IRT Model with Personal Guessing Levels

Lyubomir Christov, Dimiter Tsvetkov

Abstract: *In this paper we offer a modification of the 3-parametric logistic model by introducing personal guessing level parameters which appears to be interesting and useful in some specific test situations. We offer a* ***computer implementation*** *of the model on the basis of the well-known* ***MMLE/EM*** *parameter estimation technique. Also here we use the* ***BFGS*** *quasi-Newton method instead of the traditional Newton-Raphson method in solving the maximum likelihood equations.*
Key words: IRT *models,* ***MMLE/EM****.*

INTRODUCTION

Here we consider a test with n multiple choice items which is performed by N persons. Each answer to i-th item, $i=1,\dots,n$, represents a random variable U_i accepting value 1 with probability $P_i(\theta,\vartheta)$ and value 0 with probability $Q_i(\theta,\vartheta)=1-P_i(\theta,\vartheta)$ and the test answer joint probability distribution is

$$\Pr(\mathbf{U}\mid\theta,\vartheta)=\prod_{i=1}^{n}P_i(\theta,\vartheta)^{U_i}Q_i(\theta,\vartheta)^{1-U_i} \ . \tag{1}$$

Here it is pointed out that the probability of $\mathbf{U}=(U_1,\dots,U_n)$ is conditional both to personal ability value θ and to personal guessing level ϑ $(0\le\vartheta\le1)$. Following the well known 3-parametric model (see [1,2,3,4]), we assume that

$$P_i(\theta,\vartheta)=\vartheta+(1-\vartheta)\frac{\exp[a_i(b_i-\theta)]}{1+\exp[a_i(b_i-\theta)]}, \tag{2}$$

where a_i, $i=1,\dots,n$, is the item discrimination parameter and b_i, $i=1,\dots,n$, is the item difficulty parameter. In the traditional 3-parametric model the guessing level is also an item parameter. The main aim of this paper is to present a parameter estimation technique for the model above. We will use the **MMLE/EM** approach [1,3] in which θ is considered as a random variable with known distribution density $\varphi(\theta)$, usually normal standard. Here ϑ will be also a random variable with some known beta distribution density $\beta(\vartheta)$. Under this settings, variables $\mathbf{U}$, θ and ϑ are dependent and their joint distribution is given by the formula

$$\Pr(\mathbf{U}\mid\theta,\vartheta)\varphi(\theta)\beta(\vartheta) \tag{3}$$

and for the marginal distribution of $\mathbf{U}$ we have

$$\Pr(\mathbf{U})=\iint\Pr(\mathbf{U}\mid\theta,\vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta \ . \tag{4}$$

Here we propose an essential simplification assuming that θ and ϑ are independent but in fact we use this assumption only for the sake of simplification of the numerical integration.

One can get item parameter estimation from the maximum of the logarithmic likelihood function

$$\ln\Pr(\mathbf{u})=\ln\iint\Pr(\mathbf{u}\mid\theta,\vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta \ . \tag{5}$$

Including all persons in analysis we get the following likelihood function

$$LH=\sum_{j=1}^{N}\ln\iint\Pr(\mathbf{u}_j\mid\theta,\vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta \ . \tag{6}$$

where $\mathbf{u}_j=(u_{1j},\dots,u_{nj})$ is the answer pattern of j-th person. We assume identical θ and identical ϑ distributions for all persons.

MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

We have to estimate item parameters via marginal maximum likelihood principal that means to maximize function LH with respect to them. In fact we solve the system of maximum likelihood equations $\nabla LH = 0$, where ∇LH is the gradient of LH.

Hereafter ∂_ξ will denote differentiation with respect to some item parameter ξ. Maximum likelihood equations (**MLE**) have the form $\partial_\xi LH = 0$, $\xi = a_i$ or $\xi = b_i$, $i = 1,\dots,n$, which, using a similar way as it is shown in [5], can be written as

$$\sum_{j=1}^{N}\iint\left[\frac{\partial_\xi Q_i(\theta,\vartheta)}{Q_i(\theta,\vartheta)} + u_{ij}\frac{\partial_\xi\left(P_i(\theta,\vartheta)\big/Q_i(\theta,\vartheta)\right)}{P_i(\theta,\vartheta)\big/Q_i(\theta,\vartheta)}\right]\Pr(\theta,\vartheta\,|\,\mathbf{u}_j)\,d\theta d\vartheta,\ i = 1,\dots,n, \tag{7}$$

where

$$\Pr(\theta,\vartheta\,|\,\mathbf{U}) = \frac{\Pr(\mathbf{U}\,|\,\theta,\vartheta)\varphi(\theta)\beta(\vartheta)}{\iint \Pr(\mathbf{U}\,|\,\theta,\vartheta)\varphi(\theta)\beta(\vartheta)\,d\theta d\vartheta} \tag{8}$$

is the Bayesian posterior probability of (θ,ϑ). Explicitly

$$\Pr(\theta,\vartheta\,|\,\mathbf{u}_j) = \frac{\prod_{i=1}^{n} P_i(\theta,\vartheta)^{u_{ij}} Q_i(\theta,\vartheta)^{1-u_{ij}}\varphi(\theta)\beta(\vartheta)}{\iint\prod_{i=1}^{n} P_i(\theta,\vartheta)^{u_{ij}} Q_i(\theta,\vartheta)^{1-u_{ij}}\varphi(\theta)\beta(\vartheta)\,d\theta d\vartheta}. \tag{9}$$

Therefore $2n$ equations (7) have to be solved. These equations are independent item by item.

