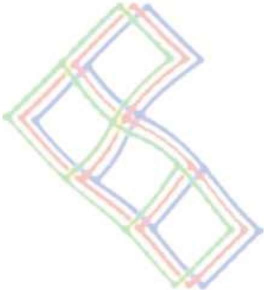
EBERHARD KARLS UNIVERSITÄT TÜBINGEN
Faculty of Science
Department of Psychology



Parameter estimation in probabilistic knowledge structures

Jürgen Heller & Florian Wickelmaier

Psychoco 2011

Outline

Introduction

Knowledge Structures

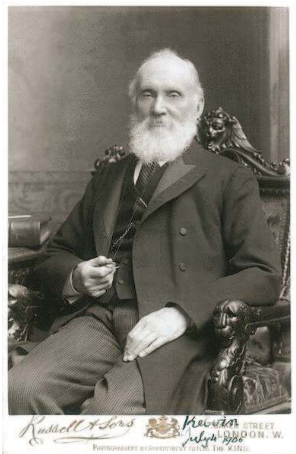
Parameter Estimation
- Maximum Likelihood Estimation
- Minimum Discrepancy Method
- Minimum Discrepancy ML Estimation

Implementation in R

Concluding Remarks

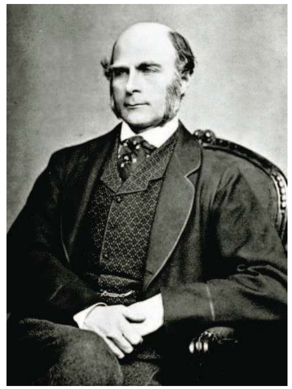
. . . Numbers in Science . . .

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you are scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be." (William Thomson Kelvin, 1889)



. . . Numbers in Psychology . . .

"Anthropometry, or the art of measuring the physical and mental faculties of human beings, enables a shorthand description of any individual by measuring a small sample of his dimensions and qualities. This will sufficiently define his bodily proportions, his massiveness, strength, agility, keenness of senses, energy, health, intellectual capacity and mental character, and will constitute concise and exact numerical values for verbose and disputable estimates." (Francis Galton, 1905)



. . . Numbers in Psychology . . .

So, imagine that some committee of experts has carefully designed an 'Athletic Quotient' or 'A.Q.' test, intended to measure athletic prowess. Suppose that three exceptional athletes have taken the test, say Michael Jordan, Tiger Woods and Pete Sampras. Conceivably, all three of them would get outstanding A.Q.'s. But these high scores equating them would completely misrepresent how essentially different from each other they are. One may be tempted to salvage the numerical representation and argue that the assessment, in this case, should be multidimensional. However, adding a few numerical dimensions capable of differentiating Jordan, Woods and Sampras would only be the first step in a sequence. Including Greg Louganis or Pele to the evaluated lot would require more dimensions, and there is no satisfactory end in sight. (Falmagne et al., 2006, p. 63)

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

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

Example

Study on Fear Symptoms (Stouffer et al., 1950)

- U.S soldiers who have been under fire report different physical reactions to the dangers of battle ($N = 93$)
- Knowledge domain $Q = \{a, b, c, d\}$ (item "solved" when options in parenthesis are chosen)

a	Violent pounding of the heart (sometimes, or often)
b	Feeling of weakness, or feeling faint (sometimes, or often)
c	Urinating in pants (sometimes, or often)
d	Losing control of the bowels (once, sometimes, or often)

Example

Study on Fear Symptoms (Stouffer et al., 1950)

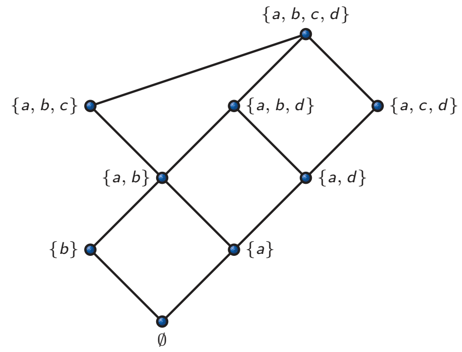
- Absolute frequencies N_R of response patterns

item				
a	b	c	d	N_R
1	0	0	0	40
0	0	0	0	7
0	1	0	0	2
1	0	0	1	3
1	1	0	0	23
1	0	1	1	1
1	1	0	1	9
1	1	1	0	1
1	1	1	1	7
84	42	9	20	

Example

Study on Fear Symptoms (Stouffer et al., 1950)

