Two-Parameter Logistic IRT Model with Latent Classes

Lyubomir Hristov

Abstract: *In this paper we consider the well known two-parameter logistic model from the Item Response Theory (**IRT**) and suppose the presence of several latent classes among the test population which classes differ each other in the item parameter structure. We give a computer implementation of this model and also empirically examine the model by means of a software test simulator. Parameters are estimated via the **MMLE/EM** method.*

Key words: *Item Response Theory, Latent Class Analysis.*

INTRODUCTION

Here it is assumed that the reader is familiar with the basis of the Item Response Theory. We consider a test with n binary items that is performed by a sample of N persons and therefore we have a certain sample observations. In the two-parameter logistic model the answer to test item i, $i = 1,2,\dots,n$, represents a random variable U_i accepting value 1 with probability

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

and value 0 with probability $Q(\theta, a_i, b_i) = 1 - P(\theta, a_i, b_i)$, where a_i, $i = 1,2,\dots,n$, is the item discrimination parameter and b_i, $i = 1,2,\dots,n$, is connected with the item difficulty. The value θ represents the person ability level. Assuming the local independence, the test answer joint probability distribution is given by

$$f(\mathbf{u} \mid \theta, \xi) = \prod_{i=1}^{n} P(\theta, a_i, b_i)^{u_i} Q(\theta, a_i, b_i)^{1-u_i} . \tag{2}$$

Here it is pointed out that the probability distribution of the random vector $\mathbf{U} = (U_1, \dots, U_n)$ with values $\mathbf{u} = (u_1, u_2, \dots, u_n)$ is conditional with respect to the personal ability level θ and depends on the set ξ of all item parameters. In the **MMLE** approach it is also assumed that persons who form the sample are drawn from some well defined population in which the abilities θ are realizations of a random variable Θ with known normal distribution of a density $\varphi(\theta \mid \mu, \sigma)$. Then the joint distribution density of $(\mathbf{U}, \Theta)$ is $f(\mathbf{u} \mid \theta, \xi)\varphi(\theta \mid \mu, \sigma)$ and the marginal distribution of $\mathbf{U}$ is given by

$$f(\mathbf{u} \mid \xi) = \int_{-\infty}^{\infty} f(\mathbf{u} \mid \theta, \xi)\varphi(\theta \mid \mu, \sigma) d\theta , \tag{3}$$

which over the sample implies the following log-likelihood function

$$l(\xi) = \sum_{j=1}^{N} \ln f(\mathbf{u}_j \mid \xi) = \sum_{j=1}^{N} \ln \int_{-\infty}^{\infty} f(\mathbf{u}_j \mid \theta, \xi)\varphi(\theta \mid \mu, \sigma) d\theta . \tag{4}$$

Here $\mathbf{u}_j = (u_{j1}, u_{j2}, \dots, u_{jn})$ is the answer pattern of person j, $j = 1,2,\dots,N$. The maximum likelihood estimate for the parameters ξ can be found in the framework of the often used **MMLE/EM** scheme.

Without loss of generality we can put $\mu = 0$ and $\sigma = 1$. In this way we avoid the evident indeterminacy of the parameter set and make further conclusions easier.

In this paper we suppose the presence of K latent classes in the sample. Each of these classes is characterized by its own item parameter set ξ_k and normal class ability distribution with density $\varphi(\theta \mid \mu_k, \sigma_k)$, $k = 1,2,\dots,K$. Let also α_k be the prior provability of a random person to belong to the corresponding class. Introduce a random variable $\boldsymbol{\kappa}$ with values $k=1,2,\ldots,K$ that points out the membership to class k. Then the conditional with respect to $\boldsymbol{\kappa}=k$ density of $(\mathbf{U},\Theta)$ is $f(\mathbf{u}\,|\,\theta,\xi_k)\varphi(\theta\,|\,\mu_k,\sigma_k)$ and the joint distribution of $(\mathbf{U},\Theta,\boldsymbol{\kappa})$ is described by $f(\mathbf{u}\,|\,\theta,\xi_k)\varphi(\theta\,|\,\mu_k,\sigma_k)\alpha_k$, therefore the marginal distribution of $\mathbf{U}$ is given by

$$f(\mathbf{u}\,|\,\xi)=\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty} f(\mathbf{u}\,|\,\theta,\xi_k)\varphi(\theta\,|\,\mu_k,\sigma_k)d\theta. \tag{5}$$

