A Maximum Likelihood Based Offline Estimation of Student Capabilities and Question Difficulties

Shana Moothedath

Prasanna Chaporkar &

Madhu N. Belur,

Department of Electrical Engineering,

Indian Institute of Technology Bombay

15-Mar-16

Introduction

1. Exam based assessment
   1. subjective exam
   2. objective exam
   3. Computerized adaptive exam (CAT) and offline exam
2. Psychometric test analysis
   1. Classical test theory (CTT)$\rightarrow$ true score.
   2. Item response theory (IRT)[1]$\rightarrow$ examinee and item characteristics

[1] Baker, The Basics of Item Response Theory, 1985

Introduction

1. Exam based assessment
   1. subjective exam
   2. objective exam
   3. Computerized adaptive exam (CAT) and offline exam
2. Psychometric test analysis
   1. Classical test theory (CTT)$\rightarrow$ true score.
   2. Item response theory (IRT)[1]$\rightarrow$ examinee and item characteristics

[1] Baker, The Basics of Item Response Theory, 1985

Item response theory (IRT)

1. Item based test theory
2. Start marked in 1916 by Binet-Simon.
3. Item characteristic curve (ICC):
   Functional relationship between probability of correct response to an item and a criterion variable.
   1. Normal ogive model
   2. Logistic ogive model
4. Maximum likelihood estimates of the parameters of ICC [2].

[2] Lawley, Proc. of the Royal Society of Edinburgh, 1943

Item response theory (IRT)

1. Item based test theory
2. Start marked in 1916 by Binet-Simon.
3. Item characteristic curve (ICC):
   Functional relationship between probability of correct response to an item and a criterion variable.
   1. Normal ogive model
   2. Logistic ogive model
4. Maximum likelihood estimates of the parameters of ICC [2].

[2] Lawley, Proc. of the Royal Society of Edinburgh, 1943

Item response theory (IRT)

1. Item based test theory
2. Start marked in 1916 by Binet-Simon.
3. Item characteristic curve (ICC):
   Functional relationship between probability of correct response to an item and a criterion variable.
   1. Normal ogive model
   2. Logistic ogive model
4. Maximum likelihood estimates of the parameters of ICC [2].

[2] Lawley, Proc. of the Royal Society of Edinburgh, 1943

Motivation

1. CAT
   1. Pre-caliberated pool of questions
   2. Self-tailored set of questions
   3. optimal test.
2. Offline exams
   1. All examinees answer same set of questions
   2. Few questions → erroneous estimation
   3. Long exam → fatigue, guessing → skewed result
   4. Multiple session exam → students answer different question paper for same discipline
   5. Score comparison across years and disciplines.

Motivation

1. CAT
   1. Pre-caliberated pool of questions
   2. Self-tailored set of questions
   3. optimal test.
2. Offline exams
   1. All examinees answer same set of questions
   2. Few questions → erroneous estimation
   3. Long exam → fatigue, guessing → skewed result
   4. Multiple session exam → students answer different question paper for same discipline
   5. Score comparison across years and disciplines.

IRT Model

1. Logistic ogive model

$$P_i(c_j) = P(d_i, a_i, c_j) = \frac{e^{a_i(c_j - d_i)}}{1 + e^{a_i(c_j - d_i)}}, \qquad (1)$$

2. Parameters:
   - Capability c_j
   - Difficulty d_i
   - Discrimination a_i
3. Objective: Given the response matrix, estimate C, D

IRT Model

1. Logistic ogive model

$$P_i(c_j) = P(d_i, a_i, c_j) = \frac{e^{a_i(c_j - d_i)}}{1 + e^{a_i(c_j - d_i)}}, \tag{1}$$

