Probabilistic Knowledge Structures I

Jürgen Heller & Florian Wickelmaier

University of Tübingen

November 2016

Outline

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

Probabilistic Theory

Motivation

- In practical applications we cannot assume that a student's response to an item is correct if and only if the student masters it (i.e. if the item is an element of the respective knowledge state)
- There are two types of response errors
  - careless error, i.e. the response is incorrect although the item is contained in the knowledge state
  - lucky guess, i.e. the response is correct although the item is not contained in the knowledge state
- In any case, we need to dissociate knowledge states and response patterns
  - R denotes a response pattern, which is a subset of Q
  - $\mathcal{R} = 2^Q$ denotes the set of all possible response patterns

Probabilistic Theory

Motivation

- In this approach
  - the knowledge state K is a latent construct
  - the response pattern R is a manifest indicator to the knowledge state
- The conclusion from the observable response pattern to the unobservable knowledge state can only be of a stochastic nature
- This requires to introduce a probabilistic framework, which may also capture the fact that the knowledge states will not occur with equal probability
  - A probability distribution on the knowledge states provides information highly relevant to knowledge assessment

Probabilistic Framework

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 (PKS) 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 $P(K)$ 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

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

Local Stochastic Independence

Assumptions

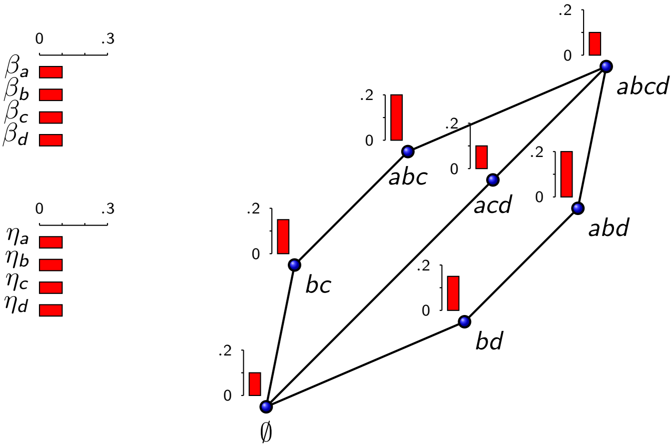
- Given the knowledge state K of a person
  - the responses are stochastically independent over problems
  - the response to each problem q only depends on the probabilities
    β_q of a careless error
    η_q of a lucky guess
- The probability of the response pattern R given the knowledge state K reads

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

- A PKS satisfying these assumptions is called a basic local independence model (BLIM)

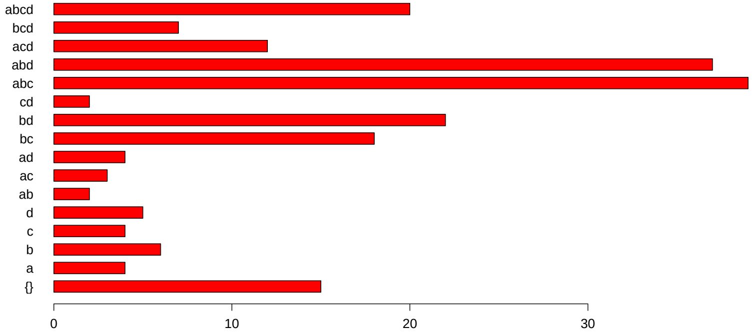
Theory

Basic local independence model on $Q = \{a, b, c, d\}$



Data

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



Exercise

Compute $P(R \mid K)$ for

- $R = \{a, b\}$ and $K = \{a, c, d\}$
- $R = \{d\}$ and $K = \{b, c\}$
- $R = \{a, d\}$ and $K = \{b, c\}$

and $Q = \{a, b, c, d\}$, $\beta_q = 0.1$ and $\eta_q = 0.05$ $(\forall q \in Q)$