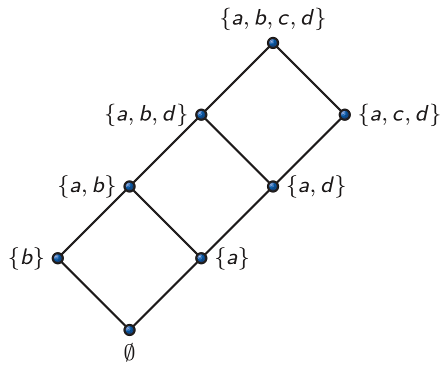
- Hasse-Diagram of response patterns

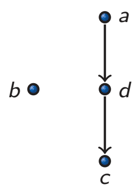


Example

Study on Fear Symptoms (Stouffer et al., 1950)

- Hasse-Diagram of response patterns (excluding $\{a, b, c\}$)





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 (Falmagne & Doignon, 1988a, 1988b)

- 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

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

Local stochastic independence

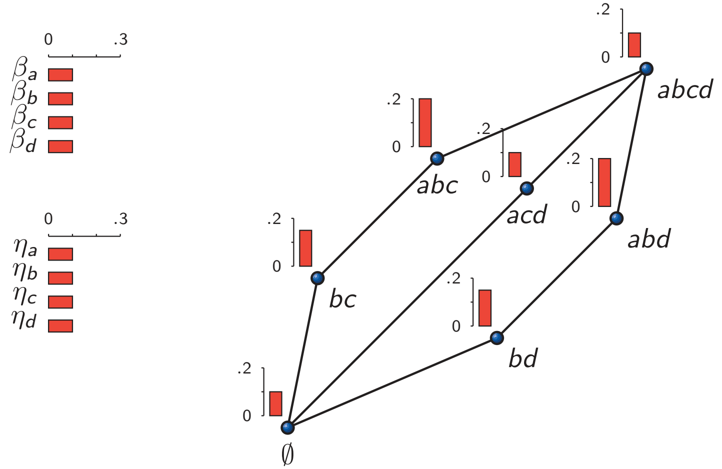
Assumptions

- 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
    β_q of a careless error
    η_q of a lucky guess
- The probability of the response pattern R given the knowledge state K reads

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

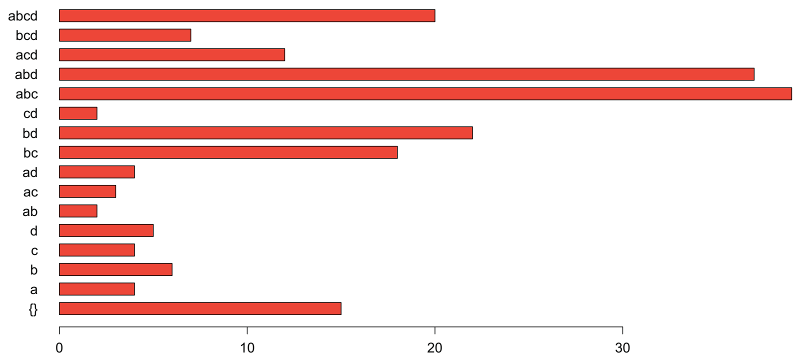
Theory

Probabilistic Knowledge Structure on $Q = \{a, b, c, d\}$



Data

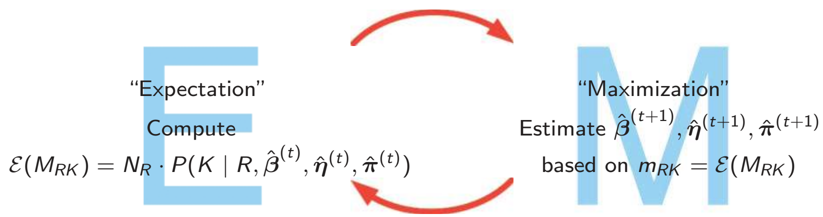
Observed frequencies N_R of response patterns $R \subseteq Q = \{a, b, c, d\}$



Maximum Likelihood Estimation

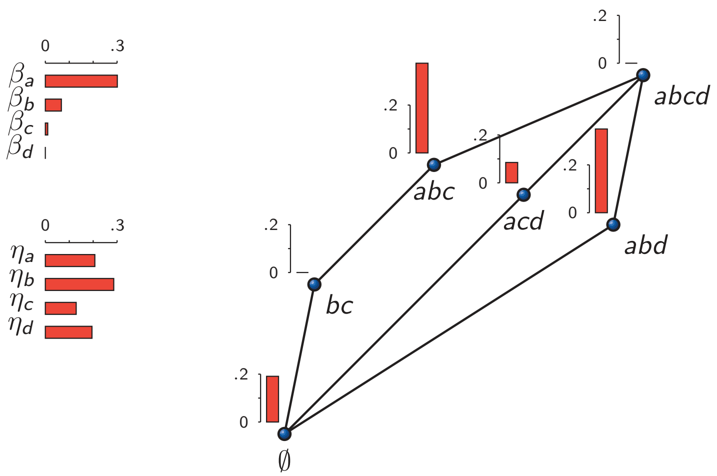
EM Algorithm

- 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)