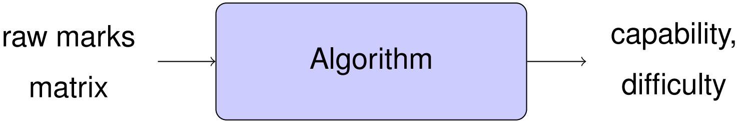
2. Parameters:
   - Capability c_j
   - Difficulty d_i
   - Discrimination a_i
3. Objective: Given the response matrix, estimate C, D

IRT Model

1. Logistic ogive model

$$P_i(c_j) = P(d_i, a_i, c_j) = \frac{e^{a_i(c_j - d_i)}}{1 + e^{a_i(c_j - d_i)}}, \tag{1}$$

2. Parameters:
   - Capability c_j
   - Difficulty d_i
   - Discrimination a_i
3. Objective: Given the response matrix, estimate C, D



Item characteristic curve (ICC)

1. Question difficulty d_i

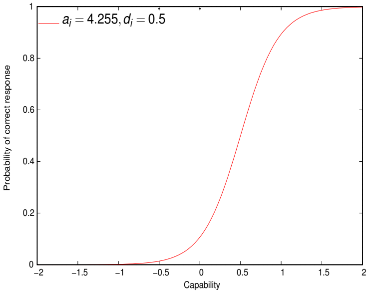


Figure : ICC for correct response with $d_i = 0.5$ and $a_i = 4.255$

Item characteristic curve (ICC)

1. Student capability c_j

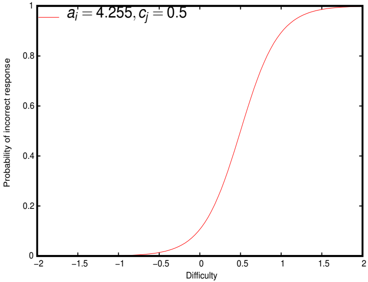


Figure : ICC for incorrect response with $a_i = 4.255$ and $c_j = 0.5$.

Item characteristic curve (ICC)

1. Question discrimination a_i

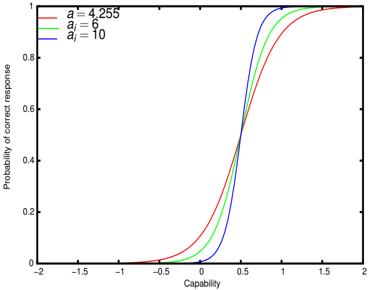


Figure : ICC for correct response with $c_j = 0.5$ and $d_i = 0.5$.

Maximum Likelihood based estimation

1. Likelihood of a set of parameter values, θ, given outcomes X, is the probability of those observed outcomes given those parameter values.

$$L(\theta|X) = P(X|\theta). \tag{2}$$

2. Maximum likelihood
   1. Given the response matrix R, estimate capability vector $C = [c_1 \dots c_{n_S}]$, difficulty vector $D = [d_1 \dots d_{n_Q}]$ and discrimination vector $A = [a_1 \dots a_{n_Q}]$.
   2. Likelihood function:

$$L = \mathrm{Prob}(\mathrm{R}|\ \mathrm{C},\mathrm{D},\mathrm{A}). \tag{3}$$

Maximum Likelihood based estimation

1. Likelihood of a set of parameter values, θ, given outcomes X, is the probability of those observed outcomes given those parameter values.

$$L(\theta|X) = P(X|\theta). \tag{2}$$

2. Maximum likelihood
   1. Given the response matrix R, estimate capability vector $C = [c_1 \dots c_{n_S}]$, difficulty vector $D = [d_1 \dots d_{n_Q}]$ and discrimination vector $A = [a_1 \dots a_{n_Q}]$.
   2. Likelihood function:

$$L = \mathrm{Prob}(\mathrm{R}|\ \mathrm{C}, \mathrm{D}, \mathrm{A}). \tag{3}$$

Maximum Likelihood based estimation

1. Assumptions
   1. All examinees are independent
   2. All test items are modelled by ICC of the same family
2. Likelihood function for the exam:

$$\mathrm{Prob}(R) = \Pi_{j=1}^{n_S} \Pi_{i=1}^{n_Q} P_{ij}^{m_{ij}} (1 - P_{ij})^{1-m_{ij}} \quad (4)$$

