國 立 交 通 大 學

統計學研究所

碩士論文

問卷中單複選題的選項排序方法的探討

Ranking Responses of a Single Response Question or a Multiple Response Question

研 究 生：林育駿

指導教授：王秀瑛　教授

中 華 民 國 一 百 零 三 年 六 月

問卷中單複選題的選項排序方法的探討

Ranking Responses of a Single Response Question or a Multiple Response Question

研 究 生：林育駿 Student: Yu-Chun Lin

指導教授：王秀瑛 教授 Advisor: Dr. Hsiuying Wang

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute of Statistics College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2014

Hsinchu, Taiwan, Republic of China

中華民國一百零三年六月

問卷中單複選題中的選項排序方法的探討

學 生: 林育駿　　　　　　　　　　　　指導教授: 王秀瑛 教授

國立交通大學統計學研究所

摘　　要

問卷調查在許多研究中是一個最常用的工具，而複選題是一個最常被設計在問卷中的題型。近年來，許多研究提出了一些方法對於複選題的資料作分析，其中，複選題的選項排序的問題是一個重要且我們感興趣的議題。在本篇文章中，我們應用了Wang (2008)，Wang和Huang (2014) 所提出的檢定方法，以及Hunter DR (2004) 中討論的Bradley-Terry模型並用Minorization—Maximization演算法來對單選題或複選題的選項做排序。我們將這些方法寫成一個R package，而這個package已經被收錄在R 軟體的程式套件，讓使用者們能更方便的用這些方法來對單選題或複選題的選項做排序。

關鍵字：問卷、單選題、複選題、排序

Ranking Responses of a Single Response Question or a Multiple Response Question

Student: Yu Chun Lin Advisor: Hsiuying Wang

Institute of Statistics

National Chiao Tung University

Abstract

In many studies, the questionnaire is a common tool for surveying. A multiple response question is a commonly-used question designed in a questionnaire. Recently, many methods had been proposed to analyze data of a multiple responses question in the literature. And ranking responses is one of important issues in analyzing data of a multiple responses question. In this thesis, we use the methods in Wang (2008), Wang and Huang (2014) and Bradley-Terry model with MM (Minorization—Maximization) algorithm in Hunter DR (2004) to develop a R package. It can provide a useful and convenient tool to rank responses in a signal response or a multiple response question.

Keywords : Questionnaire, Single response question, Multiple response question, Ranking

誌謝

很高興當初考研究所的時候可以錄取交大統研所，是我理想的學校；在這邊兩年的時間學到很多我喜歡的事物，不僅僅是在課業上，生活上也跟同學們過得很快樂。首先，要先感謝我的指導教授—王秀瑛 老師，老師在我作研究、學習的路上給予我很多的幫助，我不清楚、不知道的概念請教老師，老師都會很仔細的講解讓我真正的了解問題所在並找出解決方法，也很謝謝老師讓我早早就口試，可以為下一個人生階段做更多的準備；也要謝謝我的口試委員：蔡明田 教授、蕭金福 教授以及陳鄰安 教授，在口試的時候提多許多想法及建議，讓我的論文更加的完整。

碩士兩年的生活過得很多采多姿，跟同學們常常出遊、吃大餐，在研究室也時常有歡笑聲。還記得碩一的時候為了數理統計的期中考，幾乎有一半的人在研究室讀通宵，當時很享受大家這樣的凝聚力，讓我覺得有這群同學真的不是一般的福氣。不只在課業上大家有這樣的向心力，在享樂方面大家也都很積極的參與，雖然說沒有人討厭玩樂，但是大家也都會把事情排開就為了要跟班上的人一起出遊，這部分就要很感謝姿琪，她根本就是我們班的玩樂王，常常揪大家一起出去玩，安排行程、規劃路線全都是靠她，我能在交大過的這麼開心有一部分是她的功勞。另外還要特別感謝驛為和民翰，分別是我們碩一碩二時的班代，為我們處理許多事，讓我們只要做簡單的回報就可以把事情完成。班上的其他同學明燕、昱均、兆慶、翔之、奇煒、青樺以及其他人，也都讓我的碩士生活有更多色彩。兩年的時間過得真的很快，也很捨不得跟這群同學分開，但我相信在以後我們依然能像現在一樣，友情永遠不滅。

林育駿 謹誌于

國立交通學統計學研究所

中華民國一零三年六月

Contents

List of Tables

1 Introduction

Questionnaires are a commonly used tool for surveying in many fields. They are especially important in marketing or management studies. There are usually two kinds of questions : single-response questions and multiple-response questions. The analysis of single-response questions have been investigated in literature. Approaches of analyzing multiple-response questions have been lacking until recently. Several researchers propose some methods about analyzing the dependence between a single-response question and a multiple-response question (Umesh 1995,Loughin and Scherer 1998, Decady and Thomas 1999, Bilder, Loughin and Nettleton 2000). However, most researchers are also interested in ranking responses in a question. Wang (2008) and Wang and Huang (2014) had proposed methods for ranking responses in a multiple-response question under the frequentist or Bayesian setup. Thus, in this thesis, we provide some methods to discuss the problem. They include Wald test and Generalized score test in Wang (2008), Bayesian ranking response method in Wang and Huang (2014) and Bradley-Terry model with MM algorithm in Hunter DR (2004).

For example, a company is designing a marketing survey to help develop a drink product. The researchers will design a multiple response question and list several factors, including price, packaging, capacity, taste that could attract consumers to buy this product. Suppose that a group of individuals are surveyed on purchasing a drink product. They are asked to write questionnaires which list all the questions. The following is a multiple-response question in the questionnaire:

Question 1 : Which factor are important to you when considering the purchase of a drink? (1) taste (2) capacity (3) packaging (4) price (5) other

According to the number of each response which is chosen, many respondents more care about price and taste than other factors. Then we can easily rank the response "price" first and "taste" second according to the number of responses which are selected. But it is based on the response selected numbers to rank, is not statistically significant and we cannot clearly distinguish "price" is more important than "taste". In this thesis, we base on several methods, Wald test, Generalized Score test, Bayesian ranking response method and Bradley-Terry model with MM algorithm in the literature to develop ranking procedures. These ranking procedures has been written as a package RankResponse for R. RankResponse is available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=RankResponse`, which include code function rank.wald, rank.gs, LR, LN, L2R, btmm, btqn and btnr. We review these methods in the Preliminary section. In Section 3, we propose rules to rank responses. In Section 4, we compare these methods by a simulation study. In Section 5, we introduce details of these R codes.

