Probabilistic Knowledge Structures II

Jürgen Heller & Florian Wickelmaier

University of Tübingen

November 2016

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Outline

Introduction

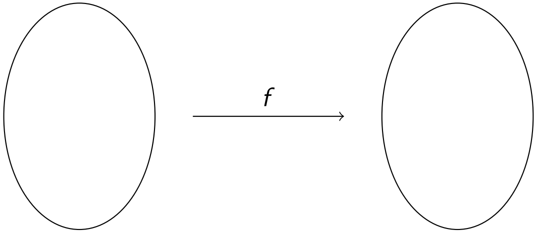
- Interpreting parameter estimates in a model requires that there is a unique set of parameters for any observed data set
- Models with this property are called identifiable
- There are some results on the identifiability of the most prominent probabilistic knowledge structure, which is known as the basic local independence model (BLIM)
  - Heller (2016); Spoto, Stefanutti, & Vidotto (2012); Stefanutti, Heller, Anselmi, & Robusto (2012)
- These results can be extended to a competence-based extension of the basic local independence model (CBLIM) and transfer to important types of Cognitive Diagnostic Models
  - Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80, 995-1019.

Parametric Model

Definition (Bamber & van Santen, 1985, 2000)

A (parametric) model consists of

- parameter space $\Theta \subseteq \mathbb{R}^n$
- outcome space $\Phi \subseteq \mathbb{R}^m$
- prediction function $f : \Theta \to \Phi$

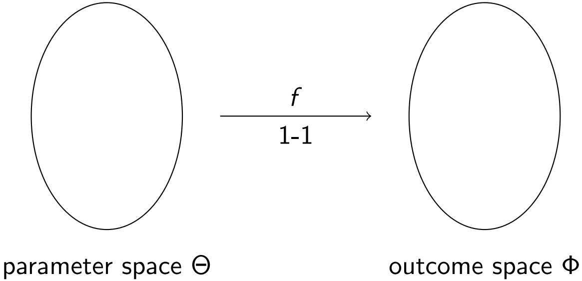


parameter space Θ outcome space Φ

Identifiability

Definition

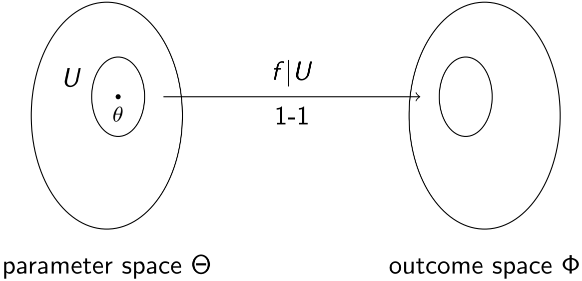
A model is identifiable if its prediction function f defined on the parameter space is a one-to-one function



Local Identifiability

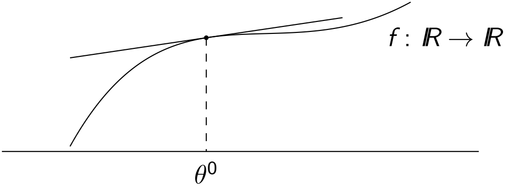
Definition

A model is locally identifiable at a point θ of the parameter space if there is an open neighborhood U of θ such that $f|U$ (i.e., f restricted to the neighborhood U) is one-to-one



Local Identifiability

- Linear approximation of the predictor function in a neighborhood of the point θ^0 (unidimensional case)



- Multidimensional case $f : \Theta \to \Phi$

$$l(\theta) = J_f(\theta^0) \cdot (\theta - \theta^0) + f(\theta^0)$$

with

$$J_f(\theta) = \left(\frac{\partial f_k(\theta)}{\partial \theta_j} \right)_{k=1,\ldots,m;\ j=1,\ldots,n}$$

the *Jacobian matrix* of partial derivatives of f

Basic Local Independence Model

- Knowledge structure $\mathcal{K}$ on a (finite) knowledge domain Q (i.e., a collection $\mathcal{K} \subseteq 2^Q$ with $\emptyset, Q \in \mathcal{K}$)
- Probability distribution $(\pi_K)_{K \in \mathcal{K}}$ over all knowledge states
- Given the knowledge state $K \in \mathcal{K}$, the response to each item q only depends on the probabilities
  - β_q of a careless error
  - η_q of a lucky guess
- Local stochastic independence: Conditional probability of response pattern $R \in \mathcal{R} = 2^Q$ given knowledge state $K \in \mathcal{K}$

$$P(R \mid K) = \left(\prod_{q \in K \setminus R} \beta_q \right) \cdot \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \cdot \left(\prod_{q \in R \setminus K} \eta_q \right) \cdot \left(\prod_{q \in \bar{R} \cap \bar{K}} (1 - \eta_q) \right)$$