3. Log-ikelihood function:

$$L(c_j, a_i, d_i) = \sum_{j=1}^{n_S} \sum_{i=1}^{n_Q} [m_{ij} a_i (c_j - d_i) - \log(1 + e^{a_i(c_j - d_i)})] \quad (5)$$

Maximum Likelihood based estimation

1. Assumptions
   1. All examinees are independent
   2. All test items are modelled by ICC of the same family
2. Likelihood function for the exam:

$$\text{Prob}(R) = \Pi_{j=1}^{n_S} \Pi_{i=1}^{n_Q} P_{ij}^{m_{ij}} (1 - P_{ij})^{1-m_{ij}} \quad (4)$$

3. Log-ikelihood function:

$$L(c_j, a_i, d_i) = \sum_{j=1}^{n_S} \sum_{i=1}^{n_Q} [m_{ij} a_i (c_j - d_i) - \log(1 + e^{a_i(c_j - d_i)})] \quad (5)$$

Maximum Likelihood based estimation

1. Assumptions
   1. All examinees are independent
   2. All test items are modelled by ICC of the same family
2. Likelihood function for the exam:

$$\text{Prob}(R) = \Pi_{j=1}^{n_S} \Pi_{i=1}^{n_Q} P_{ij}^{m_{ij}} (1 - P_{ij})^{1-m_{ij}} \tag{4}$$

3. Log-ikelihood function:

$$L(c_j, a_i, d_i) = \sum_{j=1}^{n_S} \sum_{i=1}^{n_Q} [m_{ij} a_i (c_j - d_i) - \log(1 + e^{a_i(c_j - d_i)})] \tag{5}$$

Is marks a good estimate of capability?

Lemma

If all questions are of same discrimination a, then total marks is the maximum likelihood estimate of student capability.

$$\sum_{i=1}^{n_Q} m_{ij} = \sum_{i=1}^{n_Q} \frac{e^{a(c_j - d_i)}}{(1 + e^{a(c_j - d_i)})} \tag{6}$$

Alternating optimization

1. Alternating optimization

$$\max_{U,V,W} f(U,V,W) \quad (7)$$

- fix U, V, optimize for W $\rightarrow$ $arg \max_{W} f(U^t, V^t, W)$
- fix U, W, optimize for V $\rightarrow$ $arg \max_{V} f(U^t, V, W^t)$
- fix V, W, optimize for U $\rightarrow$ $arg \max_{U} f(U, V^t, W^t)$

