time, the newly appointed President of HKBU, Professor C. F. Ng had a plan of building a liberal arts college in the Pearl River Delta in Southern China. The new College would be modeled on the public universities of Hong Kong for its structure and management, with English as its teaching medium, whole-person education in orientation, knowledge for application, and be more internationalization. I liked this challenge and thus accepted the invitation of Professor Ng to become a member of the planning team for this innovative institution.

LPS: *When did your involvement with the UIC begin?*

Fang: In the winter of year 2000, Professor C. F. Ng formed a delegation consisting of the vice-presidents, deans, faculty members and administrators of the HKBU, to begin the search for a site in China for the building of the UIC. The delegation visited many cities in the Pearl River Delta and municipals in the mainland and met with their senior officials. After much comparison and deliberation, all of us found Zhuhai to be the best location due to its green and clean environment, its emphasis on education, and the enthusiastic support of its government leaders. The delegation was given a very high profile reception by the Municipal Government led by the Municipal Party Secretary Mr. Longyuan Huang and City Vice-Mayor Mr. Ningke He at the time.

According to the Chinese law, a local partner had to be found for all local-foreign joint enterprises. The new college, as such nature, needed to find a local collaborator. For this mission, a committee was set up by the HKBU, headed by the then Vice-President (Development), to begin the search. Finally, through the connection and recommendation of Professor Jialu Xu, the then Deputy Chair of the Standing Committee of the National People's Council and a good friend of HKBU, the Beijing Normal University (BNU) was chosen. The new college was named as "Beijing Normal University-Hong Kong Baptist University United International College" (UIC for short), being the longest name of a university/college in China.

LPS: *Was the joint venture well received right from the start?*

Fang: While still on the drawing board, the College was shrouded with intense scepticism. Many people questioned its chance of success. The College, being new and financially self-supporting, had many aspects worked against its smooth birth. The negative factors included the low confidence of parents towards a new institution especially its being such new model of education in China, the high tuition fees charged (compared to the local universities), the limited academic profile with only three Science programmes (Computing Science, Statistics, and Environmental Science) and two Business programmes (Finance, and Applied Economics) on offer, and inability to lure the Associate Degree students from Hong Kong. Nevertheless, the UIC was born, out of the hard labour and perseverance of the founders/pioneers.

LPS: *How difficult was it to recruit students?*

Fang: In its first year, without the full support of the Guangdong Education Bureau, less than 200 students were recruited, picked from those who were turned down by other universities under the National College Entrance Examination (NCEE) system. After years of hard work, the UIC now has 20 Major Programmes and more than 4900 undergraduate students.

LPS: *How difficult was it to recruit teaching staff at that time/ How about staff recruitment at that time?*

Fang: As a private college, tuition fee is the only source of income for the College. In the early years, due to the low salary and short history, many newly graduated doctorates, worrying about their future prospect, turned down the College offer for elsewhere. At this difficult moment, I came up with an innovative solution: inviting Professor Yung Liang Tong who had just retired from Georgia Institute of Technology, Atlanta, USA, to come and teach for one semester, and Philip Cheng of the Statistical Institute of Taiwan for another.

LPS: *What made the College a success?*

Fang: The UIC built up its reputation by its teaching quality and excellence. The strong commitment of its staff to provide the best for their students, and their seriousness in teaching was greatly impressive to and appreciated by the students and parents. Repeat students, programme transfer students, and students returning from sick leave were given extra hours and special courses free by the teachers to help them catch up with their studies. There was a Statistics student who was in a coma after a car accident. He could not continue his study for almost a whole semester. Teachers not only donated to a fund to help pay his medical bills and health recovery, but also redesigned his study plans so that he could catch up with his academic study and graduate on time. The student of course worked very hard so not to let down those who had supported him and he successfully completed his final year project. Despite his poor family background, the student through part-time work and scholarship supported himself to further study at the Georgetown University in Washington, DC, USA, after graduated from the UIC. Another student, internally transferred from the Teaching English as a Second Language Programme to Statistics, was able to gain admission to the University of Oxford, UK, for postgraduate there after years of patient teaching and guidance of the programme teachers.

Such caring attitude and devotion of staff to the well-being of students was a living proof of the College's educational philosophies of four point education (students-parents-staff-community) and whole-person education which were clearly seen by many parents as education of a very different kind.

On average, about 65% of our Statistics graduates went to further studies in overseas, many to the USA, UK, Australia, Canada, and Hong Kong. The rest found work in mainland China and Hong Kong.

LPS: *How about the Chinese Government's support to education? It seems to be very good, maybe better than in many Western countries?*

Fang: Support from the Chinese Government to public institutions has grown on a regular basis. But as a private institution, no support comes to the UIC from either the Hong Kong or Zhuhai Governments. Tuition fee is its only income. Even then, the tuition fees cannot be raised without the permission of the Guangdong Price Bureau, the authority that regulates the level of charges for all goods and services in the province. The development path of the College has not been a smooth one. Recently, the Zhuhai Government has offered the College a piece of land for its future development. This may be deemed as a commendation for its contribution and achievement.

LPS: *Ten years ago you had some plans in mind regarding how you see the years after your "retirement". We now see that you actually did not retire at all. So, can the same question be asked again? And how did your plans work?*

Fang: In the past ten years, most of my time was spent on the establishment of the new Statistics Programme in the UIC. But my research on experimental design, data mining, statistical simulation and magic squares has never stopped. From 2005 to now I published one monograph *Design and Modeling for Computer Experiments* (2005), three textbooks in Chinese, 48 research papers, and nine papers in mathematics/statistics and culture. I have also done a lot of services in refereeing research papers, organizing sessions in some international conferences, etc.

LPS: *What is the role of Statistics in Chinese universities these days?*

Fang: In the recent 30 years, Statistics has grown enormously in status in China. The discipline used to be subsumed under Mathematics which was a Class 1 discipline, whereas Statistics was only a Class 2. Three years ago, the Chinese Ministry of Education decided to turn Statistics into a Class 1 discipline. At present, most universities and colleges already have an independent Statistics department, offering Master and Ph.D. programmes. Unfortunately, there is a lack of qualified teachers for it. In view of the teaching need, the Higher Education Press of China invited me to be the Chief Editor, to assist them in the publication of a series of statistics textbooks. Several books have been published including three of my works with other professors. These are: *Design and Modeling of Experiments* (2011) with Min-Qian Liu and Yong-Dao Zhou, *Matrix Algebra in Statistics* (2013) with Min Chen, and *Modern Basic Statistics* (2015) with Xiaoling Peng.

LPS: *What has been your main research interest during that 10 years?*

Fang: Due to my duties as the department head of the new Statistics Programme, and chairs/members of many committees, the time I could devote on research has been substantially less when I was at the HKBU. However, I kept myself closely informed of the new development of experimental designs, especially uniform design. I continued to provide help to many users of the uniform design in different countries. In 2008, Professor Yuan Wang and I were awarded the Chinese State Natural Science Award, the highest recognition for the initial concept, theory and application of the uniform design

I have a definite concern over the "Big Data". I disagree with some authors who attempted to exclude Statistics out of the research and application of Big Data. No matter what, the popularity and application of Big Data is a challenge for the statisticians who can find ample space for development from there.

Recently, I have some interest in magic squares. The concept of magic squares was originated from China. Over the years, mathematicians and its fans in China have made huge contribution to the development of magic squares. I have great respect for George P. H. Styan especially for his research and achievement in this area. The series of papers by him and his collaborators were greatly inspiring for us.

LPS: *Could you please tell some collaborators over the last ten years?*

Fang: Due to the convenient location of the UIC (in the Pearl River Delta), I have had many opportunities to visit the numerous universities in the region, such as Sun Yat-sen University, Guangzhou University, South China Normal University, Shenzhen University, Beijing Normal University at Zhuhai campus, etc. I was even able to make frequent visits to the Chinese Academy of Sciences, where I studied and worked for about 27 years. I have been working with Professors Yuan Wang, and Min Chen there on some joint projects.

The founding of the Statistics Programme at the UIC had taken up much of my time and I was unable to take part in international conference as frequent as before. However, I was invited to be a member of the Institute of Mathematical Statistics (IMS) Fellow Committee for the selection of IMS fellows from 2007 to 2009. I have also attended two International Workshops on Matrices and Statistics: one in Shanghai, China (2010) and one in Toronto, Canada (2013). I am going to attend the one in Haikou City (Hainan, China) in 2015. Those occasions enabled me to meet up with many old friends. Through my personal contact with Tam Ming, UIC and Georgetown University, USA, have started a close cooperation for Master Programmes in biostatistics and other areas. Similar cooperation will take place with the Victoria University of Wellington, New Zealand, and Illinois Institute of Technology, Chicago, USA.

PHOTOGRAPH 30: UIC teaching staff in statistics, 2010.

PHOTOGRAPH 31: Hans Joachim Werner, Jeffrey J. Hunter, Yonghui Liu, Kai-Tai Fang, Shenfang Zheng, Simo Puntanen, Tingmui Li; IWMS-Shanghai, June 2010.

LPS: *How big is the teaching staff in statistics in the UIC?*

Fang: Currently we have four full professors, three associate professors, three assistant professors and many instructors and teaching assistants.

LPS: *Importance of your wife, Tingmui Li. As we all know, her role and support in your career has been go utmost importance.*
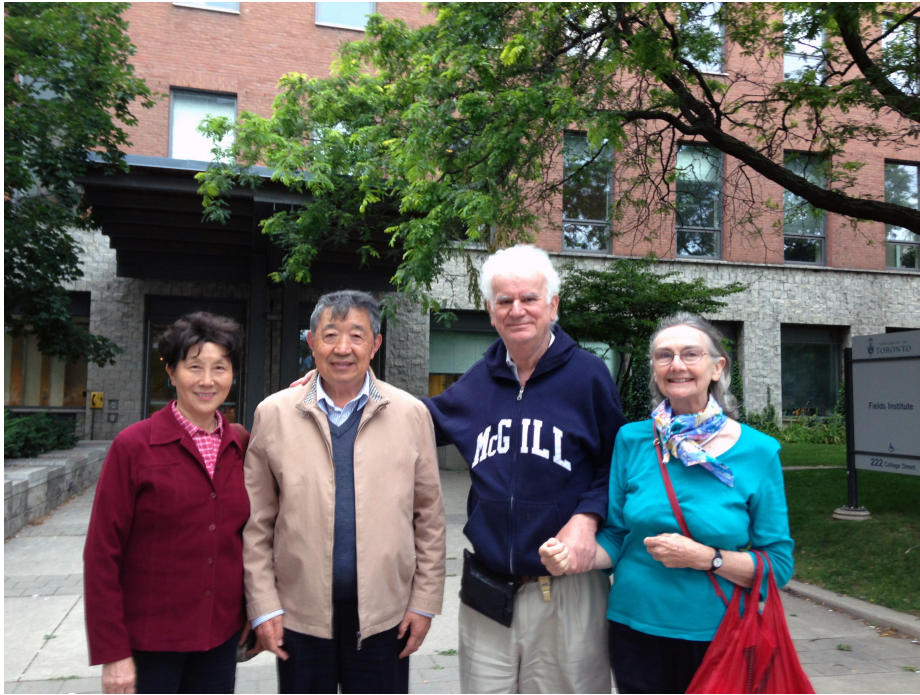
PHOTOGRAPH 32: Kai-Tai Fang with his wife Tingmui, George P. H. Styan with his wife Evelyn; IWMS-Toronto, August 2013.

PHOTOGRAPH 33: Kai-Tai with his wife Tingmui Li in Kunming (capital of Yunnan Province), 2014.

Fang: Tingmui retired from the Peoples' University of China (also known as Renmin University of China) in 1999. We married in 1968 and during the first 31 years of our marriage we lived apart for some 20 years. Soon after we got married and for seven years (from 1968 to 1975), I worked in a military farm or in in Beijing while she worked as a technician in a chemical factory in Shandong Province; from 1980 to 1986, she lived in Beijing and I spent three years in the USA and six months in Switzerland; during the nine years of the 90s (1990–1999), I was in Hong Kong and she was in Beijing. We finally got back together when Tingmui moved to Hong Kong after her retirement in 1999. Next year is our Golden Jubilee. Despite our long periods of separation, we managed to bring up two wonderful daughters: Ying Fang and Yan Fang. They successfully completed a Master of Business Administration and Master of Biostatistics, respectively. We have one grandson and two granddaughters.

LPS: *Hobbies these days?*

Fang: When I was a secondary school student, I liked playing Chinese Chess, swimming, reading all sorts of books; while at university, I liked long-distance running, watching modern drama, reading novels; in Hong Kong, my favourite sports were table tennis, swimming and hiking; nowadays, I can only take walks and watch TV.

LPS: *What do you like to do when you are not doing statistics?*

Fang: When I retire, I would like to write my own stories and through them to illuminate the changes and development of the nation.

LPS: *Any particular advice that you would like to give for a young student having the academic career in mind?*

Fang: We would like to see in our young scholars the team spirit, perseverance, tolerance, and ability to work hard and endure hardship. For research students, we would like them to pay more attention to seminars, as many good ideas come from discussion and exchange.

LPS: *Thank you very much for your time devoted to this interesting conversation, Kai-Tai! It was indeed a great pleasure for us.*

Acknowledgements

References

Chen, Dayue & Olkin, Ingram (2012). Pao-Lu Hsu (Xu, Bao-lu): The grandparent of Probability and Statistics in China. *Statistical Science*, 27, 434–445.

Fan, Jianqing & Li, Gang (eds.) (2005). *Contemporary Multivariate Analysis and Design of Experiments*. World Scientific, Singapore.

Fang, Kai-Tai (1981). The limiting distribution of linear permutation statistics and its applications. *Acta Mathematicae Applicatae Sinica*, 4, 69–82.

Fang, Kai-Tai (1982). A restricted occupancy problem. *Journal of Applied Probability*, 19, 707–711.

Fang, Kai-Tai (2006). Uniform designs. In: *Encyclopedia of Statistics, 2nd Ed.*, Samuel Kotz, ed., Volume 14, pp. 8841–8850. Wiley, New York.

Fang, Kai-Tai & Chan, Ling-Yau (2006). Uniform design and its industrial applications. In: *Springer Handbook of Engineering Statistics*, Hoang Pham, ed., pp. 229–247. Springer, New York. The Higher Education Press, Beijing. (in Chinese)

Fang, Kai-Tai & Chen, Min (2013). *Matrix Algebra in Statistics*. The Higher Education Press, Beijing. (in Chinese)

Fang, Kai-Tai; Kotz, Samuel & Ng, Kai-Wang (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall/CRC, London.

Fang, Kai-Tai; Li, Runze & Sudjianto, Agus (2005). *Design and Modeling for Computer Experiments*. Chapman and Hall/CRC, London.

Fang, Kai-Tai & Lin, Dennis K. J. (2003). Uniform designs and their application in industry. In: *Handbook of Statistics 22: Statistics in Industry*, R. Khattree & C. R. Rao, eds., pp. 131–170. North-Holland, Amsterdam.

Fang, Kai-Tai; Liu, Min-Qian & Zhou, Yong-Dao (2011), *Design and Modeling of Experiments*. The Higher Education Press, Beijing. (in Chinese)

Fang, Kai-Tai & Mukerjee, Rahul (1999). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika*, 87, 193–198.

Fang, Kai-Tai & Peng, Xiaoling (2015). *Modern Basic Statistics*. The Higher Education Press, Beijing. (in Chinese)

Fang, Kai-Tai; Quan, H. & Chen, Q. Y. (1988). *Applied Regression Analysis*. Science Press, Beijing. (In Chinese)

Fang, Kai-Tai & Wang, Yuan (1994). *Number-Theoretic Methods in Statistics*. Chapman and Hall/CRC, London.

Fang, Kai-Tai & Xu, Jian-Lun (1987). *Statistical Distributions*. Science Press, Beijing.

Fang, Kai-Tai & Zhang, Yaoting (1990). *Generalized Multivariate Analysis*. Science Press and Springer, Beijing and Berlin.

Loie, Agnes W. L. (2005). A conversation with Kai-Tai Fang. In: *Contemporary Multivariate Analysis and Design of Experiments*, Jianqing Fan & Gang Li, eds. World Scientific, Singapore, pp. 1–22.

Pólya, George (1957). *How to Solve It: A New Aspect of Mathematical Method, 2nd Ed.*. Princeton University Press, NJ. [First Edition 1945]

Zhang, Yaoting & Fang, Kai-Tai (1982). *An Introduction to Multivariate Analysis*. Science Press, Beijing. (In Chinese)

Index

Twenty-six years of the International Workshop on Matrices and Statistics (IWMS): 1990–2015

Simo Puntanen[1,*]   George P. H. Styan[2,†]

[1]School of Information Sciences, FI-33014 University of Tampere, Finland
[2]Department of Mathematics and Statistics, McGill University, Montréal, Canada H3A 0B9

Abstract   We present a short history of the International Workshop on Matrices and Statistics (IWMS). The first IWMS was held in Tampere, Finland, in 1990, and the 24th IWMS will be held at Hainan Normal University, Haikou City (Hainan Island), China (25–28 May 2015); see IWMS-2015 website and IWMS-2015 announcement.

We have established an open-access website for all twenty-four IWMS at the University of Tampere: `http://www.sis.uta.fi/tilasto/iwms/` where we intend to put all associated reports and photographs of the IWMS from 1990 onwards, including those published in *Image: The Bulletin of the International Linear Algebra Society.* Complete videos, prepared by Jarmo Niemelä and Reijo Sund, of the talks at two pre-IWMS Tampere conferences in statistics in 1987 and 1990 are on YouTube.

Keywords   IWMS; Workshop; Matrices and Statistics

1   Background

The first workshop in the International Workshop on Matrices and Statistics (IWMS) series took place at the University of Tampere in Tampere, Finland, 6–8 August 1990. This workshop was organized by a local committee from the Statistics Unit of the Department of Mathematical Sciences at the University of Tampere. The key persons in the organizing committee were Pentti Huuhtanen, Erkki Liski, Tapio Nummi, Tarmo Pukkila, Simo Puntanen, and George P. H. Styan. There was no idea at that time that this would be the beginning of an almost annual series of meetings. This first IWMS was actually called "The International Workshop on Linear Models, Experimental Designs, and Related Matrix Theory". Since 1990 the name has changed twice, and in 1998 the IWMS became the "International Workshop on Matrices and Statistics", following a suggestion by Professor C. Radhakrishna Rao.

In 1990 in Tampere there were 98 participants from 18 different countries. The Keynote Address in 1990 was given by C. Radhakrishna Rao. The invited speakers were

Jerzy K. Baksalary	Sujit Kumar Mitra	Friedrich Pukelsheim
R. Dennis Cook	Seppo Mustonen	Jagdish N. Srivastava
Yadolah Dodge	Heinz Neudecker	George P. H. Styan
Shanti S. Gupta	Ingram Olkin	

The organizers of the group meetings were

---

*Corresponding author. Email: simo.puntanen@uta.fi
†Email: geostyan@gmail.com

Jerzy K. Baksalary	Sanpei Kageyama	Kirti R. Shah
Tadeusz Caliński	Jürgen Kleffe	George P. H. Styan
R. Dennis Cook,	Sujit Kumar Mitra	Götz Trenkler
R. William Farebrother	Seppo Mustonen	Song-Gui Wang
Yasunori Fujikoshi	Friedrich Pukelsheim	Haruo Yanai
T. P. Hettmansperger	Jorma Rissanen	

Many of these persons have also been active participants in later workshops. George P. H. Styan has missed only one IWMS in 1990–2014 and thereby has the highest score of $n-1$ attended IWMS. For the schedule of the first IWMS see the last page of this article.

The following is an up-to-date version of the aims of the IWMS:

> The purpose of the IWMS is to stimulate research and, in an informal setting, to foster the interaction of researchers in the interface between statistics and matrix theory. The Workshop will provide a forum through which statisticians may be better informed of the latest developments and newest techniques in linear algebra and matrix theory and may exchange ideas with researchers from a wide variety of countries.

Quite soon after the 2nd IWMS in Auckland, New Zealand, in 1992, the organizing system for the IWMS found its form as two committees: International and Local. The International Organizing Committee (IOC) for several years comprised R. William Farebrother (UK), Simo Puntanen (Finland), George P. H. Styan (Canada), Hans Joachim Werner (Germany). Recently, also S. Ejaz Ahmed (Canada), Jeffrey J. Hunter (New Zealand), Augustyn Markiewicz (Poland), Götz Trenkler (Germany), Júlia Volaufová (USA), and Dietrich von Rosen (Sweden), have joined the IOC; in 2008 George P. H. Styan was named Honorary Chair of the IOC of the IWMS. It is of course worth emphasizing that a most demanding task and responsibility for the meeting arrangements belongs to the local organizing committee.

The IWMS series has had three *ILAS Lecturers*: Gene H. Golub (1999), Jerzy K. Baksalary (2003), Ravindra B. Bapat (2008), and a fourth Karl Gustafson will be in Haikou (2015), two *Nokia Lecturers*: Ingram Olkin (2004) and C. Radhakrishna Rao (2005). Chris Gotwalt will be an *SAS Lecturer* in 2015.

As IWMS Birthday Boys, the following have been celebrated: T. W. Anderson (80, 90), Lynn Roy LaMotte (70), Ingram Olkin (80, 90), C. Radhakrishna Rao (80), Muni S. Srivastava (75), George P. H. Styan (65, 70, 75), and a Special Session for Tarmo Pukkila (60) was held in 2006. In 2015 in Haikou there will be special birthday sessions for Kai-Tai-Fang (75) and Simo Puntanen (70).

Memorial Sessions have been held for Bernhard Flury (1999), Sujit Kumar Mitra (2004), Jerzy K. Baksalary (2005), Shayle R. Searle (2013), and Haruo Yanai (2014).

2 List of the workshops

We now present a list of the 23 Workshops that have been held from 1990 to 2014, leading up to the 2015 Workshop in Haikou City, Hainan Island, China. The photographs 1, 7, 8, 10 and 13 are taken by the the University of Tampere photographer; the photograph 2 by the University of Auckland photographer; the photographs 18–19 by the Shanghai University of Finance and Economics photographer; the photograph 5 by Hazel Hunter; the photograph 21 by the University of Ljubljana photographer; the rest of the photographs are taken by Simo Puntanen.

Figure 1: Group of participants in IWMS-1990, Tampere; C. Radhakrishna Rao inviting more people to the picture.

1990/1: International Workshop on Linear Models, Experimental Designs, and Related Matrix Theory Tampere, Finland, 6–8 August 1990, $n = 98$.
Chair of the Organizing Committee: Erkki Liski. Programme.
Videos from the conferences in statistics in Tampere in 1987 and 1990, prepared by Jarmo Niemelä and Reijo Sund.

1992/2: [2nd] International Workshop on Matrix Methods for Statistics, Auckland, New Zealand, 4–5 December 1992, $n = 23$.
Chair of the Organizing Committee: Alastair J. Scott. Report in *Image.*

1994/3: Tartu Satellite Workshop on Matrices in Statistics, Tartu, Estonia, 28 May 1994.
Local Chairs: Ene-Margit Tiit & Hannu Niemi, IOC Chair: George P. H. Styan.

1995/4: 4th International Workshop on Matrix Methods for Statistics, Montréal, Québec, Canada, 15–16 July 1995, $n = 70$.
Local and IOC Chair: George P. H. Styan. Report in *Image.*

1996/5: 5th International Workshop on Matrix Methods for Statistics, Shrewsbury, England, 18–19 July 1996, $n = 28$.
Local and IOC Chair: R. William Farebrother. Programme. Report in *Image.*

1997/6: 6th International Workshop on Matrix Methods for Statistics, Istanbul, Turkey, 16–17 August 1997, $n = 40$.
Local Chair: Fikri Akdeniz, IOC Chair: Hans Joachim Werner. Report in *Image.*

1998/7: 7th International Workshop on Matrices and Statistics, Fort Lauderdale, Florida, USA, 11–14 December 1998, in celebration of *T. W. Anderson's 80th birthday*, $n = 78$. Programme.
Local Chair: Fuzhen Zhang, IOC Chair: George P. H. Styan. Report in *Image.*

1999/8: 8th International Workshop on Matrices and Statistics, Tampere, Finland, 7–8 August 1999, $n = 95$.
■ The ILAS Lecturer: *Gene H. Golub.*
Memorial Session: *Bernhard Flury* (1951–1999).
Local and IOC Chair: Simo Puntanen. Programme.
`http://www.uta.fi/laitokset/mattiet/workshop99/`, Report in *Image.*

2000/9: 9th International Workshop on Matrices and Statistics, Hyderabad, India, 9–13 December 2000, in celebration of *C. Radhakrishna Rao's 80th birthday*, $n = 100$.
Local Chairs: S. B. Rao, P. Bhimasankaram, IOC Chair: Hans Joachim Werner.
Programme. Report in *Image.*

Figure 2: Group of participants in IWMS-2, Auckland, New Zealand, 4–5 December 1992.

2001/10: 10th International Workshop on Matrices and Statistics, Voorburg, The Netherlands, 2–3 August 2001, $n = 54$. Report in *Image*.
Local Chair: Patrick J. F. Groenen, IOC Chair: George P. H. Styan.

2002/11: 11th International Workshop on Matrices and Statistics, Lyngby, Denmark, 29–31 August 2002, in celebration of *George P. H. Styan's* 65th birthday, $n = 65$. Report in *Image*.
Local Chair: Knut Conradsen, IOC Chair: Hans Joachim Werner.

2003/12: 12th International Workshop on Matrices and Statistics, Dortmund, Germany, 5–8 August 2003, $n = 45$.
■ The ILAS Lecturer: *Jerzy K. Baksalary.*
Programme. Report in *Image*.
Local Chair: Götz Trenkler, IOC Chair: Hans Joachim Werner.

2004/13: 13th International Workshop on Matrices and Statistics, Będlewo, Poznań, Poland, 18–21 August 2004, in celebration of *Ingram Olkin's* 80th birthday, $n = 82$.
■ The Nokia Lecturer: *Ingram Olkin.*
Memorial Session: *Sujit Kumar Mitra* (1932–2004).
Local Chair: Augustyn Markiewicz, IOC Chair: Simo Puntanen.
http://matrix04.amu.edu.pl/, Programme. Report in *Image.* Poster.

2005/14: 14th International Workshop on Matrices and Statistics, Massey University, Albany Campus, Auckland, New Zealand, 30 March – 1 April 2005, $n = 50$.
■ The Nokia Lecturer: *C. Radhakrishna Rao.*
Memorial Session: *Jerzy K. Baksalary* (1944–2005).
Local Chair: Jeffrey J. Hunter, IOC Chair: George P. H. Styan.
Website. Announcement. Programme. Report in *Image.* Flyer. Poster.

2006/15: 15th International Workshop on Matrices and Statistics, Uppsala, Sweden, 13–17 June 2006, $n = 68$.
Special Session for *Tarmo Pukkila's* 60th birthday.
Local Chair: Dietrich von Rosen, IOC Chair: Hans Joachim Werner.
http://www.bt.slu.se/iwms2006/iwms06.html, Programme. Report in *Image*.

2007/16: 16th International Workshop on Matrices and Statistics, Windsor, Ontario, Canada, 1–3 June 2007, $n = 74$,

Figure 3: IWMS-4, Montréal, 15–16 July 1995.
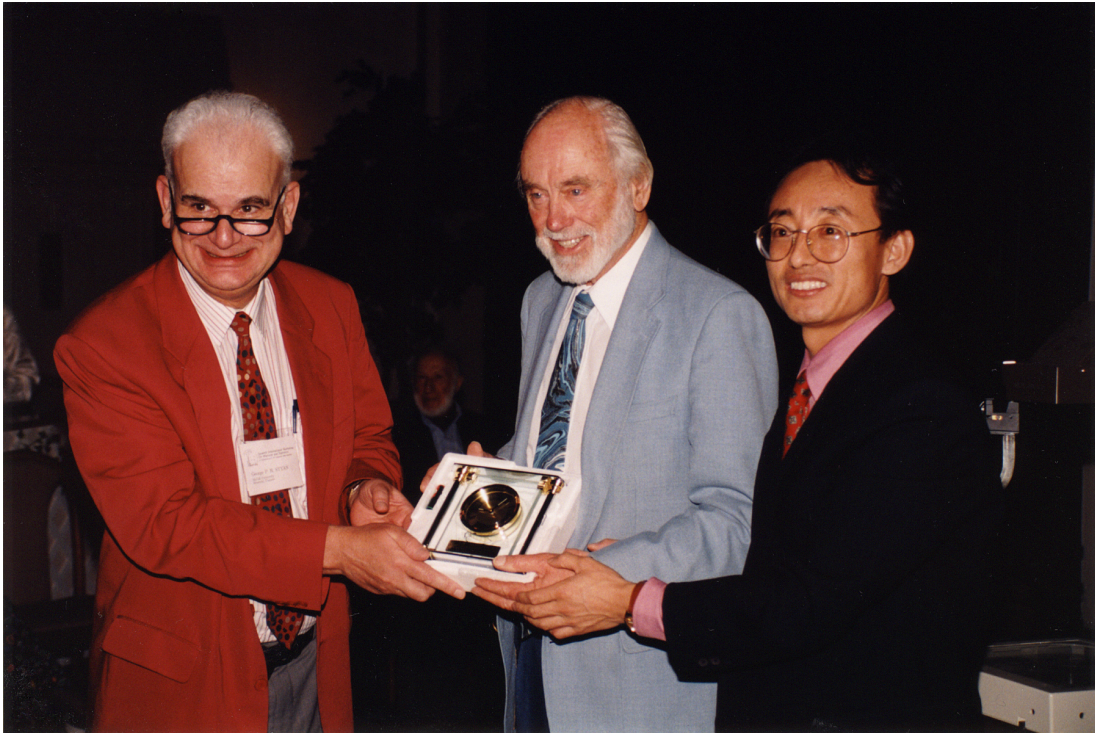
Figure 4: George P. H. Styan, T. W. Anderson, Fuzhen Zhang; Fort Lauderdale, December 1998.

in celebration of *George P. H. Styan's* 70th birthday,
Local Chair: S. Ejaz Ahmed, IOC Chair: George P. H. Styan.
`http://www.uwindsor.ca/units/iwms/main.nsf`, Programme. Poster.

2008/17: 17th International Workshop on Matrices and Statistics, Tomar, Portugal, 22–26 July 2008, $n = 80$, in celebration of *T. W. Anderson's* 90th birthday.
■ The ILAS Lecturer: *Ravindra B. Bapat.*
Local Chair: João T. Mexia, IOC Chair: Simo Puntanen.
`http://www.iwms08.ipt.pt/`, Programme. Report in *Image.* Poster. A special stamp.

2009/18: 18th International Workshop on Matrices and Statistics, Smolenice Castle, Slovakia, 23–27 June 2009, $n = 67$.
Local Chair: Viktor Witkovský, IOC Chair: Júlia Volaufová.
`http://www.um.sav.sk/en/iwms2009.html`, Programme. Poster.

2010/19: 19th International Workshop on Matrices and Statistics, Shanghai University of Finance and Economics, China, 5–8 June 2010, $n = 186$.
Local Chair: Yonghui Liu, IOC Chair: Jeffrey J. Hunter.
`http://www1.shfc.edu.cn/iwms/index.asp`, Programme. Report in *Image.*

2011/20: 20th International Workshop on Matrices and Statistics, with the Tartu 9th Conference on Multivariate Statistics, Tartu, Estonia, 27–30 June 2011, $n = 100$.
Special Session for *Muni S. Srivastava's* 75th birthday.
Local Chair: Kalev Pärna, Programme Committee Chair: Dietrich von Rosen,
Vice-Chair: Tõnu Kollo.
`http://www.ms.ut.ee/tartu11/`

2012/21: 21st International Workshop on Matrices and Statistics, with the LinStat-2012, Będlewo, Poznań, Poland, 16–20 July 2012, $n \approx 100$
Special Session for *George P. H. Styan's* 75th birthday.
Local Chair: Katarzyna Filipiak, LinStat Chair: Augustyn Markiewicz,
IWMS Chair: Simo Puntanen.
Programme. Report in *Image.*
`http://linstat2012.au.poznan.pl/index.html`.

2013/22: 22nd International Workshop on Matrices and Statistics, hosted at the University of Toronto by the Fields Institute, Toronto, Ontario, Canada, 12–15 August 2013.
IWMS Chair: S. Ejaz Ahmed.

Figure 5: IWMS-13, Będlewo, Poznań, Poland, 18–21 August 2004.

Figure 6: IWMS-15, Uppsala, Sweden, 13–17 June 2006.

Memorial Session: *Shayle R. Searle* (1928–2013)
Special Session for *Lynn Roy LaMotte's* 70th birthday. Programme. Report in *Image*. Poster.
"Shayle R. Searle: Pioneer in Linear Modelling." Article by Jeffrey J. Hunter (2015), in *Australian & New Zealand Journal of Statistics*.
`http://www.fields.utoronto.ca/programs/scientific/13-14/IWMS/`.

2014/23: 23rd International Workshop on Matrices and Statistics, Department of Mathematics, University of Ljubljana, Ljubljana, Slovenia, 9–12 June 2014, $n \approx 60$.
IWMS Chair: Hans Joachim Werner, Local Chair: Matjaž Omladič.
Special Sessions for *Ingram Olkin's* 90th birthday.
Memorial Session: *Haruo Yanai* (1940–2013)
`https://www.fmf.uni-lj.si/en/faculty/location/`
A Linear Algebra Workshop (LAW) organized by Matjaž Omladič preceded the IWMS in Ljubljana from June 4 to June 12.

2015/24: 24th International Workshop on Matrices and Statistics, Haikou City, Hainan Province, China, 25–28 May 2015.
■ The ILAS Lecturer: *Karl Gustafson*
■ The SAS Lecturer: *Chris Gotwalt*
Special Sessions for *Kai-Tai Fang's* 75th birthday and for *Simo Puntanen's* 70th birthday.
IWMS Chair: Jeffrey J. Hunter, Local Chair: Chuanzhong Chen.
Announcement.
`http://iwms2015.csp.escience.cn/dct/page/1`

3 Special issues of journals devoted to the IWMS

Selected refereed papers presented at the IWMS have been (or are about to be) published in the following journal special issues:

1992: Third Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 176 (1992), viii + 289 pp. (Includes 8 research papers presented at the Workshop held in Tampere, Finland, 6–8 August 1990.) Preface. DOI.
Jerzy K. Baksalary & George P. H. Styan, eds.

1993: *Journal of Statistical Planning and Inference*, vol. 36, no. 2–3 (1993), pp. 127–432. (24 research papers presented at the Workshop held in Tampere, Finland, 6–8 August 1990.) Preface. Author index. DOI.
Jerzy K. Baksalary & George P. H. Styan, eds.

1994: Fourth Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 210 (1994), 273 pp. Preface. DOI.
Jeffrey J. Hunter, Simo Puntanen & George P. H. Styan, eds.

1996: Fifth Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, In Celebration of C. Radhakrishna Rao's 75th Birthday. vol. 237/238 (1996), vii + 273 pp. Author index.
Ravindra B. Bapat, George P. H. Styan & Hans Joachim Werner, eds.

1997: Sixth Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 264 (1997), ix + 506 pp. Preface. DOI.
R. William Farebrother, Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

1999: Seventh Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 289 (1999), iv + 344 pp. Preface. DOI.
R. William Farebrother, Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

2000: Eighth Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 321 (2000), xi + 412 pp. Preface. DOI.
Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

2002: Ninth Special Issue on Linear Algebra and Statistics: *Linear Algebra and its Applications*, vol. 354 (2002), xii + 291 pp. Preface. DOI.
Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

2004: Tenth Special Issue on Linear Algebra and Statistics, Part 1: *Linear Algebra and its Applications*, vol. 388 (2004), 400 pp. Preface. DOI.
Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

2005: Tenth Special Issue on Linear Algebra and Statistics, Part 2: *Linear Algebra and its Applications*, vol. 410 (2005), 290 pp. Preface. DOI.
Simo Puntanen, George P. H. Styan & Hans Joachim Werner, eds.

2005: *Research Letters in the Information and Mathematical Sciences*, vol. 8 (2005), v + 228 pp. Special Issue: Proceedings of the 14th International Workshop on Matrices and Statistics, Auckland, New Zealand, 30 March–1 April 2005. Foreword. Available online.
Jeffrey J. Hunter & George P. H. Styan, eds.

2006: *Linear Algebra and its Applications*, vol. 417 (2006), Proceedings of the 13th International Workshop on Matrices and Statistics, Będlewo, Poznań, Poland, 18–21 August 2004. Preface. DOI.
Ludwig Elsner, Augustyn Markiewicz & Tomasz Szulc, eds.

Figure 7: Jerzy K. Baksalary giving a talk in Tampere, August 1990. Front row: Shanti S. Gupta, J.N. Srivastava, Song-Gui Wang, Götz Trenkler, Stanisław Gnot.

2009: *Linear Algebra and its Applications*, vol. 430, no. 10 (2009), pp. 2563–2834, Proceedings of the 16th International Workshop on Matrices and Statistics, Windsor, Ontario, Canada, 1–3 June 2007. Preface. DOI.
S. Ejaz Ahmed, Jeffrey J. Hunter, George P. H. Styan & Götz Trenkler, eds.

2010: *Acta et Commentationes Universitatis Tartuensis de Mathematica*, vol. 14, 2010. Proceedings of the 18th International Workshop on Matrices and Statistics, Smolenice Castle, Slovakia, 23–27 June 2009.
Tõnu Kollo, Dietrich von Rosen, Viktor Witkovský & Júlia Volaufová, eds.

2011: *Acta et Commentationes Universitatis Tartuensis de Mathematica*, vol. 15, nr. 1, 2011. Proceedings of the 19th International Workshop on Matrices and Statistics. Shanghai University of Finance and Economics, China, 5–8 June 2010. Shanghai IWMS.
Tõnu Kollo & Dietrich von Rosen, eds.

2012: Special volume of *Acta et Commentationes Universitatis Tartuensis de Mathematica*, vol. 16, 2012. Proceedings of The 9th Tartu Conference on Multivariate Statistics & IWMS-2011. Simo Puntanen, ed.

2013: *Multivariate Statistics: Theory and Applications*, World Scientific, devoted to The 9th Tartu Conference on Multivariate Statistics & IWMS-2011. ISBN: 978-981-4449-39-7 (hardcover). Tõnu Kollo, ed.

2013: IWMS-2012 and LinStat-2012: A special issue of *Discussiones Mathematicae - Probability and Statistics*, vol. 33, 2013, Devoted to IWMS-2012 and LinStat-2012. Preface. Coordinating Editor: Roman Zmyślony, Guest Editors: Carlos A. Coelho and Simo Puntanen.

2014: IWMS-2012 and LinStat-2012: *Communications in Statistics - Theory and Methods*, Volume 43, Issue 5, 2014. Special Issue: Advances on Linear Models and Inference. Devoted to IWMS-2012 and LinStat-2012. Preface. Guest Editors: Júlia Volaufová and Augustyn Markiewicz.

2014: IWMS-2012 and LinStat-2012: *Communications in Statistics - Simulation and Computation*, Volume 43, Issue 9, 2014. Special Issue: Advances on Linear Models and Inference – Computational Aspects. Devoted to IWMS-2012 and LinStat-2012. Preface. Guest Editors: Katarzyna Filipiak and Simo Puntanen.

2015: IWMS-2013: A special issue of *Journal of Statistical Computation and Simulation*. Editor: S. Ejaz Ahmed. [In progress]

2015: IWMS-2014: A special issue of *Operators and Matrices*. Devoted to IWMS-2014. Guest Editors: Jeffrey J. Hunter, Mitja Mastnak, Matjaž Omladič, and Simo Puntanen. [In progress]

2015: IWMS-2015: *Souvenir Booklet of the 24th International Workshop on Matrices and Statistics* (25–28 May 2015), Haikou City, Hainan Island, China. Editor: Jeffrey J. Hunter. [In progress]

2015: IWMS-2015: Special Issue of *Special Matrices*. Editors: Jeffrey J. Hunter, Simo Puntanen and Dietrich van Rosen. [In progress]

Figure 8: R. Dennis Cook, Norman Draper, Nye John, George P. H. Styan; Tampere, August 1990.

Figure 9: Ravindra B. Bapat, Tomar, Portugal, July 2008.
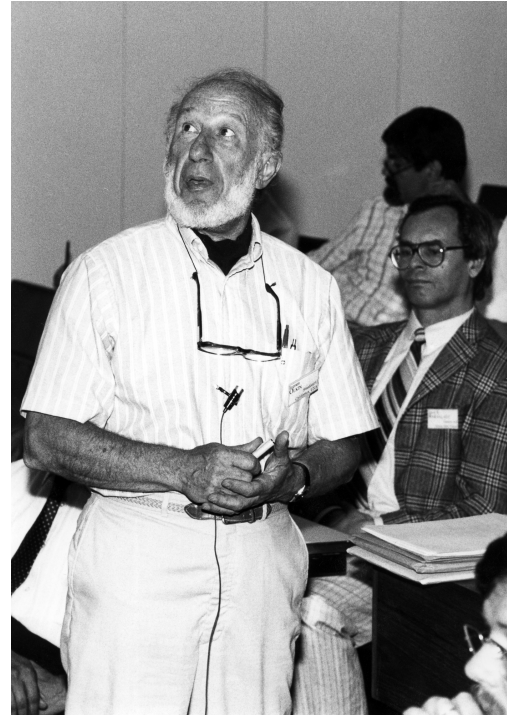
Figure 10: Ingram Olkin, Tampere, August 1990. With Jerzy K. Baksalary and Yadolah Dodge.

Figure 11: C. Radhakrishna Rao, Istanbul, August 1997.

Figure 12: C. Radhakrishna Rao and Bhargavi Rao, Hyderabad, December 2000.

Figure 13: Jerzy K. Baksalary, Tadeusz Caliński, Sujit Kumar Mitra; Tampere, August 1990.

Figure 14: Gene H. Golub, Ingram Olkin, T. W. Anderson; Montréal, July 1995.

Figure 15: Gene H. Golub, X.X., Bikas K. Sinha, Ingram Olkin, Augustyn Markiewicz, Ludvig Elsner, Yongge Tian; Będlewo, Poznań, August 2004.

Figure 16: Enjoying the IWMS banquet (and the view to Detroit) in Windsor, Canada, June 2007.

Figure 17: In an after-dinner session in Smolenice Castle in July 2009, Tõnu Kollo (smiling in the picture) tentatively agreed to organize the IWMS-2011 in Tartu. Left: Soile Puntanen, right: Miroslav Fiedler.

Figure 18: Group of participants in IWMS-2010, Shanghai.

Figure 19: More participants in IWMS-2010, Shanghai.

Figure 20: Group of participants in IWMS-2013, Toronto.

Figure 21: Group of participants in IWMS–2014, Ljubljana.

International Workshop on Linear Models, Experimental Designs & Related Matrix Theory
6–8 August 1990, University of Tampere, Tampere, Finland

Monday, 6 August	Tuesday, 7 August	Wednesday, 8 August
10:00–10:45 *Opening Session:* Tarmo PUKKILA Chair: E.P. Liski C. Radhakrishna RAO Chair: G.P.H. Styan	*Group Meetings:* 8:00–9:30 Trenkler-1 Wang Hettmansperger 9:30–11:30 Cook Baksalary	*Group Meetings:* 8:00–9:30 Shah Farebrother-5 9:30–11:00 Editorial Policy in Statistical Journals
Invited Talk: 10:45–11:30 Ingram OLKIN Chair: J.K. Baksalary	11:30–12:45 ❙❙ *Lunch*	11:00–11:30 ☕ *Coffee* 11:30–12:15 *Keynote Address:* C. Radhakrishna RAO Chair: T. Caliński
11:30–12:45 ❙❙ *Lunch*	*Group Meetings:* 12:45–14:15 Mitra Farebrother-3 14:15–15:45 Srivastava Yanai Farebrother-4	12:15–13:30 ❙❙ *Lunch*
Invited Talks: 12:45–13:30 Sujit Kumar MITRA 13:30–14:15 R. Dennis COOK Chair: F. Pukelsheim	15:45–16:15 ☕ *Coffee*	*Invited Talks:* 13:30–14:00 Jagdish N. SRIVASTAVA 14:00–14:30 Shanti S. GUPTA 14:30–15:00 Yadolah DODGE Chair: T.P. Hettmansperger
14:15–14:45 ☕ *Coffee*	*Group Meetings:* 16:15–17:45 Fujikoshi Pukelsheim Trenkler-2 17:45–19:15 Styan Kleffe Neudecker	15:00–15:30 ☕ *Coffee* 15:30–16:00 Heinz NEUDECKER 16:00–16:30 Friedrich PUKELSHEIM Chair: N.R. Draper
Group Meetings: 14:45–15:45 Gustav Elfving: 1908–1984 16:15–17:45 Mustonen Caliński Farebrother-1 17:45–19:15 Rissanen Kageyama Farebrother-2	20:30–... 🌲🌲🌲🌲🌲 *Sauna Party*	16:30–17:00 ☕ *Coffee* 17:00–17:30 George P. H. STYAN 17:30–18:00 Jerzy K. BAKSALARY Chair: T. Pukkila
20:00–... ❖❙❙ *Reception*		

The first-mentioned groups meet in Room A I, the second in C IX (except on Monday in the Main Auditorium), and the third in A III.

Figure 22: Schedule of the first IWMS; Tampere, 6–8 August 1990.

An indexed illustrated bibliography for Simo Puntanen in celebration of his 70th birthday

George P. H. Styan[1,*]   Ka Lok Chu[2,†]

[1]McGillUniversity, Montréal (Québec), Canada
[2]Dawson College, Westmount (Québec), Canada

Abstract   Many happy returns, Simo! To celebrate over 25 years of collaboration, we present an indexed and illustrated bibliography on the occasion of your 70th birthday on 20 July 2015. This bibliography, which is also annotated and hyperlinked, identifies over 60 publications with both Simo Puntanen and George P. H. Styan as co-authors, our so-called "PunStys". Some selected preprints are included.

Keywords   Simo Puntanen, 70th birthday, bibliography
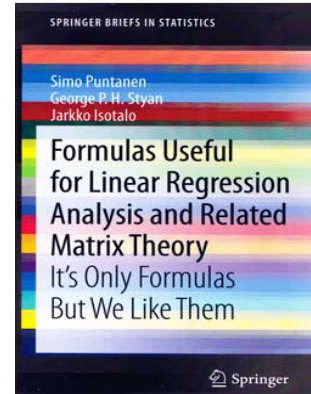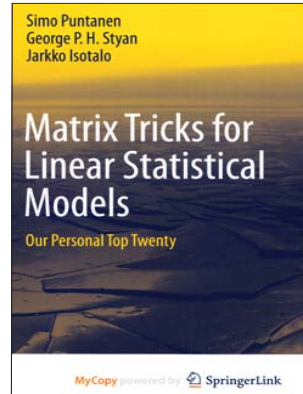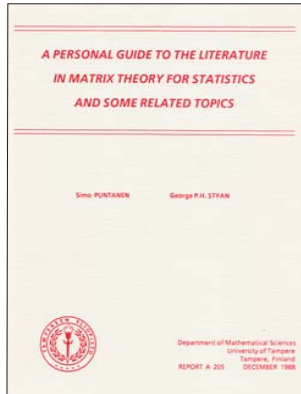
1   Introduction

Many happy returns, Simo! To celebrate over 25 years of collaboration, we present an indexed and illustrated bibliography on the occasion **of your 70th birthday on 20 July 2015.** We begin with this fully-magic most-perfect Graeco-Latin square:

$$\mathbf{S} = \begin{pmatrix} 20 & 7 & 45 & 70 \\ 45 & 70 & 20 & 7 \\ 70 & 45 & 7 & 20 \\ 7 & 20 & 70 & 45 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 20 & 7 & 45 & 70 \\ 70 & 45 & 7 & 20 \\ 7 & 20 & 70 & 45 \\ 45 & 70 & 20 & 7 \end{pmatrix}; \qquad m_{\mathbf{SP}} = 142 = 2 \times 71.$$

This bibliography identifies over 60 publications with both emeriti Simo Puntanen and George P. H. Styan as co-authors, our so-called "PunStys". Some selected preprints are included. A fully-indexed and hyperlinked version is in preparation.

References

[1] [1988-PersiGiddi] *A Personal Guide to the Literature in Matrix Theory for Statistics and Some Related Topics,* by Simo Puntanen & George P. H. Styan, Report A 205, Dept. of Mathematical Sciences, University of Tampere, ISBN 951-44-2385-2, ISSN 0356-3134. iii + 157 pp (December 1988).

[2] [1989-AmStat] "The equality of the ordinary least squares estimator and the best linear unbiased estimator", by Simo Puntanen & George P. H. Styan [with comments by Oscar Kempthorne & by Shayle R. Searle and with "Reply" by the authors], *The American Statistician,* vol. 43, no. 3, pp. 153–164 (1989). [See also "Reply" [3] by Puntanen & Styan to "Further comments by R. W. Farebrother ["Weighted least squares and the Gauss-Markov theorem"], by Ronald Christensen & by David A. Harville.]

[3] [1990-Reply] "Reply" [with "Additional References"], by Simo Puntanen & George P. H. Styan [to Letters to the Editor by R. W. Farebrother, by Ronald Christensen & by David A. Harville on "The equality of the ordinary least squares estimator and the best linear unbiased estimator" [2]], *The American Statistician,* vol. 44, no. 2, pp. 192–193 (May 1990).

[4] [1990-Sankhyā] "A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model", by Jerzy K. Baksalary, Simo Puntanen & George P. H. Styan, *Sankhyā: The Indian Journal of Statistics, Series A,* vol. 52, no. 3, pp. 279–296 (October 1990).

[5] [1990-JKB/Spectrum] "Spectrum and trace invariance criterion and its statistical applications", by Jerzy K. Baksalary & Simo Puntanen, submitted by George P. H. Styan, *Linear Algebra and its Applications,* vol. 142, pp. 121–128 (December 1990).

[6] [1990-BPS/TWA] "On T. W. Anderson's contributions to solving the problem of when the ordinary least-squares estimator is best linear unbiased and to characterizing rank additivity of matrices", by Jerzy K. Baksalary, Simo Puntanen & George P. H. Styan, in *The Collected Papers of T. W. Anderson: 1943–1985,* edited by George P. H. Styan, pub. Wiley, New York, vol. 2, ISBN 0-471-52786-6, pp. 1579–1591 (1990).

[7] [1993-Guide] *A Guide to Books on Matrices and on Inequalities, with Statistical and Other Applications,* by George P. H. Styan & Simo Puntanen, Report A 283, pub. Dept. of Mathematical Sciences, University of Tampere, 65 pp (1993).

[8] [1995-Mitra] "Shorted matrices and their applications in linear statistical models: a review", by Sujit Kumar Mitra, Simo Puntanen & George P. H. Styan, in *Multivariate Statistics and Matrices in Statistics: Proceedings of the Fifth Tartu Conference, Tartu–Pühajärve, Estonia, 23–28 May 1994,* edited by E.-M. Tiit, T. Kollo & H. Niemi, New Trends in Probability and Statistics, vol. 3, pub. VSP International Science Publishers, Zeist (Utrecht), The Netherlands, ISBN 90-6764-195-2 & TEV Ltd., Vilnius, Lithuania, ISBN 9986-546-03-6, pp. 289–311 (1995).

[9] [1996-ET1] "An equivalence relation for two symmetric idempotent matrices", First Solution (by Simo Puntanen & George P. H. Styan) to Problem 95.3.3 (proposed by Shuangzhe Liu & Wolfgang Polasek), *Econometric Theory,* vol. 12, no. 3, pp. 590–591 (August 1996). [Second Solution (by Fuzhen Zhang & George P. H. Styan) on pp. 591–592.]

[10] [1996-ET3] "Matrix results associated with Aitken's generalization of the Gauss–Markov theorem", Solution (by Simo Puntanen & George P. H. Styan) to Problem 95.3.5 (proposed by R. W. Farebrother), *Econometric Theory,* vol. 12, no. 3, vol. 593–595 (August 1996).

[11] [1996-ET4] "The Moore–Penrose generalized inverse of a symmetric matrix", Solution (by Simo Puntanen & George P. H. Styan) to Problem 95.4.3 (proposed by R. W. Farebrother), *Econometric Theory*, vol. 12, pp. 748–749 (October 1996).

[12] [1996-CR/LAA] "A brief biography and appreciation of Calyampudi Radhakrishna Rao, with a bibliography of his books and papers", by Simo Puntanen & George P. H. Styan, *Linear Algebra and its Applications*, vol. 237/238, pp. 1–40 (April 1996). [Linear Algebra and Statistics: In Celebration of C. R. Rao's 75th Birthday (10 September 1995), edited by R. B. Bapat, G. P. H. Styan & H. J. Werner.]

[13] [1996-PPS] "Some further matrix extensions of the Cauchy–Schwarz and Kantorovich inequalities, with some statistical applications", by Josip E. Pečarić, Simo Puntanen & George P. H. Styan, *Linear Algebra and its Applications*, vol. 237/238, pp. 455–477 (April 1996). [Linear Algebra and Statistics: In Celebration of C. R. Rao's 75th Birthday (10 September 1995): edited by R. B. Bapat, G. P. H. Styan & H. J. Werner.]

[14] [1997-GTR] "Orthogonal projectors", Solution (by Simo Puntanen & George P. H. Styan) to Problem 96.4.3 (proposed by Jürgen Groß & Götz Trenkler), *Econometric Theory*, vol. 13, no. 5, pp. 764–765 (October 1997).

[15] [1997-Guide2] *A Second Guide to Books on Matrices and Books on Inequalities, with Statistical and Other Applications*, by Simo Puntanen & George P. H. Styan, Preliminary Edition, in *A Bibliography and A Guide*, prepared for the Sixth International Workshop on Matrix Methods for Statistics, Istanbul, 16–17 August 1997, printed by the McGill University Printing Service, pp. 10–81 (August 1997).

[16] [1998-Guide3] *A Third Guide to Books on Matrices and Books on Inequalities, with Statistical and Other Applications*, by Simo Puntanen, Shane T. Jensen & George P. H. Styan, Preliminary Edition, in "Three Bibliographies and A Guide", printed by the Copy Center of Nova Southeastern University, Fort Lauderdale, Florida, pp. 27–100 (1998).

[17] [1998/ET-scaling] "A fundamental matrix result on scaling in multivariate analysis", Solution to Problem 97.5.3 (proposed by Heinz Neudecker, Albert Satorra & Michel van de Velden), by Simo Puntanen & George P. H. Styan, *Econometric Theory*, vol. 14, no. 5, pp. 693–695 (October 1998).

[18] [1998/ET-Gerald] "Mahalanobis distance for multinomial data", Solution to Problem 97.5.4 (proposed by Heinz Neudecker)], by Simo Puntanen, George P. H. Styan & Gerald E. Subak-Sharpe, *Econometric Theory*, vol. 14, no. 5, pp. 695–698 (October 1998).

[19] [1998-Calcutta] "Simplified matrix proofs related to the deletion of an observation in general linear model", by Markku Nurhonen, Simo Puntanen, George P. H. Styan & Haruo Yanai, in *Frontiers in Probability and Statistics* (S. P. Mukherjee, S. K. Basu & B. K. Sinha, eds.), pub. Narosa Publishing House, New Delhi, ISBN 81-7319-115-8, pp. 267–275 (1998). [One of 38 selected papers (out of 101) presented at the Second International Triennial Calcutta Symposium on Probability and Statistics (Calcutta, India, December 30, 1994–January 2, 1995).]

[20] [1999-ET5] "A determinantal identity", Solution (by Simo Puntanen, George P. H. Styan & Hans Joachim Werner) to Problem 98.4.1 (proposed by Heinz Neudecker & Michel van de Velden), *Econometric Theory*, vol. 15, no. 4, pp. 632–633 (August 1999).

[21] [2000-ET6] "The eigenvalue decomposition of a symmetric matrix", Solution No. 1 (by Simo Puntanen, George P. H. Styan & Hans Joachim Werner) to Problem 99.3.1 (proposed by R. W. Farebrother), *Econometric Theory*, vol. 16, no. 2, pp. 289–294 (April 2000). [Solution 2 by Geert Dhaene is on pp. 292–294; combined references are on page 294.]

[22] [2000-PSW/JSPI] "Two matrix-based proofs that the linear estimator Gy is the best linear unbiased estimator", by Simo Puntanen, George P. H. Styan & Hans Joachim Werner, *Journal of Statistical Planning and Inference*, vol. 88, no. 2, pp. 173–179 (1 August 2000).

[23] [2000-Elian] Letter to the Editor (by Simo Puntanen, George P. H. Styan & Hans Joachim Werner), about "Simple forms of the best linear unbiased predictor in the general linear regression model", by Silvia N. Elian [*The American Statistician*, vol. 54, pp. 25–28 (2000)], *The American Statistician*, vol. 54, no. 4, pp. 326–327 (November 2000).
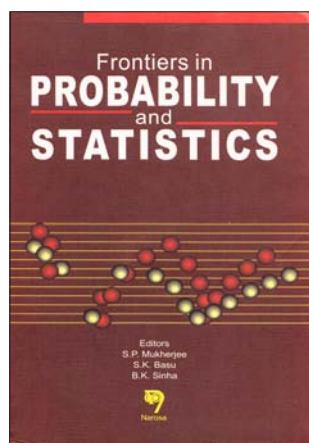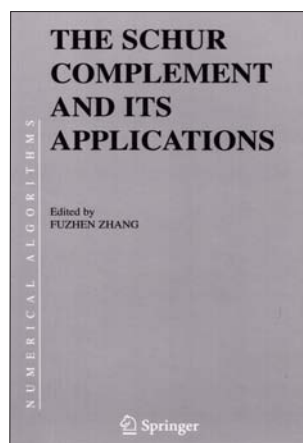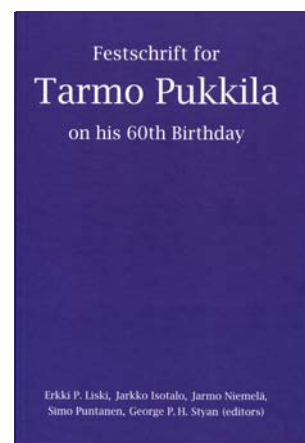
[19]   [27, 28]   [32, 33, 34, 35]

[24] [2001-ET7] "Determinant of a skew-symmetric matrix" by Simo Puntanen, George P. H. Styan & Hans Joachim Werner, Solution 1 (with Solution 2 by Geert Dhaene) to Problem 00.1.1 (proposed by Steve Lawford), *Econometric Theory,* vol. 17, no. 1, pp. 277–278 (February 2001).

[25] [2002-DLLPS] "Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments", by S. W. Drury, Shuangzhe Liu, Chang-Yu Lu, Simo Puntanen & George P. H. Styan, *Sankhyā: The Indian Journal of Statistics, Series A,* vol. 64, pp. 453–507 (June 2002). [Selected Articles from San Antonio Conference in Honour of C. R. Rao (June 2002).]

[26] [2004-Chips1] "On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model", by K. L. Chu, J. Isotalo, S. Puntanen & G. P. H. Styan, *Sankhyā: The Indian Journal of Statistics,* vol. 66, no. 4, pp. 634–651 (November 2004). Reprinted in [29].

[27] [2004-Schur1] "Historical introduction: Issai Schur and the early development of the Schur complement", by Simo Puntanen & George P. H. Styan, Chapter 0 & Bibliography in *The Schur Complement and Its Applications.* edited by Fuzhen Zhang, pub. Springer Science+Business Media, ISBN 978-0-387-50468-1, pp. 1–16 & 259–288 (2004).

[28] [2004-Schur2] "Schur complements in statistics and probability", by Simo Puntanen & George P. H. Styan, Chapter 6 & Bibliography in *The Schur Complement and Its Applications,* edited by Fuzhen Zhang, pub. Springer Science+Business Media, ISBN 978-0-387-50468-1, pp. 163–226 & 259–288 (2004).

[29] [2004-KLC/PhD] *Inequalities and equalities associated with ordinary least squares and generalized least squares in partitioned linear models,* by Ka Lok Chu, xiv + 240 pp (2004). MHB AS42 M3 2005 C499 Rare Books/Special Collections - Theses Collection (McLennan Bldg, 4th floor). Thesis includes reprints of [26, 30, 44, 48].

[30] [2005-Chips2] Ka Lok Chu, Jarkko Isotalo, Simo Puntanen & George P. H. Styan, "Some further results concerning the decomposition of of the Watson efficiency in partitioned linear models", *Sankhyā: The Indian Journal of Statistics,* vol. 67, no. 1, pp. 74–89 (February 2005). Reprinted in [29].

[31] [2005-PST/ET] "Three rank formulas associated with the covariance matrices of the BLUE and the OLSE in the general linear model", by Simo Puntanen, George P. H. Styan & Yongge Tian, *Econometric Theory,* vol. 21, no. 3, pp. 659–663 (June 2005).

[32] [2006-Tarmo4] *Festschrift for Tarmo Pukkila on his 60th Birthday,* edited by Erkki P. Liski, Jarmo Niemelä, Jarkko Isotalo, Simo Puntanen & George P. H. Styan, pub. Dept. of Mathematics, Statistics and Philosophy, University of Tampere, ISBN 978-951-44-6620-5, ISSN 1456-3134, 383 pp (2006).

[33] [2006-Tarmo1] "A conversation with Tarmo Mikko Pukkila", by Simo Puntanen & George P. H. Styan, in *Festschrift for Tarmo Pukkila on his 60th Birthday* [32, pp. 13–44 (2006)].

[34] [2006-Tarmo2] "Some comments on the research publications of Tarmo Mikko Pukkila", by Simo Puntanen & George P. H. Styan, in *Festschrift for Tarmo Pukkila on his 60th Birthday* [32, pp. 45–62 (2006)].

[35] [2006-Tarmo3] "On the role of the constant term in linear regression", by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, in *Festschrift for Tarmo Pukkila on his 60th Birthday* [32, pp. 243–259 (2006)].

[36] [2006-Schur/Laval] "Some comments about Issai Schur (1875–1941) and the early history of Schur complements", by Simo Puntanen & George P. H. Styan, invited talk [presented by George P. H. Styan] at Université Laval, Québec City (17 February 2006).

[37] [2006-Schur/Canberra] "Some comments about Issai Schur (1875–1941) and the early history of Schur complements", by Simo Puntanen & George P. H. Styan, in *Contributions to Probability and Statistics: Applications and Challenges: Proceedings of the International Statistics Workshop, University of Canberra, 4–5 April 2005,* edited by Peter Brown, Shuangzhe Liu & Dharmendra Sharma, pub. World Scientific, Singapore, ISBN 981-270-391-8, pp. 28–66 (2006).

[38] [2006-Tricks/Canberra] "Matrix tricks for linear statistical models: a short review of our personal top fourteen", by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, in *Contributions to Probability and Statistics: Applications and Challenges–Proceedings of the International Statistics Workshop, University of Canberra, 4–5 April 2005,* edited by Peter Brown, Shuangzhe Liu & Dharmendra Sharma, pub. World Scientific, Singapore, ISBN 981-270-391-8, pp. 113–128 (2006).

[39] [2007-Jarkko/PhD] *Linear estimation and prediction in the general Gauss–Markov model,* by Jarkko Isotalo, Acta Universitatis Tamperensis Series A, no. 1242, Ph.D. Thesis, Dept. of Mathematics, Statistics and Philosophy, University of Tampere. ix + 241 pp (2007).

[40] [2007/2013-Handbook1] "Random vectors and linear statistical models", by Simo Puntanen & George P. H. Styan, Chapter 52 in *Handbook of Linear Algebra,* edited by Leslie Hogben, pub. Chapman & Hall/CRC, Boca Raton, pp. 52.1–52.17 (2007); Chapter 70 in the Second edition, pp. 70.1–70.17 (2013).

[41] [2007/2013-Handbook2] "Multivariate statistical analysis", by Simo Puntanen, George A. F. Seber & George P. H. Styan, Chapter 53 in *Handbook of Linear Algebra,* edited by Leslie Hogben, pub. Chapman & Hall/CRC, Boca Raton, pp. 53.1–52.15 (2007); Chapter 71 in the Second edition, pp. 71.1–71.16 (2013).

[42] [2007-IJSS] "Effect of adding regressors on the equality of the OLSE and BLUE" by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, *International Journal of Statistical Sciences,* vol. 6, pp. 193–201 (2007). [Invited paper in the Second Special Issue in Felicitation of Professor Mir Masoom Ali on the Occasion of his 70th Birthday.]

[43] [2007-Hansbenno] "Many happy returns, George Peter Hansbenno Styan!", by Simo Puntanen, in *Program and Abstracts, MAT–TRIAD 2007: three days full of matrices* (Będlewo, Poland, 22–24 March 2007), edited by K. Filipiak, A. Markiewicz & W. Wolyński, pp. 85–87 (2007).

[44] [2007-Chips3] "The efficiency factorization multiplier for the Watson efficiency in partitioned linear models: some examples and a literature review", by Ka Lok Chu, Jarkko Isotalo, Simo Puntanen & George P. H. Styan, *Journal of Statistical Planning and Inference,* vol. 137, no. 11, pp. 3336–3351 (2007). Reprinted in [29]. [Invited paper in the Special Issue in Celebration of the Centennial of the Birth of Samarendra Nath Roy (1906–1964), edited by G. S. Mudholkar, A. D. Hutson & M. P. McDermott.]
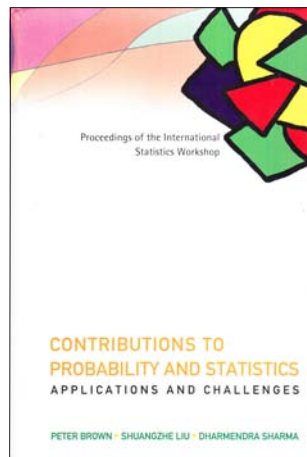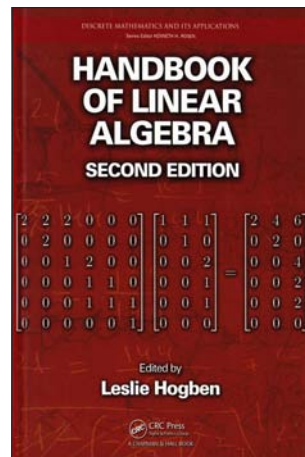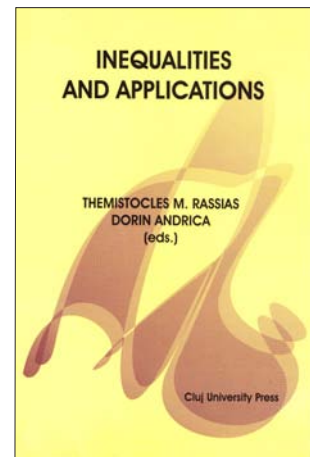
[28, 38]   [40, 41]   [48]

[45] [2008-IPS2] "A useful matrix decomposition and its statistical applications in linear regression", by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, *Communications in Statistics: Theory and Methods,* vol. 37, no. 9, pp. 1436–1457 (28 February 2008).

[46] [2008-JKB1] "Foreword [to the Special Issue in Honour of Jerzy K. Baksalary]", by Simo Puntanen & George P. H. Styan, *Discussiones Mathematicae: Probability and Statistics,* vol. 28, no. 1, pp. 85–90 (2008).

[47] [2008-JKB2] "Decomposing matrices with Jerzy K. Baksalary", by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, *Discussiones Mathematicae: Probability and Statistics,* vol. 28, no. 1, pp. 91–111 (2008).

[48] [2008-Chips4] "Inequalities and equalities for the generalized efficiency function in orthogonally partitioned linear models", in *Inequalities and Applications,* edited by Themistocles M. Rassias & Dorin Andrica, pub. Cluj University Press, Cluj, Romania, ISBN 978-973-610-793-1, pp. 13–69 (2008). Reprinted in [29].

[49] [2008-Tomar/address] "Some comments on T. W. Anderson's publications", by George P. H. Styan, special invited talk at the 17th International Workshop on Matrices and Statistics in Honour of T. W. Anderson's 90th Birthday, Tomar, Portugal (23–26 July 2008). Talk based on the handout "An annotated and illustrated bibliography with hyperlinks for T. W. Anderson in celebration of his 90th birthday" [50] by Simo Puntanen & George P. H. Styan, 12 pp. Beamer file, 35 pp (18 July 2008).

[50] [2008-Tomar/handout] "An annotated and illustrated bibliography with hyperlinks for T. W. Anderson in celebration of his 90th birthday", by Simo Puntanen & George P. H. Styan [handout for "Some comments on T. W. Anderson's publications" [49]], 12 pp (14 July 2008).

[51] [2008-Tomar/stamp] "17th International Workshop on Matrices and Statistics (IWMS-2008) in Celebration of T.W. Anderson's 90th Birthday: Tomar, Portugal (23–26 July 2008)", by Francisco Carvalho, Simo Puntanen & George P. H. Styan, *Image: The Bulletin of the International Linear Algebra Society,* no. 41, p. 8 (Fall 2008).

[52] [2008-StochStamps2] "Stochastic stamps: a philatelic introduction to chance", by Simo Puntanen & George P. H. Styan, *Chance,* vol. 21, no. 3, pp. 36–41 (September 2008).

[53] [2008-BLUE] "The BLUE's covariance matrix revisited: a review", by Jarkko Isotalo, Simo Puntanen & George P. H. Styan, *Journal of Statistical Planning and Inference,* vol. 138, no. 9, pp. 2722–2737 (2008). [Special Issue in Honor of Theodore Wilbur Anderson, Jr. on the Occasion of his 90th Birthday.]
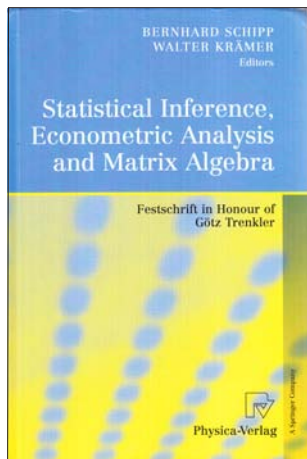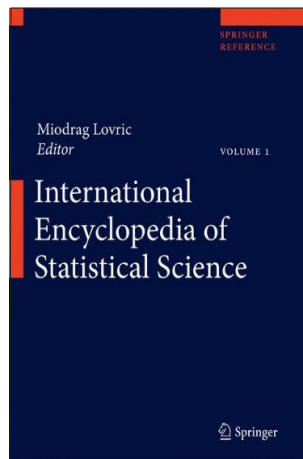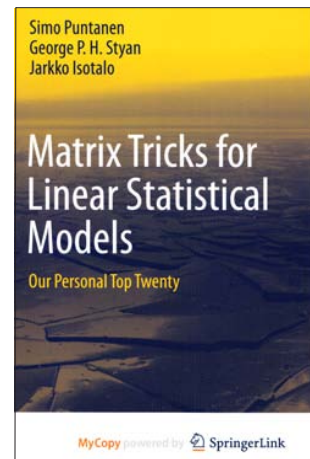
[56]   [59]   [60]

[54] [2009-ESI-1/statistical physics] "A philatelic introduction to statistical physics", by Simo Punta-nen & George P. H. Styan, *ESI News: Erwin Schrödinger Institute of Mathematical Physics,* vol. 4, no. 1, pp. 5–7 (Summer 2009).

[55] [2009-Calcutta6] "Some comments on the Watson efficiency of the ordinary least squares estimator under the Gauss–Markov model", by Jarkko Isotalo, Simo Puntanen & George Styan, *Calcutta Statistical Association Bulletin,* vol. 61, no. 241–244, pp. 1–15 (2009). ["Special 6-th Triennial Proceedings Volume" for the Sixth International Triennial Calcutta Symposium on Probability and Statistics (29–31 December 2006), edited by Nitis Mukhopadhyay & Manisha Pal.]

[56] [2009-BCPS/*Trenkler-Festschrift*] "Some comments on Fisher's α index of diversity and on the *Kazwini Cosmography*", by Oskar Maria Baksalary, Ka Lok Chu, Simo Puntanen & George P. H. Styan, in *Statistical Inference, Econometric Analysis and Matrix Algebra: Festschrift in Honour of Götz Trenkler,* edited by Bernhard Schipp & Walter Krämer, pub. Physica-Verlag, Heidelberg, pp. 369–394 (2009).

[57] [2009-Pakistan1] "Some comments on philatelic Latin squares from Pakistan", by Ka Lok Chu, Simo Puntanen & George P. H. Styan, *Pakistan Journal of Statistics*, vol. 25, no. 4, pp. 427–471 (October 2009).

[58] [2010-Pakistan2] "A note on the interpretation of the equality of OLSE and BLUE", by Augustyn Markiewicz, Simo Puntanen & George P. H. Styan, *Pakistan Journal of Statistics,* vol. 26, no. 1, pp. 127–134 (January 2010).

[59] [2011-IESS] "Best linear unbiased estimation in linear models", by Simo Puntanen & George P. H. Styan, in *International Encyclopedia of Statistical Science,* edited by Miodrag Lovric, pub. Springer, Section B (page 3 of 32), p-ISBN 978-3-642-04897-5, e-ISBN 978-3-642-04898-2, pp. 141–144 (2011). ["The 3 volumes (1671 pp) present contributions by 617 statisticians and educators from 105 countries, including 4 Nobel laureates." Reprinted and freely available in *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies.*"]

[60] [2011-*Tricks*] *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*, by Simo Puntanen, George P. H. Styan & Jarkko Isotalo, pub. Springer-Verlag, Heidelberg, p-ISBN 978-3-642-10472-5, e-ISBN 978-3-642-10473-2, xvii + 486 pp (31 August 2011).

[61] [2011-StatPapers] "Inverse and determinant of a special symmetric matrix", Solution (by Ka Lok Chu, Simo Puntanen & George P. H. Styan) to Problem 1/SP09 (proposed by Heinz Neudecker, Götz Trenkler & Shuangzhe Liu). *Statistical Papers,* vol. 52, no. 1, pp. 258–260 (February 2011).
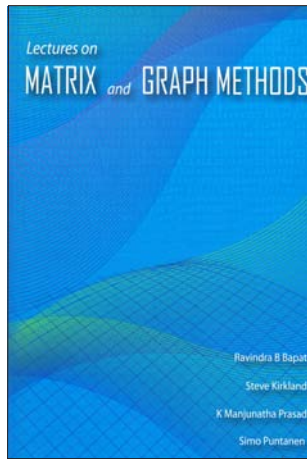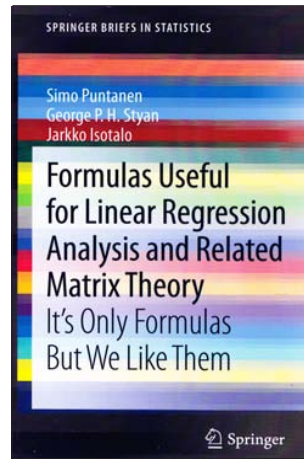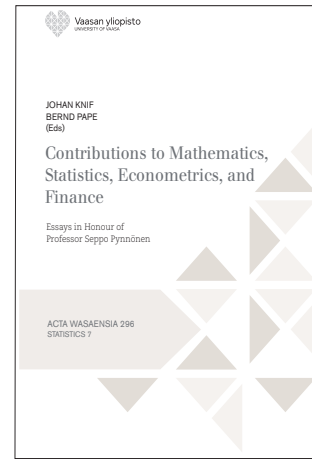
[62]   [63]   [64]

[62] [2012-SKM/Manipal] "A conversation with Sujit Kumar Mitra in 1993", by Simo Puntanen & George P. H. Styan, in *Lectures on Matrix and Graph Methods,* edited by Ravindra B. Bapat, Steve Kirkland, K. Manjunatha Prasad & Simo Puntanen, pub. Manipal University Press, ISBN 978-81-922759-6-3, pp. 221–244 (June 2012). [Lecture notes from the International Workshop and Conference on Combinatorial Matrix Theory and Generalized Inverses of Matrices, Manipal University, Manipal (Karnataka), India (2–7 & 10–11 January 2012).]

[63] [2013-Formulas] *Formulas Useful for Linear Regression Analysis and Related Matrix Theory: It's Only Formulas But We Like Them,* by Simo Puntanen, George P. H. Styan & Jarkko Isotalo, Springer Briefs in Statistics, pub. Springer, Heidelberg, ISBN 978-3-642-32930-2, xii+125 pp (2013):

[64] [2014-CPS/Pynnönen] "Properties of a special symmetric matrix associated with postage stamps", by Ka Lok Chu, Simo Puntanen & George P. H. Styan, in *Contributions to Mathematics, Statistics, Econometrics, and Finance: Essays in Honour of Professor Seppo Pynnönen,* edited by Johan Knif & Bernd Pape, Acta Wasaensia 296, Statistics 7, pub. University of Vaasa/Vaasan yliopisto, ISBN 978-952-476-523-7, pp. 33–41 (2014).

[65] [2014-KTF/conversation] "A conversation with Kai-Tai Fang", by Agnes W. L. Loie, Lucinda Li, Simo Puntanen & George P. H. Styan, in preparation, preprint 39 pp (22 March 2015). For publication in the *Souvenir Booklet of the 24th International Workshop on Matrices and Statistics (25-28 May 2015), Haikou City, Hainan, China,* edited by Jeffrey J. Hunter.

[66] [2015-IWMS] "Twenty-six years of the International Workshop on Matrices and Statistics (IWMS): 1990–2015", by Simo Puntanen & George P. H. Styan, in preparation, preprint 19 pp (14 April 2015). For publication in the *Souvenir Booklet of the 24th International Workshop on Matrices and Statistics (25-28 May 2015), Haikou City, Hainan, China,* edited by Jeffrey J. Hunter.

$$\mathbf{S} = \begin{pmatrix} 20 & 7 & 45 & 70 \\ 45 & 70 & 20 & 7 \\ 70 & 45 & 7 & 20 \\ 7 & 20 & 70 & 45 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 20 & 7 & 45 & 70 \\ 70 & 45 & 7 & 20 \\ 7 & 20 & 70 & 45 \\ 45 & 70 & 20 & 7 \end{pmatrix}; \quad m_{\mathbf{SP}} = 142 = 2 \times 71.$$

Acknowledgements

Unified Speed Estimation of Various Stabilities

Mu-Fa Chen[1,*]

[1] Beijing Normal University

Abstract   The main topic of this talk is the speed estimation of stability/instability. The word "various" comes with no surprising since there are a lot of different types of stability/instability and each of them has its own natural distance to measure. However, the adjective "unified" is very much unexpected. The talk surveys our recent progress on the topic, made in the past five years or so.

Keywords   Speed estimation; Stability; Birth–death process

In the next section, we introduce our first unified result: Theorem 1. Then, several extensions or generalizations of Theorem 1 are collected briefly in Section 2.

1   Basic estimates of the first non-trivial eigenvalue

Here is our first stability, the exponential stability in the ergodic case. Given a Markov chain on a countable E with transition probability $P(t) = (p_{ij}(t) : i, j \in E)$ $(t \geqslant 0)$, in the irreducible ergodic case, we have a stationary distribution π: $\pi P(t) = \pi$ for all $t \geqslant 0$. Then, we have

$$p_{ij}(t) \to \pi_j \quad \text{as} \quad t \to \infty \quad \text{for all } i, j.$$

We are now looking for the exponential convergence speed (rate) ε:

$$p_{ij}(t) - \pi_j = C_i e^{-\varepsilon t}, \qquad t \geqslant 0, \, i, j \in E.$$

Define the Q-matrix by

$$Q = (q_{ij} : i, j \in E) = \frac{\mathrm{d}}{\mathrm{d}t} P(t) \Big|_{t=0} \quad \text{(pointwise)}.$$

In the reversible case, we have $\varepsilon_{\max} = \lambda_1$, where λ_1 is the smallest (the first nontrivial) eigenvalue of $-Q$: $Qg = -\lambda g$ for some $g \neq$ constant.

Let us now consider a simpler birth–death Q-matrix on $E = \{0, 1, 2, \ldots\}$:

$$Q = \begin{pmatrix} -b_0 & b_0 & 0 & 0 & \ldots \\ a_1 & -(a_1 + b_1) & b_1 & 0 & \ldots \\ 0 & a_2 & -(a_2 + b_2) & b_2 & \ldots \\ \vdots & \ddots & \ddots & \ddots & \end{pmatrix},$$

where $a_k, b_k > 0$. Since the sum of each row equals 0, we have $Q\mathbb{1} = = 0 \cdot \mathbb{1}$, where $\mathbb{1}$ is the vector having elements 1 everywhere and is the zero vector. This means that the Q-matrix has a trivial eigenvalue $\lambda_0 = 0$ with eigenvector $\mathbb{1}$. Our question is what is the next eigenvalue λ_1 of $-Q$?

Actually, the story is much harder than it looks like, as shown in [3; pages 1–3], even for $E = \{0,1,2,3\}$. The reader is urged strongly to have some personal computation or have a look at the pages just mentioned.

We now show that the story is even much more complicated. Let $E = \{0,1,\ldots,N\}$ with $N < \infty$ for a moment. Consider the eigenvalue problem:

$$Qg = -\lambda g, \qquad g \neq 0$$

with Dirichlet boundary at 0: $g_0 = 0$ and Neumann boundary at N: $g_N = g_{N+1}$. Using codes 'D' and 'N', we may denote this minimal eigenvalue λ by λ^{DN}. Actually, the DN case is well studied in the history. Obviously, except the DN case, we should have three more cases: ND, DD, and NN. The last one, λ^{NN}, denotes the ergodic rate λ_1 just mentioned above, for which the constraint is not at the endpoints but is having mean zero.

In the non-ergodic case, the symmetric measure μ can not be finite. Hence, the exponential convergence rate is changed to be the exponential decay rate:

ergodic case : $p_{ij}(t) - \pi_j = C_i e^{-\varepsilon t}$, $t \geqslant 0$, $\varepsilon_{\max} = \lambda^{\mathrm{NN}}$;

non-ergodic case : $p_{ij}(t) = C_i e^{-\varepsilon t}$, $t \geqslant 0$, $\varepsilon_{\max} = \lambda^{\#}$, $\qquad i, j \in E$,

where # = DN, ND, or DD.

Altogether, there are four cases: NN, DD, DN, and ND.

To state our main result, we need a standard notion. Return to our general state space $E = \{0,1,\cdots,N\}$, $N \leqslant \infty$. Define

$$\mu_0 = 1, \quad \mu_n = \frac{b_0 b_1 \cdots b_{n-1}}{a_1 a_2 \cdots a_n}, \qquad 1 \leqslant n \leqslant N.$$

For general $N \leqslant \infty$, the principal eigenvalue $\lambda^{\#}$ defined above has to be extended to the largest λ satisfying

$$\sqrt{\lambda} \, \|f\|_{\mu,2} \leqslant \|\partial f\|_{\nu,2} \tag{1}$$

with one of the four boundary conditions, where $\|\cdot\|_{\mu,q}$ denotes the $L^q(\mu)$-norm and

$$\nu_i = \begin{cases} \nu_i^- = \mu_i a_i & i \leqslant \theta \\ \nu_i^+ = \mu_i b_i & \theta \leqslant i < N+1; \end{cases}$$

$$\partial_i f = \begin{cases} (\partial_i f)^- = f_{i-1} - f_i & i \leqslant \theta \\ (\partial_i f)^+ = f_{i+1} - f_i & \theta \leqslant i < N+1 \end{cases}$$

and $\theta \in E$ is a reference point.

The author started to study λ^{NN} in 1988 (cf. [2, 3]), but the following result (the first unified exponential rate estimation) was obtained in 2010 [5] only.

Theorem 1.
For the first non-trivial eigenvalue $\lambda^{\#}$ defined above, we have the following unified basic estimates:

$$\left(4\kappa^{\#}\right)^{-1} \leqslant \lambda^{\#} \leqslant \left(\kappa^{\#}\right)^{-1},$$

where

$$\left(\kappa^{\mathrm{NN}}\right)^{-1} = \inf_{n,m \in E, m < n} \left[\left(\sum_{i=0}^{m} \mu_i\right)^{-1} + \left(\sum_{i=n}^{N} \mu_i\right)^{-1}\right] \left(\sum_{j=m}^{n-1} \frac{1}{\mu_j b_j}\right)^{-1}$$

$$\left(\kappa^{\mathrm{DD}}\right)^{-1} = \inf_{n,m \in E, m \leqslant n} \left[\left(\sum_{i=0}^{m} \frac{1}{\mu_i a_i}\right)^{-1} + \left(\sum_{i=n}^{N} \frac{1}{\mu_i b_i}\right)^{-1}\right] \left(\sum_{j=m}^{n} \mu_j\right)^{-1}$$

$$\kappa^{\mathrm{DN}} = \sup_{n \in E} \left(\sum_{i=0}^{n} \frac{1}{\mu_i a_i} \right)^{-1} \left(\sum_{j=n}^{N} \mu_j \right)^{-1}$$

$$\kappa^{\mathrm{ND}} = \sup_{n \in E} \left[\left(\sum_{i=0}^{n} \mu_i \right)^{-1} \left(\sum_{j=n}^{N} \frac{1}{\mu_j b_j} \right)^{-1} \right].$$

In particular, $\lambda^{\#} > 0$ iff $\kappa^{\#} < \infty$.

Note that if we define $\hat{v}_k = (\mu_k b_k)^{-1}$, and in the DD and DN cases, under the sum $\sum_{k=-M}^{m}$, we modify $\hat{v}_k$ to be $(\mu_k a_k)^{-1}$ (noting that when $k \in E$, $\mu_k b_k = \mu_{k+1} a_{k+1}$), then the basic estimates given in the theorem can be described completely by two measures μ and $\hat{v}$. The upper and lower bounds are the same up to a universal constant 4 only. It is easy to see that the two endpoints 0 and N are symmetric in these two constants.

Finally, we mention that the DN and ND cases are known around 1970 in harmonic analysis, our main contribution is for the cases of DD and NN, especially the two isoperimetric constants κ^{NN} and κ^{DD} (come from [5; Corollaries 7.8 and 7.9]). In the proof of the DD and NN cases, three advanced mathematical tools are used and its proof given in [5] consists of five steps. Later, a direct elementary proof was found in [6]. It then leads to the study in the next section.

2 Generalizations

2.1 Bilateral case

Clearly, the birth-death process studied in the last section can be extended to the bilateral one with state space $E = \{i : -M - 1 < i < N + 1\}$, where $M, N \leqslant \infty$, and with evolution rates: $q_{i,i+1} = b_i$, $q_{i,i-1} = a_i$, and $q_{ij} = 0$ for other $j \neq i$, $i, j \in E$. In this case, the symmetric measure μ is defined as follows.

$$\mu_{\theta+n} = \frac{a_{\theta-1} a_{\theta-2} \cdots a_{\theta+n+1}}{b_\theta b_{\theta-1} \cdots b_{\theta+n}}, \qquad -M - 1 - \theta < n \leqslant -2,$$

$$\mu_{\theta-1} = \frac{1}{b_\theta b_{\theta-1}}, \qquad \mu_\theta = \frac{1}{a_\theta b_\theta}, \qquad \mu_{\theta+1} = \frac{1}{a_\theta a_{\theta+1}},$$

$$\mu_{\theta+n} = \frac{b_{\theta+1} b_{\theta+2} \cdots b_{\theta+n-1}}{a_\theta a_{\theta+1} \cdots a_{\theta+n}}, \qquad 2 \leqslant n < N + 1 - \theta.$$

where $\theta \in E$ is a reference point. In this bilateral case, Theorem 1 remains the same. Refer to [5].

2.2 Bilateral Hardy-type inequalities

Obviously, the Poincaré inequalities (1) can be generalized to the following

$$\|f\|_{\mu,q} \leqslant A^{\#} \|\partial f\|_{v,p}, \qquad f \in L^q(\mu) \tag{2}$$

for $p, q \in [1, \infty]$. This and the parallel inequalities with different boundary condition consist of the bilateral Hardy-type inequalities. When $q \geqslant p$, a generalization of Theorem 1 is given in [7] in the continuous context and in [14] in the discrete one.

2.3 Normed linear space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}}, \mu)$

In many applications (Sobolev inequalities, logarithmic Sobolev inequalities, Nash inequalities, and so on), the L^q-norm in (2) is not enough. This leads to the extension to a normed linear space $\mathbb{B}$ which is a linear subset of Borel measurable functions on (E, μ) with a specific norm $\|\cdot\|_{\mathbb{B}}$. In other words, instead of (2), we study the following Hardy-type inequalities

$$\||f|^q\|_{\mathbb{B}}^{1/q} \leqslant A_{\mathbb{B}}^{\#} \|\partial f\|_{v,p}, \qquad f \in \mathbb{B}$$

with different boundary conditions as before. Our result is presented in [1, 7]. For the last two topics, some popular reports are presented in [7–12].

2.4 Birth–death processes with killing

For the remainder of this section, we consider the birth–death processes with killing on $E = \{0, 1, 2, \dots, N\}$, $N \leqslant \infty$. Its Q-matrix becomes

$$
Q^c = \begin{pmatrix}
-(b_0 + c_0) & b_0 & 0 & 0 & \cdots \\
a_1 & -(a_1 + b_1 + c_1) & b_1 & 0 & \cdots \\
0 & a_2 & -(a_2 + b_2 + c_2) & b_2 & \cdots \\
\vdots & \vdots & \ddots & \ddots & \ddots
\end{pmatrix}
$$

with $a_i > 0$, $b_i > 0$, and $c_i \geqslant 0$ for every $i \in E$. Clearly, this is a special type of tridiagonal or Jacobi's matrix. Assume that $c_i \not\equiv 0$ on $(0, N)$, otherwise, we would return to Section 1. Even though the spectral problem becomes much harder than before, since a new sequence of parameter (c_i) is added, we are lucky to obtain a result in parallel to Theorem 1. Refer to [13, 9].

2.5 Discrete spectrum

We say that the matrix Q^c (or its quadratic form) on $L^2(\mu)$ has discrete spectrum if its spectrum consists of only eigenvalues with finite multiplicity. Since an operator on a finite space is compact and hence must have discrete spectrum, we need only consider an infinite state space. Next, since the whole line can be split into two half lines, without loss of generality, we assume that $E = \{0, 1, \dots\}$. In this subsection, we allow $c_i|_{(0, N-1)} \equiv 0$. This problem is solved completely by [9; Theorem 2.1], based on [13]. From the last cited paper, one finds interesting story on isospectral operators.

The references given below are only those the talk is based on. It is regretted that a large number of publications in the active research area is omitted here, otherwise, the list would be too long. For more references on the related sub-topics, the reader is urged to look at the related papers below.

References

[1] Chen, M.F. (2003). *Variational formulas of Poincaré-type inequalities for birth-death processes*. Acta Math. Sin. Eng. Ser. 19(4): 625-644.

[2] Chen, M.F. (2004). *From Markov Chains to Non-equilibrium Particle Systems*. World Scientific. 2nd ed. (1st ed., 1992).

[3] Chen, M.F. (2005). *Eigenvalues, Inequalities, and Ergodic Theory*. Springer, London.

[4] Chen, M.F. (2007). *Exponential convergence rate in entropy*. Front. Math. China, 2(3): 329–358.

[5] Chen M.F. (2010). *Speed of stability for birth–death processes*. Front Math China 5(3): 379–515.

[6] Chen, M.F. (2012). *Lower bounds of principal eigenvalue in dimension one*. Front. Math. China 7(4): 645–668.

[7] Chen, M.F. (2013a). *Bilateral Hardy-type inequalities*. Acta Math Sin Eng Ser. 29(1): 1–32.

[8] Chen, M.F. (2013b). *Bilateral Hardy-type inequalities and application to geometry*. Mathmedia 37(2): 12–32; Math. Bulletin 52(8/9) (in Chinese).

[9] Chen, M.F. (2014). *Criteria for discrete spectrum of 1D operators*. Commu. Math. Stat. 2: 279–309

[10] Chen, M.F. (2015a). *Criteria for two spectral problems of 1D operators* (in Chinese). Sci Sin Math, 44(1):

[11] Chen, M.F. (2015b). *The optimal constant in Hardy-type inequalities*. Acta Math. Sinica, Eng. Ser.

[12] Chen, M.F. (2015c). *Progress on Hardy-type inequalities*. Chapter 6 in the book "Festschrift Masatoshi Fukushima", eds: Z.Q. Chen, N. Jacob, M. Takeda, and T. Uemura, World Sci.

[13] Chen, M.F. and Zhang, X. (2014) *Isospectral operators*. Commu Math Stat 2: 17–32.

[14] Liao, Z.W. (2015). *Discrete Hardy Inequalities* (II). Preprint.

Mu-Fa Chen

School of Mathematical Sciences, Beijing Normal University, Laboratory of Mathematics and Complex Systems (Beijing Normal University), Ministry of Education, Beijing 100875, The People's Republic of China.

