

# 2017

## Hangzhou International Statistical Symposium

# Program

**Organizer:**

The Chinese Association for Statistical Computing

**Co-organizers:**

Center for Statistical Science, Tsinghua University

School of Mathematical Sciences, Zhejiang University

November 1-3, 2017

Hangzhou, China

# Contents

Conference Quick Guide.....	1
Local Information & Contact.....	2
Organizing Committee.....	6
Schedule for November 2, 2017 (Thursday) .....	7
Schedule for November 3, 2017 (Friday) .....	9
Towards the Next ICOTS10 in Kyoto .....	10
Abstracts of Invited Talks.....	11

# Conference Quick Guide

## November 1, 2017 (Thursday)

- **Morning:** Arrive in Hangzhou
- **Noon:** Arrive in Hangzhou
- **Afternoon:** Register at the lobby of Hangzhou Huagang HNA Resort
- **Evening:** Dinner at Tian Xiang Yuan (天香苑), 3<sup>rd</sup> floor.

## November 2, 2017 (Friday)

- **Morning:** Opening ceremony & Technical sessions
- **Noon:** Lunch buffet at Western Restaurant (恺撒宫), 1<sup>st</sup> floor.
- **Afternoon:** Technical sessions
- **Evening:** Welcome banquet at Tian Xiang Yuan (天香苑), 3<sup>rd</sup> floor

## November 3, 2017 (Saturday)

- **Morning:** Technical sessions & Ending mark
- **Noon:** Lunch buffet at Western Restaurant (恺撒宫), 1<sup>st</sup> floor.
- **Evening:** Dinner outside, 6 : 00pm waiting at hotel gate.

# Local Information & Contact

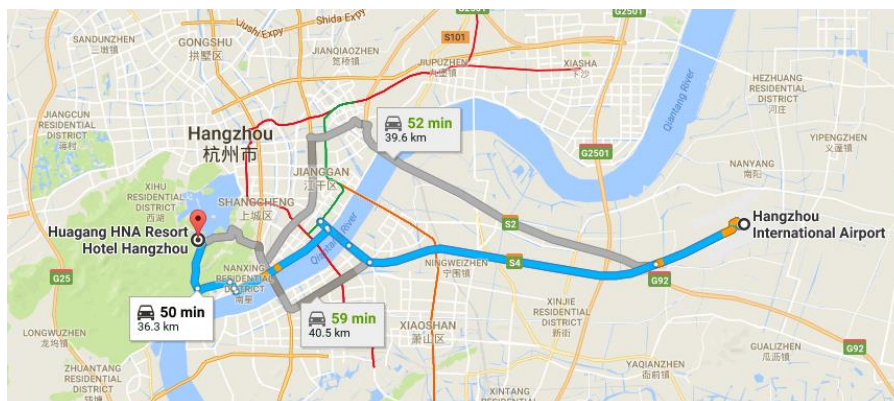
## About the Conference Hotel

- **Hotel name:** Hangzhou Huagang HNA Resort
- **Address:** 1 Yanggongdi, Xihu Qu, Hangzhou Shi, Zhejiang Sheng, China, 310007
- **Phone number:** +86 571 8799 8899



## Local Transportation from Airport to the Conference Hotel

- **Taxi Fee:** About 100 RMB



## Currency Exchange

You are recommended to **take enough Chinese Yuan\_with you from Japan,** or **exchange them in the airport.** We do not guarantee that there is currency exchange service in the conference hotel.

## Local Contact

**Name:** Miss Qian Bao

**Email:** qianbao@mail.tsinghua.edu.cn

**Mobile:** +86 18101216283

**Registration Fee:** 1200 RMB (*Cash only*)

**Hotel Price:** 450 RMB/night (*Receive credit card*)

## Attractions of the City of Hangzhou

1. **West Lake** West Lake is located to the west of Hangzhou, hence the name of "West Lake ". It is said that there are totally 36 "West Lakes "in China, the West Lake in Hangzhou is the most beautiful of 36 west lakes of the same name. It was a lagoon,once a small bay at the mouth of the Qiantang River. It was once the sea mouth that was blocked from the sea by river sediment and later became a lagoon lake.

2. **Lingyin Temple** The Lingyin (Souls Retreat) Temple is one of the most famous

Buddhist Temples in China. It was constructed in 326AD at the foot of Lingyin Mountain by an Indian monk named Huili with a history of more than 1,600 years. The temple has the Front Hall, the Hall and the Hall of Herb Masters. China's largest wood-carving of sitting Buddha Sakyamuni sits in the Magnificent Hall. The main hall is 33.6 meters in height.

3. **Hangzhou Bar Street** Just like each major city has its own unique streets of pub, as a world famous leisure city, Hangzhou is not lagged behind in this regard. Among the numerous bar and pub streets in Hangzhou, Nanshan Lu (or Nanshan Road) Bar Street is one of the best in terms of location, architecture, quality and taste.

4. **Pedestrian Streets in Hangzhou** Hangzhou has several walking only streets including Hefang Street (also named Qinghefang Street) and Zhongshan Road (also named "the Royal Street of the Southern Song Dynasty". The two main traditional and cultural streets form a pedestrian block reconstructing the former prime time of the Southern Song Dynast(1127–1279) when Hangzhou served as its capital about 800 years ago.

5. **Hiking Trail at Precious Stone Hill** There are numerous paths leading up the hill either from the lakeside, or from its north hillside. Precious Stone Hill (Baoshi Shan) is dotted with unique caves, Buddhist and Taoist shrines. The slender

Baochu Pagoda rises atop Precious Stone Hill where you can have a panoramic view of West Lake and the city as a whole, one of the two best West Lake viewing places with the other being Leifeng Pagoda.

6. **Grand Canal** The Beijing-Hangzhou Grand Canal is one of the greatest ancient projects in China and is the longest man-made canal in world. It covers a total length of 1,794 kilometers, and meets five rivers of the Yangtze, Yellow, Huaihe, Haihe, and Qiantang rivers. On June 22,2014, China's Grand Canal was listed as World Heritage. The decision is announced at the 38th session of UNESCO's World Heritage committee in Doha, Qatar

7. **Song Dynasty Theme Park** "Given me one day, I will give you back one thousand years," so promises the marketing mastermind behind the largest scale Song Dynasty theme park in China. This slogan does indeed deliver on its promise; this is one of the places around Hangzhou to experience ancient China within the walls of the park.

## Organizing Committee

### Chinese Side

Ke Deng (Chair)	Tsinghua University
Weihu Cheng	Beijing University of Technology
Wei Gao	Northeast Normal University
Zhi Geng	Peking University
Qihua Wang	Chinese Academy of Sciences
Chongqi Zhang	Guangzhou University
Lixin Zhang	Zhejiang University
Qiang Zhao	Shandong Normal University
Wensheng Zhu	Northeast Normal University

### Japanese Side

Toshinari Kamakura (Chair)	Chuo University
Hiroki Hashiguchi	Tokyo University of Science
Manabu Iwasaki	Seikei University
Masahiro Kuroda	Okayama University of Science
Yuichi Mori	Okayama University of Science
Shigekazu Nakagawa	Okayama University of Science
Michiko Watanabe	Keio University
Kazunori Yamaguchi	Rikkyo University



# Schedule for November 2, 2017 (Thursday)

Yunqi Hall (云栖厅) 3<sup>rd</sup> floor

## 08:20-08:30 Opening Ceremony

Chair: *Ke Deng*, Tsinghua University

## 08:30-09:50 Session I

Chair: *Toshinari Kamakura*, Chuo University

08:30-08:50 Estimation of treatment effects for heterogeneous matched pairs data with probit models

Speaker: *Wei Gao*, Northeastern Normal University

08:50-09:10 Visualization of relationship between group and characteristic using association rules

Speaker: *Yoshiro Yamamoto*, Tokai University

09:10-09:30 A robust statistical method based on principal Chi-squared test for assessing genetic association with multi-locus genotype data

Speaker: *Zhengbang Li*, Central China Normal University

09:30-09:50 A homogeneity test for odds ratio in 2x2 tables

Speaker: *Yoshinori Fujii*, University of Miyazaki

## 09:50-10:10 Tea Break

## 10:10-11:50 Session II

Chair: *Lixin Zhang*, Zhejiang University

10:10-10:30 Statistical methods for assessing safety risk of Imported Foods in China

Speaker: *Chongyuan Xu*, Tsinghua University

10:30-10:50 Cluster difference scaling for asymmetric dissimilarity data based on unfolding models

Speaker: *Kensuke Tanioka*, Wakayama Medical University

10:50-11:10 A novel approach for Markov random field with intractable normalizing constant on large lattices

Speaker: *Wanchuang Zhu*, Tsinghua University

11:10-11:30 Predictive inference with transferred priors

Speaker: *Jinfang Wang*, Chiba University

11:30-11:50 A Bayesian approach to real time monitor and forecast Chinese foodborne disease

Speaker: *Xueli Wang*, Beijing University of Posts and Telecommunications

## 11:50-12:10 Group Photo

12:10-13:30 Lunch Buffet Western Restaurant (恺撒宫), 1st floor

**13:30-15:30 Session III**

Chair: *Manabu Iwasaki*, Seikei University

13:30-13:50 Comparison of EER for W- and Z-rules when the dimension is large

Speaker: *Takayuki Yamada*, Kagoshima University

13:50-14:10 LPRE criterion based estimating equation approaches for the error-in-covariables multiplicative regression models

Speaker: *Qihua Wang*, Chinese Academic of Science

14:10-14:30 An incomplete-data Fisher scoring

Speaker: *Keiji Takai*, Kansai University

14:30-14:50 A generalized EMS algorithm for model selection with incomplete data

Speaker: Ping-Feng Xu, Changchun University of Technology

14:50-15:10 Bias reduction of ML estimators of Gamma distribution parameters and comparison with other methods.

Speaker: *Yumina Kodaira*, Graduate school of science and engineering, Chuo University

15:10-15:30 Random search algorithm for optimal mixture experimental design

Speaker: *Chongqi Zhang*, Guangzhou University

**15:30-15:50 Tea Break**

**15:50-17:30 Session IV**

Chair: *Qihua Wang*, Chinese Academic of Science

15:50-16:10 Measuring and analyzing intercultural competence

Speaker: *Sari Hosoya*, Kanto Gakuin University

16:10-16:30 The gap between DAR and AR Models: estimation and testing when coefficients are on the boundary

Speaker: *Feiyu Jiang*, Tsinghua University

16:30-16:50 A comparison study of rule space method and neural network model for learning diagnosis and its application

Speaker: *Atsuhiko Hayashi*, Nagoya Institute of Technology

16:50-17:10 Identifiability and estimation of causal mediation effects with missing data

Speaker: *Wei Li*, Peking University

17:10-17:30 Visualization for radiation monitoring post data using spatial interpolation

Speaker: *Fumio Ishioka*, Okayama University

**17:40-19:40 Banquet** Tian Xiang Yuan (天香苑), 3rd floor

# Schedule for November 3, 2017 (Friday)

Yunqi Hall (云栖厅) 3<sup>rd</sup> floor

## 08:30-10:10 Session V

Chair: *Kazunori Yamaguchi*, Rikkyo University

08:30-08:50 Statistical GAIT analysis based on the data set from glasses installed with IMU  
Speaker: *Yoshihiro Sone*, Graduate School of Science and Engineering, Chuo University

08:50-09:10 Proper Inference for Value Function in High-Dimensional Q-Learning for Dynamic Treatment Regimes  
Speaker: *Wensheng Zhu*, Northeastern Normal University

09:10-09:30 Item selection for impression survey  
Speaker: *Hiroko Katayama*, Graduate School of Informatics, Okayama University of Science

09:30-09:50 Sparse sliced inverse regression for high dimensional via Lasso  
Speaker: *Qian Lin*, Tsinghua University

09:50-10:10 Clustering of multivariate categorical data via penalized latent class analysis with dimension reduction  
Speaker: *Michio Yamamoto*, Okayama University

## 10:10-10:30 Tea Break

## 10:30-11:50 Session VI

Chair: *Zhi Geng*, Peking University

10:30-10:50 Discovering RNA interaction network by integrating sequence characteristics and expression profile of RNAs  
Speaker: *Qi Li*, Tsinghua University

10:50-11:10 A necessary test for sphericity based on uniformity on hyper-sphere  
Speaker: *Moe Amagai*, Graduate School of Science, Tokyo University of Science

11:10-11:30 Batch effects correction with unknown subtypes  
Speaker: *Yingying Wei*, The Chinese University of Hong Kong

11:30-11:50 Performance studies of test statistics for two sample directional data  
Speaker: *Yoshitomo Akimoto*, Graduate School of Science and Engineering, Chuo University

## 11:50-12:00 Ending Mark

Chair: *Ke Deng*, Tsinghua University

Towards the Next ICOTS10 in Kyoto

Speaker: *Kazunori YAMAGUCHI*

12:00-13:30 **Lunch Buffet** Western Restaurant (恺撒宫), 1st floor

# TOWARDS THE NEXT ICOTS10 IN KYOTO

July 8-13, 2018, Kyoto Japan

Kazunori YAMAGUCHI\* and Michiko WATANABE\*\*

\*Rikkyo University, Tokyo, Japan

\*\*Keio University, Tokyo Japan

---

## 1. Introduction

ICOTS (International Conference on Teaching Statistics) is held every four years. Its main purpose is to give statistic educators and professionals around the world the opportunity to exchange information, ideas and experiences, to present recent innovation and research in the field of statistic education, and to expand their range of collaborators.

## 2. ICOTS

We are at a critical time in statistics education where the world of data is changing rapidly. We need to be looking ahead to how as a field we will evolve and engage with the future. At the same time, we are celebrating our tenth ICOTS and this marks a time for us to look back on the past 40 years when in 1978, ISI's Education Committee Task Force was established to plan for the first ICOTS.

This is indeed an exciting time. It's clear that statistics education has matured as a field. Data have become part of everyday life, vital for professions and part of our very fabric as a society. Data are used everywhere to document, evaluate, plan and persuade. The very nature of what we call "data" is not what it was 10 years ago or even last year. Data science is emerging as a new field. And yet it is not clear if we are moving together or apart. Evidence exists that it is both. Both areas focus on variability, uncertainty and context but may approach the analysis and collection of data quite differently. In terms of education, what can we learn from each other? Where do we see ourselves going?

Next ICOTS will be held on July 8-13, 2018 in Kyoto Japan. Main theme of ICOTS-10 is "Looking back, looking forward". More than 500 international statistic educators and professionals will join the conference.

# Session I

November 2, 2017 (Thursday)

08:30-09:50

**Chair:**

Toshinari Kamakura, Chuo University

**Speaker:**

Wei Gao, Northeastern Normal University

Yoshiro Yamamoto, Tokai University

Zhengbang Li, Central China Normal University

Yoshinori Fujii, University of Miyazaki

**ESTIMATION OF TREATMENT EFFECTS FOR HETEROGENEOUS MATCHED  
PAIRS DATA WITH PROBIT MODELS**

Jun WANG, Wei GAO and Man-Lai TANG

Key Laboratory for Applied Statistics of MOE, School of Mathematics  
and Statistics

Northeast Normal University, Changchun, Jilin 130024, China

**Abstract:** Estimating the effect of medical treatments on subject responses is one of the crucial problems in medical research. Matched-pairs designs are commonly implemented in the field of medical research to eliminate confounding and improve efficiency. In this article, new estimators of treatment effects for heterogeneous matched pairs data are proposed. Asymptotic properties of the proposed estimators are derived. Simulation studies show that the proposed estimators have some advantages over the famous Heckman's estimator and inverse probability weighted (IPW) estimator. We apply the proposed methodologies to a blood lead level data set and an acute leukaemia data set.

# VISUALIZATION OF RELATIONSHIP BETWEEN GROUP AND CHARACTERISTIC USING ASSOCIATION RULES

Yoshiro YAMAMOTO and Sanetoshi YAMADA

Department of Mathematics and Graduate School of Science, Tokai University

Hiratsuka, Kanagawa 259-1292, Japan

## 1. Introduction

The information that customer data usually provides is the personal profile information including gender, age and so on. But, we can obtain the personal internal information by analyzing questionnaire data. We propose the visualization of multiple choice questionnaire to find the difference of the internal characteristic, about 6 ( $=2 \times 3$ ) layers we call the media layers. The media layers are M1 (male from 20 years old to 34 years old), M2 (male from 35 years old to 49 years old), M3 (male over 50 years old), F1 (female from 20 years old to 34 years old), F2 (female from 35 years old to 49 years old) and F3 (female over 50 years old).

Offered data have 31 multiple choice questionnaires. This time, we think about this question "Please check all appropriate items about your health worries". Figure 1-1 shows total results of people that checked "catch a cold easily" at this question. We could understand that total results are similar at all media layers. However, because the numbers of people of media layers are different from each other, if we look at each ratios of media layers, we understand that young groups have relatively at their health worries. About other health worries, M2 layer and M3 layer mind "body odor of old people", F1 layer and F2 layer mind "period pains". However, it is difficult to find out the difference of media layers about all health worries by this method because we should find out all question items.

## 2. Extraction of strong relationship between attribute and item by association rule analysis

We looked at the tendency of reaction of the media layer in Figure 1 about three question items, but we should find out all question items. There, we used association rule analysis (see [1,2,3]) as a method to find difference of tendency of answer by a group about all question items. It is basket data that association rule analyses are used well, but questionnaire data can be thought of 0-1 data as basket data. We also treated 0-1 data about media layer.

To find out about question items that reacted the media layer, condition part extracted only the rule that was the media layer. We used `apriori` function for association rule analysis in statistical analysis software R. In addition, we used `subset` function to extract the association rules. When we

set support more than 0.01, confidence more than 0.1 and lift more than 1.5 to find out strong rules, we could extract 27 rules. We can visualize association rules by `arulesViz` package ([8]). We could see the rule very well by this kind of plot, but the position between variables is unrelated to strength of relationship. Therefore we tried improvement of visualization to reflect the position relationship of variables.

Table 1. Association rules of health troubles by media layers

lhs		rhs	supp	conf	lift
F1	⇒	period pains	0.044	0.381	4.408
F2	⇒	period pains	0.039	0.255	2.946
F1	⇒	chalk mark	0.038	0.323	2.726
F1	⇒	anemia	0.019	0.164	2.389
F1	⇒	edema	0.033	0.283	2.354
F1	⇒	feeling of cold	0.055	0.471	2.230
F2	⇒	anemia	0.023	0.151	2.201
M3	⇒	high blood pressure	0.066	0.309	2.196
M3	⇒	athlete's foot	0.045	0.211	2.085
M1	⇒	nothing special	0.017	0.131	2.060
F2	⇒	edema	0.037	0.243	2.020
F3	⇒	the skin's aging	0.116	0.500	1.993
F1	⇒	tend to be constipated	0.036	0.309	1.932
F3	⇒	ache in the knee	0.058	0.250	1.918
F2	⇒	feeling of cold	0.060	0.400	1.894

### 3. Visualization of association rules using the correspondence analysis

We used the correspondence analysis (see [4,7]) to get the position relationship of variables. Correspondence analysis uses cross tabulation. And so that the correlation of the element of the row and the element of the column becomes biggest, correspondence analysis calculate row score and column score and plot them.

We explain how to make of the visualization. First, we performed the correspondence analysis of the result that performed cross tabulation by questionnaire items and media layers to set the position of each item, and we displayed the second axis of the correspondence analysis (Figure 1 left). Next, we displayed circles that size depended on the number of the check and we displayed squares that size depended on the number of each media layer. Then, if its media layer is male, color of squares are blue, if its media layer is female, color of squares are red, and if its media layer is old, color of squares are deep (Figure 1 right).



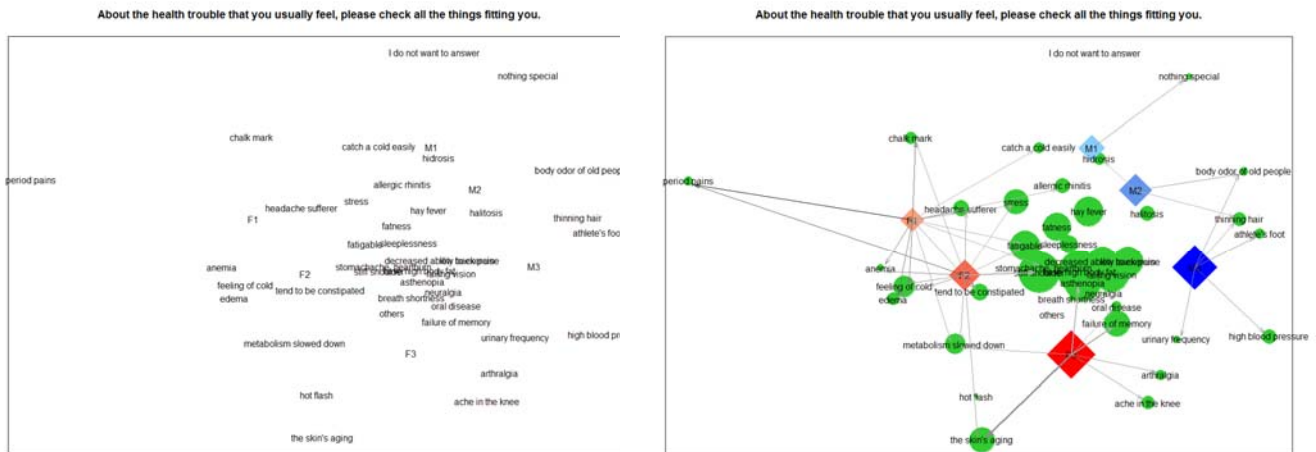


Figure 1. Visualization between group and characteristic using Biplot (left) and proposed visualization (right)

Then, if support of rule is high, the thickness of the arrow of this visualization is big, and if lift of rule is high, the depth of the arrow of this visualization is deep. In this way, we made the visualization of the questionnaire result that reflected the position relationship of items.

By Figure 2, about the question of the worry of the health, we understand that M2 layer and M3 layer are worried in "body odor of old people" and F1 layer and F2 layer are worried in "period pains". And, because there is "overweight" midmost and there is much answer number, we understand many people are worried in "fatness". But, because association rules are not shown, we understand that the relations with the specific media layer are not accepted on "fatness". In this way, we can do much consideration in one plot.

