R U T C O R

R E S E A R C H

R E P O R T

SEQUENTIAL OPTIMAL DESIGNS FOR ON-LINE ITEM CALIBRATION

Douglas. H. Jones[a] Mikhail Nediak[b]
Xiang-Bo Wang[c]

RRR 2-99, February, 1999

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/~rrr

[a] Department of Management Science and Information Systems, Rutgers, the State University of New Jersey, 180 University Av./Ackerson Hall, Newark, NJ 07102-1803, USA, dhjones@rci.rutgers.edu
[b] RUTCOR, Rutgers, The State University of New Jersey, 640 Bartholomew Road, Piscataway, NJ 08854-8003, USA, msnediak@rutcor.rutgers.edu
[c] Operations and Test Development Research, Law School Admissions Council, Box 40, 661 Penn Street, Newtown, PA, 18940-0040, USA.

RUTCOR RESEARCH REPORT
RRR 2-99, FEBRUARY, 1999

SEQUENTIAL OPTIMAL DESIGNS FOR ON-LINE ITEM CALIBRATION

Douglas H. Jones   Mikhail S. Nediak   Xiang-Bo Wang

Abstract. In an on-line adaptive test, data for calibrating new items are collected from examinees while they take an operational test. In this paper, we will assume a situation where a calibration session must collect data about several experimental items simultaneously. While past research has focused on estimation of one or more two-parameter logistic items, this research focuses on estimating several three-parameter logistic items simultaneously. We consider this problem in terms of constrained optimization over probability distributions. The probability distributions are over a two-by-two contingency table, and the marginal distributions form the constraints. We formulate these constraints for network-flow constraint, and investigate a conjugate-gradient-search algorithm to optimize the determinant of Fisher's information matrix.
Keywords and phrases: Optimal design, item spiraling, item response theory, psychological testing, computerized adaptive tasting, network flow mathematical programming, nonlinear response function.

Acknowledgements: This study received funding from the Law School Admission Council (LSAC). The opinions and conclusions contained in this report are those of the author and do not necessarily reflect the position or policy of LSAC. The work of D.H. Jones was partially supported by the Cognitive Science Program of the Office of the Chief of Naval Research under Contact N00012-87-0696.

1 Introduction

In an on-line adaptive test, data for calibrating new items are collected from examinees while they take an operational test. In this paper, we will assume a situation where a calibration session must collect data about several experimental items simultaneously. Other researchers have recognized the advantages of using optimal sampling designs to calibrate items (van der Linden, 1988) (Berger, 1991, 1992, 1994) (Berger and van der Linden, 1991) and (Jones and Jin, 1994). While past research has focused on estimation of one or more two-parameter logistic items, this research focuses on estimating several three-parameter logistic items simultaneously. This problem is considered as an unconstrained nonlinear mathematical programming model in Berger (1994). We consider this problem in terms of constrained optimization over probability distributions. The probability distributions are over a two-by-two contingency table, and the marginal distributions form the constraints. We formulate these constraints for network-flow constraints (Ahuja, Magnanti and Orlin, 1993), and investigate a conjugate-gradient-search algorithm to optimize the determinant of Fisher's information matrix (Murtaugh and Saunders, 1978).

In next section 2, we introduce D-Optimality. Section 3 states the mathematical program, constraints, and solution technique, namely projected conjugate-gradient. Section 4 described the sequential estimation schemes that employ D-optimal designs. Section 5 presents simulation results using actual LSAT item data, followed by the conclusion in Section 6. The gradient of the Fisher information matrix for the 3PL item is derived in Appendix 8.1. The projection on the null space of the constraint matrix is derived in Appendix 8.2.

2 D-optimal design criterion

Let u_i denote a response to a single item from individual i with ability level θ_i, possibly multivariate. Let $\boldsymbol{\beta}^T = (\beta_0, \beta_1 \ldots \beta_p)$ be a vector of unknown parameters associated with the item. Assume that all responses are scored either correct; $u_i = 1$, or incorrect, $u_i = 0$. An item response function, $P(\theta_i, \boldsymbol{\beta})$, is a function of θ_i, and describes the probability of correct response of an individual with ability level θ_i. The mean and variance of the parametric family are

$$E\{u_i|\boldsymbol{\beta}\} = P(\theta_i;\boldsymbol{\beta}) \tag{1}$$

$$\sigma^2(\theta_i;\boldsymbol{\beta}) = Var\{u_i|\boldsymbol{\beta}\} = P(\theta_i;\boldsymbol{\beta})\left[1-P(\theta_i;\boldsymbol{\beta})\right]. \tag{2}$$