NUMERICAL INTEGRATION

We use Gaussian numerical approximation to the integrals above with nodes x_k and weights w_k (21 nodes are used), for the integration with respect to θ and nodes xc_l and weights wc_l for the integration with respect to ϑ (11 nodes are used). These nodes and weights are easy to be calculated by various tools. Then, using the similar way as in [5], we get the following numerical form of the **MLE**

$$\sum_{k,l}\left[\frac{\partial_\xi Q_i(\theta,\vartheta)}{Q_i(\theta,\vartheta)}\mathbf{f}_{ikl} + \frac{\partial_\xi\left(P_i(\theta,\vartheta)\big/Q_i(\theta,\vartheta)\right)}{P_i(\theta,\vartheta)\big/Q_i(\theta,\vartheta)}\mathbf{r}_{ikl}\right] = 0, \tag{10}$$

where

$$\mathbf{r}_{ikl} = \sum_{j=1}^{N}\frac{u_{ij}\prod_{i=1}^{n}\left[P_i(x_k,xc_l)\right]^{u_{ij}}\left[Q_i(x_k,xc_l)\right]^{1-u_{ij}} w_k wc_l}{\sum_{k,l}\prod_{i=1}^{n}\left[P_i(x_k,xc_i)\right]^{u_{ij}}\left[Q_i(x_k,xc_i)\right]^{1-u_{ij}} w_k wc_l}, \tag{11}$$

$$\mathbf{f}_{ikl} = \sum_{j=1}^{N}\frac{\prod_{i=1}^{n}\left[P_i(x_k,xc_l)\right]^{u_{ij}}\left[Q_i(x_k,xc_l)\right]^{1-u_{ij}} w_k wc_l}{\sum_{k,l}\prod_{i=1}^{n}\left[P_i(x_k,xc_i)\right]^{u_{ij}}\left[Q_i(x_k,xc_i)\right]^{1-u_{ij}} w_k wc_l}. \tag{12}$$

Following Baker [1], for the individual guessing level we use a beta distribution with a density function

$$\beta(\vartheta) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\vartheta^{\alpha}(1-\vartheta)^{\beta} \tag{13}$$

with $\alpha = 20p - 1$ and $\beta = 20(1-p) - 1$ where $p = 1/\mathbf{nalt}$ and $\mathbf{nalt}$ is the number of response alternatives, for example if each item has 5 alternatives then $\mathbf{nalt} = 5$ and $p = 0.2$. We have to use different set of nodes and weights in the different cases of $\mathbf{nalt}$. Remember that usually $\mathbf{nalt} = 5$ or $\mathbf{nalt} = 4$.

EM SOLUTION

EM consists of two basic steps: **E** – ***expectation step*** and **M** – ***maximization step***. Choose some natural initial values for the parameters estimated after that successive steps are fulfilled until some proper stop criterion is met.

E – *step*. Calculate $\mathbf{f}_{ik}$ и $\mathbf{r}_{ik}$ using formulas (11-12) and current parameter values.

M – *step*. Solve equations (10).

These steps are repeated until stop criterion is met. For example, if the values of the left hand sides of equations (10) are very close to zero or when difference between two consequent parameter estimates becomes small enough.

We solve equations (10) by means of quasi-Newton method **BFGS**. It looks that **BFGS** works well and is not very sensitive to the precision of solving the line search problem. This lack of sensitivity appears to be the main advantage of the method.

ABILITY AND GUESSING LEVEL ESTIMATION – EAP METHOD

Having once estimated item parameters it remains to estimate the personal parameters which in our model are the personal ability $\hat{\theta}_j$ and the personal guessing level $\hat{\vartheta}_j$ for the j-th person, $j = 1, \ldots, N$. For this purpose we use **EAP** (***e**xpected **a** **p**osteriori*) approach which is more appropriate for the **MMLE** scheme. **EAP** estimates are the expectations of the θ and ϑ distribution of a given person, estimated posterior using Bayesian formula. By definition we have

$$\hat{\theta}_j = E_\theta\left[\Pr\left(\theta, \vartheta \mid \mathbf{u}_j\right)\right] = \frac{\iint \theta \Pr(\mathbf{U} \mid \theta, \vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta}{\iint \Pr(\mathbf{U} \mid \theta, \vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta}, \quad j = 1, \ldots, N, \qquad (14)$$

$$\hat{\vartheta}_j = E_\vartheta\left[\Pr\left(\theta, \vartheta \mid \mathbf{u}_j\right)\right] = \frac{\iint \vartheta \Pr(\mathbf{U} \mid \theta, \vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta}{\iint \Pr(\mathbf{U} \mid \theta, \vartheta)\varphi(\theta)\beta(\vartheta)d\theta d\vartheta}, \quad j = 1, \ldots, N. \qquad (15)$$