Maximum Likelihood Estimation

Basics

- The data consist of a vector $\mathbf{x} = (x_1, \ldots, x_N)$ specifying for each subject in a sample of size N the given response pattern
- From this we can derive the absolute frequencies N_R of the patterns $R \in \mathcal{R}$
- For a given knowledge structure $\mathcal{K}$ let $\boldsymbol{\beta} = (\beta_q)_{q \in Q}$ denote the parameter vector of careless error rates, $\boldsymbol{\eta} = (\eta_q)_{q \in Q}$ the parameter vector of all lucky guess rates, and $\boldsymbol{\pi} = (\pi_K)_{K \in \mathcal{K}}$ the parameter vector of all state probabilities $\pi_K = P(K)$
- Then the likelihood is given by

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi} \mid \mathbf{x}) = \prod_{R \in \mathcal{R}} P(R \mid \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi})^{N_R}$$

Maximum Likelihood Estimation

Basics

- Determining the maximum likelihood estimates (MLEs) requires to compute the partial derivatives of the (log-)likelihood with respect to each of the parameters collected in the vectors $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\pi}$
- The problems concerning the analytical tractability of this derivation arise from the fact that $P(R \mid \mathcal{K}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi})$ actually is the sum

$$P(R \mid \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi}) = \sum_{K \in \mathcal{K}} P(R, K \mid \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi})$$

- Doignon & Falmagne (1999) thus resort to numerical optimization techniques
- A (partial) solution to this problem is provided by the so-called EM algorithm (Heller & Wickelmaier, 2013; Stefanutti & Robusto, 2009)

Maximum Likelihood Estimation

EM algorithm

- The EM algorithm is an iterative optimization method for providing MLEs of unknown parameters, which proceeds within a so-called incomplete-data framework
  - Considering the given data as incomplete, and (artificially) extending them by including actually unobservable variables ('unknown data') often facilitates the computation of the MLEs considerably
- In the present context we assume that for each subject we not only observe the given response pattern R, but also know the respective knowledge state K
  - Let $\mathbf{y} = (y_1, \ldots, y_N)$ denote the vector of the knowledge states
- We call $\mathbf{x}$ the incomplete data, and $(\mathbf{x}, \mathbf{y})$ the complete data, for which we now have available the absolute frequencies M_{RK} of subjects who are in state K and produce pattern R

Maximum Likelihood Estimation

EM algorithm

- The likelihood of the complete data then reads

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi} \mid \mathbf{x}, \mathbf{y}) = \prod_{R \in \mathcal{R}} \prod_{K \in \mathcal{K}} P(R, K \mid \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi})^{M_{RK}}$$

  with $N = \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} M_{RK}$
- It is routine to derive the MLEs starting from

$$\ln \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi} \mid \mathbf{x}, \mathbf{y}) = \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} M_{RK} \cdot [\ln P(R \mid K, \boldsymbol{\beta}, \boldsymbol{\eta}) + \ln P(K \mid \mathcal{K}, \boldsymbol{\pi})]$$

- As the term containing $\boldsymbol{\beta}, \boldsymbol{\eta}$ and that containing $\boldsymbol{\pi}$ are not related, we can maximize them independently

Maximum Likelihood Estimation

EM algorithm

- Plugging in local independence, partial derivation with respect to β_q yields

$$\frac{\partial \ln \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\pi} \mid M_{RK}, R \in \mathcal{R}, \mathcal{K})}{\partial \beta_q} = \sum_{R,K: q \in K \setminus R} \frac{M_{RK}}{\beta_q} - \sum_{R,K: q \in R \cap K} \frac{M_{RK}}{1 - \beta_q}$$

- Setting this expression to zero and solving for β_q yields the estimate

$$\hat{\beta}_q = \frac{\sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} M_{RK}}{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}_q} M_{RK}}$$

Maximum Likelihood Estimation

EM algorithm

- In complete analogy to this derivation we obtain

$$\hat{\eta}_q = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\overline{q}}} M_{RK}}{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}_{\overline{q}}} M_{RK}}$$

- The estimates of the π_K are derived by introducing the Lagrange multiplier λ with the constraint that $\sum_{K' \in \mathcal{K}} \pi_{K'} = 1$

$$\frac{\partial}{\partial \pi_K} \left[\sum_{R \in \mathcal{R}} \sum_{K' \in \mathcal{K}} M_{RK} \cdot \ln \pi_{K'} + \lambda \left(\sum_{K' \in \mathcal{K}} \pi_{K'} - 1 \right) \right] = 0$$

which provides

$$\sum_{R \in \mathcal{R}} M_{RK} \frac{1}{\pi_K} + \lambda = 0$$

Maximum Likelihood Estimation

EM algorithm

- Rearranging terms we get

$$\lambda \pi_K = - \sum_{R \in \mathcal{R}} M_{RK}$$

and from summing both sides over $\mathcal{K}$ we conclude $\lambda = -N$

- This means that

$$\hat{\pi}_K = \frac{\sum_{R \in \mathcal{R}} M_{RK}}{N}$$

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)