Here ξ stands for the set of all parameters α_k, ξ_k, $k=1,2,\ldots,K$. Technically, again to avoid the indeterminacy we put $\mu_k=0$ and $\sigma_k=1$ for all classes. Explicitly we have

$$f(\mathbf{u}\,|\,\xi)=\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty}\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_i}Q(\theta,a_{ik},b_{ik})^{1-u_i}\varphi(\theta)d\theta, \tag{6}$$

where

$$\varphi(\theta)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2}\right) \tag{7}$$

is the density function for the normal standard distribution. Over the sample, (6) implies the following log-likelihood function

$$l(\xi)=\sum_{j=1}^{N}\ln\left(\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty}\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta\right). \tag{8}$$

Here $l(\xi)$ is the marginal log-likelihood function for the parameter set ξ. Finding its maximum with respect to ξ forms the method of the **M**arginal **M**aximum **L**ikelihood **E**stimation (**MMLE**). Our main purpose is to give a parameter estimation procedure for (8) via the **EM** algorithm. One should keep in mind that the **EM** algorithm here appears as an auxiliary tool in the **MMLE** framework nevertheless that it can be used independently without the **MMLE** scheme and such an use leads to the same results.

MMLE/EM PARAMETER ESTIMATION

At first let us find the maximum likelihood equations (**MLE**) for the item parameters a_{ik} and b_{ik}, $i=1,2,\ldots,n$, $k=1,2,\ldots,K$. Let ∂_ξ denotes differentiation with respect to some item parameter ξ. **MLE** for the item parameters have the form $\partial_\xi l(\xi)=0$ where $\xi=a_{ik}$ or $\xi=b_{ik}$, $i=1,\ldots,n$, $k=1,2,\ldots,K$, which, using a similar way as it is shown in [5], can be written as

$$\sum_{j=1}^{N}\int_{-\infty}^{\infty}\left[\frac{\partial_\xi Q(\theta,a_{ik},b_{ik})}{Q(\theta,a_{ik},b_{ik})}+u_{ij}\frac{\partial_\xi\left(P(\theta,a_{ik},b_{ik})\big/Q(\theta,a_{ik},b_{ik})\right)}{P(\theta,a_{ik},b_{ik})\big/Q(\theta,a_{ik},b_{ik})}\right]f(\theta,\boldsymbol{\kappa}=k\,|\,\mathbf{u}_j)d\theta=0, \tag{9}$$

where

$$f(\theta,\boldsymbol{\kappa}=k\,|\,\mathbf{u}_j)=\frac{\alpha_k\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)}{\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty}\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta} \tag{10}$$

is the conditional posterior density of $(\Theta,\boldsymbol{\kappa})$.

The essence of the **EM** algorithm in the case considered consists of the following. Start with some reasonable initial parameter estimate $\xi^{(0)}$. Suppose we have a current estimate $\xi^{(t)}$ at a step $t=0,1,2,\ldots$. Then by (10) we can find the posterior densities

$$f\left(\theta,\boldsymbol{\kappa}=k\mid\mathbf{u}_j,\xi^{(t)}\right)=\frac{\alpha_k^{(t)}\prod_{i=1}^{n}P\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}\varphi(\theta)}{\sum_{k=1}^{K}\alpha_k^{(t)}\int_{-\infty}^{\infty}\prod_{i=1}^{n}P\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}\varphi(\theta)d\theta}\tag{11}$$

in a fully determined form that is the content of the **E**-step (expectation). After that we can solve equations (9) by means of the Newton method and in this way we can find the next approximation for the discrimination and difficulty parameters. At this point for the parameters α_k the Lagrangue multipliers method gives

$$\alpha_k^{(t+1)}=\frac{1}{N}\sum_{j=1}^{N}\frac{\alpha_k^{(t)}\int_{-\infty}^{\infty}\prod_{i=1}^{n}P\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}\varphi(\theta)d\theta}{\sum_{k=1}^{K}\alpha_k^{(t)}\int_{-\infty}^{\infty}\prod_{i=1}^{n}P\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(\theta,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}\varphi(\theta)d\theta},\ k=1,2,\ldots,K\ ,\tag{12}$$

Finding $\xi^{(t+1)}$ forms the **M**-step (maximization). These steps are repeated until convergence criterion is met. This procedure will be described numerically in the next section.