Basic Local Independence Model

Predicted probability distribution over responses $R \in \mathcal{R} = 2^Q$

$$P(R) = \sum_{K \in \mathcal{K}} P(R \mid K) \cdot \pi_K$$

Parameter Space

The parameter space $\Theta_{\mathcal{K}}$ is defined to consist of all vectors

$$\theta_{\mathcal{K}} = (\beta, \eta, \pi),$$

with

$$\beta = (\beta_q)_{q \in Q}, \quad \eta = (\eta_q)_{q \in Q}, \quad \pi = (\pi_K)_{K \in \mathcal{K}^*}$$

and $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$, such that

- $\theta \in (0,1)^n$ with $n = 2\,|Q| + |\mathcal{K}| - 1$
- $\sum_{L \in \mathcal{K}^*} \pi_L < 1$
- $\beta_q + \eta_q < 1$ for all $q \in Q$

This implies that $\Theta_{\mathcal{K}}$ is an open convex set

Parameter Space

Assumptions

- $\theta \in (0,1)^n$ with $n = 2|Q| + |\mathcal{K}| - 1$
  - Interpretation: All probabilities are nonzero, however small they are (probabilities that are known to be zero a priori, are not part of the parameter space)
- $\sum_{L \in \mathcal{K}^*} \pi_L < 1$
  - Interpretation: π is a probability distribution on $\mathcal{K}$
- $\beta_q + \eta_q < 1$ for all $q \in Q$
  - Interpretation: A correct response is more likely if the item is mastered than if it is not mastered

$$\eta_q < 1 - \beta_q,$$

and an incorrect response is more likely if the item is not mastered than if it is mastered

$$\beta_q < 1 - \eta_q$$

Parameter Space

Implications

- The inequality $\beta_q + \eta_q < 1$ for all $q \in Q$ needs to be strict
- Otherwise the prediction function of a probabilistic knowledge structure is not one-to-one
- Consider the case $\beta_q = \eta_q = 0.5$ for all $q \in Q$
- Then we have

$$P(R) = \sum_{K \in \mathcal{K}} P(R \mid K, \boldsymbol{\beta}, \boldsymbol{\eta}) \cdot \pi_K = \sum_{K \in \mathcal{K}} 0.5^{|Q|} \cdot \pi_K$$
$$= 0.5^{|Q|} \sum_{K \in \mathcal{K}} \pi_K = 0.5^{|Q|}$$

- This means that the prediction function is independent of the distribution π_K over the knowledge states $K \in \mathcal{K}$

Outcome Space and Prediction Function

The outcome space Φ is the set of all probability distributions on $\mathcal{R} = 2^Q$

$$\Phi = \left\{ (\phi_R)_{R \in \mathcal{R}^*} \in [0,1]^m \colon \sum_{R \in \mathcal{R}^*} \phi_R \leq 1 \right\},$$

with $\mathcal{R}^* = \mathcal{R} \setminus \{Q\}$ and $m = |\mathcal{R}^*| = 2^{|Q|} - 1$

The prediction function is defined on Θ and given by

$$f(\theta) = \big(f_R(\theta)\big)_{R \in \mathcal{R}^*} = \big(P(R \,|\, \theta)\big)_{R \in \mathcal{R}^*}$$

Jacobian Matrix

The Jacobian matrix of f at $\theta \in \Theta$ is the matrix of partial derivatives

$$J_f(\theta) = \left(\frac{\partial f_R(\theta)}{\partial \theta_j}\right)_{R \in \mathcal{R}^*,\ j=1,\ldots,n}$$

$$\begin{aligned}
\frac{\partial f_R(\theta)}{\partial \pi_K} &= P(R \,|\, K) - P(R \,|\, Q) \\
\frac{\partial f_R(\theta)}{\partial \beta_q} &= \rho_q^R \cdot \left(\sum_{K \in \mathcal{K}_q^*} P(R \,|\, K) \cdot \pi_K + P(R \,|\, Q) \cdot (1 - \sum_{L \in \mathcal{K}^*} \pi_L) \right) \\
\frac{\partial f_R(\theta)}{\partial \eta_q} &= \sigma_q^R \cdot \sum_{K \in \mathcal{K}_{\bar{q}}^*} P(R \,|\, K) \cdot \pi_K
\end{aligned}$$

$$\rho_q^R = \begin{cases} -1/(1-\beta_q) & \text{if } q \in R \\ 1/\beta_q & \text{if } q \notin R \end{cases} \quad \text{and } \sigma_q^R = \begin{cases} 1/\eta_q & \text{if } q \in R \\ -1/(1-\eta_q) & \text{if } q \notin R \end{cases}$$