E

"Expectation"

Compute

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

M

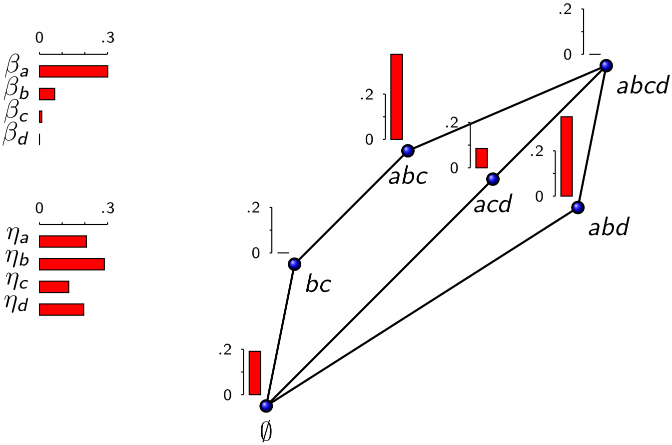
"Maximization"

Estimate $\hat{\boldsymbol{\beta}}^{(t+1)}, \hat{\boldsymbol{\eta}}^{(t+1)}, \hat{\boldsymbol{\pi}}^{(t+1)}$

based on $m_{RK} = \mathcal{E}(M_{RK})$

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?
    - Let parameter estimation be informed by the underlying knowledge structure

Minimum Discrepancy Method

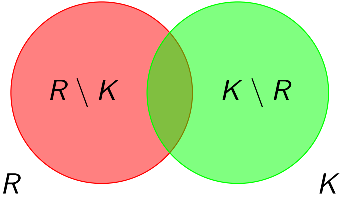
Rationale

- The method is based on a measure of discrepancy describing how closely an observed response pattern corresponds to a knowledge structure $\mathcal{K}$ (cf. Kambouri, 1991)
- The basic idea is that any response pattern is assumed to be generated by a knowledge state that is close to it (in some sense specified below)
- It can be shown that this assumption
  - minimizes the number of response errors and thus counteracts an inflation of careless error and lucky guess probabilities
  - leads to explicit (i.e. non-iterative) estimators of the error probabilities
- A previously suggested implementation of this idea by Schrepp (1999, 2001) is flawed

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

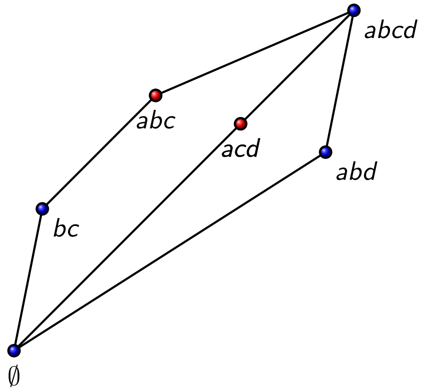
Rationale

- For a given response pattern R then consider the minimum of the symmetric distances $d(R, K)$ between R and all the knowledge states $K \in \mathcal{K}$
$$d(R, \mathcal{K}) = \min_{K \in \mathcal{K}} d(R, K)$$
- Treating $d(R, \mathcal{K})$ as the distance of the response pattern R to the knowledge structure $\mathcal{K}$ may be interpreted as taking an optimistic point of view assuming that only a minimum number of (unavoidable) response errors occur
- In fact, it can be shown that the subsequently outlined approach minimizes the expected total number of response errors

Minimum Discrepancy Method

Example

- Knowledge states minimally discrepant to the response pattern $R = \{a, c\}$



Exercise (in R)

For the example on the previous slide and $R = \{a, c\}$

- Compute $d(R, K)$ for all $K \in \mathcal{K}$
- Show that $d(R, \mathcal{K}) = 1$

