

Parameter Estimation in Probabilistic Knowledge Structures – Step by Step

Sep/3/2024

Florian Wickelmaier
University of Tübingen

Jürgen Heller
University of Tübingen

Abstract

This vignette provides some details about the parameter estimation procedures implemented in the pks package.

Keywords: Knowledge spaces, probabilistic knowledge structures, basic local independence model, simple learning model.

Doignon and Falmagne (1999) describe two probabilistic models frequently used in applications of knowledge structure theory: the basic local independence model (BLIM) and the simple learning model (SLM). This vignette provides some details about the parameter estimation procedures implemented in the pks package, which are an adaptation of the expectation-maximization algorithm. Section 1 contains methods for the BLIM. The results are taken from Heller and Wickelmaier (2013) and are repeated here (with intermediate steps) in order to motivate the analogous derivations for the SLM in Section 2.

1. The basic local independence model

1.1. Likelihood and log-likelihood

Let $q \in Q$ be an item, $R \in \mathcal{R} = 2^Q$ be a response pattern, and $K \in \mathcal{K} \subseteq 2^Q$ be a knowledge state. Let N_R be the observed frequency of response pattern R . For the BLIM, the parameter vector $\theta = (\beta, \eta, \pi)$ consists of (vectors of) careless error, lucky guess, and state probabilities. The (incomplete-data) likelihood of θ given all observed response patterns has the multinomial form

$$L(\theta) = \prod_R P(R)^{N_R} = \prod_R \left(\sum_K P(R, K) \right)^{N_R}, \quad (1)$$

which is difficult to maximize because of the sum. Therefore, let M_{RK} denote the unobservable frequency of pattern R resulting from state K , where $N_R = \sum_K M_{RK}$. Then the complete-

data likelihood can be written as

$$L(\theta) = \prod_R \prod_K P(R, K)^{M_{RK}} \quad (2)$$

$$= \prod_R \prod_K (P(R | K) \cdot P(K))^{M_{RK}} \quad (3)$$

$$= \prod_R \prod_K \left(\prod_{q \in K \setminus R} \beta_q \prod_{q \in K \cap R} (1 - \beta_q) \prod_{q \in R \setminus K} \eta_q \prod_{q \in \bar{R} \cap \bar{K}} (1 - \eta_q) \right)^{M_{RK}} \cdot (\pi_K)^{M_{RK}} \quad (4)$$

and the log-likelihood becomes

$$\begin{aligned} \ell(\theta) &= \sum_R \sum_K M_{RK} \left(\sum_{q \in K \setminus R} \log \beta_q + \sum_{q \in K \cap R} \log(1 - \beta_q) + \sum_{q \in R \setminus K} \log \eta_q + \sum_{q \in \bar{R} \cap \bar{K}} \log(1 - \eta_q) \right) \\ &+ \sum_R \sum_K M_{RK} \log \pi_K. \end{aligned} \quad (5)$$

1.2. Careless errors and lucky guesses

Let \mathcal{R}_q and $\mathcal{R}_{\bar{q}}$ denote the subset of response patterns that contain and do not contain q , respectively. Similarly, let \mathcal{K}_q and $\mathcal{K}_{\bar{q}}$ denote the subset of knowledge states that contain and do not contain q , respectively. Then the partial derivative of the log-likelihood with respect to β_q is

$$\frac{\partial \ell(\theta)}{\partial \beta_q} = \sum_R \sum_K \left(\sum_{q \in K \setminus R} \frac{M_{RK}}{\beta_q} - \sum_{q \in K \cap R} \frac{M_{RK}}{1 - \beta_q} \right) \quad (6)$$

$$= \sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} \frac{M_{RK}}{\beta_q} - \sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{M_{RK}}{1 - \beta_q} \quad (7)$$

$$= \sum_R \sum_{K \in \mathcal{K}_q} \frac{M_{RK}}{\beta_q} \cdot (1 - i) - \frac{M_{RK}}{1 - \beta_q} \cdot i, \quad (8)$$

where $i = 1$ if $R \in \mathcal{R}_q$, and $i = 0$ else. Setting the derivative to zero yields

$$\sum_R \sum_{K \in \mathcal{K}_q} \frac{M_{RK}(1 - i)(1 - \hat{\beta}_q) - M_{RK}i\hat{\beta}_q}{\hat{\beta}_q(1 - \hat{\beta}_q)} = 0 \quad (9)$$

$$\sum_R \sum_{K \in \mathcal{K}_q} M_{RK} - M_{RK}\hat{\beta}_q - M_{RK}i + M_{RK}i\hat{\beta}_q - M_{RK}i\hat{\beta}_q = 0 \quad (10)$$

$$\sum_R \sum_{K \in \mathcal{K}_q} M_{RK}(1 - i) = \sum_R \sum_{K \in \mathcal{K}_q} M_{RK}\hat{\beta}_q \quad (11)$$

$$\sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} M_{RK} = \hat{\beta}_q \sum_R \sum_{K \in \mathcal{K}_q} M_{RK} \quad (12)$$

$$\hat{\beta}_q = \frac{\sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} M_{RK}}{\sum_R \sum_{K \in \mathcal{K}_q} M_{RK}}. \quad (13)$$