Home page: http://math.bnu.edu.cn/˜chenmf/main_eng.htm

Magic Squares and Postage Stamps

Ka Lok Chu[1,∗]   George P. H. Styan[2,†]

[1]Dawson College, Westmount (Québec), Canada
[2]McGill University, Montréal (Québec), Canada

Abstract   We present a philatelic introduction to magic squares, beginning with the well-known 4×4 Dürer magic square in *Melencolia I* (1514) by Albrecht Dürer (1471–1528) depicted in sheetlets from Aitutaki/Cook Islands (1986), Mongolia (1978), Djibouti (2007), and the Comores (1978). We also found a Franklin bent-diagonal magic square on a postage stamp from the USA (2006), which also shows a whirlwind and water spouts, from Benjamin Franklin's *Experiments and Observations on Electricity* (1769). On 9 October 2014, Macau Post issued a set of several philatelic items for magic squares [2]. Two stamps feature magic-square palindromes: the 5×5 Sator–Arepo (c. 100 AD) and a 15×15 detail from the 29×29 Su Hui palindrome (c. 357 AD). The set also includes a souvenir sheet (1) featuring the 3×3 Luoshu and a first-day cover (Fig. 13) with a 9×9 Hendricks diamond-inlaid magic square.

Keywords   Dürer, *Melencolia I*, Macau, magic squares, philatelic items, Luoshu

(1)

∗This file compiled on 15 April 2015. Email: gchu@dawsoncollege.qc.ca

†Corresponding author. Research supported, in part, by the Natural Sciences and Engineering Council of Canada. Email: geostyan@gmail.com

Figure 1: (left panel): Aitutaki/Cook Islands (1986); (right panel): Mongolia (1978).

Figure 2: (left panel): Djibouti (2007); (right panel): Comores (1978).

1 *Melencolia I* (1514) by Albrecht Dürer

We begin this talk, which is based in part on the recent talks [13, 14], see also [16], with the well-known 4×4 magic square in *Melencolia I* (1514) by Albrecht Dürer (1471–1528) depicted in sheetlets from Mongolia (1978), Aitutaki/Cook Islands (1986), and Djibouti (2007). The Dürer magic square appears in the detail (in the selvage, top right) of a sheetlet (Fig. 2) featuring Dürer's painting *Oswolt Krel* (1499) from the Comores (1978).

Figure 3: Two *Marke Individuell MaPhyPhil* from Germany (2013). On the left a detail from Dürer's *Melencolia I* (see also Fig. 4) and on the right a 5×5 bordered magic square by the German monk and mathematician Michael Stifel (1487–1578) from his *Arithmetica Integra* [12, (1544)].

Figure 4: 4.00 Ptcs stamp from Macau, China (2014).

The Dürer matrix $\mathbf{D}$ with magic sum $m = 34$ in Fig. 4 is Dudeney Type III in that

$$\mathbf{DF} + \mathbf{FD} = \tfrac{1}{2}m\mathbf{E} \qquad (2)$$

where $\mathbf{F}$ is the flip (reverse-identity) matrix and every entry of $\mathbf{E}$ is equal to 1 (see also Fig. 5). Moreover, all the entries in each of the top-left, top-right, bottom-left, bottom-right and centre-heart 2×2 submatrices sum to the magic sum $m = 34$.
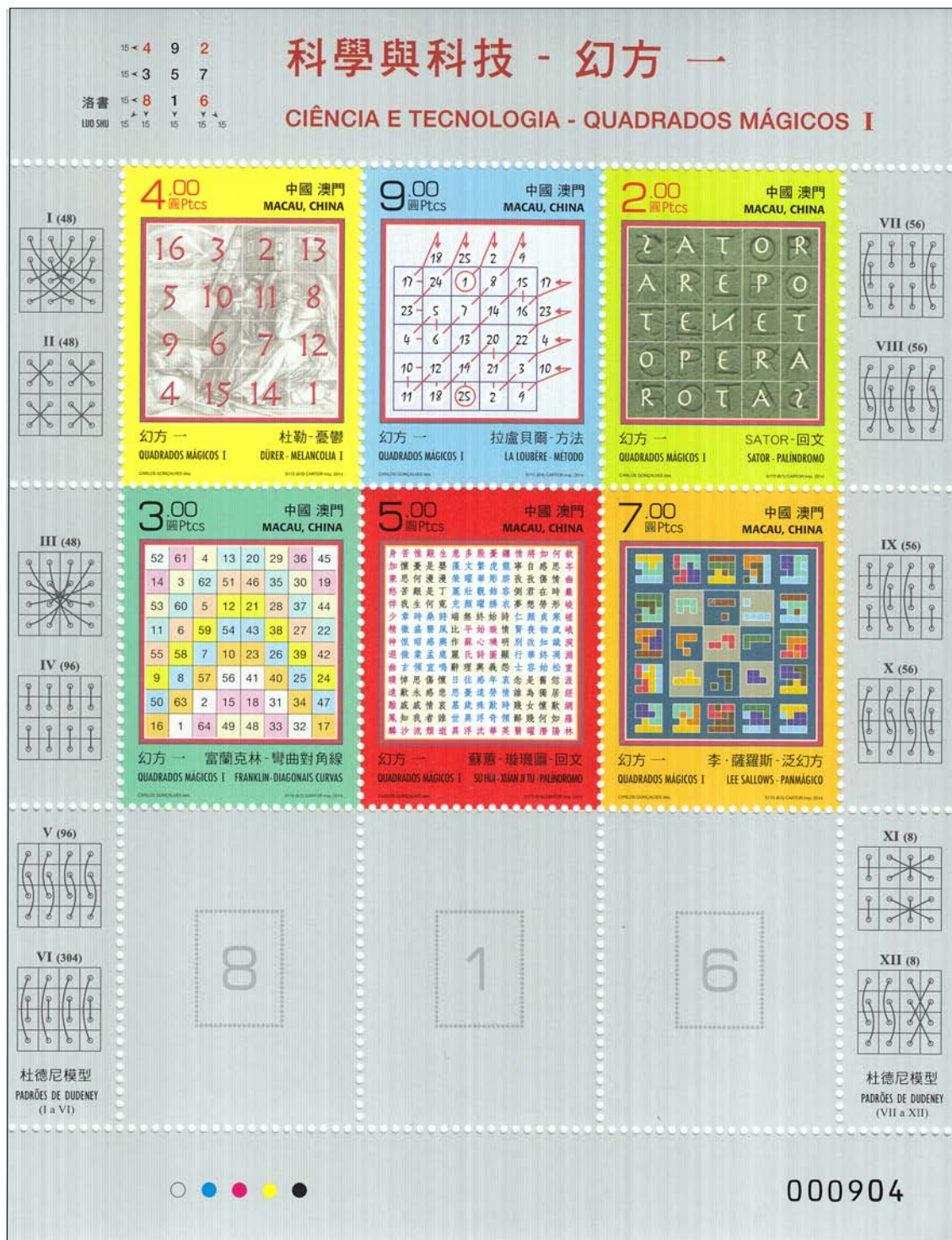
Figure 5: Souvenir sheet from Macau, China (2014) featuring 6 magic-square stamps and the 12 Dudeney groups [3] for 4 × 4 classic magic squares. See also Fig. 4, 7, 8, 11, 14. Stamps for 8 Ptcs, 1 Ptc and 6 Ptcs are scheduled to be issued by Macau Post in 2015.

2 Benjamin Franklin (1706–1790) & Simon de la Loubère (1642–1729)

Figure 6: Stamp from the USA 2006 (*Scott* 4022) with bent-diagonal Franklin magic square in a detail from *Water-spouts and Whirlwinds* (1753) by the postmaster, scientist, inventor, statesman, and diplomat Benjamin Franklin (1706–1790). See also Fig.7 (left panel). For more on bent-diagonal Franklin magic squares see the excellent book by Pasles [7].

Figure 7: (left panel) 3.00 Ptcs stamp with bent-diagonal Franklin magic square from Macau, China (2014), see also Fig. 6; and (right panel) 9.00 Ptcs stamp from Macau, China (2014) featuring a 5×5 magic square and its construction method as given by the French diplomat, writer, mathematician and poet Simon de La Loubère (1693) [4, p. 229]; see also Pherū [8, p. 171] and *Wikipedia* [11].

3   Su Hui (fl. c. 357 AD) and her poem *Xuan Ji Tu*

Figure 8: Shown on the stamp is the inner-central 15×15 detail of the 29×29 palindrome poem *Xuan Ji Tu,* literally "armillary-sphere map", by Su Hui (fl. c. 357 AD). See also Fig. 9.
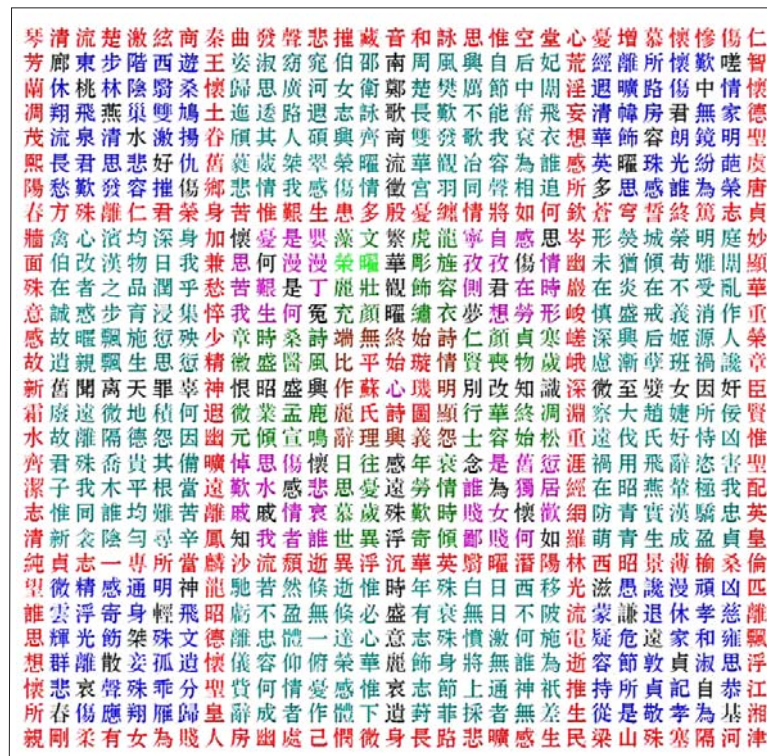
Figure 9: The complete 29×29 palindrome poem by Su Hui (c. 357 AD). See also Fig. 8.

4 Zhang Heng (78–139 AD) and his armillary sphere

Figure 10: Zhang Heng (78–139 AD) created the first armillary sphere (left panel) in c. 117 AD. He (right panel) was an astronomer, mathematician, inventor, geographer, artist, poet, statesman, and literary scholar of the Eastern Han Dynasty in ancient China.

Su Hui's poem *Xuan Ji Tu* is a palindrome that can be read in all directions. Such poems are called in Chinese *huiwen shi.* Five colours of silk threads were used. The poem contains $841 = 29 \times 29$ Chinese characters: there are 29 lines horizontally and vertically that can be read in any which way, it always makes a poem. This poem created a sensation in China, from Su Hui's time onward. The Tang empress Wu Zetian, herself a poet, made some 200 poems out of Su Hui's creation. In the Song Dynasty, one scholar said there were 10 diagrams in there and he could decipher 3,752.

Su Hui conceived of and then embroidered this "map" for her husband who was in another city and who had fallen for another woman. She sent *Xuan Ji Tu* to him. The husband understood her meaning, had her brought to him and they then spent the rest of their lives [happily?] together.

The poem is called *Xuan Ji Tu,* literally "armillary-sphere map" (Fig. 10). An armillary sphere (spherical astrolabe, armilla, or armil) is a model of objects in the sky (in the celestial sphere), consisting of a spherical framework of rings, centred on Earth, that represent lines of celestial longitude and latitude and other astronomically important features such as the ecliptic. As such, it differs from a celestial globe, which is a smooth sphere whose principal purpose is to map the constellations.

6   Lee Sallows's geometric magic square
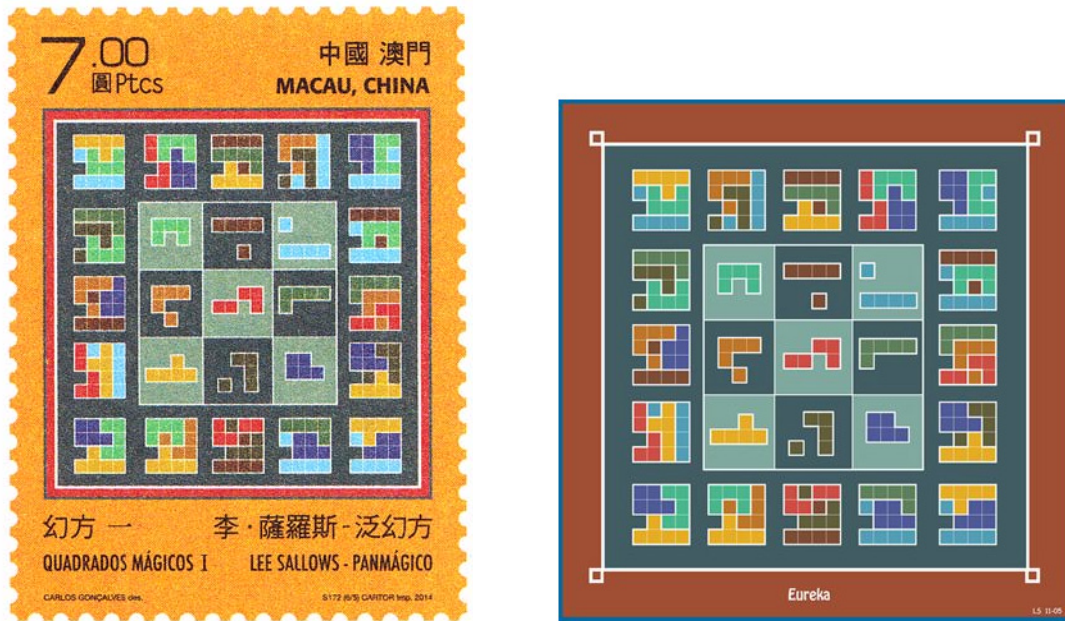
Figure 11: Sallows "Eureka" geometric magic square [9, Fig. 18 (2011)], [10, Fig 7.3 (2013)].

For more about the Sallows "Eureka" geometric magic square (geomagic) square (Fig. 11) see *The Mathematical Inteligncer* article [9] and the excellent recent Dover book *Geometric Magic Squares* [10].

5   Diamond-inlaid magic squares

Diamond-inlaid magic squares (see Fig. 12 and 13 below) were created c. 1991 by David M. Collison (1937–1991) and John Robert Hendricks (1929–2007). The Hendricks diamond-inlaid magic square (upper left panel here from the Macau Post First Day Cover) is a 9×9 bordered magic square with three inlaid magic squares, 1 each 3×3, 5×5, and 7×7. The 3×3 magic square is rotated 45 degrees and is called a "diamond inlay". We note that the smaller and larger numbers are mixed throughout the square, not just in the outside border as they would be with traditional bordered magic squares.
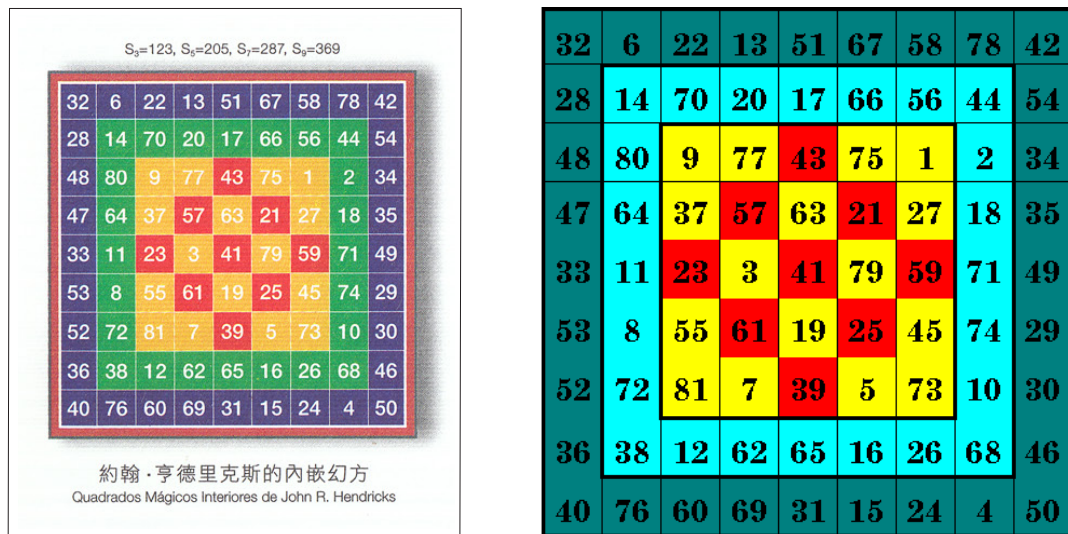
Figure 12: Hendricks diamond-inlaid magic square: (left panel) detail from Macau Post First-Day Cover (Fig. 13 below).
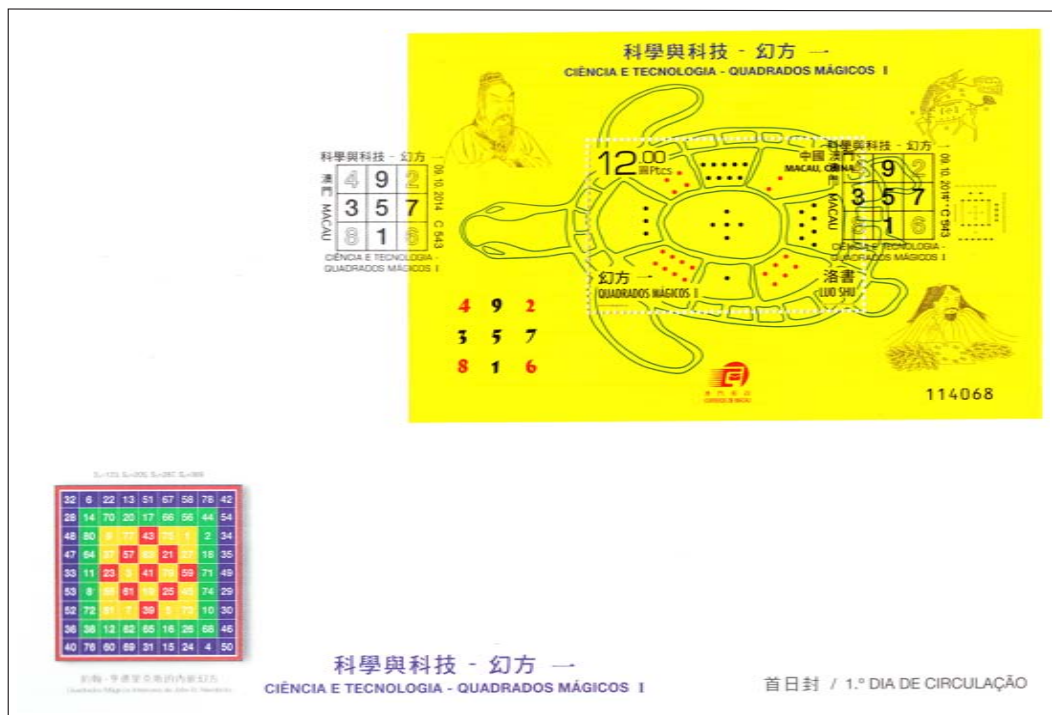
Figure 13: Macau Post First Day Cover (9 October 2014) featuring the 3×3 Luoshu magic square and a 9×9 Hendricks diamond-inlaid magic square. See also Fig. 12 above and the display (1).

7 Anton Friedrich Wilhelm von Webern (1883–1945) and the Sator-Arepo magic square

Figure 14: The Austrian composer Anton Friedrich Wilhelm von Webern (1883–1945) composed music using the Sator-Arepo magic square, which is inscribed on Webern's gravestone (left panel) in Mittersill (near Salzburg), Austria. The stamp (left panel) for Webern is from Austria 1995, *Scott* 1691. The 2.00 Ptcs Sator-Arepo stamp (right panel) is from Macau, China (2014). For more on the Sator-Arepo magic square, see Moeller [6].

An expanded version, based in part also on [1, 13, 14, 15, 16, 17], is in preparation.

Acknowledgements

References

[1] Baksalary, Oskar Maria; Chu, Ka Lok; Styan, George P. H. & Trenkler, Götz (2009). A philatelic introduction to some special matrices. Report 2009-04 from the Department of Mathematics and Statistics, McGill University, Montréal.

[2] Bellos, Alex (2014). Macau's magic square stamps just made philately even more nerdy. *The Guardian* (4 November 2014).

[3] Dudeney, Henry E. (1910). The magic square of sixteen. *The Queen, The Lady's Newspaper & Court Chronicle,* vol. 127, no. 3290, pp. 125–126 (15 January 1910).

[4] de la Loubère, Simon (1693/1986). *The Kingdom of Siam*, with an Introduction by David K. Wyatt. Oxford University Press.

[5] Miller, Jeff (2015). Images of Mathematicians on Postage Stamps, website by Jeff Miller, Gulf High School, New Port Richey, Florida: `http://jeff560.tripod.com/stamps.html` Accessed on 13 April 2015.

[6] Moeller, Walter O. (1973). *The Mithraic Origin and Meanings of the Rotas-Sator Square.* Leiden: Brill.

[7] Pasles, Paul C. (2008). *Benjamin Franklin's Numbers: An Unsung Mathematical Odyssey.* Princeton University Press.

[8] Pherū, Ṭhakkura (2009). *Gaṇitasārakaumudī: The Moonlight of the Essence of Mathematics*, in English and edited with Introduction, Translation & Mathematical Commentary, by SaKHYa [Sreeramula Rajeswara Sarma, Takanori Kusuba, Takao Hayashi and Michio Yano]. New Delhi: Manohar.

[9] Sallows, Lee C. F. (2011). Geometric magic squares. *The Mathematical Intelligencer,* vol. 30, no. 4, pp. 25–31.

[10] Sallows, Lee C. F. (2013). *Geometric Magic Squares: A Challenging New Twist Using Colored Shapes Instead of Numbers.* New York: Dover.

[11] Siamese method: `http://en.wikipedia.org/wiki/Siamese_method` online at *Wikipedia.*

[12] Stifel, Michael (1544). *Arithmetica Integra,* with a foreword by Philippi Melanchthonis, in Latin, pub. Norimbergæ apud Iohan. Petreium [Nürnberg].

[13] Styan, George P. H. (2007). An illustrated philatelic introduction to magic squares, talk presented (by George P. H. Styan) in the Session on Interdisciplinary research projects for undergraduates at the International Conference on Advances in Interdisciplinary Statistics and Combinatorics, Greensboro, North Carolina (13 October 2007).

[14] Styan, George P. H. (2014). Some comments on the set of philatelic items "Science and Technology: Magic Squares I" from Macau Post, talk presented at the Lakeshore Stamp Club, Pointe-Claire (Québec), Canada (16 October 2014).

[15] Styan, George P. H. (2014). "Magic Postal Square" puzzles with Victorian/Edwardian philatelic solutions and a philatelic introduction to Lewis Carroll and *Alice in Wonderland,* Report 2014-01 from the Department of Mathematics and Statistics, McGill University, Montréal.

[16] Styan, George P. H. (2014). An illustrated philatelic introduction to Albrecht Dürer (1471–1528) and his *Melencolia I* (1514). Report 2014-03 from the Department of Mathematics and Statistics, McGill University, Montréal.

[17] Styan, George P. H. (2015). An illustrated philatelic introduction to doubly-classic 6×6 bordered magic matrices and to 4×4 Plato-like magic talismans. In preparation, Report 2015-01 from the Department of Mathematics and Statistics, McGill University, Montréal.

Classification of Magic Squares of Order 4

Kai-Tai Fang[1,*]   Yuying Luo[2,†]   Yanxun Zheng[3,‡]

[1]Division of Science and Technology, BNU-HKBU United International College
[2]Department of Biostatistics, Georgetown University, Washington D.C., USA
[3]Otsuka Pharmaceutical Company in Princeton, NJ, USA

Abstract   With a history of more than 3000 years, magic squares still are mysterious in various aspects. We in this paper give a comprehensive review and study on classification of magic squares of order 4. There are a lot of studies on this topic. Several classification methods were proposed such as Anderson graph, Dudeney types, Frénicle - Amela pattern, transformation group, and so on. In this paper we propose two new ways for classification of magic squares of order 4. One is based on eigenvalues of the magic squares and another is to employ the theory of the majorization. The second consideration is new. Relationships among results by the different classification methods are given.

Keywords   Dudeney types, Eigenvalues of a matrix, Frénicle-quadset Magic squares, majorization theory

1   Magic Squares and Classification

A *semi-magic square* is an $n \times n$ matrix of numbers in which the sum of entries along each row and each column is a constant μ. If in addition the sum of entries along the main diagonals is the same constant μ_n, then the matrix is said to be a *magic square of order n* with *magic sum μ_n*. In this paper we always assume that elements in a magic square are n^2 consecutive integers, $\{1, 2, \ldots, n^2\}$. In this case the magic sum is $\mu_n = n(n^2+1)/2$. As a result the magic sum is $\mu_3 = 15$ for $n = 3$; $\mu_4 = 34$ for $n = 4$.

It is worldwide-accepted in the realm of mathematics that the first magic square was found by Chinese. With a history of more than 3000 years, magic squares still are mysterious and there are a lot of open questions to be solved. For example, how many possible different magic squares of order n are there? It has been pointed out: There are 8 magic squares for $n = 3$; 7040 magic squares for $n = 4$; more than 275×10^6 for $n = 5$. The number of magic squares of order n increases exponentially as n increases. An upper bound of the number of magic squares of order n is given by $(n^2)!/(8(2n+1)!)$ [23]. Therefore, classification of the magic squares becomes a very important issue.

In 1693 Bernard Frénicle de Bessy (1605-1675) published his results of enumerated 7040 magic squares of order 4 and classified them into 880 categories in his paper. And those results were confirmed by Friedrich Fitting (1862-1945) in 1931 (see [16]). Later, Ollerenshaw and Bondi [14] gave a comprehensive study on magic squares of order 4, and they used a different way to verify the number of magic squares of order 4 to be 7040.

For classification purpose some additional balance conditions in the literature are required. For example, (a) a *pandiagonal magic square* has all the broken (or co-)diagonals

---

with the same magic sum; and (b) a *regular or symmetric magic square* of order n is a magic square of order n satisfying $m_{ij} + m_{n-i+1,n-j+1} =$ constant, $i, j = 1, \ldots, n$, where m_{ij} are elements of the magic squares.

Because there is a huge number of magic squares of order $n, n \geq 5$. Many authors have focused on classification of the magic squares for $n = 3$ and 4. The classification of the magic squares of order 3 is simple. Table 1 list all the magic squares of order 3. Most authors suggest that two magic squares of order n should belong to the same class (*basic form*) if one square can be obtained from another square by anticlockwise rotation of $\pi/2$ and/or by reflection. Starting from M_1^3, we can obtain M_2^3 by the anticlockwise rotation of $\pi/2$. By the similar way we can get M_3^3, and then M_4^3. The square M_5^3 is the reflection of M_1^3, i.e., $M_5^3 = (M_1^3)^T$, then by a similar rotation to M_5^3 we can obtain M_6^3, M_7^3, and M_8^3. In the literature most authors confirm that all 8 magic squares of order 3 form one basic form. In general many authors have used the following definition:

Definition 1.
Two magic squares M_1 and M_2 of order n are said to belong to the same basic form if M_2 can be found by using one or more time operations h_1 and/or h_2, where

$$h_1(M) = M^T, \text{ where } M^T \text{ is the reflection (transpose) of } M,$$
$$h_2(M) \text{ is obtained by the anticlockwise rotation of } \pi/2.$$

Table 1: Eight Magic Squares of Order 3

4	9	2
3	5	7
8	1	6

M_1^3

2	7	6
9	5	1
4	3	8

M_2^3

6	1	8
7	5	3
2	9	4

M_3^3

8	3	4
1	5	9
6	7	2

M_4^3

4	3	8
9	5	1
2	7	6

M_5^3

8	1	6
3	5	7
4	9	2

M_6^3

6	7	2
1	5	9
8	3	4

M_7^3

2	9	4
7	5	3
6	1	8

M_8^3

There are some famous magic squares of order 4. Two below, i.e., M_1^4 and M_2^4, are generated by Yang Hui (1127-1279), a famous Chinese mathematician in Song Dynasty. He obtained a pair of magic squares for each $n, 3 \leq n \leq 8$ and called them as "Ying graph and Yang graph". Moessner in 1947 published a magic square, M_3^4 that has a lot of balance properties. In fact, this magic square can be obtained by exchanges of rows and columns of M_1^4, i.e., permuting columns of M_1^4 as $(1, 2, 3, 4) \Rightarrow (2, 4, 1, 3)$ and then exchanging first and fourth rows. Note, these operations do not appear in Definition 1.

$$M_1^4 = \begin{pmatrix} 4 & 9 & 5 & 16 \\ 14 & 7 & 11 & 2 \\ 15 & 6 & 10 & 3 \\ 1 & 12 & 8 & 13 \end{pmatrix} \quad M_2^4 = \begin{pmatrix} 2 & 16 & 13 & 3 \\ 11 & 5 & 8 & 10 \\ 7 & 9 & 12 & 6 \\ 14 & 4 & 1 & 15 \end{pmatrix} \quad M_3^4 = \begin{pmatrix} 12 & 13 & 1 & 8 \\ 6 & 3 & 15 & 10 \\ 7 & 2 & 14 & 11 \\ 9 & 16 & 4 & 5 \end{pmatrix}$$

Another famous magic square of order 4 (see M_4^4 below) appeared in Albrecht Dürer's Woodcut "Melancholia I". The square M_5^4 was found in some irons in relics of Xian,

China. Identification by the experts showed that there is a magic square of order 6 in ancient Arabic Numbers on the irons. According to more further researches, it is believed that this special order 6 magic square comes from its inner order 4 magic square. The last square M_6^4 was found in the Parshvanath Jain temple in Khajuraho, India.

$$M_4^4 = \begin{pmatrix} 16 & 3 & 2 & 13 \\ 5 & 10 & 11 & 8 \\ 9 & 6 & 7 & 12 \\ 4 & 15 & 14 & 1 \end{pmatrix} \quad M_5^4 = \begin{pmatrix} 8 & 11 & 14 & 1 \\ 13 & 2 & 7 & 12 \\ 3 & 16 & 9 & 6 \\ 10 & 5 & 4 & 15 \end{pmatrix} \quad M_6^4 = \begin{pmatrix} 7 & 12 & 1 & 14 \\ 2 & 13 & 8 & 11 \\ 16 & 3 & 10 & 5 \\ 9 & 6 & 15 & 4 \end{pmatrix}$$

For magic squares of order 4 their classification is not simple. Denote all 7040 magic squares of order 4 by $\mathcal{M}$. Considering the rotation and reflection transformations on $\mathcal{M}$, there are 880 basic forms. A complete list of these 880 basic forms can refer to [14]. Obviously, the above six famous magic squares do not belong to the same basic form.

In this paper we first give a comprehensive review on various methods for classification of 7040 magic squares of order 4. These methods can be sprit into graph visualization (Section 2), group of transformations (Section 3), eigenvalues of a magic square (Section 4), and enumeration of the Frénicle quadsets (Section 5). Secondly, we give relationships among different methods among which many results are new. We also apply the majorization theory to sort all the Frénicle quads and propose an open question that there is no magic square involved the quad (7,8,9,10). The latter is the largest guad in the sense of majorizarion. By a computational enumeration we show that the open question is correct. The final section gives conclusions.

2   Visualization Methods

The key characteristic of a magic square is its balance. Many earlier studies on classification of the magic squares are based on graphical balance of a magic square. The following graphes are proposed.

(A) *Frénicle–Amela patterns*: Frénicle (1931) put a magic square into a square with 16 subsquares. Consider all possible 4 neighborhood numbers add up to the magic sum 34. There are 5 kinds of pattern in Figure 1 with labels $\alpha, \beta, \gamma, \delta$, and ϵ. Amela (2009) gave some suggestion on Frénicle's consideration, see[3]. By a computation there are 48 basic forms in pattern α, 192 in β, 192 in γ, 328 in δ and 120 in ε.

(B) *Dudeney patterns*: Over 100 years ago Dudeney in his "Amusements in Mathematics" considered to connect two numbers with sum $17(=\mu/2)$ in a magic square of order 4. The corresponding two numbers are called as *complement* each other. Figure 2 shows the total 12 patterns, denoted by $I, II, \ldots, XII$.

Dudeney's classification has close relationships with many other considerations. For example, all the magic squares in pattern I are pandiagonal magic squares; the set of the regular magics squares of order 4 and the set of pattern III are the same. Trigg [18] pointed out that these 12 patterns can be further grouped into 4 *determinant types*: Type A includes I, II, and III patterns; patterns IV, V, and VI form type B; patterns $VII, VIII, IX$ and X form type C; and the remaining two patterns XI and XII form type D. He also showed that magic squares in each type can be obtained by some row/column exchanges. For example, interchange 2nd and 3rd rows, and then interchange 2nd and 3rd columns of a magic square in pattern II to get a magic square in pattern I; interchange 3rd and 4th rows, and then interchange 3rd and 4th columns of a magic square in pattern III to get a magic square in pattern I.
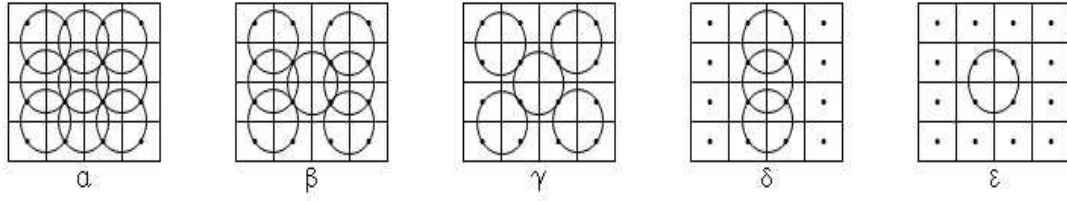
Figure 1: Frénicle - Amela pattern
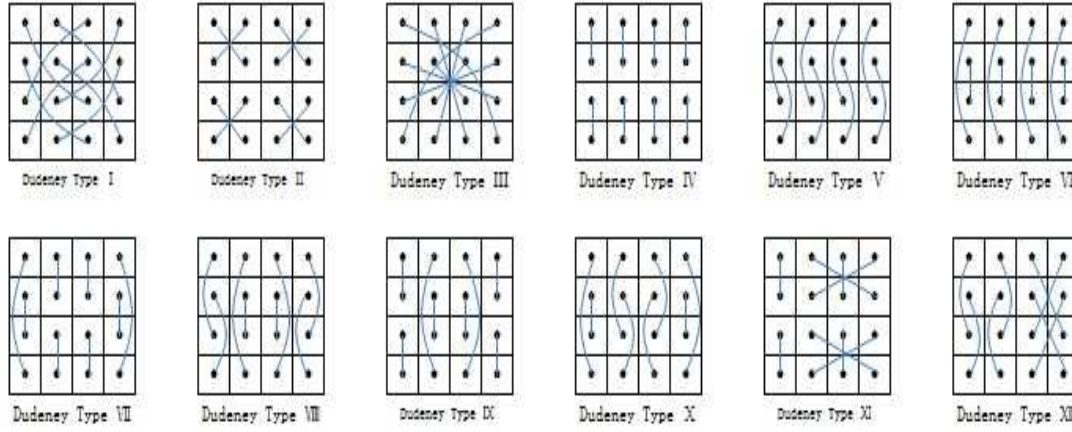
Figure 2: 12 Dudeney patterns

Table 2 gives relationships between Frénicle–Amela and Dudeney classifications, more details can refer to [3]. Note that each Dudeney pattern belongs to one and only one Frénicle - Amela pattern except pattern VI. For example, the 48 basic magic squares which belong to Frénicle-Amela α are all in Dudeney pattern I, magic squares in Dudeney pattern IV are all classified into Frénicle–Amela β, but However, the 304 magic squares in Dudeney pattern VI are not in the same category based on Frénicle–Amela classification. This means that the two methods are mostly consistent, with minor inconsistency.

(C) *Anderson graph*: F.J. Anderson (1918) suggested the graph produced by joining the consecutive numbers $1, 2, \ldots, 16$ in sequence including the line joining 16 and 1. By visualization Anderson graphs can be grouped as two classes: symmetric and unsymmetric. Anderson graph has been applied to many fields, but it does not involve a rich information

Table 2: Relation between Frénicle–Amela and Dudeney classifications

Dudeney		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Frénicle–Amela	α	48												48
	β				96		96							192
	γ		48	48		96								192
	δ						208	56		56		8		328
	ε								56		56		8	120
Total		48	48	48	96	96	304	56	56	56	56	8	8	880
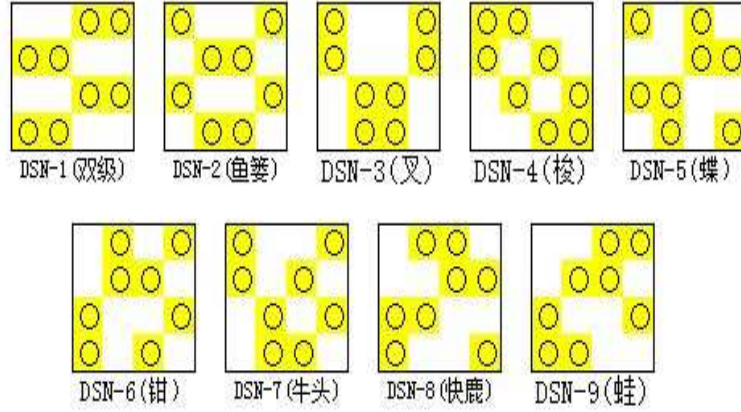
Figure 3: Distributions of Smaller Numbers of Magic Squares of order 4

for the classification purpose.

(D) *Distribution of small numbers*: For a given magic square of order 4, we split the entrances into two parts, the numbers of $\{1, \ldots, 8\}$ are called as smaller numbers and remaining are called larger numbers. These two sets should have a good balance geometrically. It is obvious that for any one of the smaller numbers, there must be another one and only one in the larger numbers, with which their sum is 17, half of the magic sum. Xu [21] proposed a plot of the distribution of smaller numbers and find there are only 9 types, see Figure 3. According to [21], at least two smaller numbers appear in each row and in each column, and at least one in any of diagonals while there are 4 in total in both diagonals.

However, Xu did not give the classification details of his method. Table 3 gives the classification of 7040. We calculate a contingency table between the smaller numbers distribution and Frénicle–Amela pattern as well as a contingency table between the smaller numbers distribution and Dudeney pattern and find that there is no clear relationships.

Table 3: Classification Based on Distribution of Small Numbers

Type	1	2	3	4	5	6	7	8	9
No	192	212	212	48	48	42	42	42	42

3 Transformation Group

The above graph methods for classification only pay attention to geometric balance of a magic square. Many authors have tried to use more algebraic tools for classification. The group theory is one of such powerful tools.

Definition 2.
Let $\mathcal{M}$ be a set of all the magic squares of order 4. Consider a set of transformations, $\mathcal{G}$ say, from $\mathcal{M}$ into itself, i.e., for each $g \in \mathcal{G}$ and $M \in \mathcal{M}$, we have $g(M) \in \mathcal{M}$. We call $\mathcal{G}$ to be a group of transformations if it satisfies two conditions: 1) if $g_1 \in \mathcal{G}$ and $g_2 \in \mathcal{G}$, then $g_1 g_2 \in \mathcal{G}$, where $g_1 g_2$ is defined as $g_1 g_2(M) = g_1(g_2(M))$; 2) if $g \in \mathcal{G}$, then $g^{-1} \in \mathcal{G}$ where g^{-1} satisfies $g g^{-1} = g^{-1} g = e$, with e being the identity transformation in $\mathcal{G}$.

By using the group of transformations for classification we need more concepts.

Definition 3.

Two magic square matrices $M_1, M_2 \in \mathcal{M}$ are say to be isomorphic under $\mathcal{G}$ if there exists a $g \in \mathcal{G}$ such that $M_2 = gM_1$. We write $M_1 \sim M_2 \pmod{\mathcal{G}}$.

It is easy to see that the isomorphic relation has the following properties: 1) $M \sim M \pmod{\mathcal{G}}$; 2) $M_1 \sim M_2 \pmod{\mathcal{G}}$ implies $M_2 \sim M_1 \pmod{\mathcal{G}}$; and 3) $M_1 \sim M_2 \pmod{\mathcal{G}}$ and $M_2 \sim M_3 \pmod{\mathcal{G}}$ implies $M_1 \sim M_3 \pmod{\mathcal{G}}$. The set $\{g(M)|g \in \mathcal{G}\}$ is called the orbit of M under $\mathcal{G}$. Clearly, two orbits are either identical or disjoint, and the orbits form a partition of $\mathcal{M}$.

Xu and Zhang [20] considered two groups of transformations, denoted by $\mathcal{H}$ and $\mathcal{N}$, for $\mathcal{M}$ classification. The group $\mathcal{H}$ involves two basic transformations: h_1 is the reflection against the main diagonal, i.e., $h_1(M) = M^T$ is the transpose of M; and h_2 is a rotation of anticlockwise $\pi/2$. Let

$$P_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Now $h_2(M) = P_1 M^T$. Based on these two basic transformations we can define other 6 transformations: $h_3 = h_2^2, h_4 = h_2^3, h_5 = h_1 h_2, h_6 = h_1 h_2^2, h_7 = h_1 h_2^3$, and $h_8 = e$, the identity transformation. There are relationships $h_2^4 = h_1^2 = e$ so that $\mathcal{H}$ is a group transformations on $\mathcal{M}$ with 8 transformations in each orbit. So there are $880 = 7040/8$ basic forms (orbits) in $\mathcal{H}$. This conclusion is consistent with that in the literature.

The group $\mathcal{N}$ involves four basic transformations $n_1 = h_1, n_2 = h_2$, n_3 is defined as: exchanging rows 1 and 4 as well as exchanging columns 1 and 4 of a magic square and n_4 is defined as exchanging rows 1 and 3, rows 2 and 4; as well as exchanging columns 1 and 3 and columns 2 and 4. Let

$$P_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

It is easy to see that $n_3(M) = P_2 M P_2^{-1}$ and $n_4(M) = P_3 M P_3^{-1}$. Based on these 4 basic transformations one can easily obtain a group with 32 transformations on $\mathcal{N}$, i.e.,

$$n_2^4 = n_1^2 = n_3^2 = n_4^2 = e, \tag{1}$$

$$n_1 n_2 = n_2^{-1} n_1, n_1 n_3 = n_3 n_1, n_1 n_4 = n_4 n_1, \tag{2}$$

$$n_2 n_3 = n_3 n_2, n_2 n_4 = n_4 n_2, n_4 n_3 = n_2^2 n_3 n_4. \tag{3}$$

Given one magic square $M \in \mathcal{M}$, we can obtain 31 magic squares that are isomorphic to M. There are 220 basic forms (orbits) under $\mathcal{N}$. But [20] did not give classification relations among $\mathcal{N}$ and Frénicle–Amela and Dudeney classifications. We shall study this issue later.

Furthermore, we find some differences between the groups $\mathcal{H}$ and $\mathcal{N}$ in the sense of eigenvalues and we propose some consideration of the classification based on eigenvalues.

4 Classification Based on Algebraic Properties

A matrix has many useful characteristic numbers, like rank, determinant, trace, eigenvalues, etc. Several authors have discussed some algebraic properties of magic squares, for example, [1], [2], [6], [9], [10],[13], [15], [17], and [18].

Definition 4.
Let $f(M)$ be a function of $M \in \mathcal{M}$. If we can find a partition of $\mathcal{M} = \mathcal{M}_1^f \cup \mathcal{M}_2^f \cup \cdots \cup \mathcal{M}_m^f$, where $f(M)$ has the same value on each $\mathcal{M}_i^f$ and $f(M_1) \neq f(M_2)$ for $M_1 \in \mathcal{M}_i^f, M_2 \in \mathcal{M}_j^f, i \neq j$. The partition $\{\mathcal{M}_1^f, \mathcal{M}_2^f, \cdots, \mathcal{M}_m^f\}$ gives a classification of $\mathcal{M}$ and we call them as f-basic forms.

The function f can be chosen as rank, determinant, trace, eigenvalues of $M \in \mathcal{M}$. Trevor [17] and Trigg [18] studied determinant of magic squares. Cleve Moler (2012) posted some results on rank of magic squares on the webpage (http://blogs.mathworks.com/cleve/2012/11/12/magic-squares-part-3-linear-algebra/). But his study is uncomplete.

As the determinant and rank are functions of the eigenvalues. Let us consider $f(M)$ to be the eigenvalues of a magic square M. There are n eigenvalues and related eigenvectors. It is obvious that the magic sum μ is an eigenvalue and $\mathbf{1} = (1, \ldots, 1)'$ is the corresponding eigenvector. Khan[9] proved that the μ is the largest eigenvalue in the sense of the absolute value of an eigenvalue. Magic squares in the same f-basic form have the same eigenvalues.

However, for the eight magic squares of order 3 in Table 1, there are two sets of eigenvalues: $\{15, 0 + 4.899i, 0 - 4.899i\}$ and $\{15, 4.899, -4.899\}$, where $i = \sqrt{-1}$ is the imaginary unit. Four magic squares have eigenvalues $\{15, 0+4.899i, 0-4.899i\}$, their $\pi/2$-rotation matrices have eigenvalues $\{15, 4.899, -4.899\}$. It seems that the 8 magic squares of order 3 should be sprit into two basic forms, not one basic form.

Therefore, the transformation h_2 should be modified as a rotation of anticlockwise π instead of $\pi/2$, denoted by this transformation by h_2^*. It is easy to find $h_2^*(M) = P_1 M P_1$. According to this modification we should define a new group of transformations, $\mathcal{Q}$ say, that involves four basic transformations $q_1 = h_1, q_2 = h_2^*, q_3 = n_3$ and $q_4 = n_4$ used in the groups of $\mathcal{H}$ and $\mathcal{N}$. Based on these 4 basic transformations one can obtain a group with 16 transformations: $q_5 = q_1 q_2, q_6 = q_1 q_3, q_7 = q_1 q_4, q_8 = q_2 q_3, q_9 = q_2 q_4, q_{10} = q_3 q_4, q_{11} = q_1 q_2 q_3, q_{12} = q_1 q_2 q_4, q_{13} = q_1 q_3 q_4, q_{14} = q_2 q_3 q_4, q_{15} = q_1 q_2 q_3 q_4$, and $q_{16} = e$. Note that $P_i^{-1} = P_i, i = 1, 2, 3$ and we can easily find the inverse of $q_j, j = 1, \ldots, 16$. So $\mathcal{Q}$ is a group of transformations with 7,040/16=440 orbits (basic forms) in $\mathcal{M}$. We shall show that this new transformation group $\mathcal{Q}$ is consistent with the classification by the eigenvalues.

Now, let us consider classification of 7040 magic squares of order 4 based on the eigenvalues. A criterion for magic square classification is reasonable if it is related some kind of balance of the magic square. Can we find some balance of a magic square of order 4 related to its eigenvalues? Let $M = (m_{ij})$ be a magic square of order 4. Its characteristic polynomial is

$$|\lambda I - M| = c_4(M)\lambda^4 + c_3(M)\lambda^3 + c_2(M)\lambda^2 + c_1(M)\lambda + c_0(M) = \prod_{i=1}^{4}(\lambda - \lambda_i), \qquad (4)$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are eigenvalues of M. It is easy to see the following facts, where

$c_i(M)$'s are in (4):

$$
\begin{aligned}
c_4(M) &= 1; \\
c_3(M) &= -(a_{11} + a_{22} + a_{33} + a_{44}) = -\mu = -\sum_{i=1} \lambda_i; \\
c_2(M) &= \text{the sum of all 2 order principal minors of } M = \sum_{i \neq j} \lambda_i \lambda_j; \\
c_1(M) &= -(\text{the sum of all 3 order principal minors of } M) = -\sum_{1 \leq i < j < k \leq 4} \lambda_i \lambda_j \lambda_k; \\
c_0(M) &= \text{the determinant of } M = \prod_{i=1}^{4} \lambda_i.
\end{aligned}
$$

The above formulas show some kinds of balance of M and relationships between the eigenvalues and the coefficients of the characteristic polynomial. It is reasonable to use the eigenvalues of $M \in \mathcal{M}$ for classification.

However, if you use the command of 'eig" in Matlab, due to the rounded rule two different magic squares with the same eigenvalues may get different eigenvalue outputs. It is better to directly use the coefficients of the characteristic polynomial for classification. As $c_4(M) = 1, c_3(M) = \mu = 34$ for all $M \in \mathcal{M}$, let $\mathbf{c}(M) = (c_2(M), c_1(M), c_0(M))$ that can be used for classification for $\mathcal{M}$.

Definition 5.
Two magic squares $M_1, M_2 \in \mathcal{M}$ are said to belong to the same E-class if $\mathbf{c}(M_1) = \mathbf{c}(M_2)$, where E-class emphasizes "eigenvalue".

It is easy to know from the coefficients of the characteristic polynomial whether there exists zero eigenvalue or not. When $c_0 = 0$, the function can be written as

$$(\lambda^3 - 34\lambda^2 + c_2\lambda + c_1)\lambda = 0,$$

which means one of λ's must be 0. Similarly, when $c_0 = c_1 = 0$, there must be two of λ's equal to 0; when $c_0 = c_1 = c_2 = 0$, three of λ's are 0; and when $c_0 = c_1 = c_2 = c_3 = 0$, all the eigenvalues are 0. By a calculation the space $\mathcal{M}$ can be spitted into 140 E-basic forms (Table 4) as follows:

(1) Group E_4: there are 108 E-basic forms with 4 non-zero different eigenvalues. This group can be further divided into two sub-groups: E_{4r}, where all the 4 eigenvalues are real; and E_{4c}, where there are 2 eigenvalues to be complex numbers. These two sub-groups involve 34 and 74 E-basic forms, respectively.

(2) Group E_3: there are 31 E-basic forms with one zero-eigenvalue and three non-zero eigenvalues, among which 21 E-basic forms with all real eigenvalues and 10 E-basic forms with two complex eigenvalues. Denote these two sub-groups by E_{3r} and E_{3c}, respectively. Obviously, each magic square in E_3 has a rank 3 and $c_0(M) = 0$.

(3) Group E_{30}: there are 1 E-basic forms with three zero eigenvalues and a nondiagonal Jordan form

$$
JF = \begin{pmatrix} 34 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.
$$

Each magic square in this group has a rank 3.

Now we discuss relationships between a function $f(M)$ on $\mathcal{M}$ and the groups $\mathcal{H}, \mathcal{N}$ and $\mathcal{Q}$.

Table 4: The number of magic squares and E-basic forms based on the eigenvalues

Category	Degenerate			Non-degenerate	
	E_3			E_4	
	E_{30}	E_{3r}	E_{3c}	E_{4r}	E_{4c}
Numbers of magic squares	192	3168	1760	576	1344
Numbers of E-basic forms	1	21	10	34	74

Definition 6.
A function $f(M)$ on $\mathcal{M}$ is said to be consistent with group $\mathcal{G}$ if $f(g(M)) = f(M)$ for each $M \in \mathcal{M}$ and each $g \in \mathcal{G}$; otherwise inconsistent.

Consider the following three functions related to the rank, determinant, of eigenvalues:

$f_1(M) = $ the rank of M;

$f_2(M) = |M|$, the determinate of M ;

$f_e(M) = 1, 2, 3, 4$ or 5 if $M \in E_{30}, E_{3r}, E_{3c}, E_{4r}$ or E_{4c}, respectively,

we have the following theorem.

Theorem 1.
We have the following conclusions:

(1) The functions $f_1(M)$ and $f_2(M)$ are consistent with groups $\mathcal{H}$ and $\mathcal{N}$. But the function $f_e(M)$ is inconsistent with groups $\mathcal{H}$ and $\mathcal{N}$.

(2) The functions $f_1(M), f_2(M)$ and $f_e(M)$ are consistent with groups $\mathcal{Q}$.

Proof. It is known that M and M^T have the same eigenvalues. Therefore, $f_i(h_1(M)) = f_i(M), i = 1, 2, 3$. Note that the matrices P_1, P_2 and P_3 are permutation matrices that have the property of $P_i^{-1} = P_i, i = 1, 2, 3$. As $h_2(M) = P_1 M^T$, $h_2(M)$ and M have the same rank and the same f_2-value. Other six transformations in $\mathcal{H}$ are generated by these two basic transformations h_1 and h_2, we conclude that the functions $f_1(M)$ and $f_2(M)$ are consistent with groups $\mathcal{H}$.

For the group $\mathcal{N}$, as $n_1 = h_1, n_2 = h_2$, $n_3(M) = P_2 M P_2 = P_2 M P_2^{-1}$, and $n_4(M) = P_3 M P_2 = P_3 M P_3^{-1}$, the functions $f_i(M), i = 1, 2, 3$ are invariant under n_1, n_3, n_4, but the function f_3 is not consistent under the transformation n_2 as M and $h_2(M) = n_2(M)$ may have different eigenvalues. Table 4 shows many such cases. By a similar statement we conclude the conclusion (1).

The group $\mathcal{Q}$ replace $h_2 = n_2$ by $q_2(M) = P_1 M P_1$ that does not change eigenvalues of M. Hence we have conclusion (2). □

Mattingly[13] pointed out that even order regular magic squares are singular. The magic squares of order 4 are of even order magic squares. It is easy to find that Dudeney pattern III of magic squares are regular and is a subset of E_{3r}.

Now, let us see the relation of classification results between f_3-basic forms and Frénicle–Amela Patterns (See Table 5). It can be easily seen that the magic squares with the first three patterns α, β, γ are degenerated, while those with the last pattern ε must be non-degenerated. Only the pattern δ have more complicated cases. It suggests that we should look at the pattern δ deeper.

Table 6 shows relationships between f_3-basic forms and Dudeney patterns. The two classifications are consistent in a certain sense:

Table 5: Classification relation between f_3-basic forms and Frénicle–Amela Patterns

Degenerated/Non-degenerated		Degenerated			Non-degenerated	
Rank		3			4	
No. of Zero Eigenvalues		3	1		0	
Real/Complex		Real	Real	Complex	Real	Complex
Frénicle–Amela Patterns	α	64	160	160	0	0
	β	0	768	768	0	0
	γ	128	704	704	0	0
	δ	0	1536	128	288	672
	ϵ	0	0	0	288	672

Table 6: Relationships between f_3-basic forms and Dudeney types

Degenerated/Non-degenerated		Degenerated			Non-degenerated	
Rank		3			4	
No. of Zero Eigenvalues		3	1		0	
Real/Complex		Real	Real	Complex	Real	Complex
Dudeney Types	$I \sim III$	64	160	160	0	0
	$IV \sim V$	0	384	384	0	0
	VI	0	1920	512	0	0
	$VII \sim X$	0	0	0	136	312
	$XI \sim VII$	0	0	0	16	48

(a) Each magic square in Dudeney patterns I, II and III are degenerated with rank 3. There are same numbers of magic squares (64,160,160) from patterns I, II and III appear in E_{30}, E_{3r} and E_{3c}, respectively.

(b) Patterns IV and V have the same frequency in E_{3r} adn E_{3c}. But pattern VI has a different behavior. We should pay more attention on this pattern.

(c) All the patterns $VII - XII$ have the full rank 4. The frequency in E_{4r} and E_{4c} of Patterns $VI, VII, VIII$ are the same (136 and 312).

(d) Patters XI and XII appear the same numbers in E_{4r} and E_{4c}.

5 Frénicle–quadset and Majorization Theory

A magic square M of order n involves $2n + 2$ n-vectors (n rows, n columns and 2 diagonal vectors) with the same sum μ_n. Sort these vectors in descending order and denote the set of the n row vectors and n column vectors of M by $\mathcal{V}(M)$ and the set of $2n + 2$ n-vectors by $\mathcal{V}^+(M)$. For example, the Ying square M_1^4 has

$$\mathcal{V}(M_1^4) = \{(16,9,5,4),(14,11,7,2),(15,10,6,3),(13,12,8,1),$$
$$(15,14,4,1),(12,9,7,6),(11,10,8,5),(16,13,3,2)\}$$

and

$$\mathcal{V}^+(M_1^4) = \mathcal{V}(M_1^4) \cup \{(13,10,7,4),(16,11,6,1)\}.$$

In this view we propose a concept of equivalence of two magic squares.

Definition 7.
Two magic squares M_1 and M_2 of order n are said to be equivalent if $\mathcal{V}^+(M_1) = \mathcal{V}^+(M_2)$, and are said to be weakly equivalent if $\mathcal{V}(M_1) = \mathcal{V}(M_2)$.

Let $\Omega = \{1, 2, \ldots, 16\}$ and

$$\mathcal{A}_4 = \{(x_1, x_2, x_3, x_4) : x_i \in \Omega, x_1 > x_2 > x_3 > x_4, x_1 + x_2 + x_3 + x_4 = 34\}.$$

Each vector in $\mathcal{A}_4$ is called as the Frénicle quad by [14] and the set $\mathcal{A}_4$ is called as the Frénicle quadset. Obviously, when $n = 4$, $\mathcal{V}^+(M)$ is a subset of $\mathcal{A}$ for any $M \in \mathcal{M}$. [14] found that there are 86 vectors in $\mathcal{A}_4$ and gave a comprehensive studies on these vectors.

Obviously, the concept of the weak equivalence can be applied to semi-magic squares. For simplicity, when $n = 4$ we use the same notations $\mathcal{V}(M)$ and $\mathcal{V}^+(M)$. It is easy to prove that each basic transformations under $\mathcal{H}, \mathcal{N}$ or $\mathcal{Q}$ applying to a $M \in \mathcal{M}$ does not change $\mathcal{V}^+(M)$. Therefore, we have the following theorem.

Theorem 2.
Any two $M_1, M_2 \in \mathcal{M}$ in the same orbit either under $\mathcal{H}, \mathcal{N}$ or $\mathcal{Q}$ must be equivalent.

By an enumerative algorithm we find that there are 86 vectors in $\mathcal{A}_4$ list in Table 7. These vectors are the same as [14]. Denote them by $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_{86}$. Let $\mathbf{a} = (a_1, a_2, a_3, a_4) \in \mathcal{A}_4$. If some permutation of $\mathbf{a}$ appear in a row of some M, we say $\mathbf{a}$ to appear in row; similarly we can define $\mathbf{a}$ appear in column and diagonal. We find the following facts:

(a) If $\mathbf{a}$ appears in row of a M, it must appear in column of another magic square; vice versa, if $\mathbf{a}$ appears in column of a M, it must appear in row of another magic square. Therefore, for a given $\mathbf{a} \in \mathcal{A}_4$ the number of magic squares where $\mathbf{a}$ appears in row equals to the number of magic squares where $\mathbf{a}$ appears in column.

(b) If $\mathbf{a}$ appears in main diagonal of a M, it must appears in anti-diagonal in another magic square. So we shall only say $\mathbf{a}$ appears in diagonal.

(c) There are 15 vectors in $\mathcal{A}_4$ each appearing in row/column, but not appearing in diagonal; there are 34 vectors on $\mathcal{A}_4$ each appearing in diagonal, not in row/column. Table 8 gives details.

(d) There is only one vector, $\mathbf{a}_{86} = (10, 9, 8, 7)$, that does not appear in row/coloum, or in diagonals.

For given two vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ in the same dimension, we can not sort them in general. If $a_i \geq 0, b_j \geq 0, i = 1, \ldots, n$, i.e., $\mathbf{a}, \mathbf{b} \in R^n_+$ and $a_1 + \ldots + a_n = b_1 + \ldots + b_n$, the majorization theory proposes the way to compare these two vectors in a certain sense. The majorization theory is an powerful tool in many fields. A comprehensive study on the majorization theory can refer to [12]. In a magic squares all the elements are positive and all row-, column- and diagonal-sums are the same. Perhaps we can employ the majorization theory to study magic squares. First, let us introduce some basic concepts in the majorization theory.

Definition 8.
For two vectors $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$ with positive elements and the same sum ($\sum_{i=1}^n a_i = \sum_{j=1}^n b_j = \mu$), we say that $\mathbf{b}$ weakly majorizes $\mathbf{a}$ and write $\mathbf{a} \prec^w \mathbf{b}$, if $\sum_{i=k}^n a_{(i)} \leq \sum_{i=k}^n b_{(i)}$, for $k = 1, \ldots, n$, where $a_{(1)} \geq \ldots \geq a_{(n)}$ and $b_{(1)} \geq \ldots \geq b_{(n)}$ are reordered of $\mathbf{a}$ and $\mathbf{b}$ in descending order, respectively. If there exists k such that $\sum_{i=k}^n a_{(i)} < \sum_{i=k}^n b_{(i)}$, we say that $\mathbf{b}$ majorizes $\mathbf{a}$ and write $\mathbf{a} \prec \mathbf{b}$.

Kai-Tai Fang,　Yuying Luo,　Yanxun Zheng

Table 7: 86 vectors in $\mathcal{A}_4$ sort by the majorization order

No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
	16	16	16	15	16	15	16	15	14	16	15	14	16	15	14	13	14	13	12	16	15	16	15	16	15	14	16	15	14
	15	14	13	14	12	13	11	12	13	10	11	12	9	10	11	12	10	11	11	13	14	12	13	11	12	13	10	11	12
	2	3	4	4	5	5	6	6	6	7	7	7	8	8	8	8	9	9	10	3	3	4	4	5	5	5	6	6	6
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
No	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
	16	15	14	13	15	14	13	13	12	16	15	14	16	15	14	16	15	14	13	16	15	14	13	14	13	12	12	16	15
	9	10	11	12	9	10	11	10	11	11	12	13	10	11	12	9	10	11	12	8	9	10	11	9	10	11	10	9	10
	7	7	7	7	8	8	8	9	9	4	4	4	5	5	5	6	6	6	6	7	7	7	7	8	8	8	9	5	5
	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
No	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	
	14	13	16	15	14	13	15	14	13	12	13	12	11	16	15	14	13	12	14	13	12	12	11	13	12	11	11	10	
	11	12	8	9	10	11	8	9	10	11	9	10	10	7	8	9	10	11	8	9	10	9	10	8	9	10	9	9	
	5	5	6	6	6	6	7	7	7	7	8	8	9	6	6	6	6	6	7	7	7	8	8	7	7	7	8	8	
	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	6	6	6	6	7	

Table 8: The number of vectors in $\mathcal{A}$ that appear in row, column and/or diagonal

	Appear in column or row	Not appear in column or row	Total
Appear in diagonal	36	34	70
Dot appear in diagonal	15	1	16
Total	51	35	86

Example 1.
Given two vectors $\mathbf{a} = (15, 2, 16, 1)$ *and* $\mathbf{b} = (9, 7, 10, 8)$ *in* $\mathcal{A}$*. They have the same sum and their descending vectors are* $\mathbf{a}^{\downarrow} = (16, 15, 2, 1)$ *and* $\mathbf{b}^{\downarrow} = (10, 9, 8, 7)$*. We have* $\sum_{i=2}^{n} a_{(i)} = 15 + 2 + 1 = 18 < \sum_{i=2}^{n} b_{(i)} = 9 + 8 + 7 = 24$, $\sum_{i=3}^{n} a_{(i)} = 2 + 1 = 3 < \sum_{i=3}^{n} b_{(i)} = 8 + 7 = 15$, $\sum_{i=4}^{n} a_i = 1 < \sum_{i=4}^{n} b_{(i)} = 7$*. Therefore,* $\mathbf{a} \prec \mathbf{b}$*.*

Consider two other vectors $\mathbf{c} = (15, 14, 4, 1)$ *and* $= (16, 12, 5, 1)$ *in* $\mathcal{A}$*. They are in descending form and their partial sums are* $\sum_{i=2}^{n} c_{(i)} = 14 + 4 + 1 = 19 > \sum_{i=2}^{n} d_{(i)} = 12 + 5 + 1 = 18$, $\sum_{i=3}^{n} c_{(i)} = 4 + 1 = 5 < \sum_{i=3}^{n} d_{(i)} = 5 + 1 = 6$, $\sum_{i=4}^{n} c_i = 1 = \sum_{i=4}^{n} d_{(i)} = 1$*. These two vectors can not be ordered.*

Applying the majorization order to the vectors in $\mathcal{A}$ we can partially sort these 86 vectors and denote them. For example, we find following relationships among $\mathbf{a}_1, \ldots, \mathbf{a}_9$ (from left to right):

$$
\begin{bmatrix} 16 \\ 15 \\ 2 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 16 \\ 14 \\ 3 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 16 \\ 13 \\ 4 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 15 \\ 14 \\ 4 \\ 1 \end{bmatrix} \begin{bmatrix} 16 \\ 12 \\ 5 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 15 \\ 13 \\ 5 \\ 1 \end{bmatrix} \begin{bmatrix} 16 \\ 11 \\ 6 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 15 \\ 12 \\ 6 \\ 1 \end{bmatrix} \prec \begin{bmatrix} 14 \\ 13 \\ 6 \\ 1 \end{bmatrix},
$$

i.e., $\mathbf{a}_1 \prec \mathbf{a}_2 \prec \mathbf{a}_3 \prec \mathbf{a}_4$, $\mathbf{a}_5 \prec \mathbf{a}_6$, $\mathbf{a}_7 \prec \mathbf{a}_8 \prec \mathbf{a}_9$. Here, $\mathbf{a}_4$ and $\mathbf{a}_5$ as well as $\mathbf{a}_6$ and $\mathbf{a}_7$ can not be ordered. We also find that $\mathbf{a}_4 \prec \mathbf{a}_j$ for $j > 5$ and $\mathbf{a}_6 \prec \mathbf{a}_j$ for $j > 7$. Therefore, we can say that $\mathbf{a}_4$ partially majorizes $\mathbf{a}_5$ and denote it as $\mathbf{a}_4 \prec^p \mathbf{a}_5$. Similarly we can write $\mathbf{a}_6 \prec^p \mathbf{a}_7$. As a result we have

$$\mathbf{a}_1 \prec \mathbf{a}_2 \prec \mathbf{a}_3 \prec \mathbf{a}_4 \prec^p \mathbf{a}_5 \prec \mathbf{a}_6 \prec^p \mathbf{a}_7 \prec \mathbf{a}_8 \prec \mathbf{a}_9. \tag{5}$$

Fortunately, all the 86 vectors in $\mathcal{A}$ can be sorted by the majorization order or partially majorization order. Table 9 lists $\mathbf{a}_i$'s according to the sort (5). This is a new result.

Furthermore, the most majorized vector $\mathbf{a}_{86} = (10, 9, 8, 7) \equiv \mathbf{m}$ does not appear in any magic square of order 4. The four elements $10, 9, 8, 7$ in $\mathbf{a}_{86}$ are nearest to the center point $(34/4, 34/4, 34/4, 34/4)$ among all vectors in $\mathcal{A}$. The key characteristic of magic squares is the balance, but the most balance vector $\mathbf{m}$ is not accepted by any $M \in \mathcal{M}$. This is an interesting fact.

6 Conclusion remark

Classification of magic squares has been studied by many authors in more than 100 years. The paper gives a comprehensive review. There are two kinds of methods for classification: visualization methods including Anderson graph, Frénicle - Amela pattern, Dudeney types, and small number pattern and algebraic methods including transformation group, eigenvalues and related functions, and majorization theory. The latter is newly proposed in this paper. We give a detailed discussion on relationships among the above classification methods. We point out that the transformation by a rotation of anticlockwise $\pi/2$ is not consistent with the eigenvalues consideration and propose a new transformation group. The methods mentioned in this paper can be extended to classification of magic square of order 5 or order 6.

We have mentioned 6 famous magic squares in section 1. The reader would be interesting how about their classification results in the above methods. Table 9 presents some partial results. We can see that all of them have rank 3 and are singular. But from the eigenvalue point of view there are different: Yang graph, Melancholia I and Arabic square have real eigenvalues, Ying graph and India square have a pair of complex eigenvalues, and Moessner square has 3 zero eigenvalues. The first 4 squares in the table belong to Frénicle–Amila III, but the last two squares belong to different Frénicle–Amila patterns VI and I, respectively. Similarly, the the 4 squares belong to γ type, i.e., they are symmetric. According to the small number method the first 4 except Moessner square belong to the same class, but the last two belong to different classes. However, these 6 squares belong to different basic forms under $\mathcal{H}, \mathcal{N}$ and $\mathcal{Q}$ groups.

Table 9: Classification to six famous magic squares

	Name	Small Number	Dudeney type	FA pattern	E type	rank
M_1^4	Ying graph	2	III	γ	E_{3c}	3
M_2^4	Yang graph	2	III	γ	E_{3r}	3
M_3^4	Moessner	3	III	γ	E_{30}	3
M_4^4	Melancholia I	2	III	γ	E_{3r}	3
M_5^4	Arabic	5	VI	δ	E_{3r}	3
M_6^4	India	1	I	α	E_{3c}	3

Acknowledgement

References

[1] Amir-Moéz, A.R. and Fredricks, G.A. (1984), Characteristic polynomials of magic squares, *Math. Mag.*, **57**, 220-221.

[2] Chu, K.L. (2007), Some comments on magic squares with three nonzero eigenvalues, invited talk at the 16th International Workshop on Matrices and Statistics, Windsor, Ontario, Canada.

[3] Chu, K.L and Styan,G. P. H. (2013). Some illustrated comments on Anderson graphs and Greek mythology tables for classic magic squares, presentation at The 22nd International Workshop on Matrices and Statistics (IWMS-2013), Toronto, Canada.

[4] Clifford A. P. (2003). *The Zen of Magic Squares, Circles, and Stars.* Princeton University Press, USA.

[5] Dudeney H. E. (1951). *Amusements in Mathematics.* London: Nelson.

[6] Essen, A.V.D, (1990), Magic squares and linear algebra, *Americal Math. Monthly*, **97**, 60-62.

[7] Fang K. T., Zheng Y. X. (2013). Magic Intergration of Mathematics and Culture ‑ Magic squares, *Mathematical Culture*, **4(3)**, 52-65 (in Chinese).

[8] Fang K. T., Zhang Y. T. (1990). *Generalized Multivariate Analysis.* Springer-Verlag & Science Press.

[9] Khan, N.A. (1957), Characteristic roots of semi-magic matrices, *Americal Math. Monthly*, **64**, 261-263.

[10] Loly, P., Cameron, I., Trump, W. and Schindel, D. (2009), Magic square spectra, *Linear Algebra and its Applications*, **430**, 2659-2680.

[11] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth, *Publications of the American Statistical Association* , **9(70)**, 209-219

[12] Marshall A. W., Olkin I.(1980). Inequalities: Theory of Majorization and Its Applications, *Academic Press*, ISBN 978-0-12-473750-1

[13] Mattingly R. B. (2000). Even order regular magic squares are singular, *American Math. Monthly*, **107**, 777-781.

[14] Ollerenshaw, D.K. and Bondi H. (1982), Magic squares of order four, *Phil. Trans. R. Sac. London*, **A 306**, 443-532

[15] Styan G. P. H. (2007), An illustrated (philatelic) introduction to magic matrices and statistics, with special emphasis on magic matrices having 3 nonzero eigenvalues, Invited talk at the Fifth Annual Meeting of the Statistical Society of Canada, St. Johns Newfoundland, Canada.

[16] Styan G. P. H., Chu K. L.(2013). Some illustrated comments on "magic squares in the country" with an extensive annotated, hyperlinked, and indexed bibliography. *Personal communication.*

[17] Trevor, J.E., (1924), Determinants whose arrays are magic squares, *Americal Math. Monthly*, **31**, 216-222.

[18] Trigg, C.W. (1948), Determinants of fourth order magic squares, *Americal Math. Monthly*, **55**, 558-561.

[19] Wu H. l. (2008). *Magic Squares and Prime Numbers: Two Classic Problems in Recreational Mathematics.* Beijing, China: Sciences Press (in Chinese).

[20] Xu D. D., Zhang X. B. (2008). Transformation groups in order 4 magic squares, *Journal of Nanjing Normal University (Natural Science Edition)*, Vol.31, No.4, pp.26-28 (in Chinese).

[21] Xu Z. Y. (1998). Patterns of number structures in order 4 magic squares, *Yan'an Educational Journal*, Vol.23, No.2, pp.81-83.

[22] Zheng Y. X. (2013). *Math on the Tortoiseshell - Magic Square*, Final Year Project Thesis, BNU-HKBU United International College.

[23] Ward, James E. III. (1980) Vector spaces of magic squares, *Mathematics Magazine* **53 (March)**: 108-11.

Conics Fitting by Least Squares

Eva Fišerová[1,*]   Sandra Donevska[1,†]

[1]Department of Mathematical Analysis and Applications of Mathematics,
Faculty of Science, Palacký University, 17. listopadu 12, 771 46 Olomouc, Czech Republic

Abstract   For the least squares methodology can be distinguished two main approaches of fitting conics (circle, ellipse, parabola, hyperbola), the algebraic and geometric fit. We focus on the best geometric fit based on minimization of distances from the observed data points to the fitted curve by the linear model approach with nonlinear restrictions on its parameters. Through this approach by an iterative estimation procedure being proposed in [12] are gained locally best linear unbiased estimates of the unknown algebraic conic parameters and also estimates of their accuracy. Subsequently, the geometric parameters and their accuracy can be also estimated. Furthermore, simulation study regarding the accuracy of the estimates of the geometric conics parameters are also presented.

Keywords   Conics fitting; Least squares; Geometric Fit; Accuracy

1   Introduction

Fitting quadratic curves (circle, ellipse, hyperbola and parabola) to given data points in the plane is a fundamental task in many fields like engineering, astronomy, physics, biology, quality control, image processing, etc.

For the least squares methodology can be distinguished two main approaches of fitting conics, the algebraic and geometric fit. The algebraic fit corresponds with the problem of minimization squares of algebraic distances from each given point (x_i, y_i) to the curve being described by the implicit equation

$$F(x_i, y_i, \boldsymbol{\theta}) = Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F = 0.$$

Thus the minimization problem in sense of the algebraic fit can be written as

$$\sum_{i=1}^{n}(Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F)^2 \to \min.$$

This approach is computationally simple and quick, nevertheless, the estimators are biased and not very accurate, see. e.g. [3, 11, 13, 14, 15]. Another disadvantage is that the algebraic fitting is not invariant to the coordinate transformations.

The other least squares approach of fitting conic is the geometric fitting based on minimization of the sum of orthogonal (geometric) distances from the observed data points to the fitted curve, i.e. the minimization problem is

$$\sum_{i=1}^{n}\left\{(x_i - x_i')^2 + (y_i - y_i')^2\right\} \to \min,$$

where (x_i', y_i') is an orthogonal projection of the given point (x_i, y_i) at certain geometric feature, see e.g. [1, 6, 7, 10, 12]. The geometric fitting is invariant under translations, rotations, and scaling.

Behind the statistical techniques for estimation of the unknown regression parameters, which are relying on the minimization of the geometric distance belongs the orthogonal regression, as well known as the total least squares [16], based on the maximum likelihood methods [4, 5]. However, the minimization of orthogonal distances requires projecting observed points onto conic what is computationally complicated. Although several efficient algorithms have been developed, such algorithms may become numerically unstable and time consuming [1, 2, 6].

In this paper we will focus on the geometric fit by the linear regression model with nonlinear restrictions on its parameters.

2   Conics description

Any conic can be described by an implicit equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

where $\boldsymbol{\theta} = (A, B, C, D, E, F)'$ is a vector of conic algebraic parameters. Obviously, any non-zero multiple of $\boldsymbol{\theta}$ corresponds to the same conic. In order to obtain real non-degenerate conic, algebraic parameters have to satisfy the inequality $A^2 + B^2 + C^2 > 0$ and the determinant Δ of the matrix

$$\begin{bmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{bmatrix}$$

has to be non-zero. For an ellipse, the relationship $\Delta(A+C) < 0$ must be also satisfied. The types of conics are classified by the discriminant (the determinant of matrix corresponding with the quadratic part $Ax^2 + Bxy + Cy^2$):

ellipse: $B^2 - 4AC < 0$
circle: $A = C,\ B = 0$
hyperbola: $B^2 - 4AC > 0$
parabola: $B^2 - 4AC = 0$

3   Conics fitting by linear model with constraints

Let us consider the problem of best geometric fit of an ellipse by the linear regression model with nonlinear restrictions on regression parameters proposed in [12]. The restrictions are represented by the implicit equation of the ellipse. The statistical model can be expressed in the form

$$x_i = \mu_i + \varepsilon_{x,i},\ i = 1, 2, \ldots, n,$$
$$y_i = \nu_i + \varepsilon_{y,i},\ i = 1, 2, \ldots, n,$$

where the unobserved points (μ_i, ν_i) stand for the errorless values of the observed points (x_i, y_i) and $\varepsilon_{x,i}$ and $\varepsilon_{y,i}$ represent measurements errors of μ_i and ν_i, respectively. Random errors are assumed to be independent with zero mean value and with equal variances $\mathrm{var}(\varepsilon_{x,i}) = \mathrm{var}(\varepsilon_{y,i}) = \sigma^2$. The errorless values (μ_i, ν_i) has to satisfy nonlinear constraints

$$\mu_i^2 + B\mu_i\nu_i + C\nu_i^2 + D\mu_i + E\nu_i + F = 0,\ i = 1, 2, \ldots, n,$$

where B, C, D, E, F represent the algebraic ellipse parameters. It is assumed $A = 1$ in order to obtain the unique estimates of algebraic parameters. Although this model is highly overparametrized with $(2n + 5)$ parameters, which are constrained by nonlinear restrictions, and $2n$ observations, the model leads to efficient estimates. The model can be written in matrix form as

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_x \\ \boldsymbol{\varepsilon}_y \end{pmatrix}, \quad \text{var}[\boldsymbol{\varepsilon}_x', \boldsymbol{\varepsilon}_y'] = \sigma^2 \mathbf{I}_{2n},$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$, $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)'$, $\boldsymbol{\varepsilon}_x = (\varepsilon_{x,1}, \ldots, \varepsilon_{x,n})'$, $\boldsymbol{\varepsilon}_y = (\varepsilon_{y,1}, \ldots, \varepsilon_{y,n})'$. The matrix form of nonlinear constraints is

$$\mathbf{B}\boldsymbol{\theta} + \mathbf{b} = \mathbf{0},$$

where $\mathbf{B} = \left[\boldsymbol{\mu}\boldsymbol{\nu} \vdots \boldsymbol{\nu}^2 \vdots \boldsymbol{\mu} \vdots \boldsymbol{\nu} \vdots \mathbf{1} \right]$, $\boldsymbol{\theta} = (B, C, D, E, F)'$, and $\mathbf{b} = \boldsymbol{\mu}^2$.

In order to obtain approximate linear regression model, the nonlinear constraints are being linearized by the first-order Taylor expansion, when the second and higher derivatives are neglected, about $\boldsymbol{\mu}_0$, $\boldsymbol{\nu}_0$ and $\boldsymbol{\theta}_0$

$$
\begin{aligned}
\mathbf{B}\boldsymbol{\theta} + \mathbf{b} \quad &\approx \quad (\mathbf{B}_0\boldsymbol{\theta}_0 + \mathbf{b}_0) + \frac{\partial (\mathbf{B}\boldsymbol{\theta} + \mathbf{b})}{\partial \boldsymbol{\mu}'}\big|_0 (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \\
&+ \quad \frac{\partial (\mathbf{B}\boldsymbol{\theta} + \mathbf{b})}{\partial \boldsymbol{\nu}'}\big|_0 (\boldsymbol{\nu} - \boldsymbol{\nu}_0) + \frac{\partial (\mathbf{B}\boldsymbol{\theta} + \mathbf{b})}{\partial \boldsymbol{\theta}'}\big|_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
&\approx \quad \mathbf{A}_0 \begin{pmatrix} \boldsymbol{\mu}_\triangle \\ \boldsymbol{\nu}_\triangle \end{pmatrix} + \mathbf{B}_0\boldsymbol{\theta}_\triangle + \mathbf{c}_0,
\end{aligned}
\tag{1}
$$

where

$$
\begin{aligned}
\mathbf{A}_0 \quad &= \left[\text{Diag}\left(\left[\boldsymbol{\nu}_0 \vdots \mathbf{0} \vdots \mathbf{1} \vdots \mathbf{0} \vdots \mathbf{0} \right] \boldsymbol{\theta}_0 + 2\boldsymbol{\mu}_0 \right) \vdots \text{Diag}\left(\left[\boldsymbol{\mu}_0 \vdots 2\boldsymbol{\nu}_0 \vdots \mathbf{0} \vdots \mathbf{1} \vdots \mathbf{0} \right] \boldsymbol{\theta}_0 \right) \right], \\
\boldsymbol{\mu}_\triangle \quad &= \boldsymbol{\mu} - \boldsymbol{\mu}_0, \quad \boldsymbol{\nu}_\triangle = \boldsymbol{\nu} - \boldsymbol{\nu}_0, \quad \boldsymbol{\theta}_\triangle = \boldsymbol{\theta} - \boldsymbol{\theta}_0, \\
\mathbf{B}_0 \quad &= \left[\boldsymbol{\mu}_0\boldsymbol{\nu}_0 \vdots \boldsymbol{\nu}_0^2 \vdots \boldsymbol{\mu}_0 \vdots \boldsymbol{\nu}_0 \vdots \mathbf{1} \right], \\
\mathbf{c}_0 \quad &= \mathbf{B}_0\boldsymbol{\theta}_0 + \mathbf{b}_0, \quad \boldsymbol{\theta}_0 = (B_0, C_0, D_0, E_0, F_0)', \quad \mathbf{b}_0 = \boldsymbol{\mu}_0^2.
\end{aligned}
$$

The resulting model

$$\begin{pmatrix} \mathbf{x}_\triangle \\ \mathbf{y}_\triangle \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}_\triangle \\ \boldsymbol{\nu}_\triangle \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_x \\ \boldsymbol{\varepsilon}_y \end{pmatrix}, \quad \text{var}[\boldsymbol{\varepsilon}_x', \boldsymbol{\varepsilon}_y'] = \sigma^2 \mathbf{I}_{2n},$$

where $\mathbf{x}_\triangle = \mathbf{x} - \boldsymbol{\mu}_0$ and $\mathbf{y}_\triangle = \mathbf{y} - \boldsymbol{\nu}_0$, together with the restrictions (1) is known in the statistical literature as linear regression model with type II constraints [9]. Type II constraints means that some of regression parameters are estimable only from the constraints, in our case these are the algebraic ellipse parameters. The locally best linear unbiased estimators (LBLUE) of the parameters $\boldsymbol{\mu}_\triangle$, $\boldsymbol{\nu}_\triangle$ and $\boldsymbol{\theta}_\triangle$ are

$$\left(\begin{pmatrix} \hat{\boldsymbol{\mu}}_\triangle \\ \hat{\boldsymbol{\nu}}_\triangle \\ \hat{\boldsymbol{\theta}}_\triangle \end{pmatrix} \right) = - \begin{pmatrix} \mathbf{A}_0'\mathbf{Q}_{11,0} \\ \mathbf{Q}_{21,0} \end{pmatrix} \mathbf{c}_0 + \begin{pmatrix} \mathbf{I} - \mathbf{A}_0'\mathbf{Q}_{11,0}\mathbf{A}_0 \\ -\mathbf{Q}_{21,0}\mathbf{A}_0 \end{pmatrix} \begin{pmatrix} \mathbf{x}_\triangle \\ \mathbf{y}_\triangle \end{pmatrix},$$

where the matrices $\mathbf{Q}_{11,0}$, $\mathbf{Q}_{12,0}$, $\mathbf{Q}_{12,0}$ and $\mathbf{Q}_{22,0}$ are blocks of the matrix given by the relation

$$\begin{pmatrix} \mathbf{Q}_{11,0} & \mathbf{Q}_{12,0} \\ \mathbf{Q}_{21,0} & \mathbf{Q}_{22,0} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_0 \mathbf{A}_0' & \mathbf{B}_0 \\ \mathbf{B}_0' & \mathbf{0} \end{pmatrix}^{-1}.$$

The corresponding variance-covariance matrix is

$$\operatorname{cov}\left(\begin{pmatrix} \hat{\boldsymbol{\mu}}_\triangle \\ \hat{\boldsymbol{\nu}}_\triangle \\ \hat{\boldsymbol{\theta}}_\triangle \end{pmatrix} \right) = \sigma^2 \begin{pmatrix} \mathbf{I} - \mathbf{A}_0' \mathbf{Q}_{11,0} \mathbf{A}_0 & -\mathbf{A}_0' \mathbf{Q}_{12,0} \\ -\mathbf{Q}_{21,0} \mathbf{A}_0 & -\mathbf{Q}_{22,0} \end{pmatrix}$$

and the unbiased estimator of the residual variance σ^2 is given by the expression

$$\hat{\sigma}^2 = \frac{1}{n-5} \sum_{i=1}^{n} \left([x_{\triangle,i} - \hat{\mu}_{\triangle,i}]^2 + [y_{\triangle,i} - \hat{\nu}_{\triangle,i}]^2 \right).$$

Finally, the LBLUE of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ are determined as

$$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_\triangle + \boldsymbol{\mu}_0, \quad \hat{\boldsymbol{\nu}} = \hat{\boldsymbol{\nu}}_\triangle + \boldsymbol{\nu}_0, \quad \hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_\triangle + \boldsymbol{\theta}_0.$$

Because the estimators of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ depend on the approximate values, we need to use an iterative estimation procedure. The iterative procedure guarantees that resulting algebraic parameters estimates satisfy given nonlinear constraints, i.e. prescribed conic. Thus it cannot happen that the resulting estimates represent different type of conic than the required one. If the procedure converges, similarly as in [8] it can be proved that the resulting estimates converge to the orthogonal least squares estimates, and moreover under normality assumption to the maximum likelihood estimates.

The proposed approach for geometric fit of ellipse can be adapted for any other type of real conics. Fitting circle is obvious since circle is a special kind of ellipse. Note that fitting parabola requires also to estimate the algebraic parameter A, which indicates whether the parabola is convex or concave. For parabola fitting it is necessary to choose different assumption on algebraic parameters in order to obtain their unique estimates.

From practical point of view, geometric conic parameters estimation is also important. Geometric parameters are more often used for conics description due to their clear interpretation. For example, the geometric ellipse parameters are the centre, angle of rotation, and the lengths of semi-axes. Let us denote the geometric conic parameters, as $\boldsymbol{g}$, than they can be taken as nonlinear functions of algebraic conic parameters, so the jth geometric parameter will be given by $g_j = f_j(A, B, C, D, E, F)$ Once we have estimates of algebraic parameters, we can plug these estimates into functions f_j to obtain plug-in estimates of geometric parameters $\hat{g}_j = f_j(\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{E}, \hat{F})$. By the law of uncertainty propagation, the estimated variance-covariance matrix of estimators $\hat{\boldsymbol{g}}$ is

$$\widehat{\operatorname{var}}(\hat{\boldsymbol{g}}) = \hat{\boldsymbol{J}} \operatorname{var}[(\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{E}, \hat{F})'] \hat{\boldsymbol{J}}', \ \hat{\boldsymbol{J}}_{j1} = \frac{\partial f_j}{\partial A}, \dots, \hat{\boldsymbol{J}}_{j6} = \frac{\partial f_j}{\partial F}.$$
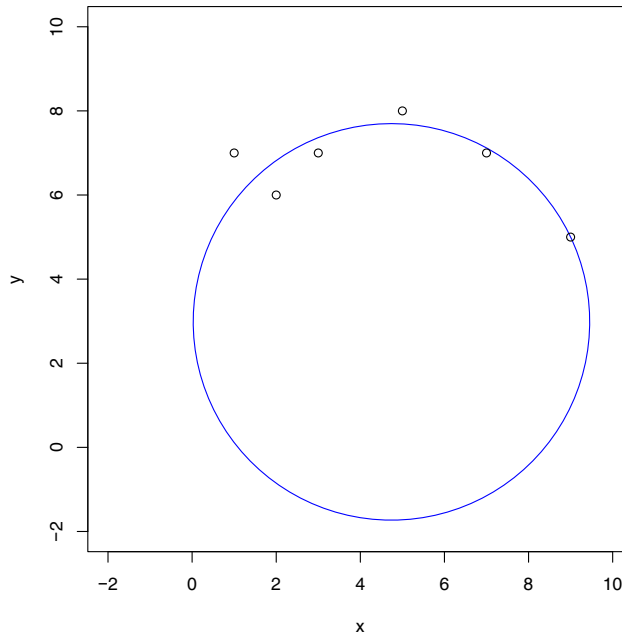
Figure 1: Fitted circle to the data from [1].

4 Examples of circle fitting

We devote to the problem of fitting a circle to six given points $(1, 7)$, $(2, 6)$, $(5, 8)$, $(7, 7)$, $(9, 5)$ and $(3, 7)$, inspired by [1, 10], with intention to compare the obtained results.

As suggested in [12] the initial values were chosen as $\boldsymbol{\mu_0} = \boldsymbol{x}$, $\boldsymbol{\nu_0} = \boldsymbol{y}$, $\boldsymbol{B_0} = (-2\boldsymbol{\mu}_0, -2\boldsymbol{\nu}_0, \boldsymbol{1})$, $\boldsymbol{b_0} = \boldsymbol{\mu}_0^2 + \boldsymbol{\nu}_0^2$, and $\boldsymbol{\theta}_0 = -(\boldsymbol{B}_0'\boldsymbol{B}_0)^{-1}\boldsymbol{B}_0'\boldsymbol{b}_0$. The convergence criterion was taken as reciprocal value of the square root of six multiplied by the Euclidean norm of differences between the estimates of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ from two consecutive steps. Convergence was reached in 19 iterations, with accuracy higher than $\varepsilon = 10^{-6}$. The fitted circle is displayed in Figure 1. Further the estimates of the circle center coordinates are $(4.740, 2.984)$ with standard errors 0.476 and 1.543; the estimate of the radius is 4.714 with the standard error 1.499. For the estimate of the residual variance we get 0.409.

Compared with the estimates of geometric circle parameters from the nonlinear least squares geometric fit as considered in [1], we found no difference; the obtained estimates of residual variance and of geometric circle parameters are the same. Although the obtained accuracy is a bit worse (almost twofold standard errors), the results show that the proposed procedure by linear model with constraints gives reasonable results even for not very good design of experiment.

For usual situation, when observed points are evenly spaced around the circle, simulations show that the accuracy of estimates of both algebraic and geometric parameters is very high. Figure 2 displays the resulting accuracy of the radius obtained from 1000 simulations with the convergence criterion higher than 10^{-12}. Simulation study was performed for each of 10, 25, 50, 100 and 500 given points. The true standard errors of measurements were chosen equidistantly from 0.01 to 0.1. Similar results were obtained for other circle parameters.
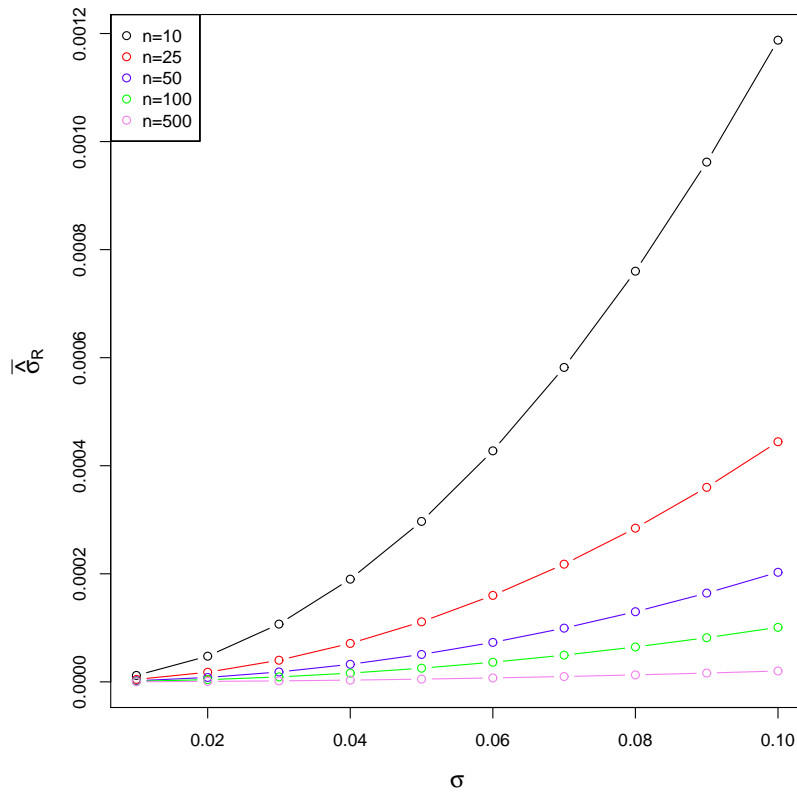
Figure 2: The average standard error of radius (n number of observed points, σ accuracy of measurement).