Now do the calculations in R

- Specify R as a binary (0/1) vector
- Specify $\mathcal{K}$ as a binary matrix
- Use the `xor()` function to compute $d(R, K)$

Minimum Discrepancy Method

Rationale

- Let the random variables T, E, and G denote the total number of errors, the number of careless errors, and the number of lucky guesses, respectively
- For all $R \in \mathcal{R}$ and $K \in \mathcal{K}$ we have

$$T(R,K) = d(R,K), \quad E(R,K) = |K \setminus R|, \quad G(R,K) = |R \setminus K|$$

- This implies

$$\begin{aligned}
\mathcal{E}(T) &= \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} d(R,K) \cdot P(R,K) \\
&= \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} |K \setminus R| \cdot P(R,K) + \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} |R \setminus K| \cdot P(R,K) \\
&= \mathcal{E}(E) + \mathcal{E}(G)
\end{aligned}$$

Minimum Discrepancy Method

Rationale

- Minimizing $\mathcal{E}(T)$ requires to restrict the sum to those pairs R, K with minimum symmetric distance $d(R, K) = d(R, \mathcal{K})$
- This restriction is obtained by constraining the conditional probabilities $P(K \mid R)$ to satisfy

$$P(K \mid R) = 0 \text{ whenever } d(R, K) > d(R, \mathcal{K})$$

- Estimating parameters subject to this constraint lies at the core of the minimum discrepancy method, which is implemented through the subsequently outlined assumptions

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

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

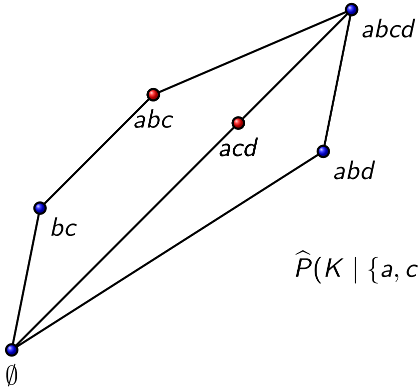
with

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

Minimum Discrepancy Method

Example

- $\widehat{P}(K \mid R)$ for response pattern $R = \{a, c\}$



$$\widehat{P}(K \mid \{a, c\}) = \begin{cases} 1/2 & \text{for } K = \{a, b, c\} \\ 1/2 & \text{for } K = \{a, c, d\} \\ 0 & \text{otherwise} \end{cases}$$

Minimum Discrepancy Method

MD estimators

- Probability of a knowledge state

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

provides the estimate

$$\widehat{\pi}_K = \sum_{R \in \mathcal{R}} \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}} \cdot \frac{N_R}{N}$$

Minimum Discrepancy Method

MD estimators

- Careless error rate

$$\beta_q = P(\mathcal{R}_{\bar{q}} \mid \mathcal{K}_q)$$

with

$$\begin{aligned}\mathcal{K}_q &= \{K \in \mathcal{K} \mid q \in K\}\\ \mathcal{R}_{\bar{q}} &= \{R \in \mathcal{R} \mid q \notin R\}\end{aligned}$$

provides the estimate

$$\widehat{\beta}_q = \frac{\sum_{K \in \mathcal{K}_q} \sum_{R \in \mathcal{R}_{\bar{q}}} \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}} \cdot N_R}{\sum_{K \in \mathcal{K}_q} \sum_{R \in \mathcal{R}} \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}} \cdot N_R}$$

Minimum Discrepancy Method

MD estimators

- Lucky guess rate

$$\eta_q = P(\mathcal{R}_q \mid \mathcal{K}_{\bar{q}})$$

with

$$\begin{aligned} \mathcal{K}_{\bar{q}} &= \{K \in \mathcal{K} \mid q \notin K\} \\ \mathcal{R}_q &= \{R \in \mathcal{R} \mid q \in R\} \end{aligned}$$

provides the estimate

$$\widehat{\eta}_q = \frac{\sum_{K \in \mathcal{K}_{\bar{q}}} \sum_{R \in \mathcal{R}_q} \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}} \cdot N_R}{\sum_{K \in \mathcal{K}_{\bar{q}}} \sum_{R \in \mathcal{R}} \frac{i_{RK}}{\sum_{K \in \mathcal{K}} i_{RK}} \cdot N_R}$$