Maximum Likelihood Estimation

ML Estimates for Example Data



Maximum Likelihood Estimation

ML Estimates for Example Data

- Global Fit
  - Number of iterations (initial values: uniform distribution on knowledge states, error rates 0.1)

$$2945$$

  - log-Likelihood (multinomial model: -477.674)

$$\mathcal{L} = -479.534$$

  - Likelihood ratio corresponds to $\chi^2(2) = 3.722$, $p = 0.156$ (asymptotic theory!)
  - Expected number of errors (minimum: 0.295)

$$\mathcal{E}(T) = 0.595, \quad \mathcal{E}(E) = 0.297, \quad \mathcal{E}(G) = 0.298$$

Maximum Likelihood Estimation

Interim Conclusions

- 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 β_q and lucky guess rates η_q, $q \in Q$
    - What we want: Good fit with small values of β_q and η_q
- 'Workaround'
  - Order constrained ML estimation (Stefanutti & Robusto, 2009)
    - Parameter estimation in a restricted parameter space by applying the EM algorithm subject to order constraints setting upper bounds to the error rates
    - How to motivate the upper bounds?
    - Problems may arise when the estimates fall on the boundary of the parameter space

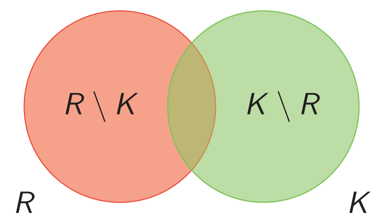
Minimum Discrepancy Method

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 and specifies the number of items that are elements of either, but not both sets R and K



Minimum Discrepancy Method

Rationale

- For a given response pattern R then consider the minimum of the symmetric distances $d(R, K)$ 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) is flawed

Minimum Discrepancy Method

Assumptions

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

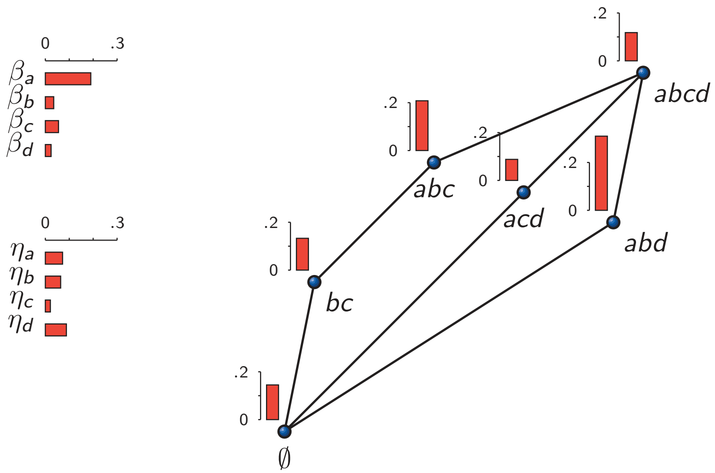
$$\widehat{P}(K \mid R) = \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}}$$

with

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

Minimum Discrepancy Method

MD Estimates for Example Data



Minimum Discrepancy Method

MD Estimates for Example Data

- Global Fit
  - Number of iterations

1

  - log-Likelihood (multinomial model: −477.674)

$$\mathcal{L} = -517.573$$

  - Expected number of errors (minimum: 0.295)

$$\mathcal{E}(T) = 0.295, \quad \mathcal{E}(E) = 0.208, \quad \mathcal{E}(G) = 0.087$$

Minimum Discrepancy ML Estimation

Modified EM Algorithm

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

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

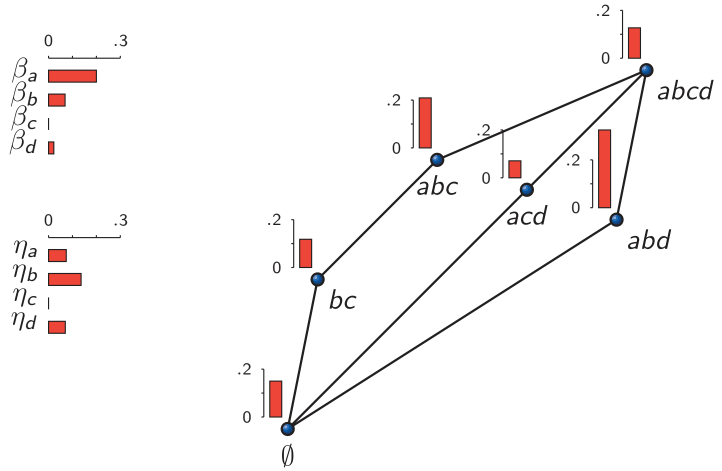
  whenever $d(R, K) > d(R, \mathcal{K})$
- This leads to