2 Preliminary

In this section, we introduce some methods in the literature to rank responses.

2.1 The Wald Test and the Generalized Score Test

First, we consider ranking two specific responses. For the general case, assume that a multiple-response question has **k** responses, $v_1, \cdots, v_k$, and we interview **n** respondents. Each respondent is asked to choose at least one and at most s answers for this question, where $0 < s \leqslant k$. If s=1, it is a single-response question. There are total of $c = C_1^k + \cdots + C_s^k$ possible kinds of answers that respondents will choose. Let $n_{i_1 \cdots i_k}$ denote the number of respondents selecting the responses v_h and not selecting $v_{h'}$ if i_h=1 and $i_{h'}$=0, and $p_{i_1 \cdots i_k}$ denotes the corresponding probability. For example, when k=5, n_{10100} denotes the number of respondents selecting the first and the third responses and not selecting the other responses. Thus, the p.m.f function of $n_{i_1 \cdots i_k}$ is

$$f_s(n_{i_1 \cdots i_k}) = I(0 < \sum_{j=1}^{k} i_j \leq s) \frac{n!}{\prod\limits_{i_j=0or1} n_{i_1 \cdots i_k}!} \prod_{i_j=0or1} p_{i_1 \cdots i_k}^{n_{i_1 \cdots i_k}}, \quad n_{i_1 \cdots i_k} \geq 0, 0 \leq p_{i_1 \cdots i_k} \leq 1, \quad (1)$$

where $I(\cdot)$ denotes the indicator function. Let m_j denote the sum of the number $n_{i_1 \cdots i_k}$ such that the jth response is selected, and π_j denote the corresponding probability, that is $m_j = \sum\limits_{i_j=1} n_{i_1 \cdots i_k}$ and $\pi_j = \sum\limits_{i_j=1} p_{i_1 \cdots i_k}$. Note π_j is called a marginal probability of response j. Also let m_{jl} denote the sum of number $n_{i_1 \cdots i_k}$ such that the jth and lth responses are selected, and π_{jl} denote the corresponding probability. Then $m_{jl} = \sum\limits_{i_j=i_l=1} n_{i_1 \cdots i_k}$ and $\pi_{jl} = \sum\limits_{i_j=i_l=1} p_{i_1 \cdots i_k}$. For ranking the important of two specified responses, say response 1 and response i in Question j from the survey data, we will consider the two sided test:

$$H_0 : \pi_i = \pi_j \;\; vs \;\; H_1 : \pi_i \neq \pi_j \quad (2)$$

which is equivalent to

$$H_0^* : \pi_i - \pi_{ij} = \pi_j - \pi_{ij} \;\; vs \;\; H_1^* : \pi_i - \pi_{ij} \neq \pi_j - \pi_{ij} \quad (3)$$

If (2) is rejected, then we ca rank the response with larger m_j first. The methods for testing (2) are given in Wang(2008).

2.1.1 Wald Test

A Wald test is a test based on a statistic of the form

$$Z_n = \frac{W_n - (\pi_i - \pi_j)}{S_n}$$

where W_n is an estimator of $\pi_i - \pi_j$, and S_n is a standard error for W_n. An unbiased estimator of $p_{i_1 \cdots i_k}$ is $n_{i_1 \cdots i_k}/n$, which is also a maximum likelihood estimator(MLE). Let $\hat{\pi}_i = m_i/n$,$\hat{\pi}_j = m_j/n$ and $\hat{\pi}_{ij} = m_{ij}/n$. We can use $\hat{\pi}_i$ and $\hat{\pi}_j$ as estimators of π_i and π_j respectively, and we have

$$Var(\hat{\pi}_i - \hat{\pi}_j) = \begin{cases} \pi_i(1-\pi_i)/n + \pi_j(1-\pi_j+2\pi_i/n) & if \;\; s = 1, \\ (\pi_i - \pi_{ij})(1-\pi_i+2\pi_j-\pi_{ij})/n + & \\ (\pi_j - \pi_{ij})(1-\pi_j+\pi_{ij})/n & otherwise. \end{cases} \tag{4}$$

Under the null hypothesis H_0 in (2) and based on central limit theorem, the statistics

$$\frac{\hat{\pi}_i - \hat{\pi}_j}{\sqrt{Var(\hat{\pi}_i - \hat{\pi}_j)}} \tag{5}$$

converges in distribution to standard normal random variable when n large. Since π_i,π_j and π_{ij} are unknown,we can use $\hat{\pi}_i$, $\hat{\pi}_j$ and $\hat{\pi}_{ij}$ to substitute π_i, π_j and π_{ij} in (4). Thus, for testing (2), H_0 is rejected if absolute value of (5) is greater than $z_{\alpha/2}$, where $z_{\alpha/2}$ is upper $\alpha/2$ cutoff point of the standard normal distribution.

2.1.2 Generalized Score Test

In section 2.1.1, π_i, π_j and π_{ij} in $Var(\hat{\pi}_i - \hat{\pi}_j)$ are replaced by $\hat{\pi}_i$, $\hat{\pi}_j$ and $\hat{\pi}_{ij}$ in the test statistic. In this section, we consider the variance under the null hypothesis in (2), that is,

$\pi_i = \pi_j$. Thus we have

$$Var_{\pi_i=\pi_j}(\hat{\pi}_i - \hat{\pi}_j) = \begin{cases} 2\pi_i/n & if \;\; s = 1 \\ 2(\pi_i - \pi_{ij})/n & otherwise. \end{cases} \quad (6)$$

By the central limit theorem, under H_0, the statistic

$$\frac{\hat{\pi}_i - \hat{\pi}_j}{\sqrt{Var_{\pi_i=\pi_j}(\hat{\pi}_i - \hat{\pi}_j)}}$$

converges to a standard normal distribution when n is large. We can use $(\hat{\pi}_i + \hat{\pi}_j)/2$ and $\hat{\pi}_{ij}$ as substitutes for π_i and π_{ij} in the variance. Hence for testing (2), the null hypothesis is rejected if

$$\begin{cases} \frac{\sqrt{n}|\hat{\pi}_i - \hat{\pi}_j|}{\sqrt{\hat{\pi}_i + \hat{\pi}_j}} > z_{\alpha/2} & if \;\; s = 1, \\ \frac{\sqrt{n}|\hat{\pi}_i - \hat{\pi}_j|}{\sqrt{\hat{\pi}_i + \hat{\pi}_j - 2\hat{\pi}_{ij}}} > z_{\alpha/2} & otherwise \end{cases}$$

This approach is similar to the score test of testing a marginal probability equal to a specified value. Hence we call this approach a generalized score test.

