МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2007
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2007
Proceedings of the Thirty Sixth Spring Conference of
the Union of Bulgarian Mathematicians
St. Konstantin & Elena resort, Varna, April 2–6, 2007

EM PARAMETER ESTIMATION FOR THE RASCH MODEL*

Lyubomir Ya. Hristov

In this paper we give parameter estimation for the well-known in the Item Response Theory Rasch model *via* the EM algorithm which is presented here in its pure form, not hidden under the framework of the often used MMLE method.

1. Introduction. Here we suppose that the reader is familiar with the base of the Item Response Theory (IRT) [1–4]. The Rasch model is the simplest yet powerful model for binary items in the IRT [2, 4]. In this model the answer to the test item i, $i = 1, 2, \ldots, n$, is considered as a random variable U_i, accepting value 1 with probability

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

and value 0 with probability $Q(\theta, b_i) = 1 - P(\theta, b_i)$, where b_i is the item difficulty parameter. Here θ represents the personal ability level. Assuming the local independence, the test answer joint probability distribution of the random vector $\mathbf{U} = (U_1, \ldots, U_n)$ is given by

$$f(\mathbf{u}|\theta, \mathbf{b}) = \prod_{i=1}^{n} P(\theta, b_i)^{u_i} Q(\theta, b_i)^{1-u_i}.$$

The notation $\mathbf{b}$ stands for the set of the item parameters. We suppose that that persons who form the sample are drawn from some well defined population in which the ability is described by a random variable Θ with known normal distribution of a density $\varphi(\theta|\mu, \sigma)$. Then, the joint distribution of $(\mathbf{U}, \Theta)$ has a density

$$f(\mathbf{u}|\theta, \mathbf{b})\, \varphi(\theta|\mu, \sigma).$$

At this point, to avoid the indeterminacy in the parameter set, we put $\mu = 0$ and for the density of Θ we simply write $\varphi(\theta|\sigma)$. For the posterior distribution of Θ we have

$$f(\theta|\mathbf{u}, \xi) = \frac{f(\mathbf{u}|\theta, \mathbf{b})\, \varphi(\theta|\sigma)}{\int\limits_{-\infty}^{\infty} f(\mathbf{u}|\theta, \mathbf{b})\, \varphi(\theta|\sigma)\, d\theta}, \tag{1}$$

where the notation ξ stands for the set of all the item parameters plus σ.

Suppose we are given a data of N observation and $\mathbf{u}^{(j)} = (u_{j1}, u_{j2}, \ldots, u_{jn})$ is the answer vector of person j, $j = 1, 2, \ldots, N$. The problem considered here refers to the incomplete data analysis. In this case the complete data consists of the pairs $\left(\mathbf{u}^{(j)}, \theta\right)$ in

*2000 Mathematics Subject Classification: 62-06
Key words: Item Response Theory, Rasch Model, EM algorithm

which θ is unobservable. Therefore, we are at the position to apply the **EM** algorithm for the parameter estimation.

2. EM parameter estimation. The complete data log-likelihood function is given by

$$l\left(\xi\right)=\sum_{j=1}^{N}\ln\left[f\left(\mathbf{u}^{(j)}|\theta,\mathbf{b}\right)\varphi\left(\theta|\sigma\right)\right]. \tag{2}$$

The basic idea of the **EM** algorithm is to replace the terms in (2) with their conditional expectations with respect to the unobservable variable Θ on the base of the current parameter estimate $\xi^{(t)}$ at successive step t (**E**xpectation) and, then, to find the better values $\xi^{(t+1)}$ for the parameters (**M**aximization) by maximizing this conditional expectation.

Therefore, the **EM** algorithm consists of the repeating of the following two local steps:

E–step (**expectation**). Having parameter estimation $\xi^{(t)}$ at step t, determine the posterior densities

$$f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)=\frac{f\left(\mathbf{u}^{(j)}|\theta,\mathbf{b}^{(t)}\right)\varphi\left(\theta|\sigma^{(t)}\right)}{\int\limits_{-\infty}^{\infty}f\left(\mathbf{u}^{(j)}|\theta,\mathbf{b}^{(t)}\right)\varphi\left(\theta|\sigma^{(t)}\right)d\theta}.$$

M–step (**maximization**). Find $\xi=\xi^{(t+1)}$ that maximizes the conditional expectation

$$\mathbf{Q}\left(\xi|\xi^{(t)}\right)=\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\ln\left[f\left(\mathbf{u}^{(j)}|\theta,\mathbf{b}\right)\varphi\left(\theta|\sigma\right)\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta.$$