#### 4. Interactive plot on RStudio and Shiny application

About the association rule, by rules extracted by setting of the lower limit of the support, confidence and lift changes, differences of plot that is obtained are seen. To display rules that are easy to characterize with the media layer, so that we can interactively coordinate the parameter of association rules about this plot, we built the indication system using the `manipulate` function on RStudio and Shiny Application (Figure 2). Then, for the improvement of indication and plots of other questions. It is possible to change and adjustment of question contents, size of fonts, size of circles and size of squares.

#### Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP15K11997. The data source was provided by Macromill, Inc.

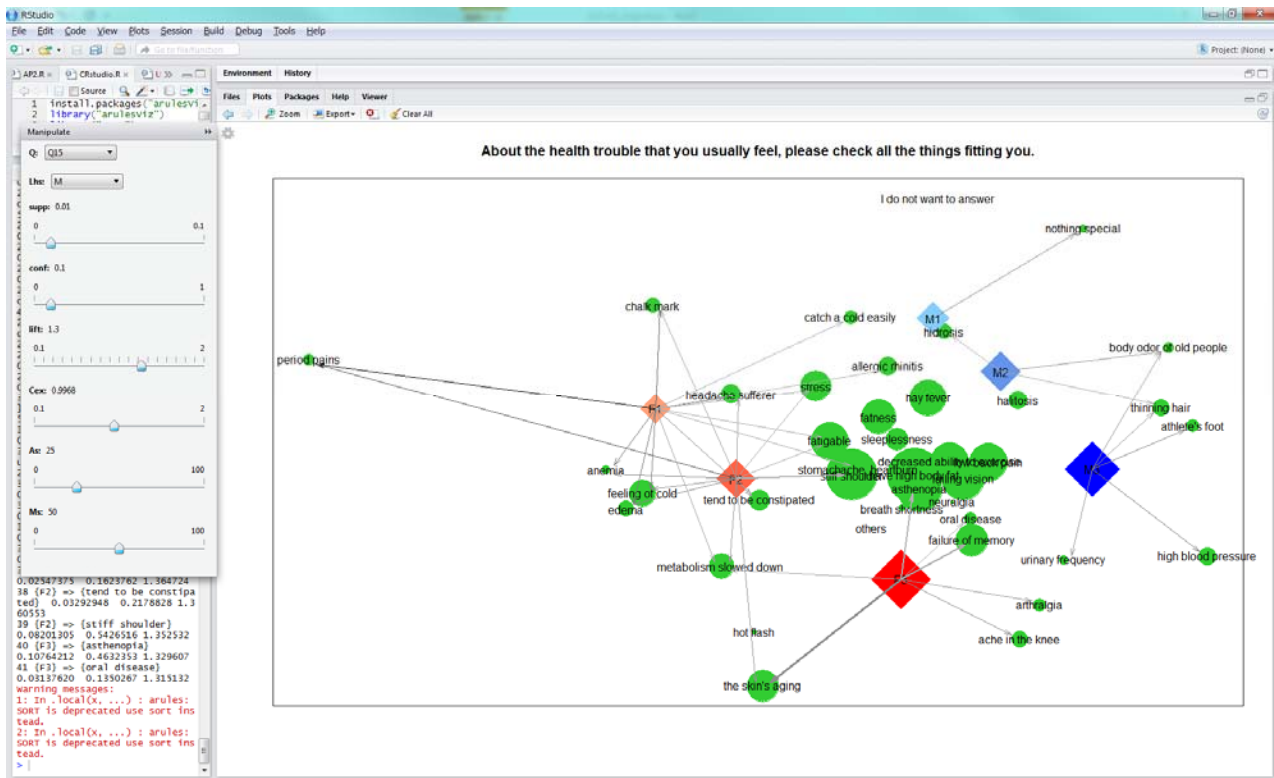


Figure 1. The interactive plot of multiple choice questionnaires by RStudio

## References

- [1] Agrawal, R., Imielinski, T., Swami, A. (1993) Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 207-216.
- [2] Agrawal, R., Srikant, R. (1994) Fast Algorithms for Mining Association Rules, *Proc. 20th int. conf. very large data bases, VLDB 1215*, 487-499.
- [3] Ito, A., Yoshikawa, T., Furuhashi, T., Ikeda, R., Kato, T. (2010) Search of an interesting rule using the visualization in the association analysis, *26th Fuzzy System Symposium*, 684-689
- [4] Gower, J.C., Hand, D.J. (1996) *Biplots*. Chapman & Hall.
- [5] Lantz, B. (2015) *Machine Learning with R. 2nd ed.* Packt Publishing.
- [6] Linoff, G.S., Berry, M.J.A. (2011) *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 3rd ed.*, Wiley.
- [7] Venables, W.N., Ripley, B.D. (2002) *Modern Applied Statistics with S. 4th ed.*, Springer.
- [8] Wong, P.C., Whitney, P., Thomas, J. (1999) *Visualizing Association Rules for Text Mining*, Pacific North-west National Laboratory.

**A ROBUST STATISTICAL METHOD BASED ON PRINCIPAL CHI-SQUARED  
TEST FOR ASSESSING GENETIC ASSOCIATION WITH MULTI-LOCUS  
GENOTYPE DATA**

Zhengbang Li

Central China Normal University

**Abstract:** Simultaneous testing of multiple genetic variants for association has been widely recognized as a valuable approach complementary to single marker tests. As such, principal components regression(PCR) has been found to have competitive power. Here, we firstly propose a new global test by the use of co-dominant codes for all markers and PCR theory. The new global test is built on an empirical Bayes type of score statistic for testing marginal associations with each single marker. The new global test gains power by effectively using linkage disequilibrium among testing markers. The new global test reduces to PCR when the genotype for each marker is coded as the number of minor alleles. This connection lends insight into the power of the new global test relative to PCR and some other popular multi-marker test methods. When the real disease causal marker adopts additive code, the new global test drop some powers with Hardy-Weinberg Equilibrium in the control population. We propose a robust test method by taking the minimum p-value of the new global test and PCR based on genotype data adopting additive codes. Through extensive simulation studies and real data analysis about the association

between pancreatic cancer and genes, we show that the proposed robust testing method can gain desirable power and can often identify association signals that may be missed by existing methods.

# A HOMOGENEITY TEST FOR ODDS RATIO IN 2 X 2 TABLES

Yoshinori FUJII

Faculty of Education, University of Miyazaki

Miyazaki 889-2192, Japan

---

## 1. Introduction

To investigate the association between two categorical variables, we often need to consider the effects of other variables. A method to solve the problem is stratifying. Summary odds ratio is used to present the association of two categorical variables based on stratified 2x2 tables. For interpreting the association it is important to check the assumption of homogeneity of odds ratios. Several tests were proposed, for example Breslow-Day test, likelihood ratio test and Cochran's Q test. These tests are constructed in the situation that the number of strata is small and the sample sizes of 2 x 2 tables are large. On the other hand Fisher's exact test are recommended when we analyses the 2 x 2 table with small sample size. In this talk we show an idea to construct a homogeneity test for odds ratios when the sample sizes of some tables are small.

## 2. Models and Methods

Consider stratified 2 x 2 tables. For k-th stratum  $x_k$  and  $y_k$  are independently distributed and the distributions of them are binomial distributions  $B(n_{1,k}, p_k)$  and  $B(n_{2,k}, q_k)$  respectively ( $k=1,2,\dots,K$ ). The odds ratio  $\psi_k$  of k-th stratum is given by  $p_k(1-q_k)/q_k(1-p_k)$ . To summarize the overall association Mantel-Haenszel estimator

$$\hat{\psi}_{MH} = \frac{\sum_{k=1}^K x_k(n_{2,k} - y_k)/(n_{1,k} + n_{2,k})}{\sum_{k=1}^K y_k(n_{1,k} - x_k)/(n_{1,k} + n_{2,k})}$$

is used.

We consider the homogeneity test for odds ratio with null hypothesis  $H_0 : \psi_1 = \psi_2 = \dots = \psi_K$  vs alternative  $H_1 : \text{not } H_0$ . In this paper following three tests are considered.

Cochran's Q-test (Kulinskaya and Dollinger, 2015)

$$T_1 = \sum_{k=1}^K \hat{w}_k (\hat{\theta}_k - \hat{\theta}_W)^2$$

where  $\hat{\theta}_k = \log \frac{x_k + 1/2}{n_{1,k} - x_k + 1/2} - \log \frac{y_k + 1/2}{n_{2,k} - y_k + 1/2}$ ,  $\hat{\theta}_W = \sum_{k=1}^K \hat{w}_k \hat{\theta}_k / \sum_{k=1}^K \hat{w}_k$  and

$$\hat{w}_k^{-1} = \frac{1}{x_k + 1/2} + \frac{1}{n_{1,k} - x_k + 1/2} + \frac{1}{y_k + 1/2} + \frac{1}{n_{2,k} - y_k + 1/2} .$$

Breslow-Day test (Breslow and Day,1980)

$$T_2 = \sum_{k=1}^K X_k^2$$

where  $X_k^2 = \frac{n_{1,k}(x_k - m_{1,k})^2}{m_{1,k}(n_{1,k} - m_{1,k})} + \frac{n_{2,k}(y_k - m_{2,k})^2}{m_{2,k}(n_{2,k} - m_{2,k})}$  and  $m_{1,k}$  and  $m_{2,k}$  are satisfied two

equations  $m_{1,k} + m_{2,k} = x_k + y_k$  and  $\frac{m_{1,k}(n_{2,k} - m_{2,k})}{m_{2,k}(n_{1,k} - m_{1,k})} = \hat{\psi}_{MH}$  .

We propose a new test for considering the small sample cases. At first we consider the testing  $\psi = \hat{\psi}_{MH}$  for the k-th 2x2 table based on non-central hypergeometric distribution and the mid-p value of the test is denoted by  $pv_k$  . The cumulative distribution function of the chi-squared

distribution with one degree of freedom is denoted by  $F_{chisq}$  . So we proposed the following test.

Proposed test

$$T_3 = \sum_{k=1}^K Y_k^2$$

where  $Y_k^2 = X_k^2$  for large sample size case and  $Y_k^2 = F_{chisq}(pv_k)$  for small sample size case.

### 3. Simulation Study

We conducted the simulation study to compare above three tests. The setting and the detailed results of the simulation is given in my talk.

### References

Breslow NE, Day NE.(1980) Statistical methods in cancer research. Int Agency Res Cancer. 136–46, 1980.

Kulinskaya E. and Dollinger M. B. (2015) An accurate test for homogeneity of odds ratios based on Cochran's Q-statistic. BMC Medical Research Methodology, 15-49.

# Session II

November 2, 2017 (Thursday)

10:10-11:50

**Chair:**

Lixin Zhang, Zhejiang University

**Speaker:**

Chongyuan Xu, Tsinghua University

Kensuke Tanioka, Wakayama Medical University

Wanchuang Zhu, Tsinghua University

Jinfang Wang, Chiba University

Xueli Wang, Beijing University of Posts and Telecommunications

# STATISTICAL METHODS FOR ASSESSING SAFETY RISK OF IMPORTED FOODS IN CHINA

Chongyuan XU

Center for Statistical Science and Department of Industrial Engineering,  
Tsinghua University, Beijing 100084, P.R. China

**Abstract:** Currently the volume of imported food to China is increasing every year, but the relevant supervision cost remains the same, so increasing the supervision efficiency is of importance. A reasonable scheme is to focus the resources mainly on the most “dangerous” foods, which is based on quantifying the safety risk of each kind of food. One important step is to estimate the rate of substandard imported food for each, categorized by any combination of a food category, a toxin to be inspected and an exporting country. It is a classical statistical problem of inferring parameters of multiple Bernoulli samples. It is challengeable to estimate the rates of substandard food because the sample size is seriously inadequate. However, this problem can be solved by using the correlation among all rates of substandard imported food. In this work, we propose a method mainly based on Hierarchical Bayes Model and Factor Model.



# CLUSTER DIFFERENCE SCALING FOR ASYMMETRIC DISSIMILARITY DATA BASED ON UNFOLDING MODELS

\*Kensuke TANIOKA and \*\*Hiroshi YADOHISA

\*School of Medicine, Wakayama Medical University, Kimiidera 811-1, Wakayama, Japan

\*\* Department of Culture and Information Science, Doshisha University,  
Tataramiyakodani 1-3, Kyoto, Japan.

---

## 1 Introduction

Asymmetric dissimilarity data are observed in various domains such as the marketing area. Given asymmetric dissimilarity data in some situations, it is important to comprehend the asymmetric relation between objects. Asymmetric Multidimensional Scaling (AMDS) is one of the approaches to understand asymmetric relations (Borg and Groenen, 2005). However, if the number of objects in the asymmetric data is large, it may be difficult for us to comprehend the results. This is because we have to interpret the relations between asymmetries, through vectors or radii, in addition to coordinates of objects. To overcome this problem, we propose simultaneous clustering and AMDS method based on Unfolding (Zielman and Heiser, 1993) and Cluster Difference Scaling (CDS) (Heiser and Groenen, 1997), called Cluster Difference Scaling for Asymmetric dissimilarity data based on Unfolding (CDSAU). In the proposed method, we can understand the asymmetric relations between clusters, visually. This method has two advantages. First, through the estimated coordinates of cluster centroids, we can easily understand the asymmetric relations between clusters. Second, the proposed method can be considered a generalization of AMDS based on Unfolding, which include existing AMDS methods.

## 2 Model and objective function of CDSAU

Here, we introduce model and the objective function of the proposed method.

### Definition 2.1 Model of CDSAU

Let  $\Delta = (\delta_{ij})$ ,  $\delta_{ij} \in \mathbb{R}_+$  ( $i, j = 1, 2, \dots, n$ ) be an asymmetric dissimilarity matrix for objects. Given  $\Delta$ , the number of dimensions  $d$ , the number of clusters for objects  $k$  and the number of slide-vectors  $m$  ( $\leq k$ ), the model of CDSAU is defined as follows: for arbitrarily  $i, j = 1, 2, \dots, n$ , there exists  $o, \ell = 1, 2, \dots, k$  and  $s = 1, 2, \dots, m$  such that

$$\delta_{ij} = d_{ols}(\mathbf{X}, \mathbf{Z}) + e_{ij} \quad (o, \ell = 1, 2, \dots, k; s = 1, 2, \dots, m) \quad (1)$$

where

$$d_{ols}(\mathbf{X}, \mathbf{Z}) = \|\mathbf{x}_o - (\mathbf{x}_\ell - \mathbf{z}_s)\|,$$

$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)^T$ ,  $\mathbf{x}_o = (x_{ot}) \in \mathbb{R}^d$  ( $i = 1, 2, \dots, n; t = 1, 2, \dots, d$ ) be coordinates matrix of cluster  $o$  ( $o = 1, 2, \dots, k$ ),  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m)^T$ ,  $\mathbf{z}_s = (z_{st}) \in \mathbb{R}^d$  be slide-vectors,  $\|\cdot\|$  is Euclidean norm and  $e_{ij} \in \mathbb{R}$  ( $i, j = 1, 2, \dots, n$ ) is error.

See Figure 1. This is the example of the interpretation for CDSAU.  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are coordinates of cluster  $i$  and  $j$ , respectively, and  $\mathbf{z}$  is a slide vector. The left part and the right part of Figure 1 represent distance from cluster  $i$  to cluster  $j$  and from cluster  $j$  to cluster  $i$ , respectively. In the case, distance from cluster  $i$  to cluster  $j$  is relatively close to the distance from cluster  $j$  to  $i$ . In short, the asymmetric relation between clusters can be described by the direction of the slide-vector in this model.

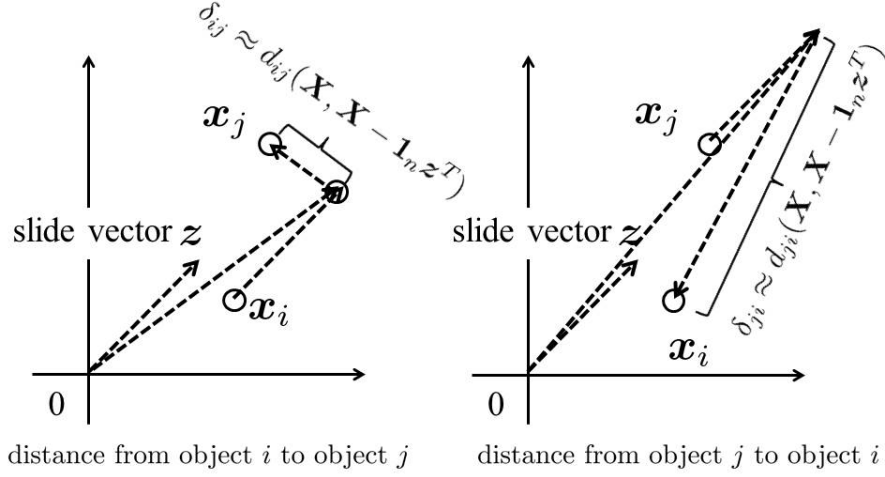


Figure 1: The example of interpretation for the CDSAU in the case of  $k = 2$  and  $m = 1$

From the model (1), in CDSAU, it can be selected for the kinds of slide-vectors. Next, we define the objective function of CDSAU based on the model (1).

**Definition 2.2** *Objective function of CDSAU*

Given asymmetric dissimilarity data  $\Delta = (\delta_{ij})$   $\delta_{ij} \in \mathbb{R}_+$ , the number of dimensions  $d$ , the number of clusters  $k$  and the number of slide-vectors  $m$  the objective function of CDSAU is defined as follows:

$$L(\mathbf{X}, \mathbf{Z}, \mathbf{U}, \Psi | \Delta) = \sum_{i=1}^n \sum_{j=1}^n \sum_{o=1}^k \sum_{\ell=1}^k \sum_{s=1}^m u_{io} u_{j\ell} \psi_{\ell s} (\delta_{ij} - d_{ols}(\mathbf{X}, \mathbf{Z}))^2. \quad (2)$$

where  $\mathbf{U} = (u_{io})$ ,  $u_{io} \in \{0, 1\}$  ( $i = 1, 2, \dots, n; o = 1, 2, \dots, k$ ) be an indicator matrix of objects, and  $\Psi = (\psi_{\ell s})$ ,  $\psi_{\ell s} \in \{0, 1\}$  ( $\ell = 1, 2, \dots, k; s = 1, 2, \dots, m$ ) be indicator matrix of these cluster centroids,

Here, examples of indicator matrix  $\mathbf{U}$  and  $\Psi$  in the case of  $n = 6$ ,  $k = 3$  and  $m = 2$  are as follows:

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \Psi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The row and column of  $\mathbf{U}$  represent objects and clusters, respectively. In the case of  $\mathbf{U}$ , object 1 and 2 belong to cluster 1, object 3 and 4 belong to cluster 2 and object 5 belong to cluster 3, respectively. The row and column of  $\Psi$  represent clusters and slide-vectors, respectively. As the same way of  $\mathbf{U}$ , in the case of  $\Psi$ , cluster 1 and cluster 2 belong to slide-vector 1, and cluster 3 belong to slide-vector 2, respectively.

The objective function (2) can be described as the special case of symmetric MDS. The relation between symmetric MDS and the proposed method is shown.

**Proposition 2.1** *The other description for objective function of CDSAU*

*Eq. (2) is redescribed as follows:*

$$L(\mathbf{X}, \mathbf{Z}, \mathbf{U}, \Psi | \Delta) = \frac{1}{2} \left\| \mathbf{W} \odot [\Delta^\dagger - D(\Phi Q)] \right\|^2 \quad (3)$$

where

$$\begin{aligned} \mathbf{W} &= \begin{bmatrix} \mathbf{O}_{n,n} & \mathbf{1}_n \mathbf{1}_n^T \\ \mathbf{1}_n \mathbf{1}_n^T & \mathbf{O}_{n,n} \end{bmatrix}, & \Delta^\dagger &= \begin{bmatrix} \mathbf{O}_{n,n} & \Delta \\ \Delta^T & \mathbf{O}_{n,n} \end{bmatrix} \\ \Phi &= \begin{bmatrix} \mathbf{U} & \mathbf{O}_{n,m} \\ \mathbf{U} & -\mathbf{U}\Psi \end{bmatrix}, & \mathbf{Q} &= \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix}, \end{aligned} \quad (4)$$

$\odot$  indicates Hadmard product such that  $\mathbf{A} \odot \mathbf{B} = (a_{ij}b_{ij})$  for any  $n$  by  $p$  matrix,  $\mathbf{A}$  and  $\mathbf{B}$ . Here,  $\mathbf{O}_{n,m} = (0)$  is  $n$  by  $m$  matrix and  $\mathbf{1}_n = (1)$  is vector with the length of  $n$ . In addition to that, for  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$   $\mathbf{y}_i \in \mathbb{R}^d$ ,

$$D(\mathbf{Y}) = (d_{ij}(\mathbf{Y})), \quad d_{ij}(\mathbf{Y}) = \|\mathbf{y}_i - \mathbf{y}_j\| \quad (i, j = 1, 2, \dots, n)$$

The objective function of the ordinal symmetric MDS can be described as follows:

$$\|\Delta - D(\mathbf{Y})\|^2,$$

where  $\Delta$  and  $\mathbf{Y}$  are symmetric dissimilarity data and coordinates of objects, respectively. Therefore, the difference between the ordinal MDS and CDSAU is depending on  $\mathbf{W}$  and coordinates of objects. The objective function of CDSAU can be considered as ordinal MDS with missing values for dissimilarity data and distance data. In addition to that, the difference between ordinal Unfolding and CDSAU is depending only on coordinates. Especially, in CDSAU, the coordinates is added to constraints such as product of indicator matrix and coordinates matrix. Finally, if  $k = m$ , CDSAU become the same method of CDU (Vera et al., 2013).

### 3 Algorithm

Here, we show the algorithm of CDSAU based on Alternative least squares criterion (ALS).