2.2 Bayesian Ranking Responses

In this section we review the Bayesian Ranking methods in Wang and Huang (2014). We assume parameters $p_{i_1 \cdots i_k}$ have a prior distribution. Thus, we consider the conjugate prior

$$\eta(p) = I(0 < \sum_{j=1}^{k} i_j \leq s) \frac{\Gamma(\sum\limits_{i_j=0 \text{ or } 1} \alpha_{i_1 \cdots i_k})}{\prod\limits_{i_j=0 \text{ or } 1} \Gamma(\alpha_{i_1 \cdots i_k})} \prod_{i_j=0 \text{ or } 1} p_{i_1 \cdots i_k}^{\alpha_{i_1 \cdots i_k}}, \quad \alpha_{i_1 \cdots i_k} > 0, 0 \leq p_{i_1 \cdots i_k} \leq 1, \quad (7)$$

which is a Dirichlet distribution. The prior information is assumed to be known here.

Under this assumption, we have the posterior distribution

$$\begin{aligned} \eta(\mathrm{p}|\mathbf{n}) &= f_s(\mathbf{n}|\mathrm{p})\eta(\mathrm{p}) \\ &= I(0 < \sum_{j=1}^{k} i_j \leq s) \frac{\Gamma\left\{\sum\limits_{i_j=0 \text{ or } 1} \alpha_{i_1 \cdots i_k} + n_{i_1 \cdots i_k}\right\}}{\prod\limits_{i_j=0 \text{ or } 1} \Gamma(\alpha_{i_1 \cdots i_k} + n_{i_1 \cdots i_k})} \prod_{i_j=0 \text{ or } 1} p_{i_1 \cdots i_k}^{\alpha_{i_1 \cdots i_k} + n_{i_1 \cdots i_k}} \end{aligned} \quad (8)$$

Through the form of the posterior distribution, we can derive the Bayes estimator for each $p_{i_1 \cdots i_k}$ under the squared error loss function. The Bayes estimator $\hat{\pi}_j$ of π_j is equal to the summation of the Bayes estimator of $p_{i_1 \cdots i_k}$, where $i_j = 1$. We can use the Bayes estimator of π_j to rank the significant of π_j. Moreover, if we can associate a testing approach with the Bayes estimator to rank π_j, this method can lead to a more accurate result. Therefore, we propose several multiple-testing methods for testing the relationship of π_j.

Assume we have k responses and we are interested in testing

$$\left.\begin{array}{rcl} H_{01} : \pi_{(2)} \leq \pi_{(1)} & \text{versus} & H_{01} : \pi_{(2)} > \pi_{(1)}, \\ H_{02} : \pi_{(3)} \leq \pi_{(2)} & \text{versus} & H_{01} : \pi_{(3)} > \pi_{(2)}, \\ & \vdots & \\ H_{0k-1} : \pi_{(k)} \leq \pi_{(k-1)} & \text{versus} & H_{0k-1} : \pi_{(k)} > \pi_{(k-1)} \end{array}\right\} \quad (9)$$

For testing expressions (9), the decision rules that are considered here are to control the posterior false discovery rate. The concept of the false discovery rate was proposed by Benjamini and Hochberg (1995) to determine optimal thresholds for a multiple-testing setting.

Then we define the false discovery rate, posterior false discovery rate, false negative rate and posterior false negative rate for the frequentist and Bayesian setting based on the literature as follows.

First, some notation and definitions are given. Let z_i denote an indicator that the ith hypothesis H_{0i} is false and let $u_i = P(z_i = 1|\mathbf{n})$ denote the marginal posterior probability of $\pi_{i+1} > \pi_i$. The rejection of H_{0i} is denoted by d_i=1; otherwise d_i=0. Let the false discovery rate and false negative rate are denoted by $FDR(d, z)$ and $FNR(d, z)$ respectively,where

$$FDR(d, z) = \frac{\sum_{i=1}^{k-1} d_i(1 - z_i)}{D + \varepsilon}$$

and

$$FNR(d, z) = \frac{\sum_{i=1}^{k-1} (1 - d_i)z_i}{n - D + \varepsilon},$$

$D = \sum_{i=1}^{k-1} d_i$ and ε is a small constant to avoid a zero denominator, so we choose ε=0.00001.

Let the posterior expected false discovery rate denoted by $\overline{FDR}(d, \mathbf{n})$ and the posterior expected false negative rate $\overline{FNR}(d, \mathbf{n})$, where

$$\overline{FDR}(d, \mathbf{n}) = \frac{\sum_{i=1}^{k-1} d_i(1 - u_i)}{D + \varepsilon}$$

and

$$\overline{FNR}(d, \mathbf{n}) = \frac{\sum_{i=1}^{k-1} (1 - d_i)u_i}{n - D + \varepsilon}$$

Let posterior expected false discovery count $\overline{FD}(d, \mathbf{n})$ and the posterior expected false negative count $\overline{FN}(d, \mathbf{n})$ are defined as

$$\overline{FD}(d, \mathbf{n}) = \sum_{i=1}^{k-1} d_i(1 - u_i)$$

and

$$\overline{FN}(d, \mathbf{n}) = \sum_{i=1}^{k-1} (1 - d_i)u_i$$

Then we propose three multiple-testing procedures from Berger(1985) and Muller *et al.*(2004) for testing expressions (9).

