Genetic Algorithm Based Optimization in the Two-Parametric Logistic Model

Svilen Stefanov, Dimiter Tsvetkov

Abstract: *Estimation of the structural parameters in the two or three parametric IRT models shows some problems as lack of convergence or obtaining of non-realistic values. These models are estimated usually by means of the MMLE/EM scheme in which the Newton-Raphson method is used during the optimization M-step. Here we propose another approach using the genetic algorithm as an optimization tool. We give a computer realization of this approach with comparison with the traditional ones.*
Key words: *IRT, MMLE/EM, Two-Parametric Model, Genetic Algorithm Optimization*

INTRODUCTION

Here we assume that the reader is familiar with the base of the Item Response Theory [1-4]. We consider a binary test of length n. Each answer to i-th item, $i = 1,2,\dots,n$, represents a random variable accepting value 1 with probability $P_i(\theta)$ and value 0 with probability $Q_i(\theta) = 1 - P_i(\theta)$ therefore the answer joint probability is

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

Here the probability of the response vector $\mathbf{U} = (U_1, U_2, \dots, U_n)$ is conditional to the ability value θ of the respondent person. In the MMLE scheme, θ is considered as a random variable with known distribution density $\varphi(\theta)$ usually normal standard. In this way the joint distribution of $\mathbf{U}$ and θ is $\Pr(\mathbf{U} \mid \theta)\varphi(\theta)$ and for the marginal distribution of $\mathbf{U}$ we have

$$\Pr(\mathbf{U}) = \int \Pr(\mathbf{U} \mid \theta)\varphi(\theta) d\theta \,. \tag{2}$$

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

$$\ln \Pr(\mathbf{u}) = \ln \int \Pr(\mathbf{u} \mid \theta)\varphi(\theta) d\theta \,, \tag{3}$$

where $\mathbf{u} = (u_1, u_2, \dots, u_n)$ is the answer pattern. Including all persons in analysis implies the following logarithmic likelihood function

$$LHF = \sum_{j=1}^{N} \ln \int \Pr(\mathbf{u}_j \mid \theta)\varphi(\theta) d\theta \,, \tag{4}$$

As usual we assume identical distribution of θ for all persons**.** Therefore the function to be optimized with respect to the all item parameters has the form

$$LHF = \sum_{j=1}^{N} \ln \int \prod_{i=1}^{n} [P_i(\theta)]^{u_{ij}} [Q_i(\theta)]^{1-u_{ij}} \varphi(\theta) d\theta \,, \tag{5}$$

where $u_{ij} = 1$ if the j-th person answers to the i-th item correctly and $u_{ij} = 0$ otherwise. The form of the probability $P_i(\theta)$ depends on the model type. Here we will use the two-parametric model of the form

$$P_i(\theta) = \frac{e^{a_i\theta - \xi_i}}{1 + e^{a_i\theta - \xi_i}} \,, \ i = 1,2,\dots,n \,, \tag{6}$$

where a_i is the item discrimination parameter and ξ_i is the second item parameter for which it holds $\xi_i = a_i b_i$ where b_i is the usual item difficulty parameter.

MARGINAL MAXIMUM LIKELIHOOD ESTIMATION

Marginal maximum likelihood principal means to find the parameter values which maximize the likelihood function LHF. In fact we solve the system of maximum likelihood equations $\nabla LHF = 0$, where ∇LHF is the gradient of LHF. Maximum likelihood equations (**MLE**) have the form $\partial_\eta LHF = 0$, $\eta = a_i$ or $\eta = \xi_i$, $i = 1,2,\ldots,n$, which (see for example [5]), can be written as

$$\sum_{j=1}^{N}\int\left[\frac{\partial_\eta Q_i(\theta)}{Q_i(\theta)} + u_{ij}\frac{\partial_\eta\left(P_i(\theta)\big/Q_i(\theta)\right)}{P_i(\theta)\big/Q_i(\theta)}\right]\Pr\left(\theta \mid \mathbf{u}_j\right)d\theta\,,\ i = 1,2,\ldots,n\,, \tag{7}$$

where

$$\Pr\left(\theta \mid \mathbf{u}_j\right) = \frac{\prod_{i=1}^{n} P_i(\theta)^{u_{ij}} Q_i(\theta)^{1-u_{ij}} \varphi(\theta)}{\int \prod_{i=1}^{n} P_i(\theta)^{u_{ij}} Q_i(\theta)^{1-u_{ij}} \varphi(\theta) d\theta}\,. \tag{8}$$

Therefore $2n$ equations (7) have to be solved.

