A regularized estimation approach for the three-parameter logistic model

Un approccio regolarizzato per la stima del modello logistico con tre parametri

Michela Battauz and Ruggero Bellio

Abstract The three-parameter logistic model is an item response theory model used with dichotomous items. It is well known that the parameters of the model are weekly identifiable and that the maximization of the likelihood, which is performed using numerical algorithms, is prone to convergence issues. In this paper, we propose the use of a penalized likelihood for the estimation of the item parameters. In particular, the penalty term shrinks the guessing parameters towards a known constant. Cross-validation is used to select such constant and the amount of shrinkage, though estimation via an empirical Bayes approach is also considered. The method is both simple and effective, and it is illustrated by means of a simulation study and a real data example.

Abstract *Il modello logistico con tre parametri è uno dei modelli item response theory usato con item dicotomici. In letteratura è noto che i parametri del modello sono debolmente identificabili e che la massimizzazione della verosimiglianza, che è condotta attraverso algoritmi numerici, è soggetta a problemi di convergenza. In questo articolo, si propone l'uso di una verosimiglianza penalizzata per la stima dei parametri del modello. In particolare, il termine di penalizzazione regolarizza le stime dei parametri di guessing verso una costante nota. L'ammontare di regolarizzazione viene determinato attraverso la validazione incrociata, e in alternativa mediante un approccio Bayesiano empirico. Il metodo è pratico ed efficace, e viene illustrato attraverso uno studio di simulazione e un esempio con dati reali.*

Michela Battauz
University of Udine - Department of Economics and Statistics, via Tomadini 30/A - Udine (Italy), e-mail: michela.battauz@uniud.it

Ruggero Bellio
University of Udine - Department of Economics and Statistics, via Tomadini 30/A - Udine (Italy) e-mail: ruggero.bellio@uniud.it

1 Introduction

The Three-Parameter Logistic (3PL) model is an Item Response Theory (IRT) model used with dichotomous responses. This model can be used for multiple-choice items, which are expected to have a non-zero probability of giving a correct response even at very low achievement levels. However, the parameters of this model are weekly identifiable [6] and the algorithms used for the maximization of the likelihood function frequently encounter convergence problems. A possible resolution is provided by regularization by means of penalized likelihood estimation [3, 10]. In particular, this paper studies the inclusion in the likelihood function of a penalty term that shrinks the guessing parameters towards a known constant. Despite a natural choice for this constant would be $1/k$, where k is the number of response options, we follow a data-driven approach for the selection of this value and of the amount of shrinkage. More specifically, the selection is performed by cross-validation [3]. An alternative route based on empirical Bayes estimation is also considered.

The paper is organized as follows. Section 2 introduces the model and the estimation methods, Section 3 shows an application to achievement data and Section 4 presents the results of a simulation study. Finally, Section 5 contains some concluding remarks.

2 Models and methods

In a 3PL model, the probability of a positive response to item j is given by

$$p_{ij} = \Pr(X_{ij} = 1|\theta_i; a_j, b_j, c_j) = c_j + (1 - c_j)\frac{\exp\{a_j(\theta_i - b_j)\}}{1 + \exp\{a_j(\theta_i - b_j)\}}, \qquad (1)$$

where θ_i is the ability of person i, a_j is the discrimination parameter, b_j is the difficulty parameter, and c_j is the guessing parameter. Here, $i = 1, \ldots, n$ and $j = 1, \ldots, J$, so that n is the sample size and J the number of items. A convenient parameterization of the model, suitable for estimation, is the following

$$p_{ij} = c_j + (1 - c_j)\frac{\exp(\beta_{1j} + \beta_{2j}\theta_i)}{1 + \exp(\beta_{1j} + \beta_{2j}\theta_i)}, \qquad (2)$$

with

$$c_j = F(\beta_{3j}) = \frac{\exp(\beta_{3j})}{1 + \exp(\beta_{3j})}. \qquad (3)$$

Let β be the vector containing all item parameters. The marginal maximum likelihood method [1], which is a commonly used estimation method of IRT models, requires the maximization of the following log likelihood function for β

$$\ell(\beta) = \sum_{i=1}^{n} \log \int_{\mathbb{R}} \prod_{j=1}^{J} p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} \phi(\theta_i) d\theta_i, \tag{4}$$

where $\phi(\cdot)$ denotes the standard normal p.d.f.