2.2.1 Method 1

The decision to accept or reject the null hypothesis according to a loss function that was proposed by Berger (1985), which is defined as

$$\left.\begin{array}{ll} 0 & \text{if the decision taken is right,} \\ \text{c} & \text{if we reject } H_{0i} \text{ when it is true,} \\ 1 & \text{if we accept } H_{0i} \text{ when it is false,} \end{array}\right\}$$

where c($\geq$0) and 1 represent the losses for making a wrong decision because of a false positive and a false negative error respectively. In this criterion, the loss function can be written as

$$L_N(d, \mathbf{n}) = c\overline{FD} + \overline{FN} \tag{10}$$

2.2.2 Method 2

The second method is consider the loss function

$$L_R(d, \mathbf{n}) = c\overline{FDR} + \overline{FNR} \tag{11}$$

2.2.3 Method 3

We consider the third loss function is a bivariate function

$$L_{2R}(d, \mathbf{n}) = (\overline{FDR}, \overline{FNR}) \tag{12}$$

We can define the optimal decisions under L_{2R} as the minimization of $\overline{FNR}$ subject to $\overline{FDR}$ $\leq e_{2R}$.

From Muller *et al.*(2004), under the three loss functions, the optimal decision that minimizes the loss function takes the form

$$d_i = I(u_i \geq t),$$

where t are $t_N = c/(c+1)$, $t_R(\mathbf{n}) = u_{(n-D^*)}$ and $t_{2R}(\mathbf{n})$=min$\{s : \overline{FDR}(s, \mathbf{n}) \leq e_{2R}\}$ under the loss function L_N,L_R and L_{2R} respectively. To obtain the value of μ_i, the code is available in `http://www.stat.nctu.edu.tw/hwang/ranking.htm`. In the expressions for t_R and t_{2R}, u_i is the ith order statistic of $\{u_1, \ldots, u_n\}$, and D^* is the optimal number of discoveries found by the function in Muller *et al.*(2004). Then we according to these d_i to decide expressions (9) whether reject or not.

2.3 Ranking Responses with Bradley-Terry Models

In this section, we introduce Bradley-Terry models with MM method (Hunter DR 2004). Then according to result of the method, we can rank responses sequentially. We review this method in Section 2.3.1 and 2.3.2.

2.3.1 Bradley-Terry Models

In a situation in which the individuals in a group are repeatedly compared with one another in pairs, Bradley-Terry(1952) suggested the model

$$P(\text{individual } i \text{ beats individual } j) = \frac{\gamma_i}{\gamma_i + \gamma_j}, \tag{13}$$

where γ_i is positive-values parameter associated with individual i, for each of the comparisons pitting individual i against individual j. As a concrete example, consider the individuals to be sports teams, where γ_i represent the overall skill of team i.

Suppose we observe a number of pairing s among m individuals or teams and we wish to estimate the parameters $\gamma_1, \ldots, \gamma_m$ using maximum likelihood estimation. If outcomes of different pairings are assumed to be independent, the log-likelihood based on the Bradley-

Terry model (13) is

$$\ell(\boldsymbol{\gamma}) = \sum_{i=1}^{m}\sum_{j=1}^{m}[w_{ij}\ln\gamma_i - w_{ij}\ln(\gamma_i + \gamma_j)], \tag{14}$$

where w_{ij} denotes the number of times individual i gas beaten individual j and we assume w_{ii}=0 by convention. Since $\ell(\boldsymbol{\gamma})$=$\ell$(a$\boldsymbol{\gamma}$) for a>0, the parameter space should be regarded as the set of equivalence classes $\mathbb{R}_+^m$, where two vectors are equivalent if one is a scalar multiple of the other. This is most easily accomplished by putting a constraint on the parameter space; to this end, we assume that $\sum_i \gamma_i$=1.

Now we describe an iterative algorithm to maximize $\ell(\boldsymbol{\gamma})$. Start with an initial parameter vector $\boldsymbol{\gamma}^{(1)}$. There are many ways to select starting points,in this paper, we assume that $\boldsymbol{\gamma}^{(1)}$ is chosen arbitrarily. For k=1,2,...,let

$$\gamma_i^{(k+1)} = W_i\left[\sum_{j\neq i}\frac{N_{ij}}{\gamma_i^{(k)}+\gamma_j^{(k)}}\right]^{-1}, \tag{15}$$

where W_i denotes the number od wins by individual i and $N_{ij} = w_{ij} + w_{ji}$ is the number of pairing between i and j. If the resulting $\boldsymbol{\gamma}^{(k+1)}$ vector does not satisfy the constraint $\sum_i \boldsymbol{\gamma}_i^{(k+1)} = 1$, it should simply be renormalized.

2.3.2 Minorizing functions and MM algorithm

The strict concavity of the logarithm function implies for positive x and y that

$$-\ln x \geq 1 - \ln y - (x/y) \qquad \text{with equality if and only if } x = y \tag{16}$$

Therefore, as shown in Lange, Hunter and Yang (2000), if we fix $\boldsymbol{\gamma}^{(k)}$ and define the function

$$Q_k(\boldsymbol{\gamma}) = \sum_{i=1}^{m}\sum_{j=1}^{m} w_{ij}\left[\ln\gamma_i - \frac{\gamma_i+\gamma_j}{\gamma_i^{(k)}+\gamma_j^{(k)}} - \ln(\gamma_i^{(k)}+\gamma_j^{(k)}) + 1\right], \tag{17}$$

we may conclude that

$$Q_k(\boldsymbol{\gamma}) \leq \ell(\boldsymbol{\gamma}) \qquad \text{with equality if } \boldsymbol{\gamma} = \boldsymbol{\gamma}^{(k)} \tag{18}$$

where $\ell(\boldsymbol{\gamma})$ is the log-likehood of (14). A function $Q_k(\boldsymbol{\gamma})$ satisfying conditions (18) is said *minorize* $\ell(\boldsymbol{\gamma})$ at the point $\boldsymbol{\gamma}^{(k)}$. It is easy to verify that, for any $Q_k(\boldsymbol{\gamma})$ satisfying the minorizing conditions (18),

$$Q_k(\boldsymbol{\gamma}) \geq Q_k(\boldsymbol{\gamma}^{(k)}) \qquad \text{implies} \qquad \ell(\boldsymbol{\gamma}) \geq \ell(\boldsymbol{\gamma}^{(k)}) \tag{19}$$

Property (19) suggests an iterative algorithm in which we let $\boldsymbol{\gamma}^{(k)}$ denote the value of the parameter vector before the kth iteration and define $\boldsymbol{\gamma}^{(k+1)}$ to be the maximizer of $Q_k(\boldsymbol{\gamma})$; thus $\boldsymbol{\gamma}^{(k+1)}$ of (15) maximizes $Q_k(\boldsymbol{\gamma})$. Since this algorithm consists of alternately creating a minorizing function $Q_k(\boldsymbol{\gamma})$ and then maximizing it, Hunter and Lange (2000) call it an MM algorithm.