Numerical integration implies the following estimation rules

$$\hat{\theta}_j = \frac{\sum_{k,l} x_k \prod_{i=1}^{n} \left[P_i\left(x_k, xc_i\right)\right]^{u_{ij}} \left[Q_i\left(x_k, xc_i\right)\right]^{1-u_{ij}} w_k wc_l}{\sum_{k,l} \prod_{i=1}^{n} \left[P_i\left(x_k, xc_i\right)\right]^{u_{ij}} \left[Q_i\left(x_k, xc_i\right)\right]^{1-u_{ij}} w_k wc_l}, \quad j = 1, \ldots, N, \qquad (16)$$

$$\hat{\vartheta}_j = \frac{\sum_{k,l} xc_l \prod_{i=1}^{n} \left[P_i\left(x_k, xc_i\right)\right]^{u_{ij}} \left[Q_i\left(x_k, xc_i\right)\right]^{1-u_{ij}} w_k wc_l}{\sum_{k,l} \prod_{i=1}^{n} \left[P_i\left(x_k, xc_i\right)\right]^{u_{ij}} \left[Q_i\left(x_k, xc_i\right)\right]^{1-u_{ij}} w_k wc_l}, \quad j = 1, \ldots, N. \qquad (17)$$

Formulas (16) and (17) are not iterative and give results even in the case when the person considered has extreme score. Some other modal algorithms are also potential to give good results.

DEMONSTRATION OF SOFTWARE

The authors propose a program in which **MMLE/EM** approach was implemented for the model above as **MS Windows** application. Consider an example with $n = 11$ items and $N = 122$ persons. Table 1 contains the results of the parameter estimates, $\mathbf{nalt} = 5$.

item	Estimate of a via traditional mode	Estimate of b via traditional mode	Estimate of a via personal guessing level mode	Estimate of b via personal guessing level mode
1	0.526	-2.122	0.575	-0.923
2	0.743	-1.149	0.915	-0.104
3	0.770	-2.891	0.857	-1.876
4	0.886	-1.146	1.011	-0.204
5	1.407	-0.122	1.631	0.776
6	0.579	-1.370	0.668	-0.197
7	1.816	-0.135	1.972	0.734
8	1.431	-1.145	1.610	-0.366
9	1.981	-0.972	1.987	-0.255
10	0.654	0.904	0.768	2.318
11	1.817	-0.508	1.848	0.292

***Table 1**: Comparison between different estimates*

Remember that in the traditional model the guessing level is an item parameter which is estimated in the **MMLE/EM**. It is clear that the estimates of the item discrimination parameters are very close. Differences arise in the estimates of the difficulty parameters which appears to be natural from the point of view of the model structure. In both cases we use our own software for calculations.

CONCLUSIONS

This work shows in particular that **MMLE/EM** approach is remarkably stable with respect to its various modifications. We especially point out that **BFGS** method is also appropriate for the equations solving during the **M**-step.

In our next works we are going to eliminate the independent assumption of ability and guessing level variables. It looks that this assumption is eligible in the coefficient estimation practice but obviously is not very consistent with the theory.

We hope that the model above is appropriate in the determining the social desirability in the psychological testing because the person social desirability can be interpreted as a person guessing level according to the interpretation of the traditional 3-parametric logistic model.

REFERENCES

[1] Baker F., Item Response Theory: Parameter Estimation Techniques, 1992.

[2] Hambleton R., and H. Swaminathan, Item Response Theory: Principles and Applications, 1984.

[3] Linden W., R. Hambleton, Handbook of Modern Item Response Theory, Springer, 1996.

[4] Lord F. M., Applications of Item Response Theory to Practical Testing Problems, 1980.

[5] Tsvetkov D., and L. Hristov, Computer Implementation of MMLE/EM Algorithm for Two-Parametric Logistic Model, CompSysTech'2003.

ABOUT THE AUTHORS

Assoc. Prof. Dimiter Petkov Tsvetkov, PhD, department of "Mathematics and Natural Sciences", NVU "Vassil Levski", Veliko Tarnovo, phone: +359 62 600 210, E-mail: dimiter99@yahoo.com.

Assist. Prof. Lyubomir Ianakiev Hristov, department of "Mathematical analysis and application", University of Veliko Tarnovo "St. st. Cyril and Methodius", Veliko Tarnovo phone: +359 62 4 74 68, E-mail: Lyubo60@yahoo.com.