Maximum Likelihood based alternating optimization algorithm

1. Response matrix:

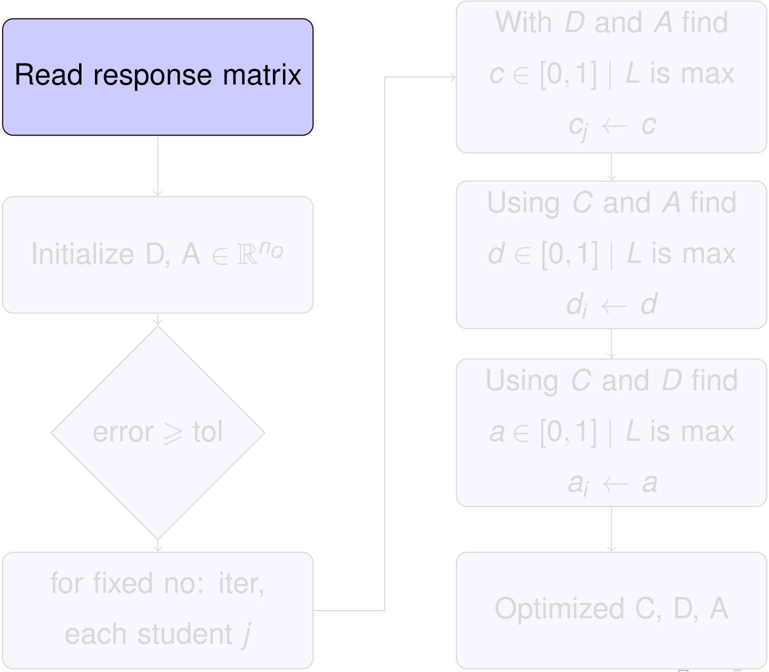
Question difficulties

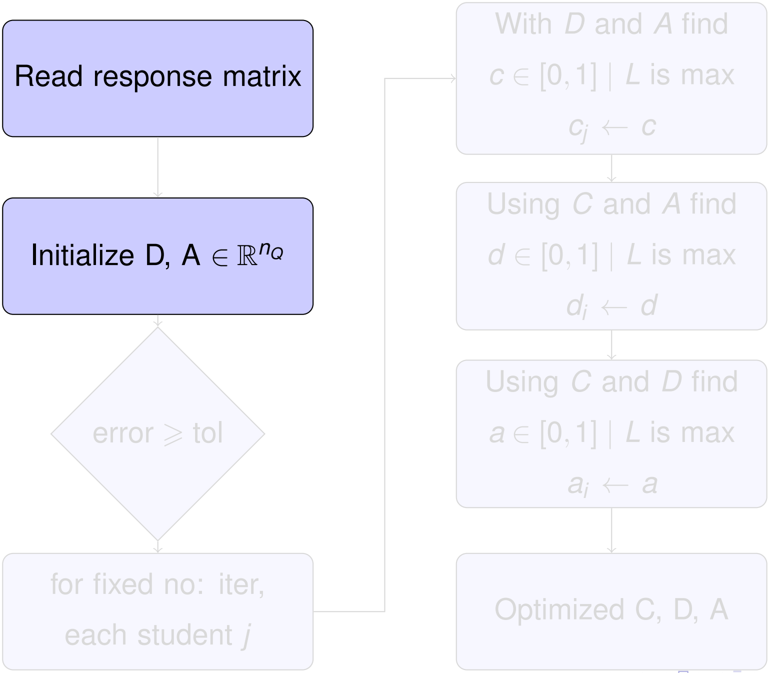
$$
\begin{array}{c c c}
 & d_1 \quad d_2 \quad d_3 & \\
\begin{array}{c} c_1 \\ c_2 \\ c_3 \\ c_4 \end{array} &
\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} &
\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} \\
 & \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} &
\end{array}
$$