(Local) Identifiablity of BLIMs

- For identifiability the number of free parameters n must not exceed the number of independent observations m, implying the restriction

$$2^{|Q|} \geq 2 \cdot |Q| + |\mathcal{K}|$$

- If this inequality holds and the Jacobian matrix of the prediction function f has full rank at some point $\theta_0 \in \Theta_{\mathcal{K}}$ then the BLIM is locally identifiable at that point
- The prediction function of a BLIM is analytic, which implies that its Jacobian matrix has full rank almost everywhere in $\Theta_{\mathcal{K}}$ and is rank deficient almost nowhere

(Local) Identifiablity of BLIMs

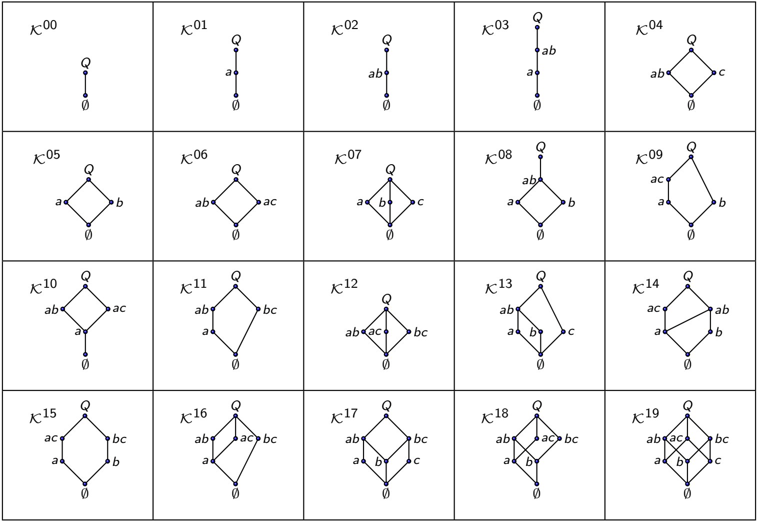
- The inequality $2^{|Q|} \geq 2 \cdot |Q| + |\mathcal{K}|$ implies that for a BLIM to be identifiable we need $|Q| \geq 3$
- In the sequel we consider the collection of all knowledge structures on the domain $Q = \{a, b, c\}$
- The following catalogue provides an overview over all equivalence classes of knowledge structures that are identical up to relabeling of the items
- For these knowledge structures the following situations are considered

Case I does not allow for guessing by defining $\eta_q = 0$ for all $q \in Q$;

Case II assumes that the error probabilities β_q and η_q are nonzero but constant over all the items $q \in Q$;

Case III combines the two previous cases by assuming that the β_q are nonzero and constant over all the items $q \in Q$, while $\eta_q = 0$ for all $q \in Q$.

Catalogue of Knowledge Structures on $Q = \{a, b, c\}$

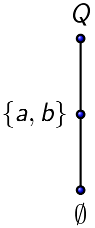


Summary of Results

		BLIM			Case I			Case II			Case III		
					$\eta_q = 0$			$\beta_q = \beta^0$, $\eta_q = \eta^0$			$\beta_q = \beta^0$, $\eta_q = 0$		
no.	size	n	rk(J_f)		n	rk(J_f)		n	rk(J_f)		n	rk(J_f)	
00	1	7	7	*	4	4	*	3	3	*	2	2	*
01	3	8	7		5	5	*	4	4	*	3	3	*
02	3	8	7		5	4		4	4	*	3	3	*
03	6	9	7		6	5		5	5	*	4	4	*
04	3	9	7		6	5		5	5	*	4	4	*
05	3	9	7		6	6	*	5	5	*	4	4	*
06	3	9	7		6	5		5	5	*	4	4	*
07	1	10	7		7	7	*	6	6	*	5	5	*
08	3	10	7		7	6		6	6	*	5	5	*
09	6	10	7		7	6		6	6	*	5	5	*
10	3	10	7		7	5		6	6	*	5	5	*
11	6	10	7		7	6		6	6	*	5	5	*
12	1	10	7		7	7	*	6	6	*	5	5	*
13	3	11	7		8	7		7	7	*	6	6	*
14	6	11	7		8	6		7	7	*	6	6	*
15	3	11	7		8	7		7	7	*	6	6	*
16	3	11	7		8	7		7	7	*	6	6	*
17	3	12	7		9	7		8	7		7	7	*
18	3	12	7		9	7		8	7		7	7	*
19	1	13	7		10	7		9	7		8	7	