**Algorithm of CDSAU**

**Initial Step:**

Set  $n, k, m, \mathbf{U}, \Psi, \mathbf{X}, \mathbf{Z}$  and  $\mathbf{H}$  as initial parameters,  
where  $\mathbf{H} = (h_{s^\dagger t})$   $h_{s^\dagger t} \in \mathbb{R}$  ( $s^\dagger = 1, 2, \dots, (k+m); t = 1, 2, \dots, d$ )

**Scaling Step:**

**Scaling Step 1:** Update  $\mathbf{X}$  and  $\mathbf{Z}$  by Eq.(5), given  $\mathbf{U}$  and  $\Psi$

**Scaling Step 2:** Update  $\mathbf{H} \leftarrow \mathbf{Q}$ , given  $\mathbf{X}, \mathbf{Z}, \mathbf{U}$ , and  $\Psi$

**Scaling Step 3:** If the stop condition of Scaling step, go to clustering step, else, back to Scaling Step 1.

**Clustering Step:**

**Clustering Step 1** Update  $\mathbf{U}$ , given  $\Phi, \mathbf{X}$  and  $\mathbf{Z}$

**Clustering Step 2** Update  $\Psi$ , given  $\mathbf{U}, \mathbf{X}$  and  $\mathbf{Z}$

**Final Step**

If stop condition is satisfied, stop this algorithm, else return to the Scaling Step.

Next, to explain the estimation way of  $\mathbf{X}$  and  $\mathbf{Z}$ , majorizing function of CDSAU is derived as the same way of SMACOF algorithm (De Leeuw and Heiser, 1980). The majorizing function can be derived based on Cauchy-Schwarz inequality.

Next, we show updating formula for  $\mathbf{X}$  and  $\mathbf{Z}$  based on the majorizing function.

**Proposition 3.1** *Updating formula of  $\mathbf{Q}$*

Given  $\Phi$  and  $\mathbf{H}$ , updating formula of  $\mathbf{Q}$  is derived as follows:

$$\mathbf{Q} = [\Phi^T \mathbf{V} \Phi]^+ \Phi^T \mathbf{B}(\Phi \mathbf{H}) \Phi \mathbf{H} \quad (5)$$

where  $[\Phi^T \mathbf{V} \Phi]^+$  is the Moore-Penrose inverse of  $\Phi^T \mathbf{V} \Phi$ .

$$\begin{aligned} \mathbf{V} &= \sum_{i^*=1}^{2n} \sum_{j^*=1}^{2n} w_{i^*j^*} (\mathbf{e}_{i^*} - \mathbf{e}_{j^*})(\mathbf{e}_{i^*} - \mathbf{e}_{j^*})^T, \\ \mathbf{e}_{i^*} &= (e_{i^*s^*}) \quad e_{i^*s^*} = \begin{cases} 1 & (i^* = s^*) \\ 0 & (i^* \neq s^*) \end{cases}, \quad (i^*, s^* = 1, 2, \dots, 2n), \\ \mathbf{B}(\Phi \mathbf{H}) &= (b_{i^*j^*}) \quad (i^*, j^* = 1, 2, \dots, 2n) \\ b_{i^*j^*} &= \begin{cases} -\frac{w_{i^*j^*} \delta_{i^*j^*}^\dagger}{d_{i^*j^*}(\Phi \mathbf{H})} & (\text{if } i^* \neq j^* \text{ and } d_{i^*j^*}(\Phi \mathbf{H}) \neq 0) \\ 0 & (\text{if } i^* \neq j^* \text{ and } d_{i^*j^*}(\Phi \mathbf{H}) = 0) \end{cases} \quad \text{and} \\ b_{i^*i^*} &= - \sum_{(j^*=1) \wedge (i^* \neq j^*)}^{2n} b_{i^*j^*} \end{aligned}$$

For the estimation of  $\mathbf{U}$  and  $\Phi$ , the solution corresponding to minimized value of the objective function is selected among all combinations, respectively.

## 4 Conclusions

In this presentation, we proposed simultaneous method and AMDS based on Unfolding. CDSAU can represent the asymmetric relations between clusters even if the number of objects is large. However, there are several future works. For example, the number of clusters and the number of slide-vectors are tuning parameters and there are no ways to choose. In addition to that, there are many local optima since there are two indicator matrices in CDSAU.

## References

- [1] Borg, I. and Groenen, P.J.F. (2005) *Modern multidimensional scaling: Theory and applications*. 2nd Ed. New York: Springer.
- [2] De Leeuw, J. and Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration, In P. Krishnaiah (Eds.), *Multivariate analysis, volume V* (pp.501-522). North Holland Publishing Company, Amsterdam, The Netherlands.
- [3] Heiser, W. J., and Groenen, P. J. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, **62**, 529-550.
- [4] Vera, J. F., Macias, R., and Heiser, W. J. (2013). Cluster differences unfolding for two-way two-mode preference rating data. *Journal of Classification*, **30**, 370-396.
- [5] Zielman, B., and Heiser, W. J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika*, **58**, 101-114.

# A NOVEL APPROACH FOR MARKOV RANDOM FIELD WITH INTRACTABLE NORMALIZING CONSTANT ON LARGE LATTICES

Wanchuang Zhu

Center for Statistical Science and Department of Industrial Engineering,  
Tsinghua University, Beijing 100084, P.R. China

**Abstract:** The pseudo likelihood method of Besag (1974) has remained a popular method for estimating Markov random field on a very large lattice, despite various documented deficiencies. This is partly because it remains the only computationally tractable method for large lattices. We introduce a novel method to estimate Markov random fields defined on a regular lattice. The method takes advantage of conditional independence structures and recursively decomposes a large lattice into smaller sublattices. An approximation is made at each decomposition. Doing so completely avoids the need to compute the troublesome normalizing constant. The computational complexity is  $O(N)$ , where  $N$  is the number of pixels in the lattice, making it computationally attractive for very large lattices. We show through simulations, that the proposed method performs well, even when compared to the methods using exact likelihoods.

# PREDICTIVE INFERENCE WITH TRANSFERRED PRIORS

Jinfang WANG and Shigetoshi HOSAKA

Department of Mathematics and Informatics, Chiba University, Japan  
Hosaka Clinic of Internal Medicine, Tokyo, Japan

---

## 1 Introduction

In this paper we introduce a novel approach for computing Bayesian predictive distributions using priors derived from a second separate data set. We illustrate this method by showing how to compute predictive distributions of HbA1c, an indicator for diabetes, using data obtained in a hospital in Tokyo.

## 2 Transferred Prior

Figure 1 is a diagram showing how to compute the transferred posterior distributions based on two sets of data. The *target data* is a set of data obtained by the researchers, with usually detailed information on many aspects of each subject. The *source data* is a set of data, usually large but coarse, such as public survey data available on the web. In our study, we shall consider source data in table forms.

Let  $\theta = (\eta', \zeta')$  be the regression parameter based on target data, which is partitioned into  $\eta$  and  $\zeta$ . Here  $\eta$  is regression parameter based on the source data, say  $D_s$ . The dimension of  $\zeta$  is usually much larger than that of  $\eta$ . We make the important conditional independence assumption of  $\zeta$  and  $D_s$ :

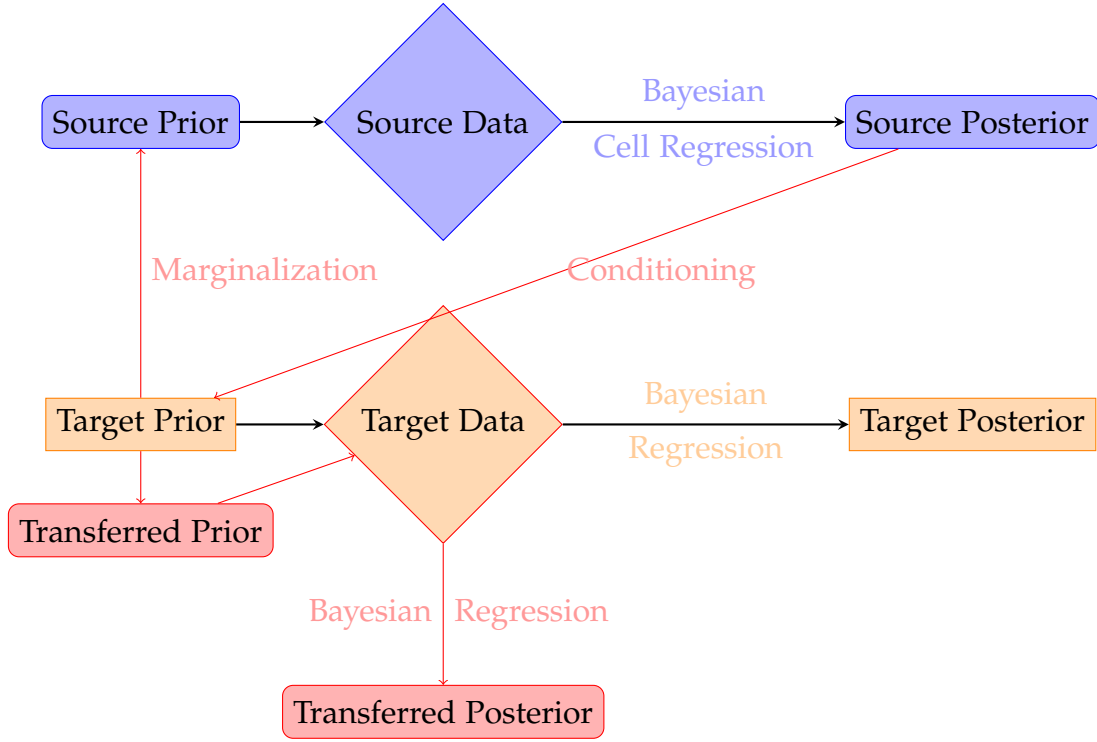
$$p(\zeta | \eta, D_s) = p(\zeta | \eta), \quad (1)$$

which states that  $\zeta$  is independent of  $D_s$  given the source regression parameter  $\eta$ . We shall introduce a framework for computing the predictive distribution of  $\theta$  using priors based on  $D_s$ .

In the rest of the paper, we outline the steps for obtaining the *transferred priors*, the prior distributions of  $\theta$  utilizing the source data  $D_s$ . Once this transferred prior distribution is obtained, we can perform traditional Bayesian analyses using techniques such as MCMC. The transferred prior can be obtained through the following steps.

1. Let  $p(\theta)$  be a prior for  $\theta$ . For instance,  $p(\theta)$  may be chosen as non-informative. We call this prior the *target prior*.
2. Next, we obtain the *source prior* for  $\eta$  by marginalizing  $\theta$ :

$$p(\eta) = \int p(\theta) d\zeta = \int p(\eta, \zeta) d\zeta.$$



**Figure 1:** A diagram showing how to compute the transferred posterior distributions using two sets of data, the target data and the source data.

3. Applying the Bayesian cell regression (Wang and Hosaka, 2017) to  $D_s$  with prior  $p(\boldsymbol{\eta})$ , we obtain the *source posterior*  $p(\boldsymbol{\eta} | D_s)$  for  $\boldsymbol{\eta}$ .
4. Finally, we obtain the desired *transferred prior*  $p(\boldsymbol{\theta} | D_s)$  as follows

$$\begin{aligned}
 p(\boldsymbol{\theta} | D_s) &= p(\boldsymbol{\eta}, \boldsymbol{\xi} | D_s) \\
 &= p(\boldsymbol{\xi} | \boldsymbol{\eta}, D_s) p(\boldsymbol{\eta} | D_s) \\
 &= p(\boldsymbol{\xi} | \boldsymbol{\eta}) p(\boldsymbol{\eta} | D_s).
 \end{aligned} \tag{2}$$

## References

- Andrew D. Martin, Kevin M. Quinn and Jong Hee Park (2011). MCMCpack: Markov Chain Monte Carlo in R, *Journal of Statistical Software*, **42**, 1–21.
- Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association*, **91**, 1343–1370.
- Wang, J. and Hosaka, S. (2017). Cell regression and reference prior, Symposium on “*Statistical Modelling and Computational Algorithms*”, at Nagoya University, Febuary 18–19, 2017.

**A BAYESIAN APPROACH TO REAL-TIME MONITOR and FORECAST  
CHINESE FOODBORNE DISEASE**

Xueli WANG<sup>1</sup>, Moqin ZHOU<sup>1</sup>, Jinzhu JIA<sup>2</sup>, Zhi GENG<sup>2</sup> and Gexin XIAO<sup>3</sup>

<sup>1</sup> School of Science, Beijing University of Posts and Telecommunications,  
Beijing, 100876, China

<sup>2</sup> School of Public Health, Peking University, Beijing, 100871, China

<sup>3</sup> China National Center For Food Safety Risk Assessment, Beijing, 100022,  
China

**Abstract:** Foodborne diseases bring big impact on public health and they are often underreported. It is because a lot of patients usually have delay to see doctors when they suffer from foodborne diseases. In this paper, we study the Chinese foodborne surveillance data. We build a Bayesian hierarchical model to monitor and forecast Chinese foodborne disease. We propose several scoring rules to assess the performance of different nowcasting procedures. We conclude that Bayesian nowcasting with consideration of right truncation of the reporting delay has the nice performance for shortterm forecasting analysis, and the Bayesian nowcasting could be a timely and valuable tool to better approximate current epidemic trends, which is crucial for public health.



# Session III

November 2, 2017 (Thursday)

13:30-15:30

**Chair:**

Manabu Iwasaki, Seikei University

**Speaker:**

Takayuki Yamada, Kagoshima University

Qihua Wang, Chinese Academic of Science

Keiji Takai, Kansai University

Ping-Feng Xu, Changchun University of Technology

Yumina Kodaira, Chuo University

Chongqi Zhang, Guangzhou University

# COMPARISON OF EER FOR W- AND Z- RULES WHEN THE DIMENSION IS LARGE

Takayuki YAMADA<sup>1</sup>, Tetsuro SAKURAI<sup>2</sup> and Yasunori FUJIKOSHI<sup>3</sup>

<sup>1</sup> Institute for Comprehensive Education, Center of General Education, Kagoshima University  
1-21-30 Korimoto, Kagoshima 890-0065 Japan

<sup>2</sup> Center of General Education, Tokyo University of Science, Suwa  
5000-1, Toyohira, Chino-shi, Nagano, 391-0292, Japan

<sup>3</sup> Department of Mathematics, Graduate School of Science, Hiroshima University  
1-3-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8626, Japan

## Summary

This paper is concerned with high-dimensional asymptotic results for  $W$ - and  $Z$ - rules when the sample size  $N$  and the dimension are large. Firstly, we obtain the EPMCs (Expected Probability of Misclassifications) of  $W$ - and  $Z$ - rules in expanded forms with errors of  $O(N^{-2})$ . It is pointed out that  $Z$ -rule has smaller EER (Expected Error Rate) than  $W$ -rule when the prior probabilities are the same, neglecting the terms of  $O(N^{-2})$ . We checked accuracies of our asymptotic results numerically by conducting a Monte Carlo simulation.

*Key Words and Phrases:* Discriminant analysis, EPMC, High-dimensional asymptotic results,  $W$ - rule,  $Z$ -rule.

## 1. Introduction

This paper is concerned with the problem of classifying an observation vector  $\mathbf{x}$  as coming from one of two populations  $\Pi_1$  and  $\Pi_2$ . Let  $\Pi_i$  have  $p$ -dimensional normal populations with mean vectors  $\boldsymbol{\mu}_i$  and the  $p \times p$  common positive definite covariance matrix  $\boldsymbol{\Sigma}$ , which are denoted as  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ . Consider the case that all parameters are unknown. Suppose that the observation vectors  $\mathbf{x}_{1i}, \dots, \mathbf{x}_{N_i, i}$  are independently and identically distributed (i.i.d.) as  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ ,  $i = 1, 2$ . Let  $W(\mathbf{x})$  be the linear discriminant function

$$W(\mathbf{x}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \left\{ \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right\},$$

where  $\bar{\mathbf{x}}_1$ ,  $\bar{\mathbf{x}}_2$  and  $\mathbf{S}$  are the sample mean vectors and the pooled sample covariance matrix defined by

$$\bar{\mathbf{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_{ij}, \quad i = 1, 2, \quad \mathbf{S} = \frac{1}{n} \sum_{i=1}^2 \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)', \quad n = N - 2 = N_1 + N_2 - 2.$$

Then, the linear discriminant rule with a cutoff point  $c$ , which is also called  $W$ -rule, classifies  $\mathbf{x}$  as  $\Pi_1$  if  $W(\mathbf{x}) > c$  for a constant  $c$ , and as  $\Pi_2$  if  $W(\mathbf{x}) < c$ . Furthermore, the discriminant rule, which is based on the likelihood ratio criterion for testing the composite null hypothesis that  $\mathbf{x}, \mathbf{x}_{11}, \dots, \mathbf{x}_{1N_1} \in \Pi_1$  against the composite alternative hypothesis that  $\mathbf{x}, \mathbf{x}_{21}, \dots, \mathbf{x}_{2N_2} \in \Pi_2$ , is called maximum likelihood rule or  $Z$ -rule (see, e.g., Anderson [1]). Let

$$Z(\mathbf{x}) = 2^{-1} \{ (1 + N_2^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) - (1 + N_1^{-1})^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \}.$$

Then the  $Z$ -rule with a cutoff point  $c$  classifies  $\mathbf{x}$  as  $\Pi_1$  if  $Z(\mathbf{x}) > c$  and  $\Pi_2$  if  $Z(\mathbf{x}) < c$ . There are two types of probability of misclassification. One is the probability of allocating  $\mathbf{x}$  into  $\Pi_2$  even though it is actually belonging

to  $\Pi_1$ . The other is the probability that  $\mathbf{x}$  is classified as  $\Pi_1$  although it is actually belonging to  $\Pi_2$ . These two types of expected probabilities of misclassifications (EPMCs) for  $W$ - and  $Z$ - rules are expressed as

$$\begin{aligned} e_w(2|1) &= P(W(\mathbf{x}) < c | \mathbf{x} \in \Pi_1) \quad \text{and} \quad e_w(1|2) = P(W(\mathbf{x}) > c | \mathbf{x} \in \Pi_2), \\ e_z(2|1) &= P(Z(\mathbf{x}) < c | \mathbf{x} \in \Pi_1) \quad \text{and} \quad e_z(1|2) = P(Z(\mathbf{x}) > c | \mathbf{x} \in \Pi_2). \end{aligned}$$

It is easily shown that  $e_w(1|2)$  (or  $e_z(1|2)$ ) is obtained from  $e_w(2|1)$  (or  $e_z(2|1)$ ) by replacing  $(c, N_1, N_2)$  with  $(-c, N_2, N_1)$ . Thus, in this paper, we only deal with  $e_w(2|1)$  and  $e_z(2|1)$ .

Note that the EPMCs of  $W$ - and  $Z$ - rules are obtained from the distribution functions of  $W$  and  $Z$ . In general, it is hard to evaluate these expected probabilities of misclassification (EPMC) explicitly, but some asymptotic results including asymptotic expansions have been obtained. It is well known that the discriminant functions  $W(\mathbf{x})$  and  $Z(\mathbf{x})$  converges in distribution to the normal distributions, i.e.,  $W(\mathbf{x})$  and  $Z(\mathbf{x}) \xrightarrow{D} N((-1)^i \Delta^2/2, \Delta^2)$ , if  $\mathbf{x} \in \Pi_i$  under the asymptotic framework A0:

$$\text{A0: } N_1 \rightarrow \infty, N_2 \rightarrow \infty, N_1/N_2 \rightarrow \gamma \in (0, \infty), p \text{ is fixed.}$$

Here,  $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ . For a review of result under A0, see e.g., Anderson [1] and Fujikoshi et al. [3]. Generally, the precision of asymptotic approximations under A0 gets worth as the dimension  $p$  becomes large. As an alternative approach to overcome this shortcoming, it has been considered to derive asymptotic distributions of discriminant functions in a high-dimensional situation where  $n$  and  $p$  tend to infinity together. Fujikoshi and Seo [2] derived the limiting distribution of a general discriminant function for a class of discriminant rules which includes both the  $W$ - rule and  $Z$ - rule under asymptotic framework A1:

$$\text{A1: } p \rightarrow \infty, \quad N_1 \rightarrow \infty, \quad N_2 \rightarrow \infty, \quad p/n \rightarrow \gamma_0 \in [0, 1) \text{ and } N_1/N_2 \rightarrow \gamma \in (0, \infty).$$

Yamada et al. [4] obtained asymptotic expansions of the EPMCs for  $W$ - and  $Z$ - rules with the errors of order  $O(n^{-2})$  under the asymptotic framework A1. Making use of these expansions, we compare EER (Expected Error Rate) asymptotically in Section 2. It is shown that  $Z$ -rule has smaller EER than  $W$ -rule when the prior probabilities are the same, neglecting the terms of  $O(n^{-2})$ . We carried out simulation to compare EER, and gave the results in Section 3.

## 2. Main results

Firstly, we give results for asymptotic expansions of EPMCs for  $W$ - and  $Z$ -rules derived in Yamada et al. [4].