5 Conclusion

Linear regression model with type II constraints has been presented as an alternative approach suitable for finding the geometric fit when conics such as circle, ellipse, hyperbola and parabola are examined. Performed simulation study has confirmed that the behaviour of the accuracy of the estimates of the both algebraic and geometric parameters give very convenient results when the data noticeably fit certain curve. Hence under these circumstances the estimates of the both algebraic and geometric parameters are immensely accurate. Additionally, the proposed iterative procedure is quite satisfactory, even if visibly the data points do not look like fitting curve is required. Furthermore, the iterative procedure has been compared with approach from [1] and has been found that the resulting geometric circle (ellipse) parameters estimates as well as the residual errors are the same. Nevertheless, the accuracy of the estimates is a bit worse.

References

[1] Ahn, S. J., Rauh, W. and Warnecke, H.J. (2001). Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola, *Pattern Recog.*, 34, 2283-2303.

[2] Ahn, S. J. (2004). *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space*, Springer, Berlin.

[3] Berman, M. (1989). Large sample bias in least squares estimators of a circular arc center and its radius, *CVGIP*, 45, 126-128.

[4] Chan, N. N. (1965). On circular functional relationships, *J. R. Statist. Soc. B.*, 27, 45-56.

[5] Chernov, N., Lesort, C. (2004). Statistical efficiency of curve fitting algorithms, *Comp. Stat. Data Anal.*, 47, 713-728.

[6] Chernov, N. (2011). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*, Chapman and Halll/CRC Monographs on Statistics and Applied Probability, London.

[7] Chernov, N. and Ma, H. (2011). Least squares fitting of quadratic curves and surfaces, *Computer Vision*, (Yoshida, S. R., eds.), Nova Science Publishers, 285-302.

[8] Donevska, S., Fišerová, E. and Hron, K. (2011). On the Equivalence between Orthogonal Regression and Linear Model with Type-II Constraints, *Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica*, 50, 19-27.

[9] Fišerová, E., Kubáček, L. and Kunderová, P. (2007). *Linear Statistical Models: Regularity and Singularities.* Academia, Praha.

[10] Gander, W., Golub, G. H. and Strebel, R. (1994). Least-squares fitting of circles and ellipses, *BIT Numer. Math.*, 34, 34-558.

[11] K. Kanatani, K., (1994) Statistical bias of conic fitting and renormalization, *IEEE Trans. PAMI*, 16, 320-326.

[12] Köning, R., Wimmer, G. and Witkovský, V. (2014). Ellipse fitting by linearized nonlinear constraints to demodulate quadrature homodyne interferometer signals and to determine the statistical uncertainty of the interferometric phase, *Meas. Sci. Technol.*, 25, 115-001.

[13] Rosin, P.L. (1993). A note on the least squares fitting of ellipses, *Pattern Recognition Lett.*, 14, 799-808.

[14] Rosin, P.L. (1996). Assessing error of fit functions for ellipses, *Graphical Models Image Process.*, 58, 494-502.

[15] Safaee-Rad, R., Tchoukanov, I., Benhabib, B. and Smith, K.C. (1991). Accurate parameter estimation of quadric curves from grey-level images, *CVGIP: Image Understanding*, 54, 259-274.

[16] Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia.

Bootstrap for Quasi Stationary Distributions

Guangbao Guo[1,*]

[1]School of Mathematics, Shandong University, Jinan, China

Abstract Research on quasi stationary distributions, is a very important research topic. When the sample size is not sufficiently large, the asymptotic results of the general parameter method, may not hold. The bootstrap method which is introduced to solve it, avoids this disadvantage. We give the approximation form of it, and study the approximation property. Simulations are presented to illustrate the method, using two examples, the method shows to be efficient when the small sample is presented. At last, an application in a pure-death chain is discussed, together with some results.

Keywords Bootstrap; Quasi stationary distributions; Confidence band

1 Introduction

1.1 Motivation

Quasi stationary distribution has become an important and a powerful tool for the statistician and economist. It provides with a modeling framework and a computationally efficient way. The quasi stationary distribution have a varied parament λ $(0 < \lambda \leq 1)$, which is not included in stationary distribution. It make quasi stationary distribution be very interested. In generally, the vector of quasi stationary distribution doesn't have a large size. When the sample size is small, the asymptotic results of the general parameter estimation method, may not hold. In these cases, the bootstrap method can be used to find approximate results. Here the most important motives are risk reduction, lack of capital and freedom of action.

The main advantage of the bootstrap is that the sampling distribution is estimated based on the original characteristics of the data, and it can provide estimation in sittings where mathematical solutions are not possible. Here the bootstrap method is proposed to solve the quasi stationary distribution, due to attracting advantages both regarding to asymptotic properties and from a practical viewpoint.

1.2 Quasi Stationary Distribution

Let $\{X_t, t \geq 0\}$ be a homogeneous ergodic Markov chain with a finite state space S, and transition probablity matrix $P = (p_{ij})$, together with a transient irreducible class $C \subseteq S$ for which the first exit time T_t from C is almost surely finite. The problem of estimating the transition probability P, the stationary probability π, arises in the areas of applied probability and statistics. The application of the bootstrap method, used to approximate the distributions of Markov chain, was considered in Basawa, et al. (1990), Kulperger and Rao (1990), Athreya and Fuh (1992), Datta and McCormick (1992), Fuh(1993) and Polunsky (2009).

---

*Corresponding author. Email: ggb11111111@163.com

We consider an ergodic Markov chain with $C = \{s_1, s_2, ..., s_l\}$. The ergodic property implies that the existence of stationary distribution $\pi = \{\pi_1, ..., \pi_l\}$ such that

$$\pi_j > 0, \sum_{j=1}^{l} \pi_j = 1, \pi_j = \sum_i \pi_i p_{ij}, j = 1, ..., l.$$

Suppose $X = \{x_0, x_1, ..., x_{n-1}\}$ is a realization of the process $\{X_0, X_1, ..., X_{n-1}\}$. Using maximum likelihood method, we can estimate the parameters π_i's and p_{ij}'s as

$$\hat{\pi}_i = n^{-1} \sum_{k=0}^{n-1} P(X_k = s_i),$$

$$\hat{p}_{ij} = \frac{\sum_{k=0}^{n-2} P(X_k = s_i, X_{k+1} = s_j)}{\sum_{k=0}^{n-1} P(X_k = s_i)}, 1 \le i, j \le l.$$

If n_{ij} is the number of ij transitions in $\{x_0, x_1, ..., x_{n-1}\}$, n_i is the number of visits to state i in $\{x_0, x_1, ..., x_{n-1}\}$. We obtain that

$$\hat{p}_{ij} = \begin{cases} n_{ij}/n_i, & \text{if } n_i > 0; \\ \delta_{ij}, & \text{otherwise}, \end{cases}$$

where $\delta_{ij} = 1$ as $i \ne j$; $\delta_{ij} = 0$ as $i = j$.

A proper probability distribution $u \equiv (u_i, i \in C)$ is said to be a quasi stationary distribution (QSD) for $\mathcal{X}$ if the distribution of X_k, is constant over k when u is the initial distribution, that is, for all $k = 0, 1, \ldots, n-1$, one has $P_u(T_t > k) > 0$ and

$$P_u(X_k = j \mid T_t > n) = u_j, j \in C,$$

where $T_t = \inf\{k \ge 0 : X_k = 0\}$ is the absorption time.

It is well known that the QSD exists and is unique whenever C is finite. In the infinite case, it is natural to ask whether the class C may be replaced by a large but finite subset $C(n)$, such that the corresponding QSD approximates one on C.

Let $\lambda = P_u(T_t > 1)$, we have

$$\sum_{i \in C} u_i P_{ij} = P_u(X_1 = j) = P_u(T_t > 1)u_j = \lambda u_j,$$

then the statements are equivalent: for some $\lambda > 0$, u is a QSD for $\mathcal{X}$. $P = (P_{ij})$ is called quasi stochastic matrix or generalized stochastic matrix.

1.3 Literature Review and Problem

The interest to bootstrap methods has largely expanded to solve some relevant problems for the dependent data. An important category of the methods is that based on the construction of time blocks such that stationary bootstrap; blocks-of-blocks bootstrap; regenerative block bootstrap; tapered block bootstrap. Another major family, developed to recognize and maintain the original data dependency, is concerned with Markov chains such that Markov chain bootstrap; sieve (Markov) bootstrap; local bootstrap (Paparoditis and Politis, 2001). Here we consider boostrap method to solve the quasi stationary distributions. Moreover, in order to implement the bootstrap, the problem is that one firstly needs to have a good estimator of the variance.

The remainder of the paper is organized as follows. In Section 2, we give the bootstrap method for the QSDs in Markov chains. The approximation theory of the resampling method for the QSDs, is described in Section 3. Section 4 presents some simulation results to illustrate the method, with two examples of the QSDs while an application in pure-death chain is given, together with some simulations in Section 5. Section 6 gives discussion.

2 Bootstrap Method for Quasi Stationary Distributions

When a model involves a Markov chain, it has proved useful to study the associated family of quasi stationary distributions (QSDs), which is commonly encountered in various scientific fields, poses great challenges to modern stochastic analysis. In this section, we introduce bootstrap method for the QSDs.

2.1 Quasi Stationary Distributions vs Maximum Likelihood Estimator

Let n_i be the number of times that state i is observed; n_{ij} be the number of observed consecutive transitions from state i to state j in $X_0, X_1, \cdots, X_{n-1}$. Given $\{X_0, \cdots, X_{n-1}\}$ from the Markov chain with the QSDs, and λ, we can estimate u_i and p_{ij} as

$$\hat{u}_i = \frac{\sum_{k=0}^{n-1} P(X_k = s_i)}{n} = \frac{n_i}{n}; \hat{p}_{ij} = \frac{\sum_{k=0}^{n-2} P(X_k = s_i, X_{k+1} = s_j)}{\sum_{k=0}^{n-1} P(X_k = s_i)} = \begin{cases} n_{ij}/n_i, & \text{if } n_i > 0; \\ \delta_{ij}, & \text{otherwise} \end{cases}$$

(1)

Here $\delta_{ij} = 1$ as $i \neq j$; $\delta_{ij} = 0$ as $i = j$.

Thus, $\hat{u} = (\hat{u}_1, \hat{u}_2, ..., \hat{u}_l)$ and $\hat{P} = (\hat{p}_{ij})$ are the parametric maximum likelihood estimator of u and P. It is, however, compulationally more challenging to obtain the non-parameter asymptotic approximation estimation of QSDs.

In general the QSDs are the probability measures. Some of the interesting problems are concerned with the search for the conditions, about the transition matrices for the existence of the QSDs; the domains of attractions of the QSDs; the numerical evaluation of the quasi stochastic matrices and the QSDs. The relationship, between the QSD u and the quasi stochastic matrix P in the chain, was identified by Nair and Pollett(1993).

The properties of $\hat{u}_i$ and $\hat{p}_{ij}$ have been given as well as the stationary distributions. The asymptotic properties of $\hat{u}_i$ and $\hat{p}_{ij}$ can be described as following (also see, Proposition 3.4).

Proposition 2.1 As $n \to \infty$, for fixed λ,

$$\sqrt{n}\left[\hat{u}_i - u_i\right] \to^P N\left(0, \hat{\sigma}_{u_i}^2\right),$$

where $\hat{\sigma}_{u_i}^2$ is a correlative covariance.

Proposition 2.2 As $n \to \infty$, for fixed λ,

$$\sqrt{n}\left[\hat{p}(i,j) - p(i,j)\right] \to^W N\left(0, \hat{\sigma}_{p_{ij}}^2\right),$$

where $\hat{\sigma}_{p_{ij}}^2$ is a correlative covariance.

where $\hat{\theta}_n$ is an estimator of θ based on $X_0, \cdots, X_{n-1}$.

Let $E(X) = \mu, Var(X) = \sigma^2$, $Y_i = \sum X_j(X_j \in B_i), Y_i^* = \sum X_j^*(X_j^* \in B_i^*)$. For $T_n = \sqrt{n}(\bar{X}_n - \mu)$, $\bar{X}_n = \sum X_i/n(n = \sum_{i=1}^{k_n} \tau_i)$, we set $T_n^* = \sqrt{n^*}(\bar{X}_{n^*}^* - \hat{\mu}_n)$, where $\bar{X}_{n^*}^*$ is the average of the n^* bootstrap variables and $\hat{\mu}_n = \sum \hat{u}_i X_i$, the expectation of X_0^*. Let $k_n/n \to p \leq 1$. k_n is the block number of bootstrap sample, $(X_1^*, ..., X_{n^*}^*)$ is the bootstrap sample, P^*, E^*, Var^* stand for conditional P, E, Var, τ_i^* is the subsample number of the ith bootstrap sample, $\sum_{i=1}^{k_n} \tau_i = \sum_{i=1}^{k_n} E^* \tau_i^* = E^* n^*$.

3.2 Weak Convergence and Bootstrap Method

Proposition 3.1 Assume that $E\tau^2 < \infty$, the length of the bootstrap block $n^* = \sum_{i=1}^{k_n} \tau_i^*$, we have the ratio $n^*/n \to 1(n, n^* \to \infty)$.

Based on the above result, we will consider $\hat{u}_i$ in this paper. By a similar argument, we use $\hat{p}_{ij}$. Setting $\hat{\mu}_{n*}^* = \sum \hat{u}_i^* X_i^* \left(\hat{u}_i^* = \frac{1}{n^*} \sum_{k=1}^{n} P(X_k^* = s_i) \right)$. The bootstrap approximation of the sampling distribution of $\sqrt{n^*}(\bar{X}^* - \hat{\mu}_n)$ and $\sqrt{n^*}(\hat{\mu}_{n*}^* - \mu)$ are based on the i.i.d. blocks $\{B_1^*, ..., B_{k_n}^*\}$.

Proposition 3.2 Assume that the chain $\{X_i\}_{i \geq 0}$ is a Markov chain with the QSDs, for fixed λ, $E\tau^2 < \infty$. Then the statistic

$$T_n = \sqrt{n}(\bar{X} - \mu) \to^W N(0, \sigma^2)$$

implies the statistic

$$T_n^* = \sqrt{n^*}(\bar{X}^* - \hat{\mu}_n) \to^W N(0, \sigma^2).$$

For $\sqrt{n^*}(\hat{\mu}_{n*}^* - \mu)$, we have the following results.

Proposition 3.3 The chain $\{X_i\}_{i \geq 0}$ is a Markov chain with the QSDs. Assume $E\tau^2 < \infty$, for fixed λ, there is a Gaussian process G_u such that

$$T_n = \sqrt{n}(\hat{\mu}_n - \mu) \to^W G_u,$$

then

$$T_n^* = \sqrt{n^*}(\hat{\mu}_{n*}^* - \mu) \to^W G_u.$$

It is seen that the bootstrap is also asymptotic approximation in the setting. In the view of Propositions 3.2, 3.3, as a special case, we now present our proposition as following.

Proposition 3.4 Assume that $E\tau^2 < \infty$, for fixed λ, u_i and $\hat{u}_i$ are the QSDs and there MLEs, respectively, using bootstrap method. p_{ij} and $\hat{p}_{ij}$ are the (quasi) transition probabilities from state i to state j and the MLEs of theirs, using bootstrap method. Then

$$\sqrt{n}\left(\hat{u}_i - u_i \right) \to^W N(0, \hat{\sigma}_{u_i}^2); \sqrt{n}\left(\hat{p}_{ij} - p_{ij} \right) \to^W N(0, \hat{\sigma}_{p_{ij}}^2)$$

implies

$$\sqrt{n^*}\left(\hat{u}_i^* - u_i \right) \to^W N(0, \hat{\sigma}_{u_i}^2); \sqrt{n^*}\left(\hat{p}_{ij}^* - p_{ij} \right) \to^W N(0, \hat{\sigma}_{p_{ij}}^2),$$

where $\hat{\sigma}_{u_i}^2$ and $\hat{\sigma}_{p_{ij}}^2$ are correlative covariances.

In the same time, if

$$\sup_i \sqrt{n}|\hat{u}_i - u_i| \to^W \sup_i |G_{u_i}|; \sup_i \sqrt{n}|\hat{p}_{ij} - p_{ij}| \to^W \sup_i |G_{p_{ij}}|,$$

we have

$$\sup_i \sqrt{n}|\hat{u}_i^* - u_i| \to^W \sup_i |G_{u_i}|; \sup_i \sqrt{n}|\hat{p}_{ij}^* - p_{ij}| \to^W \sup_i |G_{p_{ij}}|,$$

where G_{u_i} and $G_{p_{ij}}$ are appropriate Gaussian processes.

3.3 Asymptotic Variance of Bootstrap Method

Note that as $n \to \infty$, $\hat{u}_i \to u_i$ and $\hat{u}_i^\tau \to \hat{u}_i$, using Propositions 2.1-2.3. Then we only use $\hat{u}_i^\tau$ and its bootstrap estimator $\hat{u}_i^*$. By a similar argument, we consider $\hat{p}_{ij}^\tau$ and its bootstrap estimator $\hat{p}_{ij}^*$. We find τ, $g(B)$ and $h(B)$, to minimize the asymptotic variances of $\hat{u}_i^* - \hat{u}_i^\tau$ and $\hat{p}_{ij}^* - \hat{p}_{ij}^\tau$.

Proposition 3.5 Assume that the chain $\{X_i\}_{i \geq 0}$ is a Markov chain with the QSDs, for fixed λ, $E\tau^2 < \infty$. Then

$$\frac{1}{\sqrt{k_n}}\left(\hat{u}_i^* - \hat{u}_i^\tau\right) \to AN(0, \sigma_{u_i}^2); \frac{1}{\sqrt{k_n}}\left(\hat{p}_{ij}^* - \hat{p}_{ij}^\tau\right) \to AN(0, \sigma_{p_{ij}}^2),$$

where $\sigma_{u_i}^2$ and $\sigma_{p_{ij}}^2$ are correlative covariances.

Proposition 3.6 Under the conditions given in Proposition 3.5. Then the conditions of $\min \sigma_{u_i}^2$ and $\min \sigma_{p_{ij}}^2$ are $Var(\tau)E\tau = Var(g(B))Eg(B)$ and $Var(g(B))Eg(B) = Var(h(B))Eh(B)$. Moreover, we have

$$\min \sigma_{u_i}^2 = 2E(g(B))/E(\tau)^3 \cdot \left(\sqrt{Var(\tau)Var(g(B))} - Cov(\tau, g(B))\right),$$

$$\min \sigma_{p_{ij}}^2 = 2E(h(B))/E(g(B))^3 \cdot \left(\sqrt{Var((h(B))Var(g(B))} - Cov(g(B), h(B))\right).$$

3.4 Estimators of the Variances about the QSDs

In Proposition 3.6, we have good estimators of the variance about u_i and p_{ij} if one chooses suitable τ, $g(B)$ and $h(B)$. Now we compare different approaches to estimating $\sigma_{u_i}^2$ and $\sigma_{p_{ij}}^2$.

For the estimation of the variance $\sigma_{u_i}^2$ of u_i in $u = (u_1, u_2, ..., u_l)$, set

$$\hat{\sigma}_{u_i^\tau}^2 = \frac{\sum\limits_{\alpha=1}^{k_n}[\tau_\alpha(u_i^\alpha - \hat{u}_i^\tau)]^2}{(\sum\limits_{\alpha=1}^{k_n}\tau_\alpha)^2}; \hat{\sigma}_{u_i^*}^2 = \frac{\sum\limits_{\alpha=1}^{k_n}[\tau_\alpha^*(\hat{u}_i^{*\alpha} - \hat{u}_i^\tau)]^2}{(\sum\limits_{\alpha=1}^{k_n}\tau_\alpha^*)^2}; \varepsilon_\alpha = \hat{\sigma}_{u_i^\tau}^{-1}\frac{\tau_\alpha(u_i^\alpha - \hat{u}_i^\tau)}{(\sum\limits_{\alpha=1}^{k_n}\tau_\alpha)}. \quad (4)$$

where $u_i^\alpha = g(B_\alpha)/\tau_\alpha$ and $\hat{u}_i^{*\alpha} = g(B_\alpha^*)/\tau_\alpha^*$.

Set λ_α is the number of times τ_α that appears in the resample τ^*. Then $\hat{u}^* = \sum\limits_{\alpha=1}^{k_n}\lambda_\alpha\hat{u}_i^{*\alpha}$,

$$T^* = \frac{\hat{u}_i^* - \hat{u}_i}{\hat{\sigma}^*} = \frac{\hat{u}_i^* - \hat{u}_i}{\hat{\sigma}} \cdot \frac{\hat{\sigma}}{\hat{\sigma}^*} = \left(\sum_{\alpha=1}^{k_n}\lambda_\alpha\varepsilon_\alpha\right)\frac{\sum\limits_{\alpha=1}^{k_n}\tau_\alpha}{\sum\limits_{\alpha=1}^{k_n}\lambda_\alpha\tau_\alpha} \cdot \frac{\hat{\sigma}}{\hat{\sigma}^*} \approx \sum_{\alpha=1}^{k_n}\lambda_\alpha\varepsilon_\alpha.$$

Under some suitable conditions, as $n \to \infty$, $T^* \approx (\sum\limits_{\alpha=1}^{k_n}\lambda_\alpha\varepsilon_\alpha) \sim N(0,1)$.

Then

$$E\left(\hat{\sigma}_{u_i^\tau}^2\right) \to \sigma_{u_i}^2; E\left(\hat{\sigma}_{u_i^*}^2\right) \to \sigma_{u_i}^2.$$

For the estimation of the variance $\sigma^2_{p_{ij}}$ of p_{ij} in quasi stochastic matrix $P = (p_{ij})$, set

$$
\hat{\sigma}^2_{p^{\tau}_{ij}} = \frac{\sum\limits_{k=1}^{k_n} [\tau_k(\hat{p}_{ij} - p_{ij})]^2}{(\sum\limits_{j=1}^{k_n} \tau_j)^2} ; \hat{\sigma}^2_{p^*_{ij}} = \frac{\sum\limits_{k=1}^{k_n} [\tau^*_k(\hat{p}^*_{ij} - \hat{p}_{ij})]^2}{(\sum\limits_{j=1}^{k_n} \tau^*_j)^2} ; \varepsilon_j = \hat{\sigma}^{-1}_{u_i} \frac{\tau_j(u_i - \hat{u}_i)}{(\sum\limits_{j=1}^{k_n} \tau_j)} . \tag{5}
$$

As well as u_i, we have

$$
E\left(\hat{\sigma}^2_{p_{ij}}\right) \to \sigma^2_{p_{ij}} ; E\left(\hat{\sigma}^{*2}_{p_{ij}}\right) \to \sigma^2_{p_{ij}} .
$$

By (4) and (5), we have the asymptotic unbias estimators of the variances for u_i and p_{ij}. $\hat{\sigma}^2_{u^{\tau}_i}$ and $\hat{\sigma}^2_{p^{\tau}_{ij}}$ are tha same as $\hat{\sigma}^2_{u_{i=}}$ and $\hat{\sigma}^2_{p_{ij}}$ in (1). $\hat{\sigma}^2_{u^{\tau}_i}$ and $\hat{\sigma}^2_{p^*_{ij}}$ are the bootstrap variances, good estimators for the independent samples. But when we use them to estimate u_i and p_{ij}, by bootstrap-t method, a problem with this method is that our QSDs are dependent, so it will perform very poorly. In order to avoid this, we consider percentile method of the bootstrap in the next section.

4   Numerical simulation

Applied to the problem of estimating the QSDs, the bootstrap method consists of computing u and P from the original chain, and then generating k_n additional samples based on $\hat{u}_i$ and $(\hat{p}_{ij})$ ($i, j = 1, 2, ..., l$). The QSDs for the initial state are used. For each of these resamples, $\hat{u}^*_i$ and $\hat{p}^*_{ij}$ ($i = 1, \cdots, l$) are computed. The empirical distribution function for each element u_i and p_{ij} can also be computed, based on the sample $(B^*_1, \cdots, B^*_{k_n})$, denoted by $\hat{F}_{\hat{u}^*_i}$ and $\hat{F}_{\hat{p}^*_{ij}}$. Our bootstrap confidence interval is based on the percentile method of Efron (1979). Percentile method is in some ways less intuitive than those in the above section, but has the advantage of not requiring $\sigma^2_{u_i}$ (or $\sigma^2_{p_{ij}}$). By $\hat{u}_i - u_i \sim N(0, \sigma^2_{u_i})$, the confidence interval for u_i is

$$
(-\infty, \hat{u}_i - \sigma z_{\alpha}),
$$

where z_{α} is the α-level critical point of the standard normal distribution. Using Propositions 2.1-2.4, we have

$$
\hat{u}^*_i - \hat{u}_i \sim \hat{u}_i - u_i \sim N(0, \sigma^2_{u_i}) . \tag{6}
$$

Then, $\hat{u}^*_i - \hat{\sigma}_{u^*_i} z_{\alpha} = \hat{F}^{-1}_{\hat{u}^*_i}(1 - \alpha)$. The bootstrapped percentile confidence interval of u_i is

$$
(-\infty, \hat{F}^{-1}_{\hat{u}^*_i}(1 - \alpha)) .
$$

In same time, the percentile confidence interval of p_{ij} is $(-\infty, \hat{F}^{-1}_{\hat{p}^*_{ij}}(1 - \alpha))$.

A $(1-\alpha)100$ percent confidence band, for the elements u_i and p_{ij}, are given by $B_{\hat{u}^*_i}(\alpha)$ and $B_{\hat{p}^*_{ij}}(\alpha)$. Here,

$$
\begin{aligned}
B_{\hat{u}^*_i}(\alpha) &= \{(x, y) : \hat{F}_{\hat{u}^*_i}(x) \le \alpha, \hat{F}_{\hat{u}^*_i}(x) \ge (1 - \alpha)\}, \\
B_{\hat{p}^*_{ij}}(\alpha) &= \{(x, y) : \hat{F}_{\hat{p}^*_{ij}}(x) \le \alpha, \hat{F}_{\hat{p}^*_{ij}}(x) \ge (1 - \alpha)\}.
\end{aligned}
$$

Let

$$
\begin{aligned}
(\hat{F}^{-1}_{\hat{u}^*_i}(\alpha), \hat{F}^{-1}_{\hat{u}^*_i}(1 - \alpha)) &= \{(x, y) \in B_{\hat{u}^*_i}(\alpha) : \min |y - x|\}, \\
(\hat{F}^{-1}_{\hat{p}^*_{ij}}(\alpha), \hat{F}^{-1}_{\hat{p}^*_{ij}}(1 - \alpha)) &= \{(x, y) \in B_{\hat{p}^*_{ij}}(\alpha) : \min |y - x|\},
\end{aligned}
$$

we have the bootstrap simultaneous confidence interval of u_i and p_{ij}.

There are two simulation examples under proposed bootstrap method about the problem in this section. In these simulation studies, we consider two five-state Markov chain with quasi stochastic matrix and generalized stochastic matrix, respectively. Note that $uP = \lambda u$, then $P^T u^T = \lambda u^T$. Let the matrix $P = B^T$. Here we do not consider the sum of u is contained.

First, we give an example of Markov chains with the QSDs. Set the QSD $u_I = 1/\sqrt{5} \cdot (1,1,1,1,1)$, $\lambda = \frac{3}{4}$ and

$$
B_I = \begin{pmatrix}
\frac{3}{4} & 0 & 0 & 0 & 0 \\
\frac{3}{8} & \frac{3}{8} & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\
\frac{3}{16} & \frac{3}{16} & \frac{3}{16} & \frac{3}{16} & 0 \\
\frac{3}{20} & \frac{3}{20} & \frac{3}{20} & \frac{3}{20} & \frac{3}{20}
\end{pmatrix},
$$

where B_I is called column quasi stochastic matrix.

We present another example of Markov chains with the QSDs. Set the QSD $u_{II} = 1/\sqrt{7/2} \cdot (1,1,1,\frac{1}{2},\frac{1}{2}), \lambda = 1$ and

$$
B_{II} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
\frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & 0 \\
\frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{5} & \frac{1}{5}
\end{pmatrix},
$$

where B_2 is called column generalized stochastic matrix.

For these small sample studies, two different sample sizes $n = 100, 200$ are included. The size of the bootstrap samples is the same as the original coverage probabilities, based on 1000 replications trials. The results are presented in Tables 1 and 2.

The data were generated by these Markov chains (the original sample size, fixed as $n = 300$). Computation was performed, using R software. Using the bootstrap method in quantile estimation, we simulate percent confidence intervals, along with their coverage probabilities for B and u. In each situation, 500 bootstrap replications were run.

95 percent confidence intervals, their average length, coverage probability and others of (B_I, u_I) and (B_{II}, u_{II}), are shown in Tables 1 and 2.

Now with small samples, we consider bootstrap method to solve it, to obtain the large samples. If we use bootstrap method, it will be wonderfully worked out. It is that bootstrap method can pose large sample when there are small samples.

Table 1: Comparison of the approximate confidence intervals for B_I, and u_1 of u_I

		0.95 CI	AC	CP	0.95 CIu	0.975 CIu	0.99 CIu
	B_{11}	(0.381, 1.370)	0.989	0.959	1.196	1.359	1.502
	B_{22}	(0.202, 0.653)	0.451	0.967	0.615	0.640	0.785
$n = 100$	B_{33}	(0.097, 0.385)	0.288	0.964	0.368	0.382	0.395
	B_{44}	(0.0573, 0.2631)	0.2058	0.957	0.2574	0.2602	0.2753
	B_{55}	(0.0357, 0.2387)	0.203	0.971	0.2334	0.2375	0.2497
	u_1	(0.3385, 0.7325)	0.394	0.962	0.706	0.7473	0.7607
	B_{11}	(0.437, 1.258)	0.821	0.956	1.185	1.254	1.483
	B_{22}	(0.257, 0.565)	0.308	0.963	0.548	0.563	0.701
$n = 200$	B_{33}	(0.113, 0.349)	0.236	0.962	0.338	0.346	0.389
	B_{44}	(0.0784, 0.2543)	0.1759	0.955	0.2491	0.2538	0.2703
	B_{55}	(0.0679, 0.2346)	0.1667	0.968	0.2315	0.2341	0.2485
	u_1	(0.3591, 0.6829)	0.3238	0.958	0.678	0.7136	0.7535

CI: Confidence Interval; CIu: upper end point of one-sided Confidence Interval; CP: Coverage Probability; AL: Average Length; $diag\{B_1\} = \{B_{11}, B_{22}, B_{33}, B_{44}, B_{55}\}$

Table 2: Comparison of the approximate confidence intervals for B_2, and u_1 of u_{II}

		0.95 CI	AC	CP	0.95 CIu	0.975 CIu	0.99 CIu
	B_{11}	(0.432, 1.378)	0.946	0.960	1.340	1.374	1.505
	B_{22}	(0.379, 0.729)	0.35	0.971	0.674	0.723	0.836
$n = 100$	B_{33}	(0.212, 0.476)	0.264	0.967	0.463	0.474	0.623
	B_{44}	(0.173, 0.3739)	0.2009	0.963	0.3630	0.3721	0.3915
	B_{55}	(0.1458, 0.3018)	0.156	0.972	0.3011	0.3017	0.3121
	u_1	(0.2676, 0.821)	0.5534	0.9654	0.8076	0.8364	0.8571
	B_{11}	(0.457, 1.369)	0.912	0.953	1.39	1.364	1.473
	B_{22}	(0.423, 0.698)	0.275	0.964	0.673	0.692	0.754
$n = 200$	B_{33}	(0.273, 0.435)	0.162	0.961	0.429	0.432	0.558
	B_{44}	(0.1731, 0.3739)	0.2008	0.957	0.3657	0.3735	0.3891
	B_{55}	(0.1689, 0.2846)	0.1157	0.967	0.2803	0.2833	0.3087
	u_1	(0.2728, 0.8089)	0.5361	0.9578	0.8011	0.8214	0.8483

CI: Confidence Interval; CIu: upper end point of one-sided Confidence Interval; CP: Coverage Probability; AL: Average Length; $diag\{B_2\} = \{B_{11}, B_{22}, B_{33}, B_{44}, B_{55}\}$

Although it is impossible to know the true confidence interval for most problems, bootstrap is asymptotically more accurate than the standard intervals. Simplicity is the attraction of this method, and explains its continued popularity. Unlike the bootstrap-t, it does not require estimating the related variances. Further, no invalid parameter values can be included in the interval. Another advantage of this group over the pivotal methods is that they are transformation respecting.

5 Application

A pure-death chain is with killing on $C = \{1, 2, ..., s\}$ with death probability q_i and killing probability k_i in state i, and coffin state 0. The probability of death from state 1 is k_1, $r_1 = 1 - k_1$ is the probability of remaining in state 1, $0 \le k_i \le 1, 0 < q_i < 1$. Due to the complexity of the above QSDs in the progress (in general, the rate matrix is a three-diagonal matrix), we specialize the matrix, the model fits with a quasi stochastic

matrix (or a negative rate matrix)

$$
B_3 = P^T = \begin{pmatrix} r_1 & 0 & 0 & \cdots & 0 \\ q_2 & r_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & r_{s-1} & 0 \\ 0 & \cdots & \cdots & q_s & r_s \end{pmatrix}^T.
$$

(also see, Doorn and Pollett, 2009).

A special example in pure-death chains is as following,

$$
B_3 = P^T = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 0 \\ -1/3 & 1/3 & 0 & 0 & 0 \\ 0 & -1/4 & 1/4 & 0 & 0 \\ 0 & 0 & -1/5 & 1/5 & 0 \\ 0 & 0 & 0 & -1/6 & 1/6 \end{pmatrix}^T.
$$

Here $s = 5$, $u_P = (u_1, u_2, ..., u_5)$, set $\lambda = 1/6$ and $u_1 = 1$, we have the QSD $u_P = (1, 1, 2/3, 5/18, 1/18)$. Now we give some simulations of the QSD u_P as well as those above, we have some comparison results in Table 3.

Table 3: Comparison of the approximate confidence intervals for u_P

		0.95 CI	AC	CP	0.95 CIu	0.975 CIu	0.99 CIu
	u_1	(0.713, 1.323)	0.61	0.954	1.278	1.320	1.457
	u_2	(0.693, 1.316)	0.621	0.958	1.269	1.309	1.446
$n = 100$	u_3	(0.401, 0.875)	0.474	0.963	0.867	0.871	0.912
	u_4	(0.131, 0.413)	0.282	0.957	0.404	0.411	0.435
	u_5	(0.0021, 0.0079)	0.0058	0.960	0.0071	0.0077	0.0089
	u_1	(0.643, 1.220)	0.577	0.963	1.209	1.216	1.437
	u_2	(0.635, 1.201)	0.566	0.962	1.105	1.196	1.417
$n = 200$	u_3	(0.432, 0.884)	0.452	0.964	0.867	0.879	0.915
	u_4	(0.125, 0.405)	0.28	0.963	0.397	0.401	0.427
	u_5	(0.0018, 0.0076)	0.0058	0.961	0.0069	0.0074	0.0085

CI: Confidence Interval; CIu: upper end point of one-sided Confidence Interval; CP: Coverage Probability; AL: Average Length; $u_P = (1, 1, 2/3, 5/18, 1/18)$.

6 Discussion

The QSD u and the stationary distribution π have some common characteristics in Markov chain, for example, all vector and not more than one. The different aspect is that the QSD have a λ-invariant measure $(0 < \lambda \leq 1)$, i.e., the max eigenvalue of quasi stochastic matrix may be not one, but less than one. In our simulations and examples, we don't contain the sum of u to equal one. Our QSDs are different with the QSDs in continuous-time finite Markov chains. Here we only vary the boundary of π, in order to apply it.

This paper presents a bootstrap method for computing ML estimate of quasi stationary distribution from Markov chain. It was demonstrated that the use of the bootstrap method can improve the performance of ML of certain parameters of the Markov chain, in the same time, avoids its shortcomings of parameter solution, and shows the good nature. An empirical study suggested under what conditions the methods were reliable. A key idea

of the methods presented in this paper are building a large sample setting, though further study may be required to perform reasonably in a variety of situations. The methods proposed here can be easily extended to include any other parameter of a Markov chain by changing the function of the appropriate calculations.

It is the disadvantage that the bootstrap method may not perform well, when the matrix $\hat{P}$ may not have a structure that is close to that of P.

References

[1] Athreya, K. B., and Fuh, C. D. (1992). Bootstrapping Markov chains: countable case. Journal of statistical planning and inference, 33(3), 311-331.

[2] Artalejo, J. R., Economou A. and López-Herrero M. J. (2010). The maximum number of infected individuals in SIS epidemic models: Computational techniques and quasi-stationary distributions. J. Comput. Appl. Math., 233, 2563-2574 .

[3] Basawa, I. V., Green, T. A., McCormick, W. P., Taylor, R. L. (1990). Asymptotic bootstrap validity for finite Markov chains. Communications in Statistics-Theory and Methods, 19, 1493-1510.

[4] Bertail, P., and Clémenon, S. (2006). Regenerative block bootstrap for Markov chains. Bernoulli, 12(4), 689-712.

[5] Bertail, P., and Clémenon, S. (2008). Approximate regenerative-block bootstrap for Markov chains. Computational Statistics and Data Analysis, 52(5), 2739-2756.

[6] Cattiaux, P. and Méléard, S. (2010). Competitive or weak cooperative stochastic lotka-volterra systems conditioned to non-extinction. J. Math. Biology., 6: 797-829.

[7] Coolen-Schrijner, P. and Doorn, E. A. Van, (2006) Quasi-stationary distributions for a class of discrete-time Markov chains. Methodol. Comput. Appl. Probab., 8: 449-465.

[8] Datta, S., McCormick, W. P. (1992). Bootstrap for a finite state Markov chain based on I.I.D. resampling. Exploring the Limits of Bootstrap. R. LePage and L. Billard, Editors. New York: John Wiley and Sons. pp. 77-97.

[9] Doorn, E.A. Van and Pollett, P. K. (2009) Quasi-stationary distributions for reducible absorbing Markov chains in discrete time. Markov Processes Relat. Fields. 15, 191-204.

[10] Doorn, E.A. Van and Pollett, P. K. (2008). Survival in a quasi-death process. Linear Algebra Appl., 429, 776-791.

[11] Fuh, C. D. (1989). The bootstrap method for Markov chains. PHD Thesis, Iowa State University, 1989.

[12] Fuh, C. D. (1993). Statistical inquiry for Markov chains by bootstrap method. Statistica Sinica, 3, 53-66.

[13] Fuh, C. D., Fan, T. H., and Hung, W. L. (2000). Balanced importance resampling for Markov chains. Journal of statistical planning and inference, 83(1), 221-241.

[14] Guo, G. (2012). Parallel Statistical Computing for Statistical Inference. Journal of Statistical Theory and Practice. 6, 536-565.

[15] Guo, G. and Zhao, W. (2012). Schwarz Methods for Quasi Stationary Distributions of Markov Chains. Calcolo, 49 , 21-39.

[16] Hall, P., 1990. Performance of balanced bootstrap resampling in distribution function and quantile problems. Probab. Theory Related Fields 85, 239-260.

[17] Hall, P. The Bootstrap and Edgeworth Expansion. Springer-Verlag: London, 1992.

[18] Hall, P. and Miller, H. (2009). Using the bootstrp to quantify the authority of an empical ranking. The Annals of Statistics, 37, 3929-3959.

Antieigenvalue Analysis, New Applications: Continuum Mechanics, Economics, Number Theory

Karl Gustafson[1,*]

[1] University of Colorado at Boulder, USA

Abstract My recent book Antieigenvalue Analysis, World-Scientific, 2012, presented the theory of antieigenvalues from its inception in 1966 up to 2010, and its applications within those forty-five years to Numerical Analysis, Wavelets, Statistics, Quantum Mechanics, Finance, and Optimization. Here I am able to offer three further areas of application: Continuum Mechanics, Economics, and Number Theory.

1   Introduction

Antieigenvalue analysis [1] is an operator trigonometry concerned with those vectors, called antieigenvectors, which are most-turned by a matrix or a linear operator A. This is in contrast to the conventional eigenvalue analysis, which is concerned with those vectors, called eigenvectors, which are not turned at all by A. Antieigenvalue theory may be usefully thought of as a variational theory, extending the variational Rayleigh-Ritz theory which characterizes eigenvectors, to an enlarged theory also characterizing antieigenvectors.

Two key entities in the antieigenvalue theory are the first antieigenvalue

$$\mu_1 = \cos\phi(A) = \min_{x \neq 0} \frac{<Ax, x>}{\|Ax\|\|x\|} \tag{1}$$

and the related convex minimum

$$\nu_1 = \sin\phi(A) = \min_{\varepsilon > 0} \|\varepsilon A - I\| . \tag{2}$$

Here I will specialize the antieigenvalue theory to A, an $n \times n$ symmetric positive definite matrix. One generally has the fundamental relation

$$\cos^2\phi(A) + \sin^2\phi(A) = 1. \tag{3}$$

There are two maximally turned first antieigenvectors

$$x_\pm = \left(\frac{\lambda_n}{\lambda_1 + \lambda_n}\right)^{\frac{1}{2}} x_1 \pm \left(\frac{\lambda_1}{\lambda_1 + \lambda_n}\right)^{\frac{1}{2}} x_n \tag{4}$$

where $0 < \lambda_1 \leqq \lambda_2 \leqq \ldots \leqq \lambda_n$ are the eigenvalues of A, and where x_1 is any norm-one eigenvector from the λ_1-eigenspace and x_n is any norm-one eigenvector from the λ_n-eigenspace. The antieigenvectors $x_\pm$ in (4) have also been normalized to be of norm-one. For $n \times n$ symmetric positive definite A the expressions in (1) and (2) have useful explicit valuations as

$$\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_n}}{\lambda_1 + \lambda_n} \, , \; \nu_1 = \frac{\lambda_n - \lambda_1}{\lambda_n - \lambda_1}. \tag{5}$$

For further elaboration of the general antieigenvalue theory I refer to [1]. In particular, just as one may move up the successive eigenvalue ladder via the Rayleigh-Ritz variational quotient minimizations, one can analogously move up an antieigenvalue ladder $\mu_2, \mu_3, \ldots$ via the variational quotient minimizations of (1), thereby arriving at a decreasing sequence of critical turning angles $\phi_k(A)$. But we won't need those higher antieigenvalues in the discussions of this paper.

In Sections 2, 3 and 4, respectively, I will briefly summarize three new applications of the antieigenvalue analysis. A full development of each will appear in the three forthcoming papers ([2],[3],[4]), respectively.

I would like to express my thanks to Jeff Hunter and Simo Puntanen as chairs of the IWMS-2015 for inviting me and to ILAS for designating me as their lecturer for the conference.

2 Continuum Mechanics(Granular Materials)

The following extracts a main result from the forthcoming paper [2]. My investigations there lean heavily on the recent paper [5] which I would urge the reader to consult for further background on the modeling of granular materials. In [2] I also go further to make connections to sandpile theories which have been useful to model self-organizing criticality in Statistical Mechanics [6].

The paper [5] explores the notion of (maximum) angle of repose for granular materials. On the other hand, my theory of antieigenvalues [1] has as one of its essential ingredients the notion of (maximum) operator turning angle. Here is how to connect the two theories.

Following [5], the equilibrium equations for a granular pile of local slope θ are

$$\begin{array}{ccccc} \partial_x \sigma_{xx} & + & \partial_z \sigma_{xz} & = & \rho g \sin \theta \\ \partial_x \sigma_{xx} & + & \partial_z \sigma_{xz} & = & \rho g \cos \theta \end{array}. \tag{6}$$

The stress tensor Σ can be written in singular value decomposition

$$\Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{xz} & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{bmatrix} \tag{7}$$

where $\sigma_1 \geqq \sigma_2 > 0$ are the principal stresses and where ψ gives the principal directions.
By considering a plane within the material and the normal and tangential stresses upon it in terms of the coefficient of friction of the material and a corresponding angle δ of internal friction, it is deduced in [5] that the largest sustainable angle of repose θ is given by

$$\sin \theta = \frac{\tau}{\sigma}, \textit{where } \sigma = \frac{\sigma_1 + \sigma_2}{2}, \tau = \frac{\sigma_1 - \sigma_2}{2}. \tag{8}$$

Some assumptions in this modeling of the discrete by the continuous have of course been made. Among those are a linear dependence on the vertical z direction and a stress-free condition at the pile's surface at $z = 0$.

In [2] I am able to see and recast this continuum mechanical model for stable granular material piles into my antieigenvalue theory. The key is to remember that the unique ε for which the minimum in (2) is attained in known [1] to be $\varepsilon_m = \frac{2}{\sigma_1 + \sigma_2}$ for a 2×2 matrix with singular values σ_1 and σ_2. Then straightforward calculations confirm that

$$\sin \phi(\Sigma) = \|\varepsilon_m \Sigma - I\| = \frac{\tau}{\sigma}. \tag{9}$$

Theorem ([2], Thm 3.1). The (maximum) angle of repose θ, the material angle of friction δ, and the (maximum) turning angle $\phi(\Sigma)$ of the stress tensor, are the same.

3   Economics (Capital Asset Pricing Model)

The following extracts from the forthcoming paper [3]. I presented a preliminary version at the IWMS2013 conference in Toronto. The paper [7] ensuing from that conference was accepted and may be accessed online although it is not in print yet at this writing. It contains a rather complete articulation, in order, of randomness, risk, and reward in financial markets, along with a number of important bibliographical references. I refer you here to the excellent book [8] for aspects of the CAPM as it is used in portfolio design theory and in high frequency trading, topics I have followed now for a considerable number of years.

The main finding which I exposed at the IWMS2013 and in the paper [7] and a bit earlier in the book [[1], p.182] is that the Sharpe Ratio, which is a key tool used in the CAPM theory and more importantly in practice, can be related to my first antieigenvalue μ_1. I first observed this in 1994 when I entered the financial engineering mathematical world during a very small consulting task, but did not work out the details until [7]. Here, in brief, is how that relationship may be seen.

The Capital Asset Pricing model assumes the Efficient Market Hypothesis and then tells you to measure the return-to-risk of your portfolio against the market. From the assumption that the full market has optimized the return-to-risk, your Sharpe ratio

$$S = \frac{E[r]}{\sigma[r]}, \tag{10}$$

where $E[r]$ is the average return over a number of periods and $\sigma[r]$ is the corresponding standard deviation, will not be greater than that of the whole (e.g, think indexing) market's Sharpe ratio. See especially [[8], fig 5.2, p 55], to picture Sharpe ratios as mean-variance slopes. Here I am just dropping the risk-free return rate R_f from numerators, as it is effectively zero these days anyway.

Suppose now we look at the last two years of annualized returns r_1 and r_2. We may form the usual (arithmetic) Sharpe ratio $S_{AM} = \frac{r_1+r_2}{2\sigma}$ and also a (geometric) Sharpe ratio $S_{GM} = \frac{\sqrt{r_1 r_2}}{\sigma}$ and upon dividing the latter by the former we arrive at

$$G = \frac{S_{GM}}{S_{AM}} = \frac{2\sqrt{r_1 r_2}}{r_1 + r_2} \tag{11}$$

which is my first antieigenvalue μ_1 as seen from (5). Further details and refinements may be found in [3, 7]. There I also begin an accompanying treatment of geometric versus arithmetic portfolio design. A referee of [7] also suggested possible connections to the currently important financial economic issues concerning realized volatilities, and I am in the process of such investigations in the paper under preparation [3].

4   Number Theory (Pythagorean Triples)

Here is a very interesting and new explicit connection of my antieigenvalue Theory and its operator trigonometry, to number theory. More details will be given in a paper under preparation [4], where a number of important further ramifications will also be developed.

Let me first point out and emphasize that when I first originated the antieigenvalue theory almost fifty years ago, I was coming from semi-group perturbation theory which had led me to a question of when an operator product BA would remain (real) positive, given positive A under multiplicative perturbation by positive B. For rather general semi-group generators A, and bounded B, I found the operator theoretic sufficient condition

$$\sin\phi(B) \leqq \cos\phi(A). \tag{12}$$

Then by using variational techniques on the expression (1) in conjunction with convexity techniques on the expression (2) I found the explicit valuations (5) for $n \times n$ symmetric positive definite matrices. See [1] for more details and history.

Now the new connection to Number Theory, which I only recently discovered. Given two arbitrary relatively prime positive integers m and n, with $m > n$, one of them being even, the other odd, then the numbers

$$a = 2mn, \ b = m^2 - n^2, \ c = m^2 + n^2 \tag{13}$$

form a primitive Pythagorean triple:

$$a^2 + b^2 = c^2. \tag{14}$$

This sufficient condition is also necessary. For more details see [9]. This construction and characterization of Pythagorean triples is often called Euclid's Formula.

I may now form the matrix (and its similarity class of matrices with the same eigenvalues m^2 and n^2)

$$A = \begin{bmatrix} n^2 & 0 \\ 0 & m^2 \end{bmatrix}. \tag{15}$$

Immediately from my matrix operator trigonometry [1] and the expressions (5) we have

$$\cos\phi(A) = \frac{2mn}{m^2 + n^2}, \ \sin\phi(A) = \frac{m^2 - n^2}{m^2 + n^2}. \tag{16}$$

Proposition ([4]). Euclid's Formula for Pythagorean Triples is a special case of my operator trigonometry.

We may propose to call these matrices $A_{m,n}$, Pythagorean Triple Matrices. Their maximum turning angles may be called special Pythagorean turning angles $\phi_{m,n}(A)$. Their corresponding normalized Pythagorean antieigenvectors are

$$x_\pm = \left(\frac{m^2}{m^2 + n^2}\right)^{\frac{1}{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \pm \left(\frac{n^2}{m^2 + n^2}\right)^{\frac{1}{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{m^2 + n^2}} \begin{bmatrix} m \\ \pm n \end{bmatrix}. \tag{17}$$

We know of course there are an infinite number of these Pythagorean angles, which are now embedded within my antieigenvalue operator trigonometry.

This new connection of my antieigenvalue analysis to the Pythagorean triple number theory may be seen to have other interesting manifestations. Here is another one, couched in the terminology of algebraic geometry. Let $x = \left(\frac{m}{n}, 0\right)$ be a point on the x-axis. It's stereographic projection onto the unit circle becomes, now seen operator-theoretically, the point

$$P = \left(\frac{2mn}{m^2 + n^2}, \frac{m^2 - n^2}{m^2 + n^2}\right) = \left(\cos\phi(A_{m,n}), \ \sin\phi(A_{m,n})\right). \tag{18}$$

The stereographic point of view comes from a treatment of spinors and twistors [10].

References

[1] Gustafson, K. (2012). *Antieigenvalue Analysis, with Applications to Numerical Analysis, Wavelets, Statistics, Quantum Mechanics, Finance and Optimization.*, World-Scientific, Singapore.

[2] Gustafson, K. (2015). *The antieigenvalue hiding in the sandpile.*, Preprint, to appear.

[3] Gustafson, K. (2015). *A new growth-angle theory for the Capital Asset Pricing Model.*, To appear.

[4] Gustafson, K. (2015). *Antieigenvalue analysis and number theory*, To appear.

[5] Gordon-Wright, R, and Gremaud, P.A. (2013). *Granular free surfaces*, SIAM Review 55, pp. 168-183.

[6] Dhar, D.(2006). *Theoretical studies of self-organized criticality.*, Physica A 369, pp. 29-70.

[7] Gustafson, K. (2015). *A new finomcial risk ratio.*, Journal of Statistical Computation and Simulation. DOI: 10.1080/00949655.2014.918131

[8] Aldridge, I. (2010). *High Frequency Trading.*, Wiley & Sons, Hoboken, NJ.

[9] Maor, E. (2007). *The Pythagorean Theorem.*, Princeton University Press, Princeton, N.J.

[10] Trautman, A.(1998). *Pythagorean spinors and Penrose twistors. In: The Geometric Universe (eds. S.A. Huggett, L.J. Mason, K.P. Tod, S.T. Tsou, N.M.J. Woodhouse)*, Oxford University Press, Oxford, U.K., pp. 411-419.

Comparison of facial recognition methods based on extension methods of Principal Component Analysis

Guanyu Hu[1]

[1]Florida State University

Abstract   The human face recognition system becomes more and more universal in different parts of our society like governments, banks and social welfare. How to improve the efficiency of the discrimination of human face is the most significant issue in the human face recognition system. And it is also a hot issue in the high dimensional analysis.In this paper, we compared the performance of the facial recognition of three methods, EMPCA, Sparse PCA and Kernel PCA. These three methods based on the general idea of Principal Component Analysis and Fisher's Linear Discriminant. Our study based on the ORL face data base. EMPCA used same idea of the classic PCA by using the EM algorithm. The Sparse PCA extends the classic PCA by adding sparsity constraint in order to explain the variables more clear. The Kernel PCA extends the classic PCA by using kernel methods. The comparison of these three methods in facial recognition can help us know much more about the applications of the dimension reduction methods.

Keywords   Facial Recognition; Principal Component Analysis; EM algorithm; Sparse PCA; Kernel PCA; Dimension Reduction

1   Introduction

The human face recognition has brought great attention to the temporary society for both commercial and research, especially when it's application can be practically adopted in different areas, like human identifications, computer visions and criminal forensics. There are many different algorithms of the human face recognition system like Feature-based recognition algorithms, Appearance-based recognition algorithms, Template-based recognition algorithms and Recognition algorithms using neural network. It's difficult to develop a computational model of face recognition, because the human faces are complex, multidimensional, and meaningful visual stimuli. It's a big challenge to find a proper model which can discriminate human faces efficiently.

In this paper, I used most common method in face recognition, principal component analysis (PCA). The algorithm of face recognition by PCA was first introduced by Matthew Turk and Alex on 1991. They first introduced the concept of the Eigenfaces and transformed the high-dimension data sets into the low-dimension. Because the most difficult part of the face recognition is to deal with the high-dimension data. Principal Component Analysis provides us a possible way to solve this topic. However, Principal Component Analysis has some drawbacks in the dimension reduction problem, I will discuss in the section 2. In order to improve the method in the face recognition, I introduced some extension methods from Principal Component Analysis, like EM PCA, Sparse PCA and Kernel PCA. These three methods based on the core idea of Principal Component Analy-

sis, data transformation. In the following part, I will give a brief introduction about these three methods and their application in the face recognition.

2 Methodology

In this section, I will introduce my methodology for the face recognition problem. In general, I use two step to accomplish the face recognition. Step one is feature extraction. And the step two is classification. What I concentrate is the method of the feature extraction. So in the following parts I will introduce what I used in the face recognition. In the step two, I do the classification in three steps.

(a)**Form feature vector**: Compute the projection of the test image.
(b)**Compute Metric**: Compute the distant between the projection and each images using the 2-norm.
(c)**Find Nearest Neighbors**: Find the label of the images that has the smallest distance to the projection.

So, the following four parts will give a brief view of the methods I used in the feature extraction.

2.1 Principal Component Analysis

Principal component analysis (PCA) is a very traditional method which can be applied into feature extraction and data representation. The goal of PCA is to find a linear and low-dimensional subspace of the original feature space which is high dimensional. The algorithm of the PCA according to the covariance matrix is following:

According to the algorithm above, I can extract the feature from my training data.

Algorithm 1 Principal Component Analysis

1: Compute the covariance matrix of the data
2: Compute the eigenvectors and eigenvalues of the covariance matrix
3: Rearrange the eigenvectors and eigenvalues
4: Compute the cumulative energy content for each eigenvector
5: Select a subset of the eigenvectors as basis vectors
6: Convert the source data to z-scores
7: Project the z-scores of the data onto the new basis

But the PCA has some disadvantages of the feature extraction:
1) It's difficult to calculate the covariance matrix in an accurate manner.
2) If the training doesn't provide some information, PCA cannot capture even the simplest invariance.
3) It's difficult to give a brief explanation of the each components.

2.2 EM PCA

EM algorithm was introduced by Arthur Dempster, Nan Laird, and Donald Rubin in 1977. They proposed an iterative method to find the maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models. There are two steps in the EM algorithm. E-step is to find the expectation of the log-likelihood using the existence estimate of the parameters. M-step is to compute the parameters maximizing the expected log-likelihood. The algorithm of EM Algorithm is following:

Algorithm 2 EM Algorithm

1: Expectation Step: Compute:

$$Q(\theta \mid \theta_k, x_o) = E[log f(x_o, x_m \mid \theta) \mid \theta_k, x_o]$$

2: Maximization Step: Set
$$\theta_{k+1} = \arg \max_{\boldsymbol{\theta}} Q(\theta|\theta_k, x_o)$$

3: Check the convergence. If not, set $k = k+1$ and go to the first step.(x_o is the observed data, x_m is the missing data.)

Applying EM algorithm into PCA gives us three advantages: 1) Traditional PCA has trouble with high dimensional data or large numbers of datapoints. When n and p are large, there are so many difficulties in the form of the computational complexity and data scarcity. In some degree, the EM algorithm can easy the computational complexity.
2) Traditional PCA cannot deal with the problem which has the missing data. The EM algorithm for PCA can estimate the maximum likelihood values in the missing data condition.
3) Traditional PCA does not define the data's probability density. In this way, we have limited criterion to evaluate the fitness of the model. However EM PCA can solve this kind of problem.
So, the algorithm of the EM PCA(1998,Sam Roweis) is following:

Algorithm 3 EM PCA Algorithm

1: Expectation Step:
$$X = (C^T C)^{-1} C^T Y$$

2: Maximization Step:
$$C^{new} = Y X^T (X X^T)^{-1}$$

---

where Y is a $p \times n$ matrix of all the observed data and X is a $k \times n$ matrix of the unknown states. The columns of C will span the space of the first k principal components.

2.3 Sparse PCA

Sparse Principal Component Analysis(SPCA) was introduced by Dr. Zou, Dr Hastie and Dr.Tibshirani in 2006. Compared with the classic principal component analysis which is a linear combination of all the original variables, sparse principal component analysis(SPCA) uses the lasso(elastic net) to produce modified principal components with the sparse loadings. Sparse loadings can be obtained by estimating in the regression model with the lasso(elastic net) constraint. SPCA has some advantages in doing high-dimensional data reducing. Because, PCA has an obvious drawback. In the classic PCA, each principal component is just a linear combination of all the variables of the observed data. In this way, it's difficult for us to interpret the principal components. SPCA not only solves the problem of dimensionality reduction, but also reduces the number of explicitly of the used variables by using the lasso. SPCA is built on the general fact of the PCA which transformed into a optimization problem of the regression model with the quadratic penalty and lasso penalty. In order to get deep understanding of SPCA, I will introduce the general idea of lasso first.

LASSO: A linear regression model has n observations and p predictors. The LASSO estimators is following:

$$\hat{\beta}_{lasso} = \arg\max_{\beta} ||Y - \sum_{j=1}^{p} X_j\beta_j||^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

In this formula, $Y = (y_1, ..., y_n)^T$ is the response vector, $X = [X_1, .., X_p]$, where $X_j = (x_{1j}, ..., x_{nj})^T$, λ is non-negative. In 1996, Dr.Tibshirani gave a solution to this estimator. According to the general idea of LASSO. Dr.Zou proposed the general method of SPCA which transforms the traditional method into the optimization problem. The algorithm of SPCA(2006,H.Zou) is following:

Algorithm 4 SPCA Algorithm

1: Let A start at $V[,1:k]$, the loadings of first k ordinary principal components
2: Given a fixed $A = [\alpha_1, ..., \alpha_k]$, solve the following elastic net problem for $j = 1, 2, ..., k$

$$\beta_j = \arg\min_{\beta}(\alpha_j - \beta)^T X^T X(\alpha_j - \beta) + \lambda||\beta||^2 + \lambda_{1,j}||\beta||_1$$

3: For a fixed $B = [\beta_1, .., \beta_k]$, compute the SVD of $X^T X B = UDV^T$, then update $A = UV^T$
4: Repeat Steps 2-3, until convergence
5: Normalization: $\hat{V}_j = \frac{\beta_j}{||\beta_j||}, j = 1, .., p$

2.4 Kernel PCA

Dr.Schölkopf introduced the Kernel method in PCA in 1999. The kernel method using in the PCA extends the originally linear operations of PCA into kernel Hilbert space with a non-linear mapping.

First, the Kernel method must transform the data into the nonlinear mapping $\phi(x)$, and assume the projected new features have zero mean:

$$\frac{1}{N} \sum_{i=1}^{N} \phi(x_i) = 0$$

And then, calculate the covariance matrix of the projected feature:

$$C = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i)\phi(x_i)^T$$

According to the covariance matrix, we can get the eigenvalues and eigenvectors:

$$Cv_k = \lambda_k v_k$$

Get the representation of v_k:

$$\frac{1}{N} \sum_{i=1}^{N} \phi(x_i)\phi(x_i)T \sum_{j=1}^{N} a_{kj}\phi(x_j) = \lambda_k \sum_{i=1}^{N} a_{ki}\phi(x_i)$$

And we define the kernel function like:

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

Use the matrix notation and kernel function

$$\boldsymbol{K}^2 \boldsymbol{a}_k = \lambda_k N \boldsymbol{K} \boldsymbol{a}_k$$

where $K_{i,j} = k(x_i, x_j)$, $\boldsymbol{a}_k$ is $[a_{k1}, ... a_{kN}]$. $\boldsymbol{a}_k$ can be solved by

$$\boldsymbol{K} \boldsymbol{a}_k = \lambda_k N \boldsymbol{a}_k$$

Finally we can get the kernel principal components

$$y_k(\boldsymbol{x}) = \phi(\boldsymbol{x})^T \boldsymbol{v}_k = \sum_{i=1}^{N} a_{ki} k(\boldsymbol{x}_i, \boldsymbol{x})$$

If the mean of the projected datasets is not zero, we can use the Gam matrix

$$\tilde{\boldsymbol{K}} = \boldsymbol{K} - 1_N \boldsymbol{K} - \boldsymbol{k} 1_N + 1_N \boldsymbol{K} 1_N$$

where 1_N is the $N \times N$ matrix and all the elements equal $1/N$.(Bishop, 2006)
So we can get the algorithm of the kernel PCA(Q. Wang, 2011)

Algorithm 5 kernel PCA Algorithm

1: Construct the kernel matrix from the training data
2: Compute the Gam matrix $\tilde{\boldsymbol{K}}$
3: Solve the vectors $\boldsymbol{a}_i$ (substitute $\boldsymbol{K}$ with $\tilde{\boldsymbol{K}}$)
4: Compute the kernel principal components

---

3 Face Recognition Results

In this section, I will give the results of the simulation study. In the simulation, I used ORL database. In this database, each face's size is 64×64. Totally, I separate data into two sets. One is training data, the other is test data. Both two data have 40 people, each people has 5 different faces.To compute an average performance, I performed the above procedure in section 3 1000 times and computed the percentage of successful recognition. In addition to compare the speed of each algorithm, I computed the time of method in the feature extraction step. The following two figures show the results:
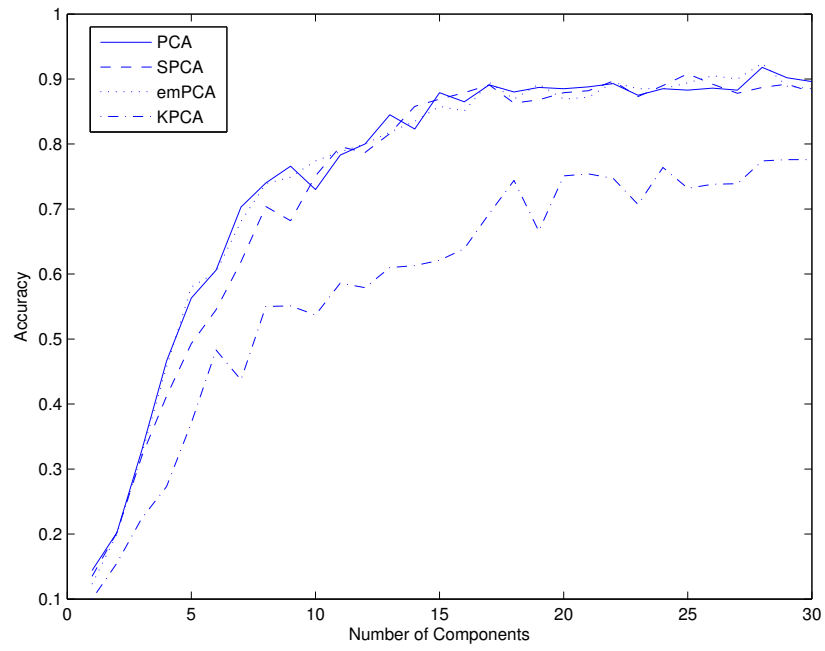
Figure 1: figure of recognition accuracy of four methods
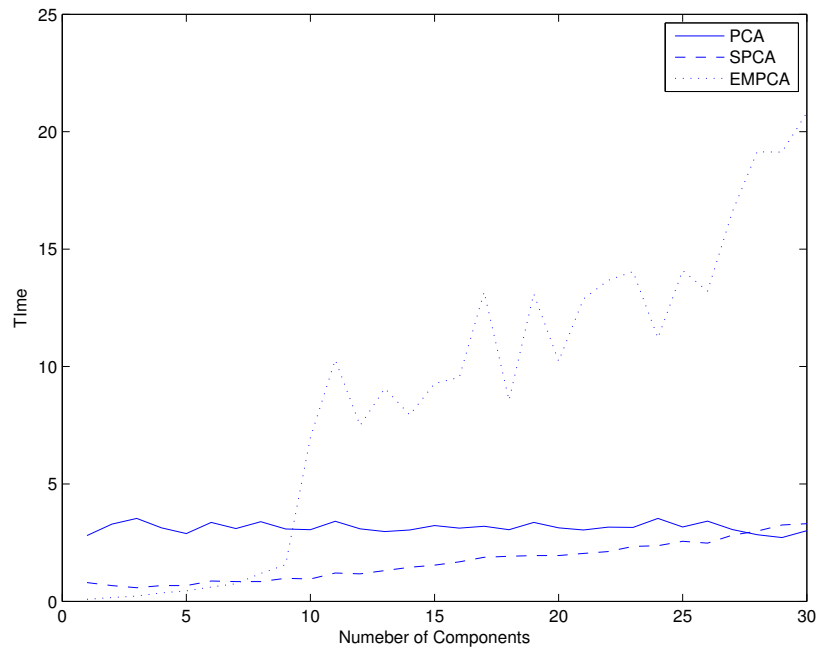
Figure 2: figure of time of four methods

4   Conclusion

In the results from the section 3, we can find that expect kernel PCA, classic PCA, Sparse PCA and EM PCA have same recognition rate for this database. But in the feature extraction time, there is difference between the three methods. Classic PCA algorithm will calculate the covariance matrix every time, so the time of extraction have significance difference in different numbers of the component. But the EM algorithm will have advantages in the low numbers of the components due to its calculation method. Sparse PCA also has this advantages. So in the problem which need low numbers of component, EM algorithm is preferred. Kernel PCA in this classification has lower accuracy than other three methods. At same time, as Kernel PCA will do some projection to other space, its time is much bigger than other three methods.

5   Future Work

In this paper, I only use the Nearest Neighbors in the procedure of classification. There are no significant differences between the sparse PCA and classic PCA. So in the future work, I need to do other research in the comparison of the classification method's effects on the feature extraction methods.

Acknowledgement

References

[1] Bishop, Christopher M ,(2006), Pattern Recognition and Machine Learning., *Springer*.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin,(1977), Maximum likelihood from incomplete data via the EM algorithm? *Journal of the Royal Statistical Society series B*, 39:1-38.

[3] Q. Hu, D. Yu, and Z. Xie (2008), Neighborhood classifiers, *Expert Systems with Applications*, 866-876.

[4] Jolliffe I.T,(2002), A Statistical PCA Method for Face Recognition, *Springer Series in Statistics*.

[5] C. Li, Y. Diao, H. Ma and Y. Li,(2008), A Statistical PCA Method for Face Recognition, *AIntelligent Information Technology Application*, 376-80.

[6] Matthew Turk and Alex Pentland,(1991), Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Volume 3.

[7] Robert Tibshirani (1996), Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society. Series B(Methodological)*, Volume 58 Issue 1, 267-288.

[8] Roweis, S,(1997), EM Algorithms for PCA and SPCA, *CNS Technical Report*.

[9] Schölkopf, B. and Smola, AJ. and Müller,(1999), Kernel principal component analysis, *Advances in Kernel Methods—Support Vector Learning*, 327-352.

[10] H. Zou, T. Hastie, and R. Tibshirani (2006), Sparse principal component analysis, *Journal of Computational and Graphical Statistics*, 265-286.

The accurate computation of the key properties of Markov chains and Markov renewal processes

Jeffrey J Hunter

School of Computer and Mathematical Sciences, Auckland University of Technology,
New Zealand

Abstract   Based upon the Grassman, Taksar and Heyman (1985) algorithm and the equivalent Sheskin (1985) State Reduction algorithm for finding the stationary distribution of a finite irreducible Markov chain, Kohlas (1986) developed a procedure for finding the mean first passage times (MFPTs) (or absorption probabilities) in Markov renewal processes. The method is numerically stable as it doesn't involve subtraction. It works well for focusing on the MFPTs from any state to a fixed state but it is not ideally suited for a global expression for the MFPT matrix. We present some refinements to the Kohlas algorithm that we specialize to the case of Markov chains. We utilise MatLab to find expressions for the MFPT matrix. A consequence of our procedure is that the stationary distribution does not need to be derived in advance but is found from the MFPTs. This also leads to an expression for the group inverse of $I - P$ where P is the transition matrix of the embedded Markov chain. Some comparisons, using some test problems from the literature, with other techniques using generalized matrix inverses and perturbation techniques are also presented.

Keywords   Markov chain, Markov renewal process, stationary distribution, mean first passage times, generalized inverse, group inverse, perturbations.

1   Introduction

Let $P = [p_{ij}]$ be the transition matrix of an irreducible, discrete time Markov chain (MC) $\{X_k\}$, $(k \geq 0)$, with finite state space $S = \{1, 2, \ldots, N\}$ i.e. $p_{ij} = P\{X_k = j | X_{k-1} = i\}$ for all $i, j \in S$.

We are interested in developing accurate ways of finding three key properties of such chains:

 (i)   the stationary probabilities $\{\pi_j\}$, $(1 \leq j \leq N)$, and

 (ii) the mean first passage times $\{m_{ij}\}, (1 \leq i, j \leq N)$.

 (iii) the group inverse of the matrix $I - P$.

2   Stationary distributions of Markov chains

Let $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \ldots, \pi_N)$ be the stationary probability vector of the Markov chain with transition matrix $P = [p_{ij}]$. We need to solve $\pi_j = \sum_{i=1}^{N} \pi_i p_{ij}$ with $\sum_{i=1}^{N} \pi_i = 1$,

i.e. $\boldsymbol{\pi}^T (I - P) = \boldsymbol{0}^T$ with $\boldsymbol{\pi}^T \boldsymbol{e} = 1$.

3   Mean first passage times

Let T_{ij} be the first passage time RV from state i to state j,

i.e. $T_{ij} = \min \{k \geq 1$ such that $X_k = j$ given that $X_0 = i\}$. T_{ii} is the first return to state i.

Let $m_{ij} = E[T_{ij} | X_0 = i]$, the mean first passage time from state i to state j.

Let $M = [m_{ij}]$ be the matrix of mean first passage times.

It is well known that $m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}$, with $m_{jj} = 1/\pi_j$.

M satisfies the matrix equation $(I - P)M = E - PD$,

where $E = [1] = \boldsymbol{e}\boldsymbol{e}^T$, and $D = M_d = [\delta_{ij} m_{ij}] = (\Pi_d)^{-1}$ (with $\Pi = \boldsymbol{e}\boldsymbol{\pi}^T$).

4 The Group Inverse

The group inverse $A^{\#}$ of $I - P$ is a special generalized matrix inverse of $I - P$ that has many desirable properties. For finite irreducible Markov chains, $A^{\#} = [I - P + \Pi]^{-1} - \Pi$. The calculation of $A^{\#}$ involves the evaluation of a matrix inverse which is prone to round off errors. Due to time constraints, we do not present results on possible calculations other than to refer the reader to the paper by Hunter ([8]) where it is shown that the elements of $A^{\#} = \left[a_{ij}^{\#} \right]$ can be expressed in terms of the stationary probabilities and the mean first passage times of the M.C. In particular, if

$$\tau_j = \sum_{k=1}^{N} \pi_k m_{kj} = \sum_{k \neq j} \pi_k m_{kj} + 1, \quad \text{then } a_{ij}^{\#} = \begin{cases} \pi_j(\tau_j - 1 - m_{ij}), & i \neq j, \\ \pi_j(\tau_j - 1), & i = j. \end{cases}$$

5 Solving for Stationary Distributions

If $G = [I - P + tu^T]^{-1}$ where u, t such that $u^T e \neq 0$, $\pi^T t \neq 0$, then $\pi^T = \dfrac{u^T G}{u^T G e}$.

(Paige, Styan,Wachter,[12]), (Kemeny, [10]), (Hunter, [6]).

In particular if $G = [I - P + eu^T]^{-1}$ then $\pi^T = u^T G$.

6 GTH Algorithm

The most accurate general procedure for finding the stationary probabilities is the GTH algorithm as developed by Grassman, Taksar and Heyman ([2]). An equivalent state reduction procedure was given by Sheskin ([13]).

Let $P_N = \left[p_{ij} \right] \equiv \left[p_{ij}^{(N)} \right]$ be the $N \times N$ transition matrix associated with a M.C. $\{X_k, k \geq 0\}$, with state space $S_N = \{1, 2, ..., N\}$, and one-step transition probabilities $p_{ij}^{(N)} = P\{X_{k+1} = j | X_k = i\}$. The general approach is to start with an N–state Markov chain and reduce the state space by one state at each stage. Thus in stages $S_N = S_{N-1} \cup \{N\}$, $S_{N-1} = S_{N-2} \cup \{N-1\}$,...., $S_2 = \{1, 2\}$. Once we get to two states we expand the state space one state at a time until we return to the final set of states S_N.

Assume that the initial M.C. with state space S_N is irreducible and that stationary probability vector is given by $\pi^T = (\pi_1, \pi_2, ..., \pi_{N-1}, \pi_N)$. Let $\pi^T = \pi^{(N)T} = (\pi_1^{(N)}, \pi_2^{(N)}, ..., \pi_{N-1}^{(N)}, \pi_N^{(N)})$.

From the stationary equations $\pi^{(N)T} = \pi^{(N)T} P_N$ or in element form $\pi_j^{(N)} = \sum_{i=1}^{N} \pi_i^{(N)} p_{ij}^{(N)}$,

$(j = 1, 2, ..., N)$, express $\pi_N^{(N)}$ in terms of $\pi_1^{(N)}$,..., $\pi_{N-1}^{(N)}$, i.e. $\pi_N^{(N)} = \dfrac{\sum_{i=1}^{N-1} \pi_i^{(N)} p_{iN}^{(N)}}{\sum_{j=1}^{N-1} p_{Nj}^{(N)}}$ and eliminate

$\pi_N^{(N)}$ from the stationary equations.

Let $P_N = \begin{bmatrix} Q_{N-1}^{(N)} & p_{N-1}^{(N)(c)} \\ p_{N-1}^{(N)(r)T} & p_{NN}^{(N)} \end{bmatrix}$.

Partition the stationary probability vector $\pi^{(N)T} = (v^{(N-1)T}, \pi_N^{(N)})$,

where $v^{(N-1)T} = (\pi_1^{(N-1)}, \pi_2^{(N-1)}, ..., \pi_{N-1}^{(N-1)})$. It is easily shown that

$v^{(N-1)T}(I_{N-1} - P_{N-1}) = 0^T$, where $P_{N-1} = Q_{N-1}^{(N)} - \dfrac{p_{N-1}^{(N)(c)} p_{N-1}^{(N)(r)T}}{p_{N-1}^{(N)(r)T} e^{(N-1)}}$.

Let $P_{N-1} = \left[p_{ij}^{(N-1)} \right]$ then $p_{ij}^{(N-1)} = p_{ij}^{(N)} + \dfrac{p_{iN}^{(N)} p_{Nj}^{(n)}}{S(N)}$, $1 \leq i \leq N-1$, $1 \leq j \leq N-1$.

Note that calculation of the $S(N)$ and the $p_{ij}^{(N-1)}$ do not involve subtractions.

Observe that P_{N-1} is a stochastic matrix with state space S_{N-1}, P_{N-1} is irreducible and that $v^{(N-1)T}$ is a scaled stationary prob vector of this $N-1$ state M.C. In particular

$$\boldsymbol{\pi}^{(N-1)T} = (\pi_1^{(N-1)}, \pi_2^{(N-1)}, ..., \pi_{N-1}^{(N-1)}) \equiv \frac{1}{1-\pi_N^{(N)}} v^{(N-1)T}, \text{ so that the first } N-1 \text{ stationary probs of the}$$

N-state M.C. are scaled versions of the $N-1$ state M.C.

We can repeat this process reducing the state space from n to $n-1$ ($n = N, N-1, ..., 2$) with the resulting M.C. with state space S_{n-1} having a stationary distribution that is a scaled version of the first $n-1$ components of the stationary distribution of the M.C. with n states.

Thus if $P_n = \left[p_{ij}^{(n)} \right]$ with $P_{n-1} = \left[p_{ij}^{(n-1)} \right]$ then $p_{ij}^{(n-1)} = p_{ij}^{(n)} + \dfrac{p_{in}^{(n)} p_{nj}^{(n)}}{S(n)}$, $1 \le i \le n-1, 1 \le j \le n-1$, where

$$S(n) = 1 - p_{nn}^{(n)} = \sum_{j=1}^{n-1} p_{nj}^{(n)}.$$

The $p_{ij}^{(n-1)}$ can be interpreted as the transition probability from state i to state j of the M.C. on S_n restricted to S_{n-1}. (For $(i, j) \in S_{n-1} \times S_{n-1}$ it is possible to jump directly from state i to state j with probability $p_{ij}^{(n)}$. Alternatively, jump from i to j via state n, being held at state n for $t (= 0, 1, 2, ...)$ steps, followed by a jump to j, with probability $p_{in}^{(n)} \left(\sum_{t=0}^{\infty} \left(p_{nn}^{(n)} \right)^t \right) p_{nj}^{(n)} = \dfrac{p_{in}^{(n)} p_{nj}^{(n)}}{1 - p_{nn}^{(n)}} = \dfrac{p_{in}^{(n)} p_{nj}^{(n)}}{S(n)}$,

leading to the general expression for $p_{ij}^{(n-1)}$.)

Note that if the M.C. with state space S_N is irreducible (i.e. each state j can be reached from state i in a finite number of steps k) then the M.C. with state space S_{N-1} is also irreducible since there will still be a path from state j can be reached from state i in k or fewer steps (either avoiding N in the original M.C. or in a reduced number of steps if passing through N in the original M.C.)

If we start with $\boldsymbol{\pi}^{(N)T} = (\pi_1^{(N)}, \pi_2^{(N)}, ..., \pi_{N-1}^{(N)}, \pi_N^{(N)}) \equiv (\pi_1, \pi_2, ..., \pi_{n-1}, \pi_n)$ then the $N-1$ elements of $\boldsymbol{\pi}^{(N-1)T}$ are scaled elements of the first $N-1$ elements of $\boldsymbol{\pi}^{(N)T}$ and hence of $\pi_1, \pi_2, ..., \pi_{n-1}$. Thus each $\boldsymbol{\pi}^{(n)T}$ is a scaled version of $(\pi_1, \pi_2, ..., \pi_{n-1}, \pi_n)$.

The process continues to $n = 2$, where we have $P_2 = \begin{bmatrix} p_{11}^{(2)} & p_{12}^{(2)} \\ p_{21}^{(2)} & p_{22}^{(2)} \end{bmatrix}$ which is a stochastic matrix.

The stationary probability vector of this M.C. is given by $\boldsymbol{\pi}^{(2)T} = (\pi_1^{(2)}, \pi_2^{(2)})$ which will be a scaled version of (π_1, π_2). The second stationary equation is $\pi_2 = \pi_1 p_{12}^{(2)} + \pi_2 p_{22}^{(2)}$,

implying $\pi_2 = \pi_1 \dfrac{p_{12}^{(2)}}{S(2)}$, where $S(2) = 1 - p_{22}^{(2)} = \sum_{j=1}^{1} p_{2j}^{(2)} = p_{21}^{(2)} = \boldsymbol{p}_1^{(2)(r)T} e^{(1)}$.

We now proceed with increasing the state space.

$$\pi_3 = \frac{\sum_{i=1}^{2} \pi_i p_{i3}^{(3)}}{\sum_{i=1}^{2} p_{3i}^{(3)}} = \pi_1 \frac{p_{13}^{(3)}}{S(3)} + \pi_2 \frac{p_{23}^{(3)}}{S(3)}, \text{ In general, } \pi_n = \frac{\sum_{i=1}^{n-1} \pi_i p_{in}^{(n)}}{\sum_{i=1}^{n-1} p_{ni}^{(n)}} = \sum_{i=1}^{n-1} \pi_i \frac{p_{in}^{(n)}}{S(n)}.$$

If $\pi_j = k r_j$ with $r_1 = 1$ then $\sum_{i=1}^{N} \pi_i = 1$ so that $k = 1 / \sum_{i=1}^{N} r_i$ with

$$r_n = \frac{\sum_{i=1}^{n-1} r_i p_{in}^{(n)}}{S(n)}, (n = 2, ..., N), \text{ implying } \pi_i = \frac{r_i}{\sum_{n=1}^{N} r_n}, \quad i = 1, 2, ..., N.$$

Thus the GTH Algorithm procedure can be summarised as

1. Start with a Markov chain with N states and transition matrix $P_N = \left[p_{ij}^{(N)} \right]$.
2. Compute for $n = N, N-1, ..., 3$,

$$p_{ij}^{(n-1)} = p_{ij}^{(n)} + \frac{p_{in}^{(n)} p_{nj}^{(n)}}{S(n)}, \ 1 \le i \le n-1, 1 \le j \le n-1, \ \text{where } S(n) = \sum_{j=1}^{n-1} p_{nj}^{(n)}.$$

3. Set $r_1 = 1$ and compute $r_n = \dfrac{\sum_{i=1}^{n-1} r_i p_{in}^{(n)}}{S(n)}$, for $n = 2,...,N$.

4. Compute $\pi_i = \dfrac{r_i}{\sum_{j=1}^{N} r_j}$, $i = 1, 2, ..., N$.

7 Solving for Mean First Passage Times

(*i*) If G is any g-inverse of $I - P$, then $M = [G\Pi - E(G\Pi)_d + I - G + EG_d]D$. (Hunter, [6]).

(*ii*) $Ge = ge$ for some g if and only if $M = [I - G + EG_d]D$. (Hunter, [9]).

The "***standard algorithm***" is $M = [I - Z + EZ_d]D$ where $Z = [I - P + e\boldsymbol{\pi}^T]^{-1}$, Kemeny and Snell's "fundamental matrix".

Hunter ([7]) presented a "***simple algorithm***" which is the simplest method to simultaneoulsy compute the stationary distribution and the MFPTs.

If $G_{eb} = [I - P + ee_b^T]^{-1} = \left[g_{ij}\right]$, then $\pi_j = g_{bj}$, $1 \le j \le N$, and $m_{ij} = \begin{cases} 1/g_{bj}, & i = j, \\ (g_{jj} - g_{ij})/g_{bj}, & i \ne j. \end{cases}$

Hunter ([9]) developed a number of "***perturbation algorithms***" where the transition matrix is successively updated row by row from an initial simple transition matrix to end up with the required transition matrix. One such algorithm, which appears to be very accurate, is the following:

(*i*) Let $K_0 = I$.

(*ii*) For $i = 1, 2, ...,N$, let $\boldsymbol{p}_i^T = \boldsymbol{e}_i^T P$, $\boldsymbol{b}_i^T = \boldsymbol{p}_i^T - \boldsymbol{e}^T/N$,

$K_i = K_{i-1}\left(I + C_i\right)$, where $k_i = 1 - \boldsymbol{b}_i^T K_{i-1}\boldsymbol{e}_i$ and $C_i = \dfrac{1}{k_i}\boldsymbol{e}_i\boldsymbol{b}_i^T K_{i-1}$.

(*iii*) At $i = N$, let $K = K_N$ then $\boldsymbol{\pi}^T = \dfrac{1}{N}\boldsymbol{e}^T K$,

$M = [I - K + EK_d]D$, where $D = (\Pi_d)^{-1}$.

8 MFPT via the Extended GTH Algorithm

We seek a computational procedure for MFPTs, utilising the GTH/State reduction procedure.
We first summarise the key properties of Markov renewal process (M.R.P.).
From Hunter ([6]), let $\{(X_k, T_k)\}$, $(k \ge 0\}$, be a M.R.P. with state space S_N and semi-Markov kernel $Q(t) = \left[Q_{ij}(t)\right]$, where $Q_{ij}(t) = P\{X_{k+1} = j, T_{k+1} - T_k \le t | X_k = i\}$, $(i, j) \in S_N$.

Observe that X_k is the state at the *k-th* transition, T_k is the time of the *k-th* transition.

Let $P = \left[p_{ij}\right]$ be the transition matrix of the embedded M.C. $\{X_k\}$, $(k \ge 0)$.

Then $p_{ij} = Q_{ij}(+\infty) = P\{X_{k+1} = j | X_k = i\}$ and $Q_{ij}(t) = p_{ij}F_{ij}(t)$ where

$F_{ij}(t) = P\{T_{k+1} - T_k \le t | X_k = i, X_{k+1} = j\}$.

$F_{ij}(t)$ is the distribution function of the "holding time" $T_{k+1} - T_k$ in state X_k until transition into state X_{k+1} given that the M.R.P. makes a transition from state X_k to state X_{k+1}.

Let $\mu_{ij} = \int_0^\infty t\, dQ_{ij}(t)$ so that $\mu_{ij} = p_{ij}E[T_{k+1} - T_k | X_k = i, X_{k+1} = j]$.

Let $P^{(1)} = \left[\mu_{ij}\right]$ then $(I - P)M = P^{(1)}E - PM_d$.

Let $\boldsymbol{\mu} = P^{(1)}\boldsymbol{e}$ then $\boldsymbol{\mu}^T = (\mu_1, \mu_2, ..., \mu_N)$ where $\mu_i = \sum_{j=1}^{N} \mu_{ij}$.

$\mu_i = E[T_{n+1} - T_n | X_n = i]$ is the "mean holding time in state i ". Thus $P^{(1)}E = P^{(1)}\boldsymbol{e}\boldsymbol{e}^T = \boldsymbol{\mu}\boldsymbol{e}^T$.

Let $\lambda_1 = \boldsymbol{\pi}^T \boldsymbol{\mu}$ be the "asymptotic mean increment", since for a M.R.P., $M_d = \lambda_1 (\Pi_d)^{-1}$ where $\Pi = \boldsymbol{e}\boldsymbol{\pi}^T$ implying $m_{jj} = \lambda_1 / \pi_j$.

Thus for a M.R.P. $\{X_n, T_n\}$, on the state space S_N, the MFPT matrix satisfies $(I - P)M = \boldsymbol{\mu}^{(N)}\boldsymbol{e}^{(N)T} - PM_d$, where we identify the state space S_N through $\boldsymbol{\mu} \equiv \boldsymbol{\mu}^{(N)}$.

For a M.C. $\{X_n\}$ with N states and transition matrix P, $\boldsymbol{\mu}^{(N)T} = (1,1,...,1) = \boldsymbol{e}^{(N)T}$ and $P^{(1)}E = E$, its mean first passage time matrix (MFPT) M satisfies $(I - P)M = E - PM_d$ where $E = [1] = \boldsymbol{e}^{(N)}\boldsymbol{e}^{(N)T}$ and $M_d = [\delta_{ij} m_{ij}] = diag(\pi_1, \pi_2, ..., \pi_N)$.

For an M.R.P. with state space S_n partition $M = M_n$ as $M_n = \begin{bmatrix} M_{n-1} & \boldsymbol{m}_{n-1}^{(n)(c)} \\ \boldsymbol{m}_{n-1}^{(n)(r)T} & m_{nn} \end{bmatrix}$,

where $M_{n-1} = \begin{bmatrix} m_{ij} \end{bmatrix}$, $(1 \le i \le n-1, 1 \le j \le n-1)$, $\boldsymbol{m}_{n-1}^{(n)(r)T} = (m_{n1}, m_{n2}, ..., m_{n,n-1})$ and $\boldsymbol{m}_{n-1}^{(n)(c)T} = (m_{1n}, m_{2n}, ..., m_{n-1,n})$.

We also partition $\boldsymbol{\mu}^{(n)T} = (\mu_1^{(n)}, ..., \mu_{n-1}^{(n)}, \mu_n^{(n)}) = (\boldsymbol{\mu}_{n-1}^{(n)T}, \mu_n^{(n)})$ where $\boldsymbol{\mu}_{n-1}^{(n)T} = (\mu_1^{(n)}, ..., \mu_{n-1}^{(n)})$.

In general, if $(I_n - P_n)M_n = \boldsymbol{\mu}^{(n)}\boldsymbol{e}^{(n)T} - P_n(M_n)_d$, then (with the details omitted)

$(I_{n-1} - P_{n-1})M_{n-1} = \boldsymbol{\mu}^{(n-1)}\boldsymbol{e}^{(n-1)T} - P_{n-1}(M_{n-1})_d$ where $\boldsymbol{\mu}^{(n-1)T} = \boldsymbol{\mu}_{n-1}^{(n)T} + \dfrac{\mu_n^{(n)} \boldsymbol{p}_{n-1}^{(n)(c)T}}{\boldsymbol{p}_{n-1}^{(n)(r)T} \boldsymbol{e}^{(n-1)}}$.

Observe that $\boldsymbol{\mu}^{(n)T}$ is a $1 \times n$ vector, and $\boldsymbol{\mu}^{(n-1)T}$ is a $1 \times (n-1)$ vector, with

$\mu_i^{(n-1)} = \mu_i^{(n)} + \dfrac{\mu_n^{(n)} p_{i,n}^{(n)}}{S(n)}$, $(1 \le i \le n-1)$ where $S(n) = \boldsymbol{p}_{n-1}^{(n)(r)T} \boldsymbol{e}^{(n-1)} = \sum_{j=1}^{n-1} p_{nj}^{(n)} = 1 - p_{nn}^{(n)}$.

Further, it can be shown that, $\boldsymbol{m}_{N-1}^{(N)(r)T} = \dfrac{\left\{ \boldsymbol{p}_{N-1}^{(N)(r)T} \left(M_{N-1} - (M_{N-1})_d \right) + \mu_N^{(N)} \boldsymbol{e}^{(N-1)T} \right\}}{\boldsymbol{p}_{N-1}^{(N)(r)T} \boldsymbol{e}^{(N-1)}}$,

implying $m_{Nj} = \dfrac{\left\{ \sum_{k=1,k\ne j}^{N-1} p_{Nk}^{(N)} m_{kj} + \mu_N^{(N)} \right\}}{S(N)}$ for $1 \le j \le N-1$,

leading to expressions for m_{Nj} in terms of $m_{1j}, .., m_{kj}, .., m_{N-1,j}$ $(k \ne j)$, i.e. expressions for m_{Nj} in terms of the remaining elements of the $j-$ th column of M.

It is more difficult to find $\boldsymbol{m}_{N-1}^{(N)(c)}$, i.e. the m_{iN} for $1 \le i \le N-1$. It can be shown

$(I_{n-1} - Q_{n-1}^{(n)})\boldsymbol{m}_{n-1}^{(n)(c)} = \boldsymbol{\mu}_{n-1}^{(n)}$, where $Q_{N-1}^{(N)} = \begin{bmatrix} p_{ij}^{(N)} \end{bmatrix}$ for $1 \le i \le N-1, 1 \le j \le N-1$, an $(n-1) \times (n-1)$ matrix derived from P_N, requires further a step by step reduction procedure by eliminating $m_{N-1,N}$ from $\boldsymbol{m}_{N-1}^{(N)(c)T}$ replacing it in the expressions for the elements $m_{1N}, m_{2N}, ..., m_{N-2,N}$.