3 Ranking Rule

In this section we introduce a criterion to rank the responses using the methods introduced in Section 2. The ranking rule for the Wald test, the Generalized Score test and the Bayesian ranking methods are similar, but they are different from the rank method with Bradley-Terry model.

Now we first illustrate the ranking rule of the Wald test, the Generalized Score test and the Bayesian ranking methods. Assume that we have k responses and corresponding m_j

value, $j = 1, \ldots, k$. Let $m_{(j)}$ be the order statistics, that is, $m_{(1)} \leq m_{(2)} \leq, \ldots, m_{(k)}$. Let $v_{(j)}$ be the response corresponding to $m_{(j)}$. It is natural to rank the importance of responses in order of $m_{(j)}$. That is, the most important response is $v_{(k)}$, and the second important response is $v_{(k-1)}$. The proposed testing methods in Section 2.1 and Section 2.2 can be used to rank the responses. If the hypothesis $\pi_{(k)} = \pi_{(k-1)}$ is rejected, we may claim that $v_{(k)}$ is the most important response. If it is not rejected, the two responses have same rank, that is, the response $v_{(k)}$ is as important as the response $v_{(k-1)}$, and next to test the hypothesis $\pi_{(k-1)} = \pi_{(k-2)}$. Similarly, if it is rejected, we rank response $v_{(k-1)}$ first and response $v_{(k-2)}$ second. If it is not rejected, response $v_{(k-1)}$ and $v_{(k-2)}$ have same rank.

We use Question 1 as an example to illustrate the rule. For example, let m_1=54, m_2=49, m_3=28, m_4=71, m_5=23, and then we have $m_{(1)}$=23, $m_{(2)}$=28, $m_{(3)}$=49, $m_{(4)}$=54, $m_{(5)}$=71. It is natural to rank the importance of responses in order of $m_{(j)}$. We may claim that price is more important than taste for consumers to purchase. However, this rank method based on the order of $m_{(j)}$ is not statistically significant. Hence, we follow the proposed rule and use one of these methods to rank all response. First, we rank response $v_{(5)}$ and response $v_{(4)}$, i.e. rank the response "taste" and response "price". If $H_{01} : \pi_{(5)} = \pi_{(4)}$ is not reject, it means that the response "taste" and the response "price" are equally important. Then we test $H_{02} : \pi_{(4)} = \pi_{(3)}$ versus $H_{12} : \pi_{(4)} \neq \pi_{(3)}$. If H_{02} is rejected ,we rank responses $v_{(5)} and v_{(4)}$ first and response $v_{(3)}$ third. And when hypothesis $H_{03} : \pi_{(3)} = \pi_{(2)}$ and $H_{04} : \pi_{(2)} = \pi_{(1)}$, we reject H_{03} and do not reject H_{04}. Then we denote the ranking notations for the above result as "taste" 1, "capacity" 3, "packaging" 4, "price" 1, "other" 4. Hence, we know that the response "taste" and the response "price" are top priorities when consumers purchase a drink, the response "capacity" is third important factor, and the response "packaging" and the response "other" are relatively unimportant for consumers.

Next, we illustrate the ranking rule of the method with Bradley-Terry model. It is easier than above rule. In this method, we compute all $\boldsymbol{\gamma}$ values and according to the size of values, we can obtain an order with descending. Hence, the order is the rank of these responses.

4 Simulation

In this section, a simulation study is conducted to evaluate the performance of these methods in this section. Because the Bayesian ranking method has prior distribution assumption, it is different from other methods. Hence, we do not discuss Bayesian ranking method in the simulation study. The true rank of these responses is according to the order of $\pi_{(j)}$. In this study, we regard two responses have the same rank if $|\pi_{(i)} - \pi_{(j)}| \leq \epsilon$ where ϵ is a constant in a tolerance region. We compare the ranks of the 5 methods in terms of consistent rate, which is defined as the proportion that the rank of these methods and the true rank are consistent for n respondents in 1000 replicates. For example, let $\pi_{10000} = 0.032$, $\pi_{01000} = 0.015$, $\pi_{00100} = 0.087$, $\pi_{00010} = 0.061$, $\pi_{00001} = 0.009$, $\pi_{11000} = 0.008$, $\pi_{10100} = 0.0082$, $\pi_{10010} = 0.068$, $\pi_{10001} = 0.002$, $\pi_{01100} = 0.002$, $\pi_{01010} = 0.00005$, $\pi_{01001} = 0.0005$, $\pi_{00110} = 0.0005$, $\pi_{00101} = 0.00006$, $\pi_{00011} = 0.00319$,others equal to 0.044 and $\epsilon = 0.01$, resulting $\pi_1 = 0.626$, $\pi_2 = 0.501$, $\pi_3 = 0.585$, $\pi_4 = 0.617$, $\pi_5 = 0.479$. Thus the true is 1 4 3 1 5. We obtain a sample, and use the Wald test to obtain the rank 1 3 3 1 5. If the rank of each response derived from the the Wald test is smaller than the true rank, we call this phenomenon is consistent. Here, the result of the example is consistent. The codes to run γ in the Bradley-Terry models can be download in `http://sites.stat.psu.edu/~dhunter/code/btmatlab/`. We apply these programs to rank responses in a multiple response question. Since there are three codes for this method in `http://sites.stat.psu.edu/~dhunter/code/btmatlab/`. The first code of using Bradly-Terry with MM method is denoted as btmm in Tables 1-6 and in the R code section. The second code of using Bradly-Terry with quasi-Newton accelerated MM method is denoted as btqn in Tables 1-6 and in the R code section. The third code of using Bradly-Terry with a Newton-Raphson method is denoted as btnr in Tables 1-6 and in the R code section.