Exercise

Let

$$
\begin{aligned}
Q &= \{a, b\} \\
\mathcal{K} &= \{\emptyset, \{a\}, Q\} \\
\mathcal{R} &= \{\emptyset, \{a\}, \{b\}, Q\},
\end{aligned}
$$

and the response frequencies $N_\emptyset = N_{\{a\}} = N_Q = 3$ and $N_{\{b\}} = 1$ $(N = 10)$

Using the minimum discrepancy method

- Estimate π_K for all $K \in \mathcal{K}$
- Show that $\hat{\beta}_a = 1/13$

Minimum Discrepancy Method

Properties of the MD estimators

- MD estimators minimize the expected number of response errors

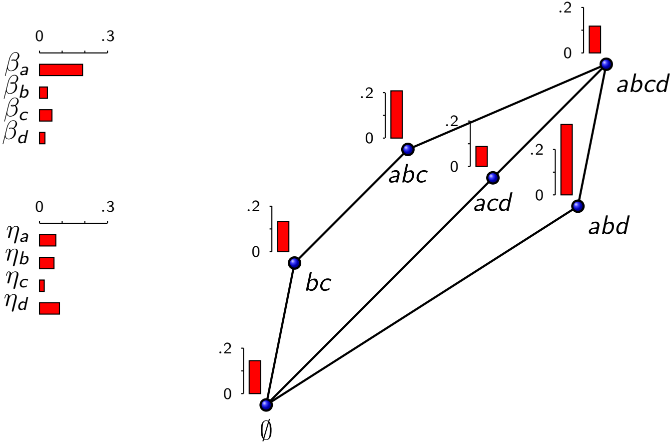
$$
\begin{aligned}
\mathcal{E}(T) &= \mathcal{E}(E) + \mathcal{E}(G) \\
&= \sum_{q \in Q} \beta_q \cdot \pi(\mathcal{K}_q) + \sum_{q \in Q} \eta_q \cdot \pi(\mathcal{K}_{\bar{q}})
\end{aligned}
$$

with $\pi(\mathcal{M}) = \sum_{K \in \mathcal{M}} \pi_K$ for all $\mathcal{M} \subseteq \mathcal{K}$

- Evaluating this expression for the MD estimators provides a lower bound to the expectation resulting for the ML estimators
- This does not mean that the individual MD estimators $\widehat{\beta}_q$ and $\widehat{\eta}_q$ of the careless error and lucky guess rates necessarily provide lower bounds for the respective ML estimators

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 Method

Interim conclusions

- Advantages
  - Explicit estimators (computationally efficient)
  - Avoids inflating the careless error and lucky guess rates
  - Minimum properties of MD estimators identified
- Drawback
  - Parameter estimation ignores likelihood
- Desired Extension
  - Integrating the MD criterion within ML estimation

Minimum Discrepancy ML Estimation

Modified EM algorithm

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

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

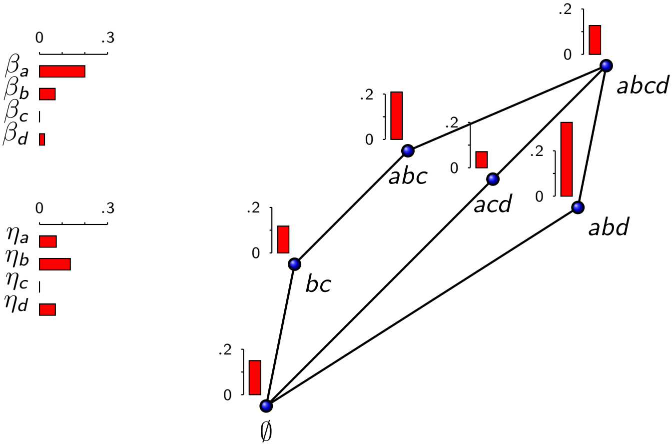
  whenever $d(R, K) > d(R, \mathcal{K})$
- This leads to

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

- 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)

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

Interim Conclusions

- The MDML estimators
  - minimize the expected total number of response errors
  - maximize the likelihood subject to the above constraint
- Comparing goodness of fit of ML with MDML estimation allows for quantifying the amount explained by lucky guesses and careless errors
- Work in progress
  - Generalize the minimum discrepancy criterion
    - Include knowledge states that are at minimum distance plus some increment, or even more general ...
    - Generalize the indicator function i_{RK}, e.g. to

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

      with a real valued function F, non-increasing in its first argument, and non-decreasing in its second argument
  - Large scale applications
- Identifiability ...