The last formula (3) can be rewritten in the form

$$\begin{aligned}\mathbf{Q}\left(\xi|\xi^{(t)}\right)&=\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\ln\left[\prod_{i=1}^{n}P\left(\theta,b_i\right)^{u_{ji}}Q\left(\theta,b_i\right)^{1-u_{ji}}\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta\\&\quad+\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\ln\left[\varphi\left(\theta|\sigma\right)\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta,\\ \mathbf{Q}\left(\xi|\xi^{(t)}\right)&=\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\sum_{i=1}^{n}\left[-\ln\left(1+\exp\left(\theta-b_i\right)\right)+u_{ji}\left(\theta-b_i\right)\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta\\&\quad+\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\left[-\frac{\theta^2}{2\sigma^2}-\ln\sigma\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta-N\sqrt{2\pi}.\end{aligned} \tag{4}$$

Therefore, to maximize (4) we have to solve the following likelihood equations $\frac{\partial}{\partial b_i}\mathbf{Q}\left(\xi|\xi^{(t)}\right)=0$, $i=1,2,\ldots,n$, and $\frac{\partial}{\partial\sigma}\mathbf{Q}\left(\xi|\xi^{(t)}\right)=0$, which equations have the form [1, 4, 5].

$$\sum_{j=1}^{N}\int\limits_{-\infty}^{\infty}\left[Q\left(\theta,b_i\right)-u_{ji}\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta=0,\quad i=1,2,\ldots,n, \tag{5}$$

and

$$\sum_{j=1}^{N}\int_{-\infty}^{\infty}\left[\frac{\theta^2}{\sigma^3}-\frac{1}{\sigma}\right]f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta=0. \tag{6}$$

Obviously,

$$\sum_{j=1}^{N}\int_{-\infty}^{\infty}u_{ji}f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta=\sum_{j=1}^{N}u_{ji}\int_{-\infty}^{\infty}f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta=\sum_{j=1}^{N}u_{ji}=s_i,$$

where s_i, $i=1,2,\ldots,n$, are the item scores. Thus, (5) reduces to

$$\sum_{j=1}^{N}\int_{-\infty}^{\infty}Q\left(\theta,b_i\right)f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta-s_i=0,\quad i=1,2,\ldots,n, \tag{7}$$

From (6) one can immediately get the solution

$$\sigma^{(t+1)}=\sqrt{\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}\theta^2 f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta.}$$

To solve equations (7) by the Newton method, we need also the second derivatives

$$\frac{\partial^2}{\partial b_i^2}\mathbf{Q}\left(\xi|\xi^{(t)}\right)=\sum_{j=1}^{N}\int_{-\infty}^{\infty}P\left(\theta,b_i\right)Q\left(\theta,b_i\right)f\left(\theta|\mathbf{u}^{(j)},\xi^{(t)}\right)d\theta.$$

3. Numerical solution. We use Gaussian numerical approximation to the integrals above with nodes x_s and weights w_s. These nodes and weights are for the case $\sigma=1$, i.e.

$$\int_{-\infty}^{\infty}g\left(\theta\right)\varphi\left(\theta|\sigma=1\right)d\theta\approx\sum_s w_s g\left(x_s\right),$$

which, in the general case, implies the following numerical formulae

$$\int_{-\infty}^{\infty}g\left(\theta\right)\varphi\left(\theta|\sigma\right)d\theta\approx\sum_s w_s g\left(\sigma x_s\right).$$

In this way equations (13) get the numerical form

$$\sum_{j=1}^{N}\sum_s Q\left(\sigma^{(t)}x_s,b_i\right)\frac{w_s f\left(\mathbf{u}^{(j)}|\sigma^{(t)}x_s,\mathbf{b}^{(t)}\right)}{\sum\limits_s w_s f\left(\mathbf{u}^{(j)}|\sigma^{(t)}x_s,\mathbf{b}^{(t)}\right)}-s_i=0,\quad i=1,2,\ldots,n. \tag{8}$$

Denote

$$\mathbf{f}_s^{(t)}=\sum_{j=1}^{N}\frac{w_s f\left(\mathbf{u}^{(j)}|\sigma^{(t)}x_s,\mathbf{b}^{(t)}\right)}{\sum\limits_s w_s f\left(\mathbf{u}^{(j)}|\sigma^{(t)}x_s,\mathbf{b}^{(t)}\right)}\,. \tag{9}$$

Now, the **EM** algorithm accepts the following numerical form. Choose initial values $\mathbf{b}^{(0)}$, usually set $b_i=0$, $i=1,2,\ldots,n$, and $\sigma^{(0)}=1$. Having estimation $\mathbf{b}^{(t)}$ and $\sigma^{(t)}$ at the global step t, perform the next two local steps.