NUMERICAL INTEGRATION, EM SOLUTION AND ABILITY ESTIMATION

We use Gaussian numerical approximation to the integrals above with nodes x_k and weights w_k with 21 nodes. For the sake of completeness we offer again these values in the next table.

x[1]:=-7.849382895113821993	w[1]:=5.2613907391655318390472575700 38*0.00000000000001
x[2]:= -6.751444718717460767	w[2]:=1.247139961980288641761962577712*0.0000000001
x[3]:=-5.829382007304471372	w[3]:=3.636268592622896655215704715827*0.00000001
x[4]:=-4.994963944782025193	w[4]:=3.071509078745049070684688201 1858*0.000001
x[5]:=-4.214343981688421350	w[5]:=0.000105760531029854805547453097 98247
x[6]:=-3.469846690475376295	w[6]:=0.001774812623944167273627024097 2564
x[7]:=-2.7505929810523730936	w[7]:=0.016141926709119123706916429168358
x[8]:=-2.0491024682571626618	w[8]:=0.085106872483854444609410120550 63
x[9]:=-1.3597658232112302657	w[9]:=0.271699167903062365024889819128 74
x[10]:=-0.6780456924406440262	w[10]:=0.539761580242589855335404849388
x[11]:=0.0000000000000000000	w[11]:=0.677441817650738318137258381269 4
x[12]:=0.6780456924406440262	w[12]:=0.539761580242589855335404849388
x[13]:=1.3597658232112302657	w[13]:=0.271699167903062365024889819128 74
x[14]:=2.0491024682571626618	w[14]:=0.085106872483854444609410120550 63
x[15]:=2.7505929810523730936	w[15]:=0.016141926709119123706916429168358
x[16]:=3.469846690475376295	w[16]:=0.001774812623944167273627024097 2564
x[17]:=4.214343981688421350	w[17]:=0.000105760531029854805547453097 98247
x[18]:=4.994963944782025193	w[18]:=3.071509078745049070684688201 1858*0.000001
x[19]:=5.829382007304471372	w[19]:=3.636268592622896655215704715827*0.00000001
x[20]:=6.751444718717460767	w[20]:=1.247139961980288641761962577712*0.0000000001
x[21]:=7.849382895113821993	w[21]:=5.2613907391655318390472575700 38*0.00000000000001

Table 1: ***Nodes and weights for the numerical integration***

Then we get the following numerical form of the **MLE**

$$\sum_k\left[\frac{\partial_\eta Q_i(\theta)}{Q_i(\theta)}\mathbf{f}_{ik} + \frac{\partial_\eta\left(P_i(\theta)\big/Q_i(\theta)\right)}{P_i(\theta)\big/Q_i(\theta)}\mathbf{r}_{ik}\right] = 0\,,\ i = 1,2,\ldots,n\,, \tag{9}$$

where

$$\mathbf{r}_{ik} = \sum_{j=1}^{N}\frac{u_{ij}\prod_{i=1}^{n}\left[P_i(x_k)\right]^{u_{ij}}\left[Q_i(x_k)\right]^{1-u_{ij}} w_k}{\sum_k\prod_{i=1}^{n}\left[P_i(x_k)\right]^{u_{ij}}\left[Q_i(x_k)\right]^{1-u_{ij}} w_k}\,,\ \mathbf{f}_{ik} = \sum_{j=1}^{N}\frac{\prod_{i=1}^{n}\left[P_i(x_k)\right]^{u_{ij}}\left[Q_i(x_k)\right]^{1-u_{ij}} w_k}{\sum_k\prod_{i=1}^{n}\left[P_i(x_k)\right]^{u_{ij}}\left[Q_i(x_k)\right]^{1-u_{ij}} w_k}\,. \tag{10}$$

For the model considered equations (9) become

$$\sum_k\left[-\frac{e^{a_i x_k}}{e^{a_i x_k} + e^{\xi_i}}\mathbf{f}_{ik} + \mathbf{r}_{ik}\right]x_k = 0 \text{ for } a_i\,,\ i = 1,2,\ldots,n\,, \tag{11}$$

$$\sum_k \left[-\frac{e^{a_i x_k}}{e^{a_i x_k} + e^{\xi_i}} \mathbf{f}_{ik} + \mathbf{r}_{ik} \right] = 0 \text{ for } b_i,\ i = 1,2,\ldots,n. \qquad (12)$$

EM algorithm [1] consists of two steps. In the **E** – step we calculate $\mathbf{f}_{ik}$ and $\mathbf{r}_{ik}$ using formulas (10) and current parameter values and in the **M** – step we solve equations (11-12). These steps are repeated until some proper stop criterion is met. Equations (11-12) can be solved by means of the Newton-Raphson or some quasi-Newton method like BFGS. Here we offer a Genetic Algorithm approach which also is appropriate to derive the solution.