NUMERICAL SOLUTION

We use Gaussian numerical approximation to the integrals above with nodes x_s and weights w_s

$$\int_{-\infty}^{\infty}g(\theta)\varphi(\theta)d\theta\approx\sum_s w_s g(x_s),\tag{13}$$

where 21 nodes are used in our implementation. These nodes and weights in fact form a finite random variable which appears as a good finite approximation to the normal standard distribution. In this way the number of nodes do not affect essentially the general idea for the prior ability distribution, but obviously the large number of nodes is a better choice.

Now using the similar way as in [5], we get the following numerical form of the item parameter **MLE** at the global successive step t

$$\sum_s\left[\frac{\partial_\xi Q(x_s,a_{ik},b_{ik})}{Q(x_s,a_{ik},b_{ik})}\mathbf{f}_{sk}^{(t)}+\frac{\partial_\xi\left(P(x_s,a_{ik},b_{ik})\big/Q(x_s,a_{ik},b_{ik})\right)}{P(x_s,a_{ik})\big/Q(x_s,a_{ik},b_{ik})}\mathbf{r}_{isk}^{(t)}\right]=0\,,\tag{14}$$

where $\xi=a_{ik}$ or $\xi=b_{ik}$, $i=1,2,\ldots,n$, $k=1,2,\ldots,K$, and

$$\mathbf{r}_{isk}^{(t)}=\sum_{j=1}^{N}\frac{u_{ji}\alpha_k^{(t)}\prod_{i=1}^{n}\left[P\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{u_{ji}}\left[Q\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{1-u_{ji}}w_s}{\sum_{k=1}^{K}\alpha_k^{(t)}\sum_s\prod_{i=1}^{n}\left[P\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{u_{ji}}\left[Q\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{1-u_{ji}}w_s}\,,\tag{15}$$

$$\mathbf{f}_{sk}^{(t)}=\sum_{j=1}^{N}\frac{\alpha_k^{(t)}\prod_{i=1}^{n}\left[P\left(x_s\mid a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{u_{ji}}\left[Q\left(x_s\mid a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{1-u_{ji}}w_s}{\sum_{k=1}^{K}\alpha_k^{(t)}\sum_s\prod_{i=1}^{n}\left[P\left(x_s\mid a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{u_{ji}}\left[Q\left(x_s\mid a_{ik}^{(t)},b_{ik}^{(t)}\right)\right]^{1-u_{ji}}w_s}\,.\tag{16}$$

By (14) for we receive the following equations for the item parameters

$$(a_{ik}):\sum_s\left[-x_sQ(x_s,a_{ik},b_{ik})\mathbf{f}_{sk}^{(t)}+x_s\mathbf{r}_{isk}^{(t)}\right]=0\,,\tag{17}$$

$$(b_{ik}):\sum_s \left[Q(x_s,a_{ik},b_{ik})\mathbf{f}_{sk}^{(t)}-\mathbf{r}_{isk}^{(t)}\right]=0\,. \quad (18)$$

Practically **EM** algorithm consists of the following. At first choose some proper initial values $\xi^{(0)}$ for the parameters to be estimated. Suppose we have a parameter estimate $\xi^{(t)}$ at the global step t. Then perform the following local steps.

E – *step*. Calculate $\mathbf{r}_{isk}^{(t)}$ and $\mathbf{f}_{sk}^{(t)}$ using formulas (15-16) for the current parameter values $\xi^{(t)}$.