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

- The M-step proceeds as usual

Minimum Discrepancy ML Estimation

MDML Estimates for Example Data



Minimum Discrepancy ML Estimation

MDML Estimates for Example Data

- Global Fit
  - Number of iterations (initial values: uniform distribution on knowledge states, error rates 0.1)

$$181$$

  - log-Likelihood (multinomial model: -477.674)

$$\mathcal{L} = -489.626$$

  - Expected number of errors (minimum: 0.295)

$$\mathcal{E}(T) = 0.295, \quad \mathcal{E}(E) = 0.212, \quad \mathcal{E}(G) = 0.083$$

Towards Package pks

Function `mdml()`

```
mdml(K, N.R, R = t(sapply(strsplit(names(N.R), ""), as.numeric)),
    pi = NULL, beta = NULL, eta = NULL,
    type = c("both", "error", "guessing"), equal = FALSE, radius.inc = 0,
    method = c("ML", "MD", "MDML"), tol=0.0000001, maxiter = 5000)
```

- `K` knowledge structure (matrix)
- `N.R` vector of absolute frequencies of observed response patterns
- `R` observed response patterns (matrix)
- `pi`, `beta`, `eta` vectors of initial parameter values
- `type` careless errors and/or lucky guesses occur
- `radius.inc` increment to include knowledge states beyond the minimum distance
- `method` ML, or MD, or MDML estimation

Towards Package pks

Example

```
> K
     [,1] [,2] [,3] [,4]
0000    0    0    0    0
0110    0    1    1    0
1110    1    1    1    0
1101    1    1    0    1
1011    1    0    1    1
1111    1    1    1    1

> N.R

0000 0001 0010 0011 0100 0101 0110 0111 1000 1001 1010
  15    5    4    2    6   22   18    7    4    4    3
1011 1100 1101 1110 1111
  12    2   37   39   20

> r.mdml <- mdml(K, N.R, type="both", method="MDML")
> print(r.mdml)
```


Towards Package pks

Example (cont'd)

```
Parameter estimation in probabilistic knowledge structures
Method: Minimum discrepancy maximum likelihood

Number of knowledge states: 6
Number of response patterns: 16
Number of respondents: 200

Minimum discrepancy distribution (Mean = 0.295)
  0   1
141  59

Number of iterations: 181

Mean number or errors (total = 0.295)
careless error    lucky guess
     0.2117973      0.0832027

log-Likelikood: -489.6255
```


Towards Package pks

Example (cont'd)

```
Distribution of knowledge states
           pi
0000 0.150000
0110 0.118203
1110 0.208634
1101 0.324999
1011 0.071367
1111 0.126797

Error and guessing parameters
        beta        eta
a 2.0060e-01 7.4570e-02
b 6.8881e-02 1.3552e-01
c 1.4925e-06 3.0358e-33
d 2.1726e-02 6.9632e-02
```


Concluding Remarks

- The MDML estimators
  - minimize the expected total number of response errors
  - maximize the likelihood subject to the above constraint
- Work in progress
  - Generalize the minimum discrepancy criterion
    - Include knowledge states that are at minimum distance plus some increment
    - Generalize the indicator function i_{RK} to

$$i_{RK} = F\,[d(R,K), d(R,\mathcal{K})]$$

      with a real valued function F, non-increasing in its first argument, and non-decreasing in its second argument
  - Large scale applications
  - identifiability in probabilistic knowledge structures
  - pks functions `summary()`, `simulate.pks()`, ...

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. *New York: Springer*.

Falmagne, J.-C., & Doignon, J.-P. (1988a). A class of stochastic procedures for the assessment of knowledge. *British Journal of Mathematical and Statistical Psychology*, **41**, 1-23.

Falmagne, J.-C., & Doignon, J.-P. (1988b). A Markovian procedure for assessing the state of a system. *Journal of Mathematical Psychology*, **32**, 232-258.

Falmagne, J.-C., Doignon, J.-P., Cosyn, E., & Thiery, N. (2006). The Assessment of Knowledge in Theory and in Practice. *Lecture Notes in Computer Science*, **3874**, 61-79.

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.