# Parameter estimation in probabilistic knowledge structures with the pks package

Florian Wickelmaier and Jürgen Heller

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Psychoco 2012, Innsbruck, February 9



## Knowledge structures

(Doignon & Falmagne, 1985, 1999)

Goals

- Characterizing the strengths and weaknesses in all parts of a knowledge domain
  - Precise, non-numerical characterization of the state of knowledge that is computationally tractable
  - Building upon results from discrete mathematics and exploiting the power of current computers
- Adaptive knowledge assessment
  - Efficiently identifying the current state of knowledge based on asking a minimal number of questions
  - Adapting to the already given responses as experienced teachers do in an oral examination
- Personalization in technology-enhanced learning
  - Automatically select content that a person is ready to learn





## A subdomain of physics: Conservation of matter (1)

(Taagepera et al., 1997)

a) When ice melts and produces water:
   (i) The water weighs more than the ice.
   (ii) The ice weighs more than the water.
   (iii) The water and ice weigh the same.
   (iv) The weight depends on the temperature.

b) After the nail rusts, its mass:
   (i) is greater than before.
   (ii) is less than before.
   (iii) is the same as before.
   (iv) cannot be predicted.

c) When 10 grams of iron and 10 grams of oxygen combine, the total amount of material after iron oxide (rust) is formed must weigh:
   (i) 10 grams.
   (ii) 19 grams.
   (iii) 20 grams.
   (iv) 21 grams.





## A subdomain of physics: Conservation of matter (2)

(Taagepera et al., 1997)

d) After 3 metal nuts and 3 metal bolts are joined together:
   (i) The total amount of metal is the same.
   (ii) There is less metal than before.
   (iii) There is more metal than before.
   (iv) The amount of metal cannot be determined.

e) Photosynthesis can be described as:

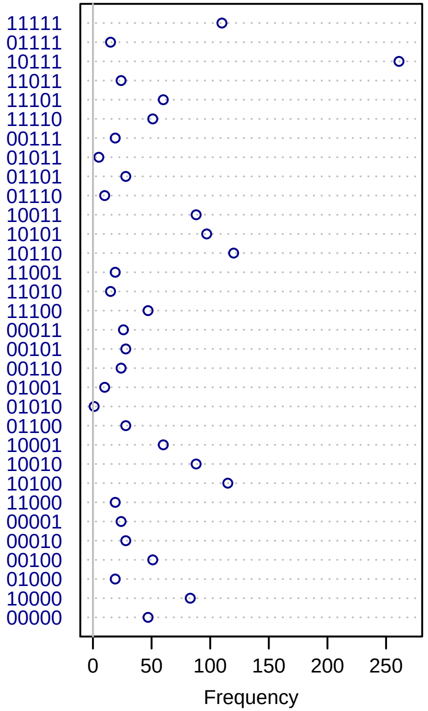
$$\text{WATER} + \text{CARBON DIOXIDE} \xrightarrow[\text{sunlight}]{\text{chlorophyll}} \text{GLUCOSE}$$

   Which of the following statements about this reaction is NOT true?
   (i) As more water and more carbon dioxide react, more glucose is produced.
   (ii) The same amount of glucose is produced no matter how much water and carbon dioxide is available.
   (iii) Chlorophyll and sunlight are needed for the reaction.
   (iv) The same atoms make up the GLUCOSE molecule as were present in WATER and CARBON DIOXIDE.

# Response patterns

(Taagepera et al., 1997)

Students from grades four through twelve
$N = 1620$








# Deterministic theory

Definitions

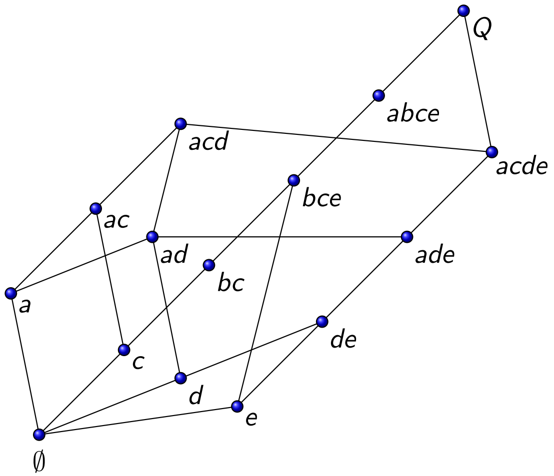
- A knowledge domain is identified with a set $Q$ of (dichotomous) items.
- The knowledge state of a person is identified with the subset $K \subseteq Q$ of problems in the domain $Q$ the person is capable of solving.
- A knowledge structure on the domain $Q$ is a collection $\mathcal{K}$ of subsets of $Q$ that contains at least the empty set $\emptyset$ and the set $Q$.
- The subsets $K \in \mathcal{K}$ are the knowledge states.