M – *step*. Solve equations (17-18) to find the next approximation for the item parameters and then calculate the next values for alphas by the numerical equivalent of formula (12)

$$\alpha_k^{(t+1)}=\frac{1}{N}\sum_{j=1}^{N}\frac{\alpha_k^{(t)}\sum_s w_s\prod_{i=1}^{n}P\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}}{\sum_{k=1}^{K}\alpha_k^{(t)}\sum_s w_s\prod_{i=1}^{n}P\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{u_{ji}}Q\left(x_s,a_{ik}^{(t)},b_{ik}^{(t)}\right)^{1-u_{ji}}},\ k=1,2,\ldots,K\,. \quad (19)$$

Thus we find $\xi^{(t+1)}$.

These steps are repeated until convergence stop criterion is fulfilled. For example, the iterations stop if the values of the left hand sides of equations (17-18) are very close to zero or when difference between two consequent parameter estimates becomes small enough.

One can solve equations (17-18) by means of the Newton or some quasi-Newton method.

ABILITY ESTIMATION

Having once estimated item parameters it remains to estimate the personal ability of person j, $j=1,\ldots,N$. For this purpose we use the **EAP** (***e****xpected* ***a*** ***p****osteriori*) method which is appropriate for the **MMLE** scheme. **EAP** estimates are the expectations of Θ, estimated posterior for a given person. By (10) we have

$$\overline{\theta}_j=\mathbf{E}\left[\Theta\,|\,\mathbf{u}_j\right]=\frac{\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty}\theta\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta}{\sum_{k=1}^{K}\alpha_k\int_{-\infty}^{\infty}\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta},\ j=1,2,\ldots,N\,. \quad (20)$$

Numerical integration implies the following estimation rules

$$\overline{\theta}_j\approx\frac{\sum_{k=1}^{K}\alpha_k\sum_s x_s\prod_{i=1}^{n}P(x_s,a_{ik},b_{ik})^{u_{ji}}Q(x_s,a_{ik},b_{ik})^{1-u_{ji}}w_s}{\sum_{k=1}^{K}\alpha_k\sum_s\prod_{i=1}^{n}P(x_s,a_{ik},b_{ik})^{u_{ji}}Q(x_s,a_{ik},b_{ik})^{1-u_{ji}}w_s},\ j=1,2,\ldots,N\,. \quad (21)$$

Also we can get conditional ability estimates for each class $(k=1,2,\ldots,K)$

$$\overline{\theta}_{jk}=\mathbf{E}\left[\Theta\,|\,\mathbf{u}_j,\boldsymbol{\kappa}=k\right]=\frac{\int_{-\infty}^{\infty}\theta\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta}{\int_{-\infty}^{\infty}\prod_{i=1}^{n}P(\theta,a_{ik},b_{ik})^{u_{ji}}Q(\theta,a_{ik},b_{ik})^{1-u_{ji}}\varphi(\theta)d\theta},\ j=1,2,\ldots,N\,, \quad (22)$$

with numerical form

$$\overline{\theta}_{jk} \approx \frac{\sum_{s} x_s \prod_{i=1}^{n} P(x_s, a_{ik}, b_{ik})^{u_{ji}} Q(x_s, a_{ik}, b_{ik})^{1-u_{ji}} w_s}{\sum_{s} \prod_{i=1}^{n} P(x_s, a_{ik}, b_{ik})^{u_{ji}} Q(x_s, a_{ik}, b_{ik})^{1-u_{ji}} w_s}, \quad j = 1,2,\ldots,N. \tag{23}$$

Formulas (21) and (23) are not iterative and give result even in the case when some of the person considered have extreme scores.

DEMONSTRATION OF SOFTWARE

The author propose a computer program in which **MMLE/EM** scheme was implemented for the described above latent class model as MS .NET application. Testing data is constructed by means of a software test data generator.

CONCLUSIONS AND FUTURE WORK

Latent class models are useful when we want to investigate the existence of different response styles in the sample. These differences may be connected with the gender or with the ethnical or cultural background of the persons. There are no essential obstacles to implement **MMLE/EM** latent class scheme to some graded response models like **PCM** or **GPCM** which is the next purpose for the author.

REFERENCES

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

[2] Hambleton R., 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., L. Hristov, Computer Implementation of MMLE/EM Algorithm for Two-Parametric Logistic Model, CompSysTech'2003.

ABOUT THE AUTHOR

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.