**Theorem 1.** Let  $y_w = v_w^{-1/2}(c + u_w)$ , where  $u_w = u_w(N_1, N_2, \Delta^2)$  and  $v_w = v_w(N_1, N_2, \Delta^2)$  are given as follows.

$$u_w = -\frac{1}{2} \frac{n}{m+1} \left\{ \Delta^2 - \left( \frac{p-2}{N_1} - \frac{p-2}{N_2} \right) \right\}, \quad v_w = \frac{n^2(n+1)}{(m+1)^2(m+2)} \left( \Delta^2 + \frac{N(p-2)}{N_1 N_2} \right).$$

Then, it holds that under A1,

$$P(W(\mathbf{x}) < c | \mathbf{x} \in \Pi_1) = \Phi(y_w) - \frac{1}{n} \sum_{k=1}^4 \ell_k H_{k-1}(y_w) \phi(y_w) + O(n^{-2}),$$

where  $\ell_k = \ell_k(N_1, N_2, \Delta^2)$  for  $k = 1, 2, 3, 4$  are given as follows.

$$\begin{aligned} \ell_1 &= -\frac{2}{\sqrt{v_w}} \frac{n}{m+1} \left\{ u_w + \frac{1}{2} \left( \frac{n}{N_1} - \frac{n}{N_2} \right) \right\}, \\ \ell_2 &= \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{\frac{Nn}{N_1 N_2}}{\Delta^2 + \frac{N(p-2)}{N_1 N_2}} + \frac{n}{2N} + \frac{1}{2v_w} \left[ \frac{2n}{m+1} u_w^2 \right. \\ &\quad \left. + \frac{n(p-2)}{2(m+1)^2} \left( \frac{n}{N_1} - \frac{n}{N_2} \right)^2 + \frac{n^2}{(m+1)^2} \frac{N_1 n}{N N_2} \Delta^2 + \frac{n^2(p-1)}{(m+1)^2(m+2)} \frac{N_2 n}{N N_1} \Delta^2 \right], \end{aligned}$$

$$\ell_3 = -\frac{1}{\sqrt{v_w}} \left\{ \frac{2n}{m+1} u_w + \frac{Nn}{N_1 N_2} \frac{u_w + \frac{1}{2} \frac{n}{m+1} \frac{N_2 - N_1}{N} \Delta^2}{\Delta^2 + \frac{N(p-2)}{N_1 N_2}} \right\},$$

$$\ell_4 = \frac{n}{m+1} + \frac{n(p-1)}{4(m+2)(n+1)} + \frac{1}{4} \frac{\frac{Nn}{N_1 N_2}}{\Delta^2 + \frac{N(p-2)}{N_1 N_2}} \left( 1 + \frac{\Delta^2}{\Delta^2 + \frac{N(p-2)}{N_1 N_2}} \right).$$

Here,  $H_k(\cdot)$  is  $k$ -th Hermite polynomial.

**Theorem 2.** Let  $y_z = v_z^{-1/2}(c + u_z)$ , where  $u_z = u_z(N_1, N_2, \Delta^2)$  and  $v_z = v_z(N_1, N_2, \Delta^2)$  are given as follows.

$$u_z = -\frac{n}{m+1} \sqrt{\frac{N}{N_1 N_2}} \omega_1^{-1} \omega_2^{-1} \Delta^2, \quad v_z = \frac{n^2(n+1)}{(m+1)^2(m+2)} \frac{N}{N_1 N_2} \omega_1^{-2} \{ \Delta^2 + (p-2)\omega_1^2 \},$$

where  $\omega_1 = \omega_1(N_1, N_2) = \sqrt{2\{(1 + N_2^{-1}) - a^{1/2}\}}$ ,  $\omega_2 = \omega_2(N_1, N_2) = \sqrt{2\{(1 + N_2^{-1}) + a^{1/2}\}}$ ,  $a = a(N_1, N_2) = (1 + N_2^{-1})/(1 + N_1^{-1})$ . Then, it holds that under A1,

$$P(Z(\mathbf{x}) < c^* | \mathbf{x} \in \Pi_1) = \Phi(y_z) - \frac{1}{n} \sum_{k=1}^4 \zeta_k H_{k-1}(y_z) \phi(y_z) + O(n^{-2}),$$

where  $c^* = \sqrt{(1 + N^{-1})/\{(1 + N_1^{-1})(1 + N_2^{-1})\}}c$ ,  $\zeta_k = \zeta_k(N_1, N_2, \Delta^2)$  for  $k = 1, 2, 3, 4$  are given as follows:

$$\zeta_1 = -\frac{2n}{m+1} \frac{u_z}{\sqrt{v_z}},$$

$$\zeta_2 = \frac{3n}{m+1} + \frac{n(p-1)}{(n+1)(m+2)} + \frac{n\omega_1^2}{\Delta^2 + (p-2)\omega_1^2} + \frac{1}{2v_z} \left\{ \frac{2n}{m+1} u_z^2 + \frac{n^2}{(m+1)^2} \frac{Nn}{N_1 N_2} \omega_2^{-2} \Delta^2 + \frac{n^2(p-1)}{(m+1)^2(m+2)} \frac{Nn}{N_1 N_2} \omega_2^{-2} \Delta^2 \right\},$$

$$\zeta_3 = -\frac{1}{\sqrt{v_z}} \left\{ \frac{2n}{m+1} u_z + \frac{n\omega_1^2 u_z}{\Delta^2 + (p-2)\omega_1^2} \right\},$$

$$\zeta_4 = \frac{n}{m+1} + \frac{n(p-1)}{4(m+2)(n+1)} + \frac{1}{4} \frac{n\omega_1^2}{\Delta^2 + (p-2)\omega_1^2} \left\{ 1 + \frac{\Delta^2}{\Delta^2 + (p-2)\omega_1^2} \right\}.$$

Let  $\pi_i$  be the prior probabilities of  $\mathbf{x}$  drawn from  $\Pi_i$  for  $i = 1, 2$ . Then, the expected error rate (EER) for  $W$ -rule with a cutoff point  $c_w$  and for  $Z$ -rule with a cutoff point  $c_z$  are expressed as

$$\text{EER}_w(c_w) = \pi_1 P(W(\mathbf{x}) < c_w | \mathbf{x} \in \Pi_1) + \pi_2 P(W(\mathbf{x}) > c_w | \mathbf{x} \in \Pi_2),$$

$$\text{EER}_z(c_z) = \pi_1 P(Z(\mathbf{x}) < c_z | \mathbf{x} \in \Pi_1) + \pi_2 P(Z(\mathbf{x}) > c_z | \mathbf{x} \in \Pi_2).$$

The following theorem is derived by using Theorem 1 and 2.

**Theorem 3.** Let  $\text{EER}_w(c_w)$  and  $\text{EER}_z(c_z)$  be the expected error rates of  $W$ -rule with a cutoff point  $c_w$  and  $Z$ -rule with a cutoff point  $c_z$ , respectively.

(1) The minimums of  $\lim_{A1} \text{EER}_w(c_w)$  and  $\lim_{A1} \text{EER}_z(c_z)$  are attained at  $c_w = c_{w,m}$  and  $c_w = c_{z,m}$ , respectively, where

$$c_{w,m} = \frac{1}{2} \frac{N}{N-p} \left[ \frac{p}{N_2} - \frac{p}{N_1} + 2 \frac{N}{N-p} \left( 1 + \frac{Np}{N_1 N_2} \frac{1}{\Delta^2} \right) \log \frac{\pi_2}{\pi_1} \right],$$

$$c_{z,m} = \left\{ \frac{1 + N^{-1}}{(1 + N_1^{-1})(1 + N_2^{-1})} \right\}^{-1/2} \left( \frac{N}{N-p} \right)^2 \left( 1 + \frac{Np}{N_1 N_2} \frac{1}{\Delta^2} \right) \log \frac{\pi_2}{\pi_1}.$$

(2) When  $\pi_1 = \pi_2$ , it holds that

$$\text{EER}_w(c_{w,m}) - \text{EER}_z(c_{z,m}) = -\frac{1}{4v_w} \frac{(n-1)(p-2)}{(m+1)^3} \left( \frac{n}{N_1} - \frac{n}{N_2} \right)^2 H_1(y_c) \phi(y_c) + O(n^{-2}),$$

where  $y_c = \{-(1/2)\Delta^2\} / \sqrt{\left\{ \Delta^2 + \frac{N(p-2)}{N_1 N_2} \right\} \frac{n+1}{m+2}}$ . Further, since  $H_1(y_c) = y_c < 0$ , we have that  $\text{EER}_z(0)$  is less than or equal to  $\text{EER}_w(c_{w,m})$ , neglecting the term of  $O(n^{-2})$ . When  $N_1 = N_2$ , the difference becomes  $O(n^{-2})$ .

### 3. Simulation result

In Table 1, we give the values of  $\text{EER}_w(0)$ ,  $\text{EER}_w(c_{w,m})$  and  $\text{EER}_z(0)$  obtained by simulation for the case  $\pi_1 = \pi_2$  obtained by simulation. We calculated these values when  $p = 8, 32$ ,  $(N_1, N_2) = (30, 10), (25, 15)$ ,  $\Delta = 1.05, 1.68, 2.56, 3.29$ , where the setting of  $\Delta$  is followed to Wyman et al. [5]. We can see a tendency from Table 1 that the magnitude of these error rate has the order “ $\text{EER}_z(0) < \text{EER}_w(c_{w,m}) < \text{EER}_w(0)$ ” for almost all simulation settings (The difference appears in less than 4th place of decimal point.).

Table 1: Comparison of EER

$(N_1, N_2)$	$\Delta$	$p = 8$			$p = 32$		
		$\text{EER}_w(0)$	$\text{EER}_w(c_{wm})$	$\text{EER}_z(0)$	$\text{EER}_w(0)$	$\text{EER}_w(c_{wm})$	$\text{EER}_z(0)$
(30, 10)	1.05	0.371	0.370	0.369	0.460	0.460	0.459
	1.69	0.262	0.262	0.261	0.410	0.410	0.409
	2.56	0.145	0.145	0.144	0.335	0.334	0.334
	3.29	0.082	0.081	0.081	0.276	0.276	0.275
(25, 15)	1.05	0.362	0.362	0.362	0.455	0.455	0.455
	1.69	0.255	0.255	0.254	0.403	0.403	0.403
	2.56	0.141	0.141	0.141	0.326	0.326	0.326
	3.29	0.079	0.079	0.079	0.269	0.269	0.269

### References

- [1] Anderson, T.W. (2003). *An Introduction of Multivariate Statistical Analysis 3rd ed.* Wiley, Hoboken.
- [2] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC’s of the linear and the quadratic discriminant functions when the sample sizes and the dimension are large. *Random Oper. Stochastic Equations*, **6**, 269–280.
- [3] Fujikoshi, Y., Ulyanov, V.V. and Shimizu, R. (2010). *Multivariate Statistics High-Dimensional and Large-Sample Approximations*, Wiley, Hoboken, NJ.
- [4] Yamada, T., Sakurai, T. and Fujikoshi, Y. (2017). High-dimensional asymptotic results for EPMCs of W- and Z- rules. TR 17-12, Hiroshima Statistical Research Group, Hiroshima University.
- [5] Wyman, F.J., Young, D.M. and Turner, D.W. (1990). A comparison of asymptotic error rate expansions for the sample linear discriminant function. *Pattern Recognition*, **23**, 775–783.

# LPRE CRITERION BASED ESTIMATING EQUATION APPROACHES FOR THE ERROR-IN-COVARIABLES

Qihua WANG

Institute of Applied Mathematics, AMSS, CAS, China

**Abstract:** In this paper, we propose two estimating equation based methods to estimate the regression parameter vector in the multiplicative regression model when a subset of covariates are subject to measurement error but replicate measurements of their surrogates are available. Both methods allow the number of replicate measurements to vary between subjects. No parametric assumption is imposed on the measurement error term and the true covariates which are not observed in the dataset. Under some regularity conditions, the asymptotic normality is proved for both the proposed estimators. Furthermore, a theoretical comparison is made for them in a special case where the distribution of the measurement error follows the normal distribution. Some simulation studies are conducted to assess the performances of the proposed methods. Real data analysis is used to illustrate our methods.

# AN INCOMPLETE-DATA FISHER SCORING

Keiji TAKAI

Faculty of Commerce and Business, Kansai University  
3-3-5, Yamate, Suita, Osaka 564-0073, Japan

---

## 1 Introduction

The EM algorithm (Dempster, Laird and Rubin 1977) is a standard method to compute the maximum likelihood estimates of incomplete-data models (models with missing data or/and latent variables). Although the EM algorithm is both stable and simple, it has major problems. One of the major problems with use of the EM algorithm is its slow convergence. Some authors have proposed variants to overcome this problem (McLachlan and Krishnan 2002).

In this talk, I propose an incomplete-data Fisher scoring method (IFS for short) to accelerate the convergence. Let  $\boldsymbol{\alpha}$  be the parameter of interest,  $\boldsymbol{\alpha}^{(t)}$  the current estimated value of  $\boldsymbol{\alpha}$ , and  $\boldsymbol{\alpha}^{(t+1)}$  the improved estimated value of  $\boldsymbol{\alpha}$ . The IFS has an iteration of the form,

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} + s \frac{1}{n} J_{\text{com}}(\boldsymbol{\alpha}^{(t)})^{-1} \nabla \ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)}) \quad (1)$$

for independent data of size  $n$ , where  $s$  is the steplength,  $J_{\text{com}}(\boldsymbol{\alpha}^{(t)})$  is the expected complete-data information matrix, evaluated at  $\boldsymbol{\alpha}^{(t)}$ , and  $\nabla \ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})$  is the first derivative of the observed-data log-likelihood function, evaluated at  $\boldsymbol{\alpha}^{(t)}$ . Note that  $J_{\text{com}}(\boldsymbol{\alpha}^{(t)})$  can be computed even in the presence of missing values.

The steplength adjustment in the IFS has three advantages over the EM algorithm. To clarify this point, I use the bivariate normal distribution data of sample size  $n$  with missing values. Let the mean and the variance be  $\boldsymbol{\mu}$  and  $\Sigma$ , respectively. The iterations for the EM algorithm can be transformed to

$$\begin{aligned} \boldsymbol{\mu}^{(t+1)} &= \boldsymbol{\mu}^{(t)} + \frac{1}{n} \Sigma^{(t)-1} \frac{\partial \ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})}{\partial \boldsymbol{\mu}}, \\ \Sigma^{(t+1)} &= \Sigma^{(t)} + \frac{1}{n} 2(\Sigma^{(t)-1} \otimes \Sigma^{(t)-1}) \frac{\partial \ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})}{\partial \text{vec}(\Sigma)}. \end{aligned}$$

Apparently, these two iterations are identified as equation (1) where  $s = 1$ . Note that the parameters  $\boldsymbol{\mu}$  and  $\Sigma$  are not the natural parameters of the exponential family. Thus, the first advantage of the IFS is that the EM algorithm and the IFS method produce an identical iteration for the unit steplength. For the other distributions and the other settings with latent variables as well, by adjusting the steplength, the EM algorithm and the IFS become identical to each other. It follows that when the steplength is greater than the certain length that makes the IFS identical to the EM algorithm, the convergence is expected to be faster than the EM algorithm, which is the second advantage. The third

advantage is that the IFS can make a smooth transition from the EM algorithm to the IFS with a faster convergence than the EM algorithm by only changing its steplength. This transition is a kind of “hybrid EM algorithm” which uses the faster convergence property of the EM algorithm during the initial stages and the faster convergence property of the IFS near the convergent point.

## 2 Derivation

In this section, we briefly show a derivation of the IFS. The basic idea is to mimic the sequence produced by the incomplete data with the complete data which are not actually available. If the complete data were available, the Fisher scoring method would provide its direction as  $\frac{1}{n}J_{\text{com}}(\boldsymbol{\alpha}^{(t)})^{-1}\nabla\ell_{\text{com}}(\boldsymbol{\alpha}^{(t)})$ , where  $\nabla\ell_{\text{com}}(\boldsymbol{\alpha}^{(t)})$  is the first derivative of the complete-data log-likelihood function. The Fisher scoring method is well known to provide a good result (Small and Wang 2003). However, in reality, the complete data (and thus  $\ell_{\text{com}}(\boldsymbol{\alpha}^{(t)})$ ) are not available. An alternative function to  $\ell_{\text{com}}(\boldsymbol{\alpha}^{(t)})$  is  $\ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})$ . I will approximate the sequence the Fisher scoring method produces for complete data by the actually available sequence, which a scaled version of  $\ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})$  produces. When the distance between variables  $\boldsymbol{x}$  and  $\boldsymbol{y}$  is measured with  $E[\|\boldsymbol{x} - \boldsymbol{y}\|^2]$ , the distance to be made as small as possible is

$$E \left[ \left\| \frac{1}{n}J_{\text{com}}(\boldsymbol{\alpha}^{(t)})^{-1}\nabla\ell_{\text{com}}(\boldsymbol{\alpha}^{(t)}) - B\nabla\ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)}) \right\|^2 \right]$$

with respect to matrix  $B$ . The solution is  $B = n^{-1}J_{\text{com}}(\boldsymbol{\alpha}^{(t)})^{-1}$ , which provides the direction of the IFS. This derivation is completely different from the principle of the EM algorithm. However, as I noted before, the sequence provided by the IFS becomes identical to that by the EM algorithm for a certain steplength. Thus, the IFS is closely related to the EM algorithm in this way. Note that the same iteration as equation (1) can be obtained through the Lower-bound algorithm (Böhning and Lindsay 1988) for  $\ell_{\text{obs}}(\boldsymbol{\alpha}^{(t)})$ .

## References

- Böhning, D. and Lindsay, B. (1988). Monotonicity of quadratic-approximation algorithms. *Annals of the Institute of Statistical Mathematics*, 40(4):641–663.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39:1–38.
- McLachlan, G., and Krishnan, T. (2002). The EM algorithm and extensions, 2nd Edition. Wiley.
- Small, C. G. and Wang, J. (2003). Numerical methods for nonlinear estimating equations (Oxford Statistical Science Series 29). Oxford University Press.



# A GENERALIZED EMS ALGORITHM FOR MODEL SELECTION WITH INCOMPLETE DATA

Ping-Feng XU<sup>1</sup>, Man-Lai TANG<sup>2</sup>, Na SHAN<sup>3</sup>, Guo-Liang TIAN<sup>4</sup>, and Lai-Xu SHANG<sup>1</sup>

<sup>1</sup>Department of Statistics, Changchun University of Technology, Changchun 130012, China

<sup>2</sup>Department of Mathematics and Statistics, Hang Seng Management College, Hong Kong

<sup>3</sup>School of Psychology, Northeast Normal University, Changchun 130024, China

<sup>4</sup>Department of Mathematics, Southern University of Science and Technology, Shenzhen, 518055, China

## Abstract

The EMS algorithm is a useful method for model selection with missing data. It performs E-step (Expectation step) and MS-step (Model Selection Step) alternately to find the minimum point of the observed generalized information criteria (GIC). However, sometimes it may not be numerically feasible to perform the MS-step, especially for high dimensional settings. In this paper, we seek only a decrease in the observed generalized information criteria in the MS-step. The resulting method is called a generalized EMS (GEMS) algorithm, which includes the EMS algorithm as a special case. We obtain several numerical convergence results of the GEMS algorithm. A useful special case is that all limit points of the EMS algorithm satisfy a necessary condition of the minimum points of the observed GIC under very weak conditions. We apply the GEMS algorithm for Gaussian graphical model selection and variable selection with missing data and compare with state of the art methods via numerical experiments.

# BIAS REDUCTION OF ML ESTIMATORS OF GAMMA DISTRIBUTION PARAMETERS AND COMPARISON WITH OTHER METHODS

Yumina KODAIRA, Ko ABE, Takenori SAKUMURA, and Toshinari KAMAKURA  
Graduate School of Science and Engineering, Chuo University  
Tokyo 112-8551, Japan

---

## 1. Introduction

The gamma distribution, denoted as  $\text{gam}(\alpha, \beta)$ , is a two-parameter distribution with probability density function,

$$f_{\text{gam}}(x) = \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\beta}\right), \quad x > 0, \quad (1)$$

where  $\alpha > 0$  is the shape parameter,  $\beta > 0$  is the scale parameter and  $\Gamma(\cdot)$  is the gamma function. The gamma distribution is a useful model in many field of applications.

Although the maximum likelihood (ML) method is the most popular parameter estimation method, there are no closed-form expressions for the ML estimators for the two-parameter gamma distribution. Therefore the numerical calculations with iteration are generally used.

The moment estimators of the two gamma parameters have closed-forms, they are not efficient under either small samples or large samples (Ye and Chen (2017)). Ye and Chen (2017) proposed the new method using the three-parameter generalized gamma distribution.

The three-parameter generalized gamma distribution is denoted as  $\text{gg}(\alpha, \beta, \gamma)$  where  $\gamma$  is a power parameter. It is a extension of the gamma distribution with probability density function

$$f_{\text{gg}}(x) = \frac{\gamma x^{\alpha\gamma-1}}{\beta^{\alpha\gamma} \Gamma(\alpha)} \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right], \quad x > 0. \quad (2)$$

The generalized gamma distribution can be obtained by a power transformation of gamma: if  $X \sim \text{gam}(\alpha, \beta^\gamma)$ , then  $X^{1/\gamma} \sim \text{gg}(\alpha, \beta, \gamma)$ . This distribution was proposed by Stacy (1962). This is a flexible model that contains the gamma, Weibull and log-normal distributions as special cases.

Ye and Chen (2017) proposed the method as follows. In the generalized gamma distribution, the likelihood equations of three parameters can be solved explicitly for  $\alpha$  and  $\beta$  when  $\gamma = 1$ . The solutions of this likelihood equations can be regarded as approximate on ML estimators for two-parameter gamma distribution. Here we call this estimators the Ye and Chen's estimators.

In this study, we investigate the properties of the Ye and Chen's estimators and propose the method to correct bias of these estimators. Also, we compare with the method of moments (MM) estimators considering the feature that their likelihood equations can be solved explicitly.

## 2. Model

We compare the ML estimators obtained by Newton's method, Ye and Chen's estimators and MM estimators. It is known that when the sample size is small, the ML estimators have large biases. Cox and Snell (1968) gave the procedure to correct the first-order biases of ML estimators in the general frameworks. This procedure was developed by Cox *et al.* (1968) and Cordeiro *et al.* (1994). We correct both biases of ML estimator and Ye and Chen's estimator calculating the asymptotic bias. We confirmed the asymptotic biases are the same as the biases

derived by Johnson (2011). Ye and Chen (2017) also proposed the bias correction of their new estimator, then we also compare these results with ML estimators and Ye and Chen's estimators.

## 2.1 Maximum Likelihood Estimators

The most popular parameter estimation method is the maximum likelihood (ML) method, but there are no closed-form expressions for the ML estimators for the two-parameter gamma distribution. Therefore we use numerical calculation by Newton's method.