Example I

$\mathcal{K}^{02} = \{\emptyset, \{a, b\}, Q\}$ on $Q = \{a, b, c\}$ with parameter vector $\theta' = (\beta_a, \beta_b, \beta_c, \pi_\emptyset, \pi_{ab})$ (no guessing!)



Example I

Prediction function of $\mathcal{K}^{02}$

$$f^{02}(\theta) =$$

$$\begin{pmatrix} \phi_\emptyset(\theta) \\ \phi_a(\theta) \\ \phi_b(\theta) \\ \phi_c(\theta) \\ \phi_{ab}(\theta) \\ \phi_{ac}(\theta) \\ \phi_{bc}(\theta) \end{pmatrix} = \begin{pmatrix} \pi_\emptyset + \beta_a\beta_b\beta_c(1-\pi_\emptyset-\pi_{ab}) + \beta_a\beta_b\pi_{ab} \\ (1-\beta_a)\beta_b\beta_c(1-\pi_\emptyset-\pi_{ab}) + (1-\beta_a)\beta_b\pi_{ab} \\ \beta_a(1-\beta_b)\beta_c(1-\pi_\emptyset-\pi_{ab}) + \beta_a(1-\beta_b)\pi_{ab} \\ \beta_a\beta_b(1-\beta_c)(1-\pi_\emptyset-\pi_{ab}) \\ (1-\beta_a)(1-\beta_b)\beta_c(1-\pi_\emptyset-\pi_{ab}) + (1-\beta_a)(1-\beta_b)\pi_{ab} \\ (1-\beta_a)\beta_b(1-\beta_c)(1-\pi_\emptyset-\pi_{ab}) \\ \beta_a(1-\beta_b)(1-\beta_c)(1-\pi_\emptyset-\pi_{ab}) \end{pmatrix}$$

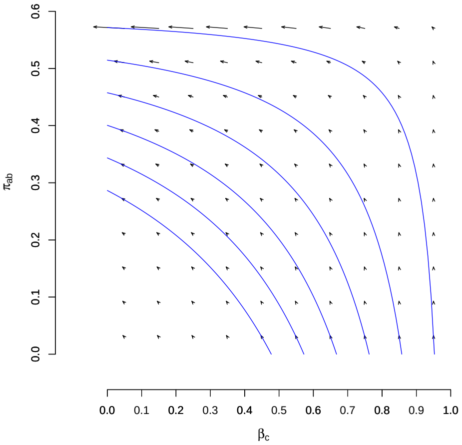
Example I

Local identifiability of $\mathcal{K}^{02}$

- Max. rank of Jacobian matrix $rk\left(J_f^{02}(\theta)\right) = 4$ $(n = 5)$
- The vector
$$\nu'(\theta) = (0, 0, -\frac{1-\beta_c}{1-\pi_\emptyset-\pi_{ab}}, 0, 1)$$
generates the null space of the Jacobian matrix
- $\nu(\theta)$ characterizes the trade-off between parameter values
  - There is a trade-off between parameters β_c and π_{ab}
  - Zero entries indicate identifiability of the corresponding parameter
- $\nu(\theta)$ defines a vector field through the differential equation
$$\frac{d\theta(t)}{dt} = \nu(\theta)$$
- Its solutions specify the curves, along which predictions are identical

Example I

Parameter trade-off in $\mathcal{K}^{02}$



Results I

Proposition

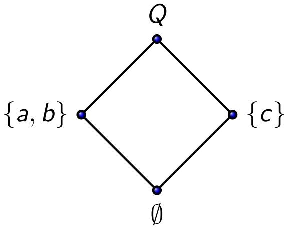
Assume that for $q \in Q$ *we have* $\mathcal{K}_q = \{Q\}$ *and* $K = Q \setminus \{q\} \in \mathcal{K}$. *Then*

$$(1 - \beta_q - \eta_q) \cdot \frac{\partial f_R(\theta)}{\partial \beta_q} = (1 - \sum_{L \in \mathcal{K}^*} \pi_L) \cdot \frac{\partial f_R(\theta)}{\partial \pi_K}$$

for all $R \in \mathcal{R}^*$.