The penalized log likelihood function considered here is given by

$$\ell_p(\beta) = \ell(\beta) + J(\beta_3), \tag{5}$$

where β_3 is the vector containing all the guessing parameters, and $J(\beta_3)$ is a quadratic penalty term

$$J(\beta_3) = -\frac{1}{2\sigma^2} \sum_{j=1}^{J} (\beta_{3j} - \mu)^2, \tag{6}$$

proportional to the log p.d.f. of the normal distribution with mean μ and standard deviation σ. Note that to constrain the guessing parameters c_j towards a constant c, it is necessary to choose μ so that $c = F(\mu)$, with F defined in (3). Another option would take a penalty derived from a beta distribution for c_j.

2.1 Parameter tuning by cross-validation

A relevant issue is the choice of the tuning parameters μ and σ. The first proposal is to adopt a typical approach followed for regularized regression [10], namely

i. Estimate the item parameter for fixed values of the tuning parameters μ and σ;
ii. Select the tuning parameters by minimizing some error rate computed by cross-validation [3].

Step i. above is simply performed, for example by recourse to IRT software which allows for the introduction of a penalty term for 3PL models. Step ii. requires the definition of a suitable cross-validation error, here taken as minus the log likelihood of the validation set evaluated at the parameter estimates.

2.2 Empirical Bayes

An alternative method consists in treating μ and σ as parameter estimated by an empirical Bayes approach. In particular, we implement this method by treating β_3 as normal random effects, and then jointly estimate $(\beta_1, \beta_2, \mu, \sigma)$ after integrating them out from joint likelihood function of the data and β_3. Here the Laplace approximation is employed to carry out the latter integration, a strategy greatly simplified by the usage of automatic differentiation software [8].

3 A real data example

The proposed methodology was applied to achievement data collected on students attending the third year of vocational high school in Italy. In particular, we used the mathematics test that was administered during the final exam. The sample is composed of 3843 students. Only multiple-choice items were included in the analysis. These were 14 items, all with four response options. All analyses were performed using the R software [7].

We started the analysis from ordinary (unpenalized) maximum likelihood estimation. Figure 1 visualizes the estimated correlation matrix among item parameter estimators for the first four items. The estimates were obtained with the R package mirt [2], and the R package ellipse [5] was used to obtain the plot.

The correlation matrix is nearly singular, since the estimated guessing parameter of each item $\widehat{\beta}_{3j}$ is negatively correlated with the estimated easiness parameter $\widehat{\beta}_{1j}$, and it is positively correlated with the estimated discrimination parameter $\widehat{\beta}_{2j}$. At times, such correlations are very high (in either direction), so that it is not surprising that parameter estimation may become cumbersome. Some sort of regularization is surely helpful.

Fig. 1 Estimated correlation matrix among item parameter estimators for the first four items.
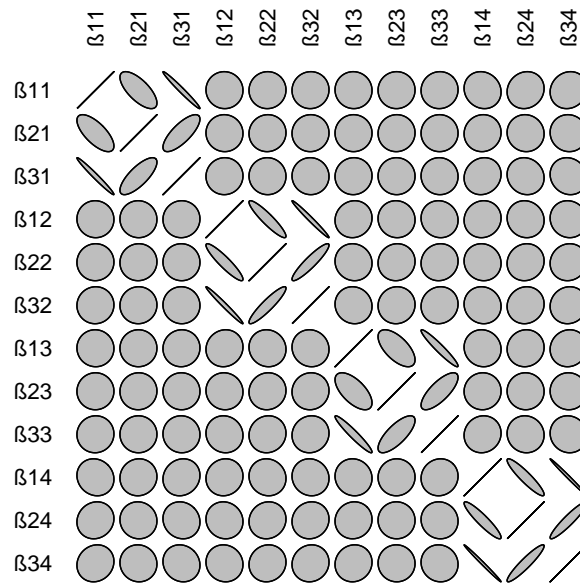
The function `mirt` to fit an IRT model in the `mirt` package has an option `parprior` to introduce the penalty for the guessing parameters, which turned out to be very handy to apply the methodology endorsed here. The tuning parameters were selected by choosing among a set of candidate values. In particular, $F(\mu)$ was selected in the set $\{0.1, 0.15, 0.2, 0.25\}$, whereas for σ a set formed by 100 values between 0.2 and 5 was considered. The selection was performed by means of 10-fold cross-validation.