## 2.2 Ye and Chen's Estimators

Let  $X_i \sim \text{gam}(\alpha, \beta^\gamma)$ ,  $i = 1, 2, \dots, n$  and  $X_1, X_2, \dots, X_n$  be in iid. The parameters  $\alpha$  and  $\beta$  are needed to estimate. Here, let us pretend that  $X$  follows the above generalized gamma distribution with unknown  $\gamma$ . Then, the log-likelihood function based on the observed  $X_1, X_2, \dots, X_n$  is

$$l_{\text{gg}}(\alpha, \beta, \gamma) = \log\gamma - \alpha\gamma\log\beta - \log\Gamma(\alpha) + \frac{1}{n} \sum_{i=1}^n \left[ (\alpha\gamma - 1)\log X_i - \left(\frac{X_i}{\beta}\right)^\gamma \right].$$

The likelihood equations are obtained by taking the partial derivatives of  $l_{\text{gg}}$  with respect to  $\alpha, \beta$  and  $\gamma$ , respectively

$$\frac{\partial l_{\text{gg}}(\alpha, \beta, \gamma)}{\partial \alpha} = -\psi(\alpha) - \gamma\log\beta + \frac{\gamma}{n} \sum_{i=1}^n \log X_i, \quad (3)$$

$$\frac{\partial l_{\text{gg}}(\alpha, \beta, \gamma)}{\partial \beta} = -\alpha + \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i}{\beta}\right)^\gamma, \quad (4)$$

$$\frac{\partial l_{\text{gg}}(\alpha, \beta, \gamma)}{\partial \gamma} = \frac{1}{\gamma} + \frac{\alpha}{n} \sum_{i=1}^n \log\left(\frac{X_i}{\beta}\right) - \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i}{\beta}\right)^\gamma \log\left(\frac{X_i}{\beta}\right), \quad (5)$$

where  $\psi(\cdot) = d\log\Gamma(x)/dx$  is the digamma function. Setting these equal to zero and solving the resulting system of equations gives the ML estimators of  $(\alpha, \beta, \gamma)$ . In particular, by setting (4) equal to zero, we can express  $\beta$  as a function of  $\alpha$  and  $\beta$  :

$$\beta(\alpha, \beta) = \left(\frac{\sum X_i^\gamma}{n\alpha}\right)^{1/\gamma}. \quad (6)$$

Substitute the above display into (5) to give

$$\alpha(\gamma) = \frac{n \sum X_i^\gamma}{n\gamma \sum X_i^\gamma \log X_i - \gamma \sum \log X_i \sum X_i^\gamma}. \quad (7)$$

Now, we consider the gamma distribution. If  $\gamma = 1$ , generalized gamma can be regarded as gamma. Use this fact in (6) and (7) to obtain the Ye and Chen's estimators for  $\alpha$  and  $\beta$  as

$$\tilde{\alpha}_Y = \frac{n \sum X_i}{n \sum X_i \log X_i - \sum \log X_i \sum X_i}, \quad (8)$$

$$\tilde{\beta}_Y = \frac{1}{n^2} (n \sum X_i \log X_i - \sum \log X_i \sum X_i). \quad (9)$$

Next, Ye and Chen's bias correction method is shown as follows. The estimator for the scale parameter  $\beta$  is

$$\tilde{\beta}_Y^\dagger = \frac{n}{n-1} \tilde{\beta}_Y = \frac{1}{n(n-1)} \left( n \sum X_i \log X_i - \sum \log X_i \sum X_i \right). \quad (10)$$

While the estimator for  $1/\alpha$  is

$$\tilde{\alpha}_Y^{-1\dagger} = \frac{n}{n-1} \tilde{\alpha}_Y^{-1} = \frac{n \sum X_i \log X_i - \sum \log X_i \sum X_i}{(n-1) \sum X_i}. \quad (11)$$

See this detailed proof on Ye and Chen (2017).

### 2.3 First-Order Bias Correction

The method of bias correction is provided by Cox *et al.* (1968) and Cordeiro *et al.* (1994). Actually we calculate the first-order bias, obtaining the following result.

$$E(\hat{\alpha}_{ML} - \alpha) = \frac{-2 + \alpha\psi(1, \alpha) - \alpha^2\psi(2, \alpha)}{2n\{-1 + \alpha\psi(1, \alpha)\}^2} + O\left(\frac{1}{n^2}\right), \quad (12)$$

$$E(\hat{\beta}_{ML} - \beta) = \frac{\beta\{\psi(1, \alpha) + \alpha\psi(2, \alpha)\}}{2n\{-1 + \alpha\psi(1, \alpha)\}^2} + O\left(\frac{1}{n^2}\right). \quad (13)$$

This result corresponds to that of Johnson's (2011).

### 2.4 Proposed estimator

We propose the estimators that corrects the Ye and Chen's estimators using bias correction given by (12) and (13), respectively. Here, we express these estimators  $\hat{\alpha}_Y^*$ ,  $\hat{\alpha}_Y^*$ .

## 3. Result

Table1 shows the mean of each estimators and Table2 shows the MSE. According to Ye and Chen (2017), they defined their estimators  $\hat{\alpha}_Y^\dagger$ ,  $\hat{\beta}_Y^\dagger$  are unbiased, but under small samples we find biases. These biases can be reduced by using the method of Cox *et al.* (1968) and Cordeiro *et al.* (1994). Especially, our proposed estimators can be expressed explicitly, the estimators can be closed-form.

## References

- Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of mathematical statistics*, 1187-1192.
- Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 248-275.
- Cordeiro, G. M., & Klein, R. (1994). Bias correction in ARMA models. *Statistics & Probability Letters*, **19**(3), 169-176.
- Johnson Jr, P. H., Qi, Y., & Chueh, Y. (2011). Bias-corrected maximum likelihood estimation in actuarial science. In Proceedings of 46th Actuarial Research Conference.
- Ye, Z. S., & Chen, N. (2017). Closed-Form Estimators for the Gamma Distribution Derived from Likelihood Equations. *The American Statistician*, **71**(2), 177-181.

Table 1: mean of estimators

$\alpha$	$n$	MLE				Ye and Chen's estimator				Proposed		MME	
		$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\alpha}_{ML}^*$	$\hat{\beta}_{ML}^*$	$\bar{\alpha}_Y$	$\bar{\beta}_Y$	$\bar{\alpha}_Y^\dagger$	$\bar{\beta}_Y^\dagger$	$\alpha_{Y^*}$	$\beta_{Y^*}$	$\alpha_{MM}$	$\beta_{MM}$
1	3	9.45	0.68	0.19	0.89	9.46	0.66	6.31	1.00	0.21	0.87	6.59	0.74
	5	2.23	0.82	1.00	0.97	2.25	0.80	1.80	1.01	1.02	0.95	2.08	0.83
	7	1.62	0.87	1.00	0.98	1.64	0.86	1.40	1.00	1.02	0.97	1.65	0.87
	9	1.40	0.91	0.99	1.00	1.41	0.90	1.26	1.01	1.01	0.99	1.47	0.91
	10	1.35	0.91	1.00	0.99	1.37	0.90	1.23	1.00	1.02	0.98	1.43	0.91
	20	1.15	0.95	1.01	0.99	1.16	0.95	1.10	0.99	1.01	0.99	1.21	0.95
	30	1.09	0.97	1.00	1.00	1.10	0.97	1.06	1.00	1.01	1.00	1.15	0.96
	40	1.06	0.98	1.00	1.00	1.07	0.98	1.04	1.00	1.00	1.00	1.11	0.97
5	3	47.97	0.68	0.21	0.90	47.98	0.67	31.98	1.01	0.22	0.89	32.22	0.90
	5	7.18	0.81	3.00	0.96	7.19	0.80	5.75	1.00	3.01	0.96	6.02	0.94
	7	5.00	0.86	2.95	0.98	5.01	0.86	4.30	1.00	2.96	0.98	4.54	0.95
	9	4.36	0.89	2.98	0.99	4.37	0.89	3.89	1.00	2.99	0.99	4.12	0.97
	10	4.20	0.90	3.01	0.99	4.22	0.90	3.80	1.00	3.02	0.99	4.01	0.97
	20	3.50	0.95	3.01	0.99	3.51	0.94	3.33	0.99	3.02	0.99	3.47	0.98
	30	3.31	0.97	3.00	1.00	3.32	0.97	3.21	1.00	3.01	1.00	3.30	0.99
	40	3.22	0.98	2.99	1.00	3.22	0.98	3.14	1.00	3.00	1.00	3.22	0.99
7	3	127.69	0.68	0.22	0.91	127.71	0.68	85.14	1.02	0.23	0.91	85.39	0.97
	5	17.15	0.80	6.99	0.96	17.16	0.80	13.73	1.00	7.00	0.96	13.99	0.98
	7	12.01	0.86	6.95	0.98	12.02	0.86	10.30	1.00	6.97	0.98	10.54	0.98
	9	10.42	0.89	7.02	0.99	10.43	0.89	9.27	1.00	7.03	0.99	9.50	0.99
	10	9.79	0.91	6.92	1.00	9.80	0.91	8.82	1.01	6.93	1.00	9.02	0.99
	20	8.20	0.95	7.00	1.00	8.20	0.95	7.79	1.00	7.01	1.00	7.93	0.99
	30	7.77	0.97	7.02	1.00	7.78	0.97	7.52	1.00	7.02	1.00	7.61	0.99
	40	7.55	0.98	7.00	1.00	7.55	0.98	7.36	1.00	7.00	1.00	7.44	1.00

\* the first-order bias correction  
 † Ye and Chen's bias correction

Table 2: MSE

$\alpha$	$n$	MLE				Ye and Chen's estimator				Proposed		MME	
		$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\alpha}_{ML}^*$	$\hat{\beta}_{ML}^*$	$\bar{\alpha}_Y$	$\bar{\beta}_Y$	$\bar{\alpha}_Y^\dagger$	$\bar{\beta}_Y^\dagger$	$\alpha_{Y^*}$	$\beta_{Y^*}$	$\alpha_{MM}$	$\beta_{MM}$
1	3	$5.37 \times 10^3$	0.72	0.65	1.03	$5.37 \times 10^3$	0.68	$2.38 \times 10^3$	1.28	0.69	0.96	$2.38 \times 10^3$	0.69
	5	20.37	0.48	3.03	0.61	20.35	0.47	12.66	0.67	3.02	0.59	13.32	0.53
	7	2.57	0.34	0.72	0.41	2.61	0.34	1.78	0.43	0.74	0.40	2.30	0.43
	9	0.92	0.27	0.34	0.32	0.94	0.28	0.68	0.34	0.36	0.32	1.06	0.39
	10	0.69	0.25	0.28	0.29	0.71	0.25	0.52	0.30	0.29	0.28	0.85	0.35
	20	0.16	0.12	0.10	0.13	0.17	0.13	0.14	0.14	0.10	0.13	0.26	0.20
	30	0.08	0.08	0.06	0.09	0.09	0.09	0.08	0.09	0.06	0.09	0.16	0.14
	40	0.05	0.06	0.04	0.06	0.05	0.06	0.05	0.07	0.04	0.07	0.11	0.11
5	3	$8.50 \times 10^5$	0.61	7.77	0.89	$8.50 \times 10^5$	0.60	$3.78 \times 10^5$	1.11	7.99	0.87	$3.78 \times 10^5$	0.84
	5	229.37	0.39	33.92	0.50	228.74	0.39	142.73	0.55	33.67	0.50	143.59	0.53
	7	23.41	0.28	6.35	0.33	23.40	0.28	15.89	0.35	6.33	0.34	16.99	0.38
	9	10.17	0.22	3.71	0.26	10.28	0.23	7.42	0.27	3.77	0.26	8.55	0.31
	10	8.01	0.20	3.22	0.23	8.09	0.21	5.98	0.24	3.26	0.24	6.90	0.28
	20	1.70	0.10	1.05	0.11	1.73	0.10	1.44	0.11	1.07	0.11	1.85	0.14
	30	0.89	0.07	0.64	0.07	0.91	0.07	0.80	0.08	0.66	0.07	1.07	0.10
	40	0.59	0.05	0.47	0.06	0.60	0.05	0.55	0.06	0.48	0.06	0.73	0.07
7	3	$6.01 \times 10^6$	0.61	45.99	0.91	$6.01 \times 10^6$	0.61	$2.69 \times 10^6$	1.14	46.71	0.90	$2.69 \times 10^6$	0.98
	5	1614.86	0.38	241.92	0.49	1612.60	0.38	1011.27	0.53	241.03	0.49	1011.55	0.52
	7	138.88	0.27	37.17	0.33	139.08	0.27	94.58	0.34	37.28	0.33	97.59	0.35
	9	65.92	0.22	24.10	0.25	66.04	0.22	48.04	0.26	24.16	0.25	50.22	0.28
	10	43.26	0.19	17.40	0.22	43.39	0.19	32.11	0.23	17.48	0.22	34.05	0.25
	20	9.65	0.10	5.94	0.11	9.73	0.10	8.11	0.11	6.00	0.11	9.09	0.12
	30	5.24	0.07	3.76	0.07	5.29	0.07	4.65	0.07	3.80	0.07	5.27	0.08
	40	3.42	0.05	2.67	0.05	3.45	0.05	3.12	0.05	2.70	0.05	3.56	0.06

\* the first-order bias correction  
 † Ye and Chen's bias correction

# Random Search Algorithm for Optimal Mixture Experimental Design

Chongqi ZHANG\_ and Guanghui LI

([cqzhang@gzhu.edu.cn](mailto:cqzhang@gzhu.edu.cn))

School of Economics and Statistics, Guangzhou University

**Abstract:** It is well known that it is difficult to obtain accurate optimal design for mixture experimental design with complex constraints. In this paper, we construct a random search algorithm which can be used to find the optimal design for mixture model with complex constraints. First, we generate an initial set by the Monte-Carlo method, and then run the random search algorithm to get the optimal set of points. After that, we explain the effectiveness of this method by using two examples. Finally, we discuss some applications of this random search algorithm in other models.

**Keywords:** Information matrix, Optimal design, Random search Random mixture points.

# Session IV

November 2, 2017 (Thursday)

15:50-17:30

**Chair:**

Qihua Wang, Chinese Academic of Science

**Speaker:**

Sari Hosoya, Kanto Gakuin University

Feiyu Jiang, Tsinghua University

Atsuhiko Hayashi, Nagoya Institute of Technology

Wei Li, Peking University

Fumio Ishioka, Okayama University

# MEASURING AND ANALYZING INTERCULTURAL COMPETENCE

Sari HOSOYA

College of Economics, Kanto Gakuin University, Japan

---

## 1. Introduction

Due to the mobility of people and interdependence in global world, our society now is characterized with diversity more than before. Although people have learned a lot from people who belong to the society with different values and knowledge, we seem to be reluctant to accept such diversities nowadays. While there are people who enjoy differences and benefit from them, others attempt to reject and exclude people with different background. It seems that people needs a certain kind of competence to live positively in the society with increased diversity. Intercultural competence and sensitivity towards diversity are needed in order to function in multicultural contexts, and this competence can be obtained through education.

In the field of education, statistical analysis is often used for analysing quantitative data, such as test scores. However, very often we also need to analyze qualitative data such as the ones referring to human quality. This research is one example of our attempt to measure and analyze intercultural competence, one of the qualitative data.

## 2. Intercultural Competence

According to UNESCO definition, Intercultural Competences are explained as follows: Having adequate relevant knowledge about particular cultures, as well as general knowledge about the sorts of issues arising when members of different cultures interact, that encourage establishing and maintaining contact with diverse others as well as having the skills required to draw upon both knowledge and attitudes when interacting with others from different cultures. (UNESCO, 2013, p.16)



Suppose we need to cultivate such competence through the process of education, it is important that we can somehow assess and evaluate such abilities. In the process of measuring the competence, we need to clarify the elements which consist this particular competence. Drawing upon the theories of intercultural/multicultural education and intercultural communication, I attempt to collect characteristics of those who seem to function well in the diverse society. In this presentation I will introduce the characteristics and will focus how qualitative data can be transformed to measurable data in order to operationalize the analysis.

## References

Hosoya, Sari; Talib, Mirja; Arslan, Hasan (2014) Finnish, Japanese and Turkish pre-service teachers' intercultural competence: The impact of pre-service teachers' culture, personal experience and education, in Leoncio Vega (Ed.) *Empires, Post-Coloniality and Interculturality: New Challenges for Comparative Education*, Rotterdam: Sense Publishers. pp. 235-250.

Hosoya, Sari., & Talib, Mirja. (2010). "Pre-service teachers' intercultural competence: Japan and Finland", in D. Mattheou (Ed.), *Changing educational landscape: Educational practice, schooling systems and higher education — a comparative perspective*. New York: Springer. pp.241-260.

UNESCO. (2013). *Intercultural Competences*.

<http://unesdoc.unesco.org/images/0021/002197/219768e.pdf>

**ESTIMATION AND INFERENCE IN DOUBLE AUTOREGRESSIVE MODELS  
WITH NULL VOLATILITY COEFFICIENTS**

Feiyu Jiang, Dong Li, Ke Zhu

Center for Statistical Science and Department of Industrial Engineering,

Tsinghua University, Beijing 100084, P.R. China

Department of Statistics and Actuarial Science, University of Hong Kong,

Hong Kong, P.R. China

**Abstract:** This paper considers the self-weighted quasi-maximum likelihood estimation (QMLE) of the double autoregressive (DAR) model with null volatility coefficients and derives the asymptotic distribution under the mild conditions, which is a projection of a normal vector onto a convex cone by a quadratic approximation. A practical implication of this result concerns the estimation of over-identified DAR models. Further, testing the nullity of volatility coefficients are also studied. This problem is nonstandard since the self-weighted QMLE is subject to positivity constraints. Three classical tests, i.e., the Wald, Lagrange multiplier and quasi-likelihood ratio tests, are studied as well as their asymptotic null and local alternative distributions. Under fixed alternatives, efficiency performance are investigated. Two cases of special interest are tests of the null hypothesis of one coefficient equal to zero and tests of the null

hypothesis of no conditional heteroscedasticity. Monte Carlo simulation studies are carried out to assess the performance of the self-weighted QMLE and three tests in finite samples. Finally, an empirical example is given for illustrating the usefulness of our proposed approach.

# A COMPARISON STUDY OF RULE SPACE METHOD AND NEURAL NETWORK MODEL FOR LEARNING DIAGNOSIS AND IT'S AN APPLICATION

Atsuhiko HAYASHI

Nagoya Institute of Technology  
Nagoya, Aichi, 466-8555, Japan

## 1. Introduction

Both methods, Rule Space Method (RSM) and Neural Network Model (NNM) are techniques of statistical pattern recognition and classification approaches developed from different fields; one is for behavioral and the other is for neural sciences.

RSM is a technique of clustering examinees into one of the predetermined latent Knowledge States (KS) that are derived logically from an expert's hypotheses about how students learn. RSM uses the multivariate decision theory to classify individuals, and NNM that is considered as a nonlinear regression method uses the middle layer of the network structure as classification results. We have found that there two methods are similarities between the results from the two approaches, and moreover they have complementary characteristics when applied in practice.

In this paper, we discuss the comparisons of both approaches by focusing on the structure of the NNM and of KSs in the RSM. And we show an application result for a reasoning test.

## 2. Rule Space Method

RSM is a technique developed in the domain of the cognitive science. It starts from the use of an incidence matrix  $Q$  that characterizes the underlying cognitive processes and knowledge (Attribute) involved in each Item. It is a grasping method of each examinee's mastered/non-mastered learning level (Knowledge State, KS) from item response patterns, and a list of all the possible KSs can be generated algorithmically by applying Boolean Algebra to the incidence matrix  $Q$ . This method is fairly new but has lately started getting some attention because it is possible to provide diagnostic scoring reports for a large-scale assessment.

Up to now, the results of examinees' performance on a test are reported by total scores or scaled scores. However, if this technique is used in educational practices, it is possible to report which attributes each student mastered or non-mastered, in addition to his/her total scores. It is often true that the same total score may have several different KSs. By reporting detailed information of his/her KS, learning can be facilitated more effectively than just providing total scores only.

### **3. Feed-Forward Neural Network Model**

In spite of that the mathematical formulization of the Feed-Forward NNM is simple, almost any nonlinear function can be approximated by selecting deferent numbers of middle layers and connections between neurons. When we apply this technique to existing data obtained from learning processes, we can use this model to search for the strategy of any joint intensity between units.

From a statistical point of view, NNM is a nonlinear regression model. In this paper Feed-Forward NNM is considered as a model-fitting procedure to estimate the optimum values of the parameters in the regression model.

This procedure is called parameter estimation in statistics, but is called a learning algorithm in NNM. One of the learning algorithms commonly used is Back Propagation (BP) that is a learning method by passing on errors to previous layers. BP is an adaptation of the steepest descent method to the NNM field. This method has a reducible faculty of the convergence to the local minimum point.

### **4. Science Reasoning Test**

The Science Reasoning Test (SR-Test) is an entrance examination test that measures the student's interpretation, analysis, evaluation, reasoning, and problem-solving skills required in the natural sciences.

Since we got the ACT's (American College Testing, Inc.) cooperation, we used one open-form of their ACT Assessment tests for our experimentation. The test is based on units containing scientific information and a set of multiple choice questions about the scientific information. Calculators are not permitted to be used for the test. The scientific information for the test is provided in one of three types of formats.

The first format, data representation, presents graphic and tabular material similar to that found in science journals and texts. The questions associated with these format measure skills such as graph reading, interpretation of scatter plots, and interpretation of information presented in tables. The second format, research summaries, provides students with descriptions of one or more related experiments. The questions focus upon the design of experiments and interpretation of experimental results. The third format, conflicting viewpoints, presents students with expressions of several hypotheses or views that, being based on differing premises or on incomplete data, are inconsistent with one another. The questions focus upon the understanding, analysis, and comparison of alternative viewpoints or hypotheses.

The SR-Test questions require students to use scientific reasoning to answer the questions. The students are required to recognize and understand the basic features of, and concepts related to, the provided information; to critically examine the relationships between the information provided and the conclusions drawn or hypotheses developed; and to generalize from given information to gain new information, draw conclusions, or make predictions.

## 5. Numerical Examples

We applied the RSM to a data of fraction addition problems, and got a tree structure for the KS. We related RSM that derives the KS from an incidence matrix  $Q$ , to the Feed-Forward NNM. For that, we designed the network of the three-layer structure in which items were assigned to the input layer and Attributes were to the output layer. The KSs in the RSM were considered to correspond to the middle layers of NNM. We applied several numerical examples to the both methods, and found close similarities in their results although they were not identical.