Example II

$\mathcal{K}^{04} = \{\emptyset, \{a, b\}, \{c\}, Q\}$ with parameter vector
$\theta = (\beta_a, \beta_b, \beta_c, \pi_\emptyset, \pi_{ab}, \pi_c)$ (no guessing!)



- Max. rank of Jacobian matrix $rk\big(J_f(\theta)\big) = 5$

Results II

Forward- and backward-gradedness

- The relevant structural property of $\mathcal{K}^{04}$ is the following (Spoto et al., 2012)

Definition

Let $\mathcal{K}$ be a knowledge structure on Q. Then

- $\mathcal{K}$ is forward-graded in $q \in Q$, i.e., $K_{+q} \subseteq \mathcal{K}$ with $\mathcal{K}_{+q} = \{K \cup \{q\} \mid K \in \mathcal{K}\}$
- $\mathcal{K}$ is said to be backward-graded in $q \in Q$ if $\mathcal{K}_{-q} \subseteq \mathcal{K}$ with $\mathcal{K}_{-q} = \{K \setminus \{q\} \mid K \in \mathcal{K}\}$

Results II

Backward- and forward-gradedness

Proposition (Heller, 2016)

Let $\mathcal{K}$ be backward-graded in $q \in Q$. Then for all $R \in \mathcal{R}^*$

$$(1 - \beta_q - \eta_q) \cdot \frac{\partial f_R(\theta)}{\partial \beta_q} - \sum_{K \in \mathcal{K}_{-q}} \pi_{K \cup \{q\}} \cdot \frac{\partial f_R(\theta)}{\partial \pi_K} + \sum_{K \in \mathcal{K}_q^*} \pi_K \cdot \frac{\partial f_R(\theta)}{\partial \pi_K} = 0$$

with $\pi_Q = 1 - \sum_{L \in \mathcal{K}^*} \pi_L$.

There is a dual result for parameter η_q if $\mathcal{K}$ is forward-graded in $q \in Q$

Results II

Non-identifiable knowledge structures due to forward- or backward-gradedness in some item

- Well-graded knowledge spaces (i.e., learning spaces), and thus ordinal knowledge spaces
- Knowledge spaces satisfying $\{q\} \in \mathcal{K}$ for some $q \in Q$, or $q \notin B$ for all atoms B at $p \neq q$
- Quasi ordinal knowledge spaces with the corresponding quasi order having unique maximal or minimal elements

There are other sources of non-identifiability

It seems that equally informative items p and q, satisfying $p \in K$ iff $q \in K$ for all $K \in \mathcal{K}$, are never involved in trade-offs due to non-identifiability

Exercise in R

Return to Exercise I (by Pasquale) on the surmise relation

1. Fit the BLIM corresponding to the surmise relation provided in the solution to the five-item subset of the fraction-subtraction data
2. Test whether the model is locally identifiable by computing the rank of the Jacobian matrix: `qr(jacobian(model))$rank`
3. Characterize the possible reasons for non-identifiability. Which parameters are involved in trade-offs?
4. Check forward- and backward-gradedness of the knowledge structure: `is.forward.graded()`, `is.backward.graded()`
5. Think of ways to remedy non-identifiability

Competence-Based Basic Local Independence Model

- Skill function (Q, S, μ)
- Competence structure $\mathcal{C}$ on the set S of skills, i.e., a collection $\mathcal{C} \subseteq 2^S$ with $\emptyset, S \in \mathcal{C}$
- Probability distribution $(\pi_C)_{C \in \mathcal{C}}$ over all competence states
- A probability distribution on the delineated knowledge structure $\mathcal{K} = p(\mathcal{C})$ is defined by

$$\pi_K = \sum_{C \in p^{-1}(\{K\})} \pi_C,$$

  which induces a BLIM
- The above assumptions together with the induced BLIM define a competence-based basic local independence model (CBLIM)

Parameter Space

The parameter space $\Theta_{\mathcal{C}}$ is defined to consist of all vectors

$$\theta_{\mathcal{C}} = (\beta, \eta, \pi),$$

with

$$\beta = (\beta_q)_{q \in Q}, \quad \eta = (\eta_q)_{q \in Q}, \quad \pi = (\pi_C)_{C \in \mathcal{C}^*}$$

and $\mathcal{C}^* = \mathcal{C} \setminus \{S\}$, such that

- $\theta \in (0,1)^k$ with $n = 2\,|Q| + |\mathcal{C}| - 1$
- $\sum_{T \in \mathcal{C}^*} \pi_T < 1$
- $\beta_q + \eta_q < 1$ for all $q \in Q$