Having once the item parameters values one can find the person ability estimate $\hat{\theta}_j$ by means of the expected posterior method following the next numerical Bayesian formula

$$\hat{\theta}_j = \frac{\sum_k x_k \prod_{i=1}^n \left[P_i(x_k)\right]^{u_{ij}} \left[Q_i(x_k)\right]^{1-u_{ij}} w_k}{\sum_k \prod_{i=1}^n \left[P_i(x_k)\right]^{u_{ij}} \left[Q_i(x_k)\right]^{1-u_{ij}} w_k},\ j = 1,2,\ldots,N. \qquad (13)$$

THE GENETIC ALGORITHMS APPROACH

The Genetic Algorithms (**GA**s) have a great advantage with respect to many other optimization tools the main of which is the avoidance of the local maxima. The reader may find an excellent introduction to **GA**s in the books [5-6]. Also these algorithms are implemented as a MATLAB toolbox. Function to be optimized with respect to $2n$ parameters a_i and ξ_i, $i = 1,2,\ldots,n$, has the following numerical form

$$LHF = \sum_{j=1}^N \ln \sum_{k=1}^{21} \prod_{i=1}^n w_k \left[P_i(x_k)\right]^{u_{ij}} \left[Q_i(x_k)\right]^{1-u_{ij}} = \sum_{j=1}^N \ln \left[\sum_{k=1}^{21} w_k \prod_{i=1}^n \frac{e^{\xi_i} e^{(x_k a_i - \xi_i) u_{ij}}}{e^{x_k a_i} + e^{\xi_i}} \right]. \qquad (14)$$

In this way one can use direct optimization of LHF as a function of $2n$ variables. Such a solution seems to be interesting in many aspects. First of all, here we use the full power of the genetic algorithms approach and perhaps (under some sophistications) this is the way to avoid unrealistic parameter values. On the second hand one can explore **GA** in the **M**-step finding the solution of the equations (11-12) by minimizing the following cost function

$$CF = \sum_{i=1}^n \left| \sum_k \left[-\frac{e^{a_i x_k}}{e^{a_i x_k} + e^{\xi_i}} \mathbf{f}_{ik} + \mathbf{r}_{ik} \right] x_k \right| + \sum_{i=1}^n \left| \sum_k \left[-\frac{e^{a_i x_k}}{e^{a_i x_k} + e^{\xi_i}} \mathbf{f}_{ik} + \mathbf{r}_{ik} \right] \right|. \qquad (15)$$

The second way is significantly faster that the first one because the main disadvantage of the **GA**s is the very frequent calculation of the cost function. It is relatively hard to compute the values of the maximum likelihood function LHF because of its long terms.

Here we present an example to make comparison between different optimization methods. We use a generated test result with $n = 11$ items and $N = 100$ persons.

item	Estimate of a via BFGS method	Estimate of a via Genetic Algorithm method	Estimate of b via BFGS method	Estimate of b via Genetic Algorithm method
1	1.226	1.219	1.945	1.858
2	1.564	1.412	1.237	1.239
3	0.172	0.198	0.467	0.308
4	0.833	0.849	1.358	1.299
5	0.919	0.868	2.878	2.802
6	0.139	0.206	8.722	5.606
7	1.085	1.101	1.005	0.948
8	1.838	2.098	-0.053	-0.121
9	1.527	1.373	1.253	1.250
10	0.800	0.822	0.999	0.936
11	3.171	2.722	0.432	0.371

*Table 2: **Comparison between different estimates***

The results are close enough each other nevertheless that there exist some differences. Remember that $ab = \xi$.

DEMONSTRATION OF SOFTWARE

The authors propose a computer realization of the method above. We would be very thankful for any comments and recommendations.

CONCLUSIONS AND FUTURE WORK

Obviously the **GA**s have many significant advantages which are not sufficiently investigated with respect to the IRT class problems. We are going to implement **GA**s to some other models also we are going to try some hybrid modifications.

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.

[6] Haupt L., and E. Haupt, Practical Genetic Algorithms, Wiley-Interscience, 2004.

[7] Mitchell M., An Introduction to Genetic Algorithms, MIT Press, 1998.

ABOUT THE AUTHORS

Assoc. Prof. Svilen Evtimov Stefanov, PhD, department of "Technical Logistics", NVU "Vassil Levski", Veliko Tarnovo, phone: +359 887 706 710, E-mail: sestefanov@abv.bg.

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.