We shall focus on the family of three-parameter logistic (3PL) response functions:

$$P(\theta_i;\boldsymbol{\beta}) = \beta_2 + (1-\beta_2)R(\theta_i;\boldsymbol{\beta}), \tag{3}$$

where

$$R(\theta_i;\boldsymbol{\beta}) = \frac{e^{\beta_0+\beta_1\theta}}{1+e^{\beta_0+\beta_1\theta}}. \tag{4}$$

Denote three IRT characteristics of a three-parameter item to be: the *discrimination power*, a = 1.7*β_1; the *difficulty*, $b = -\beta_0/\beta_1$, *guessing*, $c = \beta_2$. Note the family of two-parameter logistic (2PL)-response function is expression (3) $\beta_2 = 0$.

The expected information from ability θ is defined as Fisher's information matrix. The 3PL-information matrix is

$$\mathbf{m}(\boldsymbol{\beta};\theta) = \frac{\left[1-R(\theta;\boldsymbol{\beta})\right]^2}{P(\theta;\boldsymbol{\beta})\left[1-P(\theta;\boldsymbol{\beta})\right]}\begin{pmatrix} (1-\beta_2)^2\left[R(\theta;\boldsymbol{\beta})\right]^2 & (1-\beta_2)^2\left[R(\theta;\boldsymbol{\beta})\right]^2\theta & (1-\beta_2)R(\theta;\boldsymbol{\beta}) \\ (1-\beta_2)^2\left[R(\theta;\boldsymbol{\beta})\right]^2\theta & (1-\beta_2)^2\left[R(\theta;\boldsymbol{\beta})\right]^2\theta^2 & (1-\beta_2)R(\theta;\boldsymbol{\beta})\theta \\ (1-\beta_2)R(\theta;\boldsymbol{\beta}) & (1-\beta_2)R(\theta;\boldsymbol{\beta})\theta & 1 \end{pmatrix}. \tag{5}$$

We are interested in assigning test-takers to items so that the greatest about of information about the item parameters can be obtained. Assume point estimates of test-takers' ability are available, denoted by $\boldsymbol{\Theta} = \{\theta_i: i = 1,\ldots,n\}$. Assume initial point estimates of item vectors are available, denoted by $\mathbf{B} = \{\boldsymbol{\beta}_j: j = 1,\ldots,m\}$. Denote the information obtained from pairing ability θ_i with item vector $\boldsymbol{\beta}_j$ by

$$\mathbf{m}^{(i,j)} = \mathbf{m}\left(\boldsymbol{\beta}_j; \theta_i\right). \tag{6}$$

Introduce x_{ij} equal to the number of observations taken for item j from ability i. Then by the additive property of Fisher's information, the information matrix of item j is

$$\mathbf{M}^{(j)} \equiv \mathbf{M}^{(j)}\left(\boldsymbol{\beta}_j; \Theta, \mathbf{x}\right) = \sum_{i=1}^{n} x_{ij} \mathbf{m}^{(i,j)} \tag{7}$$

where $\mathbf{x} = (x_{11},\ldots, x_{1n}; x_{21},\ldots, x_{2n};\ldots; x_{m1},\ldots, x_{mn})$. If observations for different items are taken independently, the information matrix for all items taken together is

$$\mathbf{M}\left(\mathbf{B}; \Theta, \mathbf{x}\right) = diag\left\{\mathbf{M}^{(j)}\right\}. \tag{8}$$

We call a design X *exact* if x_{ij} is an integer for all i and j; otherwise, we call it *approximate*. An optimal design will typically be an approximate design.

The design criterion we wish to consider is

$$\log\det \mathbf{M}\left(\mathbf{B}; \Theta, \mathbf{x}\right) = \sum_{j=1}^{n} \log\det \mathbf{M}^{(j)}\left(\boldsymbol{\beta}_j; \Theta, \mathbf{x}\right). \tag{9}$$

For a single item problem, this criterion is the classical D-optimal criterion. For which there exist practical methods for deriving D-optimal designs (Donev and Atkinson, 1988), (Haines, 1987), (Mitchell, 1974), (Welch, 1982). See (Silvey, 1980) for other criteria. Approximate D-optimal designs are derived in (Ford, 1976) for a single item using the continuous design space [-1, 1], which are discrete, equal probability, two-point distributions with support points depending on the values of the item parameters. A simplification of the support points is approximately $\pm\min\{1, 1.5434(D\beta_1)^{-1}\}$ (Jones and Jin, 1994). In the present investigation, the design space is the set of all pairings between test-takers and items, and the design specifies which parings between test-takers and items will be observed. Jones, Chiang, and Jin (1997) derived solutions for this problem when the response functions were limited to the 2PL family.