The partial derivative of the log-likelihood with respect to η_q is

$$\frac{\partial \ell(\theta)}{\partial \eta_q} = \sum_R \sum_K \left(\sum_{q \in R \setminus K} \frac{M_{RK}}{\eta_q} - \sum_{q \in \bar{R} \cap \bar{K}} \frac{M_{RK}}{1 - \eta_q} \right) \quad (14)$$

$$= \sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{M_{RK}}{\eta_q} - \sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{M_{RK}}{1 - \eta_q} \quad (15)$$

$$= \sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{M_{RK}}{\eta_q} \cdot i - \frac{M_{RK}}{1 - \eta_q} \cdot (1 - i), \quad (16)$$

where, as above, $i = 1$ if $R \in \mathcal{R}_q$, and $i = 0$ else. Setting the derivative to zero yields

$$\sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{M_{RK} i (1 - \hat{\eta}_q) - M_{RK} (1 - i) \hat{\eta}_q}{\hat{\eta}_q (1 - \hat{\eta}_q)} = 0 \quad (17)$$

$$\sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK} i - \cancel{M_{RK} i \hat{\eta}_q} - M_{RK} \hat{\eta}_q + \cancel{M_{RK} i \hat{\eta}_q} = 0 \quad (18)$$

$$\sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK} i = \sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK} \hat{\eta}_q \quad (19)$$

$$\sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK} = \hat{\eta}_q \sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK} \quad (20)$$

$$\hat{\eta}_q = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK}}{\sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} M_{RK}}. \quad (21)$$

1.3. State probabilities

Including the constraint $\sum_K \pi_K = 1$ into the log-likelihood using the Lagrange multiplier λ leads to the function

$$\ell(\pi, \lambda) = \sum_R \sum_K M_{RK} \log \pi_K + \lambda \left(\sum_K \pi_K - 1 \right) \quad (22)$$

to be maximized with respect to π_K and λ . Thus,

$$\frac{\partial \ell(\pi, \lambda)}{\partial \pi_K} = \sum_R \frac{M_{RK}}{\pi_K} + \lambda, \quad (23)$$

$$\frac{\partial \ell(\pi, \lambda)}{\partial \lambda} = \sum_K \pi_K - 1. \quad (24)$$

Setting (23) to zero and solving for π_K gives

$$\pi_K = -\frac{\sum_R M_{RK}}{\lambda}. \quad (25)$$

Setting (24) to zero, substituting π_K by (25), and solving for λ gives

$$\hat{\lambda} = -\sum_R \sum_K M_{RK} = -N, \quad (26)$$

which, when substituted back into (25), leads to

$$\hat{\pi}_K = \frac{\sum_R M_{RK}}{N}. \quad (27)$$

1.4. Expectation maximization

The conditional expectation of the unobservable frequencies M_{RK} is

$$E(M_{RK} | N_R) = N_R \cdot P(K | R) \quad (28)$$

$$= N_R \cdot \frac{P(R | K)P(K)}{P(R)} \quad (29)$$

$$= N_R \cdot \frac{P(R | K)P(K)}{\sum_K P(R | K)P(K)}, \quad (30)$$

where, in each iteration, $P(R | K)$ and $P(K)$ are calculated from the current values of $\hat{\beta}_q$, $\hat{\eta}_q$, and $\hat{\pi}_K$. These values are iteratively updated when substituting the M_{RK} by their expectations in (13), (21), and (27).