# Conservation of matter: Knowledge structure

(Taagepera et al., 1997)








# Probabilistic knowledge structures

Rationale

- If there are response errors then knowledge states $K \subseteq Q$ and response patterns $R \subseteq Q$ have to be dissociated.

Definition

- A probabilistic knowledge structure is defined by specifying
  - a knowledge structure $\mathcal{K}$ on a knowledge domain $Q$ (i. e., a collection $\mathcal{K} \subseteq 2^Q$ with $\emptyset, Q \in \mathcal{K}$)
  - a marginal distribution $P_{\mathcal{K}}(K)$ on the knowledge states $K \in \mathcal{K}$
  - the conditional probabilities $P(R \mid K)$ to observe response pattern $R$ given knowledge state $K$

The probability of the response pattern $R \in \mathcal{R} = 2^Q$ is predicted by

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

## The basic local independence model (BLIM)

(Doignon & Falmagne, 1999)

Assumption: Local stochastic independence

- Given the knowledge state $K$ of a person
  - the responses are stochastically independent over problems
  - the response to each problem $q$ only depends on the probabilities
    $\beta_q$ of a careless error
    $\eta_q$ of a lucky guess
- The probability of the response pattern $R$ given the knowledge state $K$ reads

$$P(R \mid K) = \prod_{q \in K \setminus R} \beta_q \prod_{q \in K \cap R} (1-\beta_q) \prod_{q \in R \setminus K} \eta_q \prod_{q \in Q \setminus (R \cup K)} (1-\eta_q).$$





## The pks package

- Provides functionality for parameter estimation in probabilistic knowledge structures.
- Main functions

| | |
|---|---|
| `blim` | Fitting and testing basic local independence models (BLIMs) |
| `print, logLik` `plot, residuals` | Extractor functions |
| `simulate` | generate response patterns from a given BLIM |
| `as.pattern,as.binmat` | conversion functions |





## Maximum likelihood estimation

EM algorithm

- Formulate the likelihood as if we have available the absolute frequencies $M_{RK}$ of subjects who are in state $K$ and produce pattern $R$ (complete data) instead of the absolute frequencies $N_R$ of the response patterns $R \in \mathcal{R}$ (incomplete data).

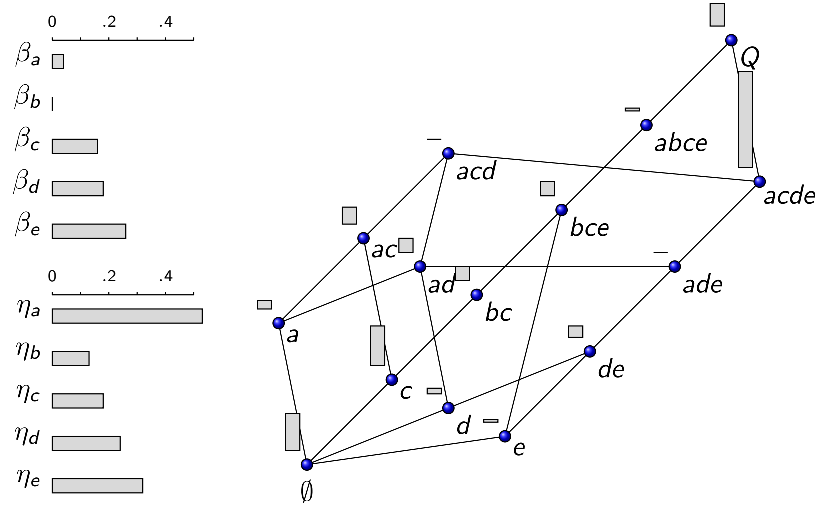
| **Expectation** | **Maximization** |
|---|---|
| Compute | Estimate $\hat{\beta}^{(t+1)}, \hat{\eta}^{(t+1)}, \hat{\pi}^{(t+1)}$ based on $m_{RK} = E(M_{RK})$ |

$$E(M_{RK}) = N_R \cdot P(K \mid R, \hat{\beta}^{(t)}, \hat{\eta}^{(t)}, \hat{\pi}^{(t)})$$