3 Mathematical Program and Solution for optimal designs

Consider the problem of selecting x_{ij}, the number of observations taken for item j from ability i, to maximize information. A mathematical programming model for finding an optimal design with marginal constraints on $\mathbf{x} = \{x_{ij}: i = 1\ldots n; j = 1\ldots m\}$ is the following:

$$\textit{Maximize} \log \det \mathbf{M}(\mathbf{B};\Theta,\mathbf{x})$$

such that

$$\sum_{j=1}^{m} x_{ij} = s; i = 1,\ldots,n \tag{10}$$

$$\sum_{i=1}^{n} x_{ij} = d; j = 1,\ldots,m \tag{11}$$

$$x_{ij} \geq 0; i = 1,\ldots,n; j = 1,\ldots,m \tag{12}$$

where $md = ns$.

Constraints (10) stipulate that there is a supply of s test-takers with ability θ_i, and all these test-testers must receive an item. Constraints (11) stipulate that each item j demands d test-takers. Constraints (12) insure that the solution is logically consistent. The requirement $md = ns$ ensures that the total item demand is equal to the total supply of test-takers.

The constraint matrix corresponding to constraints (10) and (11) is derived as follows. Denote $\mathbf{I}_m$ and $\mathbf{I}_n$ as m by m and n by n identity matrices, respectively. In addition, denote $\mathbf{e}_m$ and $\mathbf{e}_n$ as m and n column vectors of all 1's. Define the following constraint matrix:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{I}_n \\ \mathbf{e}_n^T & 0 & \cdots & 0 \\ 0 & \mathbf{e}_n^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{e}_n^T \end{pmatrix} \tag{13}$$

Define,

$$\mathbf{b} = \begin{pmatrix} s\mathbf{e}_n \\ d\mathbf{e}_m \end{pmatrix}. \tag{14}$$

Then the

$$\mathbf{Ax} = \mathbf{b} \tag{15}$$

is a restatement of the constraints (10) and (11). A point $\mathbf{x} \geq \mathbf{0}$ is feasible if it satisfies (15). If $\mathbf{x} > \mathbf{0}$ is feasible and α is an arbitrary sufficiently small scalar, then, a point $\mathbf{x} + \alpha\mathbf{r}$ is feasible if and only if $\mathbf{Ar} = 0$; that is, if $\mathbf{r}$ is in the null space of $\mathbf{A}$.

The objective function for this problem is known to be concave in $\mathbf{x}$ (Federov, 1972). We use the conjugate-gradient methods for linearly constrained problems and logarithmic penalty functions for the non-negativity constraints (12) to get approximate designs (for an overview of optimality procedures, see Gill, Murray and Wright, 1981). Thus, our objective function becomes:

$$F(\mathbf{x}) \equiv F_\mu(\mathbf{x}) = \log\det \mathbf{M}(\mathbf{B};\Theta,\mathbf{x}) + \mu \sum_{i,j} \log x_{ij} \tag{16}$$

where μ is a penalty parameter; this parameter is sequentially reduced to a sufficiently small value.

The unconstrained conjugate-gradient method is described as follows. Assume that $\mathbf{x}_k$ is an approximation to the maximum of F. Given a direction of search $\mathbf{p}_k$, let

$$\alpha^* = \arg\max_\alpha F(\mathbf{x}_k + \alpha\mathbf{p}_k). \tag{17}$$

The next approximation to the maximum of F is

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha^*\mathbf{p}_k. \tag{18}$$

Denote the gradient of F at $\mathbf{x}_k$ as $\mathbf{g}(\mathbf{x}_k)$ (see Appendix 8.1 for its derivation). The direction of search, obtained from the method of conjugate-gradient, is defined as:

$$\mathbf{p}_k = \mathbf{g}(\mathbf{x}_k) + \mathbf{p}_{k-1} \frac{\mathbf{g}(\mathbf{x}_k)^T [\mathbf{g}(\mathbf{x}_k) - \mathbf{g}(\mathbf{x}_{k-1})]}{\|\mathbf{g}(\mathbf{x}_{k-1})\|^2}. \tag{19}$$

Where $\mathbf{p}_{k\text{-}1}$ and $\mathbf{x}_{k\text{-}1}$ denote the previous search direction and approximation, respectively. The conjugate-gradient must be restarted in the direction of steepest ascent every *nm* steps.