Fig. 2 Left: cross-validation error as a function of $1/\sigma$, for given μ. Right: estimates of guessing parameters, as a function of $1/\sigma$, for given μ.
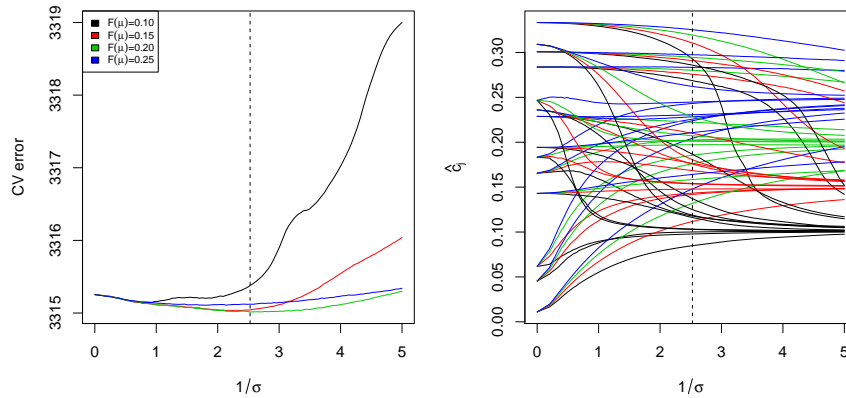

Figure 2 shows the results obtained for this data set. On the left panel of the figure, the cross-validation error is plotted against the reciprocal of σ. Thus, higher values on the *x*-axis correspond to larger amounts of shrinkage. The different colors refer to different values of $F(\mu)$. The vertical dashed lines indicate the point at which the cross-validation error is smallest. This corresponds to the values $\mu = F^{-1}(0.20)$ and $1/\sigma = 2.5$. The right panel of the figure shows the estimates of the guessing parameters at different levels of shrinkage. For all the values of μ, the smallest value of the cross-validation score is attained for σ values away from 0, pointing to the need of some shrinkage for the guessing parameters.

The estimation based on empirical Bayes has been carried out by means of the `Template Model Builder (TMB)` R package [9], which allows to define a `C++` template used to estimate the random effects model of interest. The use of the package resulted in a rather efficient implementation, the key point being the explicit integration of the ability parameters within the `C++` template by means of efficiently-coded Gaussian quadrature.

For the data set of interest, the empirical Bayes method provides estimated tuning parameters equal to $1/\widehat{\sigma} = 4.4$ and $\widehat{\mu} = F^{-1}(0.22)$, implying a higher amount of shrinkage for the guessing parameters with respect to cross validation. The estimates

of the remaining item parameters were instead quite similar for the two penalized methods.

The overall message of this example is that, even for a large sample of subjects, the estimation of the guessing parameter is challenging, and penalized maximum likelihood estimation improves the inferential results. The need of some regularization may become more striking for smaller sample sizes, where numerical problems may hamper the estimation routines of IRT software.

4 A simulation study

A small-scale simulation study has been performed to assess the performance of the various methods. In particular, the focus was on a relatively small scale setting, with $n = 500$ subjects and $J = 30$ items. Two different choices for the guessing parameters were considered. In the first setting, we took all the c_j parameters as constant and equal to 0.2, whereas for the second setting we took as guessing parameters the estimates obtained from a large scale educational assessment, with values of c_j ranging between 0.04 and 0.33, with an average value of 0.16. The real data set was employed also to set the values for the other item parameters, for either setting.

Three different methods were considered, given by ordinary maximum likelihood estimation (MLE) as implemented in the `mirt` package, and the two penalized estimation methods with tuning parameters estimated by cross validation (CV) and empirical Bayes (EB), respectively. Table 1 and 2 summarize the result of 100 simulated datasets. In particular, the tables report the average over the three groups of parameters of Root Mean Squared Error (RMSE), the squared root of the average of squared bias (B), and the average of Median Absolute Error (MAE).

Table 1 Summary of simulation results for Setting 1 (equal guessing parameters).