## Example: Maximum likelihood estimation



# Example: Maximum likelihood estimation

```
blim(matter97$K, matter97$N.R, method="ML")

Number of iterations: 9474
Goodness of fit (2 log likelihood ratio):
        G2(7) = 13.763, p = 0.055553

Minimum discrepancy distribution (mean = 0.2858)
   0    1
1157  463

Mean number of errors (total = 1.02435)
careless error    lucky guess
     0.3697625      0.6545893
```





# Maximum likelihood estimation

Problems

- 'Good fit' (w.r.t. likelihood ratio statistic) not sufficient for empirical validity of knowledge structure
  - Fit may be obtained by inflating careless error rates $\beta_q$ and lucky guess rates $\eta_q$
  - What we want: Good fit with small values of $\beta_q$ and $\eta_q$
- How to apply constraints on the error probabilities that are motivated by the knowledge structure?
  (instead of brute-force constraints, Stefanutti & Robusto, 2009)
- How much of the fit is due to inflating the error probabilities in ML estimation?





# Minimum discrepancy method

Rationale

- For a response pattern $R$ and a knowledge state $K$ consider the distance

$$d(R, K) = |(R \setminus K) \cup (K \setminus R)|,$$

  which is based on the symmetric set-difference.
- It is the number of items that are elements of either, but not both sets $R$ and $K$ (number of response errors).
- Example

$$d(10001, 10100) = 2$$





# Minimum discrepancy method

Rationale

- For a given response pattern $R$, consider the minimum of the symmetric distances between $R$ and all the knowledge states $K \in \mathcal{K}$

$$d(R, \mathcal{K}) = \min_{K \in \mathcal{K}} d(R, K).$$

- The basic idea is that any response pattern is assumed to be generated by a close knowledge state
  - leads to explicit (i. e., non-iterative) estimators of the error probabilities
  - minimizes the number of response errors and thus counteracts an inflation of careless error and lucky guess probabilities
- A previously suggested implementation of this idea by Schrepp (1999, 2001) does not allow for item specific estimates.

# Minimum discrepancy method

Assumptions

- A knowledge state $K \in \mathcal{K}$ is assigned to a response pattern $R \in \mathcal{R}$ only if the distance $d(R, K)$ is minimal
- Each of the minimal discrepant knowledge states is assigned with the same probability

$$\hat{P}(K \mid R) = \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}}$$

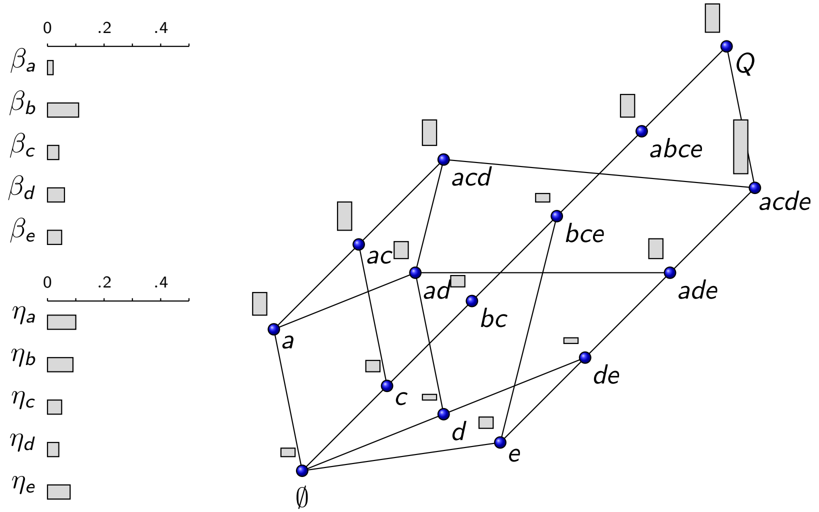
with

$$i_{RK} = \begin{cases} 1 & d(R, K) = d(R, \mathcal{K}) \\ 0 & \text{otherwise} \end{cases}$$





# Example: Minimum discrepancy estimation








# Example: Minimum discrepancy estimation

```
blim(matter97$K, matter97$N.R, method="MD")

Number of iterations: 1
Goodness of fit (2 log likelihood ratio):
        G2(7) = 384.95, p = 0

Minimum discrepancy distribution (mean = 0.2858)
   0    1
1157  463

Mean number of errors (total = 0.2858)
careless error    lucky guess
     0.1269547      0.1588477
```