1.5. Minimum-discrepancy maximum likelihood

Assuming that K is the underlying knowledge state for a response pattern R , the distance

$$d(R, K) = |(R \setminus K) \cup (K \setminus R)| \quad (31)$$

contains the number of response errors in R . Minimizing the number of response errors, a state is assigned to R if its distance takes on the smallest possible value,

$$i_{RK} = \begin{cases} 1 & \text{if } d(R, K) = \min_K d(R, K), \\ 0 & \text{else.} \end{cases} \quad (32)$$

Under this assumption of minimum discrepancy, the conditional probability $P(K | R)$ is estimated by i_{RK}/i_{R+} , where $i_{R+} = \sum_K i_{RK}$. Consequently,

$$\hat{P}(R, K) = \hat{P}(K | R) \cdot \hat{P}(R) = \frac{i_{RK}}{i_{R+}} \cdot \frac{N_R}{N}. \quad (33)$$

This leads to the minimum-discrepancy estimators

$$\hat{\pi}_K = \hat{P}(K) = \sum_R \hat{P}(R, K) = \sum_R \frac{i_{RK}}{i_{R+}} \cdot \frac{N_R}{N} \quad (34)$$

$$\hat{\beta}_q = \hat{P}(\mathcal{R}_{\bar{q}} | \mathcal{K}_q) = \frac{\hat{P}(\mathcal{R}_{\bar{q}}, \mathcal{K}_q)}{\hat{P}(\mathcal{K}_q)} = \frac{\sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} \hat{P}(R, K)}{\sum_{K \in \mathcal{K}_q} \hat{P}(K)} = \frac{\sum_{R \in \mathcal{R}_{\bar{q}}} \sum_{K \in \mathcal{K}_q} \frac{i_{RK}}{i_{R+}} \cdot N_R}{\sum_R \sum_{K \in \mathcal{K}_q} \frac{i_{RK}}{i_{R+}} \cdot N_R} \quad (35)$$

$$\hat{\eta}_q = \hat{P}(\mathcal{R}_q | \mathcal{K}_{\bar{q}}) = \frac{\hat{P}(\mathcal{R}_q, \mathcal{K}_{\bar{q}})}{\hat{P}(\mathcal{K}_{\bar{q}})} = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} \hat{P}(R, K)}{\sum_{K \in \mathcal{K}_{\bar{q}}} \hat{P}(K)} = \frac{\sum_{R \in \mathcal{R}_q} \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{i_{RK}}{i_{R+}} \cdot N_R}{\sum_R \sum_{K \in \mathcal{K}_{\bar{q}}} \frac{i_{RK}}{i_{R+}} \cdot N_R}. \quad (36)$$

An estimator that combines minimum discrepancy and maximum likelihood is obtained by augmenting the E step in (28) such that the conditional expectation can only be non-zero under minimum discrepancy,

$$E(M_{RK} | N_R) = N_R \cdot \frac{i_{RK} \cdot P(K | R)}{\sum_K i_{RK} \cdot P(K | R)}. \quad (37)$$

Substituting the M_{RK} by this expectation in (13), (21), and (27) yields the minimum-discrepancy maximum likelihood estimators.

2. The simple learning model

2.1. Likelihood and log-likelihood

The SLM constrains the state distribution $P(K)$ by introducing item-specific solvability parameters g_q . Thus, the parameter vector $\theta = (\beta, \eta, g)$ consists of (vectors of) careless error, lucky guess, and solvability probabilities. With the unobservable frequency M_{RK} as defined above, the likelihood becomes

$$L(\theta) = \prod_R \prod_K (P(R | K) \cdot P(K))^{M_{RK}} \quad (38)$$

$$= \prod_R \prod_K P(R | K)^{M_{RK}} \cdot \left(\prod_{q \in K} g_q \prod_{q \in K^{\mathcal{O}}} (1 - g_q) \right)^{M_{RK}}, \quad (39)$$

where

$$K^{\mathcal{O}} = \{q \notin K \mid K \cup \{q\} \in \mathcal{K}\} \quad (40)$$

is the set of items that can be learned from K , called the outer fringe of K . According to the SLM, $P(K)$ is the product of the probabilities of mastering all items in K , g_q , and not mastering any items accessible from K , $1 - g_q$. The log-likelihood then becomes

$$\begin{aligned} \ell(\theta) &= \sum_R \sum_K M_{RK} \log P(R | K) \\ &+ \sum_R \sum_K M_{RK} \left(\sum_{q \in K} \log g_q + \sum_{q \in K^{\mathcal{O}}} \log(1 - g_q) \right). \end{aligned} \quad (41)$$

where $P(R | K)$ is defined as for the BLIM in (4).

2.2. Careless errors and lucky guesses

The estimators for careless error β and lucky guess η parameters in the SLM are the same as for the BLIM.