Method	Easiness β_1			Discrimination β_2			Guessing c		
	RMSE	B	MAE	RMSE	B	MAE	RMSE	B	MAE
MLE	0.76	0.19	0.44	0.59	0.18	0.25	0.15	0.02	0.14
CV	0.23	0.03	0.14	0.26	0.03	0.15	0.04	0.01	0.01
EB	0.20	0.02	0.13	0.23	0.02	0.15	0.03	0.01	0.02

The tables points to some interesting results. First, the ordinary unpenalized estimation performs rather poorly in both settings, with unacceptably large variability for all the parameters, thus confirming that this method is essentially useless for datasets of this size. For Setting 1, the two penalized methods perform an excellent adjustment, with negligible bias and greatly reduced variability for all the parameters. This is the setting more relevant to the adopted penalty, so that the good performances are not surprising. For Setting 2, as expected, the two penal-

Table 2 Summary of simulation results for Setting 2 (different guessing parameters).

Method	Easiness β_1			Discrimination β_2			Guessing c		
	RMSE	B	MAE	RMSE	B	MAE	RMSE	B	MAE
MLE	0.70	0.21	0.37	0.53	0.18	0.23	0.14	0.04	0.10
CV	0.32	0.24	0.23	0.29	0.14	0.17	0.07	0.08	0.06
EB	0.30	0.28	0.25	0.24	0.16	0.17	0.08	0.07	0.07

ized methods perform less well, with bias similar to that of the MLE. At any rate, the two penalized estimators have overall better performances, and they represent a clear improvement over the MLE, with shrinkage offering a good level of protection against the large fluctuations affecting ordinary MLE. Finally, the differences between the two penalized methods are generally minor, though the method based on cross-validation seems to be slightly preferable in the most challenging setting.

5 Conclusion and ongoing research

This paper presents a procedure for the estimation of the 3PL model based on a penalized likelihood approach. The application shows that by penalizing the likelihood the error rate of prediction, assessed through cross-validation, is reduced. Even if IRT models are not usually fitted to predict new observations, the procedure can be viewed as a regularization method to obtain a model closer to the data generating process. The simulation study suggests that the penalized estimation represents a notable improvement over ordinary maximum likelihood when the guessing parameters are constant, while the improvement is less substantial when the guessing parameters exhibit large variation. A further finding is that the results obtained via cross validation are generally similar to those obtained by an empirical Bayes approach.

In the Bayesian literature [6], the prior for the guessing parameters is usually a distribution with mean equal to the reciprocal of the number of response options for the items. Despite this seems a sensible choice, the application of the proposed approach typically leads to the selection of a different value of the mean, a fact that seems worth mentioning.

The method introduced here is very practical, since it only requires the introduction of a simple penalty term in the ordinary log likelihood function for MML estimation of a 3PL model. More sophisticated approaches could be considered, such as model-based shrinkage aiming to reduce the bias of guessing parameter estimators [4]. Some investigation in this direction appears worth considering. The introduction of flexible assumptions for the ability parameters, which may be recommendable in some instances [11], appears instead more challenging.

It should be noted that this is a preliminary study. The use of regularization techniques for the estimation of the 3PL model is still under investigation by the authors. Future research will involve more extensive simulation studies to achieve a better understanding of the performance of the procedure, and the consideration of further regularization methods, targeting better inferential properties.

Acknowledgements

References

1. Bock, R. D., Aitkin, M.: Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika **46**, 443–459 (1981)
2. Chalmers, R.: mirt: A Multidimensional Item Response Theory Package for the R Environment. J. Stat. Softw. **48**, 1–29 (2012)
3. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning (2nd Edition). Springer, New York (2009)
4. Kosmidis, I., and Firth, D. (2010). A generic algorithm for reducing bias in parametric estimation. Electron. J. Statist. **4**, 1097–1112 (2010)
5. Murdoch, D. and Chow, E. D. ellipse: Functions for drawing ellipses and ellipse-like confidence regions. R package version 0.3-8 (2013)
6. Patz, R. J., Junker, B. W.: Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. J. Educ. Behav. Stat. **24**, 342–366 (1999)
7. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2018)
8. Skaug, H. J. (2002). Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. J. Comput. Graph. Statist. **11**, 458–470 (2002)
9. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J., Bell, B. TMB: Automatic differentiation and Laplace approximation. J. Stat. Softw. **70**, 1–21 (2016)
10. Tutz, G: Regression for Categorical Data. Cambridge University Press, Cambridge (2011)
11. Woods, C.: Estimating the latent density in unidimensional IRT to permit non-normality. In: Reise, S. P., and Revicki, D. A. (Eds.) Measuring Psychological Constructs: Advances in Model-Based Approaches, pp. 60–84. Routledge, New York (2015)