IDENTIFIABILITY AND GOODNESS OF RECOVERY IN THE CONSTRAINED GAIN-LOSS MODEL

Pasquale Anselmi, Luca Stefanutti and Egidio Robusto
Department of Applied Psychology, University of Padua
pasquale.anselmi@unipd.it, luca.stefanutti@unipd.it, egidio.robusto@unipd.it

Abstract

The Gain-Loss Model (GaLoM) is a probabilistic skill multimap model for assessing learning processes proposed within the framework of knowledge space theory. Model parameters are initial probabilities of the skills, effects of learning object on gaining and losing the skills, careless error and lucky guess probabilities of problems. When the skills are assessed through a small number of problems or the data are noisy, model identifiability and goodness of recovery are a concern. An extension of the GaLoM is proposed in which the parameter space of careless error and lucky guess probabilities is constrained. Different ratios between number of problems and skills, and levels of noise in the data are considered in a simulation study. Advantages of the constrained GaLoM with respect to identifiability and goodness of recovery are presented and discussed.

The Gain-Loss Model (GaLoM, Robusto, Stefanutti, & Anselmi, in press) is a probabilistic model for assessing learning processes proposed within the framework of knowledge space theory (Doignon & Falmagne, 1985, 1999). It assesses the effect of learning objects on the attainment of skills required to solve problems in a given knowledge domain. Via the competency model, a skill multimap (Doignon, 1994; Doignon & Falmagne, 1999) associates with each problem a collection of subsets of skills that are necessary and sufficient to solve it.

Let S be a finite and non empty set of discrete skills, and C be any subset of S. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be two discrete random variables whose realizations are the competence states of a student at the pretest and posttest, respectively. Let Q be a finite and nonempty set containing n dichotomous problems, and $\mathbf{R}_1$ and $\mathbf{R}_2$ be two discrete random variables whose realizations are the response patterns $\mathbf{r} \in \{0, 1\}^n$ of a student at the pretest and posttest, respectively. Let m be the number of learning objects, and $o \in \{1, 2, \dots, m\}$ be the learning object the student is presented with. The conditional probability that $\mathbf{r}_1$ and $\mathbf{r}_2$ are the response patterns of a randomly sampled student at the pretest and posttest, given learning object o, is:

$$P(\mathbf{R}_1 = \mathbf{r}_1, \mathbf{R}_2 = \mathbf{r}_2 | o) = \sum_{C \subseteq S} \sum_{D \subseteq S} P(\mathbf{R}_1 = \mathbf{r}_1 | \mathbf{C}_1 = C) P(\mathbf{R}_2 = \mathbf{r}_2 | \mathbf{C}_2 = D) P(\mathbf{C}_2 = D | \mathbf{C}_1 = C, o) P(\mathbf{C}_1 = C), \quad (1)$$

where $P(\mathbf{C}_1 = C)$ is the probability of the competence state C at the pretest, $P(\mathbf{C}_2 = D | \mathbf{C}_1 = C, o)$ is the transition probability from state C at the pretest to state D at the posttest, $P(\mathbf{R}_1 = \mathbf{r}_1 | \mathbf{C}_1 = C)$ and $P(\mathbf{R}_2 = \mathbf{r}_2 | \mathbf{C}_2 = D)$ are the emission probabilities of response patterns $\mathbf{r}_1$ and $\mathbf{r}_2$ at the pretest and posttest, respectively.

The GaLoM is characterized by five types of parameters. The parameter π_s specifies the probability that skill s belongs to the competence state at the pretest. Gain γ_{os} is the probability that students presented with learning object o gain skill s going from the pretest to the posttest, and loss λ_{os} is the probability that the same students lose it. Careless error α_q is the probability that students fail problem q given that they possess some competence for it, and lucky guess β_q is the probability that the same students solve problem q given that they do not posses any competence for it. Emission probabilities of response patterns are governed by α and β parameters according to the Basic Local Independence Model (Doignon & Falmagne, 1999; Falmagne & Doignon, 1988) and the Deterministic Inputs Noisy AND gate model (de la Torre & Douglas, 2004; Junker & Sijtsma, 2001).