We need to express $(N-1) \times (N-1)$ matrix $Q_{N-1}^{(N)}$ in block form. From this we can show that

$m_{N-1,N} = \dfrac{\left\{ \boldsymbol{p}_{N-2}^{(N-1)(N)(r)T} \boldsymbol{m}_{N-2}^{(N)(c)} + \mu_{N-1}^{(N)} \right\}}{1 - p_{N-1,N-1}^{(N-1)}} = \dfrac{\left\{ \sum_{k=1}^{N-2} p_{N-1,k}^{(N-1)} m_{kN} + \mu_{N-1}^{(N)} \right\}}{R(N)}$,

where $R(N) = 1 - p_{N-1,N-1}^{(N-1)} = \sum_{j=1,j\ne N-1}^{N} p_{N-1,j}^{(N)}$ (i.e. obtained from P_N).

Note that this requires further additional calculations of the $R(N)$ and retaining further probabilities from the sub-matrix $Q_{N-1}^{(N)}$ derived from $P^{(N)}$ which complicates the procedure.

We can carry out the state reduction process, as in the GTH procedure, reducing the state space by 1 at successive steps retaining the same MFPT's for the reduced state space.

i.e. $M_{n-1} = \begin{bmatrix} m_{ij} \end{bmatrix}$, for $1 \le i \le n-1, 1 \le j \le n-1$. Note however we do not have a M.C. at successive

state reductions even if we start with a M.C. because the calculation involves the mean holding times in the states being modified. i.e. in effect we are using the M.R.P. variant to preserve the mean first passage times for the reduced state space. This is the basis of the procedure developed by Kohlas ([11]).

We show however that using just the computations of the GTH algorithm we can find expressions for m_{i1} for $i = 1, 2, \ldots, N$, i.e. the first column of M.

If we are given $M_{n-1} = \left[m_{ij} \right]$, $(1 \leq i \leq n-1, 1 \leq j \leq n-1)$ we wish to find $m_{n-1}^{(n)(c)}, m_{n-1}^{(n)(r)T}$ and m_{nn}.

First, for a M.C. $m_{nn} = 1/\pi_n$ so that we could use the GTH algorithm from the calculation of the stationary probabilities.

For the elements of $\boldsymbol{m}_{n-1}^{(n)(c)}$, where $\boldsymbol{m}_{n-1}^{(n)(c)T} = (m_{n1},\ldots,m_{n,n-1})$, we can show that

$$m_{nj} = \frac{\left\{ \sum_{k=1,k\neq j}^{n-1} p_{nk}^{(n)} m_{kj} + \mu_n^{(n)} \right\}}{S(n)} \text{ for } 1 \leq j \leq n-1.$$

For n = 2 : $(I_2 - P_2)M_2 = \boldsymbol{\mu}^{(2)}e^{(2)T} - P_2(M_2)_d$, implying

$$\begin{bmatrix} 1-p_{11}^{(2)} & -p_{12}^{(2)} \\ -p_{21}^{(2)} & 1-p_{22}^{(2)} \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} \mu_1^{(2)} & \mu_1^{(2)} \\ \mu_2^{(2)} & \mu_2^{(2)} \end{bmatrix} - \begin{bmatrix} p_{11}^{(2)}m_{11} & p_{12}^{(2)}m_{22} \\ p_{21}^{(2)}m_{11} & p_{22}^{(2)}m_{22} \end{bmatrix},$$

leading to $M_2 = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{p_{21}^{(2)}\mu_1^{(2)} + p_{12}^{(2)}\mu_2^{(2)}}{p_{21}^{(2)}} & \dfrac{\mu_1^{(2)}}{p_{12}^{(2)}} \\ \dfrac{\mu_2^{(2)}}{p_{21}^{(2)}} & \dfrac{p_{21}^{(2)}\mu_1^{(2)} + p_{12}^{(2)}\mu_2^{(2)}}{p_{12}^{(2)}} \end{bmatrix}.$

This leads to a general procedure for finding all the elements of M.

Step 1: Start with P_N and concentrate on finding only the expressions for m_{i1} for $i = 1, 2, \ldots, N$. i.e. if $P_N = \left[p_{ij}^{(N)} \right]$ carry out the extended GTH algorithm computing in addition the $\boldsymbol{\mu}^{(N)}$.

For $n = N$, N-1, $\ldots, 3$, let $p_{ij}^{(n-1)} = p_{ij}^{(n)} + \dfrac{p_{in}^{(n)} p_{nj}^{(n)}}{S(n)}$, $1 \leq i \leq n-1$, $1 \leq j \leq n-1$, with

$\mu_i^{(n-1)} = \mu_i^{(n)} + \dfrac{\mu_n^{(n)} p_{i,n}^{(n)}}{S(n)}$, $(1 \leq i \leq n\text{-}1)$, and $S(n) = \sum_{j=1}^{n-1} p_{nj}^{(n)}$.

For M.C. variant, start at state N with $(\mu_1^{(N)},\mu_2^{(N)}\ldots,\mu_N^{(N)}) = (1,1,\ldots,1)$.

Let $m_{11} = \mu_1^{(2)} + \dfrac{p_{12}^{(2)}\mu_2^{(2)}}{p_{21}^{(2)}}$, $m_{21} = \dfrac{\mu_2^{(2)}}{S(2)}$, with $m_{n1} = \dfrac{\sum_{k=2}^{n-1} p_{nk}^{(n)}m_{k1} + \mu_n^{(n)}}{S(n)}$, $n = 3,\ldots,N$.

This provides the entries of the first column of $M = \left[m_{ij} \right]$, i.e. $\boldsymbol{m}_N^{(1)(N)}$, where

$M = (\boldsymbol{m}_N^{(1)(N)},\boldsymbol{m}_N^{(2)(N)}\ldots,\boldsymbol{m}_N^{(N)(N)})$ with $\boldsymbol{m}_N^{(1)(N)T} = (m_{11},m_{21},\ldots,m_{N1})$.

Note that we only have to retain the $p_{in}^{(n)}(1 \leq i \leq n-1)$ and $p_{nj}^{(n)}(1 \leq j \leq n-1)$, i.e. the n-th row and n-th column of P_N, as in the GTH algorithm.

Step 2: Now reorder the rows of $P_N \equiv P_N^{(1)}$ by moving the first column after the N-th column, followed by moving the first row to the last row.

$$P_N^{(1)} = \begin{bmatrix} p_{11} & p_{12} & p_{1,N-1} & p_{1,N} \\ p_{21} & p_{22} & p_{2,N-1} & p_{2N} \\ & & & \\ p_{N-1,1} & p_{N-1,2} & p_{N-1,N-1} & p_{N-1,N} \\ p_{N1} & p_{N2} & p_{N,N-1} & p_{NN} \end{bmatrix} \rightarrow \begin{bmatrix} p_{22} & p_{2,N-1} & p_{2N} & p_{21} \\ & & & \\ p_{N-1,2} & p_{N-1,N-1} & p_{N-1,N} & p_{N-1,1} \\ p_{N2} & p_{N,N-1} & p_{NN} & p_{N,1} \\ p_{12} & p_{1,N-1} & p_{1,N} & p_{11} \end{bmatrix} \equiv P_N^{(2)}$$

Step 3: Carry out the algorithm, as in Step 1, with $P_N = P_N^{(2)}$ to obtain the vector of MFPTs which we label as $\overline{\boldsymbol{m}}_N^{(2)(N)}$ where $\overline{\boldsymbol{m}}_N^{(2)(N)T} = (m_{22}, m_{32}, ..., m_{N2}, m_{12})$.

Step 4: Reorder $P_N^{(2)}$ as in step 2 to obtain $P_N^{(3)}$ and repeat Step 3 to obtain $\overline{\boldsymbol{m}}^{(3)(N)}$ where

$\overline{\boldsymbol{m}}^{(3)(N)T} = (m_{33}, m_{43}, ..., m_{N3}, m_{13}, m_{23})$.

Step k: Repeat with $P_N^{(k)}$ to obtain $\overline{\boldsymbol{m}}^{(k)(N)}$ with $\overline{\boldsymbol{m}}^{(k)(N)T} = (m_{kk}, m_{k+1,k}, ..., m_{N,k}, m_{1,k}, ..., m_{k-1,k})$

Step N: Finish with $P_N^{(N)}$ and $\overline{\boldsymbol{m}}^{(N)(N)}$ where $\overline{\boldsymbol{m}}^{(N)(N)T} = (m_{NN}, m_{1,N}, m_{2,N}, ..., m_{N-1,N})$.

Step $N+1$: Let $\overline{M} = (\boldsymbol{m}_N^{(1)(N)}, \overline{\boldsymbol{m}}_N^{(2)(N)}...., \overline{\boldsymbol{m}}_N^{(N)(N)})$.

Step $N+2$: Reorder $\overline{M}$ to obtain $M = (\boldsymbol{m}_N^{(1)(N)}, \boldsymbol{m}_N^{(2)(N)}...., \boldsymbol{m}_N^{(N)(N)})$

This last step can be performed in MatLab as follows:

```
    end
      for col=1:m
        for row= 1:m
        P_new1(mod(row+m-2,m)+1,col)=P(row,col);
        end
      end
      for col=1:m
        for row= 1:m
        P_new2(row,mod(col+m-2,m)+1)=P_new1(row,col);
        end
      end
    P=P_new2;
     PP=P;
    end
    for col=1:m
     for row=1:m
        M1(mod(row+col-2,m)+1,col)=M(row,col);
     end
```

While this procedure requires N repetitions of the extended GTH, one would have to carry N auxiliary calculations using the elements of the sub-matrices of the P_n, $Q_{n-1}^{(n)}$, for each n and retain more elements of P_n rather than just the last row and column The extended GTH, (EGTH), algorithm outlined above retains calculation accuracy with no subtractions being involved.

9 The Test Problems
The following test problems were introduced by Harrod & Plemmons ([3]) and have considered by others in different contexts.

TP1: The original transition matrix was not irreducible and was replaced (Heyman ([4]), Heyman & Reeves ([5])) by

$$\begin{bmatrix} .1 & .6 & 0 & .3 & 0 & 0 \\ .5 & .5 & 0 & 0 & 0 & 0 \\ .5 & .2 & 0 & 0 & .3 & 0 \\ 0 & .7 & 0 & .2 & 0 & .1 \\ .1 & 0 & .8 & 0 & 0 & .1 \\ .4 & 0 & .4 & 0 & 0 & .2 \end{bmatrix}.$$

TP2: (Also Benzi ([1]))

$$\begin{bmatrix} .85 & 0 & .149 & .0009 & 0 & .00005 & 0 & .00005 \\ .1 & .65 & .249 & 0 & .00009 & .00005 & 0 & .00005 \\ .1 & .8 & .09996 & .0003 & 0 & 0 & .0001 & 0 \\ 0 & .0004 & 0 & .7 & .2995 & 0 & .0001 & 0 \\ .0005 & 0 & .0004 & .399 & .6 & .0001 & 0 & 0 \\ 0 & .00005 & 0 & 0 & .00005 & .6 & .2499 & .15 \\ .00003 & 0 & .00003 & .00004 & 0 & .1 & .8 & .0999 \\ 0 & .00005 & 0 & 0 & .00005 & .1999 & .25 & .55 \end{bmatrix}.$$

TP3:

$$\begin{bmatrix} 0.999999 & 1.0\,E-07 & 2.0\,E-07 & 3.0\,E-07 & 4.0\,E-07 \\ 0.4 & 0.3 & 0 & 0 & 0.3 \\ 5.0\,E-07 & 0 & 0.999999 & 0 & 5.0\,E-07 \\ 5.0\,E-07 & 0 & 0 & 0.999999 & 5.0\,E-07 \\ 2.0\,E-07 & 3.0\,E-07 & 1.0\,E-07 & 4.0\,E-07 & 0.999999 \end{bmatrix}.$$

TP4 and variants:
TP41 : $\varepsilon = 1.0E-01$; **TP42** : $\varepsilon = 1.0E-03$; **TP43** : $\varepsilon = 1.0E-05$; **TP44** : $\varepsilon = 1.0E-07$

$$\begin{bmatrix} .1-\varepsilon & .3 & .1 & .2 & .3 & \varepsilon & 0 & 0 & 0 & 0 \\ .2 & .1 & .1 & .2 & .4 & 0 & 0 & 0 & 0 & 0 \\ .1 & .2 & .2 & .4 & .1 & 0 & 0 & 0 & 0 & 0 \\ .4 & .2 & .1 & .2 & .1 & 0 & 0 & 0 & 0 & 0 \\ .6 & .3 & 0 & 0 & .1 & 0 & 0 & 0 & 0 & 0 \\ \varepsilon & 0 & 0 & 0 & 0 & .1-\varepsilon & .2 & .2 & .4 & .1 \\ 0 & 0 & 0 & 0 & 0 & .2 & .2 & .1 & .3 & .2 \\ 0 & 0 & 0 & 0 & 0 & .1 & .5 & 0 & .2 & .2 \\ 0 & 0 & 0 & 0 & 0 & .5 & .2 & .1 & 0 & .2 \\ 0 & 0 & 0 & 0 & 0 & .1 & .2 & .2 & .3 & .2 \end{bmatrix}$$

10 Computation comparisons for MFPT

We present comparisons for the test problems for the MFPT matrix M, using the 4 algorithms:
Standard, Simple, Perturbations and Extended GTH, each under double precision, and compute
the MAXIMUM RESIDUAL ERRORS as $\max\limits_{1\le i\le m, 1\le j\le m}\left| m_{ij} - \sum_{k\ne j} p_{ik} m_{kj} - 1 \right|$.

Table 1: Maximum Residual Errors for the Four Algorithms for MFPTs

	M_Standard	M_Simple	M_Perturbation	M_EGTH
TP1	5.6843E-14	5.6843E-14	1.1369E-13	1.1369E-13
TP2	1.8190E-12	1.8190E-12	3.6380E-12	3.6380E-12
TP3	1.7027E+00	1.7594E+00	1.5073E+00	1.6188E+00
TP41	1.4211E-14	2.1313E-14	1.4211E-14	1.4211E-14
TP42	1.0370E-12	1.8190E-12	9.0950E-13	1.8190E-12
TP43	1.1642E-10	1.7462E-10	1.7462E-10	1.1642E-10
Tp44	1.4901E-08	1.4901E-08	1.4901E-08	7.4506E-09

References

[1]   Benzi, M. (2004). A direct projection method for Markov chains. *Linear Algebra Appl.*, **386**, 27-49.

[2]   Grassman, W.K., Taksar, M.I., and Heyman, D.P. (1985). Regenerative analysis and steady state distributions for Markov chains. *Operations Research*, **33**, 1107-1116.

[3]   Harrod, W.J. and Plemmons, R.J. (1984). Comparison of some direct methods for computing stationary distributions of Markov chains. *SIAM J. Sci. Stat. Comput*, **5**, 453-469.

[4]   Heyman, D.P. (1987). Further comparisons of direct methods for computing stationary distributions of Markov chains. *SIAM J. Alg. Disc. Meth.*, **8**, 226-232.

[5]   Heyman, D.P. and Reeves, A. (1989). Numerical solutions of linear equations arising in Markov chain models. *ORSA J. Computing*, **1**, 52-60.

[6]   Hunter, J.J. (1982). Generalized inverses and their application to applied probability problems. *Linear Algebra Appl.*, **45**, 157-198.

[7]   Hunter, J.J. (2007). Simple procedures for finding mean first passage times in Markov chains. *Asia- Pacific Journal of Operational Research*, **24**, 813-829.

[8]   Hunter, J. J. (2014a). Generalized inverses of Markovian kernels in terms of properties of the Markov chain. *Linear Algebra Appl.* **447**, 38-55.

[9]   Hunter, J.J. (2014b). The computation the group inverse and related properties of Markov chains via perturbations. Preprint, School of Computer and Mathematical Sciences, Auckland University of Technology, New Zealand. (Presented at the 23rd International Workshop on Matrices and Statistics, University of Ljubljana, Slovenia, June 9, 2014.)

 [10] Kemeny, J.G. (1981). Generalizations of a fundamental matrix. *Linear Algebra Appl.*, **38**, 193-206.

[11] Kohlas, J. (1986). Numerical computation of mean first passage times and absorption probabilities in Markov and semi-Markov models. *Zeit fur Oper Res*, **30,** 197-207.

[12] Paige, C. C., Styan, G. P. H., Wachter, P. G. (1975). Computation of the stationary distribution of a Markov chain. *J. Statist. Comput. Simulation*, **4**, 173-186.

[13] Sheskin, T.J. (1985). A Markov partitioning algorithm for computing steady state probabilities. Operations Research, **33**, 228-235.

Leontief's Input-Output Representation of Multiple and Simple Regression Coefficients

Eric Iksoon[*] Im[1] & Tam Bang Vu[1]

[1]College of Business and Economics, University of Hawaii at Hilo

Abstract   First, we show analytically that the nexus between Least Squares (LS) estimators of multiple and simple regression coefficients are exactly the same as between partial and total derivatives of the general function of a given number of independent variables. Second, LS estimators of multiple and simple regression coefficient vectors correspond, respectively, to net output and gross output vectors in Leontief's input/output matrix equation where the input coefficient matrix is composed of LS estimators of the coefficients in simple regression equations that can be formed from the given number of regressors. Third, we show that each element of LS estimator of multiple regression coefficient vector is represented by Cramer's rule. Finally, we show that LS estimator of each multiple regression coefficient as represented by Cramer's rule can be transformed to its counterpart in Frisch and Waugh Theorem. Thus, the fragments in regression theory are related in a unifying manner, involving the works of two Nobel laureates in Economics.

Keywords   Simple and multiple regression coefficients; Least Squares estimator; Cramer's rule; Frisch/Waugh Theorem; Input-output model; net output, gross output; input coefficient matrix; Ragnar Frisch; Wassily Leontief.

1. Introduction

Consider the differentiable general function:

$$Y = f(X_1, X_2, ..., X_K) . \tag{1}$$

Then, follows from (1)

$$\frac{dY}{dX_k} = f_1 \frac{dX_1}{dX_k} + f_2 \frac{dX_2}{dX_k} + ... + f_{K-1} \frac{dK_{K-1}}{dX_k} + f_K \frac{dX_K}{dX_k} = f_k + \sum_{\ell=1}^{K} \frac{dX_\ell}{dX_k} f_\ell \theta_{k\ell} \qquad (k, \ell = 1, 2, ..., K) \tag{2}$$

where $f_k \equiv \partial Y / \partial X_k$ and $\theta_{k\ell} \equiv 1 - \delta_{k\ell}$ with $\delta_{k\ell}$ denoting Kronecker delta (, i.e., $\delta_{k\ell} = 1$ if $k = \ell$, and $\delta_{k\ell} = 0$ if $k \neq \ell$) so that $\theta_{k\ell}$ is Kronecker delta in reverse: $\theta_{k\ell} = 0$ if $k = \ell$, and $\theta_{k\ell} = 1$ if $k \neq \ell$.

Writing Equation (2) for f_k,

$$f_k = \frac{dY}{dX_k} - \sum_{\ell=1}^{K} \frac{dX_\ell}{dX_k} f_\ell \theta_{k\ell} \quad (k, \ell = 1, 2, ..., K) . \tag{3}$$

Interpretation of Equation (3) is that in general partial effect of an independent variable X_k on dependent variable Y (f_k) is its total effect on Y (dY / dX_k) net of its indirect effects on Y through $K-1$ other independent variables accounted for by the second term of the right-hand side. However,

---

[*] The authors retain the copyright.

this entails a question of whether the general relationships between partial and total derivatives represented by (3) are identical with those between multiple and simple LS slope estimators, though it is not necessarily an *a priori* expectation in view of the fact that Equation (3) has nothing to do with the least squares. This question is fully addressed in Section 2.

2. Derivation of LS Estimators in Implicit Form without Calculus and Matrix

Except for the bivariate regression case as in Ehrenberg (1983) as an example, derivation of LS coefficients with no calculus and no matrix algebra is generally perceived as taking a tedious algebraic workout as the number of regressors increases. In this section, however, we show that it takes basically the same derivation process as in the bivariate case. To facilitate the derivation, we write the LS estimated multiple linear regression model as

$$Y_i = a + X_{1_i}b_1 + ... + X_{K_i}b_K + e_i = a + \sum_{k=1}^{K} X_{ki}b_k + e_i \quad (i = 1, 2..., n; k = 1, ..., K) \tag{4}$$

where Y_i is the regressand, X_{ki} *k-th* regressor, a and b_k LS estimators of intercept and slope coefficients, and e_i LS estimated residuals.

Let the bar on top of each variable signify the sample mean of the variable. Then,

$$Y_i = \overline{Y}_i + y_i; \quad X_{k_i} = \overline{X}_{k_i} + x_{k_i}, \tag{5}$$

and the mean of Equation (4) can be written as

$$y_i = a + \sum_{k=1}^{K} x_{ki}b_k + \overline{e}_i \quad (i = 1, 2..., n; k = 1, ..., K). \tag{6}$$

In light of (4), (5), and (6), the error sum of squares can be written as

$$
\begin{aligned}
ESS \quad &= \sum_{i=1}^{n} e_i^2 \\
&= \sum_{i=1}^{n} \left[Y_i - \left(a + \sum_{k=1}^{K} X_{ki}b_k \right) \right]^2 \\
&= \sum_{i=1}^{n} \left[\left(\overline{Y}_i + y_i \right) - \left(a + \sum_{k=1}^{K} \left(\overline{X}_k + x_{ki} \right) b_k \right) \right]^2 \\
&= \sum_{i=1}^{n} \left[\left(y_i - \sum_{k=1}^{K} x_{ki}b_k \right) + \left(\overline{Y}_i - a - \sum_{k=1}^{K} \overline{X}_k b_k \right) \right]^2 \\
&= \sum_{i=1}^{n} \left[\left(y_i - \sum_{k=1}^{K} x_{ki}b_k \right) + \overline{e}_i \right]^2 \\
&= \sum_{i=1}^{n} \left[\left(y_i - \sum_{k=1}^{K} x_{ki}b_k \right)^2 + 2 \left(y_i - \sum_{k=1}^{K} x_{ki}b_k \right) \overline{e}_i + \overline{e}_i^2 \right] \\
&= \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K} x_{ki}b_K \right]^2 + 2\overline{e}_i \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K} x_{ki}b_k \right] + \sum_{i=1}^{n} \overline{e}_i^2
\end{aligned}
$$

$$= \sum_{i=1}^{n} \left[\tilde{y}_i - \sum_{k=1}^{K} \tilde{x}_{ki} b_k \right]^2 + n\bar{e}_i^2 \equiv \lambda + n\bar{e}_i^2. \tag{7}$$

The first term of the last equality in (7) can be re-expressed, with $\theta_{k\ell}$ as Kronecker delta in reverse as defined earlier in (2), as

$$\lambda = \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K} x_{ki} b_k \right]^2$$

$$= \sum_{i=1}^{n} \left[y_i - \left(\sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} + x_{ki} b_k \right) \right]^2 \qquad (k, \ell = 1, 2, ..., K)$$

$$= \sum_{i=1}^{n} \left[\left(y_i - \sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} \right) - x_{ki} b_k \right]^2$$

$$= \sum_{i=1}^{n} \left[\left(y_i - \sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} \right)^2 - 2x_{k_i} \left(y_i - \sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} \right) b_k + x_{ki}^2 b_k^2 \right]$$

$$= \sum_{i=1}^{n} \left[y_i - \sum_{\ell=1}^{K-1} x_{\ell i} b_\ell \theta_{k\ell} \right]^2 + \left[-2\sum_{i=1}^{n} x_{ki} \left(y_i - \sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} \right) \right] b_k + \left[\sum_{i=1}^{n} x_{ki}^2 \right] b_k^2$$

$$\equiv \pi_k + \tau_k b_k + \psi_k b_k^2 \qquad (\tau_k, \ \psi_k > 0)$$

$$= \pi_k - \frac{\tau_k^2}{4\psi_k} + \psi_k \left(b_k + \frac{\tau_k}{2\psi_k} \right)^2. \tag{8}$$

Feeding (8) back into (7) yields

$$ESS = \pi_k - \frac{\tau_k^2}{4\psi_k} + \psi_k \left(b_k + \frac{\tau_k}{2\psi_k} \right)^2 + n\bar{e}^2 \geq \pi_k - \frac{\tau_k^2}{4\psi_k} \quad (\psi_k > 0; \quad n > K+1 \geq 2). \tag{9}$$

Now, it is all obvious from (9) that the minimum ESS is reached only when

$$b_k = -\frac{\tau_k}{2\psi_k}; \quad \bar{e} = 0. \tag{10}$$

Substituting τ_k and ψ_k implicitly defined by the fifth equality in (8) into b_k in (10), we obtain

$$b_k = -\frac{\tau_k}{2\psi_k}$$

$$= \frac{2\sum_{i=1}^{n} x_{ki} \left(y_i - \sum_{\ell=1}^{K} x_{\ell i} b_\ell \theta_{k\ell} \right)}{2\sum_{i=1}^{n} x_{ki}^2}$$

$$
= \frac{\sum_{i=1}^{n} x_{ki} y_i}{\sum_{i=1}^{n} x_{ki}^2} - \frac{\sum_{\ell=1}^{K} \sum_{i=1}^{n} x_{ki} x_{\ell i} b_\ell \theta_{k\ell}}{\sum_{i=1}^{n} x_{ki}^2}
$$

$$
= \frac{\sum_{i=1}^{n} x_{ki} y_i}{\sum_{i=1}^{n} x_{ki}^2} - \sum_{\ell=1}^{K} \left(\frac{\sum_{i=1}^{n} x_{ki} x_{\ell i}}{\sum_{i=1}^{n} x_{ki}^2} \right) b_\ell \theta_{k\ell} . \tag{11}
$$

Defining simplified notations as

$$
g_{ky} = \frac{\sum_{i=1}^{n} x_{ki} y_i}{\sum_{i=1}^{n} x_{ki}^2} \left(= \frac{x_k' y}{x_k' x_k} \right); \quad g_{k\ell} = \frac{\sum_{i=1}^{n} x_{ki} x_{\ell i}}{\sum_{i=1}^{n} x_{ki}^2} \left(= \frac{x_k' x_\ell}{x_k' x_k} \right), \tag{12}
$$

we can rewrite Equation (11) alternatively in exactly the same form as in (3):

$$
b_k \equiv g_{ky} - \sum_{\ell=1}^{K} g_{k\ell} b_\ell \theta_{k\ell} \qquad (k, \ell = 1, 2, ..., K) .^{\dagger} \tag{13}
$$

where g_{ky} is the simple LS slope coefficient that results when Y is regressed solely on X_k, whereas $g_{k\ell}$ defines simple LS slope coefficient when X_ℓ is regressed solely on X_k, in both cases with the intercept retained. Equation (13) is the first ever compact scalar representation of multiple regression coefficients in its unique form. In its nutshell, (13) represents multiple regression coefficients as a linear system of K simultaneous equations in K simple regression coefficients.

It is clear from comparison of (3) and (13) that multiple LS slope estimators b_k are related to simple LS slope estimators g_{ky} in exactly the same way as partial derivatives f_k are to total derivatives dY / dX_k, with correspondences of b_k and g_{ky}, respectively, to f_k and dY / dX_k. Accordingly, we can conclude that b_k and g_{ky}, respectively, measure exactly what f_k (partial impact) and dY / dX_k (total impact) measure in the general function of K independent variables. This correspondence between b_k and f_k is hardly a surprising result, given our general understanding that the multiple regression slope coefficients measure the partial impacts of regressors on the regressand, or technically in light of Frisch-Waugh Theorem in Greene (p.27, 2012) or Frisch and Waugh (1933). There are three points noteworthy, however. First, the correspondence of (3) and (13) is a bit of a surprise in the sense that b_k and g_{ky} are the solutions of the conditions for minimizing the sum of squared residuals, whereas f_k and dY / dX_k are merely the partial and total derivatives of a general function. Second, the correspondence of (3) and (13) holds regardless of functional form of (3). Third, the correspondence also hold regardless of whether (3) is a misspecification. Furthermore, as will be shown in the following section, vectors of LS estimated multiple and simple LS slope

---

† When $K = 1$ (simple regression), $k = \ell = 1$ and $\theta_{k\ell} = \theta_{11} = 0$ so that (14) is reduced to

$b_k = g_{ky} = \sum_{i=1}^{n} x_{ki} y_i \Big/ \sum_{i=1}^{n} x_{ki}^2 = x_k' y / x_k' x_k$, The LS estimator of simple regression slope

coefficient when Y_i is regressed on X_{k_i}.

coefficients collectively form a Leontief framework, namely, input-output model in which total output and net output of a *k-th* industry correspond to b_k and g_{ky}, respectively.

3. Input-Output Nexus of Simple and Multiple LS Estimators

The primary objective of the restricted derivation approach in Section 2 was to obtain the multiple LS slope coefficients in *implicit* forms consistent with partial derivative functions in general. With that mission accomplished, we now admit matrix algebra for its operational virtues in demonstrating that the LS slope coefficients exclusively define a matrix equation which is in essence Leontief input-output model as follows.

Stacking b_k $(k = 1, 2, ..., K)$ in (8), we arrive at a system of n linear equations in (13) in matrix form as

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{K-1} \\ b_K \end{pmatrix} = \begin{pmatrix} g_{1y} \\ g_{2y} \\ \vdots \\ g_{(K-1)y} \\ g_{Ky} \end{pmatrix} - \begin{bmatrix} 0 & g_{12} & \cdots & g_{1(K-1)} & g_{1K} \\ g_{21} & 0 & \cdots & g_{2(K-1)} & g_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{(K-1)} & g_{(K-1)2} & \cdots & 0 & g_{(K-1)K} \\ g_{K1} & g_{K2} & \cdots & g_{K(K-1)} & 0 \end{bmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{K-1} \\ b_K \end{pmatrix}, \tag{14}$$

or compactly as

$$b = g_{\bullet y} - Hb, \tag{15}$$

which is a Leontief static open input-output model in disguise. Writing (15) for $g_{\bullet y}$ and b, respectively,

$$g_{\bullet y} = b + H b = (I + H)b \equiv Gb; \tag{16.1}$$

$$b = (I + H)^{-1} g_{\bullet y} = G^{-1} g_{\bullet y} \equiv T g_{\bullet y}. ^{\ddagger} \tag{16.2}$$

Substituting (16.2) into the second term of the first equality of (16.1),

$$g_{\bullet y} = b + H T g_{\bullet y} \equiv b + A g_{\bullet y} \tag{17}$$

where

$$G(K \times K) = I + H = \begin{bmatrix} 1 & g_{12} & \cdots & g_{1(K-1)} & g_{1K} \\ g_{21} & 1 & \cdots & g_{2(K-1)} & g_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ g_{(K-1)1} & g_{(K-1)2} & \cdots & 1 & g_{(K-1)K} \\ g_{K1} & g_{K2} & \cdots & g_{K(K-1)} & 1 \end{bmatrix}; \quad g_{\bullet y}(K \times 1) = \begin{pmatrix} g_{1y} \\ g_{2y} \\ \vdots \\ g_{(K-1)y} \\ g_{Ky} \end{pmatrix}, \tag{18}^{\S}$$

---

‡ $T = G^{-1}$ corresponds to the so-called technology matrix in Leontief's input-output model.

Matrix equation (17) is in exactly the same form as Leontief's input-output model where $b(K \times 1)$ and $g_{\bullet y}(K \times 1)$ are analogous to net industry and gross industry output vectors, respectively, and A is analogous to input coefficient matrix.

4. **Each Multiple Regression Coefficient in the Form of Cramer's Rule**

Furthermore, analytical solution of each component of b by Cramer's Rule (Chiang (1984), p. 109) follows as:

$$b_k = |G_k|/|G| \quad (k = 1, 2, ..., K) \tag{19}$$

where $G_k(K \times K)$ defines a square matrix obtained by replacing its *i-th* column of G with $g_{\bullet y}$ both as defined in (18). Both G_k and G are special square matrices, special in the sense that both consist exclusively of LS estimators of simple regression coefficient, hence may be referred to as a special Cramer's rule.

5. **Equivalence of Cramer's and Frisch/Waugh's Representations**

If b_k in the form of (19) is correct, then it must be equal to b_k in the form of Frisch/Waugh (1933) as stated in Greene (2012, Corollary 3.3.2, p.28). That is,

$$b_k = |G_k|/|G| = (x_k' \tilde{M}_k x_k)^{-1} x_k' \tilde{M}_k y . \tag{20}$$

For a short-cut proof of (20), define a diagonal matrix D :

$$D = \begin{bmatrix} x_1'x_1 & 0 & 0 & \dots & 0 \\ 0 & x_2'x_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & x_{K-1}'x_{K-1} & 0 \\ 0 & 0 & \cdots & 0 & x_K'x_K \end{bmatrix} . \tag{21}$$

Then, LS estimator of slope coefficient vector in (16.2) can be expressed alternatively as

$$b = G^{-1} g_{\bullet y}$$
$$= (DG)^{-1} D g_{\bullet y}$$

$$= \begin{bmatrix} x_1'x_1 & x_1'x_2 & \cdots & x_1'x_{K-1} & x_1'x_K \\ x_2'x_1 & x_2'x_2 & \cdots & x_2'x_{K-1} & x_2'x_K \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{K-1}'x_1 & x_{K-1}'x_2 & \cdots & x_{K-1}'x_{K-1} & x_{K-1}'x_K \\ x_K'x_1 & x_K'x_2 & \cdots & x_K'x_{K-1} & x_K'x_K \end{bmatrix}^{-1} \begin{pmatrix} x_1'y \\ x_2'y \\ \vdots \\ x_{K-1}'y \\ x_K'y \end{pmatrix} = [x'x]^{-1} x'y . \tag{22}$$

---

§ Note that G is not symmetric, i.e., $G \neq G'$ since $g_{k\ell} \neq g_{\ell k}$ in general: $g_{k\ell} = x_k'x_\ell / x_k'x_k \neq x_\ell'x_k / x_\ell'x_\ell = g_{\ell k}$.

Partitioning $x(N \times K)$ as $x = \begin{bmatrix} x_k & \tilde{x}_k \end{bmatrix}$ in which $x_k (K \times 1)$ denotes *k-th* column vector of x whereas $\tilde{x}_k (N \times (K-1))$ denotes x with x_k excluded, b in (22) can be re-expressed as

$$b = \begin{bmatrix} x'x \end{bmatrix}^{-1} x'y = \begin{bmatrix} E_{1k} x'x \end{bmatrix}^{-1} E_{1k} x'y = \begin{bmatrix} x'_k x \\ \tilde{x}'_k x \end{bmatrix}^{-1} \begin{pmatrix} x'_k y \\ \tilde{x}_k y \end{pmatrix} = \begin{bmatrix} x'_k x_k & x'_k \tilde{x}_k \\ \tilde{x}'_k x_k & \tilde{x}'_k \tilde{x}_k \end{bmatrix}^{-1} \begin{pmatrix} x'_k y \\ \tilde{x}_k y \end{pmatrix}, \qquad (23)$$

where E_{1k} $(k = 1, 2, ..., K)$ are nonsingular elementary permutation matrices that result when the *1*-st and *k*-th rows in the identity matrix of the same order are interchanged as discussed , for example, in Searle (1982, p.173).

Applying Cramer's rule to the last expression of (23), *k*-th element of b follows as

$$b_k = \begin{vmatrix} x'_k y & x'_k \tilde{x}_k \\ \tilde{x}'_k y & \tilde{x}'_k \tilde{x}_k \end{vmatrix} \Bigg/ \begin{vmatrix} x'_k x_k & x'_k \tilde{x}_k \\ \tilde{x}'_k x_k & \tilde{x}'_k \tilde{x}_k \end{vmatrix}$$

$$= \frac{x'_k y - x'_k \tilde{x}_k \left(\tilde{x}'_k \tilde{x}_k \right)^{-1} \tilde{x}'_k y}{x'_k x_k - x'_k \tilde{x}_k \left(\tilde{x}'_k \tilde{x}_k \right)^{-1} \tilde{x}'_k x_k}$$

$$= \frac{x'_k \left[I_n - \tilde{x}_k \left(\tilde{x}'_k \tilde{x}_k \right)^{-1} \tilde{x}'_k \right] y}{x'_k \left[I_n - \tilde{x}_k \left(\tilde{x}'_k \tilde{x}_k \right)^{-1} \tilde{x}'_k \right] x_k}$$

$$\equiv (x'_k \tilde{M}_k x_k)^{-1} x'_k \tilde{M}_k y, \qquad (k = 1, 2, 3, ..., K) \qquad (24)$$

in which the second equality is based on the general formula for determinants of partitioned matrices discussed, for example, in Searle (1982, p.258), and $\tilde{M}_k = \tilde{M}_k^2$ is an idempotent matrix with rank $n - K + 1$.

6. Summaries

Our discussions up to this point may be summarized in theorems without redundant proofs as follows:

<u>Theorem 1</u>: Let $Y = f(X_1, X_2, .., X_K)$. Then, K partial derivatives of Y can be represented by

$$f_k = \frac{dY}{dX_k} - \sum_{\ell=1}^{K} \frac{dX_\ell}{dX_k} f_\ell \theta_{k\ell}, \qquad (k, \ell = 1, 2, ..., K)$$

where f_k denotes partial derivatives of Y and dY/dX_k total derivatives of Y, and $\theta_{k\ell}$ the Kronecker delta in reverse.

<u>Theorem 2</u>: Let $Y_i = \alpha + X_{1_i} \beta_1 + ... + X_{K_i} \beta_K + \varepsilon_i$ $(i = 1, 2, ..., n; \ 1 \le K < K+1 < n < \infty)$ as the multiple linear regression model of Y_i, $Y_i = \alpha_k + X_{k_i} \gamma_{ky} + u_{k_i}$ $(k = 1, 2, ..., K; \ 2 < n < \infty)$ as the simple linear regression model of Y_i on X_{ℓ_i}, and $X_{k_i} = \alpha_{k\ell} + X_{\ell_i} \gamma_{k\ell} + u_{k\ell_i}$ $(k, \ell = 1, 2, ..., K)$ as the simple linear regression model of X_{k_i} on X_{ℓ_i}. Then,

$$b_k = g_{ky} - \sum_{\ell=1}^{K} g_{k\ell} b_\ell \theta_{k\ell} \qquad (k, \ell = 1, 2, ..., K),$$

where b_k denotes LS estimator of β_k, g_{ky} LS estimator of γ_{ky}, and $g_{k\ell}$ LS estimator of $\gamma_{k\ell}$, $\theta_{k\ell}$ denoting the Kronecker delta in reverse.

Theorem 3 (Leontief's Representation): Retain $b(K \times 1)$ and $g_{\bullet y}(K \times 1)$ as denoting vectors of LS multiple slope coefficients and LS simple slope coefficients, respectively. Then, their relationship is represented precisely in the form of Leontief's input-output model:

$$g_{\bullet y} = b + A g_{\bullet y}$$

where input matrix $A(K \times K)$ consists of all simple regression coefficients that can be formed between the regressors.

Theorem 4 (Cramer's Representation): LS estimator of each multiple regression slope coefficient can be represented by Cramer's rule:

$$b_k = |G_k| / |G| \qquad (k = 1, 2, ..., K)$$

where matrices G and G_k are as defined in (19).

Theorem 5 (Equivalence of Cramer's and Frisch/Waugh's Representations):

$$b_k = |G_k| / |G| = (x_k' \tilde{M}_k x_k)^{-1} x_k' \tilde{M}_k y \qquad (k = 1, 2, ..., K).$$

7. Conclusion

In this note, we have derived the analytical scalar expression for LS estimators of multiple slope coefficients in implicit form such that their relationships are in exactly the same form as the partial and total derivatives of a general function. Subsequently, we have shown that LS estimators of the multiple and simple LS slope coefficients bear exactly the same relationships as between net and gross outputs in Leontief input-output model. We also derived an alternative analytical expression for the LS estimators of multiple regression coefficients solely in terms of the LS estimators of the simple regression coefficients, with the proof that it is equivalent to the expression by Frisch and Waugh. In its nutshell, this note interconnects some fragments in regression theory, thereby shedding some additional light on the nature of the least squares estimators that have been exercised in empirical studies for more than two centuries.

References

Chiang, A. (1984). *Fundamental Methods of Mathematical Economics*. 3rd edition, McGraw-Hill, Inc.

Ehrenberg, A. S. C. (1983). Deriving the least squares regression equation. In: American Statistician, (38, 2, p. 162).

Frisch, J. and F. Waugh. (1933). Partial Time Regression as Compared with Individual Trends. In: Econometrica, (I, pp.387-401).

Greene, W. H. (2012). *Econometric Analysis*. 7th ed., New York, NY: Prentice Hall.

Leontief, W. W. (1951). *The Structure of American Economy, 1919-1939*. 2d ed., Oxford University Press, Fair Lawn, N.J.

Searle, R. S (1982). Matrix Algebra Useful for Statistics. John Wiley & Sons.

Strong Stability Bounds for Queues

ISSAADI Badredine[1,*]   ABBAS Karim[2,†]   Djamil AISSANI[3,‡]

[1]University of M'hamed Bougara of Boumerdes
[2]University of Bejaia
[2]University of Bejaia

Abstract   This paper investigates the M/M/s queuing model to predict an estimate for the proximity of the performance measures of queues with arrival processes that are slightly different from the Poisson. Specifically, we use the strong stability method to obtain perturbation bounds on the effect of perturbing the arrival process in the M/M/s queue. Therefore, we build an algorithm based on strong stability method to predict stationary performance measures of the GI/M/s queue. Some numerical examples are sketched out to illustrate the accuracy of the proposed method.

Keywords   Queues; Strong stability method; Perturbation bounds; Algorithm.

1 Introduction

For the basic theorems of the strong stability method are given in [1]. The main tool for our analysis is the weighted supremum norm, also called v-norm, denoted by $\|.\|_v$, where v is some vector with elements $v(k) \geq 1$ for all $k \in \mathbb{Z}_+$, and for any vector f with infinite dimension

$$\|f\|_v = \sup_{k \geq 0} \frac{|f(k)|}{v(k)}. \tag{1}$$

Let μ be a probability measure on $\mathbb{Z}_+$, then the v-norm of μ is defined as

$$\|\mu\|_v = \sum_{j \geq 0} v(j)|\mu_j|. \tag{2}$$

The v-norm is extended to stochastic kernels on $\mathbb{Z}_+$ in the following way: let P the matrix with infinite dimension then

$$\|P\|_v = \sup_{k \geq 0} \frac{\|P(k,.)\|_v}{v(k)} = \sup_{k \geq 0} \frac{1}{v(k)} \sum_{j \geq 0} v(j)|P_{kj}|. \tag{3}$$

Note that v-norm convergence to 0 implies elementwise convergence to 0.

We associate to each transition kernel P the linear mappings:

$$(\mu P)_k = \sum_{i \geq 0} \mu_i P_{ik}. \tag{4}$$

$$(Pf)(k) = \sum_{i \geq 0} f(i) P_{ki}. \tag{5}$$

The strong stability method [2, 1] considers the problem of the perturbation of general state space Markov chains using operator's theory and with respect to a general class of norms. The basic idea behind the concept of stability is that, for a strongly stable Markov chain, a small perturbation in the transition kernel can lead to only a small deviation of the stationary distribution.

Definition 1. A Markov chain X with transition kernel P and stationary distribution π is said to be strongly stable with respect to the norm $\|.\|_v$ if $\|P\|_v < \infty$ and every stochastic kernel Q in some neighborhood $\{Q : \|Q - P\|_v \leq \varepsilon\}$ admits a unique stationary distribution ν and

$$\|\nu - \pi\|_v \to 0 \quad as \quad \|Q - P\|_v \to 0. \tag{6}$$

In fact, as shown in [2], X is strongly stable if and only if, there exists a positive constant $c = c(P)$ such that

$$\|\nu - \pi\|_v \leq c\|Q - P\|_v. \tag{7}$$

In the sequel we use the following results.

Theorem 1. ([2]) *The Markov chain X with the transition kernel P and stationary distribution π is strongly stable with respect to the norm $\|.\|_v$, if and only if there exists a probability measure $\alpha = (\alpha_j)$ and a vector $h = (h_i)$ on $\mathbb{Z}_+$ such that $\pi h > 0$, $\alpha \mathbf{1} = 1$, αh is a positive scalar, and*

a. *The matrix $T = P - h\alpha$ is non-negative, where $h\alpha = (a_{ij})_{ij}$ such that $a_{ij} = h_i \alpha_j$ for $i, j \in \mathbb{Z}_+$.*
b. *There exists $\rho < 1$ such that $Tv(k) \leq \rho v(k)$ for $k \in \mathbb{Z}_+$.*
c. *$\|P\|_v < \infty$.*

Here $\mathbf{1}$ is the vector having all the components equal to 1.

Theorem 2. ([1]) *Let X be a strongly v-stable Markov chain that satisfies the conditions of Theorem 1. If ν is the probability invariant measure of a stochastic kernel Q, then for $\|\Delta\|_v < (1 - \rho)/c$, we have the estimate*

$$\|\nu - \pi\|_v \leq c\|\Delta\|_v\|\pi\|_v(1 - \rho - c\|\Delta\|_v)^{-1}, \tag{8}$$

where $\Delta = Q - P$, $c = 1 + \|\mathbf{1}\|_v\|\pi\|_v$ and $\|\pi\|_v \leq (\alpha v)(1 - \rho)^{-1}(\pi h)$.

2 Analysis of the Model

2.1 Model Description

We consider a $GI/M/s$ queueing system (s servers) with infinite capacity. Customers arrive at time points $t_0 = 0, t_1, t_2, \ldots$ where the interarrival times $Z_n = t_{n+1} - t_n$, $n = 1, 2, 3, \ldots$, are independent identically distributed random variables (i.i.d r.v.'s) having a non lattice c.d.f $H(\cdot)$ with mean γ^{-1}. The sevice times $S_1, S_2, \ldots$ are i.d.d r.v.'s having a common exponentiel d.f with a finite mean μ^{-1}. Let $\gamma/s\mu$ be the traffic intensity, assumed to be strictly less than one. Let $\widetilde{Q}(t)$ be the number of customers in the system at time t and define $\widetilde{Q}(t_n - 0) = \widetilde{Q}_n$, $n = 1, 2, \ldots$. Thus $\widetilde{Q}_n$ is the number in the system just before the nth arrival. Now consider the relationship between $\widetilde{Q}_n$ and $\widetilde{Q}_{n+1}$. We have

$$\widetilde{Q}_{n+1} = \begin{cases} \widetilde{Q}_n + 1 - \widetilde{X}_{n+1} & \text{if } \widetilde{Q}_n + 1 - \widetilde{X}_{n+1} > 0, \\ 0 & \text{if } \widetilde{Q}_n + 1 - \widetilde{X}_{n+1} \leq 0, \end{cases} \tag{9}$$

where $\widetilde{X}_{n+1}$ is the total number of potential customers who can be served by s servers during an interarrival period Z_n. Due to the exponentiel service time, the process $\{\widetilde{Q}_n,\ n = 0, 1, 2, \ldots\}$ is an homogeneous Markov chain. From (9), it is found that the evolution of the homogeneous Markov chain $(\widetilde{Q}_n)_{n \geq 1}$ is governed by the transition probability matrice $\widetilde{P} = (\widetilde{P}(i,j))_{i,j \geq 0}$ described by

$$\widetilde{P}(i,j) = 0 \quad (i + 1 - j < 0).$$

$$\widetilde{P}(i,j) = \int_0^\infty \frac{(s\mu t)^{i+1-j}}{(i+1-j)!} e^{-s\mu t} dH(t) \quad (i \geq s - 1,\ j \geq s,\ i + 1 - j \geq 0).$$

$$\widetilde{P}(i,j) = \int_0^\infty \binom{i+1}{i+1-j} e^{-j\mu t}(1 - e^{-\mu t})^{i+1-j} dH(t) \quad (i \leq s - 1,\ i + 1 - j \geq 0)$$

$$\widetilde{P}(i,j) = \int_0^\infty \int_{\tau=0}^t \binom{s}{s-j} e^{-j\mu(t-\tau)}(1 - e^{-\mu(t-\tau)})^{s-j} \frac{(s\mu\tau)^{i-s}}{(i-s)!} e^{-s\mu\tau} s\mu \, d\tau \, dH(t)$$
$$(i \geq s,\ j < s,\ i + 1 - j \geq 0).$$

Consider also an system $M/M/s$, which has the same distribution of service times, where the interarrival times are independent identically distributed random variables and vary according to an exponential distribution $E_\lambda(\cdot)$ with a finite mean λ^{-1}. Further, the embedded Markov chain $(Q_n)_{n \geq 1}$, representing the number of customers in the $M/M/s$ queueing system. Denote by $P = (P(i,j))_{i,j \geq 0}$ the transition operators of the Markov chains $(Q_n)_{n \geq 1}$.

2.2 υ-Strong Stability Conditions

the main work in strong stability method is finding β such that $\|T\|_\upsilon < 1$, where T is a stochastic kernel. For that, we choose the function $\upsilon(k) = \beta^k$, $\beta > 1$, $h_i = \mathbf{I}_{i=0}$ and $\alpha_j = P_{0j}$ (see Theorem 1).

Theorem 3. *Suppose that in the $M/M/s$ queueing system the following geometric ergodicity condition, $\lambda/s\mu < 1$, holds. Then for all $\beta \in \mathbb{R}$ such that, $1 < \beta < \beta_0$, the embedded Markov chain $(Q_n)_{n \geq 1}$ is υ-strongly stable for the test function $\upsilon(k) = \beta^k$.*

Proof. We have $\pi h = \pi_0 > 0$, $\alpha \mathbf{1} = 1$ and $\alpha h = \alpha_0 = P_{00} > 0$.

$$T_{ij} = P_{ij} - h_i \alpha_j = \begin{cases} 0, & \text{if } i = 0, \\ P_{ij}, & \text{if } i \geq 1. \end{cases} \tag{10}$$

Hence, the kernel T is nonnegative.
According to Equation (5), we have:

$$T\upsilon(i) = \sum_{j \geq 0} \beta^j T_{ij}. \tag{11}$$

(a) If $i = 0$, then

$$T\upsilon(0) = \sum_{j \geq 0} \beta^j T_{0j} = 0. \tag{12}$$

(b) If $1 \leq i \leq s - 2$, then

$$T\upsilon(i) = \sum_{j=0}^{i+1} \beta^j P_{ij} = \beta^{i+1} \int_0^\infty \sum_{j=0}^{i+1} \binom{i+1}{j} \left(\frac{1 - e^{-\mu t}}{\beta}\right)^{i+1-j} (e^{-\mu t})^j \, dE_\lambda(t)$$

$$\leq \beta^i \int_0^\infty \frac{1}{\beta} \left(1 + (\beta - 1)e^{-\mu t}\right)^2 dE_\lambda(t)$$

We pose, $\rho_1 = \int_0^\infty \frac{1}{\beta}\left(1+(\beta-1)e^{-\mu t}\right)^2 dE_\lambda(t) = \int_0^\infty f(t)dE_\lambda(t)$

(c) If $i = s-1$, then

$$
\begin{aligned}
Tv(s-1) &= \sum_{j=0}^{s}\beta^j P_{(s-1)j} = \sum_{j=0}^{s-1}\beta^j P_{(s-1)j} + \beta^s P_{(s-1)s} \\
&= \beta^s \int_0^\infty \left(\frac{1-e^{-\mu t}}{\beta}+e^{-\mu t}\right)^s dE_\lambda(t) \\
&\leq \beta^{s-1}\int_0^\infty \frac{1}{\beta}\left(1+(\beta-1)e^{-\mu t}\right)^2 dE_\lambda(t) = \beta^{s-1}\rho_1
\end{aligned}
$$

(d) If $i \geq s$, then

$$
Tv(i) = \sum_{j=0}^{i+1}\beta^j P_{ij} = \sum_{j=0}^{s-1}\beta^j P_{ij} + \sum_{j=s}^{i+1}\beta^j P_{ij}
$$

$$
\begin{aligned}
\sum_{j=0}^{s-1}\beta^j P_{ij} &= \sum_{j=0}^{s-1}\beta^j \int_0^\infty \int_{\tau=0}^{t}\binom{s}{s-j}e^{-j\mu(t-\tau)}(1-e^{-\mu(t-\tau)})^{s-j}\frac{(s\mu\tau)^{i-s}}{(i-s)!}e^{-s\mu\tau}s\mu d\tau dE_\lambda(t) \\
&\leq \beta^{s-1}\int_0^\infty \left[e^{-s\mu t}\sum_{n=i+1-s}^{\infty}\frac{(s\mu t)^n}{n!}+(\beta-1)e^{-s\mu t}(\frac{s}{s-1})^{i+1-s}\sum_{n=i+1-s}^{\infty}\frac{((s-1)\mu t)^n}{n!}\right. \\
&\quad \left. -\beta e^{-s\mu t}\frac{(s\mu t)^{i+1-s}}{(i+1-s)!}\right]dE_\lambda(t)
\end{aligned}
$$

And, $\displaystyle\sum_{j=s}^{i+1}\beta^j P_{ij} = \beta^{i+1}\int_0^\infty e^{-s\mu t}\sum_{n=0}^{i+1-s}\frac{(s\mu t/\beta)^n}{n!}dE_\lambda(t)$

Therefore,

$$
\begin{aligned}
Tv(i) &= \sum_{j=0}^{s-1}\beta^j P_{ij} + \sum_{j=s}^{i+1}\beta^j P_{ij} \\
&\leq \beta^i \int_0^\infty \left(\frac{1}{\beta}-\frac{1}{\beta}e^{-s\mu t}+(\frac{\beta-1}{\beta})(\frac{s}{s-1})(e^{-\mu t}-e^{-s\mu t})+\beta e^{-s\mu t}\right)dE_\lambda(t)
\end{aligned}
$$

We pose,

$$
\rho_2 = \int_0^\infty \left(\frac{1}{\beta}-\frac{1}{\beta}e^{-s\mu t}+(\frac{\beta-1}{\beta})(\frac{s}{s-1})(e^{-\mu t}-e^{-s\mu t})+\beta e^{-s\mu t}\right)dE_\lambda(t) = \int_0^\infty g(t)dE_\lambda(t)
$$

We have, with assumption that $s \geq 2$, $\dfrac{s}{s-1} = 1 + \dfrac{1}{s-1} \leq 2$.

$$
\begin{aligned}
\text{Then, } g(t) &= \frac{1}{\beta}-\frac{1}{\beta}e^{-s\mu t}+(\frac{\beta-1}{\beta})(\frac{s}{s-1})(e^{-\mu t}-e^{-s\mu t})+\beta e^{-s\mu t} & (13) \\
&\leq \frac{1}{\beta}\left(1+(\beta-1)e^{-\mu t}\right)^2 = f(t) & (14)
\end{aligned}
$$

this shows that $\rho_2 = \int_0^\infty g(t) dE_\lambda(t) \leq \int_0^\infty f(t) dE_\lambda(t) = \rho_1$.

It suffices to take, $\rho = \max(\rho_1, \rho_2) = \int_0^\infty \frac{1}{\beta} \left(1 + (\beta - 1)e^{-\mu t}\right)^2 dE_\lambda(t)$ which is smaller then 1 for all $\beta > 1$. Now, we have $E_\lambda(t) = 1 - e^{-\lambda t}$, then

$$\rho = \frac{1}{\beta} + \frac{2\lambda(\beta - 1)}{\beta(\lambda + \mu)} + \frac{\lambda(\beta - 1)^2}{\beta(2\mu + \lambda)} \tag{15}$$

And, with assumption that $\beta > 1$, We have $\rho < 1 \Rightarrow \beta < \dfrac{2\mu^2}{\lambda(\lambda + \mu)}$.

We pose, $\beta_0 = \dfrac{2\mu^2}{\lambda(\lambda + \mu)}$. Then, for all β such that $1 < \beta < \beta_0$, we have $\rho < 1$.

Now, we verify that $\|P\|_v < \infty$. We have

$$T = P - h\alpha \Rightarrow P = T + h\alpha \Rightarrow \|P\|_v \leq \|T\|_v + \|h\|_v\|\alpha\|_v, \tag{16}$$

or, according to equation (3),

$$\|T\|_v = \sup_{i \geq 0} \frac{1}{v(i)} \sum_{j \geq 0} v(j)|T_{ij}| \leq \sup_{i \geq 0} \frac{1}{v(i)} \rho v(i) \leq \rho < 1. \tag{17}$$

According to Equations (1) and (2), we have: $\|h\|_v = \sup_{i \geq 0} \dfrac{1}{v(i)}|h_i| = 1$,

And, $\|\alpha\|_v = \sum_{j \geq 0} v(j)|\alpha_j| = \sum_{j \geq 0} \beta^j P_{0j} = P_{00} + \beta P_{01} < \beta(P_{00} + P_{01}) \leq \beta < \infty.$

Then, $\|P\|_v < \infty$.

The Markov chain $(Q_n)_{n \geq 1}$ being strongly stable then, the $\|\pi - \widetilde{\pi}\|_v$ can be bounded in terms of $\|P - \widetilde{P}\|_v$.

2.3 Bound on Perturbation

To be able to estimate numerically the margin between the stationary distributions of the Markov chains $(\widetilde{Q}_n)_{n \geq 1}$ and $(Q_n)_{n \geq 1}$ we estimate the norm of the deviation of the transition kernel.

Lemma 4. *Let $\widetilde{P}$ (respectively P) be the transition kernel of the Markov chain $(\widetilde{Q}_n)_{n \geq 1}$ (respectively of the Markov chain $(Q_n)_{n \geq 1}$). Then, for all β such that $1 < \beta < \beta_0$, we have:*

$$\|P - \widetilde{P}\|_v \leq \int_0^\infty \left(1 + (\beta - 1)e^{-\mu t}\right)|H - E_\lambda|(dt) \tag{18}$$

Proof. From Equation (3), we have

$$\|P - \widetilde{P}\|_v = \sup_{i \geq 0} \frac{\|P(i, .) - \widetilde{P}(i, .)\|_v}{v(i)} = \sup_{i \geq 0} \frac{1}{v(i)} \sum_{j=0}^{i+1} v(j)|P_{ij} - \widetilde{P}_{ij}|. \tag{19}$$

(a) For $i \leq s - 2$, we have,

$$
\sum_{j=0}^{i+1} \upsilon(j) |P_{ij} - \widetilde{P}_{ij}| \quad = \quad \sum_{j=0}^{i+1} \beta^j |P_{ij} - \widetilde{P}_{ij}|
$$

$$
\leq \quad \beta^{i+1} \sum_{j=0}^{i+1} \int_0^\infty \binom{i+1}{i+1-j} \left(\frac{1 - e^{-\mu t}}{\beta} \right)^{i+1-j} (e^{-\mu t})^j |H - E_\lambda|(dt)
$$

$$
\leq \quad \beta^i \int_0^\infty \left(1 + (\beta - 1) e^{-\mu t} \right) |H - E_\lambda|(dt)
$$

Then, $\quad \|P - \widetilde{P}\|_\upsilon \leq \int_0^\infty \left(1 + (\beta - 1) e^{-\mu t} \right) |H - E_\lambda|(dt).$ (20)

We pose, $\quad \Delta_1 = \int_0^\infty \left(1 + (\beta - 1) e^{-\mu t} \right) |H - E_\lambda|(dt) = \int_0^\infty \zeta(t) |H - E_\lambda|(dt)$

(b) For $i = s - 1$, we have,

$$
\sum_{j=0}^s \upsilon(j) |P_{(s-1)j} - \widetilde{P}_{(s-1)j}|
$$

$$
= \quad \sum_{j=0}^{s-1} \beta^j |P_{(s-1)j} - \widetilde{P}_{(s-1)j}| + \beta^s |P_{(s-1)s} - \widetilde{P}_{(s-1)s}|
$$

$$
\leq \quad \beta^s \int_0^\infty \left(\frac{1 - e^{-\mu t}}{\beta} + e^{-\mu t} \right)^s |H - E_\lambda|(dt) \quad \leq \quad \beta^{s-1} \int_0^\infty \left(1 + (\beta - 1) e^{-\mu t} \right) |H - E_\lambda|(dt)
$$

Then,

$$
\|P - \widetilde{P}\|_\upsilon \leq \int_0^\infty \left(1 + (\beta - 1) e^{-\mu t} \right) |H - E_\lambda|(dt) = \Delta_1. \tag{21}
$$

(c) For $i \geq s$, we have,

$$
\sum_{j=0}^{i+1} \upsilon(j) |P_{ij} - \widetilde{P}_{ij}| \quad = \quad \sum_{j=0}^{i+1} \beta^j |P_{ij} - \widetilde{P}_{ij}| \quad = \quad \sum_{j=0}^{s-1} \beta^j |P_{ij} - \widetilde{P}_{ij}| + \sum_{j=s}^{i+1} \beta^j |P_{ij} - \widetilde{P}_{ij}|
$$

we have,

$$
\sum_{j=0}^{s-1} \beta^j |P_{ij} - \widetilde{P}_{ij}|
$$

$$
\leq \quad \beta^s \int_0^\infty \int_0^t \left[\sum_{j=0}^{s-1} \binom{s}{j} \left(\frac{1 - e^{-\mu(t-\tau)}}{\beta} \right)^{s-j} \left(e^{-\mu(t-\tau)} \right)^j \right] \frac{(s\mu\tau)^{i-s}}{(i-s)!} e^{-s\mu\tau} s\mu d\tau |H - E_\lambda|(dt)
$$

$$
\leq \quad \beta^{s-1} \int_0^\infty \left[e^{-s\mu t} \sum_{n=i+1-s}^\infty \frac{(s\mu t)^n}{n!} + (\beta - 1) e^{-s\mu t} \left(\frac{s}{s-1} \right)^{i+1-s} \sum_{n=i+1-s}^\infty \frac{((s-1)\mu t)^n}{n!} \right.
$$

$$
\left. - \beta e^{-s\mu t} \frac{(s\mu t)^{i+1-s}}{(i+1-s)!} \right] |H - E_\lambda|(dt)
$$

And,

$$\sum_{j=s}^{i+1} \beta^j |P_{ij} - \widetilde{P}_{ij}| \leq \beta^{i+1} \int_0^\infty e^{-s\mu t} \sum_{j=s}^{i+1} \frac{(s\mu t/\beta)^{i+1-j}}{(i+1-j)!} |H - E_\lambda|(dt)$$

$$= \beta^{i+1} \int_0^\infty e^{-s\mu t} \sum_{n=0}^{i+1-s} \frac{(s\mu t/\beta)^n}{n!} |H - E_\lambda|(dt)$$

therefore,

$$\sum_{j=0}^{i+1} \upsilon(j) |P_{ij} - \widetilde{P}_{ij}| \leq \beta^i \int_0^\infty \left(\frac{1}{\beta} - \frac{1}{\beta} e^{-s\mu t} + \left(\frac{\beta-1}{\beta}\right)\left(\frac{s}{s-1}\right)(e^{-\mu t} - e^{-s\mu t}) + \beta e^{-s\mu t} \right) |H - E_\lambda|(dt)$$

Then,

$$\|P - \widetilde{P}\|_\upsilon \leq \int_0^\infty \left(\frac{1}{\beta} - \frac{1}{\beta} e^{-s\mu t} + \left(\frac{\beta-1}{\beta}\right)\left(\frac{s}{s-1}\right)(e^{-\mu t} - e^{-s\mu t}) + \beta e^{-s\mu t} \right) |H - E_\lambda|(dt) \quad (22)$$

We pose,

$$\Delta_2 = \int_0^\infty \left(\frac{1}{\beta} - \frac{1}{\beta} e^{-s\mu t} + \left(\frac{\beta-1}{\beta}\right)\left(\frac{s}{s-1}\right)(e^{-\mu t} - e^{-s\mu t}) + \beta e^{-s\mu t} \right) dE_\lambda(t) = \int_0^\infty g(t) |H - E_\lambda|(dt)$$

According to Equations (13), we have: $g(t) \leq \zeta(t)$.
This shows that $\Delta_2 \leq \Delta_1$. Finally, $\|P - \widetilde{P}\|_\upsilon \leq \Delta = \max(\Delta_1, \Delta_2) = \Delta_1$.

Lemma 5. *Let π be the stationary distribution of the embedded Markov chain $(Q_n)_{n \geq 1}$. Then, for all $1 < \beta < \beta_0$, we have:*

$$\|\pi\|_\upsilon = \pi_0 \left(\sum_{k=0}^{s-1} \frac{(s\varrho\beta)^k}{k!} + \frac{(\lambda\beta/\mu)^s}{s!(1 - \varrho\beta)} \right) = c_0 \quad (23)$$

where,

$$\varrho = \frac{\lambda}{s\mu} < 1 \ \ and \ \ \pi_0 = \left[\sum_{k=0}^{s-1} \frac{(s\varrho)^k}{k!} + \left(\frac{(s\varrho)^s}{s!}\right)\left(\frac{1}{1 - \varrho}\right) \right]^{-1}$$

Theorem 6. *Let π (respectively $\widetilde{\pi}$) be the stationary distribution of the embedded Markov chain $(Q_n)_{n \geq 1}$ (respectively of the embedded Markov chain $(\widetilde{Q}_n)_{n \geq 1}$). Then, for all $1 < \beta < \beta_0$, we have:*

$$\|\pi - \widetilde{\pi}\|_\upsilon \leq c_0 \, c \, \Delta (1 - \rho - c \, \Delta)^{-1}$$

where c_0 is given in (23) , $c = 1 + \|\pi\|_\upsilon$.

Corollary 7. *Under the conditions put forward in Theorem 6, it holds for any f such that $\|f\|_\upsilon < \infty$ that*

$$|\pi f - \widetilde{\pi} f| \leq \|f\|_\upsilon \times \boldsymbol{SSB}(\beta) = h(\beta).$$

Note that the bound in Corollary 7 has β as a free parameter. This give the opportunity to minimize the right hand side of the inequality in Corollary 7 with respect to β. For given ϱ, this leads to the following optimization problem.

$$\min_{\beta \in \mathbf{B}} h(\beta)$$

$$\text{s.t.} \quad \Delta(\beta) < \frac{1 - \rho(\beta)}{c(\beta)}.$$

By inserting $\varepsilon > 0$ small, all inequalities can be made strict and in the above optimization problem can be solved using any standard technique.

3 Numerical Example

In this section we will apply our bound put forward in Theorem 6.

3.1 Approximation Algorithm

In this subsection we elaborate an algorithm which allows us to get the domain of the approximation and to determine the error on the stationary distribution due to the approximation.

INITIALISATION: Definition of the inputs ;

The arrival mean rate λ;

Number of the servers $s \geq 2$;

The service mean rate μ;

BEGIN

STEP 1

Verification of the stability;

IF $\lambda/s\mu \geq 1$ **THEN**

the system is unstable;

Exit;

ELSE

Calculate $\beta_0 = \frac{2\mu^2}{\lambda(\lambda+\mu)}$;

END

END

STEP 2

Determine $[\beta_{min}, \beta_{max}] \subset]1, \beta_0[$ such that $\Delta < \dfrac{1-\rho}{c}$;

END

STEP 3

Determine $\beta_{opt} \in [\beta_{min}, \beta_{max}]$ which give the minimal error of the approximation $||\pi - \widetilde{\pi}||_v$;

With;

$||\pi - \widetilde{\pi}||_v \leq c_0\, c\, \Delta(1 - \rho - c\,\Delta)^{-1}$;

END

With;

$$\rho = \frac{1}{\beta} + \frac{2\lambda(\beta-1)}{\beta(\lambda+\mu)} + \frac{\lambda(\beta-1)^2}{\beta(2\mu+\lambda)};$$

$$\Delta = \int_0^\infty \left(1 + (\beta-1)e^{-\mu t}\right)|H - E_\lambda|(dt);$$

$$c = 1 + c_0;$$

$$c_0 = ||\pi||_v = \pi_0\left(\sum_{k=0}^{s-1}\frac{(s\varrho\beta)^k}{k!} + \frac{(\lambda\beta/\mu)^s}{s!(1-\varrho\beta)}\right);$$

$$\pi_0 = \left[\sum_{k=0}^{s-1}\frac{(s\varrho)^k}{k!} + \left(\frac{(s\varrho)^s}{s!}\right)\left(\frac{1}{1-\varrho}\right)\right]^{-1};$$

END.

Algorithm 1: Error of $||\pi - \widetilde{\pi}||_v$

3.2 Numerical Validation

The primary objective of this subsection is to compare our expected approximation error against results obtained from simulations. For this, we implement the algorithm 1

and simulator on a concrete case. Indeed, we apply the algorithm 1 to determine the made error (on stationary distribution) due to the approximation (when the approximation is possible) as well as the norm from which the error is obtained. This norm will be introduced into the simulator to simulate an error (on stationary distribution) with respect to the same norm.

For the simulation of the error, we used the discrete events approach and elaborated the program in the Matlab environment according to the following steps :

1. Simulate the stationary distribution $\widetilde{\pi} = (\widetilde{\pi}_i,\ i \geq 0)$;
2. Simulate the stationary distribution $\pi = (\pi_i,\ i \geq 0)$;
3. Calculate $\displaystyle\sum_{i \geq 0} \beta^i |\pi_i - \widetilde{\pi}_i|$.

We present numerical examples of M/M/s queues with perturbed arrival processes. In each case, $||\pi - \widetilde{\pi}||_\upsilon$ is compared with the error given by the simulator.

To illustrate the application of Corollary 7 to a particular performance function, we take $f(s) = s$ the identical mapping. In words, we are interested in the effect of perturbing the arrival processes on the mean queue length. It is worth noting that in this case

$$||f||_\upsilon = \frac{1}{\ln(\beta)} \beta^{-\frac{1}{\ln(\beta)}}.$$

Example 1: In the first example, the independent and identically distributed interarrival times of the M/M/s queue are perturbed by randomly inserting a few exponentially distributed intervals. The perturbed queue can be modelled by a GI/M/s queue having hyperexponentially distributed inter-arrival times with density

$$h(x) = p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x},\ \ x \geq 0, \tag{24}$$

Thus, a proportion p of inter-arrival times is exponentially distributed with mean λ_1^{-1} and a proportion $(1-p)$ is exponentially distributed with mean λ_2^{-1}. Setting the mean of the inter-arrival times with distribution (24), $p\lambda_1^{-1} + (1-p)\lambda_2^{-1}$, equal to the fixed mean, λ^{-1}. Note that as λ_1 approaches λ, p approaches 1 and the hyperexponential distribution approaches an exponential distribution with rate λ.

Hence, we may conclude that it's possible to determine approximately the characteristics of hyperexponential/M/s system using the results of the M/M/s system.

Let us choose, for example: $\lambda = \lambda_1 = 1.5$, $\lambda_2 = 1.3$, $\mu = 3$, $s = 2$ and $p = 0.95$. The approximation domain has been determined:

$$\beta \in [\beta_{\min},\ \beta_{\max}] = [1.0340,\ 2.4429]$$

So, we can give an idea about the error due to the approximation on the stationary distribution from its curve in function of β (Figure 1).
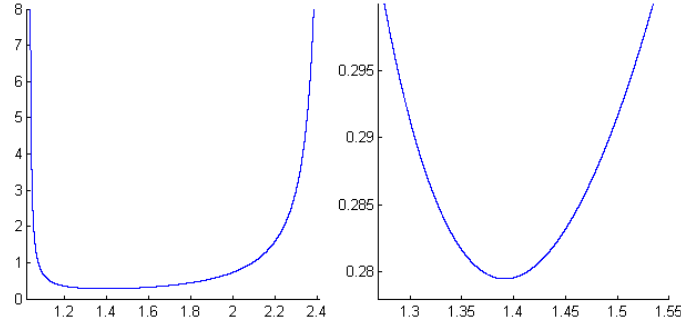
FIG. 1 − Error in function of β.

We obtain the smallest error, for $\beta = \beta_{opt} = 1.3913$, and then the algorithm 1 gives the error due to the approximation on the stationary distribution of the number of customers such that

$$||\pi - \widetilde{\pi}||_v = \sum_{i \geq 0} (1.3913)^i |\pi(i) - \widetilde{\pi}(i)| \leq \text{SSB}(1.3913) = 0.2795$$

similarly, we have

p	$\beta_{\min}$	$\beta_{\max}$	β_{opt}	$||\pi - \widetilde{\pi}||_v$	Simulated error	p	$\beta_{\min}$	$\beta_{\max}$	β_{opt}	$||\pi - \widetilde{\pi}||_v$	Simulated error
0.90	1.13	18.51	2.99	0.3053	0.0198	0.90	1.34	9.80	2.83	1.1332	0.0675
0.91	1.12	19.32	2.98	0.2674	0.0178	0.91	1.30	10.68	2.82	0.9243	0.0604
0.92	1.10	20.20	2.97	0.2315	0.0157	0.92	1.25	11.66	2.80	0.7511	0.0532
0.93	1.09	21.17	2.97	0.1974	0.0137	0.93	1.21	12.76	2.79	0.6053	0.0463
0.94	1.08	22.24	2.96	0.1649	0.0117	0.94	1.18	14.01	2.77	0.4808	0.0394
0.95	1.06	23.44	2.95	0.1341	0.0097	0.95	1.14	15.47	2.76	0.3733	0.0326
0.96	1.05	24.83	2.94	0.1047	0.0078	0.96	1.11	17.21	2.75	0.2795	0.0260
0.97	1.04	26.47	2.94	0.0767	0.0058	0.97	1.08	19.35	2.73	0.1970	0.0193
0.98	1.03	28.50	2.93	0.0500	0.0039	0.98	1.06	22.17	2.72	0.1239	0.0128
0.99	1.02	31.32	2.92	0.0244	0.0019	0.99	1.03	26.31	2.71	0.0586	0.0064

TAB. 1 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 0.2$, $\mu = 2$ and $s = 2$. **TAB.** 2 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 0.5$, $\mu = 2$ and $s = 2$.

p	$\beta_{\min}$	$\beta_{\max}$	β_{opt}	$||\pi - \widetilde{\pi}||_v$	Simulated error	p	$\beta_{\min}$	$\beta_{\max}$	β_{opt}	$||\pi - \widetilde{\pi}||_v$	Simulated error
0.90	1.55	6.08	2.69	2.8328	0.1234	0.90	1.89	3.71	2.55	12.2058	0.1909
0.91	1.45	6.89	2.67	2.0227	0.1098	0.91	1.66	4.54	2.53	5.1573	0.1694
0.92	1.38	7.80	2.65	1.4898	0.0965	0.92	1.51	5.38	2.51	2.9942	0.1485
0.93	1.31	8.82	2.63	1.1127	0.0836	0.93	1.40	6.31	2.49	1.9447	0.1282
0.94	1.25	10.01	2.61	0.8319	0.0709	0.94	1.32	7.39	2.47	1.3251	0.1084
0.95	1.20	11.43	2.60	0.6146	0.0586	0.95	1.25	8.69	2.45	0.9162	0.0891
0.96	1.15	13.16	2.58	0.4416	0.0464	0.96	1.18	10.31	2.44	0.6262	0.0706
0.97	1.11	15.37	2.57	0.3005	0.0346	0.97	1.13	12.45	2.42	0.4099	0.0523
0.98	1.07	18.39	2.55	0.1834	0.0228	0.98	1.08	15.46	2.41	0.2424	0.0346
0.99	1.04	23.08	2.54	0.0845	0.0113	0.99	1.04	20.40	2.39	0.1089	0.0171

TAB. 3 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 1$, $\mu = 2$ and $s = 2$. **TAB.** 4 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 2$, $\mu = 2$ and $s = 2$.

So, we can give an idea about the error due to the approximation on the stationary distribution from its curve in function of p (Figure 2). From these numerical results, it is easy to see that, the error decreases as the probability p increases ($p \rightarrow 1$). Besides, the

values of the both errors (algorithmic and numeric) tend to coincide in the neighborhood of lower bound ($p \to 1$). This can be explain by the way that it represents the frontier (critical value) of the stability domain.
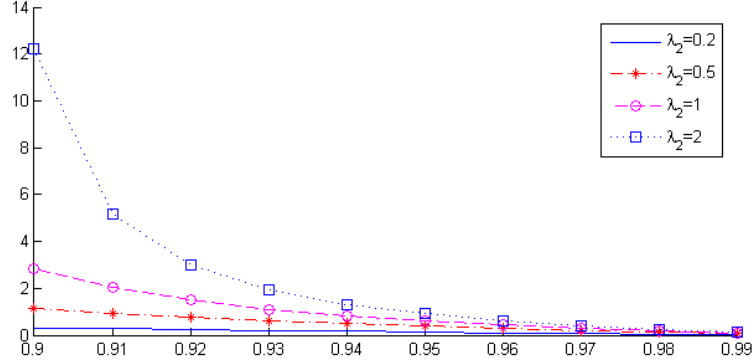
FIG. $2 - ||\pi - \widetilde{\pi}||_v$ in function of p. ($\lambda_1 = \lambda = 0.1$, $\mu = 2$ and $s = 2$)

For applying our bounds we compute the value for β_{opt} that minimizes $h(\beta)$. Then, we can compute the bounds put forward in Corollary 7 for various values for p. The numerical results are presented in the following tables

p	β_{opt}	$|\pi f - \widetilde{\pi} f|$	Simulated error	p	β_{opt}	$|\pi f - \widetilde{\pi} f|$	Simulated error
0.90	2.99	0.9128	0.0592	0.90	2.83	3.2070	0.1910
0.91	2.98	0.7969	0.0530	0.91	2.82	2.6065	0.1703
0.92	2.97	0.6876	0.0466	0.92	2.80	2.1031	0.1490
0.93	2.97	0.5863	0.0407	0.93	2.79	1.6888	0.1292
0.94	2.96	0.4881	0.0346	0.94	2.77	1.3318	0.1091
0.95	2.95	0.3956	0.0286	0.95	2.76	1.0303	0.0900
0.96	2.94	0.3078	0.0229	0.96	2.75	0.7686	0.0715
0.97	2.94	0.2255	0.0171	0.97	2.73	0.5378	0.0527
0.98	2.93	0.1465	0.0114	0.98	2.72	0.3370	0.0348
0.99	2.92	0.0712	0.0055	0.99	2.71	0.1588	0.0173

TAB. 5 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 0.2$, $\mu = 2$ and $s = 2$.   **TAB.** 6 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 0.5$, $\mu = 2$ and $s = 2$.

p	β_{opt}	$|\pi f - \widetilde{\pi} f|$	Simulated error	p	β_{opt}	$|\pi f - \widetilde{\pi} f|$	Simulated error
0.90	2.69	7.6202	0.3319	0.90	2.55	31.1248	0.4868
0.91	2.67	5.4006	0.2932	0.91	2.53	13.0480	0.4286
0.92	2.65	3.9480	0.2557	0.92	2.51	7.5154	0.3727
0.93	2.63	2.9264	0.2199	0.93	2.49	4.8423	0.3192
0.94	2.61	2.1713	0.1850	0.94	2.47	3.2730	0.2677
0.95	2.60	1.5980	0.1524	0.95	2.45	2.2447	0.2183
0.96	2.58	1.1393	0.1197	0.96	2.44	1.5279	0.1723
0.97	2.57	0.7723	0.0889	0.97	2.42	0.9920	0.1266
0.98	2.55	0.4677	0.0581	0.98	2.41	0.5842	0.0834
0.99	2.54	0.2146	0.0287	0.99	2.39	0.2603	0.0409

TAB. 7 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 1$, $\mu = 2$ and $s = 2$.   **TAB.** 8 − Evaluating at $\lambda_1 = \lambda = 0.1$, $\lambda_2 = 2$, $\mu = 2$ and $s = 2$.
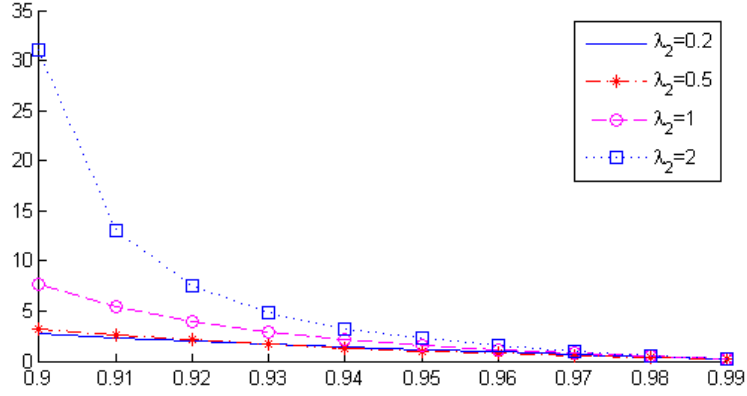
FIG. $3 - |\pi f - \widetilde{\pi} f|$ in function of p. ($\lambda_1 = \lambda = 0.1$, $\mu = 2$ and $s = 2$)

Example 2: In this exemple we examine the robustness of the M/M/s queueing model to predict accurately the operating characteristics of queues with inter-arrival times are perturbed by slightly ϵ. Such as,

$$\widetilde{\lambda} = \lambda + \epsilon$$

where, $\widetilde{\lambda}^{-1}$ is the mean of the inter-arrival times of the perturbed queue.

In a similar manner, the mapping $h(\beta)$ is minimized at β_{opt}. The numerical results are presented in Table 9, where the symbol $\times$ indicates that our bounds are not applicable.

ϵ	$\beta_{\min}$	$\beta_{\max}$	β_{opt}	$||\pi - \widetilde{\pi}||_v$		$|\pi f - \widetilde{\pi} f|$	
				Bound	Simulated error	Bound	Simulated error
0.01	1.04	4.04	1.60	0.1680	0.0072	0.2688	0.0115
0.02	1.07	3.68	1.61	0.3872	0.0145	0.6234	0.0233
0.03	1.11	3.37	1.63	0.6842	0.0221	1.1152	0.0360
0.04	1.15	3.10	1.65	1.1089	0.0298	1.8297	0.0492
0.05	1.20	2.85	1.67	1.7638	0.0377	2.9455	0.0630
0.06	1.27	2.62	1.69	2.9028	0.0459	4.9057	0.0776
0.07	1.35	2.38	1.71	5.3673	0.0542	9.1781	0.0927
0.08	1.47	2.11	1.73	14.6365	0.0628	25.3211	0.1086
$\geq$0.09	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

TAB. $9 -$ Evaluating at $\lambda = 1$, $\mu = 3$ and $s = 6$.

Figure 4 give the error due to the approximation in function of ϵ, From these numerical results, it is easy to see that, the error increases as the ϵ increases.
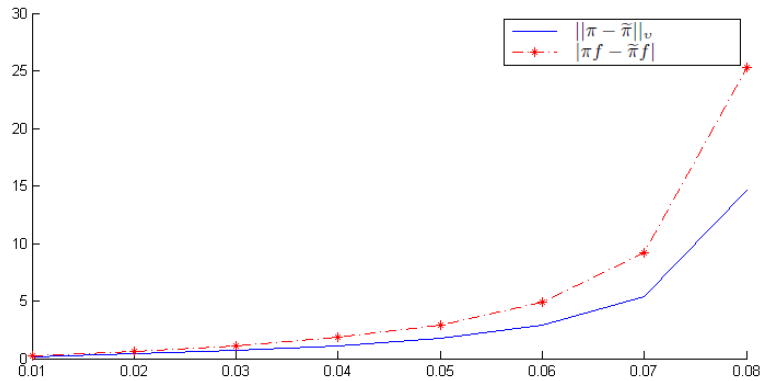
FIG. $4 -$ Bound in function of ϵ.

References

[1] N. V. Kartashov (1996), *Strong Stable Markov Chains*. VSP, Utrecht, The Netherlands.

[2] D. Aïssani and N. V. Kartashov, (1983). *Ergodicity and stability of Markov chains with respect to operator topologies in space of transition kernels*. Doklady Akad. Nauk AN Ukr. SSR, 11 (series A), 3-5.

An Extraordinary Property of The Arcsine Distribution and Its Applications

Jiajian Jiang[1]   He Ping[1,*]   Kaitai Fang[1]

[1]Division of Science and Technology, BNU-HKBU United International College,Zhuhai, 519085, China

Abstract   Finding a discrete approximation to a given continuous distribution is often requested in various fields, such as decision analysis and resampling procedure. A typical criterion for the accuracy of a discrete approximation is that it preserves as many moments of the original distribution as possible. In this paper we focus on the discrete approximation to the arcsine distribution. Representative points obtained by number-theoretic method (RPs-NTM) are used to construct the discrete approximation to the arcsine distribution. We find and prove a surprising property that the discrete approximation constructed by RPs-NTM have all the same moments as the arcsine distribution, if the number of points is larger than the order of moment. In addition, another type of representative points determined in the sense of minimizing mean squared error (RPs-MSE) is also considered as an approximation. The performance of the discrete approximations based on RPs-NTM and RPs-MSE are evaluated and compared. Results show that RPs-NTM is better than RPs-MSE as the approximation of arcsine distribution.

Keywords   Arcsine distribution, Representative points, Resampling

1   Introduction

Random variable X has the standard arcsine distribution (arcsine distribution for short) if X has probability density function given by

$$f(x) = \frac{1}{\pi \sqrt{x(1-x)}}, \ 0 \le x \le 1. \tag{1}$$

It's corresponding cumulative distribution function is equal to

$$F(x) = \frac{2}{\pi} arcsin(\sqrt{x}), \ 0 \le x \le 1. \tag{2}$$

Actually, the arcsine distribution is a special case of the beta distribution with left and right parameters $\alpha = \beta = \frac{1}{2}$. The arcsine distributions occur naturally in statistical communications theory and have played an important role in the study of Brownian motion and prime numbers. In Lee [17] and Middleton [20], the amplitude of a periodic signal in thermal noise and the limiting spectral density function of a highindex-angle modulated carrier are modeled with arcsine distribution. In addition, the arcsine distribution is useful in the study of the fluctuations of random walks (Feller [10]) and *Arcsine Law* in Brownian motion which was first proved by Lévy in [18] and [19]. Besides, the arcsine distribution has been widely used and accepted in physical and financial stochastic model, such as, economic default time and occupation times for the pricing of options are modeled by arcsine distribution (see e.g. [13, 21]).

---

*Corresponding author. Email: heping@uic.edu.hk

Jiajian Jiang, Ping He, Kai-Tai Fang

For many complex probabilistic problems involving continuous random variables, it is rare to find a closed-form solution. Consequently, it is often required that a given continuous probability distribution is approximated by a discrete distribution. Specifically, for a given continuous random variable X with pdf $f(x)$ and cdf $F(x)$, we construct a discrete random variable ξ with cdf $G(x)$ such that $G(x)$ preserves the properties of $F(x)$ as much as possible. There are many ways to construct such random variable ξ. We can simply choose a set of random sample $\{x_i, i = 1, \cdots, n\}$ and let $P(\xi = x_i) = 1/n, i = 1, \cdots, n$. Efron [5] proposed to use this ξ in the bootstrap. In general, denote a support set of ξ as $\{\xi_1, \xi_2, ..., \xi_n\}$ and the related probability mass distribution as $P(\xi = \xi_i) = p_i, i = 1, \ldots, n$. Several methods for constructing a discrete approximation are proposed, such as "Bracket-Median" method (Clemen [2]), "25-50-25" method (Keefer and Bodily [16]), "Number-Theoretic" method (Fang and Wang [8]) and "Mean Square Error" method (Cox [3] and Fang and He [7]). In the procedure of "Bracket-Median method", one chooses a number of points n and divides the continuous cumulative probability distribution $F(x)$ into n equally probable intervals. And then each interval is represented by the point corresponding to the median of the interval and each point is assigned probability $1/n$. The "25-50-25" method including three RPs (representative points) which assigns the weights 0.250, 0.500, and 0.250 to 10th(q_{10}), 50th (q_{50}), and 90th (q_{90}) percentiles of $F(x)$. Representative points generated by Number-Theoretic method are selected by using the percentiles $\xi_i = F^{-1}(q_i)$, $i = 1, \ldots, n$, where $F^{-1}(x)$ is the inverse function of $F(x)$ and $q_i = \frac{2i-1}{2n}$. For the "Mean Square Error" method, RPs are optimally selected in the sense of providing a minimum mean squared error approximation to the original continuous distribution. For a comprehensive review of a number of different discrete approximation methods, the reader can refer to Hammond and Brickel [14], Smith[22] and Fang et.al [9].

In this paper, we focus on the discrete approximation for the arcsine distribution and its application in simulation. We discover a surprising property of the arcsine distribution that the discrete distribution constructed by RPs-NTM has the same moments of all orders as the arcsine distribution if the number of points, n, is larger than the order of moment, that means, RPs-NTM provides the best approximation for the arcsine distribution in the sense of matching moments.

In addition, we also obtain a set of RPs-MSE which are optimal on the criterion of minimum mean squared error between discrete approximation and the arcsine distribution. The two different discretization methods naturally raise the question as to which approximation for the arcsine distribution is better in practice and simulation. We compare the two approximation from two aspects: L_2 distance between $F(x)$ and $F_\xi(x)$ and statistical inference by resampling. The results show that the discrete approximation constructed by RPs-NTM has a better performance than the approximation by RPs-MSE.

The next section presents an extraordinary property of the arcsince distribution. We give a brief introduction of representative points by MSE criterion (RPs-MSE) and propose an algorithm for generating RPs-MSE for the arcsine distribution in Section 3. Section 4 gives comparisons between RPs-NTM and RPs-MSE. Finally, the paper is concluded in section 4.

2 An extraordinary property

As we introduced in section 1, the arcsine distribution is a special case of beta distribution, $X \sim Beta(0.5, 0.5)$. Its pdf and cdf are given in (1) and (2), respectively. The arcsine distribution is symmetric at $x = \frac{1}{2}$. The density curve steeply rises towards infinity as x approaching to 0 and 1 and tends to be flatter around the center $x = \frac{1}{2}$ (see Figure 1).
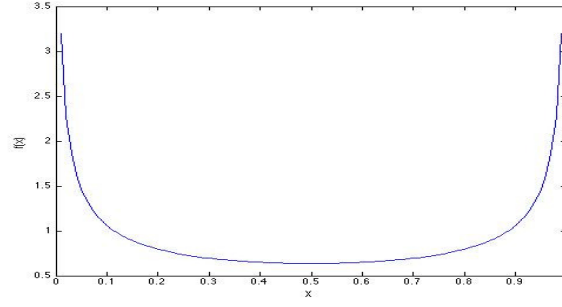
Figure 1: Probability density function of the arcsine distribution

The approximate discrete random variable ξ constructed by RPs-NTM has been introduced in section 1. It takes $P(\xi = \xi_j) = \frac{1}{n}, j = 1, \ldots, n$, where

$$\xi_j = F^{-1}(q_j), \ q_j = \frac{2j-1}{2n}, \quad j = 1, \ldots, n. \tag{3}$$

It is easy to find, for arcsine distribution, $F^{-1}(x) = \sin^2(\frac{\pi x}{2})$ and the representative points are

$$\xi_j = \sin^2(\frac{2j-1}{2n} \cdot \frac{\pi}{2}), \ j = 1, 2, ..., n. \tag{4}$$

The following theorem derives a surprising approximation property when using RPs-NTM to approximate the arcsine distribution.

Theorem 1. *Let random variable X follow the standard arcsine distribution and let ξ be an approximate random variable defined by $P(\xi = \xi_j) = \frac{1}{n}$, where $\xi_j = \sin^2(\frac{2j-1}{2n} \cdot \frac{\pi}{2})$, $j = 1, 2, ..., n$. Then X and ξ share all the same moments as follows*

$$\mu_m = E(X^m) = E(\xi^m) = \prod_{j=1}^{m} \frac{2j-1}{2j}, \ m = 1, 2, ..., \ n > m. \tag{5}$$

Proof: The mth moment of the arcsine distribution is easy to find

$$E(X^m) = \int_0^1 x^m \frac{1}{\pi\sqrt{x(1-x)}} dx = \int_0^1 \frac{2u^{2m}}{\pi\sqrt{1-u^2}} du = \int_0^{\frac{\pi}{2}} \frac{2}{\pi} \sin^{2m}(\theta) d\theta = \prod_{j=1}^{m} \frac{2j-1}{2j}.$$

The mth moment of the approximate distribution is given by

$$E(\xi^m) = \frac{1}{n} \sum_{j=1}^{n} \sin^{2m}(\frac{\pi}{2} \frac{2j-1}{2n}).$$

To prove the theorem, we need to show

$$\frac{1}{n} \sum_{j=1}^{n} \sin^{2m}(\frac{\pi}{2} \frac{2j-1}{2n}) = \prod_{j=1}^{m} \frac{2j-1}{2j}. \tag{6}$$

160

Let $\theta = \frac{\pi}{2}\frac{2j-1}{2n}$,

$$
\begin{aligned}
E(\xi^m) &= \sum_{j=1}^{n}\frac{1}{n}\sin^{2m}(\frac{\pi}{2}\frac{2j-1}{2n}) = \frac{1}{n}\sum_{j=1}^{n}\sin^{2m}(\theta) \\
&= \frac{1}{n}\sum_{j=1}^{n}[\frac{e^{i\theta}-e^{-i\theta}}{2i}]^{2m} \\
&= \frac{1}{n}\sum_{j=1}^{n}(-1)^m(2^{-2m})\sum_{k=0}^{2m}\binom{2m}{k}(-1)^k(e^{i\theta})^{2k-2m} \\
&= \frac{1}{n}(-1)^m(2^{-2m})\sum_{k=0}^{2m}\binom{2m}{k}(-1)^k\sum_{j=1}^{n}e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})(2j-1)}.
\end{aligned}
\tag{7}
$$

Note that in the right hand side of the above equation we have

$$
\sum_{j=1}^{n}\left(e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})}\right)^{(2j-1)} = \left\{ \begin{array}{ll} n, & \text{if } k=m, \\ \frac{y(1-y^{2n})}{1-y^2}, & \text{if } k\neq m. \end{array} \right.
\tag{8}
$$

where $y = e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})}$. Moreover, $1-y^{2n} = 1-e^{i(k-m)\pi} = 2$ when the value of $(k-m)$ is odd, otherwise $1-y^{2n} = 0$.