Then the following table is the consistent rate for different k and n:

Table 1: The consistent rates of the 5 methods when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.33,\ k$=5 and ϵ=0.05 for all methods and α=0.05 for the Wald test and the Generalized Score test:

Method / Sample size	Wald test	G.S. test	btmm	btqn	btnr
n=100	0.997	0.996	0.682	0.682	0.682
n=200	0.999	0.999	0.814	0.814	0.814
n=300	0.999	0.999	0.879	0.879	0.879
n=500	1	1	0.951	0.951	0.951
n=800	1	1	0.968	0.968	0.968
n=1000	1	1	0.992	0.992	0.992

Table 2: The consistent rates of the 5 methods when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ k$=6 and ϵ=0.05 for all methods and α=0.05 for the Wald test and the Generalized Score test:

Method / Sample size	Wald test	G.S. test	btmm	btqn	btnr
n=100	0.998	0.998	0.578	0.578	0.578
n=200	0.999	0.999	0.796	0.796	0.796
n=300	1	1	0.869	0.869	0.869
n=500	1	1	0.95	0.95	0.95
n=800	1	1	0.99	0.99	0.99
n=1000	1	1	0.992	0.992	0.992

Table 3: The consistent rates of the 5 methods when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ \pi_7 = 0.12,\ k$=7 and ϵ=0.05 for all methods and α=0.05 for the Wald test and the Generalized Score test:

Method / Sample size	Wald test	G.S. test	btmm	btqn	btnr
n=100	0.999	0.999	0.528	0.528	0.528
n=200	0.999	0.999	0.772	0.772	0.772
n=300	1	1	0.865	0.865	0.865
n=500	1	1	0.946	0.946	0.946
n=800	1	1	0.979	0.979	0.979
n=1000	1	1	0.99	0.99	0.99

Table 4: The consistent rates of the 5 methods when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ \pi_7 = 0.12,\ \pi_8 = 0.5$, k=8 and ϵ=0.05 for all methods and α=0.05 for the Wald test and the Generalized Score test:

Sample size \ Method	Wald test	G.S. test	btmm	btqn	btnr
n=100	0.999	0.998	0.355	0.355	0.355
n=200	0.999	0.999	0.626	0.626	0.626
n=300	1	1	0.796	0.796	0.796
n=500	1	1	0.91	0.91	0.91
n=800	1	1	0.972	0.972	0.972
n=1000	1	1	0.985	0.985	0.985

Table 5: The consistent rates of the 5 methods when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ \pi_7 = 0.12,\ \pi_8 = 0.5,\ \pi_9 = 0.9,\ \pi_{10} = 0.62$, k=10 and ϵ=0.05 for all methods and α=0.05 for the Wald test and the Generalized Score test:

Sample size \ Method	Wald test	G.S. test	btmm	btqn	btnr
n=100	0.998	0.998	0.251	0.251	0.251
n=200	0.999	0.999	0.524	0.524	0.524
n=300	1	1	0.684	0.684	0.684
n=500	1	1	0.879	0.879	0.879
n=800	1	1	0.957	0.957	0.957
n=1000	1	1	0.985	0.985	0.985

Next, we compare the Wald test and the Generalized Score test for different α. Then the following tables are the consistent rates of the Wald test and the Generalized Score test for different α:

Table 6: The consistent rates of the Wald test and the Generalized Score test when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ k$=5 and ϵ=0.05:

	Sample size			
	n=100		n=200	
Method / Significant level α	Wald test	G.S. test	Wald test	G.S. test
α=0.15	0.984	0.984	0.996	0.996
α=0.1	0.99	0.991	0.997	0.997
α=0.05	0.997	0.996	0.999	0.999
α=0.01	0.999	0.999	1	1

Table 7: The consistent rates of the Wald test and the Generalized Score test when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ \pi_7 = 0.12,\ k$=7 and ϵ=0.05:

	Sample size			
	n=100		n=200	
Method / Significant level α	Wald test	G.S. test	Wald test	G.S. test
α=0.15	0.992	0.992	0.997	0.997
α=0.1	0.994	0.994	0.997	0.997
α=0.05	0.999	0.999	0.999	0.999
α=0.01	1	1	1	1

Table 8: The consistent rates of the Wald test and the Generalized Score test when $\pi_1 = 0.77,\ \pi_2 = 0.28,\ \pi_3 = 0.56,\ \pi_4 = 0.21,\ \pi_5 = 0.34,\ \pi_6 = 0.43,\ \pi_7 = 0.12,\ \pi_8 = 0.5,\ \pi_9 = 0.9,\ \pi_{10} = 0.62$, k=10 and ϵ=0.05:

	Sample size			
	n=100		n=200	
Method / Significant level α	Wald test	G.S. test	Wald test	G.S. test
α=0.15	0.992	0.992	0.998	0.998
α=0.1	0.997	0.998	0.999	0.999
α=0.05	0.998	0.998	0.999	0.999
α=0.01	1	1	1	1

According to above results, we find that consistent rate decreases as the number of responses increase for the methods with Bradley-Terry model when n is not enough large. When the sample size is large, the results of these methods are almost consistent. In comparing the Wald test and the Generalized Score test for different α, consistent rate increases when α decreases. Although the consistent rate is not high for small sample size case, it still has good result for large sample size case. It reveals that these methods are feasible in ranking responses when the sample size is not small.

5 R code

These ranking procedures has been written as a package RankResponse for R. RankResponse is available from the Comprehensive R Archive Network at `http://CRAN.R-project.`

`org/package=RankResponse`, which include code function rank.wald, rank.gs, rank.LR, rank.LN, rank.L2R, rank.btmm, rank.btqn and rank.btnr.

---

rank.wald *Rank responses based on the Wald test*

Description

Rank responses of a single response question or a multiple response question by the Wald test procedure.