It is reasonable to expect that the data provide enough information about the skills and that they are not too noisy so that reliable statements about the skills can be made. On the contrary, when the skills are assessed through a small number of problems and/or the data are noisy, model identifiability and goodness of recovery might be problematic (Stefanutti, Anselmi, & Robusto, in press). By goodness of recovery we mean both the recovery of the true model parameters and the recognition of the knowledge structure that has generated the data. Both model identifiability and goodness of recovery might be improved by constraining the parameter space of α and β estimates. According to the procedure described in Stefanutti and Robusto (2009), a constrained version of the GaLoM (Constrained Gain-Loss Model, CoGaLoM) is proposed in which the log-likelihood of the model is maximized subject to the constraint that the α and β parameters are less or equal to an upper bound $\tau \in [0, 1]$. The GaLoM and the CoGaloM are tested in a simulation study with respect to model identifiability and goodness of recovery.

A Simulation Study

Three thousand random data sets were generated according to the GaLoM. Six conditions were produced by considering two ratios between the number of problems and underlying skills, and three levels of noise in the data. Two collections with 10 and 20 problems were generated, and 5 skills were set to underlie both. Via the conjunctive model, each problem from the two collections has been associated with the skills that were necessary and sufficient for its mastery. Each of the two resulting structures $\mathcal{K}_{\text{cor10}}$ and $\mathcal{K}_{\text{cor20}}$ contains 32 states, and was used to generate the data for the conditions with 10 and 20 problems, respectively. The noise in the data was set to be low (α_{true} and $\beta_{\text{true}} \leq .1$), medium ($\alpha_{\text{true}}$ and $\beta_{\text{true}} \leq .3$) or high ($\alpha_{\text{true}}$ and $\beta_{\text{true}} \leq .5$). For each of the six conditions, 500 random data sets were simulated. Each data set takes into account the effects of four learning objects on the skills. The first learning object was set to poorly affect both gain and loss of the skills (γ_{true} and $\lambda_{\text{true}} \leq .33$), the second to highly affect gain and poorly affect loss ($\gamma_{\text{true}} \geq .66$; $\lambda_{\text{true}} \leq .33$), the third to poorly affect gain and highly affect loss ($\gamma_{\text{true}} \leq .33$; $\lambda_{\text{true}} \geq .66$), the fourth to highly affect both gain and loss (γ_{true} and $\lambda_{\text{true}} \geq .66$). The number of response patterns was set to 1,000 for each data set (250 for each learning object).

For each of the 500 × 6 data sets, correct and incorrect models were estimated. The correct models incorporated the structures that were used to generate the data ($\mathcal{K}_{\text{cor10}}$ and $\mathcal{K}_{\text{cor20}}$). The incorrect models incorporated two structures different from the correct ones ($\mathcal{K}_{\text{incor10}}$ and $\mathcal{K}_{\text{incor20}}$). Both correct and incorrect models were estimated four times by the CoGaLoM, each time with a different choice of the upper bound of the α and β parameters ($\tau = 1, .5, .3, .1$). Note that, when $\tau = 1$, the CoGaloM corresponds to the GaLoM.

Model identifiability and goodness of recovery were tested in each of the 2 (10 and 20 problems) × 3 (noise ≤ .1, .3, .5) × 2 (correct and incorrect structures) × 4 ($\tau = 1, .5, .3, .1$) conditions. To test model identifiability, one of the 500 simulated data sets was randomly selected and the model parameters were estimated 100 times from different initial points of the parameter space. Standard deviations of parameter estimates greater or equal to .01 were taken as an indication that the model is not identifiable. Goodness of recovery was tested by considering the bias between the true parameters and the mean of the parameter estimates reproduced on the 500 simulated data sets. The proportion $P(\chi^2_{cor} < \chi^2_{incor})$ of data sets in which the Pearson's Chi-square of the model incorporating the correct structure was smaller than that of the model incorporating the incorrect structure was computed as well.