The constrained conjugate-gradient method is an extension of the foregoing method. The linearly constrained problem uses the projection of the direction (19) on the null space of the constraint matrix (13). Denote **p** as an *nm* dimensional direction vector. In the appendix we derive the following simple expression for the ij^{th} component of the projection **r** of **p** on {**x**: **Ax** = 0}:

$$r_{ij} = p_{ij} - \frac{1}{m}\sum_{l=1}^{m} p_{il} - \frac{1}{n}\sum_{k=1}^{n} p_{kj} + \frac{1}{mn}\sum_{k,l} p_{kl} \,. \tag{20}$$

This result is not very surprising in light of the theory of analysis of variance in statistics using p_{ij} for data, as these elements are known as *interactions* in the two-way model with main effects for item and ability. They are the least-squares residuals from the fitted values in the main effects model, and thus must lie in the null space of the constraints (10) and (11). Note that searching from a feasible point in the direction of **r** ensures feasibility of the solution.

4 Sequential Estimation of Item Parameters

In item response theory, optimal designs depend on knowing the item parameters. For obtaining designs for the next stage of observations, a sequential estimation scheme supplies item parameters that should converge to the true parameters. By continually improving the item parameters, the solutions of mathematical program yield better designs for obtaining data that are more efficient.

We have had experience with maximum likelihood and Baysian methods. Application of both maximum likelihood and Baysian methods are straightforward. The choice between these two methods may be decided from numerical considerations. Maximum likelihood may be unstable for the beginning of the sequential procedure, as the likelihood is nearly flat. Baysian methods enjoy stability, but the choice of a prior may be a difficult. In the following section, we study the performance of optimal design methods using Baysian estimators incorporating uninformative and informative priors.

5 Computational Results

We present some results using actual data from paper-and-pencil administration of the LSAT using the 3PL model. The data consist of the responses to 101 items from 1600 test-takers. We use this data to simulate an on-line item calibration. Five items will be calibrated. Batches of 40 test-takers are randomly chosen. Each item receives five test-takers. No test-taker receives more than one item. Initial estimates of $\boldsymbol{\beta}$ are fed to the mathematical program and a design is derived. The design specifies what records of test-takers are used to estimate the parameters of each item. This process continues for 40 batches of test-takers. Then another set of items is chosen for the next calibration round, until all items have been calibrated. The following is a summary of the simulation components:

5.1 SIMULATION STRATEGY

Items are calibrated in sets of 5 items. Each run of 40 test-takers yields 8 new observation for each item. After each run, estimates of item parameters are updated. After the initial run, each run is designed to obtain D-optimal Fisher information for item parameters. A total of 40 runs were performed, resulting in 320 observations per item

5.2 ESTIMATION STRATEGIES

Strategy R: Uniform priors; 40 runs random BB design

Strategy A: Uniform priors; one run random BB design; 39 runs optimal design

Strategy D: Priors set equal to posteriors from 40 runs BB design; 40 runs optimal design

Random BB denotes a balanced block design, where the test-takers are blocked into eight groups according to ability and then assigned randomly to items. An optimal design is derived using Bayes EAP estimators based on the accumulated data and designated prior.

5.3 NOTATION

E denotes the estimation error of β_0, β_1, β_2. The numbers 0, 1, 2 identify the parameters β_0, β_1, β_2. R, A, D identifies the estimation strategy, e.g. EA0, EA1, EA2 are estimation errors with strategy A

In addition, we use a measure of overall fit of the item response function. The Hellinger distance between two discrete probability distributions $f_1(x)$ and $f_2(x)$ is

$$H = \sum_{i=1}^{\infty} \left(f_1^{1/2}(x_i) - f_2^{1/2}(x_i) \right)^2 .$$

In our case, if we have two parameter estimates $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$, the Hellinger distance between corresponding IRF's at a particular ability level θ can be computed as (after a simple transformation):

$$H_\theta(\boldsymbol{\beta}, \boldsymbol{\beta}') = 2\left(1 - \sqrt{P(\theta, \boldsymbol{\beta}) P(\theta, \boldsymbol{\beta}')}\right).$$