# Minimum discrepancy ML estimation

Modified EM algorithm

- Modify the E-step in the EM algorithm to implement the restriction

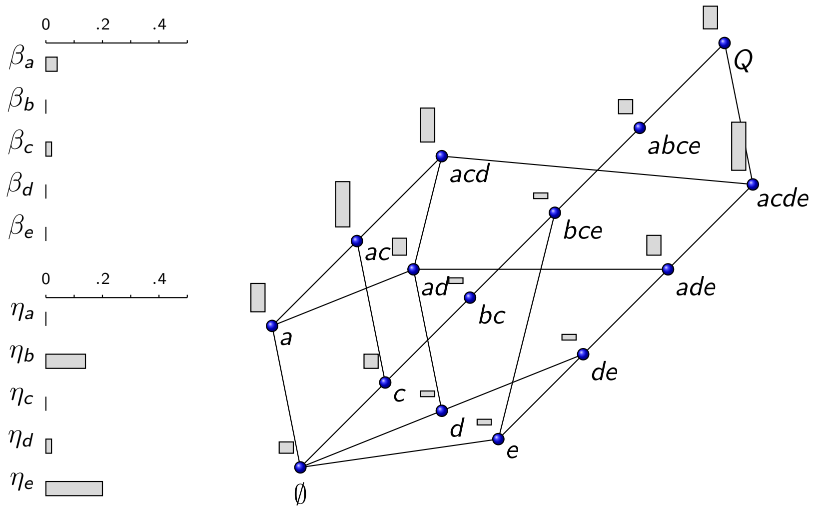
$$m_{RK} = E(M_{RK} \mid N_R, \hat{\beta}^{(t)}, \hat{\eta}^{(t)}, \hat{\pi}^{(t)}) = 0$$

  whenever $d(R, K) > d(R, \mathcal{K})$.
- This leads to

$$m_{RK} = N_R \cdot \frac{i_{RK} \cdot P(K \mid R, \hat{\beta}^{(t)}, \hat{\eta}^{(t)}, \hat{\pi}^{(t)})}{\sum_{K \in \mathcal{K}} i_{RK} \cdot P(K \mid R, \hat{\beta}^{(t)}, \hat{\eta}^{(t)}, \hat{\pi}^{(t)})}$$

- The M-step proceeds as usual.

## Example: Minimum discrepancy ML estimation








## Example: Minimum discrepancy ML estimation

```
blim(matter97$K, matter97$N.R, method="MDML")

Number of iterations: 133
Goodness of fit (2 log likelihood ratio):
        G2(7) = 310.32, p = 0

Minimum discrepancy distribution (mean = 0.2858)
   0    1
1157  463

Mean number of errors (total = 0.2858)
careless error    lucky guess
     0.0481207      0.2376818
```





## Outlook

The pks package features

- Fitting and testing basic local independence models (BLIMs)
- Response generation from a given BLIM object
- Maximum likelihood, minimum discrepancy, and MDML estimation

Work in progress

- Sampling distributions for goodness of fit tests
- Generalized MDML criterion: tradeoff between likelihood maximization and error minimization
- ...





**Thank you for your attention**

*florian.wickelmaier@uni-tuebingen.de*

`http://CRAN.r-project.org/package=pks`
`http://r-forge.r-project.org/projects/pks/`

Florian Wickelmaier [cre, aut]
Jürgen Heller [aut]
Pasquale Anselmi [ctb]

# References


Doignon, J.-P. & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*, 175–196.

Doignon, J.-P. & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin: Springer.

Schrepp, M. (1999). Extracting knowledge structures from observed data. *British Journal of Mathematical and Statistical Psychology*, *52*, 213–224.

Schrepp, M. (2001). A method for comparing knowledge structures concerning their adequacy. *Journal of Mathematical Psychology*, *45*, 480–496.

Stefanutti, L. & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, *74*, 83–96.

Taagepera, M., Potter, F., Miller, G. E., & Lakshminarayan, K. (1997). Mapping students' thinking patterns by the use of the knowledge space theory. *International Journal of Science Education*, *19*, 283–302.






# Example: Generalized MDML estimation

```
blim(matter97$K, matter97$N.R, method="MDML", incradius=1)

Number of knowledge states: 15
Number of response patterns: 32
Number of respondents: 1620

Method: Minimum discrepancy maximum likelihood
Number of iterations: 1679
Goodness of fit (2 log likelihood ratio):
        G2(7) = 47.11, p = 5.3126e-08

Minimum discrepancy distribution (mean = 0.2858)
   0    1
1157  463

Mean number of errors (total = 0.75031)
careless error    lucky guess
     0.5014542      0.2488576
```





# Example: Generalized MDML estimation