E–step. Calculate $\mathbf{f}_s^{(t)}$ using formula (9).

M–step. Solve equations (8) to find the next approximation $\mathbf{b}^{(t+1)}$ for the item parameters using the Newton iterations

$$\beta_i^{(\tau+1)} = \beta_i^{(\tau)} - \frac{\sum\limits_s Q\left(\sigma^{(t)}x_s, \beta_i^{(\tau)}\right)\mathbf{f}_s^{(t)} - s_i}{\sum\limits_s P\left(\sigma^{(t)}x_s, \beta_i^{(\tau)}\right) Q\left(\sigma^{(t)}x_s, \beta_i^{(\tau)}\right)\mathbf{f}_s^{(t)}}, \quad i = 1, 2, \ldots, n,$$

starting with $\beta_i^{(0)} = b_i^{(t)}$, and set $b_i^{(t+1)} = \beta_i^{\tau_{\text{final}}}$. Finally, calculate the next approximation value for σ by

$$\sigma^{(t+1)} = \sqrt{\frac{1}{N}\sum_s \left(\sigma^{(t)}x_s\right)^2 \mathbf{f}_s^{(t)}}.$$

These global steps are repeated until stop criterion is met.

4. Ability estimation. Suppose we have already an estimation for the item parame-rs. The next problem is to estimate the personal ability of person j, $j = 1, \ldots, N$. For this purpose we use the **EAP** method. The **EAP** estimate $\bar{\theta}_j$ are the expectation of Θ, estimated posterior for a given person j by means of the Bayes formula [1, 3]. From (1) we have

$$\bar{\theta}_j = \mathbf{E}\left[\Theta|\mathbf{u}^{(j)}\right] = \frac{\int\limits_{-\infty}^{\infty} \theta \prod\limits_{i=1}^{n} P\left(\theta|b_i\right)^{u_{ji}} Q\left(\theta|b_i\right)^{1-u_{ji}} \varphi\left(\theta|\sigma\right) d\theta}{\int\limits_{-\infty}^{\infty} \prod\limits_{i=1}^{n} P\left(\theta|b_i\right)^{u_{ji}} Q\left(\theta|b_i\right)^{1-u_{ji}} \varphi\left(\theta|\sigma\right) d\theta}, \quad j = 1, 2, \ldots, N.$$

Numerical integration produces the following results

$$\bar{\theta}_j \approx \sigma \frac{\sum\limits_s w_s x_s \prod\limits_{i=1}^{n} P\left(\sigma x_s|b_i\right)^{u_{ji}} Q\left(\sigma x_s|b_i\right)^{1-u_{ji}}}{\sum\limits_s w_s \prod\limits_{i=1}^{n} P\left(\sigma x_s|b_i\right)^{u_{ji}} Q\left(\sigma x_s|b_i\right)^{1-u_{ji}}}, \quad j = 1, 2, \ldots, N. \tag{10}$$

Formula (10) is not iterative and gives a result even in the case when the person considered has extreme score.

5. Demonstration of software. The author propose a computer program in which the **EM** algorithm was implemented for the model described above as a MS .NET application. Testing data is constructed by software test data generator. The experimental results appear to be very good because the estimated values of σ occur significantly close to the corresponding values of the simulated test data.

6. Conclusions and future work. The **EM** algorithm offers a convenient way to obtain a parameter estimation for the Rasch model. In our future work we are going to apply this approach to investigate the corresponding latent class model.

REFERENCES

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

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

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

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

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

Lyubomir Yanakiev Hristov
Department of Mathematical analysis and application
University of Veliko Tarnovo "St. st. Cyril and Methodius"
Veliko Tarnovo, Bulgaria
e-mail: Lyubo60@yahoo.com

ОЦЕНКА НА ПАРАМЕТРИТЕ ЗА МОДЕЛА НА РАШ С ПОМОЩТА НА ЕМ АЛГОРИТЪМА

Любомир Я. Христов

В статията се дава оценка на параметрите за добре познатият в Съвременната теория на тестовете модел на Раш чрез ЕМ алгоритъма. Описан е в чиста форма за разлика от MMLE метода, при който приложението му не е в явна форма.