Usage

```
rank.wald(data,alpha,type=2)
```

Argument

`data` A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.

`alpha` The significance level used in the Wald test.

`type` type=1 for a single response question;
type=2 for a multiple response question.

Value

The rank.wald returns the estimated probabilities of the responses being selected and the ranks of the responses by the Wald test procedure.

References

Wang, H. (2008). Ranking Responses in Multiple-Choice Questions. Journal of Applied Statistics, 35, 465-474.

Examples

```
## This is an example to rank three responses in a multiple response
## question when the number of respondents is 1000 and the signifi-
## cance level is 0.05.In this example,we do not use a real data, but
## generate data in the first three lines.
A <-sample(c(0,1),1000,p=c(0.21,0.79),replace=T)
B <-sample(c(0,1),1000,p=c(0.86,0.14),replace=T)
C <-sample(c(0,1),1000,p=c(0.42,0.58),replace=T)
D <-cbind(A,B,C)
data <-matrix(D,nrow=1000,ncol=3)
# or upload the true data
alpha<-0.05
rank.wald(data,alpha,type=2)
```

---

rank.gs *Rank responses based on the Generalized score test*

Description

Rank responses of a single response question or a multiple response question by the generalized score test procedure.

Usage

```
rank.gs(data,alpha,type=2)
```

Argument

`data` A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.

`alpha` The significance level used in the Generalized score test.

`type` type=1 for a single response question ;

type=2 for a multiple response question .

Value

The rank.gs returns the estimated probabilities of the responses being selected and the ranks of the responses by the Generalized score procedure.

References

Wang, H. (2008). Ranking Responses in Multiple-Choice Questions. Journal of Applied Statistics, 35, 465-474.

Examples

```
## This is an example to rank three responses in a multiple response
## question when the number of respondents is 1000 and the signifi-
## cance level is 0.05.In this example,we do not use a real data, but
## generate data in the first three lines.
A <-sample(c(0,1),1000,p=c(0.21,0.79),replace=T)
B <-sample(c(0,1),1000,p=c(0.86,0.14),replace=T)
C <-sample(c(0,1),1000,p=c(0.42,0.58),replace=T)
D <-cbind(A,B,C)
data <-matrix(D,nrow=1000,ncol=3)
# or upload the true data
alpha<-0.05
rank.gs(data,alpha,type=2)
```

rank.btmm *Rank responses based on the Bradley-Terry model with the MM method*

Description

Adopt the Bradley-Terry model to rank responses in a single response question or in a multiple response question with the MM method. This method associates each response with a value γ, and use the γ value to rank responses.

Usage

```
rank.btmm(data)
```

Argument

`data` A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.

Value

The rank.btmm returns the associated γ values in the first line and the ranks ofthe responses in the second line.

References

Hunter DR (2004). MM algorithms for generalized Bradley-Terry models. The Annals of Statistics, 32, 384-406.

Examples

This is an example to rank three responses in a multiple response

question when the number of respondents is 1000. In this example,

we do not use a real data, but generate data in the first three lines.

A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)

```
B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)

C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)

D <-cbind(A,B,C)

data <-matrix(D,nrow=1000,ncol=3)

# or upload the true data

rank.btmm(data)
```

---

rank.btqn *Rank responses based on the Bradley-Terry model with the quasi-Newton accelerated MM method*

Description

Adopt the Bradley-Terry model to rank responses in a single response question orin a multiple response question with quasi-Newton and the MM method. This method associates each response with a value γ, and use the γ value to rank responses.

Usage

```
rank.btqn(data)
```

Argument

`data` A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.

Value

The rank.btqn returns the associated γ values in the first line and the ranks of the responses in the second line.

References

Hunter DR (2004). MM algorithms for generalized Bradley-Terry models. The Annals of Statistics, 32, 384-406.

Examples

```
## This is an example to rank three responses in a multiple response
## question when the number of respondents is 1000. In this example,
## we do not use a real data, but generate data in the first three lines.
A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)
B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)
C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)
D <-cbind(A,B,C)
data <-matrix(D,nrow=1000,ncol=3)
# or upload the true data
rank.btqn(data)
```

---

rank.btnr *Rank responses based on the Bradley-Terry model with Newton Raphson method*

Description

Adopt the Bradley-Terry model to rank responses in a single response question or in a multiple response question with Newton-Raphson method. This method associates each response with a value γ, and use the γ value to rank responses.

Usage

```
rank.btnr(data)
```

Argument

`data` A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.

Value

The rank.btnr returns the associated γ values in the first line and the ranks of the responses in the second line.

References

Hunter DR (2004). MM algorithms for generalized Bradley-Terry models. The Annals of Statistics, 32, 384-406.

Examples

```
## This is an example to rank three responses in a multiple response
## question when the number of respondents is 1000. In this example,
## we do not use a real data, but generate data in the first three lines.
A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)
B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)
C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)
D <-cbind(A,B,C)
data <-matrix(D,nrow=1000,ncol=3)
# or upload the true data
rank.btnr(data)
```

---

rank.LN *Rank responses under the Bayesian framework according to the loss function* $L_N(d, n) = c\overline{FD} + \overline{FN}$

Description

Rank responses of a single response question or a multiple response question under the Bayesian framework according to the loss function $L_N(d, n) = c\overline{FD} + \overline{FN}$.

Usage

```
LN(data,response.number,prior.parameter,c)
```

Argument

`data`	A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.
`response.number`	The number of the responses
`prior.parameter`	The parameter vector of the Dirichlet prior distribution, where the vector dimension is $2^{response.number}$.
`c`	The value of c in the loss function

Value

The rank.LN returns the estimated probabilities of the responses being selected in the first line and the ranks of the responses in the second line.

References

Wang, H. and Huang, W. H. (2014). Bayesian Ranking Responses in Multiple Response Questions. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177, 191-208.