And we applied the RSM to a data of SR-Test of 286 Japanese students. The number of attributes and items are 12 and 18, respectively. Figure 1 is the tree representation of the KSs that shows the examinee's mastered/non-mastered learning level. In this figure, each circle is the KS, and the numbers in the circle are the IDs of non-mastered attribute. Or the number in the parenthesis is the number of examinee classified in this KS. We can find the fact that the main solving attribute IDs are 6, 8 and 9, and secondary attribute are 2 and 5. The total examinee classified in these KSs is 225, which is about 80% of all. The main streams to reach the full mastered state are three KSs of left-hand side in the third layer from the top.

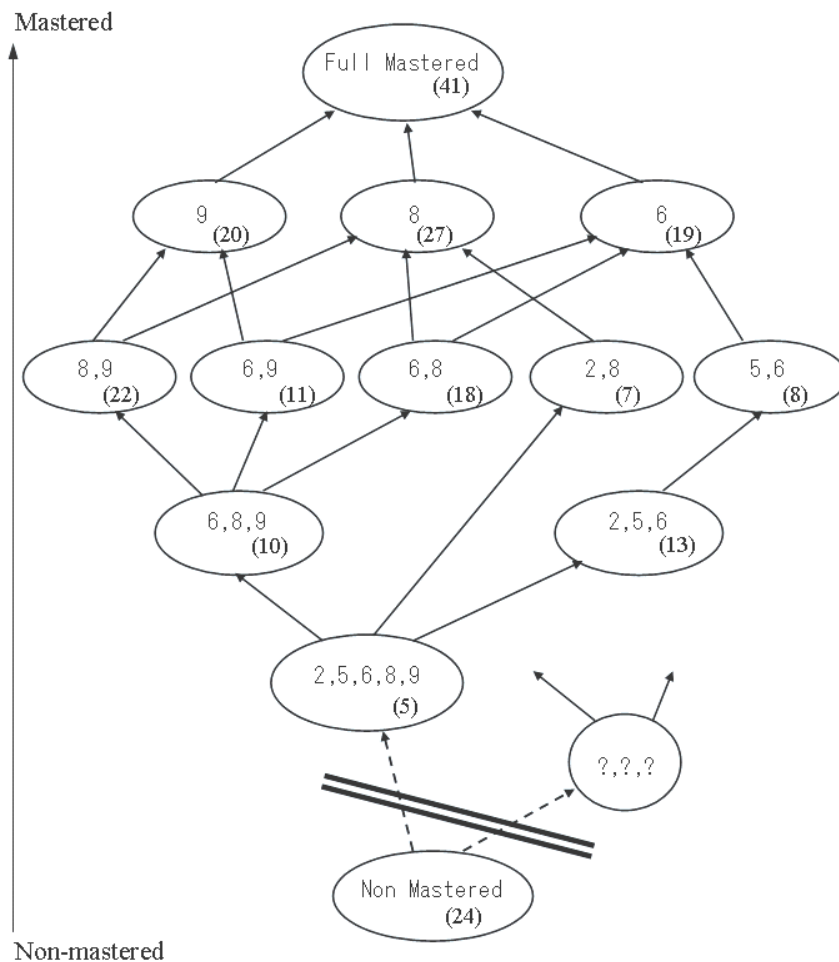


Figure 1. A tree representation of Knowledge States for the SR-Test data

## 6. Discussion and Conclusions

We investigated the relationship between the characteristics of the middle layer of NNM and the Knowledge States in the RSM, and discussed their similarities and usefulness at the weaknesses existing in the RSM.

It is well known that the composition of an incidence matrix  $Q$  in the RSM is a very laborious task, requires experts' intense cooperation. The experts identify attributes involved in each item and express them in an incidence matrix  $Q$ . It needs to investigate multiple numbers of solution strategies for each item. This is extremely hard work. If an examinee's mastering level (cluster) is known to some extent from past experiences, it is also possible to construct a network in which these clusters are assigned to the output layer of NNM. The middle layer drawn from this model is expected to correspond to Attributes. It may be possible to use this result for replacing a task analysis required in making an incidence matrix  $Q$  in RSM.

We plan to clarify the difference and similarities of the two models with numerical examples, or will apply to this problem the technique of the deep learning in the AI fields.

## References

- [1] James Maxey. (2000). *Introduction to the ACT Assessment, International comparison study for university entrance examinations*, M.Huzii et al. (Eds.), 42-55.
- [2] Kikumi K. Tatsuoka. (1995). *Architecture of Knowledge Structures and Cognitive Diagnosis: A Statistical Pattern Recognition and Classification Approach*, Paul D.Nichols et al. (Eds.), *Cognitively Diagnostic Assessment*, 327-359, Lawrence Erlbaum Associates.
- [3] Kikumi K. Tatsuoka and Atsuhiko Hayashi. (2001), *Statistical method for individual cognitive diagnosis based on latent knowledge state*, *Journal of The Society of Instrument and Control Engineers*, Vol.40, No.8, 561-567, (in Japanese).
- [4] Kikumi K. Tatsuoka(2009), *Cognitive Assessment: An Introduction to the Rule Space Method*, *Multivariate Applications Series*, Routledge.

**Acknowledgement** : We are grateful to Dr. James Maxey and ACT that gave us their permission of one open-form in their ACT Assessment tests for our experimentation. And this work was supported by JSPS KAKENHI Grant Number JP15K12169.

(e-mail: hayashi.atsuhiko@nitech.ac.jp)

# IDENTIFIABILITY AND ESTIMATION OF CAUSAL MEDIATION EFFECTS WITH MISSING DATA

Wei Li

College of mathematics and statistics, Huazhong Normal University

**Abstract:** Mediation analysis is a standard approach to understanding how and why an intervention works in social and medical sciences. However, the presence of missing data, especially missing not at random data, poses a great challenge for the applicability of this approach in practice. Current methods for handling such missingness are still lacking in causal mediation analysis. In this article, we first show the identifiability of causal mediation effects with different types of missing outcomes under different missingness mechanisms. We then provide corresponding approaches for estimation and inference. Especially for missing not at random data, we develop an estimating-equation-based approach to estimate causal mediation effects, which can easily handle different types of mediators and outcomes, and we also establish the asymptotic results of the estimators. Simulation results show good performance for the proposed estimators in finite samples. Finally, we use a real data set from the Clinical Antipsychotic Trials of Intervention Effectiveness research for Alzheimer's Disease to illustrate our approach.



# VISUALIZATION FOR RADIATION MONITORING POST DATA USING SPATIAL INTERPOLATION

Fumio ISHIOKA and Koji KURIHARA

The Graduate School of Environmental and Life Science, Okayama University  
Okayama 7008530, Japan

---

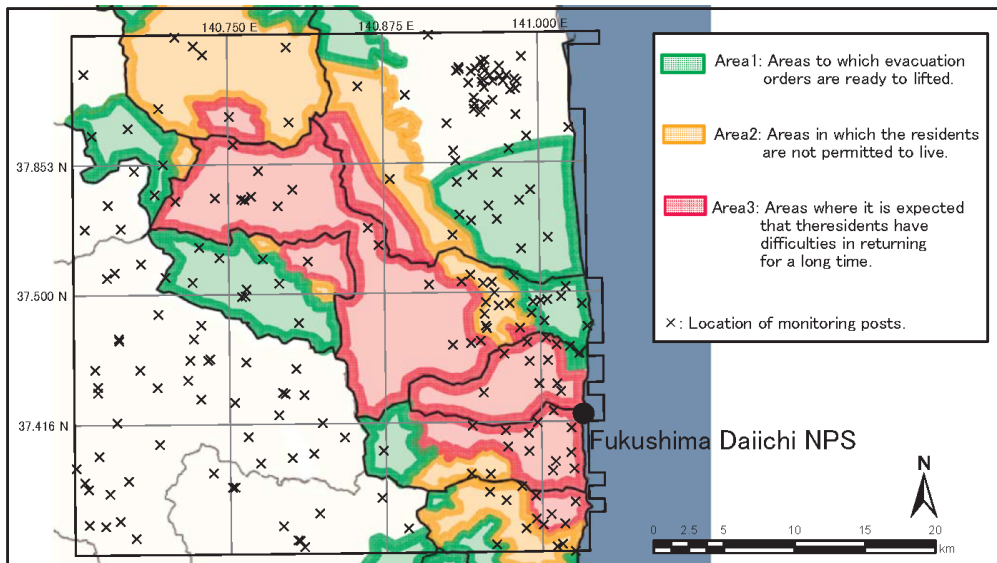
## 1. Introduction

On March 11, 2011, a massive amount of radioactive material was released from Tokyo Electric Power Company's Fukushima Daiichi Nuclear Power Station (NPS). This accident caused serious damage to both economic and social development, as well as caused a wide range of problems in the environment and in food production. Although six years have passed since the accident, there is still intense concern about the influence of radioactive contamination, and high air dose rates and high concentrations of radionuclides are still found in certain areas around the NPS [2]. The Nuclear Regulation Authority (NRA), Japan, inaugurated the Comprehensive Radiation Monitoring Plan on August 2, 2011 [1]. One aim of this project is to analyze the distribution of radiation doses and radioactive materials, principally in residential areas, on a mid- to long-term basis. Related ministries and agencies, local government bodies, nuclear operators, and related companies have been conducting radiation monitoring in a cooperative manner. In this study, we focused on air radiation dose rate data measuring by monitoring posts. These monitoring posts are equipped with air radiation dose rate measuring devices at fixed locations, with data obtained at 1 m above the ground. Approximately 4,400 monitors are installed in Japan as of October 1, 2017, and the Fukushima Prefecture accounts for over 80% more. Data is logged every ten minutes and stored continuously. The data is openly available at NRA web site. However, it is difficult to monitor areal trends in the radiation pollution, because the monitoring posts are very sparsely scattered. In this study, we apply a technique of spatial interpolation for compensating for the small number of observations and attempt to visualize the spatial distribution of air dose rate.

## 2. Materials

The study area chosen has the highest level of radiation contamination in Fukushima Prefecture. We used regularly arranged 10 km × 10 km meshes designed by the administrative organ covering the range within 37.333 degrees north latitude, 140.625 degrees east longitude, 37.667 degrees north latitude, and 141.050 degrees east longitude, containing most of the three levels of the evacuation area that the Japanese government has recognized. In addition, we set January 10, 2013 as the study period. We removed some false data caused by instrument anomalies, as announced on the NRA website, and finally the total number of monitoring posts we used was 212 at that time. Figure 1 shows the study area, the location of the Fukushima Daiichi NPS and the monitoring posts on January 10, 2013. Mean daily air dose rates with 10-min data aggregated into daily intervals were

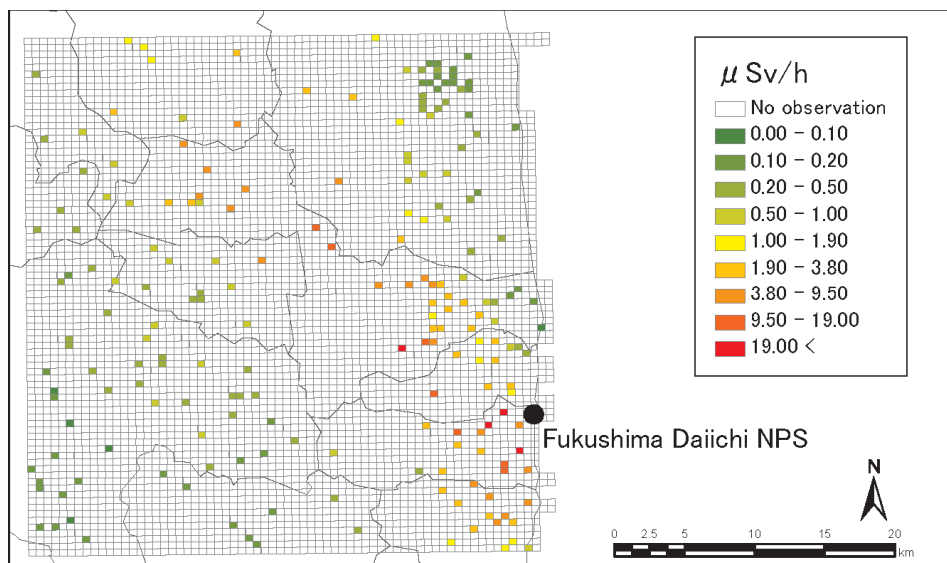
used as the analysis object in this study.



**Figure 1.** Study area divided most of the three levels of the evacuation area into 10 km × 10 km meshes, with the Fukushima Daiichi NPS and the location of each monitoring post on January 10, 2013.

### 3. Spatial interpolation

It is difficult to see the trend of the distribution of the air dose rate under these conditions, because the monitoring posts of the study area are very sparsely scattered. Accordingly, we attempt to increase the location of the analysis object by using the spatial interpolation, i.e., ordinary kriging. First, we redivided the study area into smaller 500 m × 500 m meshes and assigned an air dose rate value to each of the mesh based on the corresponding post location. If two or more posts were installed in one mesh, the mesh was assigned their mean value. Figure 2 shows the division of data into the 5,440 meshes and the assigned values of air dose rate.



**Figure 2.** Study area redivided into 500 m × 500 m meshes and the assigned value of air dose rate.

To improve kriging precision, the observations were normalized using a Box-cox transformation ( $\lambda = -0.18$ ), and we used a model based on an anisotropic variogram. That is, by assuming a geometric anisotropy, we constructed isotropic spatial processes by performing coordinate conversion based on the rotation angle of the coordinates and the anisotropy ratio. To estimate a distance between two arbitrary points without directly measuring the distances between the points (the center of gravity position of each mesh), the experimental variogram is assumed, and the theoretical variogram is applied to it. The experimental variogram is defined as

$$\gamma^*(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (z(\mathbf{x}_i) - z(\mathbf{x}_j))^2$$

where  $\gamma^*(\mathbf{h})$  is dissimilarity depending on the intercentral distance  $\|\mathbf{h}\|$  of each mesh,  $z(\mathbf{x}_i)$  and  $z(\mathbf{x}_j)$  are sample data at two points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Moreover,  $N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i - \mathbf{x}_j = \mathbf{h} \text{ for } i, j = 1, 2, \dots, n\}$  is the set of all pairs of points with  $\mathbf{h}$ , and  $|N(\mathbf{h})|$  is the number of pairs in  $N(\mathbf{h})$ . We must fit a variogram function  $\gamma$  to an experimental variogram  $\gamma^*$ , i.e., replace an experimental variogram with a theoretical variogram that is a suitable valid function. In this study, we attempted to apply the Matérn model which defined by the following parametric model,

$$\gamma(\mathbf{h}|\theta_0, \theta_1, \theta_2) = \theta_0 + \theta_1 \left( 1 - \frac{1}{2^{k-1}\Gamma(k)} \left( \frac{\|\mathbf{h}\|}{\theta_2} \right)^k K_k \left( \frac{\|\mathbf{h}\|}{\theta_2} \right) \right)$$

with the smoothness parameter  $k$  varying from 0 to  $\infty$ , gamma function  $\Gamma(\cdot)$ , and modified Bessel function  $K_k(\cdot)$ . Furthermore, the three parameters,  $\theta_0$ ,  $\theta_1$ , and  $\theta_2$  relate to *nugget*, *sill* and *range*, respectively, i.e., the nugget  $\theta_0$  is defined by  $\gamma(\mathbf{h})$  as  $\|\mathbf{h}\| = 0$ , the sill  $\theta_0 + \theta_1$  is the value  $\gamma(\infty) = \lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h})$ , and the range  $\theta_2$  is the distance at which the  $\gamma(\mathbf{h})$  exceeds the sill value for the first time. To estimate these parameters, we used likelihood-based parameter estimation methods, viz., maximum likelihood (ML) and restricted maximum likelihood (REML). These methods can be used with Gaussian random fields, and we normalized our data using the Box-cox transformation. ML and REML are available as an R-package `gEO` [3] from the statistical software R. Table 1 shows the estimated parameters and AIC values for ML and REML in isotropic and anisotropic situations. Hence, taking the models with the lowest AIC yields the Matérn model, whose parameters, estimated using the REML method in the anisotropic situation, have the “best” fit.

		$\theta_0$	$\theta_1$	$\theta_2$	$k$	Rotation angle	Anisotropy ratio	AIC
ML	Isotropic	0.16	3.34	21.45	0.98	—	—	236.0
	Anisotropic	0.13	2.09	16.75	0.85	149.89	3.57	212.3
REML	Isotropic	0.15	5.46	32.46	0.90	—	—	228.6
	Anisotropic	0.13	3.46	26.78	0.78	149.94	3.52	204.9

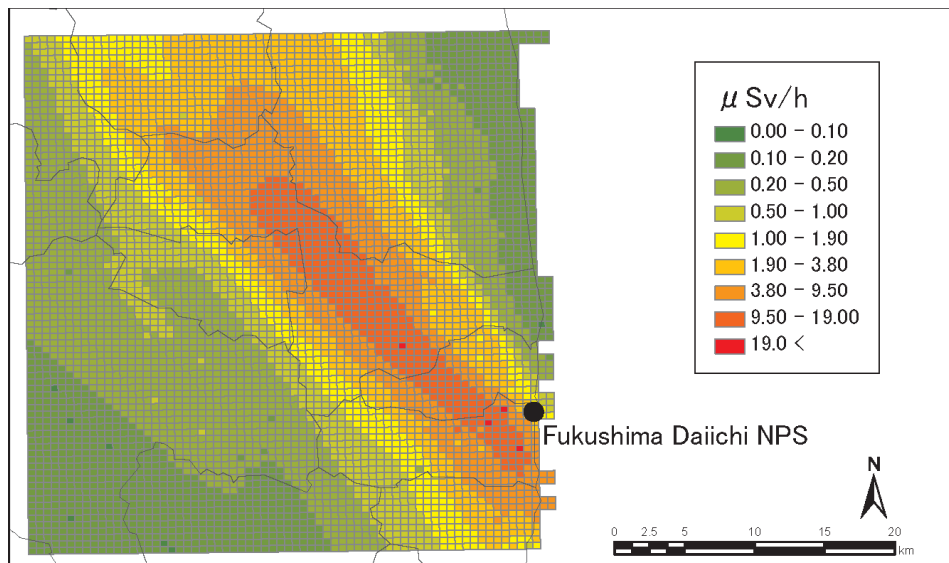
**Table 1.** Estimated parameters using ML or REML in isotropic and anisotropic situations, and their respective AICs.

In this study, we implemented an ordinary kriging based on intrinsically stationary and isotropy assumptions, which is often used as the kriging method, to predict each air dose rate value for 500 m  $\times$  500 m meshes. The ordinary kriging predictor  $Z^*(\mathbf{x}_0)$  of the value at  $\mathbf{x}_0$  is given by the linear

combination of  $Z(\mathbf{x})$  evaluated at each sample  $\mathbf{x}_i; i = 1, 2, \dots, n$ .

$$Z^*(\mathbf{x}_0) = \sum_{i=1}^n \omega_i Z(\mathbf{x}_i)$$

where  $\omega_i, i = 1, 2, \dots, n$  provides the unknown weights corresponding to the influence of the variable  $Z(\mathbf{x})$ . We can obtain the dissimilarity value between the point  $\mathbf{x}_0$  and the  $i$ -th observed point by using the estimated variogram model. Under the restriction conditions of  $\sum_{i=1}^n \omega_i = 1$ , the computation of  $Z^*(\mathbf{x}_0)$  is conducted using the Lagrange multiplier. The predicted map is shown in the Figure 3.



**Figure 3.** The predicted map of air dose rate obtained by ordinary kriging at January 10, 2013.

#### 4. Conclusion and Future work

This article has presented the radiation pollution map using the spatial interpolation technique based on the monitoring post data of air dose rate, which installed sparsely. We assumed the intrinsically stationary, and performed coordinate transformation for the isotropic spatial processes. The locations with high dose rate did not spread in concentric fashion from the NPS but the direction of northwest. However, the around of extremely high rate mesh might be estimated lower than actual rate as a possible influence. About a strict validation of the stationary in our situation needs more discussion. In the future work, we would like to apply these techniques for some different days and investigate the temporal and spatial trends of radiation pollution.

#### References

- [1] Nuclear Regulation Authority, Japan. (2015). Comprehensive Radiation Monitoring Plan. <http://radioactivity.nsr.go.jp/en/>.
- [2] METI Measures and Requests in response to the Great East Japan Earthquake. Ministry of Economy, Trade and industry. <http://www.meti.go.jp/english/>.
- [3] Ribeiro Jr. P.J. and Diggle, P.J. (2016) geoR: A package for geostatistical data analysis using the R software. Version 1.7-5.2. <http://www.leg.ufpr.br/geoR/>

# Session V

November 3, 2017 (Friday)

08:30 - 10:10

**Chair:**

Kazunori Yamaguchi, Rikkyo University

**Speaker:**

Yoshihiro Sone, Chuo University

Wensheng Zhu, Northeastern Normal University

Hiroko Katayama, Okayama University of Science

Qian Lin, Tsinghua University

Michio Yamamoto, Okayama University

# STATISTICAL GAIT ANALYSIS BASED ON THE DATA SET FROM GLASSES INSTALLED WITH IMU

Yoshihiro SONE and Toshinari KAMAKURA

Graduate School of Science and Engineering, Chuo University  
Tokyo 112-8551, Japan

---

## 1. Introduction

Walking is indispensable for human beings to behave activity in daily life. Then, gait analysis is so important to measure the activities. Gait analysis is usually based on video images or the motion capture, but it requires not only the expensive facilities also vast calculations assuming bionic model. With recent technical innovations, we can use wearable small devices installed with the gravity accelated sensors and gyro sensor.

To enable precise detection of physical states of human in daily life, we suggest to measure the body axis in gait. The body axis of human being can be known as the balance and the power of the body. In this study, we propose a system for measuring gait with body axis using glasses type device with IMU (which combined with three-axis accelerometer and three-axis gyroscope). With this system, we could know the person's gait is normal or not easily.

## 2. Methods

### 2.1 Signal Processing

The raw data have many noises. As a preprocessing for analysis, we should remove the noises. On this time, raw data is converted by fourier transform to remove the noises (unnecessary information) as Low Pass Filter(LPF).