Let $\frac{k-m}{n}\cdot\frac{\pi}{2} = \alpha\pi$, we have $y = e^{i\alpha\pi}$ and

$$
\frac{y}{1-y^2} = \frac{e^{i\alpha\pi}}{1-e^{i2\alpha\pi}} = \frac{2i(\sin\alpha\pi)}{[1-\cos(2\alpha\pi)]^2+[\sin(2\alpha\pi)]^2}.
\tag{9}
$$

From (9), for $k\neq m$ and $n>m$ we have that

$$
\text{real part of } \sum_{j=1}^{n}\left(e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})}\right)^{(2j-1)} = 0 \text{ and}
\tag{10}
$$

$$
\text{imaginary part of } \sum_{j=1}^{n}\left(e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})}\right)^{(2j-1)} = -\text{imaginary part of } \sum_{j=1}^{n}\left(e^{i(\frac{m-k}{n}\cdot\frac{\pi}{2})}\right)^{(2j-1)}
\tag{11}
$$

Therefore, from (8)-(11), we have

$$
\begin{aligned}
&\sum_{k=0}^{2m}\binom{2m}{k}(-1)^k\sum_{j=1}^{n}\left(e^{i(\frac{k-m}{n}\cdot\frac{\pi}{2})}\right)^{(2j-1)} \\
&= (-1)^m\binom{2m}{m}\cdot n + \sum_{k=0}^{m-1}\binom{2m}{k}(-1)^k\sum_{j=1}^{n}(e^{i(\frac{k-m}{n})\cdot\frac{\pi}{2}})^{(2j-1)} \\
&\quad + \sum_{k=m+1}^{2m}\binom{2m}{k}(-1)^k\sum_{j=1}^{n}(e^{i(\frac{k-m}{n})\cdot\frac{\pi}{2}})^{(2j-1)} \\
&= (-1)^m\binom{2m}{m}n.
\end{aligned}
\tag{12}
$$

Substituting (12) into (7), we obtain

$$
\begin{aligned}
E(\xi^m) &= \frac{1}{n}(-1)^m(2^{-2m})(-1)^m\binom{2m}{m}n = \binom{2m}{m}(2^{-2m}) \\
&= \frac{1 \times 3 \times 5 \times \cdots \times (2m-1)}{2^{2m} \times m!} \\
&= \prod_{j=1}^{m} \frac{2j-1}{2j} = E(X^m). \quad \square
\end{aligned}
$$

According to theorem 1, we can conclude that RPs-NTM provides the best discrete approximation for arcsine distribution in the sense of matching moments.

3   RPs-MSE of the arcsine distribution

Recently, Fang, Zhou and Wang [9] propose to use RPs-MSE, RPs-NTM and Monte Carlo method in resampling to construct approximate distributions for a given continuous distribution. They compare the three kind of approximate distributions to the standard normal distribution and find that in most cases RPs-MSE has the best performance when $F(x)$ is a normal distribution. However, from Theorem 1 we can expect that, for arcsine distribution, different conclusion on the performance of these three kinds of approximate distributions may be found. Since, so far, there is no results on RPs-MSE of the arcsine distribution, in this section, following Fang and He [7], we give an algorithm for generation RPs-MSE.

A set of representative points, denoted as RPs-MSE, which is a set of points that optimally represents a distribution in terms of mean square error (Cox [3], Bofinger[1], Fang and He [7], Flury [11, 12]). Let X be a random variable with pdf $f(x)$. Consider a set of points $\{\xi_1, \cdots, \xi_n\}$ where $\xi_1 <, \cdots, < \xi_n$ and define a loss function

$$
L(\xi_1, \xi_2, \ldots, \xi_n) = \frac{1}{\sigma^2} \int_{-\infty}^{\infty} min_{1 \le i \le n}(\xi_i - x)^2 f(x)dx, \tag{13}
$$

where σ is the standard deviation of X. Then, a set points $\{\xi_1^*, \cdots, \xi_n^*\}$ is said to be RPs-MSE of the distribution of X if it minimizes the loss function (13) for all choices of $\{\xi_1, \cdots, \xi_n\}$.

Particularly, for a random variable X following the arcsine distribution, its variance is $\frac{1}{8}$ and the mean squared distance loss function (13) is given by

$$
\begin{aligned}
L(\xi_1, \xi_2 \ldots, \xi_n) &= \frac{1}{\sigma^2} \int_0^1 min_{1 \le i \le n}(\xi_i - x)^2 f(x)dx \\
&= 8 \int_0^{\frac{1}{2}(\xi_1+\xi_2)} \frac{(\xi_1 - x)^2}{\pi\sqrt{x(1-x)}}dx \;+\; 8 \int_{\frac{1}{2}(\xi_1+\xi_2)}^{\frac{1}{2}(\xi_2+\xi_3)} \frac{(\xi_2 - x)^2}{\pi\sqrt{x(1-x)}}dx \\
&\quad + \ldots + 8 \int_{\frac{1}{2}(\xi_{n-1}+\xi_n)}^{1} \frac{(\xi_n - x)^2}{\pi\sqrt{x(1-x)}}dx.
\end{aligned} \tag{14}
$$

For the ith $(1 \le i \le n)$ term in the loss function (13)

$$8 \int_{\frac{1}{2}(\xi_{i-1}+\xi_i)}^{\frac{1}{2}(\xi_i+\xi_{i+1})} \frac{(\xi_i - x)^2}{\pi \sqrt{x(1-x)}} dx = \frac{8}{\pi} \int_{\frac{1}{2}(\xi_{i-1}+\xi_i)}^{\frac{1}{2}(\xi_i+\xi_{i+1})} \frac{\xi_i^2 - 2x\xi_i + x^2}{\sqrt{x(1-x)}} dx$$

$$= \frac{16}{\pi} \left[\left(\frac{3}{8} - \xi_i + \xi_i^2 \right) \left(arcsin \left(\sqrt{b_{i+1}} \right) - arcsin \left(\sqrt{b_i} \right) \right) - \left(\xi_i - \frac{3}{8} - \frac{1}{4} b_i \right) \left(b_i - b_i^2 \right)^{\frac{1}{2}} \right.$$

$$\left. + \left(\xi_i - \frac{3}{8} - \frac{1}{4} b_{i+1} \right) \left(b_{i+1} - b_{i+1}^2 \right)^{\frac{1}{2}} \right],$$

where

$$b_i = \begin{cases} 0, & i = 1, \\ \frac{1}{2}(\xi_{i-1} + \xi_i), & 2 \le i \le n, \\ 1, & i = n+1. \end{cases}$$

Thus, RPs-MSE $\{\xi_1^*, \cdots, \xi_n^*\}$ need to be determined by minimizing the loss function (14).

As the density function of arcsine distribution is an symmetrical function about center $\frac{1}{2}$, it is natural to assume that the RPs of arcsine distribution are also symmetric, i.e. $\xi_j = 1 - \xi_{n-j+1}$, $1 \le j \le m+1$, $m = \lfloor \frac{n}{2} \rfloor$. So, considering the symmetric, the loss function which is entirely determined by the first half of n points is displayed as $L(\xi_1, \ldots, \xi_m)$. In addition, two cases need to be dealt with respectively: when $n = 2m+1$ is odd, RPs $\{\xi_1, \cdots, \xi_m\}$ is selected under the constraint of $0 < \xi_1 < \cdots < \xi_m < \xi_{m+1} = \frac{1}{2}$; when $n = 2m$ is even, RPs $\{\xi_1, \cdots, \xi_m\}$ is determined under the constraint of $0 < \xi_1 < \cdots < \xi_m < \frac{1}{2}$.

To find a set of RPs-MSE which minimizes loss function $L(\xi_1, \ldots, \xi_m)$, we set

$$\frac{\partial L(\xi_1, \ldots, \xi_m)}{\partial \xi_i} = 0, \ i = 1, \ldots, m,$$

and then obtain the following system of equations

$$(2\xi_i - 1) \left(arcsin \left(\sqrt{b_{i+1}} \right) - arcsin \left(\sqrt{b_i} \right) \right)$$

$$= \left(b_i - b_i^2 \right)^{\frac{1}{2}} - \left(b_{i+1} - b_{i+1}^2 \right)^{\frac{1}{2}}, i = 1, \ldots, m. \quad (15)$$

In system of equations (15), when $n = 2m$ is even,

$$b_i = \begin{cases} 0 & i = 1 \\ \frac{1}{2}(\xi_{i-1} + \xi_i) & 2 \le i \le m \\ \frac{1}{2} & i = m+1, \end{cases}$$

when $n = 2m + 1$ is odd,

$$b_i = \begin{cases} 0 & i = 1 \\ \frac{1}{2}(\xi_{i-1} + \xi_i) & 2 \le i \le m \\ \frac{1}{2}(\xi_m + \frac{1}{2}) & i = m+1. \end{cases}$$

To solve the system of equations (15), we adopt an iterative numerical algorithm proposed by Fang and He [7]. The procedure is shown in Algorithm 1.

After determining the solution $\{\xi_1^*, \ldots, \xi_m^*\}$ of system of equations (15), RPs-MSE is the set of $\xi^* = \{\xi_1^*, \ldots, \xi_m^*, 1 - \xi_1^*, \ldots, 1 - \xi_m^*\}$ when n is even and the set of

---

Algorithm 1. Iterative algorithm for solving the system of equations (15)

Step 1   Choose an appropriate initial value for $\xi_1 > 0$. Set $LP = \xi_1$ and $RP = \frac{1}{2}$ such that ξ_2 must lie in the interval (LP, RP).

Step 2   Fix ξ_1, solve the first equation and obtain the value of ξ_2. And then fix ξ_1 and ξ_2, solve the second equation to get ξ_3.

Step 3   In the same manner, fix ξ_{i-1} and ξ_i, solve the ith equation and obtain the solution of ξ_{i+1}, for i from 2 to $m-1$. Thus ξ_m is found by fixing ξ_{m-2} and ξ_{m-1} and solving the second last equation.

Step 4   Fix ξ_{m-1}, solve the last equation and find the solution of ξ'_m.

Step 5   Let ϵ be the threshold of convergence (assume $\epsilon = 10^{-6}$), if

[a] $|\xi_m - \xi'_m| < \epsilon$, then the $\{\xi_1, \ldots, \xi_m\}$ are the solution for system of equations (18).

[b] $\xi_m < \xi'_m - \epsilon$, then the initial value ξ_1 is too small, modify $LP = \xi_1$, $\xi_1 = \frac{1}{2}(LP + RP)$, go to the step 2.

[c] $\xi_m > \xi'_m + \epsilon$, then the initial value ξ_1 is too large, modify $RP = \xi_1$, $\xi_1 = \frac{1}{2}(LP + RP)$, go to the step 2.

$\xi^* = \{\xi_1^*, \ldots, \xi_m^*, \frac{1}{2}, 1 - \xi_1^*, \ldots, 1 - \xi_m^*\}$ when n is odd.

In addition, the associated probability of RPs-MSE is computed as

$$
\begin{aligned}
p_i &= \int_{b_i}^{b_{i+1}} f(x)dx \\
&= \int_{b_i}^{b_{i+1}} \frac{1}{\pi\sqrt{x(1-x)}}dx \\
&= \frac{2}{\pi}\left(arcsin\left(\sqrt{b_{i+1}}\right) - arcsin\left(\sqrt{b_i}\right)\right).
\end{aligned}
\tag{16}
$$

The RPs-MSE for $n \leq 31$ and related probabilities p_i can be found in Jiang [15].

4   Comparisons between RPs-NTM and RPs-MSE

RPs-NTM and RPs-MSE are two types of representative points obtained by using different optimal criteria and can be used to constitute discrete approximation to the arcsine variate, denoted by Y_{NTM} and Y_{MSE}, respectively. In this section, we evaluate these two discrete approximations by considering L_p F-discrepancy and statistical inference based on resampling.

The L_p F-discrepancy is defined by Fang and Wang [8] as the L_p distance between the original distribution $F(x)$ and its discrete approximation, $F_\xi(x)$.

$$
D_p(F, F_\xi) = \left[\int_{-\infty}^{+\infty} |F_\xi(x) - F(x)|^p \ dx\right]^{\frac{1}{p}}.
\tag{17}
$$

It has been shown that it is a very useful measurement for evaluating discrete approximation.

To compare the performance of discrete approximation constructed by RPs-NTM and RPs-MSE, we compute the value of $D_2(F, F_{Y_{NTM}})$ and $D_2(F, F_{Y_{MSE}})$ for different number of RPs n. The results are shown in Table 1. Clearly, the distribution function of Y_{NTM} is much closer to the arcsine distribution function.

Resampling technique has been used in statistical simulation. Bootstrap method is one of the resampling methods and was proposed by Eron in 1979 [5]. Resampling method

Table 1: L_2 distance between the approximate distribution and the arcsine distribution

n	5	10	15	20	25	30	31
$D_2(F, F_{Y_{NTM}})$	0.05807	0.02891	0.01926	0.01444	0.01155	0.00962	0.00931
$D_2(F, F_{Y_{MSE}})$	0.10376	0.08621	0.07541	0.06778	0.06204	0.05751	0.05672

takes samples from an approximate population. The bootstrap uses a sample from $F(x)$ to construct an approximate distribution. Fang et al.[9] proposed to use RPs-NTM and RPs-MSE for construction of approximate distribution and found that in many cases RPs-MSE has the best performance and RPs-NTM is not as good as RPs-MSE, but both of RPs-NTM and RPs-MSE can significantly improve the traditional bootstrap method. For the normal population they pointed out $D_2(F, F_{Y_{NTM}}) > D_2(F, F_{Y_{MSE}})$, but for the arcsine distribution $D_2(F, F_{Y_{NTM}})$ is smaller and may have a better performance in statistical inference by resampling.

In the following, resampling from Y_{NTM} and Y_{MSE} are taken for estimation of the mean, variance, skewness and kurtosis of X. It is known

$$E(X) = \frac{1}{2}, \quad Var(X) = \frac{1}{8}, \quad Sk(X) = 0, \quad Ku(X) = -\frac{3}{2}.$$

Six cases with different number of representatives points $n = 20, 25, 28, 29, 30, 31$ are considered respectively. For each case, resampling are repeated $N = 1000, 2000, 5000, 10000$ times. Table 2 shows the estimation bias of four statistics by resampling form Y_{NTM} and Y_{MSE} for $n = 30$. The numbers labeled with * means the winner, with the smaller absolute value of bias, in the comparison. To save space, we only show the estimation bias when $n = 30$. Table 3 shows the number of winner for the two approximations among all the estimations. From the table 2 and 3, we observe that the discrete approximation Y_{NTM} for arcsine distribution has a better accuracy in estimating the statistics in most cases, however, the difference on accuracy between Y_{NTM} and Y_{MSE} is not very significant, especially when the number of repetition for resampling is large.

It is worth being mentioned, in previous studies about representative points, no matter for normal distribution or t distribution or other distributions, Y_{MSE} always has the best performance in approximation and resampling. However, in the case of the arcsine distribution, because of its special property (Theorem 1), Y_{NTM} often achieves the better performance than Y_{MSE}.

Table 2: n=30, estimation bias by resampling

	Category	1000	2000	5000	10000
Mean	Y_{NTM}	-0.002032*	0.000556*	-0.001091*	0.000265*
	Y_{MSE}	0.004421	-0.001554	-0.001466	-0.000752
Variance	Y_{NTM}	-0.000343*	0.000240	-0.000004*	-0.000063*
	Y_{MSE}	-0.001102	0.000162*	-0.000057	0.000091
Skewness	Y_{NTM}	0.010411*	-0.001817*	0.006044	-0.000351*
	Y_{MSE}	-0.020095	0.007899	0.005315*	0.003085
Kurtosis	Y_{NTM}	0.117548	0.111009	0.110980	0.111487
	Y_{MSE}	0.117018*	0.109467*	0.110104*	0.108305*

The * indicates the winner in each column of each statistic

Table 3: Number of winner of statistical estimation

	Y_{NTM}	Y_{MSE}
Mean	15	9
Variance	17	7
Skewness	13	11
Kurtosis	8	16
Total	53	43

5   Conclusion

This paper mainly concerns the two type of representative points (RPs-NTM and RPs-MSE) for the arcsine distribution. We discover and prove a special property of representative points selected by Number-Theoretic method for the arcsine distribution: The discrete approximation constructed by RPs-NTM shares all the same moments with the arcsine distribution. Due to this property, RPs-NTM has the best representative of the arcsine distribution in the sense of L_2 discrepancy. Another type of representative points are determined in term of minimizing mean square error (RPs-MSE). Our statistical comparisons show that RPs-NTM is often better than RPs-MSE as the discrete approximation to arcsine distribution.

Acknowledgement

References

[1] Boringer, E. (1970). Maximizing the correlation of grouped observations. *J. Amer. Statist. Theo.*, 65: 1632-1638.

[2] Clemen, R.C. (1991). *Making Hard Decision: An Introduction to Decision Analysis*, PWS-Kent Publishing Co., Boston.

[3] Cox, D.R. (1957). A note on grouping. *Journal of the American Statistical Association*, 52: 543-547.

[4] Dalenius, T. (1950). The problem of optimum stratification. *Skandinavisk Aktuarietidskrift*, 33: 203-213.

[5] Efron, B.(1979). Bootstrap methods: another look at the jacknife. *The Annals of Statistics*, 7: 1-26.

[6] Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

[7] Fang, K.T. and He, S. (1982). The problem of selecting a given number of representative points in a normal population and a generalized mill's ratio. Technical report,

[8] Fang, K.T. and Wang, Y. (1994). *Number-theoretic Methods in Statistics*, Chapman and Hall, London.

[9] Fang, K.T., Zhou, M. and Wang, W.J. (2014), Applications of the representative points in statistical simulations. *Science in China, Series A*, 57,doi: 10.1007/s11425-014-4860.

[10] Feller, W. (1967). *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons.

[11] Flury, B. (1990). Principal points. *Biometrika*, 77(l): 33-41.

[12] Flury, B. (1993). Estimation of principal points. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 42(1): 139-151.

[13] Guo, X., Jarrow, R.A., de Larrard, A. (2014). The economic default time and the arcsine law. *Journal of Financial Engineering*, 1(03).

[14] Hammond, R.K., Bickel, J.E. (2013). Reexamining discrete approximations to continuous distributions. *Decision Analysis*, 10(1): 6-25.

[15] Jiang, J.J. (2014). A surprising properties of arcsine distribution and its applications, A final year project thesis, BNU-HKBU United International College, Zhuhai, China.

[16] Keefer, D.L. and Bodily, S.E. (1983). 3-point appoximations for continuous random variables. *Management Science*, 29: 595-609.

[17] Lee, Y.W. (1960). *Statistical Theory of Communications*, New York:John Wiley & Sons.

[18] Lévy, P. (1939). Sur certains processus stochastiques homogenes, *Compositio Math.*, 7: 283-339.

[19] Lévy, P. (1965). *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris.

[20] Middleton, D. (1965). *Introduction to Statistical Communications Theory*, New York: McGraw-Hill.

[21] Pechtl, A. (1999). Distributions of occupation times of Brownian motion with drift. *Journal of Applied Mathematics and Decision Sciences*, 3(1): 41-62.

[22] Smith, J.E. (1993) Moment methods for decision analysis. *Management Sci.* 39(3): 340-358.

Direct and Indirect Effects on Road Productivity in Japan

Kazuhiko Kakamu[1,*]   Hideo Yunoue[2,†]

[1]Graduate School of Business Administration, Kobe University
[2]School of Economics, University of Hyogo

Abstract   The spatial econometrics enable us to decompose the marginal effects of the concerning variables into direct and indirect effects. LeSage and Pace [9] proposed the summary measures for them and showed that the Bayesian approach makes us implement the statistical inference for the measures. This study examines the productivity of roads in Japan from a viewpoint of the spatial econometrics using this technique and argues the direct and indirect effects from a Bayesian point of view. From the empirical results, we can find that the marginal effect is underestimated if the indirect effect is ignored.

Keywords   Direct and indirect effects; Markov chain Monte Carlo method; Spatial Durbin model

1   Introduction

The spatial econometrics enable us to decompose the marginal effects of the concerning variables into direct and indirect effects. LeSage and Pace [9] proposed the summary measures for them and showed that the Bayesian approach makes us implement the statistical inference for the measures. This study examines the productivity of roads in Japan from a viewpoint of the spatial econometrics using this technique and argues the direct and indirect effects from a Bayesian point of view. The road is widely known as typical example of public good. Since one of the purpose of subsidy is to modify the under-provision of public goods, it is important to evaluate the indirect effect, so-called spillover effect.

The rest of this paper is organized as follows. In the next section, we briefly explain the spatial econometric model, which we will use in the decomposition of the marginal effects. Section 3 provides the empirical results, which examines the productivity of roads in Japan. Section 4 gives some concluding remarks.

2   Model

In a simple linear regression model (LRM),

$$y_i = \alpha + \mathbf{x}_i \boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where y_i is a dependent variable and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})$ is a covariates, $\frac{\partial y_i}{\partial x_{ik}} = \beta_k$, where β_k is the kth element of $\boldsymbol{\beta}$, and $\frac{\partial y_i}{\partial x_{jk}} = 0 \ (i \neq j)$. However, if we focus on a spatial

---

*Corresponding author. Email: kakamu@person.kobe-u.ac.jp
†Email: yunoue@econ.u-hyogo.ac.jp

econometric model, not only $\dfrac{\partial y_i}{\partial x_{ik}} \neq \beta_k$ but also $\dfrac{\partial y_i}{\partial x_{jk}} \neq 0$ in general. Therefore, it was difficult to analyze the partial derivatives in the spatial models.

To overcome this problem, LeSage and Pace [9] proposed the summary measures for the partial derivatives and decomposed them into direct and indirect effects. This approach becomes popular and the standard method for the spatial analyses (see, for example, LeSage and Dominguez [8]). Therefore, in this section, we explain this method using the general spatial model named spatial Durbin model (SDM).

Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_n')'$ and $\mathbf{W}$ be an $n \times n$ spatial weight matrix, respectively. Then, the SDM is expressed as follows:

$$
\begin{aligned}
\mathbf{y} &= \rho \mathbf{W} \mathbf{y} + \alpha \boldsymbol{\iota}_n + \mathbf{X} \boldsymbol{\beta} + \mathbf{W} \mathbf{X} \boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n), \\
&= \rho \mathbf{W} \mathbf{y} + \mathbf{Z} \boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n),
\end{aligned} \tag{2}
$$

where $\boldsymbol{\iota}_n$ is an $n \times 1$ unit vector, $\boldsymbol{\gamma}$ is an $k \times 1$ parameter vector, $\mathbf{Z} = [\boldsymbol{\iota}_n \ \mathbf{X} \ \mathbf{W} \mathbf{X}]$, $\boldsymbol{\theta} = (\alpha \ \boldsymbol{\beta}' \ \boldsymbol{\gamma}')'$, and $\mathbf{I}_n$ is an $n \times n$ unit matrix, respectively.

To introduce the summary measures for the partial derivatives, let us introduce the matrix $\mathbf{S}_k(\mathbf{W}) = (\mathbf{I}_n - \rho \mathbf{W})^{-1} (\beta_k \mathbf{I}_n + \gamma_k \mathbf{W})$ to construct the direct and indirect effect for the kth variable from the SDM model (2). Then, $\dfrac{\partial y_i}{\partial x_{jk}} = \mathbf{S}_k(\mathbf{W})_{ij}$, where $\mathbf{S}_k(\mathbf{W})_{ij}$ is the ijth element of $\mathbf{S}_k(\mathbf{W})$. Therefore, LeSage and Pace [9] proposed the summary measures for total, direct and indirect effects as follows:

$$
\begin{aligned}
\bar{M}(k)_{total} &= n^{-1} \boldsymbol{\iota}_n' S_k(\mathbf{W}) \boldsymbol{\iota}_n, & (3) \\
\bar{M}(k)_{direct} &= n^{-1} tr(\mathbf{S}_k(\mathbf{W})), & (4) \\
\bar{M}(k)_{indirect} &= \bar{M}(k)_{total} - \bar{M}(k)_{direct}, & (5)
\end{aligned}
$$

where $\bar{M}(k)_{total}$ is the total effect, $\bar{M}(k)_{direct}$ is the direct effect, and $\bar{M}(k)_{indirect}$ is the indirect effect, respectively. Note that the summary measures take the average for these effects. However, it is difficult to judge the statistical significance for these effects directly, because these measures are the non-linear function of the parameters. As shown by Gelfand *et al.* [4], Markov chain Monte Carlo (MCMC) methods can yield valid inference on non-linear functions of the parameters such as the direct and indirect impacts in (4) and (5). All that is required is evaluation and storage of the draws reflecting the non-linear combinations of the parameters. Therefore, we utilize the MCMC to examine the direct and indirect effects (see LeSage and Pace [9]).

Then, to proceed the Bayesian inference, we will introduce the likelihood function for (2). The likelihood function of (2) is written as follows:

$$
L(\mathbf{y}|\mathbf{X}, \mathbf{W}, \boldsymbol{\theta}, \sigma^2, \rho) = \frac{1}{\sqrt{2\pi\sigma^2}^n} |\mathbf{I}_n - \rho \mathbf{W}| \exp\left\{ -\frac{\mathbf{e}'\mathbf{e}}{2\sigma^2} \right\}, \tag{6}
$$

where $\mathbf{e} = \mathbf{y} - \rho \mathbf{W} \mathbf{y} - \mathbf{Z} \boldsymbol{\theta}$.

Given the likelihood function (2), the prior distributions are required to proceed the posterior inference. Therefore, we assume the following independent prior distributions.

$$
\pi(\boldsymbol{\theta}, \sigma^2, \rho) = \pi(\boldsymbol{\theta}) \pi(\sigma^2) \pi(\rho). \tag{7}
$$

As the prior distribution for each parameter, we assume the following proper prior distributions.

$$
\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0), \quad \sigma^2 \sim \mathcal{IG}(\nu_0/2, \lambda_0/2), \quad \rho \sim \mathcal{U}(1/\lambda_{\min}, 1/\lambda_{\max}),
$$

where $\mathcal{IG}(a, b)$ denotes an inverse gamma distribution with parameters a and b, and $\mathcal{U}(a, b)$ denotes an uniform distribution with an interval between a and b. $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{W}$. As shown in Sun *et al.* [10], it is well known that $\lambda_{\min}^{-1} < 0$ and $\lambda_{\max}^{-1} > 0$ and that ρ must lie in the interval. Therefore, we restrict the prior space as $\rho \in (\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$. In addition, it is also well known that the maximum eigenvalue of a row-standardized weight matrix is one. Therefore, we set $\lambda_{\max} = 1$.

Given a prior density in (7) and the likelihood function in (2), the joint posterior distribution can be expressed as:

$$\pi(\boldsymbol{\theta}, \sigma^2, \rho | \mathbf{y}, \mathbf{X}, \mathbf{W}) \propto \pi(\boldsymbol{\theta})\pi(\sigma^2)\pi(\rho)L(\mathbf{y}|\mathbf{X}, \mathbf{W}, \boldsymbol{\theta}, \sigma^2, \rho). \tag{8}$$

We can now employ MCMC methods. The Markov chain sampling scheme can be constructed from the full conditional distributions (FCD) of $\boldsymbol{\theta}$, σ^2 and ρ. Thus, we derive the FCD for each parameter. The FCD for $\boldsymbol{\theta}$ is expressed by

$$\boldsymbol{\theta}|\mathbf{y}, \mathbf{Z}, \mathbf{W}, \sigma^2, \rho \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}), \tag{9}$$

where $\hat{\boldsymbol{\Sigma}} = \left(\sigma^{-2}\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}$ and $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Sigma}}\left\{\sigma^{-2}\mathbf{Z}'\left(\mathbf{y} - \rho\mathbf{W}\mathbf{y}\right) + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right\}$. Therefore, the standard Gibbs sampler is utilized to draw the posterior (see Gelfand and Smith [5]).

The FCD for σ^2 is given by

$$\sigma^2|\mathbf{y}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\theta}, \rho \sim \mathcal{IG}(\hat{\nu}/2, \hat{\lambda}/2), \tag{10}$$

where $\hat{\nu} = n + \nu_0$ and $\hat{\lambda} = \mathbf{e}'\mathbf{e} + \lambda_0$. This is also drawn by the Gibbs sampler.

The FCD for ρ is expressed as

$$\pi(\rho|\mathbf{y}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\theta}, \sigma^2) \propto |\mathbf{I}_n - \rho\mathbf{W}| \exp\left\{-\frac{\mathbf{e}'\mathbf{e}}{2\sigma^2}\right\}. \tag{11}$$

However, it is difficult to draw the posterior from this FCD directly. Thus, we utilize the random walk Metropolis-Hastings (MH) algorithm (see Holloway *et al.* [6]).

3 Empirical Results

Before examining empirics, we explain the dataset used in this paper. We utilize prefecture-level data in fiscal year 2010 in Japan[1]. The amount of road traffic, y, road length, x_1, and congestion rate, x_2, are derived from the *The FY2010 Road Traffic Census*. Since we consider the operation rate of the capital, we include congestion rate of the road in our production function. As the spatial weight matrix $\mathbf{W}$, we use the one, which is proposed by Kakamu *et al.* [7].

We used the following hyper-parameters to run the MCMC algorithm.

$$\boldsymbol{\theta}_0 = \mathbf{0}, \ \boldsymbol{\Sigma}_0 = 100 \times \mathbf{I}_{2K+1}, \ \nu_0 = 2.0, \ \lambda_0 = 0.01.$$

We perform the MCMC procedure by generating 1,010,000 iterations and discarding the first 10,000 iterations. Of the remaining draws, we keep every 100th draw to obtain the posterior statistics for the parameters. All the results reported here are generated using Ox version 7.0(OS_X_64/U) (see Doornik [3]).

Table 1 provides the estimation result of the SDM. To compare the result with the standard LRM, we also report the estimation result of the LRM in (1). α is the constant

---

[1]Since the survey was conducted on autumn, the results were not affected by the huge earthquake on March 2011.

Table 1: Estimation Results

	SDM			LRM		
	MEAN	95%CI		MEAN	95%CI	
α	-7.725	-10.055	-5.374	-8.395	-10.722	-6.080
β_0	1.118	0.956	1.282	1.091	0.933	1.244
β_1	2.174	1.769	2.580	2.332	1.967	2.693
γ_0	-0.245	-0.501	0.030			
γ_1	-0.249	-0.743	0.268			
ρ	0.274	-0.041	0.554			
σ^2	0.045	0.029	0.068	0.047	0.031	0.071

Note: MEAN and 95%CI stand for the posterior mean and 95% credible interval, respectively.

Table 2: Direct and Indirect Effects

	VARIABLE 1			VARIABLE 2		
	MEAN	95%CI		MEAN	95%CI	
Direct Effect	1.334	1.037	1.709	2.214	1.801	2.623
Indirect Effect	3.310	2.047	5.373	0.504	-0.172	1.538
Total Effect	4.644	3.120	7.039	2.718	1.871	3.940

term. β_1 and β_2 represent the coefficients of road capital and its operation rate, respectively. We also report γ_1, γ_2 and ρ, that represent the spatial dependency parameter. β_1 is estimated as positive and the 95% credible interval for the coefficient does not include zero. We find that the production function of road service satisfied the ordinary assumption. β_2 is also reported as positive value. We confirm that higher operated road had higher productivity. On the other hand, γ_1 and γ_2 is estimated as negative, but the 95% credible interval for the coefficient include zero. The 95% credible interval for ρ also include zero. Both output and input of neighboring prefecture do not affect the road productivity of home prefecture.

The direct and indirect effects are reported in Table 2. The direct effect of capital on road productivity is estimated positive and the 95% credible interval does not include zero. The indirect effect of capital is also estimated positive and its 95% credible interval does not include zero. The indirect effect is larger than direct effect because our method "cumulated spillover" of all regions[2].

The direct effect of operation rate is positive and the 95% credible interval does not include zero. Note that the credible interval of indirect effect of operation rate include zero. This result shows that there is direct effect of the operation rate on road productivity, but not spillover effect.

4 Conclusions

In this paper, we estimated the spillover effect on road productivity using a spatial regression model. We found the first evidence of spillover effects on road production function at the prefecture level in Japan. In particular, we observed that the indirect effects, so-called spillovers, have larger impact on the road productivity than the direct effect. The road is widely known as a public good. When the local governments put a low value on, or sometimes ignore, the effect of spillover, the road is under-provided. The

---

[2]LeSage and Dominguez (2012) made a mention of this point.

national subsidy of road construction is likely to modify the under provision of road capital in Japan.

Acknowledgement

References

[1] Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers: Dordrecht.

[2] Anselin, L. (2010). Thirty years of spatial econometrics, *Papers in Regional Science*, **89**, 3–25.

[3] Doornik, J.A. (2009). *Ox: Object Oriented Matrix Programming Language*. Timberlake Consultants Press: London.

[4] Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *Journal of the American Statistical Association*, **85**, 972–985.

[5] Gelfand, A.E., Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.

[6] Holloway, G., Shankar, B. and Rahman, S. (2002). Bayesian spatial probit estimation: A primer and an application to HYV rice adoption, *Agricultural Economics*, **27**, 384–402.

[7] Kakamu, K., Polasek, W. and Wago, H. (2008). Spatial interaction of crime incidents in Japan, *Mathematics and Computers in Simulation*, **78**, 276–282.

[8] LeSage, J.P. and Dominguez, M. (2012). The importance of modeling spatial spillovers in public choice analysis, *Public Choice*, **150**, 525–545.

[9] LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*, CRC Press: Boca Raton.

[10] Sun, D., Tsutakawa, R.K. and Speckman, P.L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions, *Biometrika*, **86**, 341–350.

Nonlinear Markov Processes in Big Networks

Quan-Lin Li*

School of Economics and Management Sciences
Yanshan University, Qinhuangdao 066004, China

Abstract   Big networks express various large-scale networks in many practical areas such as computer networks, internet of things, cloud computation, manufacturing systems, transportation networks, and healthcare systems. This paper analyzes such big networks, and applies the mean-field theory and the nonlinear Markov processes to set up a broad class of nonlinear continuous-time block-structured Markov processes, which can be applied to deal with many practical stochastic systems. Firstly, a nonlinear Markov process is derived from a large number of interacting big networks with symmetric interactions, each of which is described as a continuous-time block-structured Markov process. Secondly, some effective algorithms are given for computing the fixed points of the nonlinear Markov process by means of the UL-type *RG*-factorization. Finally, the Birkhoff center, the Lyapunov functions and the relative entropy are used to analyze stability or metastability of the big network, and several interesting open problems are proposed with detailed interpretation. We believe that the results given in this paper can be useful and effective in the study of big networks.

Keywords   Nonlinear Markov process; Big network; Mean-field theory; *RG*-factorization; Fixed point; Stability; Metastability; Lyapunov function; Relative entropy

1   Introduction

In this paper, we consider a large number of interacting big networks with symmetric interactions, each of which is described as a continuous-time block-structured Markov process, which can be applied to deal with many practical stochastic systems. As the number of nodes goes to infinite, the interactions between any two subsets of the big networks become negligible or are asymptotically independent, and the overall effect of the interactions can be replaced by an empirical measure under the mean-field setting. Based on this, the evolution of each big network is expressed as a time-inhomogeneous continuous-time Markov process, which leads to that the transient performance of any big network can be given by a system of ordinary differential equations, and its associated steady-state performance is able to be computed by any fixed point, which satisfies a system of nonlinear equations.

The purpose of this paper is to develop the mean-field computational theory both for performance evaluation and for performance optimization. During the last three decades considerable attention has been paid to studying the mean-field theory of big networks. Readers may refer to recent publications for the mean-field theory of stochastic systems, among which are Dawson [17], Shiga and Tanaka [59], Sznitman [60], Dawson and Zheng [18], Duffield and Werner [22], Duffield [21], Kipnis and Landim [38], Liggett [49], Le Boudecet at al. [43], Bordenave at al. [9], Gast and Bruno [31], Kolokoltsov at al. [41], Gast at al. [32], and Li [45].

During the last two decades the mean-field theory has been applied to studying some practical networks, such as, queue systems, computer networks, manufacturing systems and transportation networks. Readers may refer to, for example, Baccelli et al. [3], Vvedenskaya et al. [64], Vvedenskaya and Suhov [65], Mitzenmacher [53], Turner [62], Delcoigne and Fayolle [19], Karpelevich and Rybko [37], Oseledets and Khmelev [56], Bobbio et al. [8], Benaim and Le Boudec [7], Antunes et al. [2], Gast and Bruno [30], Hayden et al. [36], Fricker et al. [28], Baccelli et al. [4], Li et al. [46, 47], Li [48], and Fricker and Gast [27].

Nonlinear Markov processes play an important role in the study of big networks. Important examples include Rybko and Shlosman [58], Benaim and Le Boudec [7], Frank [25], Kolokoltsov [39], Gast and Bruno [31], Kolokoltsov [40], Kolokoltsov at al. [41], Muzychka and Vaninsky [54], Dupuis and Fischer [23], Gast at al. [32], Vaninsky et al. [63], Budhiraja et al. [14, 15], and Budhiraja and Majumder [16].

Metastability is an ubiquitous and important phenomenon of the dynamical behavior of communication networks, e.g., see Gibbens at al. [33], Antunes et al. [1, 2] and Tibi [61]. For metastability in Markov processes, readers may refer to Galves at al. [29], Bovier et al. [12, 13], Olivieri and Vares [55], Freidlin and Wentzell [26], Bovier [10, 11], den Hollander [20], and Beltran and Landim [5].

The main contributions of this paper are threefold. The first one is to set up a broad class of nonlinear continuous-time block-structured Markov processes when applying the mean-field theory to analyze a large number of interacting big networks with symmetric interactions, each of which is described as a continuous-time block-structured Markov process. The second one is to propose some effective algorithms for computing the fixed points of the nonlinear Markov processes by means of the UL-type RG-factorization, and show for some big networks that there possibly exist multiple fixed points, which lead to the metastability. The third one is to use the Birkhoff center, the Lyapunov functions and the relative entropy to analyze either stability or metastability of the big networks, and to give several interesting open problems with detailed interpretation. Based on this, this paper provides some new computational lines in the study of big networks. We believe that the results given in this paper can be useful and effective in performance evaluation and optimization of the big networks.

The remainder of this paper is organized as follows. In Section 2, we derive a class of nonlinear Markov processes through an asymptotic analysis of the weakly interacting big networks, in which each big network evolves as a continuous-time block-structured Markov process. In Section 3, we provide some effective algorithms for computing the fixed points of the system of ordinary differential equations. In Section 4, we discuss the Birkhoff center of the mean-field dynamic system, and apply the Lyapunov functions and the relative entropy to study the stability or metastability of the big network. Also, we provide several interesting open problems with detailed interpretation. Some concluding remarks are given in the final section.

2   Nonlinear Markov Processes

In this section, we derive a class of nonlinear Markov processes through an asymptotic analysis for a collection of weakly interacting big networks, in which each big network evolves as a continuous-time block-structured Markov process, which can be applied to deal with many practical stochastic systems.

To be able to discuss a system of big networks, we assume that any individual big network evolves as a continuous-time block-structured Markov process $\mathcal{X}$ whose infinitesimal

generator is given by

$$
Q = \begin{pmatrix}
Q_{0,0} & Q_{0,1} & Q_{0,2} & Q_{0,3} & \cdots \\
Q_{1,0} & Q_{1,1} & Q_{1,2} & Q_{1,3} & \cdots \\
Q_{2,0} & Q_{2,1} & Q_{2,2} & Q_{2,3} & \cdots \\
Q_{3,0} & Q_{3,1} & Q_{3,2} & Q_{3,3} & \cdots \\
\vdots & \vdots & \vdots & \vdots &
\end{pmatrix},
\tag{1}
$$

where the size of the matrix $Q_{j,j}$ is m_j for $j \geq 0$, and the sizes of other matrices can be determined accordingly. It is easy to see that the matrix $Q_{j,j}$ is also the infinitesimal generator of a continuous-time Markov process with m_j states for $j \geq 0$. We assume that the continuous-time Markov process Q is irreducible, aperiodic and positive recurrent, and its state space may be expressed as a two-dimensional structure: $\Omega = \{(k, j) : k \geq 0, 1 \leq j \leq m_k\}$. See Li [44] for more details.

From the continuous-time block-structured Markov chain $\mathcal{X}$, the system of N weakly interacting big networks is described as an $\mathcal{X}^N$-valued Markov process, where the states of the N big networks are denoted as $X^{1,N}(t)$, $X^{2,N}(t)$, ..., $X^{N,N}(t)$, respectively.

Let $\mathbf{X}^N(t) = \left(X^{1,N}(t), X^{2,N}(t), \ldots, X^{N,N}(t) \right)$. Then the empirical measure of the system of N big network system is given by

$$
\mu^N(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^{i,N}(t)},
\tag{2}
$$

where δ_x is the Dirac measure at x.

We denote by $\mathfrak{P}(\Omega)$ the space of probability vectors on the state space Ω, which is equipped with the usual topology of weak convergence. If $p \in \mathfrak{P}(\Omega)$, we write $p = (p_0, p_1, p_2, \ldots)$, where the size of the vector p_j is m_j for $j \geq 0$. At the same time, it is clear that $\mu^N(t) \in \mathfrak{P}(\Omega)$ is a random variable for $t \geq 0$, and $\{\mu^N(t) : t \geq 0\}$ is a continuous-time Markov process.

For the $\mathcal{X}^N$-valued continuous-time block-structured Markov process, we define that the probability distribution of $\mathbf{X}^N(t)$ is exchangeable, if for any level permutation $(k_{i_1}, k_{i_2}, \ldots, k_{i_N})$ of $(k_1, k_2, \ldots, k_N)$ and any phase permutation $(j_{i_1}, j_{i_2}, \ldots, j_{i_N})$ of $(j_1, j_2, \ldots, j_N)$,

$$
P\left\{ X^{1,N}(t) = (k_1, j_1), X^{2,N}(t) = (k_2, j_2), \ldots, X^{N,N}(t) = (k_N, j_N) \right\} \tag{3}
$$
$$
= P\left\{ X^{i_1,N}(t) = (k_{i_1}, j_{i_1}), X^{i_2,N}(t) = (k_{i_2}, j_{i_2}), \ldots, X^{i_N,N}(t) = (k_{i_N}, j_{i_N}) \right\}.
$$

In the system of N weakly interacting big networks, the effect of a typical big network on the dynamics of the given big network is of order $1/N$, and the jump intensity of any given big network depends on the configuration of other big networks only through the empirical measure $\mu^N(t)$. To study the system of N weakly interacting big networks in terms of Markov processes, it is seen from probability one that at most one big network will jump, i.e., change state, at a given time, and the jump intensities of any given big network depend only on its own state and the state of the empirical measure at that time. In addition, the jump intensities of the N big networks have the same functional form. Based on this, for the $\mathcal{X}^N$-valued Markov process, if the initial probability distribution of $\mathbf{X}^N(0)$ is exchangeable, then at any time $t \geq 0$, the probability distribution of $\mathbf{X}^N(t)$ is also exchangeable.

For the system of N weakly interacting big networks, if the probability distribution of $\mathbf{X}^N(t)$ is exchangeable, then the N big networks are indistinguishable, thus we apply the

mean-field theory to be able to analyze this system through only considering the Markov process of any given big network (such as, the first big network); while analysis of the total system will be completed by the propagation of Chaos (as $N \to \infty$). Based on this, the infinitesimal generator of the Markov process corresponding to the first big network can be defined as follows:

$$
\Gamma^{(N)}\left(\mu^N(t)\right) = \begin{pmatrix}
\Gamma_{0,0}^{(N)}\left(\mu^N(t)\right) & \Gamma_{0,1}^{(N)}\left(\mu^N(t)\right) & \Gamma_{0,2}^{(N)}\left(\mu^N(t)\right) & \Gamma_{0,3}^{(N)}\left(\mu^N(t)\right) & \cdots \\
\Gamma_{1,0}^{(N)}\left(\mu^N(t)\right) & \Gamma_{1,1}^{(N)}\left(\mu^N(t)\right) & \Gamma_{1,2}^{(N)}\left(\mu^N(t)\right) & \Gamma_{1,3}^{(N)}\left(\mu^N(t)\right) & \cdots \\
\Gamma_{2,0}^{(N)}\left(\mu^N(t)\right) & \Gamma_{2,1}^{(N)}\left(\mu^N(t)\right) & \Gamma_{2,2}^{(N)}\left(\mu^N(t)\right) & \Gamma_{2,3}^{(N)}\left(\mu^N(t)\right) & \cdots \\
\Gamma_{3,0}^{(N)}\left(\mu^N(t)\right) & \Gamma_{3,1}^{(N)}\left(\mu^N(t)\right) & \Gamma_{3,2}^{(N)}\left(\mu^N(t)\right) & \Gamma_{3,3}^{(N)}\left(\mu^N(t)\right) & \cdots \\
\vdots & \vdots & \vdots & \vdots
\end{pmatrix},
$$

$$(4)$$

where the size of the matrix $\Gamma_{j,j}^{(N)}\left(\mu^N(t)\right)$ is m_j for $j \geq 0$, and the sizes of other matrices can be determined (a.s.) accordingly. Since $\mu^N(t)$ is a random variable, it is clear that $\Gamma^{(N)}\left(\mu^N(t)\right)$ is a random matrix of infinite order. On the other hand, it is seen from the law of large number that the limit of the empirical measure $\mu^N(t)$ is deterministic under suitable conditions.

Let $\mu^N(t) \to p(t)$ and $\Gamma^{(N)}\left(\mu^N(t)\right) \to \Gamma(p(t))$ (a.s.) for $t \geq 0$, as $N \to \infty$. Then $p(t)$ is a probability vector, and using some probability analysis, we may obtain an infinite-dimensional dynamic system as follows:

$$
\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\,\Gamma(p(t))
\tag{5}
$$

with the initial condition

$$
p(0) = q.
\tag{6}
$$

Obviously, the mean-field dynamic system, given in (5) and (6), is related to a nonlinear Markov process whose infinitesimal generator is given by

$$
\Gamma(p(t)) = \begin{pmatrix}
\Gamma_{0,0}(p(t)) & \Gamma_{0,1}(p(t)) & \Gamma_{0,2}(p(t)) & \Gamma_{0,3}(p(t)) & \cdots \\
\Gamma_{1,0}(p(t)) & \Gamma_{1,1}(p(t)) & \Gamma_{1,2}(p(t)) & \Gamma_{1,3}(p(t)) & \cdots \\
\Gamma_{2,0}(p(t)) & \Gamma_{2,1}(p(t)) & \Gamma_{2,2}(p(t)) & \Gamma_{2,3}(p(t)) & \cdots \\
\Gamma_{3,0}(p(t)) & \Gamma_{3,1}(p(t)) & \Gamma_{3,2}(p(t)) & \Gamma_{3,3}(p(t)) & \cdots \\
\vdots & \vdots & \vdots & \vdots
\end{pmatrix}.
\tag{7}
$$

Remark: To establish the infinitesimal generator $\Gamma(p(t))$ of a nonlinear Markov process, readers may also refer to some recent publications, for example, the discrete-time Markov chains by Benaim and Le Boudec [7] and Budhiraja and Majumder [16], the Markov decision processes by Gast and Bruno [31] and Gast at al. [32], the continuous-time Markov chains by Dupuis and Fischer [23] and Budhiraja et al. [14, 15], and some nice practical examples include Mitzenmacher [53], Bobbio et al. [8], Li et al. [46, 47], and Li and Lui [48].

In what follows, it is necessary to provide some useful interpretation or proofs for how to establish the mean-field dynamic system (5) and (6).

(a) Existence and Uniqueness

Consider the infinite-dimensional ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\,\Gamma(p(t))$ with $p(0) = q$. A solution in the classical sense is a (continuously) differential function $p(t)$ such that $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\,\Gamma(p(t))$ with $p(0) = q$. A classical result is the Picard

approximation as follows. If $\Gamma(x)$ is (locally) Lipschitz on a set $E \subseteq \mathfrak{P}(\Omega)$, that is, there exists a positive constant C such that

$$\|\Gamma(x) - \Gamma(y)\| \leq C \|x - y\|, \quad x, y \in \mathfrak{P}(\Omega),$$

and $p(0) = q$ is in the interior of E, then there exists a unique global solution to the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t) \Gamma(p(t))$ with $p(0) = q$, within E.

To deduce whether the $\Gamma(x)$ is (locally) Lipschitz on a set $E \subseteq \mathfrak{P}(\Omega)$, Li et al. [46] and Li and Lui [48] gave an algorithmic method through dealing with some matrices of infinite orders.

(b) The limiting processes

To discuss the limit: $\mu^N(t) \to p(t)$ (a.s.) for $t \geq 0$, as $N \to \infty$, we need to set up some suitable conditions in order to guarantee the existence of such a limit.

Let $e_{k,j}$ be the unit vector of infinite dimension in which the (k,j)th entry is one and all the others are zero. Note that the empirical measure process $\mu^{(N)} = \{\mu^N(t) : t \geq 0\}$ is a Markov process on the state space $\mathfrak{P}_N(\Omega)$ where $\mathfrak{P}_N(\Omega) = \mathfrak{P}(\Omega) \cap \left(\frac{1}{N}\Omega\right)$, the possible jumps of $\mu^{(N)}$ are of the form $(e_{k,j} - e_{l,i})/N$ for $(k,j) \neq (l,i)$, and $(k,j), (l,i) \in \Omega$. If $\mu^N(t) = x \in \mathfrak{P}_N(\Omega)$, then at time $t \geq 0$, $Nx_{k,j}$ of the big network is in State (k,j). Hence the transition rate of the Markov process corresponding to the given big network is given by $Nx_{k,j}\Gamma_{k,j;l,i}^{(N)}(x)$. Based on this, the generator $\mathbf{A}^{(N)}$ of the Markov process $\mu^{(N)}$ is given by

$$\mathbf{A}^{(N)} f(x) = \sum_{\substack{(l,i) \in \Omega \\ (l,i) \neq (k,j)}} Nx_{k,j}\Gamma_{k,j;l,i}^{(N)}(x) \left[f\left(x + \frac{1}{N}(e_{k,j} - e_{l,i})\right) - f(x) \right],$$

where $f(x)$ is a real function on $\mathfrak{P}_N(\Omega)$. It is easy to see that as $N \to \infty$

$$\mathbf{A}^{(N)} f(x) \to \sum_{\substack{(l,i) \in \Omega \\ (l,i) \neq (k,j)}} x_{k,j}\Gamma_{k,j;l,i}(x) \left[\frac{\partial}{\partial x_{k,j}} f(x) - \frac{\partial}{\partial x_{l,i}} f(x) \right] \overset{\text{def}}{=} \mathbf{A} f(x).$$

Theorem 1. *Suppose that for $(k,j), (l,i) \in \Omega$ with $(k,j) \neq (l,i)$, there exists a Lipschitz continuous function $\Gamma_{k,j;l,i}(p) : \mathfrak{P}(\Omega) \to [0, +\infty)$ such that $\Gamma_{k,j;l,i}^{(N)}(p) \to \Gamma_{k,j;l,i}(p)$ uniformly on $\mathfrak{P}(\Omega)$. If $\{\mu^{(N)}(0)\}$ converges in probability to $q \in \mathfrak{P}(\Omega)$, then $\{\mu^{(N)}(t)\}$ converges uniformly on compact time intervals in probability to $p(t) \in \mathfrak{P}(\Omega)$ for $t \geq 0$, where the probability vector $p(t)$ is the unique global solution to the ordinary differential equation: $\frac{d}{dt}p(t) = p(t)\Gamma(p(t))$ with $p(0) = q$.*

Proof: The proof may directly follow from Theorem 2.11 in Kurtz [42]. Here, we only give a simple interpretation as follows. Firstly, we notice that

$$F^{(N)}(p) = \left(\sum_{k=0}^{\infty}\sum_{j=1}^{m_k}\right)\left(\sum_{l=0}^{\infty}\sum_{i=1}^{m_l}\right) Np_{k,j}\left(\frac{1}{N}e_{l,i} - \frac{1}{N}e_{k,j}\right)\Gamma_{k,j;l,i}^{(N)}(p)$$

and

$$F(p) = \left(\sum_{k=0}^{\infty}\sum_{j=1}^{m_k}\right)\left(\sum_{l=0}^{\infty}\sum_{i=1}^{m_l}\right) p_{k,j}(e_{l,i} - e_{k,j})\Gamma_{k,j;l,i}(p),$$

where as $N \to \infty$

$$\Gamma_{k,j;l,i}^{(N)}(p) \to \Gamma_{k,j;l,i}(p),$$

and $\Gamma(x)$ is (locally) Lipschitz on a set $E \subseteq \mathfrak{P}(\Omega)$. Then, for the sequence $\{\mu^{(N)}(t), t \geq 0\}$ of Markov processes, it follows from Equation (III.10.13) in Rogers and Williams [57] or page 162 in Ethier and Kurtz [24] that

$$\mathbf{M}^{(N)}(t) = \mu^{(N)}(t) - \mu^{(N)}(0) - \int_0^t \left\{ \mu^{(N)}(x) \Gamma\left(\mu^{(N)}(x)\right) \right\} \mathrm{d}x$$

is a martingale with respect to $N \geq 1$. Therefore, if $\{\mu^{(N)}(0)\}$ converges weakly to $q \in \mathfrak{P}(\Omega)$ as $N \to \infty$, then $\{\mu^{(N)}(t), N \geq 1\}$ converges weakly in $D_{\mathcal{F}}[0, +\infty)$ endowed with the Skorohod topology to the solution $p(t)$ to the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\Gamma(p(t))$ with $p(0) = q$, within $\mathfrak{P}(\Omega)$. This completes the proof. ∎

3   The Fixed Points

In this section, we use the UL-type RG-factorization to provide some effective algorithms for computing the fixed points of the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\Gamma(p(t))$ with $p(0) = q$. Further, we set up a nonlinear characteristic equation of the censoring matrix to level 0, which is satisfied by the fixed points.

A point $\pi \in \mathfrak{P}(\Omega)$ is said to be a fixed point of the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p(t) = p(t)\Gamma(p(t))$ with $p(0) = q$, if $p(t) \to \pi$ as $t \to +\infty$, and

$$\lim_{t \to +\infty} \left[\frac{\mathrm{d}}{\mathrm{d}t} p(t) \right] = 0.$$

In this case, it is clear that

$$\pi \Gamma(\pi) = 0, \tag{8}$$

which is an infinite-dimensional system of nonlinear equations. In general, there exist more difficulties and challenging due to its infinite dimensions when solving the fixed point equation (8) together with $\pi e = 1$, where e is a column vector of ones with a suitable size.

It is easy to check that for every $\pi \in \mathfrak{P}(\Omega)$, $\Gamma(\pi)$ is the infinitesimal generator of an irreducible continuous-time Markov process. Based on Li [44], we can develop the UL-type RG-factorization of the matrix $\Gamma(\pi)$. To that end, we partition the matrix $\Gamma(\pi)$ as

$$\Gamma(\pi) = \begin{pmatrix} T(\pi) & U(\pi) \\ V(\pi) & W(\pi) \end{pmatrix}$$

according to the level sets $L_{\leq n}$ and $L_{\geq n+1}$ for $n \geq 0$. Since the Markov chain $\Gamma(\pi)$ is irreducible, it is clear that the two truncated chains with infinitesimal generators $T(\pi)$ and $W(\pi)$ are all transient, and the matrices $T(\pi)$ and $W(\pi)$ are all invertible. Note that the inverse of the matrix $T(\pi)$ is ordinary, but the invertibility of the matrix $W(\pi)$ is different under an infinite-dimensional meaning. Although the matrix $W(\pi)$ of infinite size may have multiple inverses, we in general are interested in the maximal non-positive inverse $W_{\max}^{-1}(\pi)$ of $W(\pi)$, i.e., $W^{-1}(\pi) \leq W_{\max}^{-1}(\pi) \leq 0$ for every non-positive inverse $W^{-1}(\pi)$ of $W(\pi)$. Of course, $0 \leq [-W(\pi)]_{\min}^{-1} \leq [-W(\pi)]^{-1}$ for every nonnegative inverse $[-W(\pi)]^{-1}$ of $-W(\pi)$, that is, $[-W(\pi)]_{\min}^{-1}$ is the minimal nonnegative inverse of $-W(\pi)$. Based on this, for $n \geq 0$ we write

$$\Gamma^{[\leq n]}(\pi) = T(\pi) + U(\pi)[-W(\pi)]_{\min}^{-1} V(\pi) = \begin{pmatrix} \phi_{0,0}^{(n)}(\pi) & \phi_{0,1}^{(n)}(\pi) & \cdots & \phi_{0,n}^{(n)}(\pi) \\ \phi_{1,0}^{(n)}(\pi) & \phi_{1,1}^{(n)}(\pi) & \cdots & \phi_{1,n}^{(n)}(\pi) \\ \vdots & \vdots & & \vdots \\ \phi_{n,0}^{(n)}(\pi) & \phi_{n,1}^{(n)}(\pi) & \cdots & \phi_{n,n}^{(n)}(\pi) \end{pmatrix},$$

where the size of the matrix $\phi_{j,j}^{(n)}(\pi)$ is m_j for $0 \leq j \leq n$, and the sizes of other matrices can be determined accordingly. It is clear from Section 7 of Chapter 2 in Li [44] that for $n \geq 0$, $0 \leq i, j \leq n$,

$$\phi_{i,j}^{(n)}(\pi) = \Gamma_{i,j}(\pi) + \sum_{k=n+1}^{\infty} \phi_{i,k}^{(k)}(\pi) \left[-\phi_{k,k}^{(k)}(\pi)\right]^{-1} \phi_{k,j}^{(k)}(\pi).$$

Let

$$\Psi_n(\pi) = \phi_{n,n}^{(n)}(\pi), \quad n \geq 0;$$

$$R_{i,j}(\pi) = \phi_{i,j}^{(j)}(\pi) \left[-\phi_{j,j}^{(k)}(\pi)\right]^{-1}, \quad 0 \leq i < j;$$

and

$$G_{i,j}(\pi) = \left[-\phi_{i,i}^{(k)}(\pi)\right]^{-1} \phi_{i,j}^{(i)}(\pi), \quad 0 \leq j < i.$$

Then the UL-type RG-factorization of the matrix $\Gamma(\pi)$ is given by

$$\Gamma(\pi) = [I - R_U(\pi)] \Psi_D(\pi) [I - G_L(\pi)], \tag{9}$$

where

$$R_U(\pi) = \begin{pmatrix} 0 & R_{0,1}(\pi) & R_{0,2}(\pi) & R_{0,3}(\pi) & \cdots \\ & 0 & R_{1,2}(\pi) & R_{1,3}(\pi) & \cdots \\ & & 0 & R_{2,3}(\pi) & \cdots \\ & & & 0 & \cdots \\ & & & & \ddots \end{pmatrix},$$

$$\Psi_D(\pi) = \text{diag}(\Psi_0(\pi), \Psi_1(\pi), \Psi_2(\pi), \Psi_3(\pi), \ldots)$$

and

$$G_L(\pi) = \begin{pmatrix} 0 & & & & \\ G_{1,0}(\pi) & 0 & & & \\ G_{2,0}(\pi) & G_{2,1}(\pi) & 0 & & \\ G_{3,0}(\pi) & G_{3,1}(\pi) & G_{3,2}(\pi) & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Based on the UL-type RG-factorization (9), it follows from Subsection 2.7.3 in Li [44] that the fixed point π is given by

$$\begin{cases} \pi_0 = \tau x_0(\pi), \\ \pi_k = \sum_{i=0}^{k-1} \pi_i R_{i,k}(\pi), \quad k \geq 1, \end{cases} \tag{10}$$

where $x_0(\pi)$ is the fixed point of the censored Markov chain $\Psi_0(\pi)$ to level 0, and the scalar τ is determined by $\sum_{k=0}^{\infty} \pi_k e = 1$ uniquely.

Using the expression (10) of the fixed point π, we set up an important relation as follows:

$$\pi = \left(\tau x_0(\pi), \pi_0 R_{0,1}(\pi), \sum_{i=0}^{1} \pi_i R_{i,k}(\pi), \sum_{i=0}^{2} \pi_i R_{i,k}(\pi), \ldots\right). \tag{11}$$

In what follows we consider two special cases in order to further explain the fixed point equation (11) with R-measure.

Case one: Nonlinear Markov processes of GI/M/1 type

In this case, the infinitesimal generator $\Gamma(\pi)$ is given by

$$\Gamma(\pi) = \begin{pmatrix} B_1(\pi) & B_0(\pi) & & & \\ B_2(\pi) & A_1(\pi) & A_0(\pi) & & \\ B_3(\pi) & A_2(\pi) & A_1(\pi) & A_0(\pi) & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Let $R(\pi)$ be the minimal nonnegative solution to the nonlinear matrix equation

$$\sum_{k=0}^{\infty} R^k(\pi) A_k(\pi) = 0.$$

Then

$$\pi_k = \pi_1 R^{k-1}(\pi), \quad k \geq 1,$$

where the two vectors π_0 and π_1 satisfy the following system of nonlinear matrix equations

$$(\pi_0, \pi_1) \begin{pmatrix} B_1(\pi) & B_0(\pi) \\ \sum_{k=0}^{\infty} R^k(\pi) B_{k+2}(\pi) & \sum_{k=0}^{\infty} R^k(\pi) A_{k+1}(\pi) \end{pmatrix} = 0$$

and

$$\pi_0 e + \pi_1 [I - R(\pi)]^{-1} e = 1.$$

Thus, the fixed point equation (11) with R-measure is simplified as

$$\pi = \left(\pi_0, \pi_1, \pi_1 R(\pi), \pi_1 R^2(\pi), \ldots\right).$$

Case two: Nonlinear Markov processes of M/G/1 type

In this case, the infinitesimal generator $\Gamma(\pi)$ is given by

$$\Gamma(\pi) = \begin{pmatrix} B_1(\pi) & B_2(\pi) & B_3(\pi) & B_4(\pi) & \cdots \\ B_0(\pi) & A_1(\pi) & A_2(\pi) & A_3(\pi) & \cdots \\ & A_0(\pi) & A_1(\pi) & A_2(\pi) & \cdots \\ & & A_0(\pi) & A_1(\pi) & \cdots \\ & & & \ddots & \ddots \end{pmatrix}.$$

Let $G(\pi)$ be the minimal nonnegative solution to the nonlinear matrix equation

$$\sum_{k=0}^{\infty} A_k(\pi) G^k(\pi) = 0.$$

Then

$$\Psi_0(\pi) = B_1(\pi) + \sum_{k=2}^{\infty} B_k(\pi) G^{k-2}(\pi) G_1(\pi)$$

and for $k \geq 1$

$$\Psi(\pi) = A_1(\pi) + \sum_{k=2}^{\infty} A_k(\pi) G^{k-1}(\pi);$$

and the R-measure

$$R_{0,j}(\pi) = \left[\sum_{k=j+1}^{\infty} B_k(\pi) G^{k-1}(\pi)\right] [-\Psi(\pi)]^{-1}, \quad j \geq 1,$$

and for $i \geq 1$

$$R_j (\pi) = \left[\sum_{k=j+1}^{\infty} A_k (\pi) G^{k-1} (\pi) \right] [-\Psi (\pi)]^{-1}, \quad j \geq 1.$$

The fixed point π is given by

$$\begin{cases} \pi_0 = \tau x_0 (\pi), \\ \pi_k = \pi_0 R_{0,k} (\pi) + \sum_{i=1}^{k-1} \pi_i R_{k-i} (\pi), \quad k \geq 1, \end{cases}$$

where $x_0 (\pi)$ is the fixed point of the censored Markov chain $\Psi_0 (\pi)$ to level 0 and the scalar τ is determined by $\sum_{k=0}^{\infty} \pi_k e = 1$ uniquely. Thus, the fixed point equation (11) with R-measure is simplified as

$$\pi = \left(\tau x_0 (\pi), \pi_0 R_{0,1} (\pi), \pi_0 R_{0,2} (\pi) + \pi_1 R_1 (\pi), \pi_0 R_{0,3} (\pi) + \sum_{i=1}^{2} \pi_i R_{k-i} (\pi), \ldots \right).$$

Now, we write the fixed point equation (11) with R-measure as a functional form: $\pi = \mathbf{F} (\mathfrak{R} (\pi))$, as shown in the above two special cases. Based on this, we can provide an approximative algorithm as follows:

Algorithm I: Computation of the fixed points

Step one: Taking any initial probability vector: $\pi^{(0)} \in \mathfrak{P} (\Omega)$.

Step two: Computing the infinitesimal generator: $\Gamma (\pi^{(0)})$; and then compute the R-measure, which gives $\pi^{(1)} = \mathbf{F} (\mathfrak{R} (\pi^{(0)}))$.

Step three: For $N \geq 2$, compute $\pi^{(N+1)} = \mathbf{F} (\mathfrak{R} (\pi^{(N)}))$.

Step four: For a sufficiently small $\varepsilon > 0$, if $\left\| \pi^{(N+1)} - \pi^{(N)} \right\| < \varepsilon$, then the computation is over; otherwise we go to Step three.

Note that it is possible for some big networks that there exist multiple fixed points because the infinitesimal generator $\Gamma (\pi)$ is more general. In this case, it is a key to design a suitable initial probability vector: $\pi^{(0)} \in \mathfrak{P} (\Omega)$, for example, for any integer $m \geq 1$ we take

$$\pi^{(0)} = \left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}, 0, \ldots \right).$$

Now, we provide another algorithm for computing the fixed points. To do end, we set up a characteristic equation of the censoring matrix $\Psi_0 (\pi)$ to level 0, while the characteristic equation is satisfied by the fixed points.

Note that for the censored Markov chain $\Psi_0 (\pi)$ to level 0, we have

$$\pi_0 \Psi_0 (\pi) = 0, \quad \pi_0 e = \tau \in (0, 1).$$

Thus it is easy to see from the irreducibility of the matrix $\Gamma (\pi)$ that for the matrix $\Psi_0 (\pi)$ of size m_0, rank$(\Psi_0 (\pi)) = m_0 - 1$ according to the irreducibility of the matrix $\Psi_0 (\pi)$, and its eigenvalue with the maximal real part is equal to zero. Let the characteristic equation be $f_x (\pi) = \det (xI - \Psi_0 (\pi)) = 0$. Then the fixed points satisfy the characteristic equation $f_0 (\pi) = \det (\Psi_0 (\pi)) = 0$. Hence the fixed points satisfy the system of nonlinear equations as follows:

$$\begin{cases} \det (\Psi_0 (\pi)) = 0, \\ \text{rank} (\Psi_0 (\pi)) = m_0 - 1. \end{cases} \quad (12)$$

Note that (12) provide another algorithm for computing the fixed points as follows:

Algorithm II: Computation of the fixed points

Step one: Providing a numerical solution $\widehat{\pi}$ to the nonlinear characteristic equation: $\det\left(\Psi_0\left(\pi\right)\right) = 0$.

Step two: Check whether $\mathrm{rank}(\Psi_0\left(\widehat{\pi}\right)) = m_0 - 1$. If Yes, then $\widehat{\pi}$ is a fixed point. If No, then going to Step one.

4 Stability and Metastability

In this section, we first discuss the Birkhoff center of the mean-field dynamic system: $\frac{\mathrm{d}}{\mathrm{d}t}p\left(t\right) = p\left(t\right)\Gamma\left(p\left(t\right)\right)$ with $p\left(0\right) = q$. Then we apply the Lyapunov functions and the relative entropy to study the stability or metastability of the big networks. Furthermore, we provide several interesting open problems with detailed interpretation.

We write
$$\mathbb{S}_\pi = \left\{\pi : \pi\Gamma\left(\pi\right) = 0,\ \pi e{=}1\right\}.$$

Then it is clear that
$$\mathbb{S}_\pi = \left\{\pi : \pi = \left(\tau x_0\left(\pi\right), \pi_0 R_{0,1}\left(\pi\right), \sum_{i=0}^{1}\pi_i R_{i,k}\left(\pi\right), \sum_{i=0}^{2}\pi_i R_{i,k}\left(\pi\right), \ldots\right),\ \pi e{=}1\right\}$$

or
$$\mathbb{S}_\pi = \left\{\pi : \det\left(\Psi_0\left(\pi\right)\right) = 0, \mathrm{rank}\left(\Psi_0\left(\pi\right)\right) = m_0 - 1,\ \pi e{=}1\right\}$$

with $\pi_k = \sum_{i=0}^{k-1}\pi_i R_{i,k}\left(\pi\right)$, $k \geq 1$.

Since the vector equation $\pi\Gamma\left(\pi\right) = 0$, together with $\pi e{=}1$, is nonlinear, it is possible for some big networks that there exist multiple elements in the set $\mathbb{S}_\pi$. At the same time, an argument by analytic function can indicate that the elements of the set $\mathbb{S}_\pi$ are isolated.

To describe the isolated element structure of the set $\mathbb{S}_\pi$, we often need to use the Birkhoff center of the mean-field dynamic system, and use the Birkhoff center to check whether the fixed point is unique or not. Based on this, our discussion includes the following two cases:

Case one: $N \to \infty$. In this case, we denote by $\Phi\left(t\right)$ a solution to the system of differential equations $\frac{\mathrm{d}}{\mathrm{d}t}p\left(t\right) = p\left(t\right)\Gamma\left(p\left(t\right)\right)$ with $p\left(0\right) = q$. Thus, the Birkhoff center of the solution $\Phi\left(t\right)$ is defined as

$$\boldsymbol{\Theta} = \Big\{\overline{P} \in \mathfrak{P}\left(\Omega\right) : \overline{P} = \lim_{k\to\infty}\Phi\left(t_k\right)\ \text{for any scale sequence}$$
$$\{t_k\}\ \text{with}\ t_l \geq 0\ \text{for}\ l \geq 1\ \text{and}\ \lim_{k\to\infty}t_k = +\infty\Big\}.$$

Notice that perhaps $\boldsymbol{\Theta}$ contains the limit cycles or the equilibrium points (the local minimal points, or the local maximal points, or the saddle points). Thus it is clear that $\mathbb{S}_\pi \subset \boldsymbol{\Theta}$. Obviously, the limiting empirical Markov process $\left\{\mathbf{Y}\left(t\right) : t \geq 0\right\}$ spends most of its time in the Birkhoff center $\boldsymbol{\Theta}$, where $\mathbf{Y}\left(t\right) = \lim_{N\to\infty}\mu^N\left(t\right)$ a.s..

Case two: $t \to +\infty$. In this case, we write
$$\pi^{(N)} = \lim_{t\to+\infty}\mu^N\left(t\right),\quad \text{a.s.,}$$

since for each $N = 1, 2, 3, \ldots$, if the system of N weakly interacting big networks is stable.

Let

$$\Xi = \left\{ \overline{\pi} \in \mathfrak{P}\left(\Omega\right) : \overline{\pi} = \lim_{k \to \infty} \pi^{(N_k)} \text{ for any positive integer sequence} \right.$$

$$\left. \{N_k\} \text{ with } 1 \leq N_1 \leq N_2 \leq N_3 \leq \cdots \text{ and } \lim_{k \to \infty} N_k = \infty \right\}.$$

It is easy to see that

$$\mathbb{S}_\pi \subset \Xi \subset \boldsymbol{\Theta}.$$

At the same time, we have

$$\mathbb{S}_\pi = \{\text{the local minimal points in } \boldsymbol{\Theta}\}$$

and

$$\boldsymbol{\Theta} - \mathbb{S}_\pi = \{\text{the limit cycles in } \boldsymbol{\Theta}\} \cup \{\text{the local maximal points, or the saddle points in } \boldsymbol{\Theta}\}.$$

In what follows, we discuss stability and metastability of the big networks.

For the metastability in $\mathbb{S}_\pi$, a key is to determine a Lyapunov function for the mean-field dynamic system: $\frac{\mathrm{d}}{\mathrm{d}t} p\left(t\right) = p\left(t\right) \Gamma\left(p\left(t\right)\right)$ with $p\left(0\right) = q$. The Lyapunov function g defined on $\mathfrak{P}\left(\Omega\right)$ is constructed such that

$$y\Gamma\left(y\right) \cdot \nabla g\left(y\right) \leq 0, \quad y \in \mathfrak{P}\left(\Omega\right). \tag{13}$$

It is easy to see that if $\pi \in \mathbb{S}_\pi$, then $\pi\Gamma\left(\pi\right) \cdot \nabla g\left(\pi\right) = 0$ due to the fact that $\pi\Gamma\left(\pi\right) = 0$. On the other hand, if $\pi\Gamma\left(\pi\right) \cdot \nabla g\left(\pi\right) = 0$, then $\pi \in \mathbb{S}_\pi$.

Let $|\mathbb{S}_\pi|$ be the number of elements in the set $\mathbb{S}_\pi$. If $|\mathbb{S}_\pi| = 1$, then

$$\lim_{N \to \infty} \lim_{t \to +\infty} \mu^N\left(t\right) = \lim_{t \to +\infty} \lim_{N \to \infty} \mu^N\left(t\right) = \pi, \quad \text{a.s.}.$$

If $|\mathbb{S}_\pi| \geq 2$, then the system of big networks exhibits a metastability property, that is, the state of the given big network switches from one stable point to the other after a long residence time. In the study of metastability, it is a key to estimate the expected value of such a residence time. See Bovier [11] and Olivieri and Vares [55] for more details.

An interesting issue in the study of big networks is to analyze stability or metastability of the corresponding nonlinear Markov processes. On this line, it is a key to construct a Lyapunov function or a local Lyapunov function. Note that the relative entropy function in some sense can define a globally attracting Lyapunov function.

For $p, q \in \mathfrak{P}\left(\Omega\right)$, we define the relative entropy of p with respect to q as

$$R\left(p||q\right) = \sum_{x \in \Omega} p_x \log\left(\frac{p_x}{q_x}\right).$$

Let $\Psi\left(z\right) = z \log z - z + 1$. Then if $p\left(t\right)$ and $q\left(t\right)$ are two different solutions to the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p\left(t\right) = p\left(t\right) \Lambda,$$

where Λ is the infinitesimal generator of an irreducible continuous-time Markov process. In this case, Dupuis and Fischer [23] indicated that

$$\frac{\mathrm{d}}{\mathrm{d}t} R\left(p\left(t\right)||q\left(t\right)\right) = -\sum_{\substack{x,y \in \Omega \\ x \neq y}} \Psi\left(\frac{p_y\left(t\right) q_x\left(t\right)}{p_x\left(t\right) q_y\left(t\right)}\right) p_x\left(t\right) \frac{q_y\left(t\right)}{q_x\left(t\right)} \Lambda_{y,x} \leq 0, \tag{14}$$

and $\frac{\mathrm{d}}{\mathrm{d}t} R\left(p\left(t\right) || q\left(t\right)\right) = 0$ if and only if $p\left(t\right) = q\left(t\right)$ for $t \geq 0$. Obviously, $\frac{\mathrm{d}}{\mathrm{d}t} R\left(p\left(t\right) || \pi\right) = 0$ if and only if $p\left(t\right) = \pi$ for $t \geq 0$.

Dupuis and Fischer [23] further demonstrated that for the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p\left(t\right) = p\left(t\right) \Gamma\left(p\left(t\right)\right)$, the the relative entropy relation (14) can not be applied directly. In this case, they first defined $\mathbf{P}^{(N)}\left(t\right)$ as the state probability of the system of N big networks at time $t \geq 0$, and let $\mathbf{P}^{(N)}\left(0\right) = \otimes^N q$. Then they gave an approximate method to construct the Lyapunov function as follows:

$$F\left(q\right) = \lim_{N \to \infty} \lim_{T \to +\infty} \frac{1}{N} R\left(\mathbf{P}^{(N)}\left(0\right) || \mathbf{P}^{(N)}\left(T\right)\right) = \lim_{N \to \infty} \frac{1}{N} R\left(\otimes^N q || \otimes^N \pi\right).$$

For applying the relative entropy to construct a Lyapunov function, readers may also refer to Budhiraja et al. [14, 15] for more details.

In the remainder of this section, we provide several interesting open problems with detailed interpretation.

Open problem one: The mean drift condition.

We consider an irreducible QBD process whose infinitesimal generator is given by

$$\Gamma\left(p\right) = \left(\begin{array}{ccccc} B_1\left(p\right) & B_0\left(p\right) & & & \\ B_2\left(p\right) & A_1\left(p\right) & A_0\left(p\right) & & \\ & A_2\left(p\right) & A_1\left(p\right) & A_0\left(p\right) & \\ & & \ddots & \ddots & \ddots \end{array}\right),$$

where $\Gamma\left(p\right) e = 0$, the sizes of the matrices $B_1\left(p\right)$ and $A_1\left(p\right)$ are m_0 and m, respectively, and the sizes of other matrices can be determined accordingly. We assume that for any $p \in \mathfrak{P}\left(\Omega\right)$, the Markov process: $A\left(p\right) = A_0\left(p\right) + A_1\left(p\right) + A_2\left(p\right)$, is irreducible, aperiodic and positive recurrent. Let θ_p be the stationary probability vector of the the Markov process $A\left(p\right)$. Then it is clear that for for any $p \in \mathfrak{P}\left(\Omega\right)$, the Markov process $\Gamma\left(p\right)$ is positive recurrent if and only if $\theta_p A_2\left(p\right) e > \theta_p A_0\left(p\right) e$.

It is interesting to study how the mean drift condition: $\theta_p A_2\left(p\right) e > \theta_p A_0\left(p\right) e$ for any $p \in \mathfrak{P}\left(\Omega\right)$, can influence stability or metastability of the ordinary differential equation: $\frac{\mathrm{d}}{\mathrm{d}t} p\left(t\right) = p\left(t\right) \Gamma\left(p\left(t\right)\right)$.

Open problem two: The censoring Markov processes.

For the infinitesimal generator $\Gamma\left(p\right)$ given in (7), it is easy to give the infinitesimal generator $\Psi_0\left(p\right)$ of the censoring Markov processes to level 0. It is very interesting (but difficult) to set up some useful relations of stability or metastability between two ordinary differential equations: $\frac{\mathrm{d}}{\mathrm{d}t} p\left(t\right) = p\left(t\right) \Gamma\left(p\left(t\right)\right)$ and $\frac{\mathrm{d}}{\mathrm{d}t} p_0\left(t\right) = p_0\left(t\right) \Psi_0\left(p\left(t\right)\right)$.

5　Concluding Remarks

This paper sets up a broad class of nonlinear continuous-time block-structured Markov processes by means of applying the mean-field theory to the study of big networks, and proposes some effective algorithms for computing the fixed points of the nonlinear Markov process by means of the UL-type *RG*-factorization. Furthermore, this paper considers stability or metastability of the big network, and gives several interesting open problems with detailed interpretation. Along such a line, there are a number of interesting directions for potential future research, for example:

- providing algorithms for computing the fixed points of big networks with multiple stable points;

- studying the influence of the censoring Markov processes on the metastability;
- discussing how to apply the *RG*-factorizations given in Li [44] to compute the expected residence times in the study of metastability; and
- analyzing some big networks with a heterogeneous geographical environment, and set up their simultaneous systems of nonlinear Markov processes.

Acknowledgement

References

[1] Antunes, N., Fricker, C., Robert, P. and Tibi, D. (2006). Metastability of CDMA cellular systems. In: *Proceedings of the 12th annual international ACM conference on Mobile computing and networking*, Pages 206–214.

[2] Antunes, N., Fricker, C., Robert, P. and Tibi, D. (2008). Stochastic networks with multiple stable points, *The Annals of Probability*, Vol. 36, 255–278.

[3] Baccelli, F., Karpelevich, F.I., Kelbert, M.Y., Puhalskii, A.A., Rybko, A.N. and Suhov, Y.M. (1992). A mean-field limit for a class of queueing networks, *Journal of Statistical Physics*, Vol. 66, 803–825.

[4] Baccelli, F., Rybko, A.N. and Shlosman, S. (2013). Queuing networks with varying topology – a mean-field approach. arXiv preprint: arXiv:1311.3898.

[5] Beltran, J. and Landim, C. (2010). Tunneling and metastability of continuous time Markov chains, *Journal of Statistical Physics*, Vol. 140, 1065–1114.

[6] Beltran, J. and Landim, C. (2012). Tunneling and metastability of continuous time Markov chains II, the nonreversible case, *Journal of Statistical Physics*, Vol. 149, 598–618.

[7] Benaim, M. and Le Boudec, J.Y. (2008). A class of mean-field interaction models for computer and communication systems, *Performance Evalution*, Vol. 65, 823–838.

[8] Bobbio, A., Gribaudo, M. and Telek, M. (2008). Analysis of large scale interacting systems by mean field method. In: *The Fifth International IEEE Conference on Quantitative Evaluation of Systems*, Pages 215–224.

[9] Bordenave, C., McDonald, D. and Proutiere, A. (2010). A particle system in interaction with a rapidly varying environment: Mean-field limits and applications, *Networks and Heterogeneous Media*, Vol. 5, 31–62.

[10] Bovier, A. (2003). *Markov processes and metastability*, Lecture Notes TUB, 1–75.

[11] Bovier, A. (2006). Metastability: A potential theoretic approach. In: *Proceedings oh the International Congress of Mathematicians: Invited lectures*, Pages 499–518.

[12] Bovier, A., Eckhoff, M., Gayrard, V. and Klein, M. (2001). Metastability in stochastic dynamics of disordered mean-field models, *Probability Theory and Related Fields*, Vol. 119, 99–161.

[13] Bovier, A., Eckhoff, M., Gayrard, V. and Klein, M. (2002). Metastability and low lying spectral in reversible Markov chains, *Communications in Mathematical Physics*, Vol. 228, 219–255.

[14] Budhiraja, A., Dupuis, P., Fischer, M. and Ramanan, K. (2014). Local stability of Kolmogorov forward equations for finite state nonlinear Markov processes. arXiv preprint: arXiv:1412.5555.

[15] Budhiraja, A., Dupuis, P., Fischer, M. and Ramanan, K. (2014). Limits of relative entropies associated with weakly interacting particle systems. arXiv preprint: arXiv:1412.5553.

[16] Budhiraja, A. and Majumder, A.P. (2014). Long time results for a weakly interacting particle system in discrete time. arXiv preprint: arXiv:1401.3423.

[17] Dawson, D.A. (1983). Critical dynamics and fluctuations for a mean-field model of cooperative behavior, *Journal of Statistical Physics*, Vol. 31, 29–85.

[18] Dawson, D.A. and Zheng, X. (1991). Law of large numbers and central limit theorem for unbounded jump mean-field models, *Advances in Applied Mathematics*, Vol. 12, 293–326.

[19] Delcoigne, F. and Fayolle, G. (1999). Thermodynamical limit and propagation of chaos in polling systems, *Markov Processes and Related Fields*, Vol. 5, 89–124.

[20] den Hollander, F. (2004). Metastability under stochastic dynamics, *Stochastic Processes and their Applications*, Vol. 114, 1–26.

[21] Duffield, N.G. (1992). Local mean-field Markov processes: An application to message-switching networks, *Probability theory and related fields*, Vol. 93, 485–505.

[22] Duffield, N.G. and Werner, R.F. (1991). Local dynamics of mean-field quantum systems, *Helvetica Physica Acta*, Vol. 65, 1016–1054.

[23] Dupuis, P. and Fischer, M. (2011). On the construction of Lyapunov functions for nonlinear Markov processes via relative entropy, submitting for publication.

[24] Ethier, S.N. and Kurtz, T.G. (1986). *Markov processes: Characterization and Convergence*, John Wiley & Sons.

[25] Frank, T.D. (2008). Nonlinear Markov processes, *Physics Letters A*, Vol. 372, 4553–4555.

[26] Freidlin, M.I. and Wentzell, A.D. (1984). *Random Perturbations of Dynamic Systems*, Springer.

[27] Fricker, C. and Gast, N. (2014). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity, *EURO Journal on Transportation and Logistics*, Published Online: June 7, 2014, Pages 1–31.

[28] Fricker, C., Gast, N. and Mohamed, H. (2012). Mean field analysis for inhomogeneous bike sharing systems. In: *DMTCS Proceedings*, Pages 365–376.

[29] Galves, A., Olivieri, E. and Vares, M.E. (1987). Metastability for a class of dynamical systems subject to small random perturbations, *The Annals of Probability*, Vol. 15, 1288–1305.

[30] Gast, N. and Bruno, G. (2010). A mean field model of work stealing in large-scale systems, *ACM SIGMETRICS Performance Evaluation Review*, Vol. 38, 13–24.

[31] Gast, N. and Bruno, G. (2011). A mean field approach for optimization in discrete time, *Discrete Event Dynamic Systems*, Vol. 21, 63–101.

[32] Gast, N., Bruno, G. and Le Boudec, J.Y. (2012). Mean field for Markov decision processes: from discrete to continuous optimization, *IEEE Transactions on Automatic Control*, Vol. 57, 2266–2280.

[33] Gibbens, R.J., Hunt, P.J. and Kelly, F.P. (1990). Bistability in communication networks, *Disorder in physical systems*, 113–128.

[34] Graham, C. (2000) Chaoticity on path space for a queueing network with selection of the shortest queue among several, *Journal of Applied Probability*, Vol. 37, 198–201.

[35] Graham, C. (2004) Functional central limit theorems for a large network in which customers join the shortest of several queues, *Probability Theory and Related Fields*, Vol. 131, 97–120.

[36] Hayden, R.A., Stefanek, A. and Bradley, J.T. (2012). Fluid computation of passage-time distributions in large Markov models, *Theoretical Computer Science*, Vol. 413, 106–141.

[37] Karpelevich, F.I. and Rybko, A.N. (2000). Thermodynamical limit for symmetric closed queuing networks, *Translations of the American Mathematical Society-Series 2*, Vol. 198, 133–156.

[38] Kipnis, C. and Landim, C. (1999). *Scaling limits of interacting particle systems*, Springer Science & Business Media.

[39] Kolokoltsov, V.N. (2010). *Nonlinear Markov processes and kinetic equations*, Cambridge University Press.

[40] Kolokoltsov, V.N. (2011). Nonlinear Lévy and nonlinear Feller processes: An analytic introduction. arXiv preprint: arXiv:1103.5591.

[41] Kolokoltsov, V.N., Li, J.J. and Yang, W. (2011). Mean field games and nonlinear Markov processes. arXiv preprint: arXiv: 1112.3744.

[42] Kurtz, T.G. (1970). Solution of ordinary differential equations as limits of pure jump Markov processes, *Journal of Applied Probability*, Vol. 7, 49–58.

[43] Le Boudec, J.Y., McDonald, D. and Mundinger, J. (2007). A generic mean-field convergence result for systems of interacting objects. In: *Proc. Conf. IEEE on the quantitative evaluation of systems*, Pages 3–18.

[44] Li, Q.L. (2010). *Constructive Computation in Stochastic Models with Applications: The RG-Factorizations*, Springer and Tsinghua Press.

[45] Li, Q.L. (2014). Tail probabilities in queueing processes, *Asia-Pacific Journal of Operational Research*, Vol. 31, No. 2, 1–31.

[46] Li, Q.L., Dai, G.R., Lui, J.C.S. and Wang, Y. (2014). The mean-field computation in a supermarket model with server multiple vacations, *Discrete Event Dynamic Systems*, Vol. 24, 473–522.

[47] Li, Q.L., Du, Y., Dai, G.R. and Wang, M. (2014). On a doubly dynamically controlled supermarket model with impatient customers, *Computers & Operations Research*, Vol. 55, 76–87.

[48] Li, Q.L. and Lui, J.C.S. (2014). Block-structured supermarket models, *Discrete Event Dynamic Systems*, Published Online: June 29, 2014, Pages 1–36.

[49] Liggett, T.M. (2006). *Interacting particle systems*, Springer Science & Business Media.

[50] Luczak, M.J. and McDiarmid, C. (2006). On the maximum queue length in the supermarket model, *The Annals of Probability*, Vol. 34, 493–527.

[51] Luczak, M.J. and McDiarmid, C. (2007). Asymptotic distributions and chaos for the supermarket model, *Electronic Journal of Probability*, Vol. 12, 75–99.

[52] Martin, J.B. and Suhov, Y.M. (1999). Fast Jackson networks, *The Annals of Applied Probability*, Vol. 9, 854–870.

[53] Mitzenmacher, M.D. (1996). The Power of Two Choices in Randomized Load Balancing. Ph.D. thesis, Department of Computer Science, University of California at Berkeley, USA.

[54] Muzychka, S.A. and Vaninsky, K.L. (2011). A class of nonlinear random walks related to the Ornstein-Uhlenbeck process, *Markov Processes and Related Fields*, Vol. 17, 277–304.

[55] Olivieri, E. and Vares, M.E. (2005). *Large Deviations and Metastability*, Cambridge University Press.

[56] Oseledets, V.I. and Khmelev, D.V. (2002). Stochastic transportation networks and stability of dynamical systems, *Theory of Probability & Its Applications*, Vol. 46, 154–161.

[57] Rogers, L.C.G. and Williams, D. (1994). *Diffusions, Markov processes, and Martingales, Vol. 1: Foundations*, John Wiley & Sons.

[58] Rybko, A.N. and Shlosman, S. (2003). Poisson Hypothesis for information networks (A study in non-linear Markov processes). arXiv preprint: arXiv:0303010.

[59] Shiga, T. and Tanaka, H. (1985). Central limit theorem for a system of Markovian particles with mean field interactions, *Probability Theory and Related Fields*, Vol. 69, 439–459.

[60] Sznitman, A. (1989). *Topics in propagation of chaos*, Springer-Verlag lecture notes in mathematics 1464, Pages 165–251.

[61] Tibi, D. (2010). Metastability in communication networks. arXiv preprint: arXiv:1002.0796.

[62] Turner, S.R.E. (1996) Resource Pooling in Stochastic Networks. Ph.D. Thesis, Statistical Laboratory, Christ's College, University of Cambridge.

[63] Vaninsky, K., Myzuchka, S. and Lukov, A. (2012). A multi-agent nonlinear Markov model of the order book. arXiv preprint: arXiv:1208.3083.

[64] Vvedenskaya, N.D., Dobrushin, R.L. and Karpelevich, F.I. (1996). Queueing system with selection of the shortest of two queues: An asymptotic approach, *Problems of Information Transmission*, Vol. 32, 20–34.

[65] Vvedenskaya, N.D. and Suhov, Y.M. (1997). Dobrushin's mean-field limit for a queue with dynamic routing, *Markov Processes and Related Fields*, Vol. 3, 493–526.

Generation of All Magic Squares of Order 5 and Interesting Patterns Finding

Ziqi Lin[1,*]   Sijie Liu[1,†]   Kai-Tai Fang[1,2,‡]   Yuhui Deng[1,§]

[1]Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519085, China
[2]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Abstract   This paper aims to find an efficient way to generate all magic squares of order 5 and further to discover some interesting patterns of them. Based on Schroeppel's idea and Frénicle-quadset in order 5, we propose an enumeration efficient algorithm for generation of all magic squares of order 5. For extending Frénicle - Amela patterns from order 4 to order 5 we find four patterns to appear at the same time.

Keywords   Enumeration algorithm; Frénicle - Amela patterns; Magic square of order 5.