Prediction Function

The prediction function of a CBLIM is the composition

$$\Theta_{\mathcal{C}} \xrightarrow[\text{onto}]{g} \Theta_{\mathcal{K}} \xrightarrow{f} \Phi$$

of g and the prediction function f of the induced BLIM

Proposition

The following assertions are equivalent.

1. *The function g is locally one-to-one at some point in $\Theta_{\mathcal{C}}$;*
2. *The function p is one-to-one;*
3. *The function g is one-to-one.*

Results

Since g is onto, the composition $f \circ g$ is one-to-one if and only if both functions f and g are one-to-one

Proposition

A given CBLIM

1. *is identifiable if and only if the induced BLIM is identifiable and the problem function p is one-to-one;*
2. *is locally identifiable at a point $\theta_\mathcal{C}$ in $\Theta_\mathcal{C}$ if and only if the induced BLIM is locally identifiable at the point $g(\theta_\mathcal{C})$ in $\Theta_\mathcal{K}$ and the problem function p is one-to-one.*

Results

- In general, we cannot tell apart competence states that are mapped onto the same knowledge state, so we consider an equivalence relation $\sim_p$ such that for all $C_1, C_2 \in \mathcal{C}$

$$C_1 \sim_p C_2 \text{ if and only if } p(C_1) = p(C_2)$$

- The resulting set of equivalence classes $\mathcal{C}/{\sim_p}$ is partially ordered by

$$[C_1]_p \sqsubseteq [C_2]_p \text{ if and only if } p(C_1) \subseteq p(C_2)$$

- A probabilistic framework on $\mathcal{C}/{\sim_p}$ is obtained by identifying it with the induced BLIM via the order-isomorphism $p^* \colon \mathcal{C}/{\sim_p} \to \mathcal{K}$
- The probability distribution on the equivalence classes in $\mathcal{C}/{\sim_p}$ then is identical to that on the delineated knowledge structure $\mathcal{K}$

Fraction Subtraction Data

- Data set consists of responses of 536 middle school students to 15 fraction subtraction items and forms a subset of data originally described by K. Tatsuoka (1990)

a	$\frac{6}{7} - \frac{4}{7} =$	i	$3\frac{4}{5} - 3\frac{2}{5} =$
b	$3\frac{1}{2} - 2\frac{3}{2} =$	j	$2 - \frac{1}{3} =$
c	$\frac{3}{4} - \frac{3}{8} =$	k	$4\frac{5}{7} - 1\frac{4}{7} =$
d	$3 - 2\frac{1}{5} =$	l	$7\frac{3}{5} - \frac{4}{5} =$
e	$3\frac{7}{8} - 2 =$	m	$4\frac{1}{10} - 2\frac{8}{10} =$
f	$4\frac{4}{12} - 2\frac{7}{12} =$	n	$7 - 1\frac{4}{3} =$
g	$4\frac{1}{3} - 2\frac{4}{3} =$	o	$4\frac{1}{3} - 1\frac{5}{3} =$
h	$\frac{11}{8} - \frac{1}{8} =$		

Fraction Subtraction Data

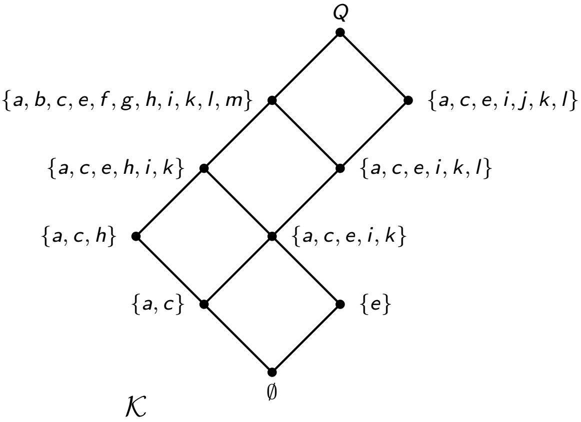
- In an analysis by de la Torre & Douglas (2008) based on the DINA model, the following skills were assumed

s	performing basic fraction subtraction operation
t	simplifying/reducing
u	separating whole number from fraction
v	borrowing one from whole number to fraction
w	converting whole number to fraction