The weighted integral Hellinger distance between IRF's is obtained by integrating out θ with respect to some suitable weight function $w(\theta)$*:*

$$H(\boldsymbol{\beta}, \boldsymbol{\beta}') = 2\int_{-\infty}^{+\infty} \left(1 - \sqrt{P(\theta, \boldsymbol{\beta}) P(\theta, \boldsymbol{\beta}')}\right) w(\theta) d\theta$$

In this particular case we have chosen $w(\theta)$ to be proportional to the probability density function of N(1,1) distribution since deviations over this distribution of ability would be most important.

These will be denoted as H with designation for estimation strategy; e.g., HD is the Hellinger distance associated with strategy D.

5.4 RESULTS

The true parameters are derived from estimation with all 1600 records for each item. The sequential estimates are based on 320 observations each, derived as explained above. In general, 320 random observations would not yield very satisfactory parameter estimates. However, the use of sequentially derived optimal designs increases the effectiveness of the relatively small number of observations, as can be seen with the percentiles of estimation errors displayed in Table 1.

Table 2 displays the percentiles of Hellinger distance over the three estimation strategies. Strategy D appears to fit better than either A or D.

Table 3 displays estimation errors and Hellinger distances of the two extreme items of each estimation strategies. Figure 1 displays the true and estimated item response functions of the these items.

6 Conclusion

D-optimal designs improve item calibration over the random design considered in this paper. The estimation is better with informative priors. Linear hierarchical priors may be employed to obtain priors that are more informative. These may be obtained from past calibrations and, possibly, information about the item. Other design criteria could be employed, e.g. Buyske (1998a,b).

Clearly, some items were calibrated more effectively than others. In actual implementation of on-line item calibration, the more easily calibrated items could be removed from sampling earlier. Computational times for deriving optimal designs were well within bounds for implementation in an actual setting.

References

Ahuja, R. K., Magnanti, T. L., & Orlin, J. B. (1993). *Network flows: Theory, algorithms, and applications*. Englewood Cliffs, NJ: Prentice-Hall.

Berger, M. P. F. (1991). On the efficiency of IRT models when applied to different sampling designs. *Applied Psychological Measurement*, *15*, 283-306.

Berger, M. P. F. (1992). Sequential sampling designs for the two-parameter item response theory model. *Psychometrika*, *57*, 521-538.

Berger, M. P. F. (1994). D-Optimal sequential designs for item response throe models. *Journal of Educational Statistics*, *19*, 43-56.

Berger, M. P. F., & van der Linden, W. J. (1991). Optimality of sampling designs in item response theory models. In M. Wilson (Ed.), *Objective measurement: Theory into practice.* Norwood, NJ: Ablex Publishing.

Buyske, S. G. (1998a). Item calibration in computerized adaptive testing using minimal information loss. Preprint.

Buyske, S. G. (1998b). *Optimal designs for item calibration in computerized adaptive testing*. Unpublished doctoral dissertation, Rutgers University of New Jersey.

Donev, A. N. & Atkinson, A. C. (1988). An adjustment algorithm for the construction of exact D-optimum experimental designs. *Technometrics*, *30*, 429-433.

Federov, V. V. (1972). *Theory of optimal experiments.* New York: Academic Press.

Ford, I. (1976). *Optimal static and sequential design: A critical review.* Unpublished doctoral dissertation, University of Glasgow.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization.* London: Academic Press.

Haines, L. M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models. *Technometrics*, *29,* 439-447.

Jones, D. H., & Jin, Z. (1994). Optimal sequential designs for on-line item estimation. *Psychometrika, 59*, 59-75.

Jones, D. H., Chiang, J, & Jin, Z. (1997). Optimal designs for simultaneous item estimation. *Nonlinear Analysis, Theory, Methods & Applications, 30,* 4051-4058.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillside, NJ: Erlbaum.

Mitchell, T. J. (1974). An algorithm for the construction of D-optimal experimental designs. *Technometrics*, *16*, 203-210.

Murtaugh, B., & Saunders, M. (1978). Large scale linearly constrained optimization. *Math. Program.*, *14,* 42-72.

Nemhauser, G., & Wolsey, L. (1988). *Integer and combinatorial optimization.* New York: Wiley.