#### 2.1.1 Fourier Transform

The fourier transform decomposes a function of time (a signal) into the frequencies that constitute it. The fourier transform of a function of time is a complex-valued function of frequency. The Fourier transform is given in the following expressions.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt,$$

where  $\omega$  can express frequency in angular frequencies then with  $\omega = 2\pi f$ .

There is also inverse transform. The inverse transform is given in the following expressions.

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} dt.$$

### 2.2 Logistic Regression

The logistic regression model is used to predict binary (0/1) response: normal or avnormal walking. The *logit* fuction can be written as following form.

$$\text{logit}(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right) = \alpha + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i},$$
$$(i = 1, \dots, n),$$

let  $x_i$  be the explanatory variable. The regression coefficients  $\alpha$  and  $\beta$  are estimated using maximum likelihood estimation.

### 3. Experiment

We used JINS MEME to measure the gait. It is a measurement device with IMU. This device is a pair of glasses so that easy to bring and to use. The IMU mounted in the eye-wear measures the body motion. As the sensor located near the head is on the top of the body axis, this eye-wear is suitable to measure movements or shift of center of gravity during physical exercises with a high precision. We take normal walking 3 times per person. Subject's face is forward to its front. The abnormal experiments were conducted for three subjects as follows: walking with smartphone used (the abnormal walking A), dragging a foot (the abnormal walking B) and watching somewhere (the abnormal walking C). Each walking is about 10 steps, and for confirmation, we also used the video to take movies for recording.

### 4. Results and Discussions

The results of analyzing data processed by logistic regression are shown in the following table,

Table 1: classification table

Prediction \ Observation	Normal	Abnormal
Normal	8	1
Abnormal	1	8

From the above results, we were able to discriminate whether walking is abnormal or not. The normal walking and the abnormal walking B are quite different, but we cannot tell the abnormal walking C from the normal walking. That is the reason of misclassification. The result from the head moving data set are comparatively good considering that the error rate is so small in our experiment.

### 5. Conclusion

In this study, we used the eye-wear device with IMU for gait analysis and get good results. We must develop the system to be able to analyze the streaming data in real time in the future.

### References

- Press, S. J., & Wilson, S. (1978). *Choosing between logistic regression and discriminant analysis*. Journal of the American Statistical Association, **73** (364), 699-705.
- JINS MEME ES (2017). <https://jins-meme.com/ja/products/es/>.

# COVARIATE-ADJUSTED ASSOCIATION STUDIES BASED ON MULTIPLE TESTING UNDER DEPENDENCE

Tingting CUI, Pengfei WANG, and Wensheng ZHU\*

School of Mathematics and Statistics, Northeast Normal University

Changchun 130024, China

## Abstract

It is more and more important to consider the dependence structure among multiple testings, especially for the genome-wide association studies (GWAS). Sun and Cai (JRSSB, 2009) proposed a large-scale multiple testing procedure to test HMM-dependent hypotheses under the framework of compound decision-theory, which was successfully applied to GWAS by Wei et al. (Bioinformatics, 2009). However, the fact that the etiology of complex diseases not only with respect to the genetic effect, but also the environmental factors. Failure to account for the covariates can produce misleading bias of the association of interest, or suffer from loss of testing efficiency. In this paper, we develop a covariate-adjusted multiple testing procedure, called covariate-adjusted local index of significance (CALIS), to account for the effects of environmental factors in GWAS via a double-chain HMM. Theoretic results show that our CALIS can control the false discovery rate (FDR) at the nominal level and has the smallest false negative rate (FNR) among all valid FDR procedures. We further demonstrate the advantage of our CALIS over the procedure LIS of Sun and Cai (2009) by simulation studies and a real data analysis.



# ITEM SELECTION FOR IMPRESSION SURVEY

Hiroko KATAYAMA<sup>1)</sup> and Yuichi MORI<sup>2)</sup>

1) Graduate School of Informatics, Okayama University of Science

2) Okayama University of Science

Okayama 700-0005, Japan

---

## 1 Introduction

There are many surveys that research impressions such as brand image, global attitude and university ranking. We sometime meet problems in which we wish to reduce the number of items (questions/variables) to save time and cost or select a reasonable subset of questions to find core items in the survey. For such cases, in general, because collected data in such surveys consist of categorical variables, we consider item (variable) selection for qualitative data.

To select a subset of categorical variables, we can consider the following two approaches: first is a variable selection in nonlinear principal component analysis (NLPCA), which has been proposed by Mori et al. (2017). This is a selection method for mixed measurement level data and then can be easily applied to categorical data in surveys; second is to select a subset by using Item Response Theory (IRT). Since there are several studies that analyze characteristics of items and latent abilities of respondents in impression survey by IRT as all responses are converted to binary data (0, 1), we may select a subset so as to keep information on items or responses estimated by IRT as much as possible. So we will propose item selection using IRT.

In this paper we apply these two selection methods to survey data and demonstrate the performance of the methods in numerical examples.

## 2 Item selection methods for impression survey

### 2.1 Variable selection in nonlinear PCA

Mori et al. (2017) proposed variable selection in NLPCA for mixed measurement level data. This method combines NLPCA using the alternating least squares with optimal sealing (e.g., Young et al., 1978; Gifi, 1990) and Modified PCA (M.PCA, Tanaka and Mori, 1997), that is, the eigenvalue problem (EVP) in ordinary NLPCA is replaced by that in M.PCA.

Suppose we have an  $n \times p$  data matrix  $\mathbf{Y}$  (that consists of categorical variables). Let  $\mathbf{Y}$  be decomposed into an  $n \times q$  submatrix  $\mathbf{Y}_1$  and  $n \times (p - q)$  submatrix  $\mathbf{Y}_2$  ( $1 \leq q \leq p$ ), and the covariance matrix of  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$  is denoted as  $\begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$ . M.PCA is intended to derive  $r$  principal components (PCs)  $\mathbf{Z}$  that are computed using  $\mathbf{Y}_1$  and represent all variables as much as possible, i.e.,  $\mathbf{Z} = \mathbf{Y}_1 \mathbf{A}$  ( $1 \leq r \leq q$ ), where  $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_r)$  is derived by an EVP,

$[(S_{11}^2 + S_{12}S_{21}) - \lambda S_{11}] \mathbf{a} = 0$ , so as to maximize the prediction efficiency based on Rao (1964) (criterion 1) or the  $RV$ -coefficient based on Robert and Escoufier (1976) (criterion 2). This naturally leads to specifying the best subset of  $q$  variables. Here we implement this M.PCA selection procedure to NLPCA algorithm, PRINCIPALS (Young et al., 1978) as follows: for a given initial data matrix  $\mathbf{Y}^{*(0)} = (\mathbf{Y}_1^{*(0)}, \mathbf{Y}_2^{*(0)})$ , where the super script  $*$  denotes an appropriately quantified data matrix and  $(t)$  is the  $t$ -th iteration, the following two steps are iterated until convergence:

- *Model estimation step*: From  $\mathbf{Y}^{*(t)} = (\mathbf{Y}_1^{*(t)}, \mathbf{Y}_2^{*(t)})$ , obtain  $\mathbf{A}^{(t)}$  by solving the above EVP and compute  $\mathbf{Z}^{(t)}$  from  $\mathbf{Z}^{(t)} = \mathbf{Y}_1^{*(t)} \mathbf{A}^{(t)}$ .
- *Optimal scaling step*: Calculate  $\hat{\mathbf{Y}}^{(t+1)} = \mathbf{Z}^{(t)} \mathbf{A}^{(t)T}$  using  $\mathbf{A}^{(t)}$  and  $\mathbf{Z}^{(t)}$  found in *Mode estimation step* and find  $\mathbf{Y}^{*(t+1)}$  such that  $\mathbf{Y}^{*(t+1)} = \text{argmin tr}(\mathbf{Y}^{*(t)} - \hat{\mathbf{Y}}^{(t+1)})^T (\mathbf{Y}^{*(t)} - \hat{\mathbf{Y}}^{(t+1)})$  for fixed  $\hat{\mathbf{Y}}^{(t+1)}$ .

$\mathbf{Y}^* = (\mathbf{Y}_1^*, \mathbf{Y}_2^*)$  obtained after convenience is an optimally scaled matrix of  $\mathbf{Y}$  and variables corresponding to  $\mathbf{Y}_1^*$  are the best subsets of size  $q$ . We can apply this variable selection procedure, so called nonlinear M.PCA (NL-M.PCA), to impression survey data to select reasonable items.

## 2.2 Item selection in IRT

There are some studies in which IRT is applied to impression survey data. For test data we ordinarily select items from an item pool, but there is no item pool for impression survey because we cannot easily make many different types of questions about the same image. Therefore we have to use collected responses in each survey to select a subset of items. Here we focus on latent abilities of respondents estimated by IRT, that is, we select a subset of items which represents information on latent abilities estimated from all items as much as possible. There are two possibilities as selection criteria: latent ability scores and rank of the scores.

To estimate the latent ability of the  $i$ -th respondent ( $1 \leq i \leq n$ ),  $\theta_i$ , we use the 2-Parameter Logistic IRT model,  $P_j(\theta_i) = \frac{1}{1 + \exp[-1.7a_j(\theta_i - b_j)]}$ , where  $P_j(\theta_i)$  is the probability of correct response for the  $j$ -th item on  $\theta_i$  ( $1 \leq j \leq p$ ), and  $a_j$  item discrimination and  $b_j$  item difficulty of the  $j$ -th item. Let  $\theta_{i(p)}$  and  $\theta_{i(q)}$  denote the latent ability based on all  $p$  items and  $q$  items selected among  $p$  items, respectively. We can find the best subset of  $q$  items among all possible  ${}_p C_q$  subsets, whose information on  $\theta_{i(q)}$ 's is mostly closed to that on  $\theta_{i(p)}$ 's. We here measure the closeness between  $\theta_{i(p)}$  and  $\theta_{i(q)}$  by the difference of scores,  $d_q = \sum_{i=1}^n (\theta_{i(p)} - \theta_{i(q)})^2$ , and the rank correlation of scores,  $\rho_q = \rho(\theta_{i(p)}, \theta_{i(q)})$  where  $\rho(\cdot)$  is the rank correlation coefficient. We call this item selection in IRT.

We find a reasonable subset which has the smallest  $d_q$  or the largest  $\rho_q$ .

### 3 Numerical examples

We have two types of data: first one is “Global Attitudes and Trends in Spring 2015” (Pew Research Center, 2015) which collects information on social issues, public opinion and other trends from 40 countries. We extracted 800 individuals and 27 questions with 4 to 7 levels from the original data; second is “Department Image survey” (Mori et al., 2014) which consists of 258 university students and 12 questions asking their images on curriculum and support of the department (all items have four levels but they were transformed to binary data so as to be analyzed by IRT).

We apply NL-M.PCA to Global Attitude data because this data have multiple levels and are suitable for analysis by NL-M.PCA. Table 1 is the result of variable selection using criterion 1 with  $r=5$  by backward elimination. Fig1. shows the change of proportion  $P$ . The results illustrate that the proportion changes slightly until the number of items is 12. When  $q=21$ , for example, 21 items except for items {8, 9, 10, 12, 14, 18} are suggested to be selected, because they have the closest information on the proportion to all items (0.6273 for selected 21 items and 0.6337 for all items). The result suggests that these items which have a relation to daily life, climate, economics and political situation can be selected a reasonable subset of size 21.

Table 1 Selection results (NL-M.PCA, Global Attitude data, Backward,  $r=5$ , Proportion  $P$ )

$q$	$Y_1 \mid Y_2$																										$P$	
27	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	0.6337
26	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	10	0.6334
25	1	2	3	4	5	6	7	8	9	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	10	12	0.6325
24	1	2	3	4	5	6	7	9	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	8	10	12	0.6314
23	1	2	3	4	5	6	7	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	8	9	10	12	0.6304
22	1	2	3	4	5	6	7	11	13	14	15	16	17	19	20	21	22	23	24	25	26	27	8	9	10	12	18	0.6290
21	1	2	3	4	5	6	7	11	13	15	16	17	19	20	21	22	23	24	25	26	27	8	9	10	12	14	18	0.6273
20	1	2	3	4	5	7	11	13	15	16	17	19	20	21	22	23	24	25	26	27	6	8	9	10	12	14	18	0.6251
19	1	2	3	4	5	7	11	13	15	17	19	20	21	22	23	24	25	26	27	6	8	9	10	12	14	16	18	0.6227
18	1	2	3	4	5	7	11	15	17	19	20	21	22	23	24	25	26	27	6	8	9	10	12	13	14	16	18	0.6201
17	1	2	3	4	5	7	11	15	17	19	21	22	23	24	25	26	27	6	8	9	10	12	13	14	16	18	20	0.6172
16	1	2	3	5	7	11	15	17	19	21	22	23	24	25	26	27	4	6	8	9	10	12	13	14	16	18	20	0.6141
15	1	2	3	5	7	11	15	17	19	21	22	23	24	25	26	4	6	8	9	10	12	13	14	16	18	20	27	0.6108
14	1	2	3	5	7	11	15	17	19	22	23	24	25	26	4	6	8	9	10	12	13	14	16	18	20	21	27	0.6074
13	2	3	5	7	11	15	17	19	22	23	24	25	26	1	4	6	8	9	10	12	13	14	16	18	20	21	27	0.6019
12	2	3	5	7	11	15	17	19	23	24	25	26	1	4	6	8	9	10	12	13	14	16	18	20	21	22	27	0.5962
11	2	3	5	7	11	15	19	23	24	25	26	1	4	6	8	9	10	12	13	14	16	17	18	20	21	22	27	0.5881
10	2	5	7	11	15	19	23	24	25	26	1	3	4	6	8	9	10	12	13	14	16	17	18	20	21	22	27	0.5793
9	2	5	7	11	15	19	23	24	25	1	3	4	6	8	9	10	12	13	14	16	17	18	20	21	22	26	27	0.5703
8	2	5	11	15	19	23	24	25	1	3	4	6	7	8	9	10	12	13	14	16	17	18	20	21	22	26	27	0.5577
7	2	5	11	19	23	24	25	1	3	4	6	7	8	9	10	12	13	14	15	16	17	18	20	21	22	26	27	0.5328
6	2	5	11	19	23	24	1	3	4	6	7	8	9	10	12	13	14	15	16	17	18	20	21	22	25	26	27	0.5062
5	2	5	11	19	24	1	3	4	6	7	8	9	10	12	13	14	15	16	17	18	20	21	22	23	25	26	27	0.4377

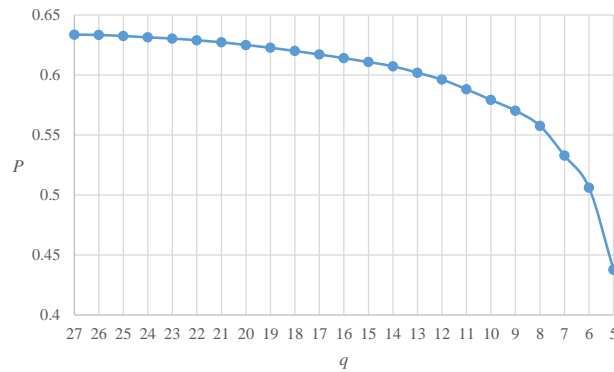


Fig 1. Change of  $P$

Table 2 Selection results (Item selection in IRT, Department Image data, Backward, left:  $d_q$ , right:  $\rho_q$ )

$q$	$Y_1$   $Y_2$												$d_q$	$q$	$Y_1$   $Y_2$												$\rho_q$
12	1	2	3	4	5	6	7	8	9	10	11	12	-	12	1	2	3	4	5	6	7	8	9	10	11	12	-
11	1	2	3	4	5	6	7	8	10	11	12	9	1.6077	11	1	2	3	4	5	6	7	8	10	11	12	9	0.9954
10	1	2	3	4	5	6	7	8	10	12	9	3.8940	10	1	2	3	4	5	6	7	8	10	12	9	11	0.9883	
9	2	3	4	5	6	7	8	10	12	1	9	11	7.1607	9	1	2	3	4	5	6	7	10	12	8	9	11	0.9773
8	2	3	4	5	6	7	10	12	1	8	9	11	10.6146	8	2	3	4	5	6	7	10	12	1	8	9	11	0.9659
7	2	3	4	5	6	7	12	1	8	9	10	11	14.9702	7	2	3	4	5	6	7	12	1	8	9	10	11	0.9541
6	3	4	5	6	7	12	1	2	8	9	10	11	19.4241	6	2	3	5	6	7	12	1	4	8	9	10	11	0.9361
5	3	4	5	6	12	1	2	7	8	9	10	11	25.3764	5	2	3	5	6	12	1	4	7	8	9	10	11	0.9109
4	3	4	5	6	1	2	7	8	9	10	11	12	35.0666	4	2	3	6	12	1	4	5	7	8	9	10	11	0.8736
3	3	5	6	1	2	4	7	8	9	10	11	12	46.9073	3	3	6	12	1	2	4	5	7	8	9	10	11	0.8239
2	3	6	1	2	4	5	7	8	9	10	11	12	68.0698	2	3	6	1	2	4	5	7	8	9	10	11	12	0.7706

Next, we apply item selection in IRT to Department Image data. The left table in Table 2 shows the selection results based on  $d_q$  and the right  $\rho_q$ . Both criteria selected similar items. When  $q=8$ , for example, items {2, 3, 4, 5, 6, 7, 10, 12} are selected which are questions about curriculum in the department while deleted four items are about departmental support although one item is about curriculum.

#### 4 Concluding remarks

We proposed two selection methods of item selection for impression survey data. One is a variable selection in nonlinear PCA with M.PCA criteria. We applied the method to data which have multiple levels. We obtained reasonable results of item selection from impression survey data with multiple levels. Other is an item selection in IRT that we proposed from the idea of analysis of impression survey on university department by IRT. We applied this method to binary response data and selected items based on respondent ability and its rank. The results suggested which items are selected in the context of respondent ability.

Future problems are to interpret the results in detail, to evaluate the performance in more real examples, and consider reasonable recommendations on how to use the methods.

#### References

- Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley, Chichester, England.
- Mori, Y., Kuroda, M., Matsuda, S. (2014). An analysis of images of Department of Socio Information by using data-mining methods. *Journal of Socio Information Studies* **13**, 87-101.
- Mori, Y., Kuroda, M., Makino, N. (2017). Variable selection in nonlinear principal component analysis. *Nonlinear Principal Component Analysis and Its Applications (JSS Research Series in Statistics)*, 31-45, Springer.
- Pew Research Center (2015). Global Attitude Spring 2015 Survey data (40-Nation survey conducted March 25 – May 27, 2015). <http://www.pewglobal.org/dataset/spring-2015-survey-data/> (2017.7.19)
- Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya* **A26**, 329-358.
- Robert, P., Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Appl. Statist* **25**, 257-65.
- Tanaka, Y., Mori, Y. (1997). Principal component analysis based on a subset of variables: Variable selection and sensitivity analysis, *American Journal of Mathematics and Management Sciences* **17** (1&2), 61-89.
- Young, F.W., Takane, Y., de Leeuw, J. (1978). Principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features. *Psychometrika* **43**, 279-281.

# ON THE OPTIMAL DETECTION BOUNDARY OF SINGLE INDEX MODELS

Qian LIN

Center for Statistical Science and Department of Industrial Engineering,

Tsinghua University, Beijing 100084, P.R. China

**Abstract:** Inspired by our previous work in sliced inverse regression, we have advocated that single/multiple index models could be studied in depth as the linear regression. In this talk we will report a further support evidence, i.e., we show that the minimal level of the generalized signal to noise ratio defined for index models matches that of SNR for the linear regression.

# CLUSTERING OF MULTIVARIATE CATEGORICAL DATA VIA PENALIZED LATENT CLASS ANALYSIS WITH DIMENSION REDUCTION

Michio YAMAMOTO

Department of Environmental and Mathematical Sciences, Okayama University  
Okayama, Japan

## 1 Introduction

Multivariate categorical data are commonly observed and analyzed in many scientific application fields: behavioral and social research, biosciences, and document classification. One of the purposes of analyzing categorical data, as well as continuous data, is the partitioning objects that have categorical features into several unpredefined homogeneous groups (i.e., clusters). In clustering objects with many variables, it is quite important to know if some of the variables do not contribute much to the cluster structure because redundant information for clusters can make it difficult to recover the latent cluster structure. The dimension reduction of variables is very useful to avoid the issue. Also, a lower-dimensional representation of the cluster structure, which is given by the dimension reduction, provides huge benefits for interpretation and visualization of the data. Recently, several techniques that achieve simultaneously clustering and dimension reduction of categorical data have been proposed (Hwang et al, 2006; Iodice D'Enza et al., 2014; van de Velden et al., 2017). These methods combine clustering and dimension reduction for categorical data by simultaneously assigning individuals to clusters and optimal scaling values to categories in various ways. Although the methods may provide good clustering recovery of the latent cluster structure and low-dimensional representation, the all methods are based on the descriptive model such as multiple correspondence analysis. As is well known, the probabilistic model for clustering, e.g., the mixture modeling, has many good points to analyze the data. In this paper, we propose a novel simultaneous analysis of clustering and dimension reduction for categorical data using the mixture modelling approach.

## 2 Proposed model

Let  $\mathbf{X} = (X_1, \dots, X_D)^\top$  be a random vector of  $D$  categorical variables with  $C$  categories.  $\mathbf{M}^\top$  denotes the transpose of a matrix (or a vector)  $\mathbf{M}$ . Here, for notational simplicity, we assume that all variables can have  $C$  categories, though it is possible to deal with data in which each variable has different categories,  $C_d$ . Suppose there are, in fact,  $K$  latent (i.e., unobservable) classes in a population and let  $\mathbf{U} = (U_1, \dots, U_K)^\top$  be an allocation variable that allocates each observation to one out of the  $K$  classes. We assume the allocation variable follows a multinomial distribution, i.e., the probability that  $\mathbf{U}$  takes the value  $\mathbf{u} = (u_1, \dots, u_K)^\top$  is

$$f(\mathbf{U} = \mathbf{u}) = \prod_{k=1}^K \xi_k^{u_k}, \quad (1)$$

where  $\xi_k = P(U_1 = 0, \dots, U_k = 1, \dots, U_K = 0)$ .