Skill Function

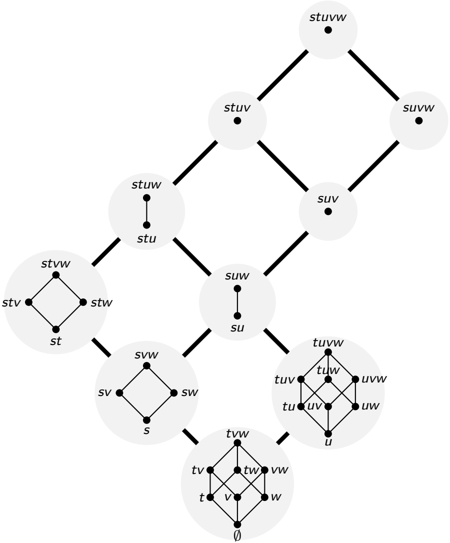
$$\mu(a) = \{\{s\}\}$$
$$\mu(b) = \{\{s, t, u, v\}\}$$
$$\mu(c) = \{\{s\}\}$$
$$\mu(d) = \{\{s, t, u, v, w\}\}$$
$$\mu(e) = \{\{u\}\}$$
$$\mu(f) = \{\{s, t, u, v\}\}$$
$$\mu(g) = \{\{s, t, u, v\}\}$$
$$\mu(h) = \{\{s, t\}\}$$
$$\mu(i) = \{\{s, u\}\}$$
$$\mu(j) = \{\{s, u, v, w\}\}$$
$$\mu(k) = \{\{s, u\}\}$$
$$\mu(l) = \{\{s, u, v\}\}$$
$$\mu(m) = \{\{s, t, u, v\}\}$$
$$\mu(n) = \{\{s, t, u, v, w\}\}$$
$$\mu(o) = \{\{s, t, u, v\}\}$$

Delineated Knowledge Structure



- The induced BLIM is locally identifiable, because its Jacobian matrix has full rank

Equivalence Classes of Competence States



R Package CDM

DINA parameter estimation

```
library(CDM)
data(fraction.subtraction.data)

item.idx <- c(6, 4, 8, 7, 9:12, 14:20)        # select 15 items
dat <- fraction.subtraction.data[, item.idx]
Q <- matrix(c(
    1, 0, 0, 0, 0,
    1, 1, 1, 1, 0,
    1, 0, 0, 0, 0,
    1, 1, 1, 1, 1,
    ...
    1, 0, 1, 1, 0,
    1, 1, 1, 1, 0,
    1, 1, 1, 1, 1,
    1, 1, 1, 1, 0), nrow = 15, ncol = 5, byrow = TRUE)
```

R Package CDM

DINA parameter estimation

```
dina <- din(dat, Q, rule = "DINA")
dina$attribute.patt

       class.prob class.expfreq
00000 0.016950612     9.0855280
10000 0.004822363     2.5847866
01000 0.016950612     9.0855280
00100 0.007978159     4.2762933
00010 0.016950612     9.0855280
...
11110 0.104401479    55.9591927
11101 0.116538136    62.4644411
11011 0.005853909     3.1376950
10111 0.001705161     0.9139663
01111 0.007978159     4.2762933
11111 0.368634688   197.5881925
```

R Package CDM

DINA parameter estimation

```
map2emptyK <- c("00000", "01000", "00010", "00001",
                "01010", "01001", "00011", "01011")
dina$attribute.patt[map2emptyK, ]

      class.prob class.expfreq
00000 0.01695061      9.085528
01000 0.01695061      9.085528
00010 0.01695061      9.085528
00001 0.01695061      9.085528
01010 0.01695061      9.085528
01001 0.01695061      9.085528
00011 0.01695061      9.085528
01011 0.01695061      9.085528
```

R Package pks

BLIM parameter estimation

```
library(pks)

skillfun <- read.table(header = TRUE, text = "
    item s t u v w
    a 1 0 0 0 0
    b 1 1 1 1 0
    c 1 0 0 0 0
    d 1 1 1 1 1
    e 0 0 1 0 0
    ...
    k 1 0 1 0 0
    l 1 0 1 1 0
    m 1 1 1 1 0
    n 1 1 1 1 1
    o 1 1 1 1 0
")
```