Prakasa Rao, B. L. S. (1987). *Asymptotic theory of statistical inference.* New York: Wiley.

Silvey, S. D. (1980). *Optimal design.* New York: Chapman and Hall.

Van der Linden, W. J. (1988). *Optimizing incomplete sampling designs for item response model parameters* (Research Report No. 88-5). Enschede, The Netherlands: University of Twente.

Welch, W. J. (1982). Branch-and-bound search for experimental designs based on D optimality and other criteria. *Technometrics*, *24,* 41-48.

Appendix

1. DERIVATION OF THE GRADIENT

In the following, we derive expressions for the gradient. In general, consider the matrix

$$\mathbf{H}(\mathbf{y}) = \sum_{j=1}^{N} y_j \mathbf{h}^{(j)} , \tag{21}$$

where $h^{(j)}$ is a k by k matrix. Let $(i_1,\ldots,i_k)$ denote the transposition of the numbers from 1 to k and $\sigma(i_1,\ldots,i_k)$ its sign. By definition,

$$\det \mathbf{H}(\mathbf{y}) = \sum_{(i_1,\ldots,i_k)} (-1)^{\sigma(i_1,\ldots,i_k)} \left[\left(\sum_{j_1=1}^{N} y_{j_1} \mathbf{h}_{1i_1}^{(j_1)} \right) \cdots \left(\sum_{j_k=1}^{N} y_{j_k} \mathbf{h}_{1i_k}^{(j_k)} \right) \right]. \tag{22}$$

Applying the chain rule for differentation,

$$\begin{aligned}
\frac{\partial \det \mathbf{H}(\mathbf{y})}{\partial y_j} &= \sum_{(i_1,\ldots,i_k)} (-1)^{\sigma(i_1,\ldots,i_k)} \left\{ \sum_{l-1}^{k} \left[\left(\sum_{j_1=1}^{N} y_{j_1} \mathbf{h}_{1i_1}^{(j_1)} \right) \cdots g_{li_l}^{(j)} \cdots \left(\sum_{j_k=1}^{N} y_{j_k} \mathbf{h}_{1i_k}^{(j_k)} \right) \right] \right\} \\
&= \sum_{l=1}^{k} \sum_{(i_1,\ldots,i_k)} (-1)^{\sigma(i_1,\ldots,i_k)} \left[\left(\sum_{j_1=1}^{N} y_{j_1} \mathbf{h}_{1i_1}^{(j_1)} \right) \cdots g_{li_l}^{(j)} \cdots \left(\sum_{j_k=1}^{N} y_{j_k} \mathbf{h}_{1i_k}^{(j_k)} \right) \right] \\
&= \sum_{l=1}^{k} \det \mathbf{C}^{(l,j)},
\end{aligned} \tag{23}$$

where $\mathbf{C}^{(l,j)}$ is a k by k matrix of the form

$$\mathbf{C}^{(l,j)} = \begin{pmatrix} \mathbf{H}_{11}(\mathbf{y}) & \cdots & \mathbf{H}_{1k}(\mathbf{y}) \\ \vdots & & \vdots \\ \mathbf{h}_{l1}^{j} & \cdots & \mathbf{h}_{lk}^{j} \\ \vdots & & \vdots \\ \mathbf{H}_{k1}(\mathbf{y}) & \cdots & \mathbf{H}_{kk}(\mathbf{y}) \end{pmatrix}. \tag{24}$$

Applying this result to our objective function, we get:

$$\frac{\partial \log \det \mathbf{M}(\mathbf{B};\Theta,\mathbf{X})}{\partial x_{ij}} = \frac{1}{\det \mathbf{M}^{(j)}} \left\{ \begin{array}{c} \det\begin{pmatrix} \mathbf{m}_{11}^{(i,j)} & \mathbf{m}_{12}^{(i,j)} & \mathbf{m}_{13}^{(i,j)} \\ \mathbf{M}_{21}^{(j)} & \mathbf{M}_{22}^{(j)} & \mathbf{M}_{23}^{(j)} \\ \mathbf{M}_{31}^{(j)} & \mathbf{M}_{32}^{(j)} & \mathbf{M}_{33}^{(j)} \end{pmatrix} + \det\begin{pmatrix} \mathbf{M}_{11}^{(j)} & \mathbf{M}_{12}^{(j)} & \mathbf{M}_{13}^{(j)} \\ \mathbf{m}_{21}^{(i,j)} & \mathbf{m}_{22}^{(i,j)} & \mathbf{m}_{23}^{(i,j)} \\ \mathbf{M}_{31}^{(j)} & \mathbf{M}_{32}^{(j)} & \mathbf{M}_{33}^{(j)} \end{pmatrix} + \\ + \det\begin{pmatrix} \mathbf{M}_{11}^{(j)} & \mathbf{M}_{12}^{(j)} & \mathbf{M}_{13}^{(j)} \\ \mathbf{M}_{21}^{(j)} & \mathbf{M}_{22}^{(j)} & \mathbf{M}_{123}^{(j)} \\ \mathbf{m}_{31}^{(i,j)} & \mathbf{m}_{32}^{(i,j)} & \mathbf{m}_{33}^{(i,j)} \end{pmatrix} \end{array} \right\}. \tag{25}$$