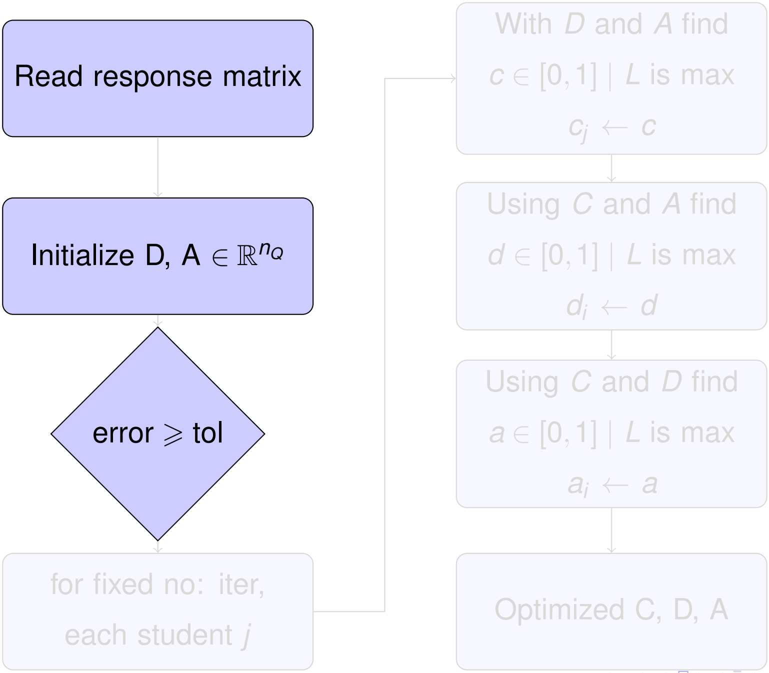
Student Capabilities

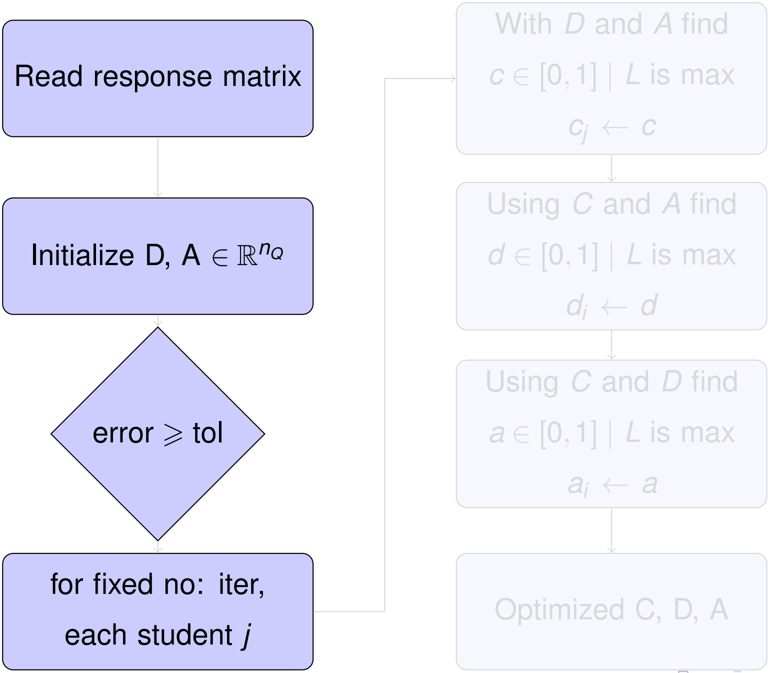
Row-wise totals

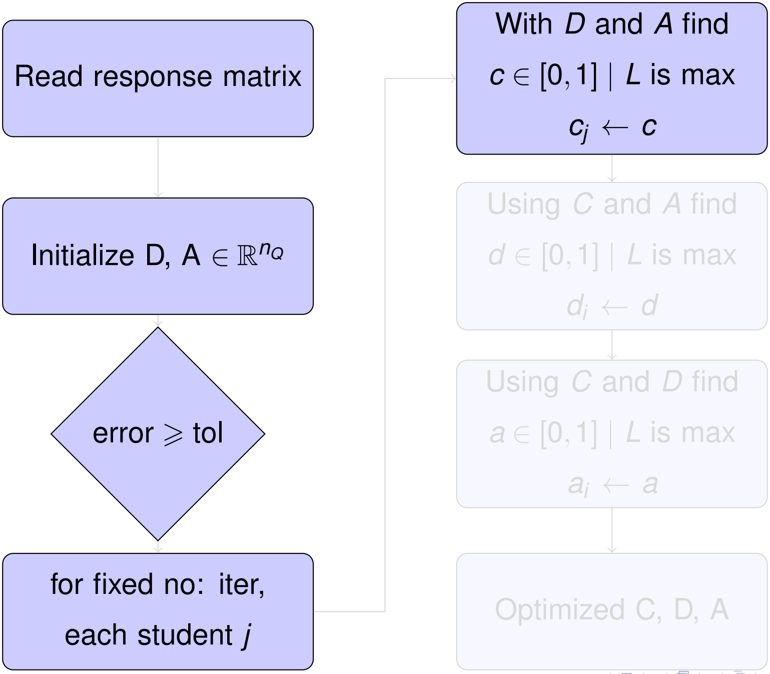
Column-wise totals

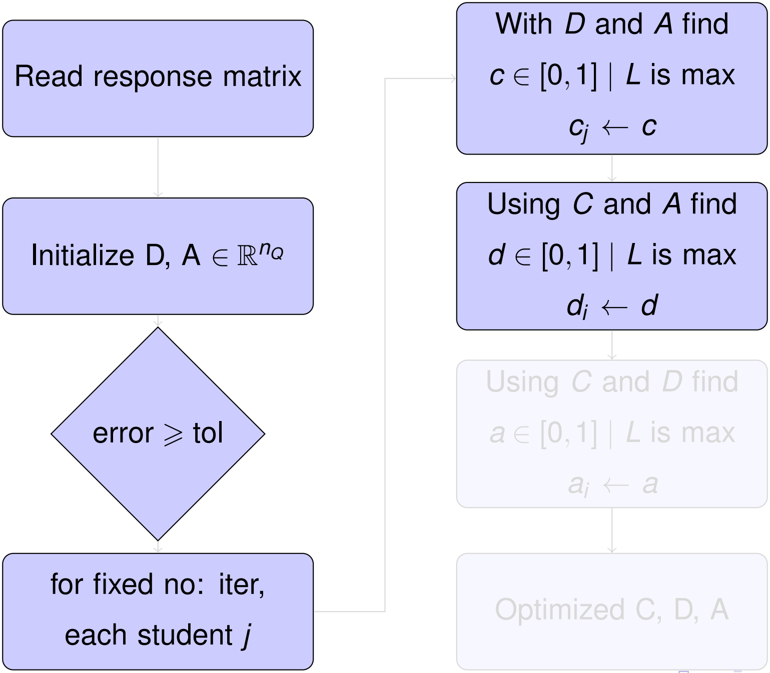

Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A

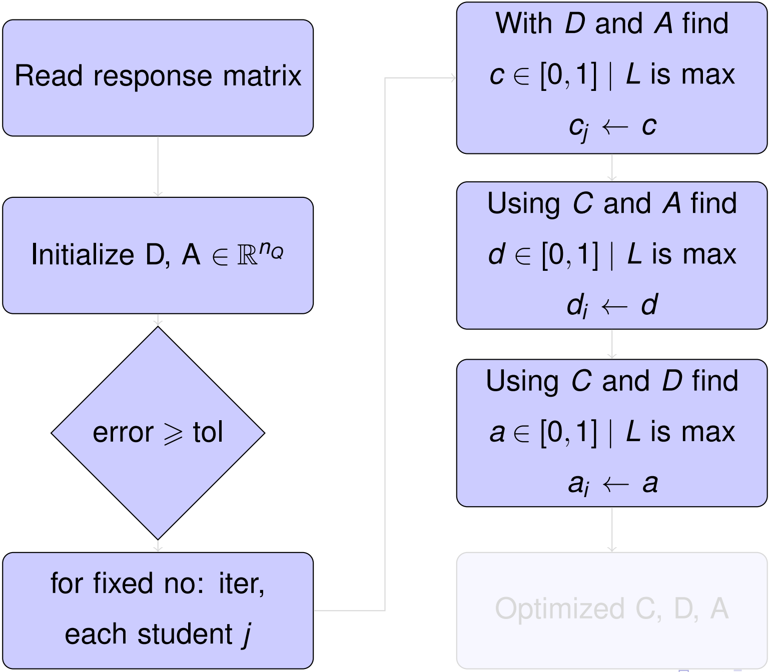

Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A

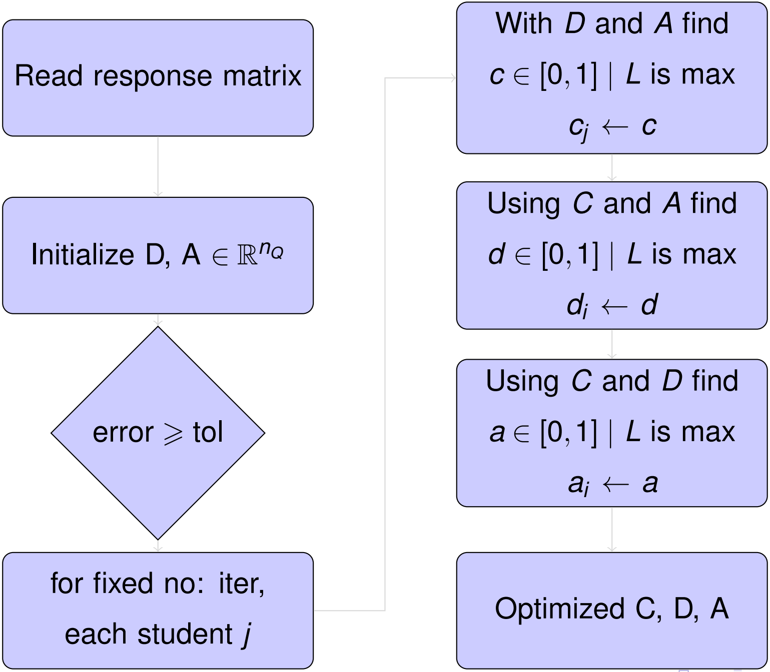

Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_j \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_j \leftarrow a$
Optimized C, D, A

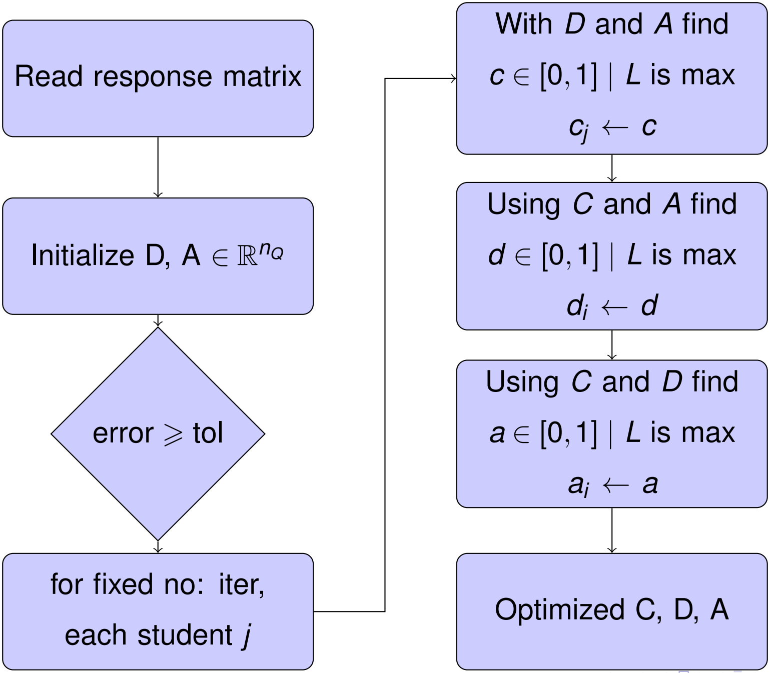

Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A


Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A


Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A


Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A


Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A


Read response matrix
Initialize D, A $\in \mathbb{R}^{n_Q}$
error $\geqslant$ tol
for fixed no: iter,
each student j
With D and A find
$c \in [0,1] \mid L$ is max
$c_j \leftarrow c$
Using C and A find
$d \in [0,1] \mid L$ is max
$d_i \leftarrow d$
Using C and D find
$a \in [0,1] \mid L$ is max
$a_i \leftarrow a$
Optimized C, D, A

Convergence of the proposed algorithm

Lemma

The log-likelihood function $L = \sum_{j=1}^{n_S} \sum_{i=1}^{n_Q} [m_{ij} a_i (c_j - d_i) - \log(1 + e^{a_i(c_j - d_i)})]$ *is concave individually in C, D and A.*