R Package pks

(Wickelmaier [cre, aut], Heller [aut], Anselmi [ctb])

Features

- Fitting and testing basic local independence models (BLIMs)
- Response generation from a given BLIM
- 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
- ...

Download

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

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
`is.forward.graded` `is.backward.graded`	forward- or backward-gradedness
`delineate`	knowledge structure delineated by skill function

Example: Maximum Likelihood Estimation

```
blim1 <- blim(endm$K2, endm$N.R, method = "ML")

Number of iterations: 2945
Goodness of fit (2 log likelihood ratio):
G2(2) = 3.7215, p = 0.15556

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

Mean number of errors (total = 0.59533)
careless error    lucky guess
     0.2967356      0.2985958

logLik(blim1)
'log Lik.' -479.5344 (df=13)
```

Example: Minimum Discrepancy Estimation

```
blim2 <- blim(endm$K2, endm$N.R, method = "MD")

Number of iterations: 1
Goodness of fit (2 log likelihood ratio):
G2(2) = 29.491, p = 3.9455e-07

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

Mean number of errors (total = 0.295)
careless error    lucky guess
        0.2075         0.0875

logLik(blim2)
'log Lik.' -492.4191 (df=13)
```

Example: Minimum Discrepancy ML Estimation

```
blim3 <- blim(endm$K2, endm$N.R, method = "MDML")

Number of iterations: 181
Goodness of fit (2 log likelihood ratio):
G2(2) = 23.904, p = 6.4469e-06

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

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

logLik(blim3)
'log Lik.' -489.6255 (df=13)
```

Exercise in R

Look at the help file for the endm data in the pks package

1. Simulate responses using a BLIM based on the knowledge structure `endm$K` and the true values for π_K, β_q, and η_q as provided in the help-file example
2. Re-estimate the parameters from the simulated responses using ML estimation. How big does the sample size have to be in order for the ML estimates to become close to the true values?
3. Specify your own BLIM: make up a knowledge structure K and true parameter values; generate responses from this BLIM
4. Re-estimate the parameters from the simulated responses using ML estimation

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.

A Subdomain of Physics: Conservation of Matter (2)

(Taagepera et al., 1997)

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.

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.

A Subdomain of Physics: Conservation of Matter (3)

(Taagepera et al., 1997)

e) Photosynthesis can be described as:

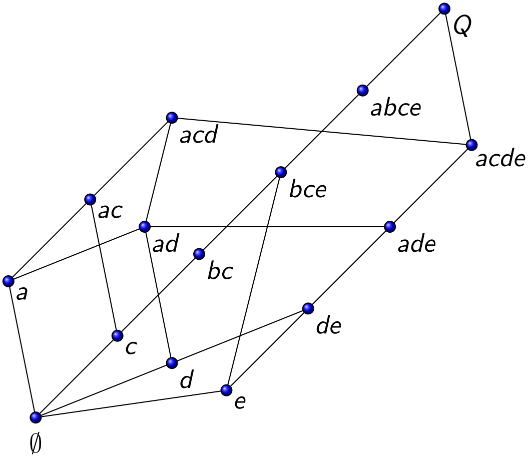
$$\text{WATER} + \text{CARBON DIOXIDE} \xrightarrow[\text{sunlight}]{\text{chlorophyll}} \text{GLUCOSE}$$

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.

Conservation of Matter: Knowledge Structure

(Taagepera et al., 1997)



Exercise in R

Analyze the Taagepera data in the pks package

1. Display the response frequencies `matter97$N.R` graphically and try to identify probable knowledge states
2. Fit the BLIM based on the knowledge structure `matter97$K` to the responses using ML, MD, and MDML estimation (ML estimation may take about 1 min)
3. Is there an indication of inflated error probabilities in the ML estimates?
4. How much of the goodness of fit of the ML-estimated BLIM can be maximally explained by lucky guesses and careless errors?