2. PROJECTIONS ON THE NULL SPACE OF THE CONSTRAINT MATRIX

Denote $\mathbf{p}$ as an nm dimensional direction vector. Denote $\mathbf{I}_m$ and $\mathbf{I}_n$ as m by m and n by n identity matrices, respectively. In addition, denote $\mathbf{e}_m$ and $\mathbf{e}_n$ as m and n column vectors of all 1's. Define the following constraint matrix:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{I}_n \\ \mathbf{e}_n^T & 0 & \cdots & 0 \\ 0 & \mathbf{e}_n^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{e}_n^T \end{pmatrix} \tag{26}$$

The projection $\mathbf{r}$ of $\mathbf{p}$ on $\{\mathbf{y}: \mathbf{Ay} = 0\}$, the null space of $\mathbf{A}$, is given by

$$\mathbf{r} = \mathbf{p} - \mathbf{A}^T\mathbf{y}, \tag{27}$$

where $\mathbf{y}$ is a solution to

$$\left(\mathbf{A}\mathbf{A}^T\right)\mathbf{y} = \mathbf{Ap}. \tag{28}$$

Introduce the following notation:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}; \mathbf{Ap} = \begin{pmatrix} \tilde{\mathbf{p}}_1 \\ \tilde{\mathbf{p}}_2 \end{pmatrix}; \mathbf{y}_1, \tilde{\mathbf{p}}_1 \in \mathbf{R}^n; \mathbf{y}_2, \tilde{\mathbf{p}}_2 \in \mathbf{R}^m. \tag{29}$$

We have the following simplifications:

$$\tilde{\mathbf{p}}_1 = \left(p_{1\bullet}, \ldots, p_{n\bullet}\right)^T; \tilde{\mathbf{p}}_2 = \left(p_{\bullet 1}, \ldots, p_{\bullet m}\right)^T, \tag{30}$$

where

$$p_{i\bullet} = \sum_{j=1}^{m} p_{ij}\,;\, p_{\bullet j} = \sum_{i=1}^{n} p_{ij}\,, \tag{31}$$

and

$$\mathbf{A}\mathbf{A}^{T} = \begin{pmatrix} m\mathbf{I}_n & \mathbf{E}^{T} \\ \mathbf{E} & n\mathbf{I}_m \end{pmatrix}, \tag{32}$$

where E is an m by n matrix of all ones. Consequently, equation (28) can be written as:

$$\begin{aligned} m\mathbf{y}_1 + \mathbf{e}_n\left(\mathbf{e}_m^{T}\mathbf{y}_2\right) &= \tilde{\mathbf{p}}_1 \\ \mathbf{e}_m\left(\mathbf{e}_n^{T}\mathbf{y}_1\right) + n\mathbf{y}_2 &= \tilde{\mathbf{p}}_2 \end{aligned} \tag{33}$$

implying:

$$\begin{aligned} \frac{\mathbf{e}_n^{T}\mathbf{y}_1}{n} + \frac{\mathbf{e}_m^{T}\mathbf{y}_2}{m} &= \frac{\sum_{i,j} p_{ij}}{nm} \\ \mathbf{y}_1 &= \frac{1}{m}\left[\tilde{\mathbf{p}}_1 - \mathbf{e}_n\left(\mathbf{e}_m^{T}\mathbf{y}_2\right)\right] \\ \mathbf{y}_2 &= \frac{1}{n}\left[\tilde{\mathbf{p}}_2 - \mathbf{e}_m\left(\mathbf{e}_n^{T}\mathbf{y}_1\right)\right] \end{aligned} \tag{34}$$