1   Introduction

A *Magic Square of order* n is an $n \times n$ array consisting numbers in some specific order, so that its sums of the elements in each horizontal rows, vertical columns, and two diagonals are all equal. If elements in a magic square are consecutive integers starting from 1 to n^2, the magic square is call a *Classical Magic Square of order* n with magic sum $\mu_n = n(n^2+1)/2$. In this study we focus only on the classical magic squares (magic squares for short). The magic squares have been studied for thousands years. There are a huge of literature on the magic square, for example, [2, 3]. But there are still so many open problems. For example, for given n what is the number of magic squares of order n? Table 1 shows the number of magic squares of order 3, 4 and 5, and [9] gives a list of the number of magic squares for $n = 3, 4, \cdots, 10$. We can see that the exact number of magic squares of order n are unknown for most n. It also shows that the number of magic squares of order n increases exponentially when n increases.

If a magic square can be obtained from another magic square by some transformations in a certain sense, it is said that these two magic squares are *isomorphic*. For different n the isomorphism employs different group of transformations. All the isomorphic magic squares of a magic square form a *basic form*. Next question is how generate all the magic squares of order n, or how generate all the basic magic squares. This question was solved for $n = 3, 4$ as the total number is not large (see [6]). For $n = 5$, Schroeppel [7] proposed an algorithm for generating all magic squares of order 5. He defined 32 transformations for the basic form. However, we cannot find his computer code and a list of all magic squares of order 5. [10] proposed a computational program which was written by an anonymous. The author claims that all magic squares of order n ($n \geq 3$) can be generated.

But this author did not give any explanation on his program. For classification purpose we need all (at least a certain percentage) of magic squares of order 5. In Section 2 we propose a new algorithm for producing all magic squares of order 5. Various methods for classification of magic squares of order 4 have been proposed. Fang, Luo and Zheng [4] gave an comprehensive review on this topic. Can we extend those methods from order 4 to order 5? Section 3 gives some results on Frénicle - Amela-like patterns and points out that there are 4 Frénicle - Amela-like patterns to appear in the same magic square. The final section gives conclusion.

Table 1: The Number of Magic Squares of Order $n, n = 3, 4, 5$

Order	Classical Magic Square	Basic Magic Square
3	8	1
4	7,040	880
5	2,202,441,792	68,826,306

2   An Enumeration Algorithm of Magic Squares of Order 5

There are several existing methods for construction of magic squares of odd order. For example, the Siamese method, the Lozenge method and matrix addition method can generate some magic squares of an odd order including order 5. These methods aim to provide only one or few magic squares of order 5 and cannot generate all magic squares of order 5. Until now, we cannot find the whole set of magic square of order 5.

Schroeppel's method [7, 5] provides a counting algorithm without the computational code. On the other hand the exhaustive method [10] provides computational code without any explanation. We are not sure whether this method can enumerate all magic squares of order 5 or not. For the classification purpose of magic squares of order 5 we really need to know the way for enumerating the all squares so that we can find some way for choosing representative magic squares of order 5 in further study. This is the main motivation of the paper. Obviously, this new algorithm will generate only all the basic magic squares of order 5. Let $\mathcal{M}$ be the set of magic squares of order 5. Schroeppel [7] considered the following transformations for any $M \in \mathcal{M}$:

s_1 : rotate M 90 degrees clockwise; repeat to use s_1 we can obtain other 3 M's in $\mathcal{M}$;

s_2 : flip M about the main diagonal which means transpose the magic square;

s_3 : exchange the first and fifth rows and columns of M, respectively;

s_4 : exchange rows and columns of M as follows: row 1 $\Leftrightarrow$ row 2, row 4 $\Leftrightarrow$ row 5, column 1 $\Leftrightarrow$ column 2, and column 4 $\Leftrightarrow$ column 5, respectively.

By the above transformations we can obtain 31 other magic squares from a $M \in \mathcal{M}$. The total 32 squares form a basic form of magic squares of order 5. Schroeppel [7] announced that there are 68,826,306 basic magic squares of order 5.

A magic square M of order 5 involves 12 5-vectors (5 rows, 5 columns and 2 diagonal vectors) with the same sum $\mu = 65$. Sort these vectors in descending order and denote the set of these 12 vectors by $\mathcal{V}(M)$. Define a set

$$\mathcal{A} = \{(x_1, \cdots, x_5) : x_i \in \Omega, x_1 > x_2 > x_3 > x_4 > x_5, \sum_{i=1}^{5} x_i = 65\}, \tag{1}$$

where $\Omega = \{1, 2 \cdots, 25\}$. This is an extension of Frénicle-quadset of order 5 (see [6]). Obviously, $\mathcal{V}(M) \subseteq \mathcal{A}$ for any $M \in \mathcal{M}$. It is easy to find that there are 1394 vectors in $\mathcal{A}$. The new algorithm is based on how to choose vector from $\mathcal{M}$ in each step.

Let us put 25 English letters from $A, B, \cdots, Y$ into a 5×5 array (cf. Figure 1). We shall decide values of these 25 letters so that the final square belongs to $\mathcal{M}$.

$$M_5 = \begin{pmatrix} B & J & K & L & F \\ V & C & T & H & X \\ R & M & A & P & S \\ W & I & U & D & Y \\ G & N & Q & O & E \end{pmatrix}$$

Figure 1: 25 letters in 5×5 array

2.1 An Enumeration Algorithm

In this subsection we propose a new enumeration algorithm for generating all basic magic squares of order 5. A flowchart of this algorithm is given in Figure 2. In this flowchart we also use a magic square as an example for illustration. In this algorithm we pick up some idea from Schroeppel [7] and each step depends on a subset of $\mathcal{A}$. In the begin we consider the full set of $\mathcal{A}$. Denote a subset of $\mathcal{A}$ by $\mathcal{A}_x$, where $x \in \Omega$ and

$$\mathcal{A}_x = \{\mathbf{x} = (x_1, \cdots, x_5) : \mathbf{x} \in \mathcal{A}, \text{some } x_i = x\}.$$

Similarly we can define a subset of $\mathcal{A}$, denoted by $\mathcal{A}_{x,y}$, where each vector should involve integers x and y.

The new algorithm involves the following steps:

Step 1 Choose an integer A from the set $\{1, 2, \cdots, 13\}$ as the center number. Suppose we choose $A = 13$.

Step 2 Choose two vectors from $\mathcal{A}_A$ for two diagonals, then we set values for $B, C, \cdots, I$ in Figure 1. Obviously, these two vectors have no overlapping elements except A. For avoiding to generate magic squares belonging to the same basic form, we follow Schroeppel's consideration to put constraints $D > C > B$, $G > F > B$, $E > B$, $I > B$, and $H > B$. For example, two vectors $(25, 24, 13, 2, 1)$ and $(23, 17, 13, 9, 3)$ are chosen and set $B = 1, C = 2, D = 24, E = 25$, and $F = 3, G = 23, H = 9, I = 17$ for satisfying the above constraints.

Step 3 Choose vectors from $\mathcal{A}$ for filling the central vertical circle in columns 2,3, and 4 so that we can assign values to J, K, L, N, Q, O, M, P. There are three substeps:

3a) Choose one vector from $\mathcal{A}_{B,F}$ for the first row of the square. As the values of B and F are fixed already, this vector should have two integers B, F and cannot involve any other integers appearing in the two diagonals. In our example, assign values to $J = 22, K = 21$, and $L = 18$, respectively.

3b) Choose one vector from $\mathcal{A}_{J,C,I}$ and another one from $\mathcal{A}_{L,H,D}$ to complete columns 2 and 4 of the square, respectively. The results are $M = 16, N = 8$, and $P = 10, O = 4$ in this example.

3c) For the last row of the square, assign values to $Q = 5$ so that sum of the last row equals to 65.

Step 4 Assign values to $R = 14, S = 12$ in row 3 and $T = 15, U = 11$ in column 3 by choosing one vector from $\mathcal{A}_{M,A,P}$ and another from $\mathcal{A}_{K,A,Q}$, respectively.

Step 5 Assign values to $V = 20, W = 7, X = 19, Y = 6$ to complete this construction so that sum of each column and row equals to 65.

2.2 Flowchart and example

The details can refer to the flowchart below. It is clear, each step may have many choices. The user needs to run all choice combinations for obtaining all basic magic squares of order 5.

Find pairs include B, F in **Pairs** and store as **JKL**.

JKL is not empty.

No

Yes

The central vertical circle:

B	J	K	L	F
	C		H	
	M	A	P	
	I		D	
G	N	Q	O	E

Store corresponding A to L in matrix **A2L**.

Find pairs include (J,C,I), (L,H,D) and (G,E) in **Pairs** and store as **MN**, **OP** and **NQO**, respectively.

M, N≠O,P and Q≠ 0.

No

Yes

1	22	21	18	3
	2		9	
		13		
	17		24	
23				25

Store in matrix **MNOPQ**.

Store corresponding A to Q in matrix **Cir1**.

1	22	21	18	3
	2		9	
	16	13	10	
	17		24	
23	8	5	4	25

The central horizontal circle:

B	J	K	L	F
V	C	T	H	X
R	M	A	P	S
W	I	U	D	Y
G	N	Q	O	E

1	22	21	18	3
	2	15	9	
14	16	13	10	12
	17	11	24	
23	8	5	4	25

1	22	21	18	3
20	2	15	9	19
14	16	13	10	12
7	17	11	24	6
23	8	5	4	25

Flowchart (left to right, top to bottom):

- Find pairs include (M,A,P) and (K,A,Q) in **Pairs** and store as **RS** and **TU**, respectively.
- R,S,T,U exist and R,S≠T,U → No / Yes
- Store in matrix **RSTU**.
- Store corresponding A to U in matrix **Cir2**.
- Find V, W, X, Y from 1 to 25, which have not assigned to A to U, also store in **VWXY**.
- **VWXY** is not empty; column sums and row sums are equal to 65. → No / Yes
- Store corresponding A to Y in matrix **MS**.
- Output magic square of order 5.
- End

Figure 2: Flowchart and example

Notice that we choose center number only from the set $\{1, 2, \cdots, 13\}$. Can we choose center number from the set $\{14, \cdots, 25\}$? It is known that if $M = (m_{ij}) \in \mathcal{M}$, then $M^c = (26 - m_{ij})$ is also a magic square. We call M^c as complement magic square of M. Therefore, magic squares of order 5 with center number $C, 13 < C \leq 25$, can be obtained from their complement magic squares.

3   Special patterns of Magic Squares of Order 5

There are a lot of studies on classification of magic square of order 4. There are two kinds of methods: graph visualization and algebraic methods. Fang, Luo and Zheng [4] gives a comprehensive review on classification of magic squares of order 4. Candy [1] gave a comprehensive study on classification of magic squares of order 5 from algebraic point of view. In this section we focus only on extension of *Frénicle - Amela patterns*. Frénicle put a magic square of order 4 into a square with 16 subsquares and considered all possible 4 neighborhood numbers adding up to the magic sum $\mu_4 = 34$. There are 5 kinds of pattern in Figure 3 with labels $\alpha, \beta, \gamma, \delta$, and ϵ.
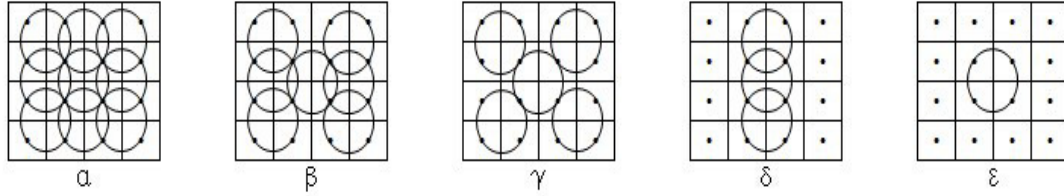
Figure 3: Frénicle - Amela pattern

How to extend Frénicle - Amela patterns from order 4 to order 5 is not a straight forward job. We consider four patterns in Figure 4 such that sum of the integers in one pattern to be the magic sum $\mu_5 = 65$. It is interesting to find these four patterns are strongly dependent.
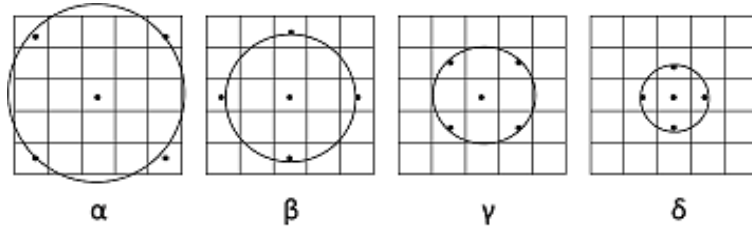
Figure 4: Four patterns

Theorem 1.
If a magic square of order 5 presents one of the four patterns in Figure 4, the other three patterns must appear in this magic square.

Proof
Use letters to stand for different patterns. Let L denote the center number, **a**, **b**, **c**, **d** denote the four patterns, where $\mathbf{a} = (a_1, a_2, a_3, a_4)$, $\mathbf{b} = (b_1, b_2, b_3, b_4)$, $\mathbf{c} = (c_1, c_2, c_3, c_4)$, and $\mathbf{d} = (d_1, d_2, d_3, d_4)$. Then a magic square of order 5 can be written as a array in Figure 5. The four patterns in Figure 4 are equivalent to satisfy the following equations by using the above letters in Figure 5.

a1	A	b1	B	a2
C	c1	d1	c2	D
b2	d2	L	d3	b3
E	c3	d4	c4	F
a3	G	b4	H	a4

Figure 5: Patterns

$$a_1 + a_2 + a_3 + a_4 + L = 65, \tag{2}$$
$$b_1 + b_2 + b_3 + b_4 + L = 65, \tag{3}$$
$$c_1 + c_2 + c_3 + c_4 + L = 65, \tag{4}$$
$$d_1 + d_2 + d_3 + d_4 + L = 65. \tag{5}$$

We need to prove if one of the above equation (2)-(5) holds, then other 3 equations are also hold. Now assume that equation (2) holds and we are going to prove (3),(4) and (5) to be true. According to the definition of the magic square of order 5, each row sum, each column sum and each diagonal sum of a magic square are all equal to 65. Thus, we can obtain 12 equations as follows.

$$
\begin{aligned}
a_1 + A + b_1 + B + a_2 &= 65 & (6)\\
C + c_1 + d_1 + c_2 + D &= 65 & (7)\\
b_2 + d_2 + L + d_3 + b_3 &= 65 & (8)\\
E + c_3 + d_4 + c_4 + F &= 65 & (9)\\
a_3 + G + b_4 + H + a_4 &= 65 & (10)\\
a_1 + C + b_2 + E + a_3 &= 65 & (11)\\
A + c_1 + d_2 + c_3 + G &= 65 & (12)\\
b_1 + d_1 + L + d_4 + b_4 &= 65 & (13)\\
B + c_2 + d_3 + c_4 + H &= 65 & (14)\\
a_2 + D + b_3 + F + a_4 &= 65 & (15)\\
a_1 + c_1 + L + c_4 + a_4 &= 65 & (16)\\
a_2 + c_2 + L + c_3 + a_3 &= 65 & (17)
\end{aligned}
$$

From (16) and (17) we have $a_1 + a_2 + a_3 + a_4 + L + c_1 + c_2 + c_3 + c_4 + L = 65 + 65$. By using (2), we obtain (4), i.e., pattern γ.

Adding equations (6) and (10) together, we have $a_1 + a_2 + a_3 + a_4 + A + B + G + H + b_1 + b_4 = 65 + 65$. Replace $a_1 + a_2 + a_3 + a_4$ by $65 - L$ we obtain

$$A + B + G + H = 65 + L - b_1 - b_4. \tag{18}$$

Adding the condition (12) and (14) together, we obtain $c1 + c2 + c3 + c4 + A + B + G + H + d_2 + d_3 = 65 + 65$. Replacing $c_1 + c_2 + c_3 + c_4$ by $65 - L$ and $A + B + G + H$ by

$65 + L - b_1 - b_4$ we have

$$d_2 + d_3 = b_1 + b_4. \tag{19}$$

By a similar way we are going to show $d_1 + d_4 = b_2 + b_3$. Firstly, adding (11) and (15) together, we obtain $a_1 + a_2 + a_3 + a_4 + C + D + E + F + b_2 + b_3 = 65 + 65$. Replacing $a_1 + a_2 + a_3 + a_4$ by $65 - L$ we obtain

$$C + D + E + F = 65 + L - b_2 - b_3. \tag{20}$$

Secondly adding (7) and (9) implies $c_1 + c_2 + c_3 + c_4 + C + D + E + F + d_1 + d_4 = 65 + 65$. Replace $c_1 + c_2 + c_3 + c_4$ by $65 - L$ and replace $C + D + E + F$ by $65 + L - b_2 - b_3$, we have

$$d_1 + d_4 = b_2 + b_3. \tag{21}$$

From (19) and (21) we have

$$d_1 + d_2 + d_3 + d_4 = b_1 + b_2 + b_3 + b_4. \tag{22}$$

Thirdly, adding the condition (8) and (13) together, we obtain $b_1 + b_2 + b_3 + b_4 + d_1 + d_2 + d_3 + d_4 + 2L = 2 \times 65$. Since $b_1 + b_2 + b_3 + b_4 = d_1 + d_2 + d_3 + d_4$, replace one of them by another one we find

$$\begin{aligned} b_1 + b_2 + b_3 + b_4 + L &= 65, \\ d_1 + d_2 + d_3 + d_4 + L &= 65. \end{aligned} \tag{23}$$

Equations in (23) show that patterns β and δ exist. By a similar way we can prove from pattern δ to find patterns α, β and γ; and other cases. It completes the proof. $\quad\square$

4 Conclusion

In this paper we propose an algorithm for generating all the basic magic squares of order 5. Millions of basic magic squares of order 5 have been generated already. It needs time to obtain all the magic squares of order 5. We plan to put these squares into a database for public use in future. Classification of magic squares of order 5 is a challenging topic to study. Candy [1] gave a comprehensive study from algebraic point of view. In this paper we use graphical method for classification. We believe that there are a lot of interesting results to be discovered along this direction in near future.

Acknowledgement

References

[1] Candy, A.L. (1939), *Construction, Classification and Census of Magic Squares of Order 5* (2nd Edition), Albert L. Candy, Author and Publisher.

[2] Chinese Magic Square, Retrieved 12 June 2014 from `http://www.zhghf.net/`

[3] Clifford, A. P. (2003). *The Zen of Magic Squares, Circles, and Stars.* US: Princeton University Press.

[4] Fang, K.T., Y.Y. Luo adn Y.X. Zheng (2014). Classification of magic squares of order 4, 24rd International Workshop on Matrices and Statistics (IWMS-2015), Haikou, China.

[5] Garder, M. (1975), Mathematical games, A breakthrough in magic squares, and the first perfect magic cube, *Scientific American, Inc.*, 118-123.

[6] Ollerenshaw, K., & Bondi, H. (1982). Magic squares of order four. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 443-532.

[7] Schroeppel, R. (1973). The Order 5 Magic Squares Program proposed by Richard Schroeppel, published in *Scientific American* in January 1976.

[8] Styan, G. P. H. (2014). *Some illustrated comments on 5 × 5 golden magic mattices and on 5 × 5 Stifelsche Quadrate.* Presented as an invited paper at the 23rd International Workshop on Matrices and Statistics (IWMS-2014), Ljubljana, Slovenia.

[9] Trump, W. (2012). *How many magic squares are there?*. Retrieved May 1, 2013, from http://www.trump.de/magic-squares/howmany.html

[10] Baidu Tieba, Retrieved 11 June 2014 from
`http://tieba.baidu.com/p/957776994?pid=10675158083&cid=0&from=prin`
`\sharp10675158083?from=prin`

Sensitivity analysis in linear models

Shuangzhe Liu[1],[*]   Tiefeng Ma[2],[†]   Yonghui Liu[3],[‡]

[1]Faculty of Education, Science, Technology and Mathematics, University of Canberra, Australia
[2]School of Statistics, Southwestern University of Finance and Economics, China
[3]School of Business Information, Shanghai University of International Business and Economics, Shanghai, China

Abstract   In this work we consider the general linear model and some of its extensions. We study their sensitivity matrix results, with theoretical developments and numerical comparisons. We include illustrated examples.

Keywords   Least squares; Maximum likelihood; Mixed estimation, Sensitivity matrix

1   Introduction

The local sensitivities of the least squares and other estimators of the regression coefficients in linear and mixed models have been studied and applied to several areas. The local sensitivities of the posterior mean and precision matrix in the Bayesian context have been established. For an introductory account of matrix differential calculus with applications in statistics and econometrics, see Magnus and Neudecker [7]. For a motivation and earlier results from the Bayesian point of view, see Polasek [9]. For a fundamental treatment of local sensitivity and diagnostic tests with applications to linear and random effects models, see Magnus and Vasnev [8] and Vasnev [12]. For the sensitivity matrices of least squares estimators and their relevant uses in spatial and panel-spatial autoregressive models, see Liu et al. [6, 4].

In this paper we make a systematic study on the local sensitivities of the generalized least squares, maximum likelihood and other estimators in the general linear model and its variants including those with instrumental variables and those with linear restrictions. We include the variance estimators in these models. We consider both normal and no-normal distribution assumptions for these models. In section 2, we introduce basic matrix differential calculus results and the definition of local sensitivity matrix in the context of linear models we need for the later sections. In section 3, we present the local sensitivity matrix results in general linear model and conduct their numerical comparisons in a number of scenarios. In section 4, we consider the different variations of the general linear model. In section 5, we discuss possible extensions of the models and results presented in the two previous sections. We make concluding remarks in section 6.

2   Matrix calculus and local sensitivity

2.1   Matrix calculus

For $n \times p$ matrix X we use vecX denote the vectorization vector of X. Let $\otimes$ denote the (right) Kronecker product of two matrices. We have the following definitions.

---

[*]Corresponding author. Email: shuangzhe.liu@canberra.edu.au
[†]Email: matiefeng@swufe.edu.cn
[‡]Email: liuyh@lsec.cc.ac.cn

Definition 1. Let $g(x)$ be a scalar function of an $n \times 1$ vector x. The derivative of $g(x)$ is

$$\mathrm{D}g(x) = \partial g(x)/\partial x'. \tag{1}$$

Definition 2. Let $f(x)$ be an $m \times 1$ vector function of an $n \times 1$ vector x. The derivative (or Jacobian matrix) of $f(x)$ is

$$\mathrm{D}f(x) = \partial f(x)/\partial x'. \tag{2}$$

Definition 3. If $F(X)$ is a differential $m \times q$ real matrix function of an $n \times p$ matrix X, then the Jacobian matrix of $F(x)$ at X is the $mq \times np$ matrix

$$\mathrm{D}F(X) = \partial F(X)/\partial (\mathrm{vec}X)'. \tag{3}$$

2.2 Local sensitivity

Consider the general linear model, as given in e.g. Magnus and Neudecker [7] and Rao et al. [10]

$$y = X\beta + \epsilon \tag{4}$$

where y is an $n \times 1$ vector of observable random variables, X is a non-stochastic $n \times q$ matrix and ϵ is an $n \times 1$ vector of random disturbances with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{E}(\epsilon\epsilon') = \sigma^2 V$, where V is a known positive definite $n \times n$ matrix and σ^2 is unknown. The $p \times 1$ vector β of regression coefficients is supposed to be fixed but unknown, and needs to be estimated.

The least squares estimator of β is

$$b = (X'X)^{-1}X'y. \tag{5}$$

The local sensitivity $p \times np$ matrix of $\hat{\beta}$ with respect to X is

$$\partial b/\partial (\mathrm{vec}X)' = (X'X)^{-1} \otimes (y - Xb)' - b' \otimes (X'X)^{-1}X'. \tag{6}$$

This matrix reflects the effects of small changes in X on the least squares estimator b.

The general linear model, with alternatives including the generalized least squares estimator of β, and some of its variations are considered in the next section. One of the purposes of this present work is to establish their local sensitivity results.

3 General linear model and variations

3.1 General linear model

For model (1), the generalized least squares estimator of β is

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y. \tag{7}$$

Its local sensitivity $\hat{\beta}$ with respect to X is $\partial\hat{\beta}/\partial (\mathrm{vec}X)'$.

As we assume σ^2 is unknown and needs to be estimated, an unbiased estimator is

$$\hat{\sigma}^2 = (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})/(n - p). \tag{8}$$

We may then find the sensitivity matrix of $\hat{\sigma}^2$ with respect to X i.e. $\partial\hat{\sigma}^2/\partial (\mathrm{vec}X)'$.

3.2 General linear model with linear restrictions

For the general linear regression model as given by (1)

$$y = X\beta + \epsilon, \tag{9}$$

where we consider we have prior information about β in the form of a set of k independent exact linear restrictions expressed as

$$r = R\beta, \tag{10}$$

where R is a $k \times p$ known matrix of rank $k \leq p$ and r is a $k \times 1$ vector of known elements.

The restricted least squares estimators of the elements of the parameters in the formulation (9) and (10) are well-known to be

$$\hat{\beta}_{\mathrm{RLS}} = \hat{\beta} - (X'V^{-1}X)^{-1}R'[R(X'V^{-1}X)^{-1}R']^{-1}(Rb - r), \tag{11}$$

$$\hat{\sigma}^2_{\mathrm{RLS}} = \frac{(y - X\hat{\beta}_{\mathrm{RLS}})'V^{-1}(y - X\hat{\beta}_{\mathrm{RLS}})}{n - p + k}, \tag{12}$$

where $\hat{\beta} = (X'V^{-1}X)^{-1}XV^{-1}y$ is the (unrestricted) generalized least squares estimator of β.

We present their local sensitivity matrices, extending a result by Liu and Neudecker [5] for $V = I$.

3.3 Numerical analysis

We conduct simulation studies to examine the sensitivity results and make numerical comparisons.

4 Possible extensions

4.1 Extensions

We discuss possible extensions of those models considered. We may consider the maximum likelihood estimation for those models under elliptical distributions studied by e.g. Fang et al. [1]. We may consider related models including those with stochastic restrictions, seemingly unrelated regression and growth curve models studied by e.g. Theil and Goldberger [11], Liu et al. [3] and Gruber [2].

4.2 Discussion

We may apply some sensitivity results to the first-order approximations of those possible estimators in an approach as taken in e.g. Liu et al. [6, 4].

5 Concluding remarks

We have treated the models in a systematic manner.

Acknowledgement

References

[1] Fang, K.T., Kotz, S., Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman, London.

[2] Gruber, M.H.J. (2010). *Regression Estimators: A Comparative Study*, John Hopkins University Press, Baltimore, MD.

[3] Liu, S., Ahmed, S.E., Ma, L.Y. (2009). Influence diagnostics in the linear regression model with linear stochastic restrictions, *Pakistan J. Statist.*, 25, 647-662.

[4] Liu, S., Ma, T., Polasek, W. (2014). Spatial system estimators for panel models: a sensitivity and simulation study. *Math. Comput. Simul.*, 101, 78-102.

[5] Liu, S., Neudecker, H. (2007). Local sensitivity of the restricted least squares estimator in the linear model. *Statist. Pap.*, 48, 525.

[6] Liu, S., Polasek, W., Sneller, R. (2012) Sensitivity analysis of SAR estimators: a numerical approximation. *J. Statist. Comput. Simul.*, 82(2), 325-342.

[7] Magnus, J.R., Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, Chichester.

[8] Magnus, J.R., Vasnev, A.L. (2007). Local sensitivity and diagnostic tests, *Econometrics J.*, 10(1), 166-192.

[9] Polasek, W. (1986). Local sensitivity analysis and Bayesian regression diagnostics, In Goel, P.K. and Zellner, A. (Eds.), *Bayesian Inference and Decision Techniques*, North-Holland, Amsterdam. p. 375-387.

[10] Rao, C.R., Toutenburg, H., Shalabh, Heumann, C. (2010). *Linear Models and Generalizations*, Springer, Berlin.

[11] Theil, H., Goldberger, A.S. (1961). On pure and mixed estimation in econometrics, *Int. Econ. Rev.*, 2, 65-78.

[12] Vasnev, A.L. (2010). Sensitivity of generalized least squares estimators in random effects models, *J. Multivariate Anal.*, 101(5), 1252-1262.

Semiparametric Analysis of Recurrent Event Data with Cure Rate

Ni Li*   Xiaofeng Zeng

School of Mathematics and Statistics, Hainan Normal University

Abstract   Recurrent event data usually occur in long-term studies which concern recurrence rates of the disease. In studies of medical sciences, patients who have infected with the disease, like cancer, were conventionally regarded as impossible to be cured. However, with the development of medical sciences, recently those patients were found to be possibly recovered from the disease. The recurrence rate of the events, which is of primary interest, may be affected by the cure rate that may exist. Therefore, we proposed semiparametric statistical analysis for recurrent event data with subjects possibly being cured. In our approach, we present a proportional rate model for recurrence rate with the cure rate adjusted through a logistic regression model, and develop some estimating equations for estimation of the regression parameters, with their large sample properties, including consistency and asymptotic normality established. Numerical studies under different settings were conducted for assessing the proposed methodology and the results suggest that they work well for practical situations. The approach is applied to a bladder cancer dataset which motivated our study.

Keywords   Recurrent Event; Proportional Rate Model; Cure Rate; Logistic Model; Estimating Equation

1   Introduction

Recurrent event data usually occur in long-term studies which concern recurrence rates of certain events. Difficulty may arise due to the inherent order and dependence of the times of recurrences. Besides the recurrence times, censor time may exist. What's more, the recurrent times and censor time may also be dependent. All above brings challenges in statistical analysis and relevant inference for recurrent event data [1, 2, 3]. The results of research on recurrent event data therefore will serve as new methodologies with wide applications.

In the analysis of recurrent event data, it is often of primary interest to study the effect of covariates on recurrence rate. Let $N^*(t)$ be the number of events occurred in $[0,t]$, then $\mu(t)$ is referred as rate function of $N^*(t)$, if $E\{dN^*(t)\} = \mu(t)dt$, where $dN^*(t) = N^*\{(t+dt)^-\} - N^*(t^-)$, the increment of N^* in the interval $[t, t+dt)$ when $dt \to 0$. Suppose there are n independent subjects, and denote $N_i^*(t) = \int_0^t dN_i^*(s)$ by the number of recurrences that subject i has experienced up to time t. In real data analysis, the observation usually terminates at some time point, that is, the subject i may be censored at time C_i. The underlying recurrent event process $N_i^*(\cdot)$ after the censor time C_i was unobserved, and the actual process we observed was $N_i(t) = \int_0^t I(C_i \geq s)\, dN_i^*(s)$. Furthermore, let $X_i(t)$ denote the covariates which may affect the recurrence rate, and assume the

---

independence of the censor time C_i with $N_i^*(t)$ given $X_i(\cdot)$. That is, $E\{dN_i^*(t)|X_i(t), C_i \geq t\} = E[dN_i^*(t)|X_i(t)]$. For the rate function of the recurrent event given $X_i(\cdot)$, what is well known as the proportional rate model is listed as follows:

$$E\{dN_i^*(t)|X_i(t)\} = \lambda_0(t)\exp\{\gamma_0'X_i(t)\}dt \tag{1}$$

where γ_0 is a vector of unknown parameter, and $\lambda_0(\cdot)$ is the baseline rate function which is also unknown. About the model (1), Pepe and Cai[4] established the large sample theory of the estimate for γ_0, Lawless, et al. [5, 6] studied the estimates of γ_0 and $\Lambda_0(t)$ and their asymptotic properties with assuming the recurrence times are discrete, among others, Lin et al. [7] made improvement to [5, 6] methods and proposed the confidence interval estimation for the mean function.

With recent development of medical sciences, some diseases that conventionally regarded as impossible to be cured and have the tendency of recurring were found to be likely cured. Those patients who have been recovered are called long-term survivors [8]. After some time point, the cured subjects would not be infected with the disease anymore and withdraw from the recurrent event process. Therefore it somehow decrease the recurrence rate of the disease. If we apply traditional methods to analyze recurrent event data with long-term survivors, the results may not be convincing and get difficult to interpret [9].

For survival data, there are some existing methods on how to take into account the cure rate. Boag [10] firstly defined mixture models and proposed estimation procedures for survival proportions of both cured and uncured patients. Farewell [11] generalized Cox model and used mixture models to analyze survival data with long-term survivors. To study the hazard function, Farewell took logistic regression model for cured subjects and proportional hazard models for subjects under risk, that is, uncured subjects. Kuk and Chen[12], and Taylor [13] estimated the regression parameters and cured probability using semiparametric framework. However, since the variable which indicates cured or not is actually unobserved, Sy and Taylor[14] viewed the cured variable as a latent variable, and then manipulated the estimation by EM algorithm. All the methods aforementioned are based on the survival data, while in the recurrent event data, subjects are assumed to be infected recurrently and there is no subject cured. Few research exists for the context of recurrent event analysis with concern of cure rate. Cook and Lawless[15] mentioned modeling the cure rate by logistic model though, more details are needed.

In section 2, we proposed a logistic regression model for cure rate and then present a proportional rate model for recurrence rate (2), with the primary interest in the effects of covariates on recurrent rate. Further, estimating equations were developed for the estimation of regression parameters. Consistency and asymptotic normality of estimation were established in section 3. In section 4, we conducted numerical studies under different settings for assessing the proposed models and estimations, and the results suggest that they work well. In addition, we applied the approach to a real data set of bladder tumor, and the results were given in section 5.

2   Models and Estimation Methods

Suppose there are n subjects in our study and they are independent. $W(t)$ and $X(t)$ denote the covariates which may affect the recurrent rate. Specifically, $W(t)$ affects the recurrent rate through cure rate, that is, $W(t)$ doesn't affect the rate of recurrences directly, but affects it through the cure rate. Comparatively $X(t)$ affects recurrent rate directly, and the affect is fitted by Cox proportional rate model (1) . $W(t)$ and $X(t)$ are exclusively different covariate vectors. Let $N^*(t)$ be the underlying counting process up to time t,

since a subject could be censored at time C, hence $N^*(t)$ is only observable by time t. Take $N(t) = N^*(t \wedge C)$, where $a \wedge b = \min(a, b)$, then $N(t)$ is the actual observed recurrent process.

Let π denote the cure probability for some subject, then uncured probability is equal to $(1 - \pi)$. If the subjects consist of cure patients, then we propose semiparametric proportional regression model with cure rate as follows:

$$E[dN_i^*(t)|W_i(t), X_i(t)] = [1 - \pi(W_i(t))] \exp\{\gamma_0' X_i(t)\} \lambda_0(t) dt \qquad (2)$$

$$\pi(W_i(t)) = \frac{\exp\{\beta_0' W_i(t)\}}{1 + \exp\{\beta_0' W_i(t)\}} \qquad (3)$$

In recurrent event data, the observed data consist of $\{N_i(\cdot), X_i(\cdot), W_i(\cdot), C_i\}$ $(i = 1, \cdots, n)$ and indicator variables $Y_i(t) = I(C_i \geq t)$. If we define the process:

$$M_i(t; \theta_0) = N_i(t) - \int_0^t Y_i(s)[(1 + \exp\{\beta_0' W_i(s)\})^{-1} \exp\{\gamma_0' X_i(s)\} d\Lambda_0(s)]$$

then we can easily show that $M_i(t; \theta_0)$ is a zero-mean process. The proof will be given in the Appendix.

To simplify, let $Z(t) = (W(t)', X(t)')'$, then for given $\theta = \{\beta', \gamma'\}'$, we estimate $\Lambda_0(t)$ by

$$\hat{\Lambda}_0(t; \theta) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s)(1 + \exp\{\beta' W_i(s)\})^{-1} \exp\{\gamma' X_i(s)\}} \qquad (4)$$

and to estimate $\theta_0 = \{\beta_0', \gamma_0'\}'$, we construct estimating equations

$$\sum_{i=1}^n \int_0^\tau Z_i(s)\{dN_i(s) - Y_i(s)[(1 + \exp\{\beta' W_i(s)\})^{-1} \exp\{\gamma' X_i(s)\} d\Lambda_0(s; \theta)]\} = 0 \qquad (5)$$

Define

$$S_z(t; \theta) = n^{-1} \sum_{i=1}^n Y_i(t)(1 + \exp\{\beta' W_i(t)\})^{-1} \exp\{\gamma' X_i(t)\} Z_i(t)$$

$$S_0(t; \theta) = n^{-1} \sum_{i=1}^n Y_i(t)(1 + \exp\{\beta' W_i(t)\})^{-1} \exp\{\gamma' X_i(t)\}$$

$$\bar{Z}(t; \theta) = \frac{S_z(t; \theta)}{S_0(t; \theta)}$$

and denote the limits of $S_z(t; \theta), S_0(t; \theta), \bar{Z}(t; \theta)$ by $s_z(t; \theta), s_0(t; \theta), \bar{z}(t; \theta)$, respectively, then if we plug (4)into (5), we have

$$\sum_{i=1}^n \int_0^\tau \{Z_i(s) - \bar{Z}(s; \theta)\} dN_i(s) = 0 \qquad (6)$$

If we denote the solution to (6) by $\hat{\theta} = (\hat{\beta}', \hat{\gamma}')'$, then the estimator of $\Lambda_0(t)$ can be given by $\hat{\Lambda}_0(t; \hat{\theta})$. In the following section, we will illustrate with asymptotic results.

3 Asymptotic Results

We assume the following holds:

($\Delta 1$) $\{N_i(\cdot), Y_i(\cdot), W_i(\cdot), X_i(\cdot)\}$, $(i = 1, \cdots, n)$ are identically independently distributed.

($\Delta 2$) $P(Y_i(\tau) = 1) > 0$, and $N_i(\tau) < \eta < \infty, i = 1, \cdots, n$, a.e., where η is a constant.

($\Delta 3$) $W_i(\cdot)$ and $X_i(\cdot)$ can be bounded by a constant.

($\Delta 4$) A is nonsingular matrix, where

$$A = E\{\int_0^\tau \{Z_i(s) - \bar{z}(s; \theta_0)\}^{\otimes 2} Y_i(s) [(1 + \exp\{\beta_0' W_i(s)\})^{-1} \exp\{\gamma' X_i(s)\} d\Lambda_0(s)] \} \quad (7)$$

Denote $b^{\otimes 2} = bb'$, and let

$$U(t; \theta) = \sum_{i=1}^n \int_0^t \{Z_i(s) - \bar{Z}(s; \theta)\} dN_i(s).$$

The main asymptotic results are listed as follows, with their proofs given in the Appendix.

Theorem 1. *(Asymptotics on $U(t; \theta)$)* *If $(\Delta 1)$-$(\Delta 4)$ holds, then $n^{-1/2}U(t; \theta)$ asymptotically follows a multivariate normal distribution with mean of 0 and variance-covariance matrix of $\xi(s, t)$, where*

$$\xi(s, t) = E[\int_0^s \{Z_i(u) - \bar{z}(u; \theta_0)\} dM_i(u; \theta_0) \int_0^t \{Z_i(v) - \bar{z}(v; \theta_0)\}' dM_i(v; \theta_0)] \quad (8)$$

By strong law of large number and lemma 1 in [7] we can know that the consistent estimator of $\xi(s, t)$ is $\hat{\xi}(s, t)$, here

$$\hat{\xi}(s, t) = n^{-1} \sum_{i=1}^n [\int_0^s \{Z_i(u) - \bar{Z}(u; \hat{\theta})\} d\hat{M}_i(u; \hat{\theta}) \int_0^t \{Z_i(v) - \bar{Z}(v; \hat{\theta})\}' d\hat{M}_i(v; \hat{\theta})]$$

$$\hat{M}_i(t; \hat{\theta}_0) = N_i(t) - \int_0^t Y_i(s)[(1 + \exp\{\hat{\beta}' W_i(s)\})^{-1} \exp\{\hat{\gamma}' X_i(s)\} d\hat{\Lambda}_0(s; \hat{\theta})]$$

Theorem 2. *(Asymptotics on $\hat{\theta}_0$)* *If $(\Delta 1)$-$(\Delta 4)$ holds, then $\hat{\theta}$ consistently converges to θ_0, and $n^{\frac{1}{2}}\{\hat{\theta} - \theta_0\}$ asymptotically follows a Gaussian process with mean of 0 and variance-covariance matrix of $A^{-1}\Sigma A^{-1}$, where A is defined in $(\Delta 4)$, and $\Sigma = \xi(\tau, \tau)$. And the consistent estimator of $n^{\frac{1}{2}}\{\hat{\theta} - \theta_0\}$ is given by $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$, where $\hat{\Sigma} = \hat{\xi}(\tau, \tau)$, and*

$$\hat{A} = n^{-1} \sum_{i=1}^n \{\int_0^\tau \{Z_i(s) - \bar{Z}(s; \hat{\theta})\}^{\otimes 2} Y_i(s)[(1 + \exp\{\hat{\beta}' W_i(s)\})^{-1} \exp\{\hat{\gamma}' X_i(s)\} d\hat{\Lambda}_0(s)]\}$$

Theorem 3. *(Asymptotics on $\hat{\Lambda}_0(t)$)* *If $(\Delta 1)$-$(\Delta 4)$ hold, then $\hat{\Lambda}_0(t)$ for $t \in [0, t]$ consistently converges to $\Lambda_0(t)$ a.e., and $n^{\frac{1}{2}}\{\hat{\Lambda}_0(t; \hat{\theta}) - \Lambda_0(t)\}$ asymptotically follows a Guassian process with mean of 0 and variance-covariance matrix of $\Gamma(s, t) = E\{\Phi_i(s)\Phi_i(t)\}$, where*

$$\Phi_i(t) = \int_0^t \frac{dM_i(u; \theta_0)}{s_0(u; \theta_0)} - B(t; \theta_0)' A^{-1} \int_0^\tau Y_i(u)\{Z_i(u) - \bar{z}(u; \theta_0)\}' dM_i(u; \theta_0) \quad (9)$$

$$B(t; \theta) = \int_0^t \bar{z}(u; \theta)\lambda_0(u) du \quad (10)$$

We can get a consistent estimator of $\Gamma(s,t)$ as

$$\hat{\Gamma}(s,t) = n^{-1} \sum_{i=1}^{n} \hat{\Phi}_i(s)\hat{\Phi}_i(t), \tag{11}$$

where

$$\hat{\Phi}_i(t) = \int_0^t \frac{d\hat{M}_i(u;\hat{\theta})}{S_0(u;\hat{\theta})} - \hat{B}(t)'\hat{A}^{-1} \int_0^\tau Y_i(u)\{Z_i(u) - \bar{Z}(u;\hat{\theta})\}'d\hat{M}_i(u;\hat{\theta}),$$

$$\hat{B}(t;\theta) = \int_0^t \bar{Z}(u;\hat{\theta})d\hat{\Lambda}_0(u;\hat{\theta})$$

4 Simulation Studies

To assess the performance of our proposed estimation procedures under small sample, we conducted a set of simulations. First we generated X_i and W_i from a Bernoulli distribution with a probability of success 0.5. The longest follow-up time was τ, and the censor time C_i was generated from a uniform distribution in $(\tau/2, \tau)$. Let $\lambda_0(t) = c/\tau$, where c is a constant, and $N_i(t)$ is a Poisson process with the mean function of

$$\Lambda(C_i|Z_i) = \Lambda_0(C_i)(1 + \exp\{\beta W_i\})^{-1}\exp\{\gamma X_i\} = cC_i(1 + \exp\{\beta W_i\})^{-1}\exp\{\gamma X_i\}/\tau.$$

The observation times are the order statistics of $(t_{i1}, \cdots, t_{ik_i})$, with k_i generated from the uniform distribution of $U(0, C_i)$.

Table 1: Simulation results for proportional rate model with cure rate

	n=100		n=100		n=200		n=200	
	$\tau = 1$		$\tau = 2$		$\tau = 1$		$\tau = 2$	
	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\beta}$
$\theta = (0,0)$								
BIAS	0.0020	-0.0123	0.0016	-0.0053	-0.0034	-0.0063	-0.0054	-0.0063
SSE	0.1078	0.2359	0.1075	0.2324	0.0794	0.1573	0.0789	0.1593
SEE	0.1386	0.2428	0.1394	0.2438	0.0649	0.1665	0.0650	0.1665
CP	0.9140	0.9260	0.9240	0.9320	0.9410	0.9430	0.9260	0.9380
$\theta = (0,0.1)$								
BIAS	-0.0044	-0.0532	-0.0048	-0.0604	-0.0041	-0.0460	-0.0057	-0.0456
SSE	0.1126	0.2181	0.1090	0.2187	0.0755	0.1442	0.0630	0.1725
SEE	0.1495	0.2548	0.1478	0.2534	0.0828	0.1745	0.0728	0.1747
CP	0.9280	0.9360	0.9280	0.9400	0.9330	0.9470	0.9420	0.9500
$\theta = (0.1,0)$								
BIAS	-0.0399	-0.0154	-0.0372	-0.0134	-0.0358	-0.0042	-0.0380	-0.0132
SSE	0.1105	0.2283	0.1111	0.2320	0.0785	0.1639	0.0786	0.1618
SEE	0.1380	0.2428	0.1383	0.2427	0.0657	0.1672	0.0645	0.1660
CP	0.9310	0.9340	0.9230	0.9290	0.9260	0.9330	0.9270	0.9340
$\theta = (0.1,0.1)$								
BIAS	-0.0408	-0.0540	-0.0417	-0.0613	-0.0415	-0.0565	-0.0433	-0.0575
SSE	0.1078	0.2220	0.1133	0.2218	0.0766	0.1559	0.0686	0.1579
SEE	0.1500	0.2549	0.1486	0.2532	0.0781	0.1739	0.0827	0.1579
CP	0.9270	0.9310	0.9170	0.9290	0.9150	0.9270	0.9240	0.9320

Table 1 showed the estimation results of (β_0, γ_0) under the settings when (β_0, γ_0) take $(0,0), (0.1,0), (0,0.1)$ and $(0.1,0.1)$, and τ takes 1 or 2 . The sample size $n = 100$ or 200. The cure rate is 10%. We make 1000 replications of simulations. The table displays Bias,

SSE, SEE and CP, where Bias represents the sample mean of $\hat{\theta}$ minus the true value of parameter, SSE is sampling standard deviation of $\hat{\theta}$, SEE is the sample mean of $\hat{\theta}$ standard deviation estimates, and CP is the 95% empirical confidence level of θ_0. From the results we can see that the Bias is nearly 0, SSE and SEE are quite close to each other, and CP is also reasonable. All above showed that our proposed approaches work well.

Table 2: Comparison to the results of $\hat{\gamma}$ under proportional rate model without cure rate

	$n = 100$	$n = 200$
	BIAS BIAS* SSE SSE*	BIAS BIAS* SSE SSE*
$\tau = 1$		
$\gamma = -0.1$	-0.0054 -0.1176 0.2238 0.2972	-0.0179 -0.2085 0.1749 0.3623
$\gamma = 0$	-0.0072 -0.1255 0.3152 0.3371	0.0019 -0.1575 0.2382 0.3722
$\gamma = 0.1$	0.0241 0.2184 0.2871 0.3865	0.0197 -0.1722 0.2247 0.4174
$\tau = 2$		
$\gamma = -0.1$	-0.0137 0.1452 0.2774 0.4012	-0.0109 -0.1528 0.2492 0.3458
$\gamma = 0$	0.0013 0.2015 0.1982 0.3208	0.0027 0.1346 0.2587 0.3635
$\gamma = 0.1$	0.0126 -0.1528 0.2498 0.4054	0.0265 0.2343 0.3481 0.4723

5 Real Data Analysis

Table 3: Estimations of Bladder Tumor Data

	Coefficient Estimates	Standard Deviation	95% Confidence Interval	P-value
Initial number of tumors	-0.0070	0.0450	(-0.0952, 0.0813)	0.8773
Cumulative number of tumors	-1.4487	0.1589	(-1.7602, -1.1372)	0.0000

Acknowledgement

Appendix

Proof of $M_i(t; \theta_0)$ By conditional expectation, we have

$$
\begin{aligned}
E\{dN_i(t)\} &= E\{E[dN_i(t)|Z_i(t)]\} \\
&= E\{E[Y_i(t)dN_i^*(t)|Z_i(t)]\} \\
&= E\{E[Y_i(t)|Z_i(t)]E[dN_i^*(t)|Z_i(t)]\} \\
&= E\{[Y_i(t)][(1 - \pi(W_i(t)))\exp\{\gamma_0' X_i(t)\}\lambda_0(t)dt]\} \\
&= E\{Y_i(t)(1 + \exp\{\beta_0' W_i(t)\})^{-1}\exp\{\gamma_0' X_i(t)\}\lambda_0(t)dt\}
\end{aligned}
$$

Therefore $E\{M_i(t; \theta_0)\} = 0$, that is, $M_i(t; \theta_0)$ is a zero mean stochastic process.

Proof of 1 Since $\sum_{i=1}^{n}\{Z_i(s) - \bar{Z}(s;\theta)\}Y_i(s)(1 + \exp\{\beta'W_i(s)\})^{-1}\exp\{\gamma'X_i(s)\} = 0$, hence $U(t;\theta_0)$ can be written as

$$U(t;\theta_0) = \sum_{i=1}^{n}\int_0^t \{Z_i(s) - \bar{Z}(s;\theta_0)\}dM_i(s;\theta_0)$$

$$= \bar{M}_Z(t) - \int_0^t \bar{Z}(s;\theta_0)d\bar{M}(s)$$

where $\bar{M}(t) = \sum_{i=1}^{n} M_i(t;\theta_0)$, $\bar{M}_Z(t) = \sum_{i=1}^{n}\int_0^t Z_i(s)dM_i(s;\theta_0)$ Similar to the proof of [7] Appendix A.2, we know that $n^{-1/2}U(t;\theta_0)$ converges weakly, and its limiting distribution has a mean of zero and variance-covariance matrix given in (8).

Proof of 2 For the existence, uniqueness, and consistency of $\hat{\theta}$: let $\hat{A}(\theta) = -n^{-1}\partial U(\tau;\theta)/\partial\theta'$, since

$$U(\tau;\theta) = \sum_{i=1}^{n}\int_0^\tau \{Z_i(s) - \bar{Z}(s;\theta)\}dM_i(s;\theta)$$

then we have

$$\begin{aligned}A^*(\theta) =& n^{-1}\sum_{i=1}^{n}\int_0^\tau \{Z_i(s) - \bar{Z}(s;\theta)\}^{\otimes 2}Y_i(s)\\ & [(1 + \exp\{\beta'W_i(s)\})^{-1}\exp\{\gamma'X_i(s)d\Lambda_0(s)]\}\\ & + n^{-1}\sum_{i=1}^{n}\int_0^\tau \frac{\partial\bar{Z}(s;\theta)}{\partial\theta'}dM_i(s;\theta)\end{aligned} \quad (12)$$

For asymptotic normality, we have

$$n^{1/2}(\hat{\theta} - \theta_0) = A^{-1}n^{-1/2}U(\tau;\theta_0) + o_p(1) \quad (13)$$

Therefore by (1), we have $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution to a normal vector with a mean of zero and variance-covariance matrix of $A^{-1}\Sigma A^{-1}$, with the variance-covariance matrix can be consistently estimated by $\hat{A}^{-1}\hat{\Sigma}\hat{A}^{-1}$.

Proof of 3 First for the consistency of $\hat{\Lambda}_0(t)$:

$$\hat{\Lambda}_0(t) - \Lambda_0(t) = n^{-1}\sum_{i=1}^{n}\int_0^t \frac{dM_i(u;\theta_0)}{S_0(u;\hat{\theta})} - \int_0^t \frac{S_Z(u;\theta^*)'}{S_0(u;\hat{\theta})}du_i(\hat{\theta},\theta_0) \quad (14)$$

By the consistency of $\hat{\theta}$, we know that the last part of (14) consistently converges to 0. Hence $\hat{\Lambda}_0(t)$ in $t \in [0,\tau]$ consistently converges a.e. to $\Lambda_0(t)$.

For the weak convergence of $\hat{\Lambda}_0(t)$, we found that

$$n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} = n^{1/2}\{\hat{\Lambda}_0(t;\hat{\theta}) - \hat{\Lambda}_0(t;\theta_0)\} + n^{1/2}\{\hat{\Lambda}_0(t;\theta) - \Lambda_0(t)\} \quad (15)$$

By Taylor expansion of $\hat{\Lambda}_0(t)$ at θ_0, the first term on the right-hand side of (15) can be written as

$$n^{1/2}\{\hat{\Lambda}_0(t;\hat{\theta}) - \hat{\Lambda}_0(t;\theta_0)\} = \frac{\partial\hat{\Lambda}_0(t;\theta_0)}{\partial\theta'}n^{1/2}(\hat{\theta} - \theta_0) + o_p(n^{1/2}\|\hat{\theta} - \theta_0\|)$$

By lemma 1 of [7] and strong law of large numbers, we have a.e.

$$\sup_{0 \le t \le \tau} \| \frac{\partial \hat{\Lambda}_0(t; \theta_0)}{\partial \theta} + B(t; \theta_0) \| \to 0$$

where $B(t; \theta_0)$ is given by (10). By (1) and (13) we can know

$$n^{1/2}(\hat{\theta} - \theta_0) = A^{-1} n^{-1/2} \sum_{i=1}^{n} \int_0^t Y_i(u) \{ Z_i(u) - \bar{z}(u; \theta_0) \} dM_i(u; \theta_0) + o_p(1)$$

then, for t we consistently have

$$n^{1/2} \{ \hat{\Lambda}_0(t; \hat{\theta}) - \hat{\Lambda}_0(t; \theta_0) \} = -B(t; \theta_0)' A^{-1} n^{-1/2}$$
$$\sum_{i=1}^{n} \int_0^t Y_i(u) \{ Z_i(u) - \bar{z}(u; \theta_0) \} dM_i(u; \theta_0) + o_p(1) \tag{16}$$

On the other hand, for $0 \le t \le \tau$ we can get

$$n^{1/2} \{ \hat{\Lambda}_0(t; \theta_0) - \Lambda_0(t) \} = n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\partial M_i(u; \theta_0)}{S_0(u; \theta_0)}$$

Hence by the lemma 1 in [7], for t it consistently holds that

$$n^{1/2} \{ \hat{\Lambda}_0(t; \theta_0) - \Lambda_0(t) \} = n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{\partial M_i(u; \theta_0)}{s_0(u; \theta_0)} + o_p(1) \tag{17}$$

Finally, by (15)-(17), for t we have

$$n^{1/2} \{ \hat{\Lambda}_0(t; \hat{\theta}) - \Lambda_0(t) \} = n^{-1/2} \sum_{i=1}^{n} \Phi_i(t) + o_p(1) \tag{18}$$

By (15), for fixed t, $n^{1/2} \{ \hat{\Lambda}_0(t; \hat{\theta}) - \Lambda_0(t) \}$ asymptotically follows the sum of zero mean identically independently distributed random variables. By multivariate central limit theorem, $n^{1/2} \{ \hat{\Lambda}_0(t; \hat{\theta}) - \Lambda_0(t) \}$ weakly converges to a zero mean Gaussian process, and the covariance function at (s, t) is given by $\Gamma(s, t) = E\{ \Phi_i(s) \Phi_i(t) \}$. At the same time, consistent estimation of $\Gamma(s, t)$ is given by (11).

References

[1] Andersen, P.K. and Gill, R.D. (1982). Cox's regression model counting process: A large sample study, *Annals Of Statistics*, 10, 1100-1120.

[2] Dai J J, Sun L Q, Yang Z H. (2009). A general additive-multiplicative rates model for recurrent event data, *Sci China Ser A.*, 52, DOI:10.1007/s11425-009-0095-6.

[3] Liu H B, Miao R, Sun L Q. (2011). Analysis of panel data with informative observation and censoring times under biased sampling, *Sci Sin Math.*, 41(4), 365-376.

[4] Pepe, M.S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates, *Journal of the American Statistical Association*, 88, 811-820.

[5] Lawless, J F and Nadeau, C (1995). Some simple robust methods for the analysis of recurrent events, *Technometrics*, 37, 158-168.

[6] Lawless, J F, Nadeau, C and Cook R J. (1997). Analysis of mean and rate functions for recurrent events, *Proc 1st Seattle Symp Biostatistcs: Survival Analysis*. New York: Springer, 37-49.

[7] Lin D Y, Wei L J, Yang I, et al. (2000). Semiparametric regression for the mean and rate functions of recurrent events, *J Roy Statist Soc Ser B*, 62(4), 711-730.

[8] Maller, R. A. and Zhou, S. (1994). Testing for sufficient follow-up outliers in survival data, *Journal of the American Statistical Association*, 89, 1499-1506.

[9] Lai X. (2009). Extensions on long-term survivor model with random effects, University of Science and Technology of China.

[10] J.W. Boag (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy, *Journal of the Royal Statistical Society Series B (Methodological)*, 11(1), 15-53.

[11] Farewell VT. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics*, 38, 1041-1046.

[12] Kuk A.Y.C. and Chen C.H. (1992). A mixture model combining logistic regression with proportional hazards regression, *Biometrika*, 79, 531-541.

[13] Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models, *Biometrics*, 51, 899-907.

[14] Sy, J.P. and Taylor, J.M.G. (2000). Estimation in a Cox proportional hazards cure model, *Biometrics*, 56, 227-236.

[15] Cook RJ and Lawless JF (2007), *The Statistical Analysis of Recurrent Events*. Springer: New York, NY.

On Bias Reduction for the Total Least-Squares Estimate of a Conic Section Within an EIV-Model

Burkhard Schaffrin[1,*]   Kyle Snow[1,2,†]

[1]Geodetic Science Program, School of Earth Sciences, The Ohio State University, Columbus, Ohio, USA
[2]Topcon Positioning Systems, Inc., Columbus, Ohio, USA

Abstract   For a straight line in 2D, it is well known that the Total Least-Squares (TLS) estimator within an Errors-In-Variables (EIV) model generates the same solution as orthogonal regression, which turns out to be (locally) unbiased.

Unfortunately, however, the adjustment of a conic section (such as a parabola in 2D) will, if based on orthogonal regression, be biased within the standard EIV model. Here, an approach will be investigated where TLS estimation is applied to a modified EIV model in order to account for the bias that must otherwise be expected. The effect of such "bias reduction" will be illuminated in a realistic example from geodesy, taken from a popular textbook.

Keywords   Total Least-Squares; Errors-In-Variables modeling; orthogonal regression; bias reduction; conic-section fitting

1   Introduction

In an early paper, Pearson [1], had shown how best-fitting lines or planes can be obtained for point clouds in 2D or 3D with all their coordinates observed and thus random. For more general curves, Deming [2, 3] found a best-fit solution by applying iterative linearization to a nonlinear Gauss-Helmert Model (GHM). Such a solution became well-known under the label "orthogonal regression."

Much later, Linkwitz [4] showed that Pearson's straight-line fit problem can equivalently be described by an eigenvalue problem, thereby essentially anticipating the famous Total Least-Squares (TLS) approach by Golub and van Loan [5] who used the framework of an Errors-In-Variables (EIV) model. This equivalence between TLS adjustment and "orthogonal regression" was also confirmed by Schaffrin and Wieser [6] for the case of data with a fairly general dispersion matrix. More recently, Schaffrin [7] established more formally certain equivalencies between the EIV-Model adjustment and various more classical models, among them the nonlinear Gauss-Helmert Model (GHM).

Surprisingly, however, this equivalence seems to break down in the case of a relatively trivial generalization, namely the best fit of a conic section (such as a parabola) within the GHM versus a TLS adjustment within an EIV-Model, which turns out to contain a certain bias unless it is taken into account by modifying the model equation accordingly. This procedure will, however, lead to an adjusted curve with the residuals no longer being perpendicular, thus different from the "orthogonal regression" solution.

---

*Email: schaffrin.1@osu.edu
†Corresponding author. Email: ksnow6378137@gmail.com

In the following, a brief overview will be given, with many more details (and examples) to follow in a longer contribution.

2 Orthogonal regression for a parabola within a Gauss-Helmert Model

For a point cloud in 2D, the x-, resp. y-coordinate values may be collected in the $n \times 1$ vectors

$$\boldsymbol{x} := \boldsymbol{\mu}_x + \boldsymbol{e}_x \quad \text{and} \quad \boldsymbol{y} := \boldsymbol{\mu}_y + \boldsymbol{e}_y, \tag{1a}$$

where vectors $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ denote the "true" coordinates, and $\boldsymbol{e}_x$ and $\boldsymbol{e}_y$ are vectors of random errors with

$$\begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_x \end{bmatrix} \sim (\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \Sigma_y & 0 \\ 0 & \Sigma_x \end{bmatrix}). \tag{1b}$$

For the sake of simplicity, it is assumed that no correlations exist so that both Σ_x and Σ_y are (nonsingular) diagonal matrices, i.e.,

$$\Sigma_x := \text{Diag}(\boldsymbol{\sigma}_x^2) \quad \text{and} \quad \Sigma_y := \text{Diag}(\boldsymbol{\sigma}_y^2), \tag{1c}$$

where the $n \times 1$ vectors $\boldsymbol{\sigma}_x^2$ and $\boldsymbol{\sigma}_y^2$ contain the squared standard deviations (variances) of the random errors associated with the data vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

Then, for each $i \in \{1, \dots, n\}$, the parabola can be expressed as

$$y_i - e_{y_i} = \xi_0 + \xi_1 \cdot (x_i - e_{x_i}) + \xi_2 \cdot (x_i - e_{x_i})^2, \tag{2a}$$

and, after introducing the 3×1 parameter vector $\boldsymbol{\xi} := [\xi_0, \xi_1, \xi_2]^T$, in vector form as

$$\boldsymbol{b}(\boldsymbol{\xi}, \boldsymbol{e}_x, \boldsymbol{e}_y) := \boldsymbol{y} - \boldsymbol{e}_y - [\boldsymbol{\tau} \mid \boldsymbol{x} - \boldsymbol{e}_x \mid (\boldsymbol{x} - \boldsymbol{e}_x) * (\boldsymbol{x} - \boldsymbol{e}_x)] \cdot \boldsymbol{\xi} = \boldsymbol{0}, \tag{2b}$$

where $\boldsymbol{\tau} := [1, \dots, 1]^T$ denotes the $n \times 1$ "summation vector" and the symbol $*$ denotes the element-wise Hadamard product. Obviously, (2b) in conjunction with (1b) represent a nonlinear Gauss-Helmert Model (GHM) that can be solved by iterative linearization until convergence to the LEast-Squares Solution (LESS) occurs, using $P_x = \Sigma_x^{-1}$ and $P_y = \Sigma_y^{-1}$ as weight matrices. Such a procedure was recently reviewed by Schaffrin and Snow [8] in the context of Total Least-Squares regularization of Tykhonov type. As a result, the residuals will be P-orthogonal to the adjusted parabola.

3 The corresponding Errors-In-Variables (EIV) Model

In this case, equation (2b) will be given the form

$$\boldsymbol{y} - \boldsymbol{e}_y = (A - E_A)\boldsymbol{\xi} \tag{3a}$$

with

$$A := \begin{bmatrix} \boldsymbol{\tau} \mid \boldsymbol{x} \mid \boldsymbol{x} * \boldsymbol{x} \end{bmatrix} \tag{3b}$$

and

$$E_A := \begin{bmatrix} \boldsymbol{0} \mid \boldsymbol{e}_x \mid 2(\boldsymbol{e}_x * \boldsymbol{x}) - (\boldsymbol{e}_x * \boldsymbol{e}_x) \end{bmatrix}, \tag{3c}$$

assuming zero-expectation for E_A, and a *linear* error propagation for its covariance matrix, leading to

$$Q_A^{\text{approx}} := D\{\boldsymbol{e}_A := \text{vec}\, E_A\} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_x & 2 \cdot \text{Diag}(\boldsymbol{\mu}_x) \cdot \Sigma_x \\ 0 & 2 \cdot \text{Diag}(\boldsymbol{\mu}_x) \cdot \Sigma_x & 4 \cdot \text{Diag}(\boldsymbol{\mu}_x * \boldsymbol{\mu}_x) \cdot \Sigma_x \end{bmatrix} \tag{4}$$

as proposed by Hu et al. [9] for a related application. Here, "vec" denotes the vectorial form of the $n \times 3$ matrix E_A, so that $\boldsymbol{e}_A$ has size $3n \times 1$.

As it turns out, the resulting Total Least-Squares (TLS) solution has to use a singular dispersion matrix, a case first described by Snow [10]. However, it will not coincide with the "orthogonal regression" solution of section 2. So far, it is still unknown whether "orthogonal regression" can be achieved in model (3a)–(3c) by a modified choice of the dispersion matrix (4) (which is not exact anyway).

4 Modifying the EIV-Model for bias reduction

Not only is (4) only a first-order approximation, but also the assumption of zero-expectation for E_A is invalid, as a quick analysis, under the assumption of normality, shows:

$$E\{\boldsymbol{e}_A = \operatorname{vec} E_A\} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ 2 \cdot E\{\boldsymbol{e}_x * (\boldsymbol{\mu}_x + \boldsymbol{e}_x)\} - E\{\boldsymbol{e}_x * \boldsymbol{e}_x\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ E\{\boldsymbol{e}_x * \boldsymbol{e}_x\} = \operatorname{vecd} \Sigma_x \end{bmatrix},$$
(5a)

where "vecd" denotes the vector that collects the diagonal elements of a matrix. In addition, the dispersion matrix results in its exact form:

$$Q_A := D\{\boldsymbol{e}_A := \operatorname{vec} E_A\} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_x & 2 \cdot \operatorname{Diag}(\boldsymbol{\mu}_x) \cdot \Sigma_x \\ 0 & 2 \cdot \operatorname{Diag}(\boldsymbol{\mu}_x) \cdot \Sigma_x & 4 \cdot \operatorname{Diag}(\boldsymbol{\mu}_x * \boldsymbol{\mu}_x) \cdot \Sigma_x + 2 \cdot \Sigma_x^2 \end{bmatrix},$$
(5b)

where the block in the lower-right corner differs from (4). But the most serious fact here is that a bias is caused by not taking the nonzero-expectation from (5a) into account. This can be done most easily by modifying the third column of the matrix A so that

$$A_{\text{mod}} := \begin{bmatrix} \boldsymbol{\tau} \mid \boldsymbol{x} \mid (\boldsymbol{x} * \boldsymbol{x}) - \Sigma_x \cdot \boldsymbol{\tau} \end{bmatrix}$$
(6a)

is used instead of (3b), and

$$(E_A)_{\text{mod}} := \begin{bmatrix} \boldsymbol{0} \mid \boldsymbol{e}_x \mid 2(\boldsymbol{e}_x * \boldsymbol{x}) - (\boldsymbol{e}_x * \boldsymbol{e}_x) - \Sigma_x \cdot \boldsymbol{\tau} \end{bmatrix},$$
(6b)

in lieu of (3c). In the *modified EIV-Model*, due to (6a)–(6b), the target function for Total-Least Squares adjustment is now based on quadratic forms of both $\boldsymbol{e}_y$ and $\boldsymbol{e}_A$, and not on those of $\boldsymbol{e}_y$ and $\boldsymbol{e}_x$ as in the GHM; therefore, the Total Sum of Squared Residuals (TSSR) may not be directly comparable as it does not necessarily express the quality of fit in the same way. In any case, it cannot be expected that the new residuals are P-orthogonal to the adjusted parabola.

5 Conclusions

In the extended version of this contribution, two examples will be investigated that use data from Neri et al. [11] and from Mikhail and Gracie [12], respectively. Numerical comparisons will be made for the above described three models in regard to P-orthogonality of the residuals, the TSSR in the form of $\Omega = \boldsymbol{e}_x^T P_x \boldsymbol{e}_x + \boldsymbol{e}_y^T P_y \boldsymbol{e}_y$, and the reduction of bias as a consequence of using the modified EIV-Model.

Acknowledgement

References

[1] Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space, *Philosophical Magazine Series 6*, 2:11, 559–572, Taylor & Francis.

[2] Deming, W. E. (1931). XI. The application of least squares , *Philosophical Magazine Series 7*, 11:68, 146–158, Taylor & Francis.

[3] Deming, W. E. (1934). LXVII. On the application of least squares II, *Philosophical Magazine Series 7*, 17:114, 804–829, Taylor & Francis.

[4] Linkwitz, K. (1976). On certain adjustment problems and their solution by means of matrix eigenvalues (in German) in: Festschrift Ernst Gotthardt, *Publ. of the German Geodetic Commission (DGK)*, Vol. B-216, Munich, pp. 111–126.

[5] Golub, G. H. and van Loan, C. F. (1980). An analysis of the total least squares problem, *SIAM Journal on Numerical Analysis*, 17:6, 883–893.

[6] Schaffrin, B. and Wieser, A. (2008). On weighted total least-squares adjustment for linear regression, *J. of Geodesy*, 82:7, 415–421.

[7] Schaffrin, B. (2015). Adjusting the Errors-In-Variables Model: Linearized Least Squares vs. nonlinear Total Least-Squares procedures, Proc. of the VIII Hotine-Marussi Symp. on Math. Geodesy (Rome/Italy 2013), *Springer IAG Symp. Series*, forthcoming.

[8] Schaffrin, B. and Snow, K. (2010). Total Least-Squares regularization of Tykhonov type and an ancient racetrack in Corinth, *Linear Algebra and its Applications*, 432:8, 2061–2076, Elsevier Inc.

[9] Hu, C., Chen Y, and Peng Y. (2014). On weighted total least squares adjustment for solving the nonlinear problems, *J. of Geodetic Sci.*, 4:1, 49–56.

[10] Snow, K. (2012). *Topics in Total Least-Squares Adjustment within the Errors-In-Variables Model: Singular Cofactor Matrices and Prior Information*, Report No. 502, Div. of Geodetic Science, School of Earth Sciences, The Ohio State Univ., Columbus, Ohio, USA.

[11] Neri, F., Saitta, G., and Chiofalo, S. (1989). An accurate and straightforward approach to line regression analysis of error-affected experimental data, *J. Phys. E: Sci. Instrum.*, 22, 215–217, IOP Publishing, Ltd.

[12] Mikhail, E. M. and Gracie, G. (1981). *Analysis & Adjustment of Survey Measurements*, New York, Van Nostrand Reinhold.

An illustrated philatelic introduction to doubly-classic $6{\times}6$ bordered magic matrices and to $4{\times}4$ Plato-like magic talismans

George P. H. Styan[1,*]   Ka Lok Chu[2,†]

[1]McGill University, Montréal (Québec), Canada
[2]Dawson College, Westmount (Québec), Canada

Abstract   Our motivation in this talk is the 13th-century Anxi iron-plate doubly-classic $6{\times}6$ bordered magic square discussed by Kai-Tai Fang in [5] at the 22nd International Workshop on Matrices and Statistics in Toronto (IWMS-2013), and which we define by the matrix **A** as in display (2). We also study the magic square in the "16th-century magical square in a manuscript" given on the "Bedouin Silver: magical squares" website, and which we define by the matrix **B** as in display (3).

In his "Bordered magic squares" website, Harry White [23] observed that for such magic squares, in all there are 140 "border groups". Of these 140 we found 93 displayed in the literature. We identify all 140 and find that they occur in 70 pairs, with the matrices **A** and **B** forming one such pair. Moreover, we find that the rank depends only on the rank of the inner 4×4 heart submatrix **H**, which is Plato-like in that $\mathbf{H} = \mathbf{P} + h\mathbf{E}$, where **P** is the well-known Plato magic matrix and **E** has every entry equal to 1; here $h = 10$. We end our talk by displaying some seals and talismans with Plato-like magic squares.

Keywords   Anxi iron-plate doubly-classic 6×6 bordered magic square, magic matrices
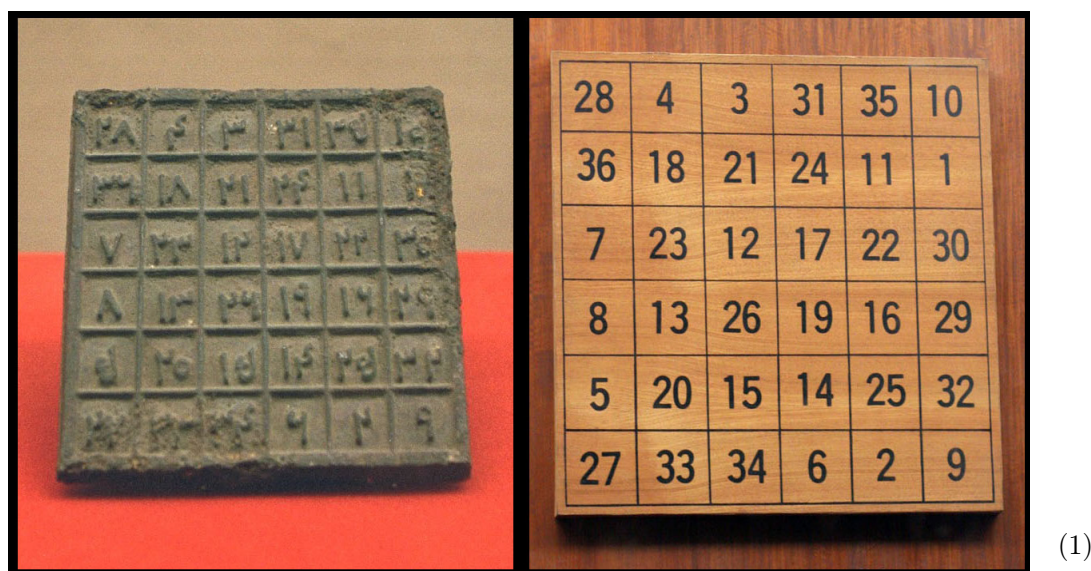
(1)

1 Anxi iron-plate and Bedouin-silver bordered magic squares

Our motivation in this talk is the 13th-century Anxi iron-plate doubly-classic 6×6 bordered magic square discussed by Kai-Tai Fang [5] at the 22nd International Workshop on Matrices and Statistics in Toronto (IWMS-2013). The Anxi iron-plate bordered 6×6 magic square as in display (1) from Shea [18] and discussed by Baocong Qian [3, (1966/1985)] and by Ho Peng Yoke [12, (1985)]. Our results build also on those by Hendricks [10, 11, (1991/2005)], Saenz de Cabezon [17, (2003/2011)], and White [23, (2009/2015)]; see also Stifel [21, (1544)] and van Roode [14, 15, (2009/2011)], [16, 2010]. See also Styan & Chu [22, (2015)].

We define the Anxi iron-plate bordered 6×6 magic square, as shown in display (1), by the Anxi iron-plate bordered 6×6 magic matrix

$$
\mathbf{A} = \begin{pmatrix}
28 & 4 & 3 & 31 & 35 & 10 \\
36 & 18 & 21 & 24 & 11 & 1 \\
7 & 23 & 12 & 17 & 22 & 30 \\
8 & 13 & 26 & 19 & 16 & 29 \\
5 & 20 & 15 & 14 & 25 & 32 \\
27 & 33 & 34 & 6 & 2 & 9
\end{pmatrix}. \tag{2}
$$

We understand from Baidu [1] that the original Anxi iron-plate bordered 6×6 magic square is in the Shaanxi History Museum in Xi'an, China, where it is called the "Yuan magic square iron Atlas". Xi'an is the capital of Shaanxi province.

From *Song Yuan shuxue shi lunwen ji* [*Collected Theses on the History of Chinese Mathematics in the Song and Yuan Dynasties*], by Baocong Qian [3, p. 264 (1966/1985)] we find that "According to the book *Mi Shu Jian Zhi*, the astronomer Zhamaluding [Jamāl al-Dīn (fl. c. 1255–1291)] calculated the calendar for the King of An Xi in 1278, with three accompanying astronomers. These four astronomers apparently manufactured the iron plate with the magic square on it. The iron plate was probably used to exorcise town demons and evil spirits. At that time, Arabic numbers were not in widespread use and had not been widely applied." See also Ho Peng Yoke [12, (1985)].

Another bordered 6×6 magic square was given by Sigrid van Roode as a "16th-century magical square in a manuscript" (Fig. 1 below) in her "Bedouin Silver: magical squares" website [14, (2009/2011)]. We have found no other presentation of this "16th-century magical square" in the literature.

We believe some entries are typos (see Fig. 2 below), and adjusting for these yields our bordered 6×6 Bedouin-silver magic matrix

$$
\mathbf{B} = \begin{pmatrix}
2 & 31 & 33 & 34 & 10 & 1 \\
29 & 18 & 21 & 24 & 11 & 8 \\
30 & 23 & 12 & 17 & 22 & 7 \\
5 & 13 & 26 & 19 & 16 & 32 \\
9 & 20 & 15 & 14 & 25 & 28 \\
36 & 6 & 4 & 3 & 27 & 35
\end{pmatrix}. \tag{3}
$$

٢	٣١	٣٣	٣٤	١٠	١
٢٩	١٨	٢١	٢٤	١١	٨
٣	١٣	١٢	١٧	٢٢	٧
٣	١٣	٢٦	٢٩	١٦	٣٢
٩	٢	١٩	١٤	٢٥	٢٨
٣٥	٦	٤	٣	٢٧	٣٦

Figure 1: "16th century magical square in a manuscript" [14]. See also Fig. 2.

Figure 2: Annotated detail from "16th century magical square in a manuscript" (Fig. 1).

We recall the Anxi iron-plate bordered 6×6 magic matrix $\mathbf{A}$ as introduced in (2) above

$$\mathbf{A} = \begin{pmatrix} 28 & 4 & 3 & 31 & 35 & 10 \\ 36 & 18 & 21 & 24 & 11 & 1 \\ 7 & 23 & 12 & 17 & 22 & 30 \\ 8 & 13 & 26 & 19 & 16 & 29 \\ 5 & 20 & 15 & 14 & 25 & 32 \\ 27 & 33 & 34 & 6 & 2 & 9 \end{pmatrix} \tag{4}$$

and we see that both the Anxi iron-plate matrix $\mathbf{A}$ in (4) and the Bedouin-silver matrix $\mathbf{B}$ introduced in (3) above are "doubly-classic" bordered magic matrices in that the 36 entries are the 36 consecutive integers $1, 2, \ldots, 36$ and in addition, the 16 entries in the 4×4 "heart" magic matrix (inner 4×4 submatrix) are the 16 consecutive integers $11, 12, \ldots, 26$, and so the entries in the "frame" (or "border") are the 10 integers $1, 2, \ldots, 10$ and the 10 integers $27, 28, \ldots, 36$. See also (6) below.

2 Plato magic matrtix and the Shanghai Museum

Figure 3: Plato magic matrix on a jade hanging and the Shanghai Museum.
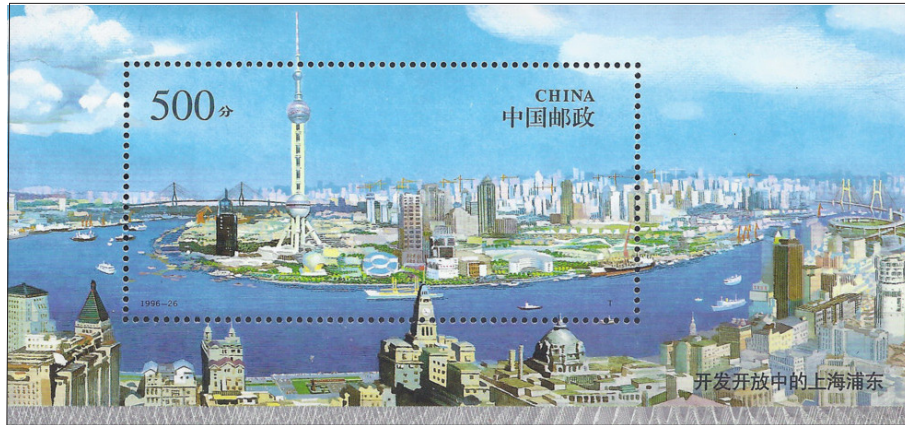
Figure 4: Shanghai World Financial Center.

The Shanghai Museum, while excavating the Ming-dynasty tomb relics in 1980 from the Shanghai Pudong Lujiazui (see Fig. 4), discovered a piece of jade (Fig. 3, left panel), with a magic square as defined by our Plato magic matrix

$$\mathbf{P} = \begin{pmatrix} 8 & 11 & 14 & 1 \\ 13 & 2 & 7 & 12 \\ 3 & 16 & 9 & 6 \\ 10 & 5 & 4 & 15 \end{pmatrix}. \tag{5}$$

We found several instances of the Plato matrix **P** as defined in (5) above in the seminal book *Qanoon-e-Islam* [19] first published in 1832. Moreover,

Cammann [4, Part 1, p. 202 (1969)] says that

with the possible exception of the magic square of [order] three [our Plato matrix **P** (5)] became the most popular and best-known magic square in the Islamic world, while

Hayashi [8, p. 166 (1987)] found [our Plato matrix **P** (5)] to be

the famous pandiagonal magic square of order four in Islam.

In addition, Cammann [4, Part 1, p. 202 (1969)] observed that

The ascription of magic squares to famous figures of the Classical World was very common. al-Qazvini (d. 1283 AD) said that Archimedes invented magic squares, and the well-known Sufi writer on mathematics Ahmad ibn 'Ali ibn Yusuf al-Buni (d. 1225) ascribed the principal Islamic magic square of four [our **P** (5)] to the philosopher and mathematician Plato (c. 428–348 BC). We have no evidence to show that the classical Greeks had magic squares.

Abu Yahya Zakariya'ibn Muhammad al-Qazwini or Kazwini (1203–1283) was an Arab or Persian physician, astronomer, geographer and proto-science fiction writer. Qazvini's famous *Cosmography* entitled *Marvels of Creatures and the Strange Things Existing* was immensely popular and is preserved today in many copies. It was translated into his native Persian language, and later also into Turkish. For some comments on magic squares in "Kazwini's *Cosmography*" see Baksalary, Chu, Puntanen & Styan [2, (2009)].

The Bedouin-silver matrix **B** defined in our (3) and the Anxi iron-plate **A** defined in our (2) have the same "heart matrix" (6)

$$\mathbf{H} = \begin{pmatrix} 18 & 21 & 24 & 11 \\ 23 & 12 & 17 & 22 \\ 13 & 26 & 19 & 16 \\ 20 & 15 & 14 & 25 \end{pmatrix} = \mathbf{P} + 10\mathbf{E} = \begin{pmatrix} 8 & 11 & 14 & 1 \\ 13 & 2 & 7 & 12 \\ 3 & 16 & 9 & 6 \\ 10 & 5 & 4 & 15 \end{pmatrix} + 10\mathbf{E} \qquad (6)$$

where **P** is the so-called "Plato magic matrix" introduced in (5) above and **E** here is the 4×4 matrix with every entry equal to 1.

We will say that the heart matrix **H** defined in our (6) is Plato-like Type E. More generally, we say that a magic matrix **M** is "Plato-like Type E" whenever

$$\mathbf{M} = \mathbf{P} + k\mathbf{E} \qquad (7)$$

for some nonzero scalar k.

Figure 5: This sheetlet from Israel featuring Plato and Aristotle (detail from Raphael's *School of Athens*) was apparently printed on 21.1.2009 (top right corner) but released only in 2015.

In a diagram for a "*puleeta* (or lamp-charm) for casting out devils" in Shurreef's *Qanoon-e-Islam* [19, facing p. 224 (1832)] we found the Plato magic matrix $\mathbf{P}$, as well as the Plato-like Type E magic matrix

$$\begin{pmatrix} 17 & 20 & 23 & 10 \\ 22 & 11 & 16 & 21 \\ 12 & 25 & 18 & 15 \\ 19 & 24 & 13 & 24 \end{pmatrix} = \mathbf{P} + 9\mathbf{E}. \tag{8}$$

3 Doubly-classic bordered 6×6 magic matrices

3.1 Hendricks-normalized form and 10 small numbers

Following Hendricks [11] we define the "Hendricks-normalized" form for a doubly-classic bordered 6×6 magic matrix to be

$$
\mathbf{N} = \begin{pmatrix}
u & a_1 & \overline{a}_4 & \overline{a}_3 & \overline{a}_2 & v \\
b_1 & h_{11} & h_{12} & h_{13} & h_{14} & \overline{b}_1 \\
b_2 & h_{21} & h_{22} & h_{23} & h_{24} & \overline{b}_2 \\
\overline{b}_4 & h_{31} & h_{32} & h_{33} & h_{34} & b_4 \\
\overline{b}_3 & h_{41} & h_{42} & h_{43} & h_{44} & b_3 \\
\overline{v} & \overline{a}_1 & a_4 & a_3 & a_2 & \overline{u}
\end{pmatrix}
\tag{9}
$$

where the 10 "small numbers"

$$
a_1, \ a_2, \ a_3, \ a_4, \ b_1, \ b_2, \ b_3, \ b_4, \ u, \ v
\tag{10}
$$

are an arrangement of the first 10 positive integers 1, 2, $\ldots$, 10 with

$$
a_2 < a_3 < a_4, \quad b_1 < b_2, \quad b_3 < b_4, \quad u < v.
\tag{11}
$$

In (9) the overline denotes the complement from 37 and so

$$
\overline{x} = 37 - x; \quad x = a_1, \ a_2, \ a_3, \ a_4, \ b_1, \ b_2, \ b_3, \ b_4, \ u, \ v.
\tag{12}
$$

The heart matrix $\mathbf{H}$ in a doubly-classic bordered 6×6 fully-magic matrix (such as the Anxi iron-plate matrix $\mathbf{A}$ (2) and the Bedouin-silver matrix $\mathbf{B}$ (3)) is $10\mathbf{E}$ plus any classic 4×4 magic matrix, for which there are 8×880 choices. The corner entry u in the frame of $\mathbf{N}$ (9) may be placed in any one of the 4 corners and then $\overline{u} = 37 - u$ diagonally opposite; the corner entry v may then be placed in any one of 2 corners with $\overline{v} = 37 - v$ diagonally opposite. The non-corner entries $a_1, \overline{a}_4, \overline{a}_3, \overline{a}_2$ between u and v may be placed in $4! = 24$ different ways and similarly for the non-corner entries $b_1, b_2, \overline{b}_4, \overline{b}_3$ between u and $\overline{v}$. The remaining non-corner entries follow at once. We find, therefore, $8 \times 880 \times 4 \times 2 \times 24 \times 24 = 32,440,320$ possible doubly-classic bordered 6×6 magic matrices for each Hendricks-normalized form frame. There are 140 such frames and hence $140 \times 32,440,320 = 4,541,644,800$ possible doubly-classic bordered 6×6 magic matrices in all.

3.2 Anxi iron-plate A and Bedouin-silver B are partners

The Hendricks-normalized form for the Anxi iron-plate bordered 6×6 magic matrix $\mathbf{A}$ is

$$
\mathbf{N}^{(A)} = \begin{pmatrix}
9 & 1 & 29 & 30 & 32 & 10 \\
2 & h_{11} & h_{12} & h_{13} & h_{14} & 35 \\
6 & h_{21} & h_{22} & h_{23} & h_{24} & 31 \\
33 & h_{31} & h_{32} & h_{33} & h_{34} & 4 \\
34 & h_{41} & h_{42} & h_{43} & h_{44} & 3 \\
27 & 36 & 8 & 7 & 5 & 28
\end{pmatrix}
\tag{13}
$$

with

$$
a_1 = 1,\ a_2 = 5,\ a_3 = 7,\ a_4 = 8; \quad b_1 = 2,\ b_2 = 6;\ b_3 = 3,\ b_4 = 4; \quad u = 9,\ v = 10, \tag{14}
$$

while for the Bedouin-silver magic matrix $\mathbf{B}$ and the Hendricks-normalized form $\mathbf{N}^{(B)}$ are

$$
\mathbf{B} = \begin{pmatrix}
2 & 31 & 33 & 34 & 10 & 1 \\
29 & 18 & 21 & 24 & 11 & 8 \\
30 & 23 & 12 & 17 & 22 & 7 \\
5 & 13 & 26 & 19 & 16 & 32 \\
9 & 20 & 15 & 14 & 25 & 28 \\
36 & 6 & 4 & 3 & 27 & 35
\end{pmatrix}, \quad
\mathbf{N}^{(B)} = \begin{pmatrix}
1 & 10 & 31 & 33 & 34 & 2 \\
7 & h_{11} & h_{12} & h_{13} & h_{14} & 30 \\
8 & h_{21} & h_{22} & h_{23} & h_{24} & 29 \\
28 & h_{31} & h_{32} & h_{33} & h_{34} & 9 \\
32 & h_{41} & h_{42} & h_{43} & h_{44} & 5 \\
35 & 27 & 6 & 4 & 3 & 36
\end{pmatrix}
\tag{15}
$$

with

$$
a_1 = 10,\ a_2 = 3,\ a_3 = 4,\ a_4 = 6; \quad b_1 = 7,\ b_2 = 8;\ b_3 = 5,\ b_4 = 9; \quad u = 1,\ v = 2. \tag{16}
$$

At first glance the two matrices

$$
\mathbf{N}^{(A)} = \begin{pmatrix}
9 & 1 & 29 & 30 & 32 & 10 \\
2 & h_{11} & h_{12} & h_{13} & h_{14} & 35 \\
6 & h_{21} & h_{22} & h_{23} & h_{24} & 31 \\
33 & h_{31} & h_{32} & h_{33} & h_{34} & 4 \\
34 & h_{41} & h_{42} & h_{43} & h_{44} & 3 \\
27 & 36 & 8 & 7 & 5 & 28
\end{pmatrix}, \quad
\mathbf{N}^{(B)} = \begin{pmatrix}
1 & 10 & 31 & 33 & 34 & 2 \\
7 & h_{11} & h_{12} & h_{13} & h_{14} & 30 \\
8 & h_{21} & h_{22} & h_{23} & h_{24} & 29 \\
28 & h_{31} & h_{32} & h_{33} & h_{34} & 9 \\
32 & h_{41} & h_{42} & h_{43} & h_{44} & 5 \\
35 & 27 & 6 & 4 & 3 & 36
\end{pmatrix}
\tag{17}
$$

appear to be quite different, even though their parents have the same heart $\mathbf{H} = \mathbf{P} + 10\mathbf{E}$.

We find, however, that the Anxi iron-plate matrix $\mathbf{A}$ (2) and the Bedouin-silver matrix $\mathbf{B}$ (3)) are "partners" in that

$$u^{(B)} = \bar{\bar{v}}^{(A)}, \quad v^{(B)} = \bar{\bar{u}}^{(A)} \tag{18}$$

$$a_1^{(B)} = \bar{\bar{a}}_1^{(A)}, \quad a_2^{(B)} = \bar{\bar{a}}_4^{(A)}, \quad a_3^{(B)} = \bar{\bar{a}}_3^{(A)}, \quad a_4^{(B)} = \bar{\bar{a}}_2^{(A)} \tag{19}$$

$$b_1^{(B)} = \bar{\bar{b}}_4^{(A)}, \quad b_2^{(B)} = \bar{\bar{b}}_3^{(A)}, \quad b_3^{(B)} = \bar{\bar{b}}_2^{(A)}, \quad b_4^{(B)} = \bar{\bar{b}}_1^{(A)} \tag{20}$$

where the double-overbar indicates subtraction from 11, e.g., for (18)

$$u^{(B)} = 1 = 11 - 10 = \bar{\bar{v}}^{(A)}, \quad v^{(B)} = 2 = 11 - 9 = \bar{\bar{u}}^{(A)}. \tag{21}$$

We find that there are 70 such partner-pairs making a total of 140 doubly-classic bordered 6×6 magic matrices. In particular we find that the Anxi matrix $\mathbf{A}$ (our ID G-139) and the Bedouin matrix $\mathbf{B}$ (our ID G-022) form such a partner-pair. Let $w = \max(u+v, 22-u-v)$. Then

ID	matrix	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	u	v	w
G-139	Anxi $\mathbf{A}$	1	5	7	8	2	6	3	4	9	10	19
G-022	Bedouin $\mathbf{B}$	10	3	4	6	7	8	5	9	1	2	19
G-131		5	1	9	10	2	6	3	4	7	8	15
G-071		6	1	2	10	7	8	5	9	3	4	15
G-140		2	6	7	8	3	4	1	5	9	10	19
G-019		9	3	4	5	6	10	7	8	1	2	19
G-133		6	2	9	10	3	4	1	5	7	8	15
G-068		5	1	2	9	6	10	7	8	3	4	15
G-118		8	2	9	10	3	4	1	5	6	7	13
G-094		3	1	2	9	6	10	7	8	4	5	13

$$(22)$$

This confirms that the Anxi matrix $\mathbf{A}$ (G-139) and the Bedouin matrix $\mathbf{B}$ (G-022) are partners and together form a partner-pair. In addition we find that there is another partner-pair (G-131, G-071) with the same sets of b_1, b_2, b_3, b_4; in fact there is no other such partner-pair. But we do find three partner-pairs with coincident b_1, b_2, b_3, b_4 though we found no other such set of three partner-pairs. The remaining 130 (= 140 - 10, listed in (22)), doubly-classic bordered 6×6 magic matrices comprise 14 double partner-pairs (56 = $14 \times 2 \times 2$ matrices) and 37 single partner-pairs (74 = 37×2 matrices).

Figure 6: Quartz-rock crystal two-sided doubly Plato-like magic seal.