Results

Model Identifiability

When noise is ≤ .1, correct and incorrect models are identifiable in the conditions with 10 and 20 problems without having to constrain the estimates of α and β parameters ($\tau = 1$). When noise is ≤ .3, correct and incorrect models are still identifiable in the condition with 20 problems without having to constrain α and β estimates, but an upper bound of .1 is required so that they are identifiable in the condition with 10 problems. When noise is ≤ .5, correct and incorrect models are identifiable if an upper bound of .1 is specified (the correct model is identifiable in the condition with 20 problems also with an upper bound of .3). This result suggests that both the ratio between the number of problems and underlying skills, and the noise in the data contribute to the non-identifiability of the model. Constraining the parameter space of α and β estimates improves model identifiability.

Goodness of Recovery

This part concerns the possibility of recovering the knowledge structure that has generated the data. With respect to the conditions with 20 problems, Figure 1 depicts the proportion of simulated data sets in which the Chi-square of the correct model χ^2_{cor} was smaller than the Chi-square of the incorrect model χ^2_{incor} (*y* axis) for each level of noise (*x* axis) and each value of upper bound (the four lines in the diagram). A first thing to notice is that, regardless of the value of the upper bound, the higher the noise in the data, the lower the probability of identifying the correct model. Moreover, regardless of the level of noise in the data, the smaller the value of the upper bound, the better the separation between the correct model and incorrect one. Generally, the probability of identifying the correct model decreases in the conditions with 10 problems.

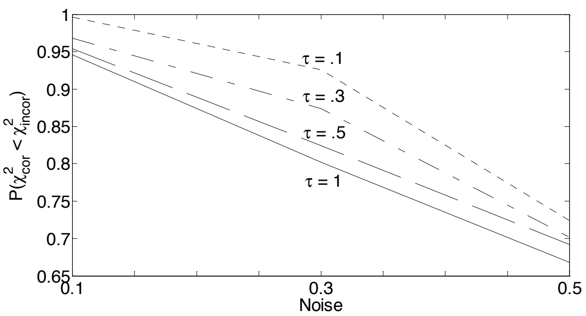

Figure 1. Proportion of simulated data sets in which the Chi-square of the correct model χ^2_{cor} turned out to be smaller than the Chi-square of the incorrect model χ^2_{incor} (*y* axis) for each level of noise (*x* axis) and each value of upper bound (the four lines in the diagram). The figure refers to the condition with 20 problems.

The rest of this section concerns how well the true model parameters are recovered. Only the conditions in which the model is identifiable are considered. In both correct model and incorrect model, the true model parameters (x axis) are plotted versus the mean of the parameter estimates (and the related standard errors) reproduced on the 500 simulated data sets (y axis). Figure 2 depicts the condition with 20 problems, noise ≤ .1 and upper bound = 1. In the correct model, the empirical biases of the estimates are negligible for most of the parameters, except for the estimates of some of the γ and λ parameters of the condition $\gamma_{\text{true}} \leq .33$ and $\lambda_{\text{true}} \geq .66$. In the incorrect model, empirical biases are high for some α and β parameters. In this condition the correct model is identified in the 95% of simulated data sets (see Figure 1). Figure 3 depicts the same condition when the upper bound is set to .1. In the correct model, recovery of model parameters does not differ from the situation in which the upper bound is 1. This is not the case with the incorrect model. Empirical biases of the estimates increase for some parameters, with the estimates of some α and β parameters lying on the upper bound. By setting the upper bound to .1, the correct model is identified in almost the 100% of data sets (see Figure 1). Figure 4 depicts the condition with 20 problems, noise ≤ .5 and upper bound = .1. Parameter estimates are biased for both models, and α and β parameters lie on the upper bound even when their true values are smaller than .1. In this condition of noise and upper bound, the correct model is identified in the 72% of data sets. Recovery of model parameters behaves in a similar way when 10 problems are considered, but empirical biases of the estimates are larger.