Skill Function

- Set S of (abstract) skills, which are relevant to the solution of the items
- To each item $q \in Q$ we assign a nonempty collection $\mu(q)$ of nonempty pairwise incomparable subsets of skills (w.r.t. $\subseteq$)

$$q \mapsto \mu(q)$$

- Each subset

$$C \in \mu(q)$$

  is called a competency, and identifies a subset of skills sufficient for solving item q
- (Q, S, μ) is called a skill function

Skill Function

Example

$$
\begin{aligned}
\mu(a) &= \{\{s, t\}, \{s, u\}\} \\
\mu(b) &= \{\{u\}\} \\
\mu(c) &= \{\{s\}, \{t\}\} \\
\mu(d) &= \{\{t\}\}
\end{aligned}
$$

Available skills

$t \quad u$

Items

a	b	c	d
✗	✓	✓	✓

Problem Function

- To each subset $T \subseteq S$ of skills there corresponds a subset of items which can be solved within T

$$T \mapsto \{q \in Q : \text{ there is a } C \in \mu(q) \text{ such that } C \subseteq T\}$$

- This defines a problem function $p \colon \mathcal{C} \to 2^Q$ with $\mathcal{C} = 2^S$
- The range of a problem function forms (i.e., delineates) a knowledge structure

Proposition (Düntsch & Gediga, 1995)

For a given domain Q and a set of skills S there exists a one-to-one correspondence between the collection of all skill functions μ on Q and the collection of all problem functions p

Competence-Based Basic Local Independence Model

- Let the skill function (Q, S, μ) delineate the knowledge structure $(Q, \mathcal{K})$
- Consider a probability distribution $P_{\mathcal{C}}$ on the powerset $\mathcal{C} = 2^S$
- With p the problem function induced by the skill function μ, a probability distribution on the delineated knowledge structure $\mathcal{K} = p(\mathcal{C})$ is defined by

$$P(K) = \sum_{T \in p^{-1}(\{K\})} P_{\mathcal{C}}(T)$$

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

Competence-Based Basic Local Independence Model

- A CBLIM satisfies the equation

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

for all $R \in \mathcal{R}$, where

$$P(R \mid T) = P(R \mid p(T))$$

for all $T \in \mathcal{C}$

- The right hand side of the latter equation is given by the local independence assumption
- All this may be easily generalized to conceiving $\mathcal{C}$ as an arbitrary subset of 2^S containing $\emptyset$ and S, for which the notion competence structure has been coined

Competence-Based Basic Local Independence Model

Proposition

Let $\mu \colon Q \to 2^{2^S}$ be a skill function and p its induced problem function. Then the following two statements are equivalent.

- *For all $q \in Q$ there is $\mu(q) = \{C\}$ for some $C \subseteq S$,*
- *$p(T_1 \cap T_2) = p(T_1) \cap p(T_2)$ for all $T_1, T_2 \subseteq S$.*

- A skill function μ is said to be a conjunctive skill function if these conditions hold

Competence-Based Basic Local Independence Model

Proposition

Let $\mu\colon Q \to 2^{2^S}$ be a skill function and p its induced problem function. Then the following two statements are equivalent.

- *For all $q \in Q$ each of the competencies $C \in \mu(q)$ is a singleton set,*
- *$p(T_1 \cup T_2) = p(T_1) \cup p(T_2)$ for all $T_1, T_2 \subseteq S$.*

- A skill function μ is said to be a disjunctive skill function if these conditions hold

Correspondences to Cognitive Diagnostic Models

It can be shown that prominent so-called cognitive diagnostic models (Rupp, Templin & Henson, 2010; Tatsuoka, 1990) are equivalent to particular CBLIMs (Heller, Stefanutti, Robusto & Anselmi, 2015)

model		equivalent to
DINA	Deterministic Input Noisy AND-gate	CBLIM on conjunctive skill function
DINO	Deterministic Input Noisy OR-gate	CBLIM on disjunctive skill function
Multiple Strategy DINA		CBLIM

References

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

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., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42, 49-56.

Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press.

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.

Tatsuoka, K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (p. 453-488). Hillsdale, NJ: Lawrence Erlbaum Associates.