Examples

##This is an example to rank three responses in a multiple response

```
##question when the number of respondents is 1000 and the value c is
##1. In this example, we do not use a real data, but generate data in
##the first three lines.
A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)
B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)
C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)
D <-cbind(A,B,C)
data <-matrix(D,nrow=1000,ncol=3)
# or upload the true data
response.number <-3
prior.parameter <- c(5,98,63,7,42,7,7,7)
c <-1
rank.LN(data,response.number,prior.parameter,c)
```

---

rank.LR *Rank responses under the Bayesian framework according to the loss* $L_R(d,n) = c\overline{FDR} + \overline{FNR}$

Description

Rank responses of a single response question or a multiple response question under the Bayesian framework according to the loss function $L_R(d,n) = c\overline{FDR} + \overline{FNR}$.

Usage

```
rank.LR(data,response.number,prior.parameter,c)
```

Argument

`data`	A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.
`response.number`	The number of the responses
`prior.parameter`	The parameter vector of the Dirichlet prior distribution, where the vector dimension is $2^{response.number}$.
`c`	The value of c in the loss function

Value

The rank.LR returns the estimated probabilities of the responses being selected in the first line and the ranks of the responses in the second line.

References

Wang, H. and Huang, W. H. (2014). Bayesian Ranking Responses in Multiple Response Questions. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177, 191-208.

Examples

```
##This is an example to rank three responses in a multiple response
##question when the number of respondents is 1000 and the value c is
##0.33. In this example, we do not use a real data, but generate data
##in the first three lines.
A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)
B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)
C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)
D <-cbind(A,B,C)
```

```
data <-matrix(D,nrow=1000,ncol=3)

# or upload the true data

response.number <-3

prior.parameter <- c(5,98,63,7,42,7,7,7)

c <-0.33

rank.LR(data,response.number,prior.parameter,c)
```

---

rank.L2R *Rank responses under the Bayesian framework according to the loss* $L_{2R}(d, n) = (\overline{FDR}, \overline{FNR})$

Description

Rank responses of a single response question or a multiple response question under the Bayesian framework according to the loss function $L_{2R}(d, n) = (\overline{FDR}, \overline{FNR})$.

Usage

```
rank.L2R(data,response.number,prior.parameter,e)
```

Argument

`data`	A $m \times n$ matrix (d_{ij}), where $d_{ij} = 0$ or 1. If the ith respondent selects the jth response, then $d_{ij} = 1$, otherwise $d_{ij} = 0$.
`response.number`	The number of the responses

`prior.parameter` The parameter vector of the Dirichlet prior distribution, where the vector dimension is $2^{response.number}$.

`e` A cut point used in the loss function which depends on the economic costs.

Value

The rank.L2R returns the estimated probabilities of the responses being selected in the first line and the ranks of the responses in the second line.

References

Wang, H. and Huang, W. H. (2014). Bayesian Ranking Responses in Multiple Response Questions. Journal of the Royal Statistical Society: Series A (Statistics in Society), 177, 191-208.

Examples

##This is an example to rank three responses in a multiple response

##question when the number of respondents is 1000 and the value e is

##0.15. In this example, we do not use a real data, but generate data

##in the first three lines.

A <-sample(c(0,1),1000,p=c(0.37,0.63),replace=T)

B <-sample(c(0,1),1000,p=c(0.71,0.29),replace=T)

C <-sample(c(0,1),1000,p=c(0.22,0.78),replace=T)

D <-cbind(A,B,C)

data <-matrix(D,nrow=1000,ncol=3)

or upload the true data

response.number <-3

```
prior.parameter <- c(5,98,63,7,42,7,7,7)

e <-0.15

rank.L2R(data,response.number,prior.parameter,e)
```

6 Conclusion

In this thesis, our goal is ranking responses of a single response question or a multiple response question. We proposed some methods to solve this problem. For ranking responses, the simulation results in Section 4 show that the proposed methods have good performance. Although these methods are not consistent for small sample size, their performances are very good for large sample size. In real applications, the sample size of responses is usually not very small, these methods are feasible in apply to real applications. According to the simulation results, the consistent rates of the Wald test and the generalized score test are larger than the method with Bradley-Terry model. Hence, we conclude that the Wald test and the generalized score test are more powerful than the method with Bradley-Terry model on ranking responses. We do not discuss Bayesian ranking responses method in the simulation study, because the other methods are under frequentist setup, but the Bayesian method is under the Bayesian framework. The codes of these methods has been written as an R package such that it is more convenient for readers to use them.

7 References

[1] Agresti, A. and Liu, I.-M. (1999) *Modeling a categorical variable allowing arbitrarily many category choices.* Biometrics 55, 936-943.

[2] Benjamini, Y. and Hochberg, Y. (1995) *Controlling the false discovery rates: a practical and powerful approach to multiple testing.* J. R. Statist. Soc. B, 57, 289—300.

[3] Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis.* 2nd edn. New York: Springer.

[4] Bilder, C. R., Loughin, T. M.and Nettleton, D. (2000) *Multiple marginal independence testing for pick any/c variables.* Communications in Statistics-Simulation and Computation, 29(4), 1285-1316.

[5] Bradley, R. A. and Terry, M. E. (1952). *Rank analysis of incomplete block designs. I. Themethod of paired comparisons.* Biometrika 39, 324—345.

[6] Decady, Y. J. and Thomas, D. H. (2000). *A simple test of association for contingency tables with multiple column responses.* Biometrics 56, 893-896.

[7] Hunter DR (2004), *MM algorithms for generalized Bradley-Terry models*, Annals of Statistics , 32, 386-408

[8] J.-L. Ke (2011) *Algorithms for Ranking Responses in Multiple Response Questions.* National Chiao Tung University

[9] Lange, K., Hunter, D.R., Yang I. (2000) *Optimization transfer using surrogate objective functions (with discussion).* J. Comput. Graphical Stat. 9: 1-59.

[10] Loughin, T. M. and Scherer, P. N. (1998). *Testing for association in contingency tables with multiple column responses.* Biometrics 54, 630-637.

[11] Umesh, U. N. (1995). *Predicting nominal variable relationships with multiple responses.* Journal of Forecasting 14, 585-596.

[12] Wang, H. (2008). *Ranking responses in multiple-choice questions.* Journal of Applied Statistics, 35, 465-474

[13] Wang, H. and Huang, W. H. (2014). *Bayesian Ranking Responses in Multiple Response Questions.* Journal of the Royal Statistical Society, Series A (Statistics in Society), 177, 191-208