2.3. Solvability parameters

The partial derivative of the log-likelihood with respect to g_q is

$$\frac{\partial \ell(\theta)}{\partial g_q} = \sum_R \sum_K \left(\sum_{q \in K} \frac{M_{RK}}{g_q} - \sum_{q \in K^\circ} \frac{M_{RK}}{1 - g_q} \right) \quad (42)$$

$$= \sum_R \sum_{K \in \mathcal{K}_q} \frac{M_{RK}}{g_q} - \sum_R \sum_{K \in \mathcal{K}_q^\circ} \frac{M_{RK}}{1 - g_q} \quad (43)$$

$$= \sum_R \sum_K \frac{M_{RK}}{g_q} \cdot i - \frac{M_{RK}}{1 - g_q} \cdot j, \quad (44)$$

where

$$i = \begin{cases} 1 & \text{if } q \in K \\ 0 & \text{if } q \notin K \end{cases}, \quad j = \begin{cases} 1 & \text{if } q \in K^\circ \\ 0 & \text{if } q \notin K^\circ \end{cases}, \quad (45)$$

and \mathcal{K}_q° is the subset of states whose outer fringe contains q . Setting the derivative to zero yields

$$\sum_R \sum_K \frac{M_{RK} i (1 - \hat{g}_q) - M_{RK} j \hat{g}_q}{\hat{g}_q (1 - \hat{g}_q)} = 0 \quad (46)$$

$$\sum_R \sum_K M_{RK} i - M_{RK} i \hat{g}_q - M_{RK} j \hat{g}_q = 0 \quad (47)$$

$$\sum_R \sum_K M_{RK} i = \sum_R \sum_K \hat{g}_q (M_{RK} i + M_{RK} j) \quad (48)$$

$$\sum_R \sum_{K \in \mathcal{K}_q} M_{RK} = \hat{g}_q \left(\sum_R \sum_{K \in \mathcal{K}_q} M_{RK} + \sum_R \sum_{K \in \mathcal{K}_q^\circ} M_{RK} \right) \quad (49)$$

$$\hat{g}_q = \frac{\sum_R \sum_{K \in \mathcal{K}_q} M_{RK}}{\sum_R \sum_{K \in \mathcal{K}_q} M_{RK} + \sum_R \sum_{K \in \mathcal{K}_q^\circ} M_{RK}}. \quad (50)$$

2.4. Expectation maximization

The E step has the same form as for the BLIM. In each iteration, $P(R | K)$ and $P(K)$ are calculated from the current values of $\hat{\beta}_q$, $\hat{\eta}_q$, and \hat{g}_q .

2.5. Minimum-discrepancy maximum likelihood

The minimum-discrepancy estimators $\hat{\beta}_q$ and $\hat{\eta}_q$ are the same as for the BLIM. For the

solvability parameters,

$$\hat{g}_q = \hat{P}(\mathcal{K}_q | \mathcal{K}_q \cup \mathcal{K}_q^{\mathcal{O}}) = \frac{\hat{P}(\mathcal{K}_q, \mathcal{K}_q \cup \mathcal{K}_q^{\mathcal{O}})}{\hat{P}(\mathcal{K}_q \cup \mathcal{K}_q^{\mathcal{O}})} = \frac{\hat{P}(\mathcal{K}_q)}{\hat{P}(\mathcal{K}_q \cup \mathcal{K}_q^{\mathcal{O}})} \quad (51)$$

$$= \frac{\sum_R \sum_{K \in \mathcal{K}_q} \hat{P}(R, K)}{\sum_R \sum_{K \in \mathcal{K}_q} \hat{P}(R, K) + \sum_R \sum_{K \in \mathcal{K}_q^{\mathcal{O}}} \hat{P}(R, K)} \quad (52)$$

$$= \frac{\sum_R \sum_{K \in \mathcal{K}_q} \frac{i_{RK}}{i_{R+}} \cdot N_R}{\sum_R \sum_{K \in \mathcal{K}_q} \frac{i_{RK}}{i_{R+}} \cdot N_R + \sum_R \sum_{K \in \mathcal{K}_q^{\mathcal{O}}} \frac{i_{RK}}{i_{R+}} \cdot N_R}. \quad (53)$$

Minimum-discrepancy maximum likelihood estimators for β , η , and g are obtained from augmenting the E step as in (37).

Acknowledgments

F.W. would like to thank Alice Maurer for her thoughtful comments on the derivation of the estimators.

References

- Doignon JP, Falmagne JC (1999). *Knowledge Spaces*. Springer, Berlin.
- Heller J, Wickelmaier F (2013). “Minimum discrepancy estimation in probabilistic knowledge structures.” *Electronic Notes in Discrete Mathematics*, **42**, 49–56. doi:10.1016/j.endm.2013.05.145.

Affiliation:

Florian Wickelmaier
 Department of Psychology
 University of Tübingen
 Schleichstr. 4
 72076 Tübingen, Germany
 E-mail: wickelmaier@web.de
 URL: <https://www.mathpsy.uni-tuebingen.de/wickelmaier/>

Jürgen Heller
 Department of Psychology
 University of Tübingen
 Schleichstr. 4
 72076 Tübingen, Germany