4   Further results

In our report [22] we present several more results concerning the 140 doubly-classic bordered 6×6 magic matrices. In particular we prove that the rank is always 2 more than the rank of the heart matrix. Moreover, we present images of various seals and talismans with the Plato or with a Plato-like magic square. A discussion of the so-called [7] auspicious number set 2 4 6 8 is also given and its use in the expeditious mailing of letters in the early 19th century [20].

In her beautiful book on *Arabic and Persian Seals and Amulets in the British Museum* [13], Venetia Porter presents several seals and amulets with magic squares. In particular, she presents images [13, Fig. A117, p. 167] of both sides of a quartz-rock crystal two-sided doubly Plato-like magic seal in The British Museum, see our Fig. 6 above: "Oval, flattish with bevelled sides. Damaged in places. Engraved both sides with a 4×4 magic square."

Acknowledgements

References

[1] Baidu (2012). Yuan magic square iron Atlas, by Baidu, a Chinese web services company headquartered in the Baidu Campus in Haidian District, Beijing [for the Shaanxi History Museum, Xi'an], last updated (25 April 2012).

[2] Baksalary, Oskar Maria; Chu, Ka Lok; Puntanen, Simo & Styan, George P. H. (2009). Some comments on Fisher's α index of diversity and on the *Kazwini Cosmography*", in *Statistical Inference, Econometric Analysis and Matrix Algebra: Festschrift in Honour of Götz Trenkler,* edited by Bernhard Schipp & Walter Krämer, pub. Physica-Verlag, Heidelberg, pp. 369–394.

[3] Baocong Qian (1966/1985). *Song Yuan shuxue shi lunwen ji = Collected Theses on the History of Chinese Mathematics in the Song and Yuan Dynasties,* Kexue chubanshe (1966), 2nd edition (1985).

[4] Cammann, Schuyler (1969). "Islamic and Indian magic squares, Part I. *History of Religions,* vol. 8, no. 3, pp. 181–209.

[5] Fang, Kai-Tai & Zheng, Yanxun (2013). The magic square: historical review and recent development". Keynote address presented (by Kai-Tai Fang on 12 August 2013) at the 22nd International Workshop on Matrices and Statistics (IWMS-2013), Fields Institute for Research in Mathematical Sciences, Toronto (Ontario), Canada.

[6] `fqjdwig` (2009). About magic squares. Appendix: wonderful magic square. `fqjdwig`'s blog. Publication online at NetEase, 10 pp (12 November 2009).

[7] Gallop, Annabel T. (2013). The amuletic cult of Ma'rūf al-Karḫī in the Malay world, *Writings and Writing: Investigations in Islamic Text and Script in Honour of Januarius Justus Witkam,* edited by Robert M. Kerr & Thomas Milo, pub. Archetype, Cambridge, pp. 167–196.

[8] Hayashi, Takao (1987). Varāhamihira's pandiagonal magic square of order four'. *Historia Mathematica,* vol. 14, no. 2, pp. 159–166.

[9] Heinz, Harvey D. (2011). *Magic Hypercube Omnibus: All about magic squares, cubes, tesseracts, etc.* HDH Publishing, Surrey BC.

[10] Hendricks, John R. (1991/1992). Catalogue of borders, pp. 90–93 in *The Magic Square Course* by John R. Hendricks.

[11] Hendricks, John R. (2005). Magic square borders. Unpublished 16 pp, with copy in [9].

[12] Ho Peng Yoke (1985). *Li, Qi and Shu: An Introduction to Science and Civilization in China.* Hong Kong University Press.

[13] Porter, Venetia (2011). *Arabic and Persian Seals and Amulets in the British Museum,* British Museum Research Publication (Book 160). British Museum Press, London (15 January 2011).

[14] van Roode, Sigrid (2009/2011). Bedouin Silver: magical squares: online at `bedouinsilver.com`

[15] van Roode, Sigrid (2011). Magical squares in jewelry from the Middle East: online at `artwis.com`

[16] van Roode, Sigrid (2011). *Desert Silver: Nomadic and Traditional Silver Jewellery [sic] from the Middle East and North Africa.* KIT Publishers, Amsterdam.

[17] Saenz de Cabezon, Eduardo (2011). Bordered magic squares: elements for a comprehensive approach. *Ars Combinatoria,* vol. 100 [Volume C], pp. 205–224.

[18] Shea, Marilyn (2010). Yuan Dynasty: Double Six Magic Square, online in the Department of Psychology, University of Maine at Farmington. Farmington, Maine: Last update (March 2010).

[19] Shurreef, Jaffur (1832). *Qanoon-e-Islam, or The Customs of the Moosulmans of India; Comprising a Full and Exact Account of Their Various Rites and Ceremonies, from the Moment of Birth Till the Hour of Death,* by Jaffur Shurreef [Ja'far Sharif] (A Native of the Deccan), composed under the direction of, and translated by G. A. Herklots, pub. Parbury, Allen & Co., London (1832).

[20] Baron Silvestre de Sacy (1827). *Chrestomathie arabe, ou Extraits de divers écrivains arabes, tant en prose qu'en vers, avec une traduction française et des notes, A l'usage des Élèves de l'École royale et spéciale des Langues orientales vivantes,* Tome III, seconde édition, corrigée et augmentée, imprimé par autorisation du roi, à l'Imprimerie royale [Paris].

[21] Stifel, Michael (1544). *Arithmetica Integra,* with a foreword by Philippi Melanchthonis, in Latin, pub. Norimbergæ apud Iohan. Petreium [Nürnberg].

[22] Styan, George P. H. (2015). An illustrated philatelic introduction to doubly-classic 6×6 bordered magic matrices and to 4×4 Plato-like magic talismans. In preparation, Report 2015-01 from the Department of Mathematics and Statistics, McGill University, Montréal.

[23] White, Harry (2009/2015). Bordered magic squares, online at
`http://users.eastlink.ca/~harrywhite/BorderedMagicSquares.html` and
`http://users.eastlink.ca/~harrywhite/howMany.html`

Professor Yanai and Multivariate Analysis

Yoshio Takane[1],[*]

[1]Department of Psychology, University of Victoria, British Columbia, Canada

Abstract Late Professor Yanai has contributed to many fields ranging from aptitude diagnostics, epidemiology, and nursing to psychometrics and statistics. This paper reviews some of his accomplishments in multivariate analysis through his collaborative work with the present author, along with some untold episodes for the inception of key ideas underlying the work. The various topics covered include constrained principal component analysis, extensions of Khatri's lemma, the Wedderburn-Guttman theorem, ridge operators, decompositions of the total association between two sets of variables, and ideal instruments. A common thread running through all of them is projectors and singular value decomposition (SVD), which are the main subject matters of a recent monograph by Yanai, Takeuchi, and Takane [35].

Keywords Projectors: Singular value decomposition (SVD); Constrained principal component analysis (CPCA); Khatri's lemma; The Wedderburn-Guttman theorem; Ridge operators; Generalized constrained canonical correlation analysis; Confounding variables; Propensity scores; Instrumental variables

1   Introduction

Professor Yanai passed away due to prostate cancer in December, 2013 at the age of 73. A quick glance at his home page reveals that his contributions extend over 7 broad categories, including aptitude diagnostics, test theories, educational psychology, epidemiology, nursing, linear algebra, statistics, and multivariate analysis (MVA). Here we focus on his contributions in the last category, namely multivariate analysis, through his collaborative works with me. Professor Yanai has been the most influential person in my career. In particular, if I had not met him when I was in the third year of college, I would not have been a statistician. We have 15 joint publications, including two books one in English [35] and one other in Japanese. The specific topics we cover today are:

(1) Constrained principal component analysis (CPCA)
(2) Khatri's lemma
(3) The Wedderburn-Guttman theorem
(4) Ridge operators
(5) Generalized constrained canonical correlation analysis
(6) Causal inference

Professor Yanai's idea about MVA can be most succinctly summarized as "partitioning the space of dimensionality n (the number of cases) into meaningful subspaces" identified by some external information or by some internal criterion (Takeuchi, Yanai, and Mukherjee [27]). Two major tools for partitioning are:

(1) Projectors

---

[*]Corresponding author. Email: yoshio.takane@mcgill.ca

(2) Singular value decomposition (SVD)

which are the main subject matters of a recent monograph by Yanai, Takeuchi, and Takane
[35]. As is well known, projectors are used to partition the space of observation vectors on
criterion variables into subspaces that can and cannot be explained by predictor variables,
and SVD seeks to find the subspace most representative of the original subspace.

Before we begin, let us introduce some basic notations we use throughout this paper:
Let $\text{Sp}(\mathbf{X})$ denote the space spanned by column vectors of $\mathbf{X}$, and let $\text{Ker}(\mathbf{X}')$ denote the
orthogonal complement subspace to $\text{Sp}(\mathbf{X})$. Let

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' \tag{1}$$

denote the orthogonal projector onto $\text{Sp}(\mathbf{X})$, and let

$$\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X \tag{2}$$

denote the orthogonal projectors onto $\text{Ker}(\mathbf{X}')$. Then,

$\mathbf{P}'_X = \mathbf{P}_X$, $\mathbf{Q}'_X = \mathbf{Q}_X$ (symmetric).
$\mathbf{P}^2_X = \mathbf{P}_X$, $\mathbf{Q}^2_X = \mathbf{Q}_X$ (idempotent).
$\mathbf{P}_X\mathbf{Q}_X = \mathbf{Q}_X\mathbf{P}_X = \mathbf{O}$ (orthogonal).

These projectors are useful in partitioning $\mathbf{y}$, the vector of observations on the dependent
variable in regression analysis, into $\mathbf{P}_X\mathbf{y}$, the portions of $\mathbf{y}$ that can be accounted for by
the predictor variables $\mathbf{X}$, and $\mathbf{Q}_X\mathbf{y}$, the portions of $\mathbf{y}$ that cannot be accounted for by $\mathbf{X}$.

Slight generalizations of the I-orthogonal projectors above lead to K-orthogonal pro-
jectors, which are useful in weighted least squares (LS) estimation in regression analysis:
Let $\mathbf{K}$ be an *nnd* matrix such that $\text{rank}(\mathbf{KX}) = \text{rank}(\mathbf{X})$. Then,

$$\mathbf{P}_{X/K} = \mathbf{X}(\mathbf{X}'\mathbf{KX})^-\mathbf{X}'\mathbf{K}, \tag{3}$$

and

$$\mathbf{Q}_{X/K} = \mathbf{I} - \mathbf{P}_{X/K} \tag{4}$$

are called K-orthogonal projectors onto $\text{Sp}(\mathbf{X}$ and $\text{Ker}(\mathbf{X}')$, respectively, with respect to
the metric matrix $\mathbf{K}$.

These projectors have properties similar to those of the I-orthogonal projectors:

$(\mathbf{KP}_{X/K})' = \mathbf{KP}_{X/K}$, $(\mathbf{KQ}_{X/K})' = \mathbf{KQ}_{X/K}$ (K-symmetric).
$\mathbf{P}^2_{X/K} = \mathbf{P}_{X/K}$, $\mathbf{Q}^2_{X/K} = \mathbf{Q}_{X/K}$ (idempotent).
$\mathbf{P}'_{X/K}\mathbf{KQ}_{X/K} = \mathbf{Q}'_{X/K}\mathbf{KP}_{X/K} = \mathbf{O}$ (K-orthogonal).

These projectors are useful in weighted LS (WLS) estimation in regression analysis. When
$\mathbf{K}$ is set to $\mathbf{K} = \mathbf{P}_Z$, the K-orthogonal projectors effect instrumental variable estimation.
See Yanai [33] for other types of projectors.

2 Constrained Principal Component Analysis (CPCA)

In as early as 1970, Professor Yanai (Yanai [32]) proposed so-called partial principal
component analysis (PPCA) to extract components unrelated to certain prescribed effects
such as differences in gender, age, levels of education, etc., which amounts to SVD of $\mathbf{Q}_G\mathbf{Y}$
(where $\mathbf{Y}$ is the matrix of criterion variables, and $\mathbf{G}$ the matrix of predictor variables whose
effects are to be eliminated). This process consists of two phases, decomposing $\mathbf{Y}$ int $\mathbf{P}_G\mathbf{Y}$
and $\mathbf{Q}_G\mathbf{Y}$, and applying SVD to the latter. While Yanai himself did not explicitly suggest

the SVD of $\mathbf{P}_G\mathbf{Y}$, it was known as redundancy analysis, a special case of reduced-rank regression. The two phases may be called External and Internal Analyses.

Similarly, CPCA consists of two major phases: External Analysis and Internal Analysis. External Analysis decomposes the main data matrix according to the external information about the row and columns of a data matrix, which amounts to projections. Internal Analysis further analyses the decomposed matrices into components, which is equivalent to SVD (singular value decomposition).

In CPCA, we consider not only the row-side constraints, $\mathbf{G}$, but also the column-side constraints $\mathbf{H}$, analogously to growth curve models (Potthoff and Roy [12]). This leads to a four-way decomposition of the main data matrix $\mathbf{Y}$ (Takane and Shibayama [20]):

$$\mathbf{Y} = \mathbf{P}_G\mathbf{Y}\mathbf{P}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{P}_H + \mathbf{P}_G\mathbf{Y}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{Y}\mathbf{Q}_H. \tag{5}$$

A similar decomposition is also possible with K-orthogonal projectors.

The decomposition above is a very basic one. When $\mathbf{G}$ and/or $\mathbf{H}$ consist of more than one set of variables, finer decompositions of $\mathbf{Y}$ are possible, corresponding to analogous decompositions of $\mathbf{P}_G$ and.or $\mathbf{P}_H$ (e.g., Takane [17]; Takane and Yanai [21]):

Let $\mathbf{G} = [\mathbf{M}, \mathbf{N}]$, for example. Then,

(1) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N \Leftrightarrow \mathbf{M}'\mathbf{N} = \mathbf{O}$. ($\mathbf{M}$ and $\mathbf{N}$ are mutually orthogonal.)

(2) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_N - \mathbf{P}_M\mathbf{P}_N \Leftrightarrow \mathbf{P}_M\mathbf{P}_N = \mathbf{P}_N\mathbf{P}_M$. ($\mathbf{M}$ and $\mathbf{N}$ are mutually orthogonal, except their common space, e.g., ANOVA w/o interactions). (3) $\mathbf{P}_G = \mathbf{P}_M + \mathbf{P}_{Q_M N} = \mathbf{P}_N + \mathbf{P}_{Q_N M}$. (Fit one first and the other to the residuals from the fist).

(4) $\mathbf{P}_G = \mathbf{P}_{M/Q_N} + \mathbf{P}_{N/Q_M} \Leftrightarrow \text{rank}(\mathbf{G}) = \text{rank}(\mathbf{M}) + \text{rank}(\mathbf{N})$. ($\mathbf{M}$ and $\mathbf{N}$ are disjoint. Fit both simultaneously).

(5) $\mathbf{P}_G = \mathbf{P}_{GA} + \mathbf{P}_{G(G'G)^- B} \Leftrightarrow \mathbf{A}'\mathbf{B} = \mathbf{O}$, and $\text{Sp}(\mathbf{A}) \oplus \text{Sp}(\mathbf{B}) = \text{Sp}(\mathbf{G}')$. (A matrix of regression coefficients $\mathbf{C}$ constrained by $\mathbf{C} = \mathbf{A}\mathbf{C}^*$ or by $\mathbf{B}'\mathbf{C} = \mathbf{O}$).

The first four decompositions above were noted in Rao and Yanai [13], while (5) is due to Yanai and Takane [34]. Analogous decompositions are possible for $\mathbf{P}_H$, $\mathbf{P}_{G/K}$, and $\mathbf{P}_{H/L}$.

In Internal Analysis, on the other hand, we apply PCA to terms obtained by the external analysis, e.g., $\mathbf{P}_G\mathbf{Y}\mathbf{P}_H$, which amounts to $\text{SVD}(\mathbf{P}_G\mathbf{Y}\mathbf{P}_H)$, whose computation time can be economized considerably by the following procedure:

A theorem on $\text{SVD}(\mathbf{P}_G\mathbf{Y}\mathbf{P}_H)$ (Takane and Hunter [18]): Let $\mathbf{F}_G$ and $\mathbf{F}_H$ be columnwise orthogonal matrices such that $\text{Sp}(\mathbf{G}) = \text{Sp}(\mathbf{F}_G)$ and $\text{Sp}(\mathbf{H}) = \text{Sp}(\mathbf{F}_H)$. Then, $\mathbf{P}_G\mathbf{Y}\mathbf{P}_H = \mathbf{F}_G\mathbf{F}_G'\mathbf{Y}\mathbf{F}_H\mathbf{F}_H'$. Let $\text{SVD}(\mathbf{F}_G'\mathbf{Y}\mathbf{F}_H)$ be denoted as $\mathbf{U}\mathbf{D}\mathbf{V}'$, and let $\text{SVD}(\mathbf{F}_G\mathbf{F}_G'\mathbf{Y}\mathbf{F}_H\mathbf{F}_H')$ be denoted as $\mathbf{U}^*\mathbf{D}^*\mathbf{V}^{*'}$. Then, $\mathbf{U}^* = \mathbf{F}_G\mathbf{U}$, $\mathbf{V}^* = \mathbf{F}_H\mathbf{V}$, and $\mathbf{D}^* = \mathbf{D}$.

3   Khatri's Lemma

Toward the end of 1980's, I was interested in the relationships among various methods of constrained correspondence analysis (CCA), a special case of CPCA. When I looked through the literature on CCA, I found that there were two ways of incorporating the constraints. Let $\mathbf{U}$ denote the row representation matrix. (For explanation, we consider only the row side constraints.) Two equivalent ways of constraining $\mathbf{U}$ are: (1) $\mathbf{U} = \mathbf{A}\mathbf{U}^*$ (e.g., ter Braak [28]), and (2) $\mathbf{B}'\mathbf{U} = \mathbf{O}$ (e.g., Böckenholt and Böckenholt [1]), where $\mathbf{A}$ and $\mathbf{B}$ are mutually orthogonal, and jointly span the entire row space of a contingency table. The relationship is rather trivial, i.e.,

$$\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}' = \mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^-\mathbf{B}' = \mathbf{Q}_B, \tag{6}$$

if the identity metric is used. I was not sure what would happen if non-identity metric $\mathbf{K}$ is used. Khatri's lemma states the exact relationship for this case (Takane, Yanai, and Mayekawa [26]):

Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p - r))$ be matrices such that $\text{rank}(\mathbf{A}) = r$, $\text{rank}(\mathbf{B}) = p - r$, and $\mathbf{A}'\mathbf{B} = \mathbf{O}$. Then (Khatri [9]),

$$\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-1}\mathbf{A}\mathbf{K} + \mathbf{K}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}', \tag{7}$$

where $\mathbf{K}$ is a symmetric pd (positive definite) matrix.

Several remarks are in order on Khtri's original lemma given above. Khatri's lemma may sometimes be expressed in an alternative form:

$$\mathbf{K} = \mathbf{K}\mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-1}\mathbf{A}\mathbf{K} + \mathbf{B}(\mathbf{B}'\mathbf{K}^{-1}\mathbf{B})^{-1}\mathbf{B}'. \tag{8}$$

Note also that $\mathbf{K}$ and $\mathbf{K}^{-1}$ are interchangeable. Khatri's lemma is useful for rewriting P-type projectors into Q-type projectors (LaMotte [11]; Shapiro [15]; Seber [14]; Takane and Zhou [24]; Verbyla [30]). Khatri's lemma has been generalized in various ways, e.g., let $\mathbf{K}$ be square, but not necessarily symmetric or nonsingular, but $\text{Sp}(\mathbf{B}) \subset \text{Sp}(\mathbf{K})$ and $\text{Sp}(\mathbf{B}) \subset \text{Sp}(\mathbf{K}')$. Then (Khatri [10]),

$$\mathbf{K} = \mathbf{K}\mathbf{A}(\mathbf{A}'\mathbf{K}\mathbf{A})^{-}\mathbf{A}'\mathbf{K} + \mathbf{B}(\mathbf{B}'\mathbf{K}^{-}\mathbf{B})^{-}\mathbf{B}'. \tag{9}$$

Professor Yanai (Yanai and Takane [34]) further extended Khatri's lemma as follows; Let $\mathbf{A}$ $(p \times r)$ and $\mathbf{B}$ $(p \times (p - r))$ be matrices such that $\text{rank}(\mathbf{A}) = r$ and $\text{rank}(\mathbf{B}) = p - r$, and let $\mathbf{M}$ and $\mathbf{N}$ be nnd matrices such that
  (i) $\mathbf{A}'\mathbf{M}\mathbf{N}\mathbf{B} = \mathbf{O}$,
  (ii) $\text{rank}(\mathbf{M}\mathbf{A}) = \text{rank}(\mathbf{A})$,
  (iii) $\text{rank}(\mathbf{N}\mathbf{B}) = \text{rank}(\mathbf{B})$.
Then,

$$\mathbf{I} = \mathbf{A}(\mathbf{A}'\mathbf{M}\mathbf{A})^{-}\mathbf{A}'\mathbf{M} + \mathbf{N}\mathbf{B}(\mathbf{B}'\mathbf{N}\mathbf{B})^{-}\mathbf{B}'. \tag{10}$$

This reduces to the original lemma when $\mathbf{M} = \mathbf{K}$ and $\mathbf{N} = \mathbf{K}^{-1}$. Takane [17] further extends it to a rectangular $\mathbf{K}$.

4   The Wedderburn-Guttman (WG) Theorem

The Wedderburn-Guttman (WG) theorem is stated as follows: Let $\mathbf{Y}$ $(n \times p)$ be of rank r, and let $\mathbf{A}$ $(n \times s)$ and $\mathbf{B}$ $(p \times s)$ be such that $\mathbf{A}'\mathbf{Y}\mathbf{B}$ is invertible. Then,

$$\text{rank}(\mathbf{Y}_1) = \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}) \tag{11}$$
$$= \text{rank}(\mathbf{Y}) - \text{rank}(\mathbf{A}'\mathbf{Y}\mathbf{B}) = r - s, \tag{12}$$

where

$$\mathbf{Y}_1 = \mathbf{Y} - \mathbf{Y}\mathbf{B}(\mathbf{A}'\mathbf{Y}\mathbf{B})^{-1}\mathbf{A}'\mathbf{Y}. \tag{13}$$

Wedderburn [31] first proved the theorem for $s = 1$. Guttman [5] extended it for $s > 1$. Guttman [6] further proved the reverse, i.e., $\mathbf{Y}_1$ must be of the above form to satisfy the rank condition stated above.

Guttman [5] used the matrix rank method for a proof of the above theorem. In this method, we apply a series of elementary block matrix operations to a matrix to derive a rank formula. We apply another series of elementary block matrix operations to the same matrix to derive another rank formula. Neither operations change the rank of the original

matrix, so the two must be equal. Guttman's proof is given in the appendix. Yongge Tian (many papers) derived many interesting rank formula based on this method. It is intriguing to find that Guttman [5] already used the method in 1944 (cf. Khatri [8]).

My initial interest in this theorem stemmed from Hubert's talk (Hubert, Meulman, and Heiser [7]) at the 1989 Meeting of the Psychometric Society at Illinois. This talk was to criticise the ignorance of numerical analysts (e.g., Chu, Funderlic, and Golub [2]) about Guttman's contributions (Guttman [5, 6]) in the WG theorem. When the talk was over, I asked a question: When $\mathbf{A}'\mathbf{YB}$ is not invertible, can we replace it by a generalized inverse? I had a feeling that it was possible, while Hubert said it was probably impossible. It has turned out that both of us are only half correct. The answer is yes, but it requires a condition. I initially thought this was purely a rank additivity (subtractivity) problem. That is, we are to prove that

$$
\begin{aligned}
\operatorname{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) \\
= \operatorname{rank}(\mathbf{Y}) - \operatorname{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}).
\end{aligned} \tag{14}
$$

This supposition also included that

$$
\operatorname{rank}(\mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \operatorname{rank}(\mathbf{A}'\mathbf{YB}) \tag{15}
$$

always holds. However, Tian and Styan [29] showed the following always holds:

$$
\operatorname{rank}(\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}) = \operatorname{rank}(\mathbf{Y}) - \operatorname{rank}(\mathbf{A}'\mathbf{YB}). \tag{16}
$$

This implies that (15) requires a condition, as does (14), and that the two conditions are equivalent.

The necessary and sufficient (ns) condition is stated as follows (Takane and Yanai [22]): Let $\mathbf{C} = \mathbf{B}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'$. Then, the ns condition for (14) and (15) to hold is:

$$
\mathbf{YCYCY} = \mathbf{YCY}. \tag{17}
$$

There are a number of equivalent conditions, e.g., $(\mathbf{YCYY}^{-})^2 = \mathbf{YCYY}^{-} \Leftrightarrow (\mathbf{Y}^{-}\mathbf{YCY})^2 = \mathbf{Y}^{-}\mathbf{YCY}$. There are also a number of interesting sufficient (but not necessary) conditions, e.g., $(\mathbf{YC})^2 = \mathbf{YC}$ or $(\mathbf{CY})^2 = \mathbf{CY}$, and $\mathbf{CYC} = \mathbf{C}$ (Cline, Funderlic, and Golub [3]; Galantai [4]). The latter is even stronger than the idempotency of $\mathbf{YC}$ or $\mathbf{CY}$.

The WG theorem states the rank condition for the residual matrix. However, from a data analytic viewpoint, the decomposition of the data matrix $\mathbf{Y}$ the theorem implies is even more interesting:

$$
\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + (\mathbf{Y} - \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y}). \tag{18}
$$

Takane and Hunter [19] developed a new family of CPCA almost exclusively based on this decomposition. The second term of the above decomposition involves a Q-type projector, but it can be replaced by a P-type projector as follows (Takane [17]): Let $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ be matrices such that
   (i) $\operatorname{Sp}(\tilde{\mathbf{A}}) \subset \operatorname{Sp}(\mathbf{Y})$,
   (ii) $\operatorname{Sp}(\tilde{\mathbf{B}}) \subset \operatorname{Sp}(\mathbf{Y}')$,
   (iii) $\operatorname{rank}(\mathbf{A}'\mathbf{YB}) + \operatorname{rank}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}}) = \operatorname{rank}(\mathbf{Y})$,
   (iv) $\mathbf{A}'\mathbf{YY}^{-}\tilde{\mathbf{A}} = \mathbf{A}'\tilde{\mathbf{A}} = \mathbf{O}$,
   (v) $\tilde{\mathbf{B}}'\mathbf{Y}^{-}\mathbf{YB} = \tilde{\mathbf{B}}'\mathbf{B} = \mathbf{O}$.
Then,

$$
\mathbf{Y} = \mathbf{YB}(\mathbf{A}'\mathbf{YB})^{-}\mathbf{A}'\mathbf{Y} + \tilde{\mathbf{A}}(\tilde{\mathbf{B}}'\mathbf{Y}^{-}\tilde{\mathbf{A}})^{-}\tilde{\mathbf{B}}'. \tag{19}
$$

5 Ridge Operators

In the mid 2000's, I was interested in extending the ridge-type of regularized least squares (RLS) estimation to various multivariate (MV) techniques. These extensions were rather straightforward, and I wrote most of the papers on them with my graduate students. I did not have to bother Professor Yanai. However, as I applied the RLS to so many MV procedures, I thought it would be important to write a paper on ridge operators, which was a common thread running through all of them (Takane [16, 23]).

The simplest form of ridge operators is defined as:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda \mathbf{P}_{X'})^- \mathbf{X}', \tag{20}$$

where $\mathbf{P}_{X'} = \mathbf{X}'(\mathbf{X}\mathbf{X}')^- \mathbf{X}$ is the orthogonal projector onto $\mathrm{Sp}(\mathbf{X}')$. ($\mathbf{P}_{X'} = \mathbf{I}$ if $\mathbf{X}$ is columnwise nonsingular.) This operator arises in the RLS estimation $\min_{\mathbf{c}} = \phi_\lambda(\mathbf{c})$ in regression analysis, where $\phi_\lambda(\mathbf{c}) = \mathrm{SS}(\mathbf{e}) + \lambda \mathrm{SS}(\mathbf{c})_{P_{X'}}$, and $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{c}$. (We assume, w/o loss of generality, that $\mathrm{Sp}(\mathbf{c}) \subset \mathrm{Sp}(\mathbf{X}')$.)

The $\mathbf{R}_X(\lambda)$ and $\mathbf{S}_X(\lambda)$ have properties similar to those of $\mathbf{P}_X$ and $\mathbf{Q}_X$, where $\mathbf{S}_X(\lambda) = \mathbf{I} - \mathbf{R}_X(\lambda)$. For example:

$\mathbf{R}_X(\lambda)$ and $\mathbf{S}_X(\lambda)$ are symmetric and invariant over the choice of a g-inverse of $(\mathbf{X}'\mathbf{X} + \lambda \mathbf{P}_{X'})$.

$\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda)\mathbf{R}_X(\lambda) = \mathbf{R}_X(\lambda)$ (i.e., $\mathbf{K}_X(\lambda) = \mathbf{R}_X(\lambda)^+$.).

$\mathbf{R}_X(\lambda) - \mathbf{R}_X(\lambda)^2 = \mathbf{R}_X(\lambda)\mathbf{S}_X(\lambda) = \mathbf{S}_X(\lambda)\mathbf{R}_X(\lambda) \geq \mathbf{O}$.

$\mathbf{R}_X(\lambda)\mathbf{K}_X(\lambda) = \mathbf{P}_X$, etc.

Similar decompositions of $\mathbf{R}_X(\lambda)$ to those of $\mathbf{P}_X$ are also possible.

The ridge operators defined above can be rewritten as follows using a ridge metric matrix defined below: Let

$$\mathbf{K}_X(\lambda) = \mathbf{P}_X + \lambda(\mathbf{X}\mathbf{X}')^+ \quad \text{(Ridge Metric Matrix)}. \tag{21}$$

Then, $\mathbf{R}_X(\lambda)$ can be rewritten as:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{K}_X(\lambda)\mathbf{X})^- \mathbf{X}'. \tag{22}$$

The simple ridge operators introduced above can be generalized into generalized ridge operators:

$$\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X} + \lambda \mathbf{L})^- \mathbf{X}'\mathbf{W}, \tag{23}$$

where $\mathbf{L}$ is an nnd matrix such that $\mathrm{Sp}(\mathbf{L}) \subset \mathrm{Sp}(\mathbf{X}')$, and $\mathbf{W}$ is an nnd matrix such that $\mathrm{rank}(\mathbf{W}\mathbf{X}) = \mathrm{rank}(\mathbf{X})$. As before, the generalized ridge operators can be rewritten as follows using a generalized ridge metric matrix defined below: Let

$$\mathbf{K}_X^{(W,L)}(\lambda) = \mathbf{P}_X + \lambda \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^- \mathbf{L}(\mathbf{X}'\mathbf{W}\mathbf{X})^- \mathbf{X}'\mathbf{W}. \tag{24}$$

Then,

$$\mathbf{R}_X^{(W,L)}(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{K}_X^{(W,L)}(\lambda)\mathbf{X})^- \mathbf{X}'\mathbf{W}. \tag{25}$$

6 Generalized Constrained Canonical Correlation Analysis

In the external analysis of CPCA, a data matrix is decomposed into several components by external information. I initially thought we could do the same in generalized constrained canonical correlation analysis (CANO). We decompose $\mathbf{X}$ and $\mathbf{Y}$ (the matrix of observations on the two sets of variables) separately into several orthogonal components,

and then choose one term from each decomposition, and apply CANO to the pair, which amounts to SVD of the product of the orthogonal projectors. It has turned out that this strategy will not do.

CANO analyzes total association between $\mathbf{X}$ and $\mathbf{Y}$, i.e., $\operatorname{tr}(\mathbf{P}_X \mathbf{P}_Y)$. However, $\mathbf{X} = \mathbf{M} + \mathbf{N}$, where $\mathbf{M'N} = \mathbf{O}$ does not guarantee $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$. This may be contrasted with a similar situation in which $\mathbf{X} = [\mathbf{M}, \mathbf{N}]$, where $\mathbf{M'N} = \mathbf{O}$, in which case we indeed have $\mathbf{P}_X = \mathbf{P}_M + \mathbf{P}_N$. This suggests that we need orthogonal decompositions of orthogonal projectors to derive additive decompositions of the total association.

Takane, Yanai, and Hwang [25] derived the following two orthogonal decompositions of $\mathbf{P}_{[X,G]}$ by combining two orthogonal decompositions ((3) and (5)) of the orthogonal projector given in the CPCA section:

(1) Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{W}$ be matrices such that $\operatorname{Sp}(\mathbf{A}) = \operatorname{Ker}(\mathbf{H'X'P}_G\mathbf{X})$, $\operatorname{Sp}(\mathbf{B}) = \operatorname{Ker}(\mathbf{H'X'Q}_G\mathbf{X})$, and $\operatorname{Sp}(\mathbf{W}) = \operatorname{Ker}(\mathbf{X'G})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_G XH} + \mathbf{P}_{P_G XA} + \mathbf{P}_{Q_G XH} + \mathbf{P}_{Q_G XB} + \mathbf{P}_{GW}. \tag{26}$$

(2) Let $\mathbf{K}$, $\mathbf{U}$, and $\mathbf{V}$ be matrices such that $\operatorname{Sp}(\mathbf{K}) = \operatorname{Ker}(\mathbf{H'X'X})$, $\operatorname{Sp}(\mathbf{U}) = \operatorname{Ker}(\mathbf{G'XH})$, and $\operatorname{Sp}(\mathbf{V}) = \operatorname{Ker}(\mathbf{G'XK})$. Then,

$$\mathbf{P}_{[X,G]} = \mathbf{P}_{P_{XH} G} + \mathbf{P}_{XHU} + \mathbf{P}_{P_{XK} G} + \mathbf{P}_{XKV} + \mathbf{P}_{Q_x G}. \tag{27}$$

We can derive similar decompositions of $\mathbf{P}_{[Y,G_Y]}$ (The subscript Y is put on G to indicate that this is a $\mathbf{G}$ for $\mathbf{Y}$.) We take one term each from a decomposition of $\mathbf{P}_{[X,G_X]}$ and that of $\mathbf{P}_{[Y,G_Y]}$, and apply SVD to the product of the two, e.g.,

$$\operatorname{SVD}(\mathbf{P}_{Q_{G_X} XH_X} \mathbf{P}_{Y H_Y U_Y}). \tag{28}$$

7   Causal Inference

Causal inference is one of the most important roles of statistics. This was the topic of our conversation when I met him last in the fall of 2013. When randomization is unavailable, there are a lot of pitfalls in establishing causal relationships based on correlational relationships alone. One crucial aspect of the problem is how to eliminate the effects of confounding variables.

The easiest way is to include the effects of the confounding variables in regression analysis along with the predictor variable of interest, although this is easier said than done. Identifying the set of confounding variables is not so easy, although here we assume that they are known. Let $\mathbf{y}$ denote the criterion variable, let $\mathbf{x}$: denote the predictor variable of interest, and let $\mathbf{U}$ denote the matrix of confounding variables. The suggested regression model can be written as:

$$\mathbf{y} = \mathbf{x}a_1 + \mathbf{U}\mathbf{c} + \mathbf{e}_1. \tag{29}$$

The ordinary least squares (OLS) estimate of $\mathbf{x}a_1$ is given by

$$\mathbf{x}\hat{a}_1 = \mathbf{P}_{x/Q_u}\mathbf{y} \tag{30}$$

.

Consider next the regression of $\mathbf{x}$ onto $\mathbf{U}$, i.e.,

$$\mathbf{x} = \mathbf{U}\mathbf{d} + \mathbf{e}_2. \tag{31}$$

The OLS estimate of $\mathbf{Ud}$ is given by

$$\mathbf{U}\hat{\mathbf{d}} = \mathbf{P}_U\mathbf{x}. \tag{32}$$

We call $\mathbf{P}_U\mathbf{x}$ linear propensity scores. Residuals from the above regression $\mathbf{Q}_U\mathbf{x}$ represent the portions of $\mathbf{x}$ left unaccounted for by $\mathbf{U}$.

We now consider using $\mathbf{P}_U\mathbf{x}$ instead of $\mathbf{U}$ in the first regression, i.e.,

$$\mathbf{y} = \mathbf{x}a_2 + \mathbf{P}_U\mathbf{x}b + \mathbf{e}_3. \tag{33}$$

The OLS estimate of $\mathbf{x}a_2$ is given by

$$\mathbf{x}\hat{a}_2 = \mathbf{P}_{x/Q_{P_U x}}\mathbf{y}, \tag{34}$$

where $\mathbf{Q}_{P_U x} = \mathbf{I} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U$.

Since

$$\mathbf{Q}_{P_U x}\mathbf{x} = \mathbf{x} - \mathbf{P}_U\mathbf{x}(\mathbf{x}'\mathbf{P}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{P}_U\mathbf{x} = \mathbf{Q}_U\mathbf{x}, \tag{35}$$

we obtain

$$\mathbf{P}_{x/Q_{P_U x}}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{36}$$

This means (30) and (34) are equivalent. This gives the rationale for replacing $\mathbf{U}$ by $\mathbf{P}_U\mathbf{x}$. The latter is more convenient because it is a single variable, and matching on a single variable is much easier than matching on multiple variables.

More recently, methods of of causal inference based on instrumental variables are getting popular. An instrumental variable $\mathbf{z}$ has the following properties:

(1) $\mathbf{z}'\mathbf{U} = \mathbf{0}$ ($\mathbf{z}$ and $\mathbf{U}$ are uncorrelated),
(2) $\mathbf{z}'\mathbf{x} \neq 0$ ($\mathbf{z}$ and $\mathbf{x}$ are correlated),
(3) $\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0$ (i.e., $\mathbf{z}$ has a predictive power on $\mathbf{Y}$ only through $\mathbf{x}$).

How is $\mathbf{z}$ related to $\mathbf{P}_U\mathbf{x}$ or $\mathbf{Q}_U\mathbf{x}$?

Assume $\mathbf{z} = c\mathbf{Q}_U\mathbf{x}$, where c is a normalization factor. This $\mathbf{z}$ satisfies (1) and (2) above. That it also satisfies (3) can be seen from:

$$(1/c)\mathbf{z}'\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_U\mathbf{Q}_{[U,x]}\mathbf{y} = \mathbf{x}'\mathbf{Q}_{[U,x]}\mathbf{y} = 0. \tag{37}$$

Consider the regression model:

$$\mathbf{y} = \mathbf{x}a_3 + \mathbf{e}_4. \tag{38}$$

The IV estimate of $\mathbf{x}a_3$ is given by

$$\mathbf{x}\hat{a}_3 = \mathbf{P}_{x/P_z}\mathbf{y} = \mathbf{P}_{x/Q_U}\mathbf{y}. \tag{39}$$

Since $\mathbf{P}_z = \mathbf{Q}_U\mathbf{x}(\mathbf{x}'\mathbf{Q}_U\mathbf{x})^{-1}\mathbf{x}'\mathbf{Q}_U$ and $\mathbf{x}'\mathbf{P}_z = \mathbf{x}'\mathbf{Q}_U$, this is identical to (30) and (34). This implies that the $\mathbf{z}$ defined above is an ideal IV.

8   Conclusions

This paper overviewed Professor Yanai's contributions to MV analysis. He adamantly emphasized linear algebraic aspects of MV analysis. His framework was grad, yet easy to understand. After almost half a century since I got to know him, I am still working within the framework of Professor Yanai.

9 Appendix: The Matrix Rank Method used by Guttman

The following is the proof of the original WG theorem by Guttman [5]. Let

$$\mathbf{C} = \left[\begin{array}{cc} \mathbf{I}_s & (\mathbf{A}'\mathbf{YB})^{-1}\mathbf{A}'\mathbf{Y} \\ \mathbf{YB} & \mathbf{Y} \end{array} \right], \quad \mathbf{E} = \left[\begin{array}{cc} \mathbf{I} & \mathbf{O} \\ -\mathbf{YB} & \mathbf{I} \end{array} \right],$$

$$\mathbf{F} = \left[\begin{array}{cc} \mathbf{I} & -(\mathbf{A}'\mathbf{YB})^{-1}\mathbf{A}'\mathbf{Y} \\ \mathbf{O} & \mathbf{I} \end{array} \right].$$

Then,

$$\mathbf{ECF} = \left[\begin{array}{cc} \mathbf{I}_s & \mathbf{O} \\ \mathbf{O} & \mathbf{Y}_1 \end{array} \right],$$

so that

$$\mathrm{rank}(\mathbf{C}) = s + \mathrm{rank}(\mathbf{Y}_1). \tag{40}$$

On the other hand, let

$$\mathbf{G} = \left[\begin{array}{cc} \mathbf{I} & -(\mathbf{A}'\mathbf{YB})^{-1} \\ \mathbf{O} & \mathbf{I} \end{array} \right], \quad \mathbf{H} = \left[\begin{array}{cc} \mathbf{I} & \mathbf{O} \\ -\mathbf{B} & \mathbf{I} \end{array} \right].$$

Then,

$$\mathbf{GCH} = \left[\begin{array}{cc} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{Y} \end{array} \right],$$

so that

$$\mathrm{rank}(\mathbf{C}) = \mathrm{rank}(\mathbf{Y}). \tag{41}$$

We obtain the WG theorem by combining (40) and (41).

Acknowledgement

References

[1] Böckenholt, U., and Böckenholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55, 633-639.

[2] Chu, M. T., Funderlic, R. E., and Golub, G. H. (1995). A rank-one reduction formula and its applications to matrix factorizations. *SIAM Review*, 37, 512-530.

[3] Cline, R. E., and Funderlic, R. E. (1979). The rank of a difference of matrices and associated generalized inverses. *Linear Algebra and its Applications*, 24, 185-215.

[4] Galantai, A. (2007). A note on generalized rank reduction. *Act Math Hung*, 116, 239-246.

[5] Guttman, L. (1944). General theory and methos for matric factoring. *Psychometrika*, 9, 1-16.

[6] Guttman, L. (1957). A necessary and sufficient formula for matric factoring. *Psychometrika*, 22, 79-81.

[7] Hubert, L., Meulman, J., and Heiser, W. J. (2000). Two purposes of matrix factorization: A historical appraisal. *SIAM Review*, 42, 68-82.

[8] Khatri, C. G. (1961). A simplified approach to the derivation of the theorems on the rank of a matrix. *Journal of the Maharaja Sayajirao University of Baroda*, 10, 1-5.

[9] Khatri, C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Ann I Stat Math*, 18, 75-86.

[10] Khatri, C. G. (1990). Some properties of BLUE in a linear model and canonical correlations associated with linear transformations. *J Multivariate Anal*, 34, 211-226.

[11] LaMotte, L. R. (2007). A direct derivation of the REML likelihood function. *Stat Pap*, 48, 321-327.

[12] Potthoff, R. F., and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.

[13] Rao, C. R., and Yanai, H. (1979). General definition and decomposition of projectors and some applications to statistical problems. *J Stat Plan Infer*, 3, 1-17.

[14] Seber, G. F. A. (1984). *Multivariate Observations*. New York: Wiley.

[15] Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *J Am Stat Assoc*, 81, 142-149.

[16] Takane, Y. (2008). More on regularization and (generalized) ridge operators. In K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (Eds.), *New Trends in Psychometrics*, (pp. 443-452). Tokyo: University Academic Press.

[17] Takane, Y. (2013). *Constrained Principal Component Analysis and Related Techniques*. Boca Raton, FL: Chapman and Hall/CRC Press.

[18] Takane, Y., and Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Appl Algebr Eng Comm*, 12, 391-419.

[19] Takane, Y., and Hunter, M. A. (2011). New family of constrained principal component analysis (CPCA). *Linear Algebra Appl*, 434, 2539-2555.

[20] Takane, Y., and Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psyhometrika*, 56, 97-120.

[21] Takane, Y., and Yanai, H. (1999). On oblique projectors. *Linear Algebra Appl*, 289, 297-310.

[22] Takane, Y., and Yanai, H. (2005). On the Wedderburn-Guttman theorem. *Linear Algebra Appl*, 410, 267-278.

[23] Takane, Y., and Yanai, H. (2008). On ridge operators. *Linear Algebra Appl*, 428, 1778-1790.

[24] Takane, Y., and Zhou, L. (2012). On two expressions of the MLE for a special case of the extended growth curve models. *Linear Algebra Appl*, 436, 2567-2577.

[25] Takane, Y., Yanai, H., and Hwang, H. (2006). An improved method for generalized constrained canonical correlation analysis. *Comp Stat Data An*, 50, 221-241.

[26] Takane, Y., Yanai, H., and Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, 56, 667-684.

[27] Takeuchi, K., Yanai, H., and Mukherjee, B. N. (1982). *The Foundation of Multivariate Analysis*. New Delhi: Wiley Eastern and New York: Halsted Press.

[28] ter Braak, C. J. F. (1986). Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67, 1167-1179.

[29] Tian, Y., and Styan, G. P. H. (2009). On some matrix equalities for generalized inverese with applications. *Linear Algebra Appl*, 430, 2716-2733.

[30] Verbyla, A. P. (1990). A conditional derivation of residual maximum likelihood. *Aust J Stat*, 32, 227-230.

[31] Wedderburn, J. H. M. (1934). *Lectures on Matrices*. Colloquium Publication, Vol. 17, Providence: American Mathematical Society.

[32] Yanai, H. (1970). Factor analysis with external criteria. *Jpn Psychol Res*, 12, 143-153.

[33] Yanai, H. (1990). Some generalized forms of least squares g-inverse, minimum norm g-inverse and Moore-Penrose inverse matrices. *Comp Stat Data An*, 10, 251-260.

[34] Yanai, H., and Takane, Y. (1992). Canonical correlation analysis with linear constraints. *Linear Algebra Appl*, 176, 75-82.

[35] Yanai, H., Takeuchi, K., Takane, Y. (2011). *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. New York: Springer.

Predictions under general linear random-effects model with coefficient restrictions and new observations

Yongge Tian*

CEMA, Central University of Finance and Economics, Beijing 100081, China

Abstract   Assume a general linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is given with coefficient vector $\boldsymbol{\beta}$ satisfying $\boldsymbol{\beta} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}$ and $\mathbf{B}\boldsymbol{\beta} = \mathbf{b}$, and new observations in the future follow the linear model $\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f$. This note aims at establishing the best linear unbiased predictor (BLUP) of the vector $\boldsymbol{\phi} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f$ of all unknown parameters in the two models from the original observation vector $\mathbf{y}$ under a most general assumption on the covariance matrix among the random vectors $\boldsymbol{\gamma}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}_f$. Through solving a constrained quadratic matrix-valued function optimization problem, we obtain a group of analytical formulas for calculating the BLUP of $\boldsymbol{\phi}$ and its covariance matrix operations. In particular, many special cases of the BLUP, as well as some fundamental decomposition equalities of the BLUP are established under various assumptions.

Keywords   Random-effects model; BLUP; covariance matrix; decomposition equality

1   Introduction

Consider a general linear random-effects model with parameter restrictions

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}, \quad \mathbf{B}\boldsymbol{\beta} = \mathbf{b}, \tag{1.1}$$

where in the first-stage model,

$\mathbf{y} \in \mathbb{R}^{n \times 1}$ is a vector of observable random variables,

$\mathbf{X} \in \mathbb{R}^{n \times p}$ is a known matrix of arbitrary rank,

$\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times 1}$ is a vector of unobservable random variables (measurement errors),
in the second-stage model,

$\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$ is a vector of unobservable random variables,

$\mathbf{A} \in \mathbb{R}^{p \times k}$ is known matrix of arbitrary rank,

$\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$ is a vector of fixed but unknown parameters (fixed effects),

$\boldsymbol{\gamma} \in \mathbb{R}^{p \times 1}$ is a vector of unobservable random variables (random effects),
and in the third matrix equation,

$\mathbf{B} \in \mathbb{R}^{p \times k}$ is known matrix of arbitrary rank,

$\mathbf{b} \in \mathbb{R}^{q \times 1}$ is a known vector with $\mathbf{b} \in \text{range}(\mathbf{BA})$.
Further, assume that new observations in the future follow the model

$$\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f, \tag{1.2}$$

---

*Email: yongge.tian@gmail.com

where

$\mathbf{y}_f \in \mathbb{R}^{n_f \times 1}$ is a vector of new observations (unknown),

$\mathbf{X}_f \in \mathbb{R}^{n_f \times p}$ is a known model matrix associated with the new observations,

$\boldsymbol{\varepsilon}_f \in \mathbb{R}^{n_f \times 1}$ is a vector of measurement errors associated with new observations,

$\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$ is the same as in (1.1).

In the investigation of restricted linear models, (1.1) is usually handled by transforming into an implicitly restricted model. The most popular transformation is combining the three parts in (1.1) as

$$\widehat{\mathbf{y}} = \widehat{\mathbf{X}}\boldsymbol{\beta} + \widehat{\boldsymbol{\varepsilon}} = \widehat{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} + \widehat{\mathbf{X}}\boldsymbol{\gamma} + \widehat{\boldsymbol{\varepsilon}}, \quad \widehat{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{b} \end{bmatrix}, \quad \widehat{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{B} \end{bmatrix}, \quad \widehat{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{0} \end{bmatrix}; \tag{1.3}$$

while combining (1.2) and (1.3) gives

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta} + \widetilde{\boldsymbol{\varepsilon}} = \widetilde{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} + \widetilde{\mathbf{X}}\boldsymbol{\gamma} + \widetilde{\boldsymbol{\varepsilon}}, \quad \widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{b} \\ \mathbf{y}_f \end{bmatrix}, \quad \widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{B} \\ \mathbf{X}_f \end{bmatrix}, \quad \widetilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{0} \\ \boldsymbol{\varepsilon}_f \end{bmatrix}. \tag{1.4}$$

In order to establish a unified theory on statistical inferences of (1.4), we assume that the expectation and covariance matrix of the combined random vector in (1.4) are given by

$$E\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_f \end{bmatrix} = \mathbf{0}, \quad Cov\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_f \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{bmatrix} := \boldsymbol{\Sigma}, \tag{1.5}$$

where we don't attach any further restrictions to the patterns of the submatrices $\boldsymbol{\Sigma}_{ij}$ in (1.5), although they are usually taken as certain prescribed forms for a specified random-effects model in statistical literatures. Correspondingly, the covariance matrix of the combined random vector $\widehat{\mathbf{y}}$ in (1.3) is given by

$$Cov(\widehat{\mathbf{y}}) = Cov(\widehat{\mathbf{X}}\boldsymbol{\gamma} + \widehat{\boldsymbol{\varepsilon}}) = \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}', \quad \mathbf{Z} = [\,\widehat{\mathbf{X}}, \widehat{\mathbf{I}}_n, \mathbf{0}\,], \quad \widehat{\mathbf{I}}_n = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix}. \tag{1.6}$$

This $Cov(\widehat{\mathbf{y}})$ is a known matrix under the assumptions in (1.1) and (1.2), and will occur in the statistical inference of (1.3) and (1.4).

Random-effects models are statistical models of parameters that vary at more than one level, which have different names in data analysis according to their origination, such as, multilevel models, hierarchical models, nested models, random parameter models, split-plot designs, etc. The first two equations in (1.1) is also called a nested linear model or two-stage hierarchical linear model in statistical literature, both of which are called the first-stage model and the second-stage model, respectively. Note from (1.4) and (1.5) that under the general assumptions in (1.1)–(1.5), $\widehat{\mathbf{y}}$ and $\mathbf{y}_f$ are correlated. Hence, it is possible to give predictions of $\mathbf{y}_f$, $\mathbf{X}_f\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_f$ in (1.2) from the original observation vector $\widehat{\mathbf{y}}$ in (1.3) under the assumptions in (1.1)–(1.5). In order to obtain some general results on BLUPs of unknown parameters under random-effects models, we construct a vector involving all unknown parameters in (1.4) as follows

$$\boldsymbol{\phi} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f, \tag{1.7}$$

where $\mathbf{F} \in \mathbb{R}^{s \times k}$, $\mathbf{G} \in \mathbb{R}^{s \times p}$, $\mathbf{H} \in \mathbb{R}^{s \times n}$, and $\mathbf{H}_f \in \mathbb{R}^{s \times n_f}$ are known matrices. In this case,

$$E(\boldsymbol{\phi}) = \mathbf{F}\boldsymbol{\alpha}, \quad Cov(\boldsymbol{\phi}) = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}', \quad Cov\{\widehat{\mathbf{y}}, \boldsymbol{\phi}\} = \mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}', \quad \mathbf{J} = [\,\mathbf{G}, \mathbf{H}, \mathbf{H}_f\,]. \tag{1.8}$$

Eq. (1.7) includes all vector operations in (1.1)–(1.4) as its special cases. For instance, if
$\mathbf{F} = \mathbf{T}\widetilde{\mathbf{X}}\mathbf{A}$, $\mathbf{G} = \mathbf{T}\widetilde{\mathbf{X}}$, $\mathbf{H} = \begin{bmatrix} \mathbf{T} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$, and $\mathbf{H}_f = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{T} \end{bmatrix}$, then (1.7) becomes $\boldsymbol{\phi} = \mathbf{T}\widetilde{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} +$
$\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\gamma} + \mathbf{T}\widetilde{\boldsymbol{\varepsilon}} = \mathbf{T}\widetilde{\mathbf{y}}$, which includes $\mathbf{y}$, $\mathbf{y}_f$, and $\widetilde{\mathbf{y}}$ as its special cases. Thus, statistical inference
of $\boldsymbol{\phi}$ is a comprehensive work, and will play prescriptive role for various special statistical
inference problems under (1.4) from both theoretical and applied points of view. Note
that there are 14 given matrices in (1.1)–(1.5) and (1.7). Hence, statistical inference of $\boldsymbol{\phi}$
is not easy task, and we will encounter many tedious matrix operations for the given 14
matrices, as demonstrated in Section 3 below.

The present author recently developed an algebraic method in [7] for deriving best
linear unbiased predictors (BLUPs) of general vectors of all unknown parameters in linear
random-effects models. Based this method, we are now able to give analytical formulas
for calculating the BLUP of $\boldsymbol{\phi}$ in (1.7) under the general assumptions as in (1.1)–(1.5),
and use the formulas to deal with various classic and new statistical inference problems on
random-effects models. As an ongoing approach on BLUPs under general linear random-
effects models, this paper aims at establishing a general theory on the BLUP of $\boldsymbol{\phi}$ in (1.7).
In particular, we shall derive

(I) analytical expression of the BLUP of $\boldsymbol{\phi}$;
(II) additive decomposition the BLUP of $\boldsymbol{\phi}$;
(III) various formulas related to the covariance matrices of the BLUP of $\boldsymbol{\phi}$.

2 Preliminaries

Before proceeding, we introduce the notation to the reader and explain its usage in
this paper. $\mathbb{R}^{m \times n}$ stands for the collection of all $m \times n$ real matrices. The symbols $\mathbf{A}'$,
$r(\mathbf{A})$ and $\mathscr{R}(\mathbf{A})$ stand for the transpose, the rank and the range (column space) of a
matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, respectively. $\mathbf{I}_m$ denotes the identity matrix of order m. The Moore–
Penrose inverse of $\mathbf{A}$, denoted by $\mathbf{A}^+$, is defined to be the unique solution $\mathbf{X}$ satisfying
the four matrix equations $\mathbf{AGA} = \mathbf{A}$, $\mathbf{GAG} = \mathbf{G}$, $(\mathbf{AG})' = \mathbf{AG}$, and $(\mathbf{GA})' = \mathbf{GA}$. $\mathbf{P_A}$,
$\mathbf{E_A}$, and $\mathbf{F_A}$ stand for the three orthogonal projectors (symmetric idempotent matrices)
$\mathbf{P_A} = \mathbf{AA}^+$, $\mathbf{E_A} = \mathbf{A}^{\perp} = \mathbf{I}_m - \mathbf{AA}^+$, and $\mathbf{F_A} = \mathbf{I}_n - \mathbf{A}^+\mathbf{A}$, where $\mathbf{E_A}$ and $\mathbf{F_A}$ satisfy
$\mathbf{E_A} = \mathbf{F_{A'}}$ and $\mathbf{F_A} = \mathbf{E_{A'}}$. Two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size are said to
satisfy the Löwner partial ordering $\mathbf{A} \succcurlyeq \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is nonnegative definite.

The following lemma is well known; see [4].

Lemma 1. *The linear matrix equation $\mathbf{AX} = \mathbf{B}$ is consistent if and only if $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A})$, or equivalently, $\mathbf{AA}^+\mathbf{B} = \mathbf{B}$. In this case, the general solution of the equation can be written in the following parametric form $\mathbf{X} = \mathbf{A}^+\mathbf{B} + (\mathbf{I} - \mathbf{A}^+\mathbf{A})\mathbf{U}$, where $\mathbf{U}$ is an arbitrary matrix.*

We also need the following formulas on ranks of matrices; see, e.g., [2, 6].

Lemma 2. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times k}$, and $\mathbf{C} \in \mathbb{R}^{l \times n}$. Then*

$$r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) + r(\mathbf{E_A}\mathbf{B}) = r(\mathbf{B}) + r(\mathbf{E_B}\mathbf{A}), \tag{2.1}$$

$$r\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) + r(\mathbf{CF_A}) = r(\mathbf{C}) + r(\mathbf{AF_C}), \tag{2.2}$$

$$r\begin{bmatrix} \mathbf{AA}' & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r[\mathbf{A}, \mathbf{B}] + r(\mathbf{B}). \tag{2.3}$$

If $\mathscr{R}(\mathbf{A}_1') \subseteq \mathscr{R}(\mathbf{B}_1')$, $\mathscr{R}(\mathbf{A}_2) \subseteq \mathscr{R}(\mathbf{B}_1)$, $\mathscr{R}(\mathbf{A}_2') \subseteq \mathscr{R}(\mathbf{B}_2')$ *and* $\mathscr{R}(\mathbf{A}_3) \subseteq \mathscr{R}(\mathbf{B}_2)$, *then*

$$r(\mathbf{A}_1\mathbf{B}_1^+\mathbf{A}_2) = r\begin{bmatrix} \mathbf{B}_1 & \mathbf{A}_2 \\ \mathbf{A}_1 & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1), \tag{2.4}$$

$$r(\mathbf{A}_1\mathbf{B}_1^+\mathbf{A}_2\mathbf{B}_2^+\mathbf{A}_3) = r\begin{bmatrix} \mathbf{0} & \mathbf{B}_2 & \mathbf{A}_3 \\ \mathbf{B}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1) - r(\mathbf{B}_2). \tag{2.5}$$

The following result on analytical solutions of a constrained matrix-valued function optimization problem was given in [7].

Lemma 3. *Let*

$$f(\mathbf{L}) = (\mathbf{LC} + \mathbf{D})\mathbf{M}(\mathbf{LC} + \mathbf{D})' \quad s.t. \quad \mathbf{LA} = \mathbf{B}, \tag{2.6}$$

where $\mathbf{A} \in \mathbb{R}^{p \times q}$, $\mathbf{B} \in \mathbb{R}^{n \times q}$, $\mathbf{C} \in \mathbb{R}^{p \times m}$, $\mathbf{D} \in \mathbb{R}^{n \times m}$ *and* $\mathbf{N} = \mathbf{N}' \in \mathbb{R}^{n \times n}$ *are given,* $\mathbf{M} \in \mathbb{R}^{m \times m}$ *is nnd, and the matrix equation* $\mathbf{LA} = \mathbf{B}$ *is consistent. Then there always exists a solution* $\mathbf{L}_0$ *of* $\mathbf{L}_0\mathbf{A} = \mathbf{B}$ *such that*

$$f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0) \tag{2.7}$$

holds for all solutions of $\mathbf{LA} = \mathbf{B}$*. In this case, the matrix* $\mathbf{L}_0$ *satisfying* (2.7) *is determined by the following consistent matrix equation*

$$\mathbf{L}_0[\,\mathbf{A},\, \mathbf{CMC}'\mathbf{A}^\perp\,] = [\,\mathbf{B},\, -\mathbf{DMC}'\mathbf{A}^\perp\,]. \tag{2.8}$$

In this case, the general expression of $\mathbf{L}_0$ *and the corresponding* $f(\mathbf{L}_0)$ *are given by*

$$\mathbf{L}_0 = \operatorname*{argmin}_{\mathbf{LA}=\mathbf{B}} f(\mathbf{L}) = [\,\mathbf{B},\, -\mathbf{DMC}'\mathbf{A}^\perp\,][\,\mathbf{A},\, \mathbf{CMC}'\mathbf{A}^\perp\,]^+ + \mathbf{V}[\,\mathbf{A},\, \mathbf{CMC}'\,]^\perp, \tag{2.9}$$

$$f(\mathbf{L}_0) = \min_{\mathbf{LA}=\mathbf{B}} f(\mathbf{L}) = \mathbf{KMK}' - \mathbf{KMC}'(\mathbf{A}^\perp\mathbf{CMC}'\mathbf{A}^\perp)^+\mathbf{CMK}', \tag{2.10}$$

$$f(\mathbf{L}) - f(\mathbf{L}_0) = (\mathbf{LCMC}'\mathbf{A}^\perp + \mathbf{DMC}'\mathbf{A}^\perp)(\mathbf{A}^\perp\mathbf{CMC}'\mathbf{A}^\perp)^+(\mathbf{LCMC}'\mathbf{A}^\perp + \mathbf{DMC}'\mathbf{A}^\perp)', \tag{2.11}$$

where $\mathbf{K} = \mathbf{BA}^+\mathbf{C} + \mathbf{D}$*, and* $\mathbf{V} \in \mathbb{R}^{n \times p}$ *is arbitrary.*

3 Main results

In order to establish the theory of the BLUP of (1.7) under (1.3), we need the following concepts. A linear statistic $\mathbf{L}\widehat{\mathbf{y}}$ under (1.3), where $\mathbf{L} \in \mathbb{R}^{s \times (n+q)}$, is said to have the same expectation with $\boldsymbol{\phi}$ in (1.7) if and only if $E(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}) = \mathbf{0}$ holds. In this case, if there exists an $\mathbf{L}_0$ such that

$$E(\mathbf{L}_0\widehat{\mathbf{y}} - \boldsymbol{\phi}) = \mathbf{0} \quad \text{and} \quad Cov(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}) \succcurlyeq Cov(\mathbf{L}_0\widehat{\mathbf{y}} - \boldsymbol{\phi}) \text{ s.t. } E(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}) = \mathbf{0} \tag{3.1}$$

hold, then the linear statistic $\mathbf{L}_0\widehat{\mathbf{y}}$ is defined to be the BLUP of $\boldsymbol{\phi}$ in (1.7), and is denoted by $\mathbf{L}_0\widehat{\mathbf{y}} = \mathrm{BLUP}(\boldsymbol{\phi})$.

Lemma 4. *The vector* $\boldsymbol{\phi}$ *in* (1.7) *is predictable by* $\widehat{\mathbf{y}}$ *in* (1.3) *if and only if*

$$\mathscr{R}[(\widehat{\mathbf{X}}\mathbf{A})'] \supseteq \mathscr{R}(\mathbf{F}'). \tag{3.2}$$

Proof. It is obvious that $E(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}) = \mathbf{0} \Leftrightarrow \mathbf{L}\widehat{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} - \mathbf{F}\boldsymbol{\alpha} = \mathbf{0}$ for all $\boldsymbol{\alpha} \Leftrightarrow \mathbf{L}\widehat{\mathbf{X}}\mathbf{A} = \mathbf{F}$. From Lemma 1, the matrix equation is consistent if and only if (3.2) holds. $\qquad \square$

Theorem 5. *Assume that $\boldsymbol{\phi}$ in (1.7) is predictable by $\widehat{\mathbf{y}}$ in (1.3), namely, (3.2) holds, and let $\widehat{\mathbf{I}}_n$, $\mathbf{Z}$, and $\mathbf{J}$ be as given in (1.6) and (1.8). Then*

$$E\left(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}\right) = \mathbf{0} \ and \ Cov\left(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}\right) = \min \Leftrightarrow \mathbf{L}[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = [\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,].$$
$$(3.3)$$

The matrix equation in (3.3) is consistent as well under (3.2). In this case, the general solution of $\mathbf{L}$ and $\mathrm{BLUP}(\boldsymbol{\phi})$ can be written as

$$\mathrm{BLUP}(\boldsymbol{\phi}) = \mathbf{L}\widehat{\mathbf{y}} = \left([\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+} + \mathbf{U}[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{\perp}\right)\widehat{\mathbf{y}},$$
$$(3.4)$$

where $\mathbf{U} \in \mathbb{R}^{s \times (n+q)}$ is arbitrary. Further, the following results hold.

(a) $r[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = r[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\,]$, $\mathscr{R}[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = \mathscr{R}[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\,]$, *and* $\mathscr{R}(\widehat{\mathbf{X}}\mathbf{A}) \cap \mathscr{R}[\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}] = \{\mathbf{0}\}$.

(b) $\mathbf{L}$ *is unique if and only if* $r[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\,] = n$.

(c) $\mathrm{BLUP}(\boldsymbol{\phi})$ *is unique with probability 1 if and only if* $\widehat{\mathbf{y}} \in \mathscr{R}[\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\,]$, *i.e., (1.3) is consistent.*

(d) *The covariance matrix of $\mathrm{BLUP}(\boldsymbol{\phi})$ is*

$$Cov[\mathrm{BLUP}(\boldsymbol{\phi})] = [\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'$$
$$\times([\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+})', \qquad (3.5)$$

the covariance matrix between $\mathrm{BLUP}(\boldsymbol{\phi})$ and $\boldsymbol{\phi}$ is

$$Cov\{\mathrm{BLUP}(\boldsymbol{\phi}), \ \boldsymbol{\phi}\} = [\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}', \qquad (3.6)$$

the difference of the covariance matrices of $\boldsymbol{\phi}$ and $\mathrm{BLUP}(\boldsymbol{\phi})$ is

$$Cov(\boldsymbol{\phi}) - Cov[\mathrm{BLUP}(\boldsymbol{\phi})] = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}'$$
$$-[\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'([\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+})',$$
$$(3.7)$$

the covariance matrix of the difference of $\boldsymbol{\phi}$ and $\mathrm{BLUP}(\boldsymbol{\phi})$ is

$$Cov[\boldsymbol{\phi} - \mathrm{BLUP}(\boldsymbol{\phi})] = \left([\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z} - \mathbf{J}\right)\boldsymbol{\Sigma}$$
$$\times \left([\,\mathbf{F}, \ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A}, \ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z} - \mathbf{J}\right)'. \quad (3.8)$$

Proof. Note that

$$\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi} = \mathbf{L}\widehat{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} + \mathbf{L}\widehat{\mathbf{X}}\boldsymbol{\gamma} + \mathbf{L}\widehat{\boldsymbol{\varepsilon}} - \mathbf{F}\boldsymbol{\alpha} - \mathbf{G}\boldsymbol{\gamma} - \mathbf{H}\boldsymbol{\varepsilon} - \mathbf{H}_f\boldsymbol{\varepsilon}_f$$
$$= (\mathbf{L}\widehat{\mathbf{X}}\mathbf{A} - \mathbf{F})\boldsymbol{\alpha} + (\mathbf{L}\widehat{\mathbf{X}} - \mathbf{G})\boldsymbol{\gamma} + \left(\mathbf{L}\widehat{\mathbf{I}}_n - \mathbf{H}\right)\boldsymbol{\varepsilon} - \mathbf{H}_f\boldsymbol{\varepsilon}_f.$$

Hence,

$$Cov\left(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}\right) = Cov\left((\mathbf{L}\widehat{\mathbf{X}} - \mathbf{G})\boldsymbol{\gamma} + \left(\mathbf{L}\widehat{\mathbf{I}}_n - \mathbf{H}\right)\boldsymbol{\varepsilon} - \mathbf{H}_f\boldsymbol{\varepsilon}_f\right)$$
$$= [\mathbf{L}\widehat{\mathbf{X}} - \mathbf{G}, \ \mathbf{L}\widehat{\mathbf{I}}_n - \mathbf{H}, \ -\mathbf{H}_f\,]\boldsymbol{\Sigma}[\mathbf{L}\widehat{\mathbf{X}} - \mathbf{G}, \ \mathbf{L}\widehat{\mathbf{I}}_n - \mathbf{H}, \ -\mathbf{H}_f\,]'$$
$$= (\mathbf{L}[\,\widehat{\mathbf{X}}, \ \widehat{\mathbf{I}}_n, \ \mathbf{0}\,] - [\,\mathbf{G}, \ \mathbf{H}, \ \mathbf{H}_f\,])\boldsymbol{\Sigma}(\mathbf{L}[\,\widehat{\mathbf{X}}, \ \widehat{\mathbf{I}}_n, \ \mathbf{0}\,] - [\,\mathbf{G}, \ \mathbf{H}, \ \mathbf{H}_f\,])'$$
$$= (\mathbf{L}\mathbf{Z} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L}\mathbf{Z} - \mathbf{J})' := f(\mathbf{L}). \qquad (3.9)$$

In this case, we see from Lemma 1 that the first part of (3.3) is equivalent to finding a solution $\mathbf{L}_0$ of the consistent matrix equation $\mathbf{L}_0\widehat{\mathbf{X}}\mathbf{A} = \mathbf{F}$ such that

$$f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0) \quad \text{s.t.} \quad \mathbf{L}\widehat{\mathbf{X}}\mathbf{A} = \mathbf{F} \tag{3.10}$$

holds in the Löwner partial ordering. Further from Lemma 3, there always exists s solution $\mathbf{L}_0$ of $\mathbf{L}_0\widehat{\mathbf{X}}\mathbf{A} = \mathbf{F}$ such that (3.10) holds, and the $\mathbf{L}_0$ is determined by the matrix equation $\mathbf{L}_0[\,\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = [\,\mathbf{F},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]$, establishing the matrix equation in (3.3). Solving the equation by Lemma 1 gives the $\mathbf{L}$ in (3.4). Also from (2.10),

$$f(\mathbf{L}_0) = Cov(\mathbf{L}_0\widehat{\mathbf{y}} - \boldsymbol{\phi}) = Cov(\boldsymbol{\phi} - \mathbf{L}_0\widehat{\mathbf{y}}) = (\mathbf{L}_0\mathbf{Z} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L}_0\mathbf{Z} - \mathbf{J})',$$

establishing (3.8). Result (a) is well known. Result (b) follows directly from (3.4). Taking covariance matrix of (3.4) yields (3.5). From (1.7) and (3.4),

$$\begin{aligned} Cov\{\mathrm{BLUP}(\boldsymbol{\phi}),\ \boldsymbol{\phi}\} = Cov\{\mathbf{L}\widehat{\mathbf{y}},\ \boldsymbol{\phi}\} &= \mathbf{L}Cov\{\widehat{\mathbf{y}},\ \boldsymbol{\phi}\} \\ &= [\,\mathbf{F},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,][\,\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}', \end{aligned}$$

establishing (3.6). Eq. (3.7) follows from (1.8) and (3.5). $\qquad\square$

Because the BLUP of $\boldsymbol{\phi}$ in (1.7) and the covariance matrix operations in Theorem 5 are derived directly from definition and analytical operations of the 14 given matrices, a solid mathematical foundation is really established for the statistical inference of BLUPs under random-effects models with with coefficient restrictions and new observations. In the light of the basic formulas in Theorem 5, we are now able to derive many new and valuable consequences on mathematical and properties of BLUPs under various conditions.

Corollary 6. *The following results hold.*

(a) *If $\boldsymbol{\phi}$ in (1.7) is predictable by $\widehat{\mathbf{y}}$ in (1.3), then $\mathbf{T}\boldsymbol{\phi}$ is predictable by $\widehat{\mathbf{y}}$ in (1.3) as well for any matrix $\mathbf{T} \in \mathbb{R}^{t \times s}$, and $\mathrm{BLUP}(\mathbf{T}\boldsymbol{\phi}) = \mathbf{T}\mathrm{BLUP}(\boldsymbol{\phi})$ holds.*

(b) *If $\boldsymbol{\phi}$ in (1.7) is predictable by $\widehat{\mathbf{y}}$ in (1.3), then $\mathbf{F}\boldsymbol{\alpha}$ is estimable by $\widehat{\mathbf{y}}$ in (1.3) as well, and the BLUP of $\boldsymbol{\phi}$ can be decomposed as the sum*

$$\mathrm{BLUP}(\boldsymbol{\phi}) = \mathrm{BLUE}(\mathbf{F}\boldsymbol{\alpha}) + \mathrm{BLUP}(\mathbf{G}\boldsymbol{\gamma}) + \mathrm{BLUP}(\mathbf{H}\boldsymbol{\varepsilon}) + \mathrm{BLUP}(\mathbf{H}_f\boldsymbol{\varepsilon}_f), \tag{3.11}$$

and the following formulas for covariance matrices hold

$$Cov\{\mathrm{BLUE}(\mathbf{F}\boldsymbol{\alpha}),\ \mathrm{BLUP}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f)\} = \mathbf{0}, \tag{3.12}$$

$$Cov[\mathrm{BLUP}(\boldsymbol{\phi})] = Cov[\mathrm{BLUE}(\mathbf{F}\boldsymbol{\alpha})] + Cov[\mathrm{BLUP}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f)]. \tag{3.13}$$

(c) *If $\boldsymbol{\alpha}$ in (1.7) is estimable under (1.3), then the $\boldsymbol{\phi}$ in (1.7) is predictable by $\widehat{\mathbf{y}}$ in (1.3). In this case,*

$$\mathrm{BLUP}\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon}_f \end{bmatrix} = \begin{bmatrix} \mathrm{BLUE}(\boldsymbol{\alpha}) \\ \mathrm{BLUP}(\boldsymbol{\gamma}) \\ \mathrm{BLUP}(\boldsymbol{\varepsilon}) \\ \mathrm{BLUP}(\boldsymbol{\varepsilon}_f) \end{bmatrix}, \tag{3.14}$$

$$\mathrm{BLUP}(\boldsymbol{\phi}) = \mathbf{F}\mathrm{BLUE}(\boldsymbol{\alpha}) + \mathbf{G}\mathrm{BLUP}(\boldsymbol{\gamma}) + \mathbf{H}\mathrm{BLUP}(\boldsymbol{\varepsilon}) + \mathbf{H}_f\mathrm{BLUP}(\boldsymbol{\varepsilon}_f). \tag{3.15}$$

Proof. The estimability of $\mathbf{T}\boldsymbol{\phi}$ follows from $\mathscr{R}[(\widehat{\mathbf{X}}\mathbf{A})'] \supseteq \mathscr{R}(\mathbf{F}') \supseteq \mathscr{R}(\mathbf{F}'\mathbf{T}')$. Also from (3.4),

$$\text{BLUP}(\mathbf{T}\boldsymbol{\phi}) = \left([\mathbf{TF},\ \mathbf{TJ}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+} + \mathbf{U}[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{\perp}\right)\widehat{\mathbf{y}}$$

$$= \mathbf{T}\left([\mathbf{F},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+} + \mathbf{U}_1[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{\perp}\right)\widehat{\mathbf{y}}$$

$$= \mathbf{T}\text{BLUP}(\boldsymbol{\phi}),$$

where $\mathbf{U} = \mathbf{T}\mathbf{U}_1$, establishing the equality in (a).

Note that $[\mathbf{F},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}$ in (3.4) can be decomposed as

$$[\mathbf{F},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}$$

$$= [\mathbf{F},\ \mathbf{0}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+} + [\mathbf{0},\ [\mathbf{G},\ \mathbf{0},\ \mathbf{0}]\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}$$

$$+ [\mathbf{0},\ [\mathbf{0},\ \mathbf{H},\ \mathbf{0}]\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}$$

$$+ [\mathbf{0},\ [\mathbf{0},\ \mathbf{0},\ \mathbf{H}_f]\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}$$

$$= \text{BLUE}(\mathbf{F}\boldsymbol{\alpha}) + \text{BLUP}(\mathbf{G}\boldsymbol{\gamma}) + \text{BLUP}(\mathbf{H}\boldsymbol{\varepsilon}) + \text{BLUP}(\mathbf{H}_f\boldsymbol{\varepsilon}_f),$$

establishing (3.11). We also obtain from (3.4) the covariance matrix between $\text{BLUE}(\mathbf{F}\boldsymbol{\alpha})$ and $\text{BLUP}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f)$ as follows

$$Cov\{\text{BLUE}(\mathbf{F}\boldsymbol{\alpha}),\ \text{BLUP}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f)\}$$

$$= [\mathbf{F},\ \mathbf{0}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'([\mathbf{0},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+})'. \quad (3.16)$$

Applying (2.5) to (3.16) and simplifying, we obtain

$$r(Cov\{\text{BLUE}(\mathbf{F}\boldsymbol{\alpha}),\ \text{BLUP}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon} + \mathbf{H}_f\boldsymbol{\varepsilon}_f)\})$$

$$= r\left([\mathbf{F},\ \mathbf{0}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'([\mathbf{0},\ \mathbf{J}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}][\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]^{+})'\right)$$

$$= r\begin{bmatrix} \mathbf{0} & \begin{bmatrix} (\widehat{\mathbf{X}}\mathbf{A})' \\ (\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}' \end{bmatrix} & \begin{bmatrix} \mathbf{0} \\ (\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}' \end{bmatrix} \\ [\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}] & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}' & \mathbf{0} \\ [\mathbf{F},\ \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$-2r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}]$$

$$= r\begin{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \end{bmatrix} & \begin{bmatrix} (\widehat{\mathbf{X}}\mathbf{A})' \\ \mathbf{0} \end{bmatrix} & \begin{bmatrix} \mathbf{0} \\ (\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}' \end{bmatrix} \\ [\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{0}] & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}' & \mathbf{0} \\ [\mathbf{F},\ \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$-2r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}']$$

$$= r\begin{bmatrix} \mathbf{0} & (\widehat{\mathbf{X}}\mathbf{A})' \\ \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}' \\ \mathbf{F} & \mathbf{0} \end{bmatrix} + r[(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp},\ (\widehat{\mathbf{X}}\mathbf{A})^{\perp}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}'] - 2r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}']$$

$$= r\begin{bmatrix} \widehat{\mathbf{X}}\mathbf{A} \\ \mathbf{F} \end{bmatrix} + r\begin{bmatrix} (\widehat{\mathbf{X}}\mathbf{A})' \\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}' \end{bmatrix} + r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}']$$

$$-r(\widehat{\mathbf{X}}\mathbf{A}) - 2r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'] \quad \text{(by (2.3))}$$

$$= r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}',\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}'] - r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}']$$

$$= r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'] - r[\widehat{\mathbf{X}}\mathbf{A},\ \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'] = 0,$$

which implies that $Cov\{\,\mathrm{BLUE}(\mathbf{F}\boldsymbol{\alpha}),\,\mathrm{BLUP}(\mathbf{G}\boldsymbol{\gamma}+\mathbf{H}\boldsymbol{\varepsilon}+\mathbf{H}_f\boldsymbol{\varepsilon}_f)\,\}$ is a zero matrix, establishing (3.12). Eq. (3.13) follow from (3.11) and (3.12). Eqs. (3.14) and (3.15) follow from (a) and (3.11). □

Corollary 7. *Let* $\boldsymbol{\Sigma}$ *and* $\mathbf{Z}$ *be as given in* (1.5) *and* (1.6), *and assume that*

$$\boldsymbol{\phi}_1 = \mathbf{F}_1\boldsymbol{\alpha} + \mathbf{G}_1\boldsymbol{\gamma} + \mathbf{H}_1\boldsymbol{\varepsilon} + \mathbf{H}_{f1}\boldsymbol{\varepsilon}_f, \quad \boldsymbol{\phi}_2 = \mathbf{F}_2\boldsymbol{\alpha} + \mathbf{G}_2\boldsymbol{\gamma} + \mathbf{H}_2\boldsymbol{\varepsilon} + \mathbf{H}_{f2}\boldsymbol{\varepsilon}_f$$

are predictable under (1.4), *where* $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^{s\times k}$, $\mathbf{G}_1, \mathbf{G}_2 \in \mathbb{R}^{s\times p}$, $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{R}^{s\times n}$, *and* $\mathbf{H}_{f1}, \mathbf{H}_{f2} \in \mathbb{R}^{s\times n_f}$ *are known matrices, and denote* $\mathbf{J}_i = [\,\mathbf{G}_i, \mathbf{H}_i, \mathbf{H}_{f_i}\,]$, $i = 1, 2$. *Then, the following results hold.*

(a) *The sum* $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2$ *is predictable under* (1.4), *and its BLUP satisfies*

$$\mathrm{BLUP}(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2) = \mathrm{BLUP}(\boldsymbol{\phi}_1) + \mathrm{BLUP}(\boldsymbol{\phi}_2). \tag{3.17}$$

(b) $\mathrm{BLUP}(\boldsymbol{\phi}_1) = \mathrm{BLUP}(\boldsymbol{\phi}_2) \Leftrightarrow \mathbf{F}_1 = \mathbf{F}_2$ *and* $\mathscr{R}(\mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}_1' - \mathbf{Z}\boldsymbol{\Sigma}\mathbf{J}_2') \subseteq \mathscr{R}(\widehat{\mathbf{X}}\mathbf{A})$.

Proof. Eq. (3.17) follows from Corollary 6(a) and (3.11). From Theorem 5, the two equations for the coefficient matrices of $\mathrm{BLUP}(\boldsymbol{\phi}_1) = \mathbf{L}_1\widehat{\mathbf{y}}$ and $\mathrm{BLUP}(\boldsymbol{\phi}_2) = \mathbf{L}_2\widehat{\mathbf{y}}$ are given by

$$\mathbf{L}_1[\,\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = [\,\mathbf{F}_1, \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,], \ \ \mathbf{L}_2[\,\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,] = [\,\mathbf{F}_2, \mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}\,].$$

The pair of matrix equations have a common solution iff

$$r\begin{bmatrix} \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} & \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \\ \mathbf{F}_1 & \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} & \mathbf{F}_2 & \mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \end{bmatrix}$$
$$= r[\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}, \widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}], \tag{3.18}$$

where

$$r\begin{bmatrix} \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} & \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \\ \mathbf{F}_1 & \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} & \mathbf{F}_2 & \mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \end{bmatrix}$$
$$= r\begin{bmatrix} \widehat{\mathbf{X}}\mathbf{A} & \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_2 - \mathbf{F}_1 & (\mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}' - \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}')(\widehat{\mathbf{X}}\mathbf{A})^{\perp} \end{bmatrix}$$
$$= r[\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}] + r[\mathbf{F}_2 - \mathbf{F}_1, (\mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}' - \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}')(\widehat{\mathbf{X}}\mathbf{A})^{\perp}],$$
$$r[\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}, \widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}] = r[\widehat{\mathbf{X}}\mathbf{A}, \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'(\widehat{\mathbf{X}}\mathbf{A})^{\perp}].$$

Hence (3.18) is equivalent to $[\mathbf{F}_2 - \mathbf{F}_1, (\mathbf{J}_2\boldsymbol{\Sigma}\mathbf{Z}' - \mathbf{J}_1\boldsymbol{\Sigma}\mathbf{Z}')(\widehat{\mathbf{X}}\mathbf{A})^{\perp}] = \mathbf{0}$, which is further equivalent to (b). □

Finally, we give some decomposition equalities of the BLUP of $\widetilde{\mathbf{y}}$ in (1.4).

Corollary 8. *The vector* $\widetilde{\mathbf{y}}$ *in* (1.4) *is predictable by* $\widehat{\mathbf{y}}$ *in* (1.3) *if and only if* $\mathscr{R}[(\widehat{\mathbf{X}}\mathbf{A})'] \supseteq \mathscr{R}[(\mathbf{X}_f\mathbf{A})']$. *In this case,* $\mathbf{B}\boldsymbol{\beta}$ *is predictable by* $\widehat{\mathbf{y}}$ *in* (1.3), *and the following decomposition equalities hold*

$$\mathrm{BLUP}(\widetilde{\mathbf{y}}) = \begin{bmatrix} \mathbf{y} \\ \mathbf{b} \\ \mathrm{BLUP}(\mathbf{y}_f) \end{bmatrix} = \begin{bmatrix} \mathrm{BLUP}(\mathbf{X}\boldsymbol{\beta}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}) \\ \mathrm{BLUP}(\mathbf{B}\boldsymbol{\beta}) \\ \mathrm{BLUP}(\mathbf{X}_f\boldsymbol{\beta}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_f) \end{bmatrix}$$
$$= \begin{bmatrix} \mathrm{BLUE}(\mathbf{X}\mathbf{A}\boldsymbol{\alpha}) + \mathrm{BLUP}(\mathbf{X}\boldsymbol{\gamma}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}) \\ \mathrm{BLUE}(\mathbf{B}\mathbf{A}\boldsymbol{\alpha}) + \mathrm{BLUP}(\mathbf{B}\boldsymbol{\gamma}) \\ \mathrm{BLUE}(\mathbf{X}_f\mathbf{A}\boldsymbol{\alpha}) + \mathrm{BLUP}(\mathbf{X}_f\boldsymbol{\gamma}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_f) \end{bmatrix}. \tag{3.19}$$

All these decomposition equalities can also be regarded as built-in restrictions to BLUPs under random-effects models, which demonstrate some elegant properties of BLUP-s. Some previous discussions on built-in restrictions to BLUPs can be found, e.g., in [1, 3, 5]. The decomposition equalities in (3.19) are so fundamental that they will play control role in statistical inferences of general linear regression models. It is expected that decomposition equalities like (3.19) can always be established for observed response random vectors with respect to BLUEs/BLUPs under various types of linear regression model, and many new conclusions on statistical inferences of regression models can be derived under the control of these decomposition equalities.

References

[1] Brumback, B.A. (2010). On the built-in restrictions in linear mixed models, with Application to smoothing spline analysis of variance. *Commun. Statist. Theor. Meth.*, 39, 579–591.

[2] Marsaglia, G., Styan, G.P.H. (1974). Equalities and inequalities for ranks of matrices. *Linear Multilinear Algebra*, 2, 269–292.

[3] McLean, R.A., Sanders, W.L., Stroup, W.W. (1991). A unified approach to mixed linear models. *Amer. Statist.*, 45, 54–64.

[4] Penrose, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Phil. Soc.*, 51, 406–413.

[5] Searle, S.R. (1997). Built-in restrictions on best linear unbiased predictors (BLUP) of random effects in mixed Models. *Amer. Statist.*, 51, 19–21.

[6] Tian, Y. (2004). More on maximal and minimal ranks of Schur complements with applications. *Appl. Math. Comput.*, 152, 675–692.

[7] Tian, Y. (2015). A new derivation of BLUPs under random-effects model. *Metrika*, DOI: 10.1007/s00184-015-0533-0.

Bayesian Covariance Modelling of Large Tensor-Variate Data Sets & Inverse Non-parametric Learning of the Unknown Model Parameter Vector

Kangrui Wang[1,*]   Dalia Chakrabarty[1,†]

[1]Department of Mathematics, University of Leicester

Abstract   Tensor-valued data are being encountered increasingly more commonly, in the biological, natural as well as the social sciences. The learning of the unknown model parameter vector given such data, involves covariance modelling of such data, though this can be difficult owing to the high-dimensional nature of the data, where the numerical challenge of such modelling can only be compounded by the largeness of the available data set. Assuming such data to be modelled using a correspondingly high-dimensional Gaussian Process ($\mathcal{GP}$), the joint density of a finite set of such data sets is then a tensor normal distribution, with density parametrised by a mean tensor $\boldsymbol{M}$ (that is of the same dimensionality as the k-tensor valued observable), and the k covariance matrices $\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k$. When aiming to model the covariance structure of the data, we need to estimate/learn $\{\boldsymbol{\Sigma}_1...\boldsymbol{\Sigma}_k\}$ and $\boldsymbol{M}$, given tha data. We present a new method in which we perform such covariance modelling by first expressing the probability density of the available data sets as tensor-normal. We then invoke appropriate priors on these unknown parameters and express the posterior of the unknowns given the data. We sample from this posterior using an appropriate variant of Metropolis Hastings. Since the classical MCMC is time and resource intensive in high-dimensional state spaces, we use an efficient variant of the Metropolis-Hastings algorithm–Transformation based MCMC–employed to perform efficient sampling from a high-dimensional state space. Once we perform the covariance modelling of such a data set, we will learn the unknown model parameter vector at which a measured (or test) data set has been obtained, given the already modelled data (training data), augmented by the test data.

Keywords   Bayesian inference; Tensor-normal distribution; High-dimensional data

1   Introduction

Let the causal relationship between observable $\boldsymbol{V}$ and model parameter $\boldsymbol{S}$ be defined as $\boldsymbol{V} = \boldsymbol{\xi}(\boldsymbol{S})$, where $\boldsymbol{V}$ is tensor-variate: $\boldsymbol{V} \in \mathbb{R}^{m_1 \times m_2 ... \times m_k}$. We want to estimate value $\boldsymbol{s}^{(test)}$ of $\boldsymbol{S}$ at which test data $\mathbf{D}^{(test)}$–i.e. measured value(s) of $\boldsymbol{V}$–is (are) realised. To do this, we need to learn function $\boldsymbol{\xi}(\cdot)$, which in this case is a tensor-variate function–of the model parameter vector $\boldsymbol{S} \in \mathbb{R}^d$. In the presence of training data $\mathbf{D}$, such supervised learning can be possible by fitting known parametric forms (such as splines/wavelets) to the training data to learn the form of $\boldsymbol{\xi}(\cdot)$, which can thereafter be inverted and operated upon the test data to yield $\boldsymbol{s}^{(test)}$. Here, training data $\mathbf{D}$ is this set of n values of $\boldsymbol{V}$, each generated at a design point, i.e. a chosen value $\boldsymbol{s}^{(*)}$ of $\boldsymbol{S}$. Thus, $\mathbf{D} := \{(\boldsymbol{v}_1, \boldsymbol{s}_1^{(*)}), \dots, (\boldsymbol{v}_n, \boldsymbol{s}_n^{(*)})\}$. However, fitting with splines/wavelets is inadequate in that it does not capture the correlations between

---

*Email: kw202@le.ac.uk
†Email: dc252@le.ac.uk

the components of a high-dimensional function; also, the computational complications of such fitting–and particularly of inversion of the learnt $\boldsymbol{\xi}(\cdot)$–increases rapidly with increase in dimensionality. Thus, we resort to the modelling of this high-dimensional data using a correspondingly high-dimensional Gaussian Process (GP), i.e. a tensor-variate GP.

2 Method

Thus, the joint probability distribution of a set of n realisations of the $k-1$-variate $\boldsymbol{V}$ is a k-variate normal distribution with mean $\boldsymbol{M}$ and k covariance matrices:

$$p(\boldsymbol{V}|\boldsymbol{M}, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_K) \propto \exp(-\|(\boldsymbol{V} - \boldsymbol{M}) \times_1 \boldsymbol{A}_1^{-1} \times_2 \boldsymbol{A}_2^{-1}... \times_k \boldsymbol{A}_k^{-1}\|^2/2) \qquad (1)$$

where the covariance matrix $\boldsymbol{\Sigma}_p = \boldsymbol{A}_p \boldsymbol{A}_p^T$, $p = 1, ..., k$. Tensor-variate normal distribution is extensively discussed in the literature, (Xu, Yan & Yuan [4]; Hoff [1])

Equation 1 implies that the likelihood of n values $(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n)$ of $\boldsymbol{V}$ given the unknown tensor-variate parameters of the GP is k-tensor variate normal. So, we write this likelihood and thereafter the posterior probability density of these unknown tensor-variate parameters given the training data (subsequent to the invoking of the priors on each unknown). Once this is achieved, we then sample from the posterior using an MCMC technique to achieve marginal density distributions of each uknown. The learning of $\boldsymbol{s}^{(test)}$ could be undertaken by writing the posterior predictive distribution of $\boldsymbol{S}$ given the test data $\boldsymbol{v}^{(test)}$, and given the tensor-variate parameters learnt using the training data. However, we decide to write the joint posterior probability density of $\boldsymbol{s}^{(test)}$ and all the other tensor-variate parameters given training+test data, and sample from this density to obtain the marginals of all the unknowns.

The first step is to write the likelihood of $\mathbf{D}$ given the tensor-variate mean and co-variance matrices of the GP. Here the mean matrix is $\boldsymbol{M} \in R^{m_1 \times m_2 ... \times m_k}$. It may be possible to estimate the mean as a function of s and be removed from the non-zero mean model. Under these circumstance, a general method of estimation, like maximum likelihood estimation or least square estimation, can be used. Then, the Gaussian Process can be converted into a zero mean GP. However, if necessary, the mean tensor itself can be regarded as a random variable and learnt from the data [3]. The modelling of the covariance structure of this GP is discussed in the following subsection.