Thus

$$\begin{aligned} \mathbf{A}^{T}\mathbf{y} = \mathbf{A}^{T}\begin{pmatrix} \frac{1}{m}\tilde{\mathbf{p}}_1 \\ \frac{1}{n}\tilde{\mathbf{p}}_2 \end{pmatrix} - \mathbf{A}^{T}\begin{pmatrix} \mathbf{e}_n\left(\frac{\mathbf{e}_m^{T}\mathbf{y}_2}{m}\right) \\ \mathbf{e}_m\left(\frac{\mathbf{e}_n^{T}\mathbf{y}_1}{n}\right) \end{pmatrix} = \\ = \mathbf{A}^{T}\begin{pmatrix} \frac{1}{m}\tilde{\mathbf{p}}_1 \\ \frac{1}{n}\tilde{\mathbf{p}}_2 \end{pmatrix} - \mathbf{e}_{n+m}\left(\frac{\mathbf{e}_m^{T}\mathbf{y}_2}{m} + \frac{\mathbf{e}_n^{T}\mathbf{y}_1}{n}\right) \end{aligned} \tag{35}$$

Denote $\mathbf{z} = \mathbf{A}^{T}\mathbf{y}$, then by (35),

$$z_{ij} = \frac{1}{m}p_{i\bullet} + \frac{1}{n}p_{\bullet j} - \frac{1}{nm}\sum_{k,l} p_{kl}\,. \tag{36}$$

Since **r** = **p - z**, we have

$$\mathbf{r}_{ij} = p_{ij} - \frac{1}{m} p_{i\bullet} - \frac{1}{n} p_{\bullet j} + \frac{1}{nm} \sum_{k,l} p_{kl} \,. \tag{37}$$

Tables

Table 1. Percentiles of estimation errors over three estimation strategies.

		Percentiles						
		5	10	25	50	75	90	95
Estimation Errors	ER0	-.58	-.48	-.27	-.08	.12	.36	.46
	EA0	-.79	-.67	-.39	-.08	.13	.30	.38
	ED0	-.71	-.50	-.27	-.05	.12	.25	.37
	ER1	-.25	-.19	-.09	.03	.18	.33	.45
	EA1	-.25	-.22	-.11	.06	.22	.32	.44
	ED1	-.25	-.17	-.08	.06	.22	.36	.45
	ER2	-.07	-.05	-.01	.02	.08	.11	.16
	EA2	-.05	-.04	-.01	.03	.07	.16	.19
	ED2	-.07	-.04	-.02	.01	.07	.13	.15

Table 2.
Percentiles of Hellinger distance over three estimation strategies.

		Percentiles						
		5	10	25	50	75	90	95
Hellinger	HR	0.00+	0.00+	0.00+	0.00+	0.01	0.02	0.04
	HA	0.00+	0.00+	0.00+	0.00+	0.02	0.04	0.06
	HD	0.00+	0.00+	0.00+	0.00+	0.01	0.02	0.03

Table 3. Estimation errors and Hellinger distances of extreme items of each estimation strategy.

	ITEM67	ITEM25	ITEM53	ITEM100	ITEM57
β_0	-2.47	-0.32	0.17	-0.52	-0.12
ER0	**0.99**	**-1.48**	-0.29	-0.49	-0.57
EA0	0.28	**-1.20**	**-0.92**	-0.79	-0.41
ED0	0.02	-0.37	-0.20	**-0.84**	**-1.13**
β_1	1.74	0.72	0.95	0.73	0.49
ER1	**-0.57**	**0.23**	0.41	0.19	0.10
EA1	-0.25	**0.31**	**0.23**	0.50	0.13
ED1	0.10	0.27	0.00	**0.57**	**0.36**
β_2	0.24	0.11	0.10	0.18	0.20
ER2	**-0.09**	**0.22**	0.06	0.06	0.20
EA2	0.00	**0.19**	**0.25**	0.15	0.14
ED2	-0.01	0.13	0.07	**0.17**	**0.26**
HR	**0.04**	**0.09**	**0.02**	**0.02**	**0.02**
HA	**0.00+**	**0.08**	**0.08**	**0.07**	**0.02**
HD	**0.00+**	**0.03**	**0.00+**	**0.08**	**0.09**

Figure 1. Comparison of true and estimated item response functions of extreme items of each estimation strategies