R Package pks

BLIM parameter estimation

```
t <- delineate(skillfun)
t$K
                a b c d e f g h i j k l m n o
000000000000000 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
000010000000000 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
101000000000000 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0
101000010000000 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0
101010001010000 1 0 1 0 1 0 0 0 1 0 1 0 0 0 0
101010011010000 1 0 1 0 1 0 0 1 1 0 1 0 0 0 0
101010001011000 1 0 1 0 1 0 0 0 1 0 1 1 0 0 0
101010001111000 1 0 1 0 1 0 0 0 1 1 1 1 0 0 0
111011111011101 1 1 1 0 1 1 1 1 1 0 1 1 1 0 1
111111111111111 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

t$classes  # equivalence classes of competence states
```

R Package pks

BLIM parameter estimation

```
N.R <- as.pattern(dat, freq = TRUE)
blim1 <- blim(K, N.R, method = "ML")
blim1$P.K

000000000000000 000010000000000 101000000000000 101000010000000
    0.135555248     0.061339710     0.020318955     0.023637502
101010001010000 101010011010000 101010001011000 101010001111000
    0.043298112     0.234080988     0.007094979     0.001588763
111011111011101 111111111111111
    0.104539508     0.368546234
```

R Package pks

BLIM local identifiability

```
J <- jacobian(blim1)

dim(J)
32767    39

qr(J)$rank
39
```

R Package pks

Compare BLIM to DINA

```
logLik(dina)
-3463.212

logLik(blim)
-3463.195

sum(dina$attribute.patt[map2emptyK, "class.prob"])
0.1356049

blim1$P.K["000000000000000"]
0.1355552
```

Conclusions

- There are two sources of non-identifiability of a CBLIM defined by a skill function (Q, S, μ) and a competence structure $\mathcal{C}$
  - The corresponding problem function p is not one-to-one
  - The induced BLIM on the delineated knowledge structure $\mathcal{K} = p(\mathcal{C})$ is not identifiable
- There is a close correspondence between probabilistic knowlegde structures and cognitive diagnostic models
  - The CBLIM is equivalent to the Multiple Strategy DINA
- (Local) identifiability matters as long as we want to interpret the parameters (as in probabilistic assessments based on knowledge structures)
- (Local) identifiability does not matter whenever we are interested in the global fit of a particular BLIM/CBLIM to given data only

Exercise in R

Return to Exercise 4 (by Pasquale) on the skill function

1. Fit the CBLIM defined by the skill function provided in the exercise to the probability data
   - Determine the delineated knowledge structure
   - Fit the induced BLIM
2. Test the identifiability of the CBLIM
   - Does the Jacobian matrix of the BLIM have full rank?
   - Is the problem function one-to-one?
3. Fit the CBLIM defined by the refined skill function provided in the solution to Exercise 4
4. Test the identifiability of the CBLIM

References

Bamber, D., & van Santen, J.P.H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, 29, 443-473.

Bamber, D., & van Santen, J.P.H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, 44, 20-40.

Heller, J. (2016). Identifiability in probabilistic knowledge structures. *Journal of Mathematical Psychology*. doi: 10.1016/j.jmp.2016.07.008

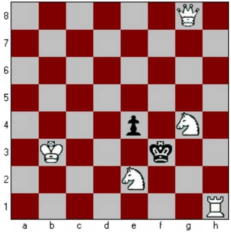
Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80, 995-1019.

Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2016). Erratum to: On the Link between Cognitive Diagnostic Models and Knowledge Space Theory. *Psychometrika*, 81, 250-251. Spoto, A., Stefanutti, L., & Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *Journal of Mathematical Psychology*, 56, 248-255.

Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E. (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behavior Research Methods*, 44, 1197-1211.

Example

Chess problems (Held, Schrepp, and Fries, 1995)



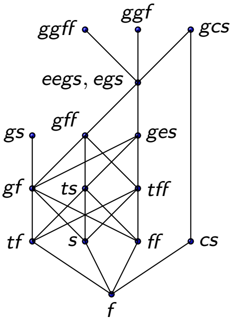
$Q = \{s, gs, egs, eegs, cs, gcs, ts, ges, f, gf, gff, ggff, ggf, ff, tf, tff\}$

- Motives
  - g (guidance), f (fork), e (elimination), c (clearing), s (stalement), t (promotion)

Example

Chess problems (Held, Schrepp, and Fries, 1995)

- Precedence relation of knowledge space DST_1 (75 states)



- 106 parameters, $\mathrm{rk}(J_f) = 101$
- forward-graded in f – backward-graded in $gs, ggff, ggf, gcs$