Given that an observation is in the  $k$ th latent class, the probability that the random vector  $\mathbf{X}$  takes the value  $\mathbf{x} = (x_1, \dots, x_D)^\top$ , where each of  $x_d$  takes one of  $1, \dots, C$ , is  $P(\mathbf{X} = \mathbf{x} \mid U_k = 1)$ , depending on a set of parameters. The unconditional probability of the response  $\mathbf{x}$ , when we do not know the latent class of the observation, is

$$P(\mathbf{X} = \mathbf{x}) = \sum_{k=1}^K \xi_k P(\mathbf{X} = \mathbf{x} \mid U_k = 1). \quad (2)$$

Here we need to specify how the probability  $P(\mathbf{X} = \mathbf{x} \mid U_k = 1)$  depends on parameters. We postulate that, given the latent class to which an observation belongs, the responses on the categorical variables are

independent:

$$P(\mathbf{X} = \mathbf{x} \mid U_k = 1) = \prod_{d=1}^D P(X_d = x_d \mid U_k = 1). \quad (3)$$

This assumption of *conditional independence* has been widely used in latent class modelling in sociology, and is directly analogous to the assumption, in the factor analysis model, that observed variables are conditionally independent given the factors.

Finally, to specify the model completely, we need to specify a set of parameters that define the conditional probability of  $\mathbf{X}$ , with the value of  $\mathbf{U}$  given. Suppose that  $\mathbf{X}_1, \dots, \mathbf{X}_N$  are i.i.d. copies of  $\mathbf{X}$ , and entries of  $\mathbf{X} = (x_{nd})$  are those realizations. We assume that, given the class  $k$ ,  $X_d$  follows the multinomial distribution with success probabilities  $\pi_{kd1}, \dots, \pi_{kdC}$ . Here, we consider a canonical parameter vector  $\boldsymbol{\theta}_k = (\boldsymbol{\theta}_{k1}^\top, \dots, \boldsymbol{\theta}_{kD}^\top)^\top$ , where  $\boldsymbol{\theta}_{kd} = (\theta_{kd1}, \dots, \theta_{kdC})^\top$ . Let  $X_{nd}$  be the  $d$ th element of  $\mathbf{X}_n$ . Then, the individual data generating probability given the class becomes

$$P(X_{nd} = x_{nd} \mid U_k = 1) = P(X_{nd} = x_{nd} \mid U_k = 1, \boldsymbol{\theta}_{kd}) = \prod_{c=1}^C \pi_{kdc}^{I(x_{nd}=c)}, \quad (4)$$

where

$$\pi_{kdc} = \frac{\exp(\theta_{kdc})}{\sum_{l=1}^C \exp(\theta_{kdl})}, \quad (5)$$

and  $I(\cdot)$  denotes an indicator function.

We aim to obtain a low-dimensional representation of multivariate categorical data, in which the true cluster structure exists. Thus, we assume that canonical parameter  $\theta_{kdc}$  has a low-rank representation as following

$$\theta_{kdc} = \mu_{dc} + \mathbf{f}'_k \mathbf{a}_{dc}, \quad (6)$$

where,  $\mu_{dc} \in \mathbb{R}$ , and for some positive integer  $L \in \mathbb{Z}^+$ ,  $\mathbf{f}_k \in \mathbb{R}^L$  and  $\mathbf{a}_{dc} \in \mathbb{R}^L$ . Note that We write  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_K)^\top$ ,  $\boldsymbol{\mu}_d = (\mu_{d1}, \dots, \mu_{dC})^\top$ ,  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_D^\top)^\top$ ,  $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_K)^\top$ ,  $\mathbf{A}_d = (\mathbf{a}_{d1}, \dots, \mathbf{a}_{dC})^\top$ , and  $\mathbf{A} = (\mathbf{A}_1^\top, \dots, \mathbf{A}_D^\top)^\top$ . For identifiability purpose, we require that  $\mathbf{F}$  has orthonormal columns. Then the log likelihood can be written as

$$\ell(\boldsymbol{\xi}, \boldsymbol{\mu}, \mathbf{F}, \mathbf{A}) = \sum_{n=1}^N \log \left( \sum_{k=1}^K \xi_k \prod_{d=1}^D \prod_{c=1}^C \left( \frac{\exp(\theta_{kdc})}{\sum_{l=1}^C \exp(\theta_{kdl})} \right)^{I(x_{nd}=c)} \right) \quad (7)$$

The proposed model has rotational indeterminacy as same as the ordinary factor analysis model and principal component analysis. Thus, to obtain an interpretable low-dimensional structure, it is needed to conduct any orthogonal rotation methods like the Varimax rotation. Instead of the classical rotation methods, we adopt sparse penalized maximum likelihood (ML) estimation to obtain the interpretable structure for  $\mathbf{A}$ . As described in Hirose and Yamamoto (2015) and Yamamoto et al. (2017), sparse penalization approach can be considered as a generalization of the ordinary ML estimation with rotation methods. In addition, the penalization can deal with high-dimensional problem if there is a sparse structure, i.e., most of the elements of the true  $\mathbf{A}$  are exactly zero.

Let  $P(\cdot)$  be a sparsity-inducing function and  $\rho$  is a tuning parameter for sparsity. Then, we obtain sparse cluster components by maximizing the following penalized log likelihood:

$$S(\boldsymbol{\xi}, \boldsymbol{\mu}, \mathbf{F}, \mathbf{A}) = \ell(\boldsymbol{\xi}, \boldsymbol{\mu}, \mathbf{F}, \mathbf{A}) - N \sum_{d=1}^D \sum_{c=1}^C \sum_{l=1}^L \rho P(|a_{dcl}|). \quad (8)$$

We call this procedure the Clustering of CAtegorical data with Reducing the Dimensionality (CCARD). We can interpret penalized maximization as the device for generating a suitable optimization function, but not a realistic representation of the actual data-generating process. Thus, in this sense, the conditional independence given the latent class for obtaining the likelihood in Eq. (7) is assumed.

### 3 Summary and conclusion

In this paper, a novel simultaneous procedure of clustering and dimension reduction for categorical data has been proposed. The proposed method is based on the ordinary latent class analysis model, and provides the low-dimensional representation of cluster structure. The low-dimensional representation of individuals is given by the convex combination of the scores for cluster representatives. It has been shown that the difference in the scores between individuals is related to the Jeffreys divergence for the response probability.

The parameters are estimated by maximizing the penalized ML. The EM-based optimization algorithm has been proposed to solve the maximization problem. In the M-step, each parameter is estimated sequentially by the block-wise coordinate descent. In addition, the convergence rate of the proposed algorithm has been investigated.

From the simulation study and the real-world data example, it can be seen that the proposed model provides better recoveries of latent cluster structures compared with existing methods. In the real-world data example, the estimated weights  $\mathbf{A}$  and the low-dimensional representation of clusters were actually helpful to understand the latent cluster structure of the data.

In conclusion, the proposed simultaneous analysis model can help researchers in various research fields to obtain and interpret the latent cluster structure in the categorical data.

### References

- [1] Iodice D'Enza, A., van de Velden, M., and Palumbo, F. (2014). On joint dimension reduction and clustering of categorical data. In D. Vicari, A. Okada, G. Ragozini, and C. Weihs (Eds.), *Analysis and modeling of complex data in behavioral and social sciences*. Berlin: Springer.
- [2] Hirose, K., and Yamamoto, M. (2015). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, 25:863–875.
- [3] Hwang, H., Dillon, W.R., and Takane, Y. (2006). An extension of multiple correspondence analysis for identifying heterogeneous subgroups of respondents. *Psychometrika*, 71:161–171.
- [4] van de Velden, M., Iodice D'Enza, A., and Palumbo, F. (2017). Cluster correspondence analysis. *Psychometrika*, 82:158–185.
- [5] Yamamoto, M., Hirose, K., and Nagata, H. (2017). Graphical tool of sparse factor analysis. *Behaviormetrika*, 44:229–250.



# Session VI

November 3, 2017 (Friday)

10:30 - 11:50

**Chair:**

Zhi Geng, Peking University

**Speaker:**

Qi Li, Tsinghua University

Moe Amagai, Tokyo University of Science

Yingying Wei, The Chinese University of Hong Kong

Yoshitomo Akimoto, Chuo University

# DISCOVERING POST-TRANSCRIPTIONAL REGULATORY NETWORKS OF RNAS BY MODELING GLOBAL MOLECULAR COMPITITION

Qi LI<sup>1,3</sup>, Pan WANG<sup>2,3</sup>, Jie HE<sup>2,4</sup>, Ke DENG<sup>1,4</sup>

1 Center for Statistical Science & Department of Industry Engineering, Tsinghua University, Beijing 100084, China

2 Department of thoracic surgery, Cancer Institute & Hospital, Chinese Academy of Medical Sciences

3 co-first authors

4 co-corresponding authors

## Introduction

Post-transcriptional regulatory network of RNAs (micro-RNAs and mRNAs) plays a critical role in a biological system, and is one of the keys to understand how cells work. However, discovering RNA interactions is a big challenge due to significant experimental difficulties and lack of high quality predicting algorithms. Integrating sequence characteristics and expression profile of RNAs with a probabilistic model that fully considers competition among different RNA molecules, we propose WePro, a computational method that provides precise predictions on RNA interactions.

## References

BARTEL D P. MicroRNAs: genomics, biogenesis, mechanism, and function [J]. Cell, 2004, 116(2): 281-97.

SALMENA L, POLISENO L, TAY Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? [J]. Cell, 2011, 146(3): 353-8.

LI Y, LIANG C, WONG K C, et al. Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach [J]. Nucleic Acids Res, 2014, 42(9): e76.

# A NECESSARY TEST FOR SPHERICITY BASED ON UNIFORMITY ON HYPERSPHERE

Moe AMAGAI<sup>a</sup>, Hiroki HASHIGUCHI<sup>b</sup> and Toshiya IWASHITA<sup>c</sup>

<sup>a</sup> Graduate School of Science, Tokyo University of Science

<sup>b</sup> Department of Applied Mathematics, Faculty of Science, Tokyo University of Science

<sup>c</sup> Department of Liberal Arts, Faculty of Science and Technology, Tokyo University of Science

## 1 Summary

This paper addresses a necessary test for sphericity of which a distribution has invariance under transformations by orthogonal matrices. Spherical distributions can be seen in many statical models for high dimensional analysis. Fang, Zhu and Bentler (1993) and Cuesta-Albertos, Cuevas and Fraiman (2009) proposed random projection methods for a sphericity test. In our presentation, we clarify that these methods have less power for some alternative hypotheses and we also propose a necessary test based on non-random projection. Numerical experiments are carried out for comparison between random and non-random projection methods.

A  $p$ -dimensional random vector  $\mathbf{x}$  is said to have a spherical distribution if  $H\mathbf{x}$  and  $\mathbf{x}$  have the same distribution for every  $p \times p$  orthogonal matrices such that  $H'H = HH' = I_p$ . We consider the null hypothesis  $H_0$ :  $\mathbf{x}$  has a spherical distribution, versus the alternative hypothesis  $H_1$ :  $\mathbf{x}$  doesn't have a spherical distribution. The null hypothesis can be changed into  $H_0^*$ :  $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$  is uniformly distributed on  $p$ -dimensional sphere, as mentioned in Cuesta-Albertos et al.(2009). Based on a goodness-of-fit test, we test  $H_0^*$  by projection of  $\mathbf{u}$  on a fix direction  $\mathbf{v} = (1, \dots, 1)'/\sqrt{p}$ . We use  $\mathbf{u}'\mathbf{v}$  as a test statistic of which the exact distribution  $F(z)$  is given as

$$F(z) = \begin{cases} 0 & z < -1, \\ (1 + \operatorname{sgn}(z) G_p(z^2 | 1/2, (p-1)/2)) / 2 & |z| \leq 1, \\ 1 & z > 1, \end{cases}$$

where  $\operatorname{sgn}(x)$  is the sign function and  $G_p(x|\alpha, \beta)$  is the cumulative distribution function of beta distribution with parameters  $(\alpha, \beta)$ . See Iwashita, Klar, Amagai and Hashiguchi (2017) for details. Now, we consider  $\mathbf{x} \sim N_p(\mathbf{0}, \Sigma)$  and  $H_0$ :  $\Sigma = I_p$ . An alternative test  $H_1$ :  $\Sigma = B'B$  is established by varying  $b$  in  $B = (1-b)I_p + bJ$ , where  $0 < b < 1$  and  $J$  is a matrix of ones. Tables 1 and 2 show Type I errors ( $\alpha = 0.05$ ) and powers of FP Test (proposed), RP1-S and RP10-S based on  $10^4$  replicants. Here, RP $k$ -S ( $k = 1, 10$ ) are refered in Cuesta-Albertos et al.(2009).

Table 1:  $p = 3, n = 20$

	$H_0$	$b = 0.2$	$b = 0.5$
FP test	0.0501	0.3755	0.9927
RP1-S	0.0120	0.0148	0.0409
RP10-S	0.0174	0.0225	0.0640

Table 2:  $p = 50, n = 100$

	$H_0$	$b = 0.2$	$b = 0.5$
FP test	0.0510	1.0000	1.0000
RP1-S	0.0405	0.3803	0.8056
RP10-S	0.0278	0.7746	0.9998

## References

- [1] Cuesta-Albertos, J.A., Cuevas, A., Fraiman, R., (2009). On projection-based tests for directional and compositional data. *Stat. Comput.*
- [2] Fang, K.T., Zhu, L.X., Bentler, P.M., (1993). A necessary test of goodness of fit for sphericity. *J. Multivariate Anal.* 45, 34–55.
- [3] Iwashita, T., Klar, B., Amagai, M. and Hashiguchi, H. (2017). A test procedure for uniformity on the Stiefel manifold based on Projection. *Statist. Probab. Lett.*, 128, 89–96.

# **BATCH EFFECTS CORRECTION WITH UNKNOWN SUBTYPES**

Xiangyu LUO and Yingying Wei

Department of Statistics, The Chinese University of Hong Kong

Hong Kong, China

High-throughput experimental data are accumulating exponentially in public databases. Unfortunately, however, mining valid scientific discoveries from these abundant resources is hampered by technical artifacts and inherent biological heterogeneity. The former are usually termed batch effects and the latter is often modeled by subtypes. Existing methods either tackle batch effects provided that subtypes are known or cluster subtypes assuming that batch effects are absent. Consequently, there is a lack of research on the correction of batch effects with the presence of unknown subtypes. Here, we combine a location-and-scale adjustment model and model-based clustering into a novel hybrid one, the Batch-effects-correction-with-Unknown-Subtypes model (BUS). BUS is capable of correcting batch effects explicitly, grouping samples that share similar characteristics into subtypes, identifying features that distinguish subtypes, and enjoying a linear-order computational complexity. We prove the identifiability of BUS and provide conditions for study designs under which batch effects can be corrected. Finally, BUS is evaluated by simulation studies and a real breast cancer dataset.

# PERFORMANCE STUDIES OF TEST STATISTICS FOR TWO-SAMPLE DIRECTIONAL DATA

Yoshitomo AKIMOTO and Toshinari KAMAKURA  
 Graduate School of Science and Engineering, Chuo University  
 Tokyo 112-8551, Japan

## 1. Introduction

Directional statistics has been studied as one of the major statistical topics in wind data or monthly sales data that sometimes handle the directional objects in 2-dimensional space. In the directional statistics, there are many test statistics.

Then, the von Mises distribution is commonly assumed as the sample distribution. It has 2 parameters: concentration  $\kappa$  and mean direction  $\mu$ .

$$f_{vm}(\theta|\mu, \kappa) = \frac{\exp\{\kappa \cos(\theta - \mu)\}}{2\pi I_0(\kappa)}, \quad (1)$$

$$0 \leq \theta < 2\pi, 0 \leq \mu < 2\pi, \kappa \geq 0,$$

where  $I_j(\kappa)$  is the modified Bessel function of the first kind and order  $j$ . noted eq(2).

$$I_j(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(jx) \exp(\kappa \cos x) dx \quad (2)$$

The follows show the distributions of  $\kappa = 0, 0.5, 1, 2$  and  $4$ .

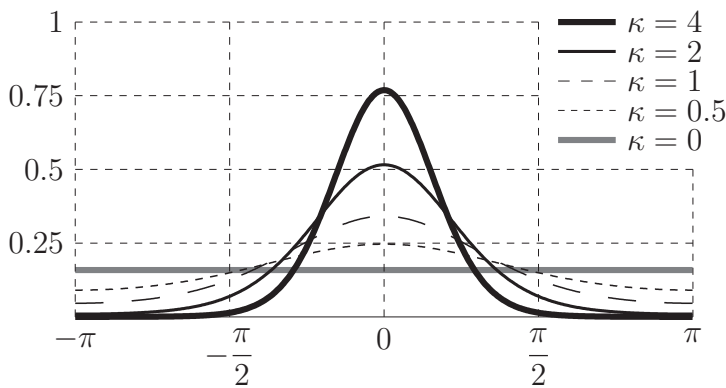


Figure 1: The behavior of  $\kappa$

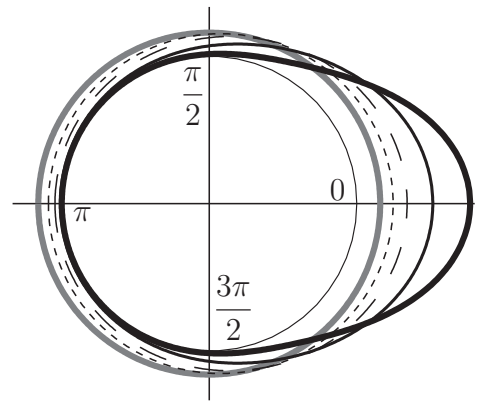


Figure 2: The behavior of  $\kappa$  on circular projection

For instance, Rayleigh test and V-test as tests for uniformity is often used. They are score test statistic of this distribution of  $\kappa = 0$ . As the further improvement, we have derived the LM test statistics for angular data. Our LM test statistic (Akimoto, Sakumura, & Kamakura, 2016) coincide with V-test when the support is on all semi-circumference.

In this paper, we derive the LM test statistic of the other hypothesis for some directional data analysis.

## 2. Two-sample test

As the one of the most used testing, the two-sample test is also required. Then, let  $\theta_{11}, \theta_{12}, \dots, \theta_{1n_1}$  and  $\theta_{21}, \theta_{22}, \dots, \theta_{2n_2}$ , the independent random samples of sizes  $n_1, n_2$  from von Mises distribution  $(\mu_1, \kappa_1)$  and  $(\mu_2, \kappa_2)$  respectively, the sample mean directions be  $\bar{\theta}_1, \bar{\theta}_2$  and the resultant lengths be  $R_1, R_2$  respectively. The  $R_j$  is denoted that

$$R_j = \sqrt{\left(\sum_{i=1}^{n_j} \cos \theta_{ji}\right)^2 + \left(\sum_{i=1}^{n_j} \sin \theta_{ji}\right)^2} = n_j \bar{R}. \quad (3)$$

As the two-sample testing from von Mises distribution  $(\mu, \kappa)$ , we focused on the mean directions in this paper.

### 2.1 Test for equality of mean directions

The testing hypothesis  $H_0 : \mu_1 = \mu_2$ , against  $H_1 : \mu_2 \neq \mu_2$ . Since each mean direction makes resultant vector of the mean directions, cosine rule gives

$$R^2 = R_1^2 + R_2^2 + 2R_1R_2 \cos(\bar{\theta}_1 - \bar{\theta}_2).$$

From this equation, Mardia(2009) described the likelihood test statistics with the known  $\kappa$ ,

$$\omega = 2\kappa(R_1 + R_2 - R).$$

On the other hand, when  $\kappa$  is unknown, the distribution of statistic depends on  $\kappa$ . The likelihood ratio statistic is

$$\omega = 2 \{ \hat{\kappa}_{12}(R_1 + R_2) - \hat{\kappa}R - \log I_0(\hat{\kappa}_{12}) + \log I_0(\hat{\kappa}) \},$$

where  $\hat{\kappa}$  and  $\hat{\kappa}_{12}$  are the MLEs of  $\kappa$  under  $H_0$  and  $H_1$  respectively.

## 3. Proposed statistic

In this section, we propose the LM test statistic for equality of mean directions. Suppose that  $\mu = \mu_1 = \mu_2$ , the MLEs of  $\mu_j$  and  $\kappa$  be  $\hat{\mu}_j = \bar{\theta}_j$  and  $\hat{\kappa} = A^{-1}(\bar{R})$ , where  $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ . Thus, We calculated the LM test statistic for equality of mean directions,

$$A'(\hat{\kappa})\hat{\kappa}^2 \left\{ \frac{n_1 \left[ \sum_{i=1}^{n_1} \sin(\theta_{1i} - \hat{\mu}) \right]^2}{A'(\hat{\kappa})\hat{\kappa}n_1 \sum_{i=1}^{n_1} \cos(\theta_{1i} - \hat{\mu}) - \left[ \sum_{i=1}^{n_1} \sin(\theta_{1i} - \hat{\mu}) \right]^2} + \frac{n_2 \left[ \sum_{i=1}^{n_2} \sin(\theta_{2i} - \hat{\mu}) \right]^2}{A'(\hat{\kappa})\hat{\kappa}n_2 \sum_{i=1}^{n_2} \cos(\theta_{2i} - \hat{\mu}) - \left[ \sum_{i=1}^{n_2} \sin(\theta_{2i} - \hat{\mu}) \right]^2} \right\}.$$

## References

- Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics*. World Scientific.  
Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics*. John Wiley & Sons.

Rao, J. S., & Sengupta, S. (1972). Mathematical techniques for paleocurrent analysis: treatment of directional data. *Journal of the International Association for Mathematical Geology*, **4**(3), 235-248.

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge University Press.

Akimoto, Y., Sakumura, T., & Kamakura, T. (2016). Testing for Uniformity of Line Segments. *IAENG International Journal of Computer Science*, **43**(2).



2017

Hangzhou

International Statistical Symposium

**Local Contact**

Name: Miss Qian Bao

Email: [qianbao@mail.tsinghua.edu.cn](mailto:qianbao@mail.tsinghua.edu.cn)

Mobile: +86 18101216283