2.1 Covariance structure

In this context, it is relevant that a k-dimensional random tensor $\boldsymbol{\Sigma} \in R^{m_1 \times m_2 ... \times m_k}$ can be decomposed to a unit random k dimensional tensor ($\boldsymbol{Z}$) and k number of covariance matrix by **Tucker product** [1]:

$$\boldsymbol{\Sigma} = \boldsymbol{Z} \times_1 \boldsymbol{\Sigma}_1 \times_2 \boldsymbol{\Sigma}_2... \times_k \boldsymbol{\Sigma}_k \qquad (2)$$

where the p-th covariance matrix is $m_p \times m_p$ matrix and $m_p \in \mathbb{Z}_{>0}$, $m_p \in \{m_1, m_2, \dots, m_k\}$ for the tensor $\boldsymbol{\Sigma}$ that is $m_1 \times m_2 \times \dots \times m_k$-dimensional .

We choose to model the covariance structure of the GP with a Squared Exponential (SQE) covariance function. The implementation of this can be expressed in different ways, but in this initial phase of the project, we perform parametrisation of the covariance structure using the Tucker Product that has been extensively studied[1]. It is recalled that the SQE form can be expressed as

$$p(\boldsymbol{V}|\boldsymbol{M}, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k) = (2\pi)^{-m/2}(\prod_{i=1}^{k} |\boldsymbol{\Sigma}_i|^{-m/2m_i}) \times \exp(-\|(\boldsymbol{V}-\boldsymbol{M}) \times_1 \boldsymbol{A}_1^{-1} \times_2 \boldsymbol{A}_2^{-1}... \times_k \boldsymbol{A}_k^{-1}\|^2/2)$$

$$(3)$$

where $m = \prod_{i=1}^{k} m_i$ and $\boldsymbol{\Sigma}_p = \boldsymbol{A}_p \boldsymbol{A}_p^T$.

Although this probability density function is well structured and can in principle be used to model high dimensional data, the computational complicacy increases with large and/or high-dimensional data sets. If we do not implement a particular parametric model for the covariance kernels but aim to learn each element of each covariance matrix, the total number of parameters in the covariance structure to be then learnt, ends up as $\sum_{p=1}^{k} m_p^2$. This could be a big number for a large data set and the computational demand on such learning can be formidable. Also, the computational task of inverting the covariance matrix $\boldsymbol{\Sigma}_p$ is in itself highly resource intensive, with the demand on time and computational resources increasing with the dimensions of $\boldsymbol{\Sigma}_p$, $p = 1, \ldots, k$.

3 Application

We perform an empirical illustration of our method, to first learn the covariance structure of a large astronomical training data set, and thereafter, employ such learning towards the prediction of the value of the unknown model parameter at which the test data is realised. The training data comprises has 216 observations, where an observation constitutes a sequence of 2-dimensional vectors. In fact, each such 2-dimensional vector is a 2-dimensional velocity vector of a star that is a neighbour of the Sun, as tracked within an astronomical simulation [2] of the disk of our Galaxy. There are 50 stars tracked at each design point i.e. at each assigned value of the unknown model parameter vector, that is in this application is the location of the Sun in the two-dimensional, (by assumption), Milky Way disk. In other words, $\boldsymbol{S}$ itself is a 2-dimensional vector. There are 216 design points used to generate this (simulated) training data that then constitutes 216 number of 50×2-dimensional velocity matrices, with each velocity matrix generated at each of the 216 design points in this training data. Thus, the training data $\mathbf{D}$ in this application is $216 \times 50 \times 2$-dimensional 3-tensor

To reduce the difficulty of MCMC algorithm, the mean tensor is estimated by the maximum likelihood estimation.

When building the covariance structure of this training data set, the likelihood of which is now 3-tensor-normal, we consider three covariance matrices. Of these, the 216×216-dimensional covariance matrix $\boldsymbol{\Sigma}_1$ bears information about the correlation between velocity matrices generated at the 216 different values of $\boldsymbol{S}$, i.e. at the 216 different solar locations in the Milky Way disk. The 50×50 covariance matrix $\boldsymbol{\Sigma}_2$ illustrates the correlation between any pair of the 50 stars at a given $\boldsymbol{s}$, that are tracked in the astronomical simulation and the last covariance matrix $\boldsymbol{\Sigma}_3$ represents the correlation between the 2 components of the velocity vector of a star that is tracked at a given $\boldsymbol{s}$ for its velocty in the astronomical simulation. If we learn the elements of each covarianc matrix directly, we will have $216 \times 216 + 50 \times 50 + 2 \times 2$ number of parameters to learn, which is too many given limits of time and computational resourse. Thus, we model the covariance kernels using known forms, the parameters of which we then learn from the data.

In particular, we use the Squared Exponential (SQE) covariance function to model the 216×216 matrix $\boldsymbol{\Sigma}_1$ and learn the correlation lengths–or rather their reciprocals, the smoothing parameters–using the training data. As the 216 velocity matrices are each generated at a respective value of $\boldsymbol{S}$, $\boldsymbol{\Sigma}_1$ can be written as $\boldsymbol{\Sigma}_1 = [a_{ij}]$ where $i, j = 1, \ldots, 216$ with

$$a_{ij} = \exp\left[-\left(\boldsymbol{s}_i - \boldsymbol{s}_j\right)^T \boldsymbol{Q}_1 \left(\boldsymbol{s}_i - \boldsymbol{s}_j\right)\right],$$

where $\boldsymbol{Q}_1$ is a $d \times d$ square diagonal matrix, with $\boldsymbol{S} \in \mathbb{R}^d$. As $d = 2$ in our application, we learn 2 smoothness parameters.

The covariance matrix $\boldsymbol{\Sigma}_2$ quantifies correlation amongst the different stellar velocity vectors generated at a given $\boldsymbol{s}$. The 50 stellar velocity vectors that are recorded at a given $\boldsymbol{s}$ are chosen over other values of stellar velocity vectors. Given that the velocity vector of each star is 2-dimensional, we again learn 2 smoothness parameters (diagonal elements of matrix $\boldsymbol{Q}_2$), using an SQE model.

In addition we learn the 4 parameters of the covariance matrix $\boldsymbol{\Sigma}_3$.

Thus, we will have 8 parameters $(q_{11}^{(1)}, q_{22}^{(1)}, q_{11}^{(2)}, q_{22}^{(2)}, \sigma_{11}^{(3)}, \sigma_{12}^{(3)}, \sigma_{21}^{(3)}, \sigma_{22}^{(3)})$ of the covariance structure to learn from the data, where these parametersare defined as in:

$$\boldsymbol{Q}_1 = \begin{pmatrix} q_{11}^{(1)} & 0 \\ 0 & q_{22}^{(1)} \end{pmatrix}; \boldsymbol{Q}_2 = \begin{pmatrix} q_{11}^{(2)} & 0 \\ 0 & q_{22}^{(2)} \end{pmatrix}; \boldsymbol{\Sigma}_3 = \begin{pmatrix} \sigma_{11}^{(3)} & \sigma_{12}^{(3)} \\ \sigma_{21}^{(3)} & \sigma_{22}^{(3)} \end{pmatrix}$$

In the initial phase of the project that is currently underway, we write the joint posterior probability density of the unknown parameters and sample from it using a variant of the metropolis-Hastings algorithm, referred to as Transformation-based MCMC (TMCMC). To write the posterior, we impose uniform priors on each of our unknowns.

Table 1: Priors for parameters

Parameters		Prior
$q_{11}^{(1)}$	Uniform	$\pi(q_{11}^{(1)}) \propto 1$
$q_{22}^{(1)}$	Uniform	$\pi(q_{22}^{(1)}) \propto 1$
$q_{11}^{(2)}$	Uniform	$\pi(q_{11}^{(2)}) \propto 1$
$q_{22}^{(2)}$	Uniform	$\pi(q_{22}^{(2)}) \propto 1$
$\boldsymbol{\Sigma}_3$	Non-informative	$\pi(\boldsymbol{\Sigma}_3) \propto |\boldsymbol{\Sigma}_3|^{-1/2}$

The proposal density that we use in our MCMC scheme, to generate updates for each of our parameters is tabulated within the section in which TMCMC is described. The results of our learning and estimation of the mean and covariance structure of the GP used to model this tensor-variate data, is discussed below in Section 5. Once this phase of the work is over, we will proceed to include the test and training data both, to write the joint posterior probability density of $\boldsymbol{s}^{(test)}$ and the 8 unknowns in $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{\Sigma}_3$, and learn all these parameters.

4 Transformation based MCMC

We are using the Transformation based MCMC algorithm to estimating the parameters. Although the TMCMC method will lose some of the information, the method is efficient in high dimensional distributions.

- 1.Set initial value $s_0, q_0^{(1)}, \ldots, q_0^{(k)}$, counter $n = 1$ and a forward probability $p_0, \ldots, p_k$
- 2.Generate $e \sim Gamma(1,1)$ and $u \sim \mathcal{U}(0,1)$ independently.
- 3.If $u < p_0$, let $s' = s_{n-1} + \beta_0 e$. Else, let $s' = s_{n-1} - \beta_0 e$
- 4.Repeat step 2 and step 3 for $q_1', \ldots, q_k'$.
- 5.Calculate the acceptance rate:

$$\alpha = \frac{\prod_{i \in D} p_i \times \prod_{j \in D^c}(1 - p_j)}{\prod_{i \in D}(1 - p_i) \times \prod_{j \in D^c} p_j} \times \frac{posterior(s', q_1', \ldots q_k')}{posterior(s_{n-1}, q_{n-1}^{(1)}, \ldots q_{n-1}^{(k)})}$$

where, set D is the elements which has the backward transform($u > p$) and set D^c is the elements which has the forward transform($u <= p$).

- 6.Accept $s', q_1', \ldots, q_k'$ as $s_n, q_n^{(1)}, \ldots, q_n^{(k)}$ with probability α or drop $s', q_1', \ldots, q_k'$ with probability $1 - \alpha$
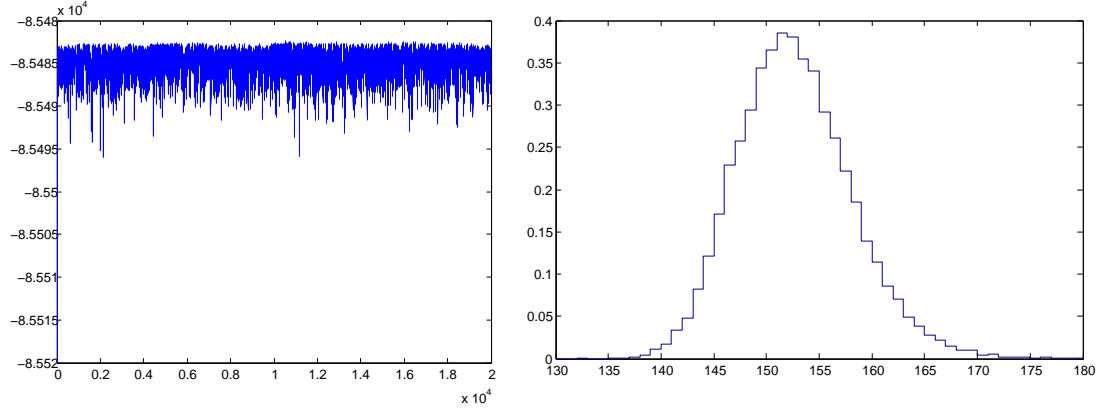- 7.Repeat 2 to 6 until the chain get convergence.

5 Results

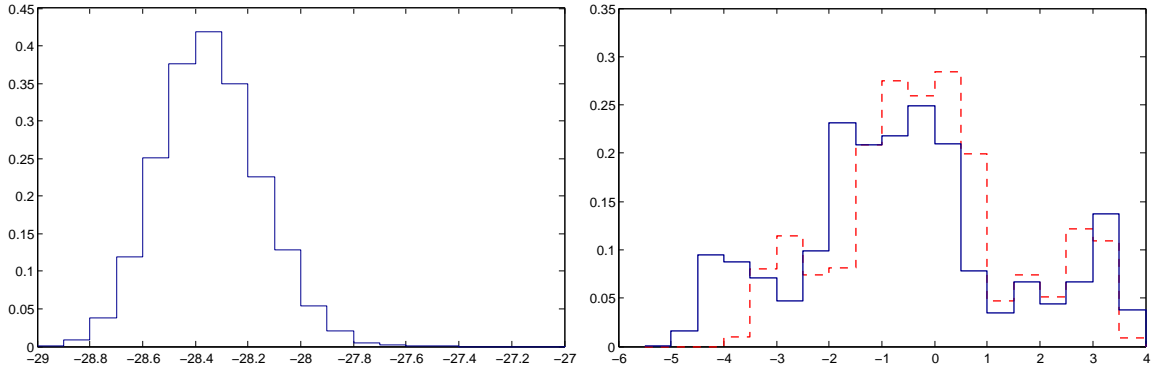Figure 1: trace of the likelihood generated by Figure 2: marginal probability density for $q_{11}^{(1)}$ TMCMC

Figure 3: marginal probability density for $\boldsymbol{\sigma}_{11}^{(3)}$

Figure 4: marginal probability density for $\boldsymbol{\sigma}_{12}^{(3)}$ in full line and $\boldsymbol{\sigma}_{21}^{(3)}$ in broken line

In the top left panel of Figure 1, we present the trace of the likelihood of the training data given the 8 unknowns in $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{\Sigma}_3$, with 2×10^4 of iterations. The stationarity of the trace betrays the achievement of convergence of the chain.

The marginal posterior probability densities of each unknown parameter is also learnt using TMCMC. The same for parameters $q_{11}^{(1)}$ (Figure 2), $\boldsymbol{\sigma}_{11}$ (Figure 3) and $\boldsymbol{\sigma}_{12}$ (Figure

4)are shown in the top right and bottom left and right panels. As noticed in the inequality of the marginals of the non-diagonal elements of $\mathbf{\Sigma}_3$ shown in the bottom panels of this figure, the covariance structure for this astronomical data set does not appear to adhere to stationarity. Had the covariance been stationary, the $1, 2$-th and $2, 1$-th elements would be equal, i.e. their marginals would coincide. But such is not the case as evident from comparing the two density in figure 4 which shows a drift from $\boldsymbol{\sigma}_{12}$ to $\boldsymbol{\sigma}_{21}$. This further suggests that our modelling of the $\mathbf{\Sigma}_2$ matrix using SQE covariance function is pre-matured. We are exploring the implementation of non-stationary covariance modelling of $\boldsymbol{s}$.

References

[1] P. D. Hoff (2011). *Bayesian Analysis*, 6, Number 2, pp. 179-196.

[2] S.Banerjee, A.Basu, S.Bhattacharya, S.Bose, D.Chakrabarty, S.S.Mukherjee (2015). *Minimum Distance Estimation of Milky Way Model Parameters and Related Inference*, SIAM/ASA Jl. of Uncertainty Quantification, arXiv:1309.0675.

[3] D.Chakrabarty, M.Biswas ,S.Bhattacharya (2013). *Bayesian Nonparametric Estimation of Milky Way Model Parameters Using a New Matrix-Variate Gaussian Process Based Method*, arXiv:1304.5967

[4] Z. Xu ,F. Yan, Y.Qi(2011). *Infinite Tucker Decomposition: Nonparametric Bayesian Models for Multiway Data Analysis*, arXiv:1108.6296.

A New Sampler: Randomized Likelihood Sampling

Yuchung J. Wang[1,*]   Jian-Hui Ning[2,†]   Yong-Dao Zhou[3,‡]
Kai-Tai Fang[4]

[1]Department of Mathematical Sciences, Rutgers University, Camden, NJ 08102, USA
[2]School of Mathematics and Statistics, Center China Normal University, Wuhan 430079, CHINA
[3]College of Mathematics, Sichuan University, Chengdu 610064, CHINA
[4]Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519085, CHINA

Abstract   In this paper, a new sampler is proposed to obtain independent and identically distributed
samples from a given univariate or multivariate distribution kernel. The sampler is called randomized
likelihood sampling (RLS) because it randomizes the likelihood and it selects samples according to the
likelihood of the target distribution. RLS utilizes the uniformity of a quasi-random sequence to search
the support,and applies randomization to achieve independence. Because the likelihood is computed
from a kernel of the target distribution, RLS has wide applicability. For example, RLS can sample
multimodal kernels without getting stuck in localities. A bootstrap procedure is proposed to compute
the Monte Carlo error. Some numerical comparisons are reported.

Keywords   Bootstrap; Good lattice points; Markov chain Monte Carlo; Monte Carlo error; Sampling
importance resampling.

1   Background and Introduction

Generating independent and identically distributed (IID) samples from a probability
density function (pdf) is an ancient but also modern topic for both statistics and computer
science, see the recent developments of Markov Chain Monte Carlo (MCMC). The algo-
rithms of such tasks are called samplers, which can be categorized as either pdf-specific
or general-purpose. For example, the Box-Muller algorithm is specific for one-dimensional
normal pdf, and the Chelosky decomposition applies only for the multivariate normal distri-
butions. On the other end, the accept-reject method (ARM) is applicable to any univariate
pdf, but its performance, such as accuracy, rejection rate and consistency, will depend on
the proposal pdf being selected. Distribution-inversion method is another general-purpose
algorithm conditioned on that the distribution function can be explicitly inverted.

Sampling multivariate pdf is much more involved than sampling univariate pdf, and
the task is further complicated by the cases that the normalizing constants may not be
available, only density kernels are known. For example, the predictive function is unknown
in many Bayesian models. When only kernel, not density, is available, pdf-specific samplers
are limited to distributions that are commonly used, such as the normal family and those
related to the multivariate exponential distribution. Metropolis-Hastings algorithm (MH)

---

is the first general-purpose sampler, and it has advanced the Bayesian calculations by leaps and bounds. Similar to ARM, MH requires a proposal pdf that proposes **one** candidate to be accepted as a sample or rejected. A large amount of research has been devoted to the design of a proposal pdf which must balance between a decent acceptance rate and fast exploration of the support. Often, decent proposal distribution is problem-dependent. Another general-purpose sampler is Rubin's [3] sampling importance/resampling (SIR) sampler, and it re-samples from a large pool of candidates that are generated from some proposal distribution. Poor choice of proposal can make SIR inadequate. The fundamental differences between MH and SIR are (a) MH is a MCMC that generates Markov-dependent samples, while SIR generates IID samples; and (b) MH accepts/rejects one candidate, while SIR accepts samples from many candidates. Both samplers are considered to be general-purpose, nevertheless custom-made proposals are required to generate efficient samples.

Here, we propose a general-purpose sampler such that (a) it does not require a proposal pdf, hence, it is problem-independent; (b) it selects multiple samples from a pool of candidates; (c) it generates IID samples; (d) the bias of a fixed sample sizes can be corrected; and (e) the Monte Carlo Error (MCE) can also be estimated. We name it randomized likelihood sampling (RLS) with bootstrap. The key ingredients for RLS are uniformly distributed points (UD) over the support and random-shift of UD. RLS begins with a quasi-random sequence as the candidates' pool, then iterates among the following three steps: (i) it computes the likelihood of all the candidates; (ii) then it randomly selects multiple samples from the candidates' pool via their likelihoods; and last (iii) it creates a new pool of candidates by randomly shifting the current pool. The likelihoods are a discrete approximation to the continuous pdf, and UD is known to be efficient for such applications, see Fang and Wang [1] and Hua and Wang [2]. With a good UD, the size of the pool needs not be large. From every pool, a batch (multiple) of samples are selected, and one sample from every batch constitutes an IID sample.

In section 2, we detail the approach. In section 3, we use a mixture of univariate normal distribution to illustrate the bias-correction and MCE of RLS. Section 4 gives some discussions.

2   Randomized Likelihood Sampling

Let $\pi(\boldsymbol{x})$ be the target pdf whose support $\mathcal{S}$ is d-dimensional. Function f is a density kernel of π if and only if $f(\boldsymbol{x})/f(\boldsymbol{y}) = \pi(\boldsymbol{x})/\pi(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}$. RLS constitutes of the following steps:

1. (compact region). Properly choose $a_i < b_i, i = 1, \ldots d$, such that probability $Pr(\boldsymbol{x} \in D \cap \mathcal{S}) = 1$ or is close to 1, where $D = \bigotimes_{i=1}^{d}[a_i, b_i]$. Let $L_i = b_i - a_i$.
2. (UD). Generate a set of N quasi-random numbers, $Q^0 = \{q^w = (q_{w_1}, \ldots, q_{w_d}), w = 1, 2, \cdots, N\}$ that are uniformly scattered over D. Set t=0.
3. (random rotation). Generate d mutually independent v_i, where every v_i is uniformly distributed over $[0, L_i]$. Use $v^t = (v_1, \ldots, v_d)$ to perturb Q^0 into $X = \{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N\}$, where $\boldsymbol{x}^w = (x_{w_1}, \ldots, x_{w_d})$, $x_{w_i} = q_{w_i} + v_i$ if $a_i \le q_{w_i} + v_i \le b_i$; otherwise, $x_{w_i} = q_{w_i} + v_i - L_i + a_i$. Set X is the candidates' pool, and let t=t+1.
4. (likelihood). For $\boldsymbol{x}^w \in \mathcal{S}$, compute $O_w = f(\boldsymbol{x}^w)$, and for $\boldsymbol{x}^w \notin \mathcal{S}$, set $O_w = 0$. Then scale $\{O_w, 1 \le w \le N\}$ into likelihood $(p_1, \cdots, p_N)$, and $p_i = O_i / \sum_{1}^{N} O_w$.
5. (sample selection). Select $m < N$ random samples from X based on the weights $(p_1, \cdots, p_N)$. Denote the chosen sample batch as $\mathcal{E}^t = \{y^t(1), \cdots y^t(m)\}$, where t represents the iteration clock.

6. (iteration) After repeating Step 3 to Step 5 n times, we have n batches of samples, $\mathcal{E}^t, 1 \leq t \leq n$.

Because quasi-random numbers are deterministic, samples of $\mathcal{E}^t$ are dependent. However, every candidates' pool is independent of other candidates pools, hence taking one $y^t(w_j)$ from every $\mathcal{E}^t$ constitute an IID samples: $\mathcal{B}_b = \{y^1(w_1), \cdots, y^n(w_n), y^j(w_j) \in \mathcal{E}^j\}$. There are m^n different such $\mathcal{B}_b$, which is called a RLS bootstrap sample. Let $F_n(\mathcal{B}_b)$ be the empirical distribution of $\mathcal{B}_b$ and let $\hat{\theta}_b = \hat{\theta}(F_n(\mathcal{B}_b))$ be the estimate of θ based on $F_n(\mathcal{B}_b)$, where $\theta = \theta(\pi)$ is any parameter of interest. After a large number, say B, $\hat{\theta}_b$ have been obtained, the θ and its Monte Carlo error are estimated, respectively by $\overline{\hat{\theta}}$ and BSE, where

$$\overline{\hat{\theta}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_b, \text{ and } \text{BSE} = \sqrt{(1/B)\sum_{b=1}^{B}(\hat{\theta}_b - \overline{\hat{\theta}})^2}. \tag{1}$$

There are many good options for choosing Q^0, and we settle with sets of Good Lattice Points (GLP). A d-dimensional GLP set $G = \{(g_{ij}), 1 \leq i \leq N\}$ is generated by a generating vector, $(N; h_1, \ldots, h_d)$, as follows:

$$g_{ij} = ih_j(\text{mod } N), 1 \leq j \leq d,$$

where $1 = h_1 < h_2 < \ldots < h_d < N$, and N is mutually prime with every h_j. The modulo operation ensures that $1 \leq g_{ij} \leq N$. A linear transformation then converts G into Q^0 over D. One-dimensional GLP set is $\{(2i-1)/2N, 1 \leq i \leq N\}$.

3 Numerical Examples

We use two examples to illustrate the performance of RLS in comparison with other samplers. The first one is the equally weighted mixture of two univariate normal pdf: $\mathcal{N}(1,1)$ and $\mathcal{N}(3,2)$.

Example 1.
We compare RLS with two IID samplers: (1) the pdf-specific rnorMix of R package "normalmix", and (2) the general-purpose SIR using t-distribution of 2 degree of freedom as the proposal pdf. In addition, the slice sampler (a MCMC) is also compared.

For RLS, the size of GLP is $N = 800$ with $m = 60$ samples per $\mathcal{E}^t$. The sample size is 300 for all four samplers. Once 300 samples are drawn by a sampler, they are bootstrapped $B = 500$ times to compute the $\overline{\hat{\theta}}$ (mean column) and BSE (SD column) from the B estimates. The parameters of interest are mean, variance, median, first quartile and third quartile. The results are shown in Table 1.

In terms of biases, RLS is the best among four competing samplers. The variance estimate of SIR is not acceptable, though its other biases are not this bad; explanations of SIR's poor performance include inefficient proposal pdf plus small sample size. The SDs of SIR are also too small to be correct. Because norMix is a well-documented sampler, its SDs are considered to be correct. The SDs of RLS are on par or 10% smaller than those of norMIX. The accuracy of slice is quite on par with norMIX.

Example 2.
Next we consider an 8-dimensional symmetric Kotz-type distribution, and there is no known algorithm to sample from its kernel, which is as follows:

$$f(\boldsymbol{x}) \propto |\Sigma|^{-1/2}[(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu)]^{I-1}\exp\{-b[(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu)^s]\}.$$

When $I = 1$, $b = 0.5$, and $s = 1$, it is the multivariate normal distribution. Here we set $I = 3$, $s = 2$, $b = 0.5$, $\mu = (0, 0, 0, 0, 0, 0, 0, 0)^T$, and

$$\Sigma = \begin{pmatrix} 7.1 & 0.4 & -0.3 & -0.3 & -0.3 & -0.2 & -0.4 & -0.3 \\ 0.4 & 6.3 & -0.8 & -0.2 & -0.1 & -0.2 & 0.0 & -0.3 \\ -0.3 & -0.8 & 6.8 & -0.7 & -0.3 & -0.6 & -0.4 & 0.4 \\ -0.3 & -0.2 & -0.7 & 5.8 & 0.1 & 0.1 & 0.0 & -0.5 \\ -0.3 & -0.1 & -0.3 & 0.1 & 6.8 & -0.3 & -0.2 & -0.4 \\ -0.2 & -0.2 & -0.6 & 0.1 & -0.3 & 6.3 & -0.4 & 0.0 \\ -0.4 & 0.0 & -0.4 & 0.0 & -0.2 & -0.4 & 5.6 & 0.0 \\ -0.3 & -0.3 & 0.4 & -0.5 & -0.4 & 0.0 & 0.0 & 5.3 \end{pmatrix}.$$

As shown in Fang, Kotz, and Ng (1990), μ is the mean vector and the covariance matrix is $\frac{r^{-1/s}}{rank(\Sigma)} \frac{\Gamma((2I+d)/2s)}{\Gamma((2I+d-2)/2s)} \Sigma$.

Let the compact cube D have lower bounds $\{-5.1, -4.8, -4.9, -4.6, -4.9, -4.8, -4.5, -4.4\}$ and upper bounds $\{5.1, 4.8, 4.9, 4.6, 4.9, 4.8, 4.5, 4.4\}$, respectively. The generating vector $(3997; 1, 3888, 3564, 3031, 2311, 1417, 375, 3211)$ is chosen from Hua and Wang [2]. For this example, RLS selects one sample from every pool, thus it generate IID samples. From $R = 200$ replications of 500 samples of RLS, we compute the average and standard deviation of the estimates. Average biases of RLS are listed in Table 2, and the MCE is listed in Table 3. With only 500 samples to estimate 64 covariances and 8 means, RLS performs very well. RLS's average biases for the mean vector are ± 0.004 and the average biases for the covariances matrix are all within ± 0.056. It would be difficult to implement MCMC algorithms for this example. For MH, a good 8-dimensional proposal distribution is hard to find. For SIR, we could use an 8-dimensional t-distribution as the proposal pdf without knowing the suitability. Moreover, since the univariate full conditionals are unknown, the Gibbs sampler must use either slice within Gibbs or MH within Gibbs; both implementations are complicated. Only the slice sampler can be implemented for a comparison, and the results are listed in Table 4 and Table 5. Based on Table 2-5, RLS achieves about the same average accuracy as the slice sampler, nevertheless, RLS is four times more consistent

Table 1: The comparison among RLS, slice sampler, SIR and R-sampler norMix

	Sampler RLS				Slice Sampler		
	mean	SD	bias	true values	mean	SD	bias
mean	1.9786	0.1018	-0.0214	2	2.0117	0.1094	0.0117
var	3.4978	0.3057	-0.0022	3.5	3.5633	0.2898	0.0633
0.25Q	0.6747	0.1023	-0.0104	0.6852	0.5272	0.1062	-0.158
med	1.6703	0.1229	0.0036	1.6667	1.6704	0.1174	0.0038
0.75Q	3.0646	0.1915	-0.0269	3.0915	3.3375	0.249	0.246
	SIR				R-sample norMix		
	mean	SD	bias		mean	SD	bias
mean	1.7248	0.0562	-0.2752		1.9725	0.1054	-0.0275
var	1.0098	0.1048	-2.4902		3.4102	0.3202	-0.0898
0.25Q	1.0983	0.0725	0.4132		0.6772	0.0853	-0.008
med	1.742	0.0725	0.0754		1.6652	0.143	-0.0015
0.75Q	2.2968	0.0684	-0.7947		2.9313	0.2937	-0.1602

Table 2: Average bias of the mean vector and the covariance matrix by RLS

Cov	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0.025	0.050	-0.001	-0.006	0.048	-0.056	0.002	0.039
x_2	0.050	0.030	0.001	-0.006	0.049	-0.056	-0.002	0.040
x_3	-0.001	0.001	-0.020	0.002	-0.003	0.006	0.006	-0.004
x_4	-0.006	-0.006	0.002	-0.020	-0.009	0.006	0.000	-0.006
x_5	0.048	0.049	-0.003	-0.009	0.026	-0.051	-0.002	0.039
x_6	-0.056	-0.056	0.006	0.006	-0.051	0.036	0.000	-0.039
x_7	0.002	-0.002	0.006	0.000	-0.002	0.000	-0.016	0.000
x_8	0.039	0.040	-0.004	-0.006	0.039	-0.039	0.000	0.012
Mean	0.000	-0.001	0.004	0.000	0.001	0.001	0.001	-0.001

The average biases are calculated from 200 repetitions of RLS simulation, each of size 500.

Table 3: MCE of the estimates of the mean vector and the covariance matrix by RLS

Cov	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0.137	0.113	0.093	0.082	0.115	0.123	0.084	0.098
x_2	0.113	0.121	0.082	0.078	0.110	0.113	0.076	0.094
x_3	0.093	0.082	0.124	0.085	0.088	0.090	0.086	0.083
x_4	0.082	0.078	0.085	0.104	0.079	0.077	0.074	0.077
x_5	0.115	0.110	0.088	0.079	0.126	0.114	0.077	0.094
x_6	0.123	0.113	0.090	0.077	0.114	0.128	0.078	0.101
x_7	0.084	0.076	0.086	0.074	0.077	0.078	0.098	0.079
x_8	0.098	0.094	0.083	0.077	0.094	0.101	0.079	0.105
Mean	0.086	0.079	0.064	0.059	0.081	0.086	0.060	0.070

The MCEs are calculated from 200 repetitions of RLS simulation, each of size 500.

Table 4: Average bias of the mean vector and the covariance matrix by the slice sampler

Cov	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	-0.050	-0.003	0.002	0.002	0.007	0.002	0.008	0.003
x_2	-0.003	-0.032	0.015	0.000	-0.005	-0.004	0.004	0.002
x_3	0.002	0.015	-0.036	0.001	0.007	0.004	0.004	-0.008
x_4	0.002	0.000	0.001	-0.031	-0.003	0.002	0.002	0.006
x_5	0.007	-0.005	0.007	-0.003	-0.043	0.002	-0.002	-0.001
x_6	0.002	-0.004	0.004	0.002	0.002	-0.042	0.004	0.001
x_7	0.008	0.004	0.004	0.002	-0.002	0.004	-0.032	0.003
x_8	0.003	0.002	-0.008	0.006	-0.001	0.001	0.003	-0.028
mean	-0.001	0.003	-0.003	0.001	0.002	-0.006	0.002	-0.002

The average biases are calculated from 200 repetitions of the slice sampler, each of size 500.

than the slice sampler because the MCEs of RLS are about a quarter of those of the slice sampler.

Table 5: MCE of the estimates of the mean vector and the covariance matrix by the slice sampler

Cov	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0.302	0.194	0.202	0.180	0.202	0.188	0.171	0.170
x_2	0.194	0.262	0.188	0.174	0.186	0.172	0.165	0.154
x_3	0.202	0.188	0.291	0.177	0.192	0.184	0.170	0.165
x_4	0.180	0.174	0.177	0.231	0.169	0.165	0.154	0.147
x_5	0.202	0.186	0.192	0.169	0.289	0.184	0.168	0.163
x_6	0.188	0.172	0.184	0.165	0.184	0.266	0.162	0.156
x_7	0.171	0.165	0.170	0.154	0.168	0.162	0.221	0.141
x_8	0.170	0.154	0.165	0.147	0.163	0.156	0.141	0.211
mean	0.213	0.190	0.204	0.180	0.204	0.193	0.174	0.168

The MCEs are calculated from 200 repetitions of slice samples each of size 500.

4   Discussion

In terms of sampling approach, RSL is similar to SIR but with some critical differences. SIR uses a fixed and huge set of pseudo-random samples as the pool of candidates, while RLS selects from much smaller and randomly rotated quasi-random sequences. Another advantage of RLS is that its pool size does not depend on the sample size, while SIR requires the pool size to be 80 to 100 folds of the sample size. For convergence, SIR requires that the ratio of pool size to sample size tends to infinity ($N/n \to \infty$), while RLS needs only $N \to \infty$. In theory, the uniformity of UD enables RLS to perform a thorough exploration of the support, while the efficiency of SIR will depend on the efficacy of the proposal pdf that produces the candidates' pool. Rotating the pool by an independent increment not only maintains the uniformity of the candidates but also makes the candidates' pools independent from iteration to iteration. Computations of both the GLP set and the random rotation are less costly than generating multivariate pseudo-random numbers. In short, RLS combines the idea of randomized quasi-random numbers, with the IID samples of SIR into a general-purpose sampler. The unique features of RLS include:

1. It can sample from probability kernels, that is, unnormalized densities.
2. It generates IID samples.
3. Every sample is a result of a global search over the entire support, which makes RSL not getting stuck in a locality. Therefore, sampling multimodal kernels becomes efficient.
4. It can sample multivariate distributions whose components are highly correlated because RLS dose not sample componentwise.
5. The sampling task can be easily partitioned and executed in a cluster of computers. Every computer can independently rotate the GLP and sample from them.

Taking multiple samples from every candidates' pool rather than selecting only one sample, RLS approximates the MCEs and reduces biases straightforwardly. In addition, RLS does not need burn-in and it is very easy to program. Because low-discrepancy sequences exit only over compact regions, RLS only generates samples within a finite region. Remedies for this issue exist. Monte Carlo always represents a finite approximation with the hope that the error of finiteness is acceptable; numerical results have shown that RLS achieves respectable accuracy and its reproducibility is on par with the best pdf-specific samplers.

References

[1] Fang, K.T. and Wang, Y. (1994). *Number-theoretic Methods in Statistics*, Chapman and Hall, London.

[2] Hua, L K and Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis*, Berlin and Beijing: Springer-Verlag and Science Press.

[3] Rubin, D. B. (1987). A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. Discussion of Tanner and Wong. *Journal of the American Statistical Association*, 82: 543-546.

Comparison of unbiased estimators using Pitman's measure of closeness

Jibo Wu[1,2,*]

[1] Key Laboratory of Group & Graph Theories and Applications, Chongqing University of Arts and Sciences, Chongqing 402160, China
[2] School of Mathematics and Finances, Chongqing University of Arts and Sciences, Chongqing 402160, China

Abstract   In this paper, under the assumption of normality, we give the comparison of unbiased estimator under the Pitman's measure of closeness (PMC). Its applications in linear regression are also discussed.

Keywords   Pitman's measure of closeness; Unbiased estimator; Mean squared error matrix; Linear regression model

1   Introduction

It is customary to apply mean squared error (MSE) and Mean squared error matrix (MSEM) as the criteria of preference of one estimator over the other for estimating a parameter θ. However, [1] have show that an estimator may be superior over another estimator in the sense of MSE but may be poorly in the sense of more intrinsic criteria such as Pitman's measure of closeness (PMC) [2]. So as a meaningful alterative criterion to MSE (MSEM), PMC has been discussed by many researches, such as, Keating and Mason [3], Ghosh and Sen [4], Nayak [5] and Keating et al. [3].

Many researches discussed the linear estimators in the sense of PMC; such as; Mason et al. [6], Peddada and Khattree [7], Fountain and Keating [8], Fountain [9], Wang and Yang [10], and Yan [12]. Some sufficient and necessary conditions for the determination of the PMC of two linear estimators were got in Mason et al. [6]. Yan [12] studied the comparison between the two linear estimators and obtained some simple and clear arguments for them. However, in practice, there are a few estimators which are satisfied the conditions which Mason et al. and Yan [12] have given in their's paper.

Fountain and Keating [8] compared two unbiased linear estimator in the PMC sense. Yan [12] compared two estimators in the PMC sense, but there results is not very good. To the best of our knowledge, there has not been any comparison between the non-linear unbiased estimator yet. Our goal in this paper is to present the comparison between the two unbiased estimator in the PMC sense.

The rest of the paper is organized as follows. The comparison between two unbiased estimators under the PMC sense is given in Section 2. Some applications for the theorem given in Section 2 in linear regression model are given in Section 3. Some conclusion remarks are given in Section 4.

2   Main results

Firstly, we give definitions of PMC, MSE, and MSEM below.

Definition 2.1. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of the unknown p-dimensional vector θ. The PMC of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ in estimating θ under a loss function $L(.,\theta)$ is defined as $P_r(\hat{\theta}_1,\hat{\theta}_2,\theta) = P_r(\Delta(\hat{\theta}_1,\hat{\theta}_2) \geq 0)$, where

$$\Delta(\hat{\theta}_1,\hat{\theta}_2) = L(\hat{\theta}_2,\theta) - L(\hat{\theta}_1,\theta) \tag{1}$$

In this paper, we consider the quadratic loss function $L(\hat{\theta},\theta) = (\hat{\theta}-\theta)'(\hat{\theta}-\theta)$.

Definition 2.2. $\hat{\theta}_1$ is said to dominate $\hat{\theta}_2$ for all $\theta \in \Theta$ in PMC (under the loss function $L(.,\theta)$, for some parameter space Θ), if

$$P_r(\hat{\theta}_1,\hat{\theta}_2,\theta) = P_r(\Delta(\hat{\theta}_1,\hat{\theta}_2) \geq 0) \geq \frac{1}{2}, \; for \; all \; \theta \in \Theta \tag{2}$$

Definition 2.3. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of the unknown p-dimensional vector θ, then the mean squared error matrix (MSEM) and mean squared error (MSE) are defined as follows:

$$MSEM(\hat{\theta}) = E\{(\hat{\theta}-\theta)(\hat{\theta}-\theta)'\} \tag{3}$$

and

$$MSE(\hat{\theta}) = E\{(\hat{\theta}-\theta)'(\hat{\theta}-\theta)\} \tag{4}$$

respectively.

Definition 2.4. $\hat{\theta}_1$ is said to dominate $\hat{\theta}_2$ for all $\theta \in \Theta$ in MSEM and MSE criteria, if

$$MSEM(\hat{\theta}_1) - MSEM(\hat{\theta}_2) < 0, \; MSE(\hat{\theta}_1) - MSE(\hat{\theta}_2) < 0 \tag{5}$$

Now we give the main results of this paper.

Theorem 2.1. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be two unbiased estimators of the unknown p-dimensional vector β, $\hat{\beta}_1 \sim N(\beta,\sigma^2 M)$, $\hat{\beta}_2 \sim N(\beta,\sigma^2 N)$, σ^2 is a known constant, M and N are two known positive definite matrix and $M < N$. Then we have estimator $\hat{\beta}_1$ is superior over estimator $\hat{\beta}_2$ in the PMC sense.

Proof. By the definition 2.1, we obtain the PMC of $\hat{\beta}_1$ relative to $\hat{\beta}_2$ as follows:

$$P_r(\hat{\beta}_1,\hat{\beta}_2,\beta) = P_r\{(\hat{\beta}_1-\beta)'(\hat{\beta}_1-\beta) \leq (\hat{\beta}_2-\beta)'(\hat{\beta}_2-\beta)\} \tag{6}$$

Define $\zeta = (\sigma^2 M)^{-1/2}(\hat{\beta}_1-\beta)$ and $\eta = (\sigma^2 N)^{-1/2}(\hat{\beta}_2-\beta)$, then we obtain $\zeta \sim N(0,I)$ and $\eta \sim N(0,I)$. Now we can simply Equation (6) as follows:

$$\begin{aligned}
P_r(\hat{\beta}_1,\hat{\beta}_2,\beta) &= P_r\{\zeta'M\zeta \leq \eta'N\eta\} \\
&= P_r\{\zeta'M\zeta - \eta'M\eta - \eta'(N-M)\eta \leq 0\} \\
&\geq P_r\{\zeta'M\zeta - \eta'M\eta \leq 0\}
\end{aligned} \tag{7}$$

Since M is a positive definite matrix, there exists an orthogonal matrix H such that

$$M = H'\Lambda H, \; \Lambda = diag(\lambda_1,\lambda_2,...,\lambda_p) \tag{8}$$

where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ is the eigenvalues of M.

Then define $\tilde{\zeta} = H\zeta$, $\tilde{\eta} = H\eta$, we have $\tilde{\zeta} \sim N(0,I)$ and $\tilde{\eta} \sim N(0,I)$. Thus, we obtain

$$P_r(\hat{\beta}_1, \hat{\beta}_2, \beta) \geq P_r\{\tilde{\zeta}'\Lambda\tilde{\zeta} - \tilde{\eta}'\Lambda\tilde{\eta} \leq 0\} \tag{9}$$

Now we calculate the probability of

$$P_r\{\tilde{\zeta}'\Lambda\tilde{\zeta} - \tilde{\eta}'\Lambda\tilde{\eta} \leq 0\} \tag{10}$$

Firstly, denote $\tilde{\zeta} = \delta_2$ and $\tilde{\eta} = -\delta_1$, then by $\tilde{\zeta} \sim N(0,I)$ and $\tilde{\eta} \sim N(0,I)$, we obtain $\delta_1 \sim N(0,I)$ and $\delta_2 \sim N(0,I)$. Then Equation (10) is equal to

$$P_r\{\tilde{\zeta}'\Lambda\tilde{\zeta} - \tilde{\eta}'\Lambda\tilde{\eta} \leq 0\} = P_r\{\delta_2'\Lambda\delta_2 - \delta_1'\Lambda\delta_1 \leq 0\} \tag{11}$$

Since $\delta_1 \sim N(0,I)$ and $\delta_2 \sim N(0,I)$, we may denote $\tilde{\zeta} = \delta_1$ and $\tilde{\eta} = \delta_2$, then we have

$$P_r\{\tilde{\zeta}'\Lambda\tilde{\zeta} - \tilde{\eta}'\Lambda\tilde{\eta} \leq 0\} = P_r\{\delta_1'\Lambda\delta_1 - \delta_2'\Lambda\delta_2 \leq 0\} \tag{12}$$

Thus By Equations (10)-(12), we have

$$\begin{aligned} P_r\{\delta_2'\Lambda\delta_2 - \delta_1'\Lambda\delta_1 \leq 0\} &= P_r\{\delta_1'\Lambda\delta_1 - \delta_2'\Lambda\delta_2 \leq 0\} \\ &= 1 - P_r\{\delta_1'\Lambda\delta_1 - \delta_2'\Lambda\delta_2 \geq 0\} \end{aligned} \tag{13}$$

So we have

$$P_r\{\tilde{\zeta}'\Lambda\tilde{\zeta} - \tilde{\eta}'\Lambda\tilde{\eta} \leq 0\} = P_r\{\delta_2'\Lambda\delta_2 - \delta_1'\Lambda\delta_1 \leq 0\} = \frac{1}{2} \tag{14}$$

Thus, we obtain

$$P_r(\hat{\beta}_1, \hat{\beta}_2, \beta) \geq \frac{1}{2} \tag{15}$$

That is to say estimator $\hat{\beta}_1$ is superior over estimator $\hat{\beta}_2$ in the PMC sense.

Remark 2.1. The conditions in Theorem 2.1. satisfied, if we added $M = N$, then we can say that the estimator $\hat{\beta}_1$ is equivalence to estimator $\hat{\beta}_2$ in the PMC sense, that is

$$P_r\{(\hat{\beta}_1 - \beta)'(\hat{\beta}_1 - \beta) \leq (\hat{\beta}_2 - \beta)'(\hat{\beta}_2 - \beta)\} = \frac{1}{2}$$

and

$$P_r\{(\hat{\beta}_2 - \beta)'(\hat{\beta}_2 - \beta) \leq (\hat{\beta}_1 - \beta)'(\hat{\beta}_1 - \beta)\} = \frac{1}{2}$$

Remark 2.2. The conditions given in Theorem 2.1 is a sufficient condition of the estimator $\hat{\beta}_1$ is superior over $\hat{\beta}_2$ in the PMC sense.

Remark 2.3. When the conditions in Theorem 2.1. is satisfied, then the estimator is $\hat{\beta}_1$ is superior over the estimator $\hat{\beta}_2$ in the MSEM and MSE sense.

Using this Theorem we will easy proof the following theorem.

Theorem 2.2. Suppose that the random vector $\hat{\beta}$ has a multivariate normal distribution with mean β and covariance. Then $a'\hat{\beta}$ is the best PMC linear unbiased estimator of β if and only if $a'\hat{\beta}$ is the best linear unbiased estimator of β.

Remark 2.4. The proof of Theorem 2.2 can also obtained by Fountain and Keating [8]. This Theorem is shown in the linear estimators, the best linear unbiased estimator is a best estimator under the PMC sense. But this only limit in the linear estimators.

Theorem 2.3. Let T_1 be a $p \times 1$ statistic and T_2 be a $q \times 1$ statistics, the joint PDF is $EC_{p+q}(u, \Sigma, \Phi)$, where $u = (\beta', 0')'$, β is a $p \times 1$ vector, Σ is a known positive definite matrix and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then for all $\beta \in R^p$, we have $P_r(\beta^*, T_1, \beta) > 1/2$, that is the covariance improved estimator β^* is better than T_1 in the PC criterion, where $\beta^* = T_1 - \Sigma_{12}\Sigma_{22}^{-1}T_2$ is the covariance improved estimator of β.

Remark 2.5. The proof of this Theorem can also obtained in Wang and Yang [11].

Theorem 2.4. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be two unbiased estimators of the unknown p-dimensional vector β, $\hat{\beta}_1 \sim N(\beta, \sigma^2 M)$, $\hat{\beta}_2 \sim N(\beta, \sigma^2 N)$, σ^2 is a known constant, M and N are two known positive definite matrix. If $\hat{\beta}_1$ is superior over the estimator $\hat{\beta}_2$ in the MSEM sense, then we have estimator $\hat{\beta}_1$ is superior over estimator $\hat{\beta}_2$ in the PMC sense.

Proof. Since
$$MSEM(\hat{\beta}_1) = E\{(\hat{\beta}_1 - \beta)(\hat{\beta}_1 - \beta)'\} = Cov(\hat{\beta}_1) \tag{16}$$
similarly, we obtain
$$MSEM(\hat{\beta}_2) = E\{(\hat{\beta}_2 - \beta)(\hat{\beta}_2 - \beta)'\} = Cov(\hat{\beta}_2) \tag{17}$$
Since $\hat{\beta}_1$ is superior over the estimator $\hat{\beta}_2$ in the MSEM sense, we have
$$MSEM(\hat{\beta}_1) - MSEM(\hat{\beta}_2) < 0 \tag{18}$$
Then by Equations (16)-(18), we have
$$\sigma^2 M = Cov(\hat{\beta}_1) < Cov(\hat{\beta}_2) = \sigma^2 N \Rightarrow M < N \tag{19}$$

Thus using Theorem 2.1., we proof this theorem.
Remark 2.6. The inverse of this theorem is not true.

3 Applications in linear regression models

In this section, we will apply theorem 2.1 to compare the unbiased estimator in linear regression models.

1. General linear regression model

Consider the general linear regression model
$$Y = X\beta + \varepsilon \tag{20}$$
where Y is an $n \times 1$ vector of observation, X is an $n \times p$ known matrix of rank p, β is a $p \times 1$ vector of unknown parameters, ε is an $n \times 1$ vector of disturbances with expectation $E(\varepsilon) = 0$ and variance-covariance matrix $Cov(\varepsilon) = \sigma^2 V$, $V > 0$ is a known positive definite matrix.

The ordinary least squares estimator (OLSE) is given as follows:
$$\hat{\beta}_{OLSE} = (X'X)^{-1}X'Y \tag{21}$$

By Wang and Show [17], they show that β^* is the best linear unbiased estimator of β, where β^* is given as follows:

$$\hat{\beta}^* = (X'V^{-1}X)^{-1}X'V^{-1}Y \tag{22}$$

Since β^* is the best linear unbiased estimator of β, then using Theorem 2.1, we can see that β^* is superior over OLSE ($\hat{\beta}_{OLSE}$) in the PMC sense.

Theorem 3.1. For the linear regression model (20), $P_r(\beta^*, \hat{\beta}_{OLSE}, \beta) \geq \frac{1}{2}$, that is for every β and σ^2, the best linear unbiased estimator (β^*) is superior to the OLSE ($\hat{\beta}_{OLSE}$) in the PMC sense.

2. Restricted linear regression model

In model (20) we consider $V = I$, and consider Equation (20) has following linear restrictions:

$$R\beta = r \tag{23}$$

where the matrix R is $m \times p$ and of full row rank $m < p$, r is an $m \times 1$ vector and both R and r are known.

The restricted least squares estimator (OLSE) is given as follows:

$$\hat{\beta}_{RLSE} = \hat{\beta}_{OLSE} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta}_{OLSE}) \tag{24}$$

It is easy to compute that

$$E(\hat{\beta}_{OLSE}) = \beta, \ Cov(\hat{\beta}_{OLSE}) = \sigma^2(X'X)^{-1} \tag{25}$$

$$E(\hat{\beta}_{RLSE}) = \beta \tag{26}$$

and

$$Cov(\hat{\beta}_{RLSE}) = \sigma^2(X'X)^{-1} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1} \tag{27}$$

respectively.

Obviously,

$$Cov(\hat{\beta}_{RLSE}) < Cov(\hat{\beta}_{OLSE}) \tag{28}$$

Thus we have the following theorem:

Theorem 3.2. For the linear regression model (20) and $V = I$, $P_r(\hat{\beta}_{RLSE}, \hat{\beta}_{OLSE}, \beta) \geq \frac{1}{2}$, the RLSE ($\hat{\beta}_{RLSE}$) is superior to the OLSE ($\hat{\beta}_{OLSE}$) in the PMC sense.

3. Linear regression model with prior information

For the linear regression model (20) let $V = I$ and β has following prior information

$$r = R\beta + e, \ e \sim (0, \sigma^2W) \tag{29}$$

where R is a $j \times p$ known matrix of rank j, e is a $j \times 1$ vector of disturbances with mean 0 and dispersion matrix σ^2W, W is supposed to be known and positive definite, the $j \times 1$ vector r can be interpreted as a random variable with expectation $E(r) = R\beta$. Therefore the restriction (11) does not hold exactly but in the mean, and we suppose r to be known, that is to be realized value of the random vector, so that all the expectations are conditional on r [13]. In the following discussions, we do not mention this separately. Furthermore, it is also supposed that the random vector ε is stochastically independent of e.

The mixed estimation (ME) proposed by Durbin [16], Theil and Goldberger [14] and Theil [15] is given as follows

$$\hat{\beta}_{ME} = (X'X + R'W^{-1}R)^{-1}(X'Y + R'W^{-1}r) \tag{30}$$

It is easy to compute that

$$E(\hat{\beta}_{ME}) = \beta, \ Cov(\hat{\beta}_{ME}) = \sigma^2(X'X + R'W^{-1}R)^{-1} \tag{31}$$

By [13], we have

$$
\begin{aligned}
& Cov(\hat{\beta}_{ME}) - Cov(\hat{\beta}_{OLSE}) \\
= \ & -\sigma^2(X'X)^{-1}R'[(X'X)^{-1} + RW^{-1}R']^{-1}R(X'X)^{-1} < 0
\end{aligned} \tag{32}
$$

Then we have the following theorem

Theorem 3.3. For the linear regression model (20) and $V = I$, $P_r(\hat{\beta}_{ME}, \hat{\beta}_{OLSE}, \beta) \geq \frac{1}{2}$, the ME ($\hat{\beta}_{ME}$) is superior to the OLSE ($\hat{\beta}_{OLSE}$) in the PMC sense.

References

[1] C. R. Rao, Some comments on the minimum mean square error as a criterion of estimation, Statistics and related topics, North Holland Amsterdam (1981) 123-143.

[2] E. J. G. Pitman, The closest estimates of statistical parameters, Proc Cambridge Philosop Soc 33 (1937) 212-222.

[3] J. P. Keating, R. L. Mason, Practical relevance of an alternative criterion in estimation, Am Stat 39 (1985) 203-205.

[4] Ghosh, M. and Sen, P.K. (1989). Median unbiasedness and Pitman closeness. J. Amer. Statist. Assoc., 84, 1089-1091.

[5] Nayak, T.K. (1990). Estimation of location and scale parameters using generalized Pitman nearness criterion. J. Statist. Planning and Inference., 24, 259-268.

[6] Mason, R. L., Keating, J. P., Sen, P. K., Blaylock, N. W. (1990). Comparison of linear estimators using pitman measure of closeness. J Amer Statist Assoc 85: 579-581.

[7] Peddada, S. D., Khattree, R. (1991). Comparison of estimators of the location parameter using Pitman s closeness criterion. Commun Statist Theor Meth 20(11): 3525-3534.

[8] Fountain, R. L., Keating, J. P., (1994). The pitman comparison of unbiased linear estimators. Statist Probab Lett 19 : 131-136.

[9] Fountain, R. L. (1991). Pitman closeness comparison of linear estimators: A canonical form. Commun Statist Theor Meth 20(11): 3535-3550.

[10] Wang, S. G., Yang, H. (1994). Linear estimators umder pitman nearness criterion. Chinese Science Bulletin 40 (23): 1944-1948.

[11] Wang, S. G., Yang, Z. H. (1995). Pitman optimality of covariance-improved estimators. Chinese Science Bulletin 40 (1): 12-15.

[12] Yan, L. Q. (2007). A Note on Comparison of (Non) Linear Estimators Using Pitman's Measure of Closeness. Commun Statist Theor Meth 36: 2687-2694.

[13] Rao, C. R., Toutenburg, H., 1995. Linear models: Least squares and Alternative. Springer-Verlag, New York.

[14] Theil, H., Goldberger, A. S., 1961. On pure and mixed statistical extimation in economics. Int. Econ. Rev 2, 65-78.

[15] Theil, H., 1963. On the use of incomplete prior information in regression analysis. J. Am. Stat. Assoc. 58, 401-414.

[16] Durbin, J., 1951. A note on regression when there is extraneous information about one of the coefficients. J. Am. Stat. Assoc. 48, 799-808.

[17] Wang, S.G., Show, S.C., Advanced Linear models. New York, Marcell Dekker Inc. 1993.

Representative Points of Student's t_n Distribution and The Applications In Statistical Simulation

Min Zhou[1,3,*]   Wenjun Wang[4,5,†]   Kaitai Fang[1,2,‡]

[1]Division of Science and Technology, BNU-HKBU United International College, Zhuhai, 519085, China
[2]Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China
[3]Department of Mathematics, Hong Kong Baptist University, Hongkong, 999077, China
[4]Department of Mathematics, Hainan Normal University, Haikou, 510000, China
[5]Department of Mathematics, Yunnan University, Kuiming, 650000, China

Abstract   The paper considers the representative points (RP) of student's t_n distribution for minimizing the mean of square error. Fang and He's algorithm is applied to find its RP. We study the convergence of this algorithm, the existence and uniqueness of RP of t_n distribution when $n \geq 3$ and discuss applications of RP in statistical simulation. Traditional Monte Carlo, Bootstrap and Resampling are the basic methods in statistical simulation based on a random sample. Fang *et al.*[7] firstly proposed to use RP instead of i.i.d. random samples to construct an approximate distribution and take the sample from the approximation for statistical inference. This paper continues to study this issue. The statistical inference of this paper focuses on two cases: classical estimation of parameters(mean, variance, skewness and kurtosis) and robust estimation of parameters (location parameter: mean and median; scalar parameter: MAD and IQR). Our results indicate that the new method once again can significantly improve the accuracy of the estimator of the statistics, and accelerates the converging speed of the statistics.

Keywords   Quasi-Monte Carlo methds; Representative points; Statistical simulation; t distribution; Resampling, Robust estimation

1   Introduction

One interesting problem, that is how to select a given number of representative points (RP) to retain as much information of the population as possible, arises in many fields. For example, Cox[2] was the first one to study this problem and he suggested a loss function, given in (2), to measure the distortion. Fang and He[5] mentioned a project: in order to make clothes standardization in China, they took p measurements of the body of each of n individuals and projected these p dimensional data onto a q-dimensional region ($q = 1, 2, 3$) by some methods in statistical multivariate analysis. They wanted to select m points that best represent the data in the q-dimension region. When $q = 1$ the above problem is to choose m points that give as much as possible information about a normal distribution. The solution of points are called representation points. A similar project was given by Flury[10]. This project concerned with the Swiss army to replace existing with newly designed protection masks. To put the construction of new types of masks on solid

empirical grounds, a group of anthropologists was hired to measure the heads of about 900 Swiss soldiers. 25 variables were taken, which were supposed to be of importance for the fit of masks. This project needs the concept of representative points of a given distribution.

Suppose that a continuous random variable X has a distribution function $F(x)$ that has a density function $f(x)$ and finite second moment $\sigma^2 = Var(X)$. In many practical applications, for example, signal transaction[14], we hope to use a discrete random variable Y to approximate X. Let Y have a support set of points $S(Y) = \{x_1 < x_2 < \cdots < x_m\}$ with a probability mass function $P(Y = x_j) = p_j, \ j = 1, \cdots, m$. We choose Y such that X and Y are closest in a certain sense. In the literature, the following MSE criterion has been widely used. Let

$$Q_m(x) = x_j, \ \ a_j < x \le a_{j+1}, \ j = 1, \cdots, m, \tag{1}$$

where $a_1 = -\infty, \ a_{m+1} = \infty, \ a_j = (x_j + x_{j+1})/2, \ j = 2, \cdots, m$. The loss function, which measures the loss when we use $Y_{MSE} = Q_m(X)$ to approximate X, is defined by

$$
\begin{aligned}
L_m &= L(x_1, \ x_2, \cdots, \ x_m) = \frac{1}{\sigma^2} E(X - Q_m(X))^2 \\
&= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \min_{1 \le i \le m} (x - x_i)^2 f(x) dx = \frac{1}{\sigma^2} \sum_{j=1}^{m} \int_{a_j}^{a_{j+1}} (x - x_j)^2 f(x) dx,
\end{aligned}
\tag{2}
$$

where $P(Y_{MSE} = x_j) = P(Q_m(X) = x_j) = p_j$, and

$$
\begin{aligned}
p_1 &= \int_{-\infty}^{(x_1+x_2)/2} f(x) dx = \int_{-\infty}^{a_2} f(x) dx, \\
p_i &= \int_{(x_{i-1}+x_i)/2}^{(x_i+x_{i+1})/2} f(x) dx = \int_{a_i}^{a_{i+1}} f(x) dx, \ i = 2, \ \cdots, \ m-1, \\
p_m &= \int_{(x_{m-1}+x_m)/2}^{\infty} f(x) dx = \int_{a_m}^{\infty} f(x) dx.
\end{aligned}
\tag{3}
$$

If the support set of Y enables the loss function to arrive at its minimum, this support set is called as representative point under MSE, denoted by RP-MSE.

RP-MSE was firstly proposed by Cox[2], and then Max[14] and Bofinger[1]. Fang and He[5] demonstrated independently and respectively the need of RP-MSE from a different statistical background. They proposed an algorithm for searching RP-MSE of the standard normal distribution and gave a comprehensive study on the proposed algorithm. The RP-MSE is also called Quantizer in the signal transaction (Max[14]). Flury[10] called RP-MSE as Principal Points and gave some results on elliptically contoured distributions. Fu [11][12] considered RP of the Gamma distribution and the Weibull distribution, and Fei[8][9] discussed more general cases: the distribution family of Pearson, and gave sets of RP-MSE for several important distributions (Negative exponential distribution, χ^2-distribution and F distribution) and some properties of RP-MSE. Yamamoto and Shinozaki [18] discussed RP-MSE (only for the case of $m = 2$) for the t distribution. Fang and Wang [6] introduced the definition and applications of RP in detail in Chapter 4 of the book. Fang *etal*.[7] recalculated the RP of standard normal distribution and further studied its application in the statistical simulation. However, so far there is no paper to give a systematical discussion on RP-MSE of the t distribution. It is well known that the t distribution plays an important role in Statistics. In the remaining sections, we will discuss RP of the t distribution and applications in statistical simulation.

Ross[17] gives some background knowledge on stochastic simulation. Bootstrap method proposed by Efron[3] is a pioneer of resampling technique widely used in statistical simulation. See Efron and Tibshiranier [4] for the details. The resampling techniques also use a discrete random variable Y to approximate a given continuous random variable X, and resampling is from the population Y. Denote the support set of Y by $S(Y) = \{x_1 < x_2 < \cdots < x_m\}$, and the probability mass function of Y is given by $P(Y = x_j) = p_j$, $j = 1, \cdots, m$. In this paper we concern with three different kinds of methods to construct Y, which were indicated by Fang *et al.*[7].

(A) Monte Carlo methods: Generate a random samples, $\{x_1, \cdots, x_m\}$, from the population $F(x)$ by the Monte Carlo method, here $x_1, \cdots, x_m$ are an i.i.d. and $x_i \sim F(x)$. Let a random variable Y_{MC} be uniformly distributed on the sample, i.e., $P(Y_{MC} = x_j) = 1/m$, $j = 1, \cdots, m$. This sample can be regarded as a set of RP, denoted by RP-MC.

(B) Quasi-Monte Carlo methods: the QMC method has been successfully used in high-dimension integration (Niederreiter[15]) and experimental design (Fang and Wang [6]). It is known that the set of $\{\frac{2j-1}{2m}, j = 1, \cdots, m\}$ is uniformly distributed on (0,1). By the inverse transformation method the set $\{x_j = F^{-1}(\frac{2j-1}{2m}), j = 1, \cdots, m\}$, denoted by RP-QMC, can be regarded as a set of representative points for the population, where F^{-1} is the inverse function of F. Let Y_{QMC} be uniformly distributed on $x_j, j = 1, \cdots, m$.

(C) RP-MSE: it has been introduced before.

Fang *et al.*[7]'s innovation is to address the issue of randomness about RP. RP generated by (B) and (C), are not random, but fixed. Whether can RP-QMC and RP-MSE be applied into stochastic simulation? This question cannot be solved perfectly in the long run.Fang *et al.*[7] discussed the case of standard normal distribution and pointed out that RP-QMC or RP-MSE can perform better. A natural question is whether we can obtain same conclusion for other distribution, especially for another important distribution-t_n distribution.

Our present discussion is organized as follows: Section 2 deals with the existence and uniqueness of RP-MSE for t distribution. The third section concerns with estimation of the population mean, variance, skewness and kurtosis of the t distribution by the use of the mean, variance, skewness and kurtosis of Y. Then we employ the resampling technique to estimate the above population parameters and compare their performance among $Y = Y_{MC}, Y_{QMC}$ and Y_{MSE}, in resampling; and we further discuss their performance in the robust estimation. The last section gives conclusion and proposes some further studies.

2   MSE Representative Points of Student's t-distribution

In this section we discuss how to find the representative points (RP-MSE) for the t distribution with n degrees of freedom, i.e. $X \sim t_n$. The probability density function of X is

$$t(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \equiv c(n)(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}. \tag{4}$$

Let $T(x)$ be the cumulative distribution function of t_n. It is well known that the mean, variance, skewness and kurtosis of X are

$$E(X) = 0, Var(X) = \frac{n}{n-2}, Sk(X) = 0, \text{ and } Ku(X) = \frac{6}{n-4}, \text{ respectively.} \tag{5}$$

Consequently, necessary conditions for existence of variance, skewness and kurtosis of X are $n > 2$, $n > 3$ and $n > 4$, respectively. In the following we always assume $n > 2$ as the definition of RP-MSE needs existence of the variance.

Li and Flury [13] proposed a sufficient condition, $f * f'' - 2(f')^2 < 0$, for the uniqueness of RP (they called as principal points) about any symmetric distribution with a density function $f(x)$. For t-distribution, it is easy to find that $t * t'' - 2(t')^2 = -n\frac{(n+1)(x^2+1)}{(n+x^2)^2}t(x)^2 < 0$, which implies that the t-distribution has a unique set of RP with degrees of freedom $n > 2$. Furthermore, as the density function (4) is an even function, it must have $x_j = -x_{m-j+1}, 1 \leq j \leq l+1$. The loss function can be denoted as $L(x_1, \cdots, x_l)$. When $n = 2l + 1$ is odd, we need to find $x_0 = 0 < x_1 < \cdots < x_l$; when $n = 2l$ is even we need to find $0 < x_1 < x_2 < \cdots < x_l$. Let $\partial L(x_1, \cdots, x_l)/\partial x_i = 0, i = 1, \cdots, x_l$ and we obtain two kinds of system of equations as follows: When $m = 2l$ is even, the system of equations are

$$\begin{cases} G(0) - G(\frac{x_1+x_2}{2}) = x_1[T(\frac{x_1+x_2}{2}) - T(0)] \\ G(\frac{x_{k-1}+x_k}{2}) - G(\frac{x_k+x_{k+1}}{2}) = x_k[T(\frac{x_k+x_{k+1}}{2}) - T(\frac{x_{k-1}+x_k}{2})], \ 2 \leq k \leq l-1 \\ G(\frac{x_{l-1}+x_l}{2}) = x_l[1 - T(\frac{x_{l-1}+x_l}{2})]. \end{cases} \qquad (6)$$

When $m = 2l + 1$ is odd, the system of equations are

$$\begin{cases} G(\frac{x_1}{2}) - G(\frac{x_1+x_2}{2}) = x_1[T(\frac{x_1+x_2}{2}) - T(\frac{x_1}{2})] \\ G(\frac{x_{k-1}+x_k}{2}) - G(\frac{x_k+x_{k+1}}{2}) = x_k[T(\frac{x_k+x_{k+1}}{2}) - T(\frac{x_{k-1}+x_k}{2})], \ 2 \leq k \leq l-1 \\ G(\frac{x_{l-1}+x_l}{2}) = x_l[1 - T(\frac{x_{l-1}+x_l}{2})], \end{cases} \qquad (7)$$

where

$$G(x) = \frac{c(n)n}{n-1}(1 + \frac{x^2}{n})^{-\frac{n-1}{2}} = \frac{nt(x)}{n-1}(1 + \frac{x^2}{n}). \qquad (8)$$

We adopt the ideas from Fang and He[5], and use a numerical method to solve the system of equations (6) and (7), respectively. Firstly, we choose an appropriate initial value $x_1 > 0$, and find the solution x_2 from the first equation; then for these two values x_1 and x_2 we obtain x_3 from the second equation; similarly, based on x_2 and x_3 we get x_4 from the third equation; finally, we obtain x_l from the penultimate equation. On the other hand, we can obtain another solution x_l^* by the last equation when x_{l-1} is given. If the difference between x_l and x_l^* is very small, x_1, x_2, $\cdots$, x_l are the solutions; otherwise, we modify the initial value x_1 and repeat the above processes. Fang and He[5] have proved the convergence of this method for the normal population. We now testify existence and uniqueness about the solutions of system of equations (6) and (7).

Being similar to the standard normal case, the system of equations (6) and (7) have only four types of the equation. For given m, denote by $x_{m1} < x_{m2} < \cdots < x_{ml}$ and $x_{m0}(=0)$, $x_{m1} < x_{m2} < \cdots < x_{ml}$ the solutions of the system of equations (6) and (7), respectively. When $m = 2(l = 1)$, the system of equations (6) only has one equation, i.e., $G(0) = x_1(1 - \frac{1}{2}) = \frac{1}{2}x_1$, we have $x_{21} = 2G(0) = \frac{2n}{n-1}c(n)$. For instance, when $n = 10$, $x_{21} = 0.8646852977$. The following theorems point out the existence and uniqueness about the solutions of system of equations (6) and (7).

Theorem 1. *For any given $x_1 \geq 0$, the first type of equation*

$$G(0) - G(\frac{x_1 + x_2}{2}) = x_1[T(\frac{x_1 + x_2}{2}) - T(0)] \qquad (9)$$

has a unique solution $x_2 \equiv g_2(x_1)$ if and only if $x_1 < x_{21}$. The function $g_2(x_1)$ is strictly increasing, when $x_1 < x_{21}$.

Theorem 2. *When $n > 2$, for any given $x_1 \geq 0$, the second type of equation*

$$G(\frac{x_1 + x_2}{2}) = x_2[1 - T(\frac{x_1 + x_2}{2})] \qquad (10)$$

has a unique solution $x_2 \equiv h_1(x_1)$, and $h_1(x_1)$ is strictly increasing.

Theorem 3. *For any given $x_1 \geq 0$, the third type of equation*

$$G(\frac{x_1}{2}) - G(\frac{x_1 + x_2}{2}) = x_1[T(\frac{x_1 + x_2}{2}) - T(\frac{x_1}{2})] \tag{11}$$

has a unique solution $x_2 \equiv h_2(x_1)$ if and only if $x_1 < x_{31}$. Furthermore, $h_2(x_1)$ is strictly increasing with respect to x_1 if $x_1 < x_{31}$.

Theorem 4. *For any fixed $0 < x_1 < x_{21}$, we can obtain the solution $x_2 = g_2(x_1)$ from Theorem 1, then the fourth type of equation*

$$G(\frac{x_1 + x_2}{2}) - G(\frac{x_2 + x_3}{2}) = x_2[T(\frac{x_2 + x_3}{2}) - T(\frac{x_1 + x_2}{2})] \tag{12}$$

has a unique solution $x_3 = g_3(x_1)$ if and only if $0 < x_1 < x_{41}$.

According to the above four theorems, for given x_1, we can obtain in turn $x_2 = g_2(x_1)$, $x_3 = g_3(x_1)$, ..., $x_{k+1} = g_{k+1}(x_1)$, $k = 2, 3, \ldots, l-1$ from the 1st, 2nd, ..., k^{th} equation of (6), then the equation

$$G(\frac{x_{k-1} + x_k}{2}) - G(\frac{x_k + x_{k+1}}{2}) = x_k[T(\frac{x_k + x_{k+1}}{2}) - T(\frac{x_{k-1} + x_k}{2})] \tag{13}$$

has a unique solution $x_{k+1} = g_{k+1}(x_1)$ if and only if $x_1 < x_{2k,1}$, $k = 2, 3, \ldots, l-1$. A similar conclusion is true for the system of equations (7).

Based on the above discussion, we can numerically search the value of $g_k(x_1)$, $k = 2, 3, \ldots, l$ by the use of the Newton method or the method of bisection. After getting $x_{l-1} = g_{l-1}(x_1)$, we can use the same methods to determine $x_l = g_l(x_1)$ and $x_l^* = h(x_{l-1})$ on the basis of the last two equations of (6), respectively. Figure 1 shows the figure of $g_k(x_1)$, $k = 2, 3, 4, 5, 6$ for the case of $m = 12$ ($l = 6$), which indicates that $x_k = g_k(x_1)$ is a strictly increasing function of x_1.

Figure 2 expresses the figures about (x_{l-1}, x_l) and (x_{l-1}, x_l^*) for the case of $m = 2l$, $l = 3, 4, 10, 13, 17$. According to Figure 2, we have the following conclusions: (1) system of equations (6) has a unique solution; (2) for any given $x_1 > 0$, when x_l is significantly less than x_l^*, we ought to increase the initial value x_1 appropriately because x_{l-1} is a strictly increasing function of x_1^*. Similarly, when x_l is larger than x_l^*, we should decrease the initial value x_1 appropriately. Obviously, a similar conclusion can be given for system of equations (7).

3　Stochastic Simulation and Resampling

Statistical simulation[17] has played an important role in statistical research and applications. Monte Carlo methods and resampling technique are the basic methods in statistical simulation. Let $X \sim F(x)$, where $F(x)$ is the population distribution function and let Y be an approximate population to X with a probability mass distribution $P(Y = x_j) = p_j, j = 1, \cdots, m$. We choose the L_2-distance between $F_Y(x)$ and $F(x)$ as a suitable criterion (refers to [6], section 1.4.2) to measure their closeness, which was proposed by Fang *et al.*[7]. The L_2-distance is defined by

$$D_2(F, \ F_Y) = \left[\int\limits_{-\infty}^{+\infty} |F_Y(x) - F(x)|^2 dx\right]^{1/2}. \tag{14}$$
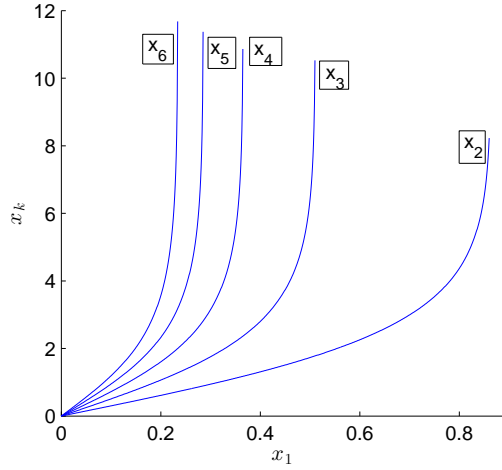
Figure 1: $x_k = g_k(x_1)$
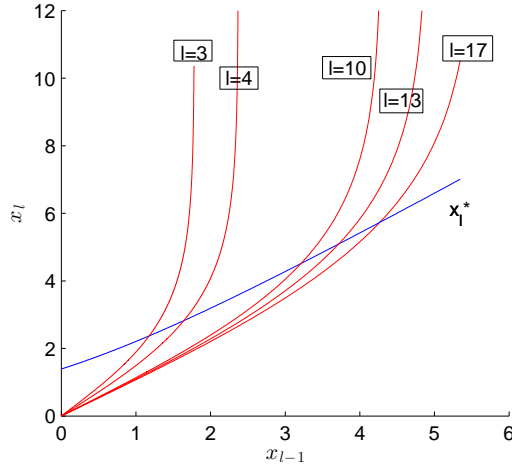
Figure 2: (x_{l-1}, x_l) and (x_{l-1}, x_l^*)

3.1 Preliminary Comparisons among three kinds of representative points

As we employ three discrete variables Y_{MC}, Y_{QMC} and Y_{MSE} to be approximate populations, their statistical properties should be close to the population ones. In particular their mean, variance, skewness and kurtosis should be close to the population ones. We choose the population $X \sim t_n, n = 10$. It is known that mean $E(X) = 0$, variance $Var(X) = \sigma^2 = \frac{n}{n-2} = 1.25$, the coefficient of Skewness (Sk) $Sk(X) = 0$ and the coefficient of Kurtosis (Ku) $Ku(X) = \frac{6}{n-4} = 1$.

Let Y be a discrete distribution with $P(Y = b_i) = p_i, i = 1, \cdots, m$. Then the above

statistics of Y can be calculated by

$$E(Y) = \sum_{i=1}^{m} b_i p_i \equiv \mu_b, \ Var(Y) = \sum_{i=1}^{m} (b_i - \mu_b)^2 p_i \equiv \sigma_Y^2,$$

$$Sk(Y) = \frac{1}{\sigma_Y^3} \sum_{i=1}^{m} (b_i - \mu_b)^3 p_i, \ Ku(Y) = \frac{1}{\sigma_Y^4} \sum_{i=1}^{m} (b_i - \mu_b)^4 p_i - 3. \tag{15}$$

As RP-QMC and RP-MSE of t-distribution are symmetric about the origin, it is easy to see that $E(Y_{QMC}) = 0, E(Y_{MSE}) = 0, Sk(Y_{QMC}) = 0, Sk(Y_{MSE}) = 0$. Therefore, we only need to consider comparisons of variance and kurtosis between X and Y, where $Y = Y_{MC}, Y_{QMC}, Y_{MSE}$. Choosing $m = 5, 10, 20, 30, 31, 32, 33, 34, 35$, we use the variance and kurtosis of Y to estimate $Var(X) = 1.25$ and $Ku(X) = 1$.

Table 1: Bias of variance and kurtosis between the approximate distribution and t_{10}

m	Category	5	10	20	30	31	32	33	34	35
	RP-MC	-1.0129	-0.4716	-0.0294	-0.2880	0.2022	0.7861	-0.5098	-0.2070	-0.1695
Variance	RP-QMC	-0.3795	-0.2213	-0.1283	-0.0931	-0.0907	-0.0885	-0.0863	-0.0843	-0.0824
	RP-MSE	-0.1238	-0.0377	-0.0106	-0.0049	-0.0046	-0.0044	-0.0041	-0.0039	-0.0037
	RP-MC	-2.4832	-2.2750	-0.9661	-1.0062	0.2069	-0.9665	-0.9440	-1.4461	-1.7677
Kurtosis	RP-QMC	-2.0830	-1.6406	-1.2743	-1.0892	-1.0751	-1.0616	-1.0487	-1.0362	-1.0242
	RP-MSE	-1.2002	-0.5277	-0.1991	-0.1060	-0.1006	-0.0956	-0.0910	-0.0867	-0.0827

Table 1 shows the biases of variance and kurtosis among X and Y for nine sizes of the approximate populations, where $Y = Y_{MC}, Y_{QMC}, Y_{MSE}$. According to the results in Table 1, we may draw the following conclusions: (1) the estimator is more accurate if m increases, so we should choose a suitable m; (2) for RP-MC method, the testing results are rather poor and Y_{MC} is not a good choice for approximate population; (3) for estimation of the variance and kurtosis, RP-MSE has the prominent performance.

RP-MSE has a better performance than others. Table 2 presents the values of $D_2(F, \ F_{Y_{QMC}})$ and $D_2(F, \ F_{Y_{MSE}})$ (refer to (14)) based on the different m. Obviously, the distribution function of Y_{MSE} is closer to the population distribution function F. As Y_{MC} strongly depends on a random sample and most of $D_2(F, \ F_{Y_{MC}})$ are significantly larger than $D_2(F, \ F_{Y_{QMC}})$ and $D_2(F, \ F_{Y_{MSE}})$, we do not list $D_2(F, \ F_{Y_{MC}})$ in the table.

Table 2: Square L_2-distance between the approximation and population distribution

m	5	10	20	30	31	32	33	34	35
$D_2(F, \ F_{Y_{QMC}})$	0.173341	0.039896	0.009127	0.003836	0.003576	0.003341	0.003128	0.002934	0.002757
$D_2(F, \ F_{Y_{MSE}})$	0.041251	0.006392	0.001702	0.000789	0.000741	0.000697	0.000657	0.000620	0.000587

3.2 Resampling on Representative Points-Classical Estimation

In traditional stochastic simulation samples are taken from the population and are i.i.d.. Resampling techniques takes a sample, $y_1, \cdots, y_N$ from an approximate distribution, F_Y. The bootstrap employs a random sample to form an approximate population. However, if the samples are not good representative for the population, the bootstrap may provide a unrobust result. Therefore, we propose to use the approximate population constructed by RP-QMC and RP-MSE for resmapling, and the estimation accuracy for a given statistic should be improved.

We still consider the population distribution $X \sim t_{10}$ and estimate its mean, variance, kurtosis and skewness. In the simulation, we choose eight sizes of RP for construction of approximate populations, that are $m = 25, 28, 30, 31, 32, 33, 34, 35$ and resampling takes from related populations of Y_{MC}, Y_{QMC} and Y_{MSE}. In order to save some space, we only display the estimation biases when $m = 35$. Table3 lists respectively estimation biases of four statistics by resampling from Y_{MC}, Y_{QMC} and Y_{MSE} for $m = 35$ and $N = 1000, 2000, 5000, 10000$, where N is the resample size. From the results in Tables, we may address the following conclusions: (1) RP-MSE occupies absolute predominance for estimation of variance and kurtosis; (2) RP-QMC has a good performance for estimation of the mean and skewness. Summarize the comparisons including the results of $m = 25, 28, 30, 31, 32, 33$ that do not appear in the tables, into Table 4, where we can see that RP-MSE provides more accurate estimation for the statistics in most cases. Furthermore, the performance of RP-QMC and RP-MSE are consistently superior to that of RP-MC.

Table 3: $m = 35$, Classical estimation biases by resampling

	Category	1000	2000	5000	10000
	RP-MC	-0.178981	-0.182136	-0.181027	-0.181247
Mean	RP-QMC	0.000074	-0.005019	0.002974	0.000982
	RP-MSE	-0.005945	0.001409	-0.002814	0.000370
	RP-MC	-0.150406	-0.158475	-0.164127	-0.162118
Variance	RP-QMC	-0.083006	-0.087393	-0.083298	-0.079996
	RP-MSE	0.007502	-0.002900	0.000532	0.000039
	RP-MC	-1.265060	-1.251296	-1.260116	-1.275396
Kurtosis	RP-QMC	-1.126946	-1.094984	-1.106769	-1.106414
	RP-MSE	-0.596670	-0.701559	-0.614099	-0.655111
	RP-MC	0.405035	0.412940	0.405345	0.400693
Skewness	RP-QMC	-0.001272	0.013655	-0.001483	0.000420
	RP-MSE	0.014134	-0.000159	-0.012804	0.001138

Table 4: The number of winner in statistical estimation

	RP-MC	RP-QMC	RP-MSE
Mean	0	14	18
Variance	0	0	32
Kurtosis	0	0	32
Skewness	0	21	11
Total	0	35	93

3.3   Resampling on Representative Points-Robust Estimation

Classical estimation such as the sample mean, the sample variance or sample covariance, can be significantly influenced by outliers, even by a single one. It often does not provide good fits to lots of data. Robust statistics is able to address this kind of problems. The concepts and methods of robust statistics originated in the 1950s. The technical term "robust statistics" was coined by G. E. P. Box in 1953. Later, many statisticians such as Huber [16], did systematically abundant works about robust estimation of parameters. In the robust estimation, the estimations of location parameter and scalar parameter are always important issues. Here, we focus on these estimations by using RP.

Assume that the population distribution $X \sim t_6$, which is a heavy-tail distribution. We use the resmapling technique to obtain samples and estimate its location parameters(median and mean) and scalar parameters(median absolute deviation and interquartile range). Finally, the bias is calculated to be used for comparing the performance about three kinds of RP. The above four population parameters can be estimated by the following robust statistics respectively:

(1) Sample median(MD):

$$MD = median\{X_i, i = 1, 2, \cdots, m\},$$

where $\{X_i, i = 1, 2, \cdots, m\}$ is one sample based on resampling.

(2) $\alpha-$Trimmed mean(TM):

$$TM = \frac{1}{m - 2k} \sum_{k+1}^{m-k} X_{(i)},$$

where $k = [\frac{m\alpha}{2}]$ and $X_{(i)}$ is the order statistic. In the statistical simulation, $\alpha = 5\%, 25\%$.

(3) Sample median absolute deviation(MAD):

$$MAD = median\{|X_i - MD|, \ i = 1, 2, \cdots, m\},$$

where $\{X_i, i = 1, 2, \cdots, m\}$ is one sample based on resampling and MD is the corresponding sample median.

(4) Interquartile range: $IQR = Q_3 - Q_1$, where Q_1 and Q_3 are the first quartile and third one, respectively.

Table 5: $N = 1000$, Robust estimation biases by resampling

	m	10	20	30	31	32	33	34
	RP-MC	0.410752	0.385452	0.120457	0.124145	0.118420	-0.262859	-0.130328
MD	RP-QMC	-0.013042	-0.004435	-0.004146	-0.014642	0.011663	0.015406	-0.002288
	RP-MSE	-0.005829	0.008455	-0.009243	0.015802	0.001426	-0.011904	-0.003484
	RP-MC	0.610134	0.288904	0.201475	0.094749	0.181069	-0.201385	-0.04942
$TM(5\%)$	RP-QMC	-0.006269	0.004060	-0.005305	-0.011542	0.012802	0.013611	-0.006398
	RP-MSE	-0.002848	0.005912	-0.003555	0.008602	0.003393	-0.009126	0.001689
	RP-MC	0.556580	0.277243	0.199287	0.103936	0.202323	-0.239128	-0.023882
$TM(25\%)$	RP-QMC	-0.007931	0.002517	-0.005849	-0.014111	0.011961	0.012959	-0.005329
	RP-MSE	-0.005362	0.009324	-0.007436	0.008017	0.001862	-0.007631	0.001402
	RP-MC	0.422028	0.079398	-0.018518	-0.095632	0.086759	-0.001717	0.022720
MAD	RP-QMC	-0.051457	-0.020547	-0.013319	-0.016528	-0.010320	-0.012012	-0.007686
	RP-MSE	-0.055038	-0.033779	-0.016635	-0.006079	-0.008595	-0.021954	-0.015021
	RP-MC	1.225935	0.396350	0.057603	-0.186916	0.082307	0.054144	0.109492
IQR	RP-QMC	-0.017032	0.014189	0.005119	-0.005498	0.013381	0.005289	0.004970
	RP-MSE	-0.025983	-0.013482	-0.010439	0.010187	0.008229	-0.015862	-0.010028

Table 5 displays the estimation biases of four statistics by resampling from Y_{MC}, Y_{QMC} and Y_{MSE} for $m = 10$, 20, 30, 31, 32, 33, 34 and $N = 1000$, where N is the resample size. From the results in Table, we may draw the conclusions that the performance of RP-QMC and RP-MSE are superior to that of RP-MC, and the estimation biases of RP-QMC and RP-MSE are very close, whose difference are almost 0.001. Therefore, from robust estimation, we can see that RP-QMC and RP-MSE improve accuracy of estimation and speed the convergent rate up for the statistics in most cases.

4 Conclusion

The paper concerns with the representative point (RP-MSE) of t-distribution with n degrees of freedom, $n \geq 3$. Firstly, we employ Fang and He[5]'s algorithm and prove that this algorithm can also be applied to the t distribution perfectly. The paper proves the existence and uniqueness of RP-MSE when $n \geq 3$. In addition, we study the application of RP in statistical simulation by using the new method that was proposed by Fang $et\ al.$[7]. We take t_6 and t_{10} for example and consider classical estimation of the parameters (mean, variance, skewness and kurtosis) and robust estimation of the parameters(location and

scalar parameters). Our results once again show that the new method proposed by Fang *et al.*[7], can significantly improve accuracy in the estimation of statistics, or accelerates the converging speed of statistics. This method can be extended to other various population distributions, including multivariate distributions population and open a new research direction for statistical simulation.

Acknowledgement

References

[1] Bofinger, E. (1970). Maximizing the Correlation of Grouped Observations *J. American Statistical Association*, **65**, 1632-1638.

[2] Cox, D. R. (1957). Note on Grouping, *J. American Statistical Association*, **52**, 543-547.

[3] Efron, B. (1979), Bootstrap methods: another look at the jacknife, *The Annals of Statistics*, **7**, 1-26.

[4] Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*, New York, Chapman & Hall.

[5] Fang, K.T. and He, S.D. (1982), The problem of selecting a given number of representative points in a normal population and a generlized Mill's ration, Technical Report, No. 5, Department of Statistics, Stanford University, California, USA.

[6] Fang, K. T. and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, Chapman & Hall, London.

[7] Fang, K. T. Zhou, M and Wang, W. J.(2014). Applications of the representative points in statistical simulations.*Science China Mathematics(A).*,2014, **57 (12)**, 2609-2620.

[8] Fei, R.C. (1990), The problem of selecting the representative points in the family of *Pearson* distribution, *Journal of Wuxi Institute of Light Industry, China*, **9 (1)**, 71-78.

[9] Fei, R.C. (1991), The statistical relationship between the representative points and population distribution, *Journal of Wuxi Institute of Light Industry, China*, **10 (1)**, 123-134.

[10] Flury, B.A. (1990) Principal points, *Biometrika*, **77 (1)**, pp. 33-41.

[11] Fu, H.H. (1985), The problem of selecting a given number of representative points in a Γ distribution, *Jornual of China University of Mining and Technology* o.4,107-117.

[12] Fu, H.H. (1993), The problem of the optimal representative points about Weibull distribution, *Journal of Wuxi Institute of Light Industry, China*, **22 (3)**, 78-83.

[13] Li, L., Flury, B.D. (1995). Uniqueness of principal points for univariate distributions. Statist. Probab. Lett. 25, 323-327.

[14] Max, J. (1960). Quantizing for minimum distortion, *IRE Transactions on Information Thery*, **6 (1)**, 7-12.

[15] Niederreiter, H. (1992). *Random number generaion and quasi-Mente Carlo methods*, SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia.

[16] Peter, J. Huber and Elvezio, M.Ronchetti.(2009). Robust statistics (Second Edition). John Wiley & Sons, Inc., Hoboken, New Jersey.

[17] Ross, S.M. (1990). *A Course in Simulation*, Macmillan Publishing Company, New York.

[18] Yamamoto, W. and Shinozaki, N. (2000) On uniqueness of two principal points for univariate location mixtures. *Statist. Probab. Lett.*, **46**, 33-42.

SPECIAL MATRICES

SCOPE

Special Matrices publishes original articles of wide significance and originality in all areas of research involving structured matrices present in various branches of pure and applied mathematics and their noteworthy applications in physics, engineering, and other sciences.

JOURNAL EDITOR

HOW TO SUBMIT:

 http://www.editorialmanager.com/spma

JOURNAL'S HOMEPAGE:

www.degruyter.com/view/j/spma

EDITORIAL ADVISORY BOARD

TOPICAL ISSUE ON PROCEEDINGS OF THE 24TH INTERNATIONAL WORKSHOP ON MATRICES AND STATISTICS

DESCRIPTION

This special issue will contain a selection of papers based on the presentations of invited speakers at the **24th International Workshop on Matrices and Statistics** held at Hainan Normal University, Haikou, China over 25-28 May 2015.

The topics of the special issue include but are not limited to:
▶ Model Selection and Post Estimation
▶ Statistical Simulation
▶ Magic Matrices
▶ Matrices in Applied Probability
▶ Statistical Modeling and Computation
▶ Matrices with Economic and Financial
▶ Design and Analysis of Experiments
▶ Linear and Mixed Models
▶ Matrices Useful for Modelling Multi-level Models
▶ Teaching Matrices within Statistics
▶ G-Inverses, Linear Models and Multivariate Analysis

GUEST EDITORS:

Prof. Jeffrey Hunter
Auckland University of Technology, New Zealand
Dr Simo Puntanen
University of Tampere, Finland
Prof. Dietrich Von Rosen
Swedish University of Agricultural Sciences, Sweden

THE SCIENTIFIC PROGRAM COMMITTEE:

Jeffrey J. Hunter (New Zealand) [Chair], **Simo Puntanen** (Finland) [Vice-Chair], **Chuanzhong Chen** (China), **S. Ejaz Ahmed** (Canada), **Kai-Tai Fang** (China), **Shuangzhe Liu** (Australia), **Yonghui Liu** (China), **Tsung-I Lin** (Taiwan), **Augustyn Markiewicz** (Poland), **Jianxin Pan** (United Kingdom), **George P. H. Styan** (Canada), **Kimmo Vehkalahti** (Finland), **Julia Volaufova** (United States of America), **Dietrich von Rosen** (Sweden), **Hans Joachim** Werner (Germany)

HOW TO SUBMIT:

Before submission authors should carefully read the Instructions for Authors, which are located at **http://www.degruyter.com/view/j/spma**.

All submissions to the Topical Issue must be made electronically at **http://www.editorialmanager.com/spma**

and will undergo the standard single-blind peer review system. When entering your submission choose the option *Topical Issue on "Proceedings of the 24th International Workshop on Matrices and Statistics"*.

Submissions for the special issue are now open.
The deadline for the submissions is August 1, 2015.

Contributors to the Topical Issue will benefit from:
▶ fair and constructive peer review provided by recognized experts in the field,
▶ Open Access to your article for all interested readers,
▶ fast online publication of articles,
▶ no publication fees,
▶ convenient, web-based paper submission and tracking system – Editorial Manager,
▶ free language assistance for authors from non-English speaking regions;

We are looking forward to your submission. If you have any question, please contact us at **spma.editorial@degruyteropen.com**