Discussion

The GaLoM and the CoGaloM were tested with respect to model identifiability and goodness of recovery. Different ratios between number of problems and underlying skills, and levels of noise in the data have been considered.

Model identifiability depends on the ratio between the number of problems and underlying skills, and on the level of noise in the data. When the data are lowly noisy (e.g., .1), the GaLoM is identifiable even if the skills are assessed by means of a small number of problems. When the noise increases (e.g., .3), a greater number of problems is required for the

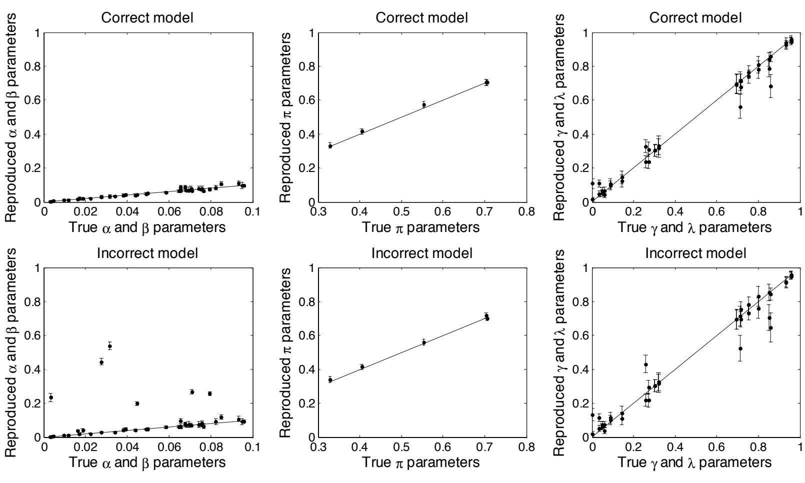

Figure 2. True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) in both correct model (upper diagrams) and incorrect model (lower diagrams). Condition with 20 problems, noise ≤ .1 and upper bound = 1. The straight line $x = y$ is added for reference.

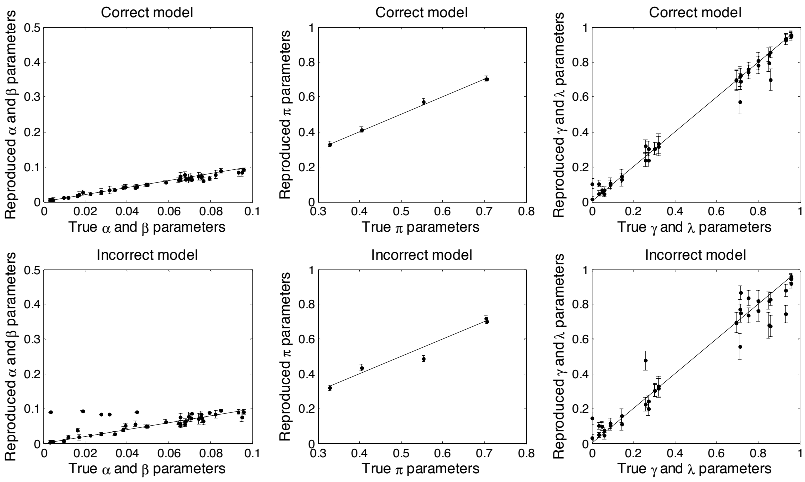

Figure 3. True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) in both correct model (upper diagrams) and incorrect model (lower diagrams). Condition with 20 problems, noise ≤ .1 and upper bound = .1. The straight line $x = y$ is added for reference.