Lemma

The log-likelihood function
$L = \sum_{j=1}^{n_S} \sum_{i=1}^{n_Q} [m_{ij} a_i (c_j - d_i) - \log(1 + e^{a_i(c_j - d_i)})]$ *is jointly concave in C, D.*

Results for fixed n_S and n_Q varied

Parameters	$n_S = 20000$							
	$n_Q = 20$		$n_Q = 30$		$n_Q = 50$		$n_Q = 70$	
	CA	MA	CA	MA	CA	MA	CA	MA
qualified	2001	2205	2001	2699	2001	2555	2001	2197
crashers	1093	1288	1465	2068	875	1307	774	965
desired	908	916	536	631	1126	1248	1227	1232
90	0.38787	0.38982	0.40384	0.40163	0.39325	0.39053	0.39065	0.39205
99	0.38886	0.39102	0.40583	0.40231	0.39104	0.38740	0.39345	0.39213

Results for fixed n_Q and n_S varied

Parameters	$n_Q = 30$							
	$n_S = 2000$		$n_S = 5000$		$n_S = 10000$		$n_S = 20000$	
	CA	MA	CA	MA	CA	MA	CA	MA
qualified	201	225	501	691	1001	1203	2001	2699
crashers	115	137	272	423	513	692	1465	2068
desired	86	88	229	268	489	511	536	631
90	0.39116	0.39253	0.38588	0.38805	0.38681	0.38916	0.40384	0.40163
99	0.39568	0.39512	0.38253	0.38853	0.38675	0.39048	0.40583	0.40231

Results for different question distributions

Parameters	$n_S = 5000, n_Q = 30$							
	beta		gamma		uniform		triangular	
	CA	MA	CA	MA	CA	MA	CA	MA
qualified	501	534	501	794	501	691	501	576
crashers	214	254	288	514	272	423	241	310
desired	287	280	213	280	229	268	260	266
90	0.38772	0.39331	0.38546	0.38731	0.38588	0.38805	0.38426	0.39272
99	0.38822	0.39927	0.38264	0.38750	0.38253	0.38853	0.38842	0.38385

Results for multiple session exam

Parameters	$n_S = 5000, n_Q = 30$							
	session 1		session 2		session 3		session 4	
	CA	MA	CA	MA	CA	MA	CA	MA
qualified	501	689	501	706	501	562	501	553
crashers	280	429	239	389	286	351	274	321
desired	221	260	249	311	215	211	227	232
90	0.38601	0.38791	0.37494	0.38001	0.41180	0.40426	0.39201	0.39073
99	0.38287	0.388591	0.36767	0.38217	0.40573	0.40247	0.39161	0.39077

Results for multiple session exam

Parameters	$n_S = 5000, n_Q = 30$							
	session 1		session 2		session 3		session 4	
	CA	MA	CA	MA	CA	MA	CA	MA
qualified	501	689	501	706	501	562	501	553
crashers	280	429	239	389	286	351	274	321
desired	221	260	249	311	215	211	227	232
90	0.38601	0.38791	0.37494	0.38001	0.41180	0.40426	0.39201	0.39073
99	0.38287	0.388591	0.36767	0.38217	0.40573	0.40247	0.39161	0.39077

Summary

$$\begin{bmatrix} 1 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix} \rightarrow$$

ML based alternating optimization algorithm

$$\rightarrow \quad C = [c_1 \dots c_{n_S}], \quad D = [d_1 \dots d_{n_Q}]$$

Conclusion

1. Convergence of the proposed algorithm to the unique optimum value is guaranteed.
2. Algorithm outperforms the conventional marks based scheme in filtering out the most deserving candidates.
3. The number of qualified candidates is atmost one extra than cutoff percentage.
4. The performance is better than normalized marks based method in case of multiple session exam.
5. Algorithm is invariant of the type of question paper distribution.

Thank You