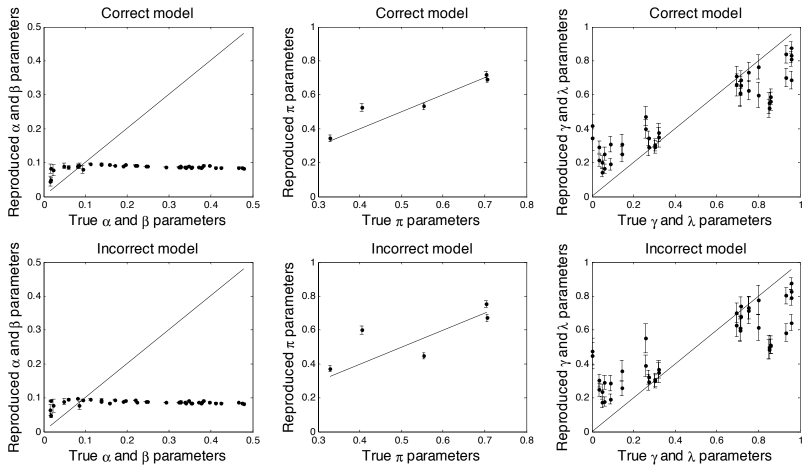

Figure 4. True parameters (x axis) versus mean of the parameter estimates (and related standard errors) reproduced on the simulated data sets (y axis) in both correct model (upper diagrams) and incorrect model (lower diagrams). Condition with 20 problems, noise ≤ .5 and upper bound = .1. The straight line $x = y$ is added for reference.

GaLoM to be identifiable. When the noise is high (e.g., .5), constraining the error parameters is needed in order to improve model identifiability. Simulations show that, when the skills are assessed by means of a small number of problems and the data are noisy, compensations among model parameters are not excluded. This is the case, for example, of the condition with 10 problems and noise .3 in which the correct model was not identifiable because of compensations among the parameters π, γ, and λ of a skill, and the β of a problem that was associated with it. Constraining the parameter space of α and β estimates is a way to improve model identifiability, even if it produces biased estimates for the parameters whose true value lies above the upper bound. A possible way of constraining the error parameter is the following. Model parameters can be estimated more times by the GaLoM in order to identify the ones that compensate each other. If error parameters of many problems are high and compensate with the skill parameters, the data set should be classified as too noisy and discarded. If error parameters of a few problems are high, the CoGaLoM can be used to specify an upper bound only for the noisy problems. This might resolve model identification problems.

Also goodness of recovery depends on both the ratio between number of problems and skills, and the level of noise. The higher the level of noise, the smaller the probability of identifying the correct model and the worse the recovery of its parameters. Moreover, recovering the correct model and its parameters gets worse when the skills are assessed by means of a small number of problems. It is interesting to note that, regardless of the level of noise in the data, the smaller the value of the upper bound, the better the separation between the correct model and the incorrect one. This happens in spite of the fact that an upper bound smaller than the level of noise in the data will produce biased estimates for the α and β parameters whose true value lies above the upper bound. This result is understood by considering that, unlike the correct model, the incorrect one tends to inflate the α and β estimates until the model likelihood reaches its maximum value. By imposing an upper bound to such estimates, the CoGaLoM constraints the incorrect model more than the correct model. As a consequence, the likelihood of the incorrect model decreases faster than that of the correct one.

Identifying the correct structure is particularly useful in practical applications in which there is not much theory about the knowledge structure on a given set of problems. In such cases, the first aim is to establish which of the models at hand is a better approximation of the "true" knowledge structure underlying the data. By specifying an upper bound to the error parameters, the CoGaLoM helps to identify the knowledge structure that best approximates the "true" one.

References

de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*, 333–353.

Doignon, J.-P. (1994). Knowledge spaces and skill assignments. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 111–121). New York, NY: Springer-Verlag.

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. (Eds.). (1999). *Knowledge spaces*. Berlin, Germany: Springer-Verlag.

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.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*, 258–272.

Robusto, E., Stefanutti, L., & Anselmi, P. (in press). The Gain-Loss Model: A probabilistic skill multimap model for assessing learning processes. *Journal of Educational Measurement*.

Stefanutti, L., Anselmi, P., & Robusto, E. (in press). Assessing Learning Processes with the Gain-Loss Model. *Behavior Research Methods*.

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