Holonomic Approach for Item Response Theory Parameter Estimation

Noguchi, Kazuhisa
Department of library Science, Graduate School of Integrated Frontier Science, Kyushu University

Ito, Eisuke
Research Institute for Information Technology, Kyushu University

https://hdl.handle.net/2324/1936750

出版情報：2018-07-10
バージョン：
権利関係：

Holonomic Approach for Item Response Theory Parameter Estimation

Kazuhisa Noguchi
Grad. School of Library Science
Kyushu University
Fukuoka, Japan 819-0395
Email: noguchi.k.567@s.kyushu-u.ac.jp

Eisuke Ito
Research Institute for IT
Kyushu University
Fukuoka, Japan 819-0395
ito.eisuke.523@m.kyushu-u.ac.jp

***Abstract*—The IRT (item response theory) is a theory for scoring of tests, and it is used for some test systems such as TOEFL. Classical test scoring is the total of raw scores for each items (questions). On the other hand, IRT can make fine performance finer than the raw score method. IRT is based on the relationship between individual test takers' performances on a test item (question), and the test takers' levels of performance. Even with the same number of correct answers, the IRT score may be different depending on the degree of difficulty of item (question). Computational complexity of parameter estimation of IRT is one of important issue. To solve this issue, we propose a parameter estimation method of IRT using the Holonomic approach. Generate differential equations which solve the likelihood function of IRT. By solving this differential equation, it is able to estimate the ability parameter.**

***Index Terms*—Item Response Theory; IRT; parameter estimation; holonomic function; ability**

I. Introduction

IRT (Item Response Theory) is a theory for scoring of tests, and it used for some test systems such as TOEFL [4]. Classical test scoring is the total of raw scores for each items (questions). On the other hand, IRT can make fine performance finer than the raw score method. IRT is based on the relationship between individual examinees' (test takers') performances on a test item (question), and the examinees' levels of performance. Even with the same number of correct answers, the IRT score may be different depending on the degree of difficulty of item (question).

In IRT, estimate item parameter for each question, and then estimate examinee's ability parameter. Conventional IRT, examinee's ability is estimated by Newton-Raphson method [5]. However, Newton-Raphson method always diverge and fail about estimation. To solve the problem of convergence and solution, we propose a new method to estimate ability parameter of IRT using the Holonomic approach. The Holonomic approach is an optimization algorithm which generates a differential equation, and solves the target function $f(x)$, and performs numerical calculation of x [11, 9, 10, 7, 4]. In this case, we consider IRT's likelihood function to be computed, generate the differential equation and estimate examinee's ability parameter.

II. Item Response Theory

This section describes the theory and operation method of IRT [1, 2, 8]. We also describe problem of parameter estimation.

A. IRT Logistic Model

At first, let us consider 1 parameter logistic model (1PLM) of IRT with b_j and θ_i, where b_j is the difficulty parameter for item j, and θ_i is the ability parameter of examinee i. Let $\mathrm{P}_j(\theta_i)$ be the probability of a correct response to an item j by examinee i. $\mathrm{P}_j(\theta_i)$ obeys the following logistic model.

$$\mathrm{P}_j(\theta_i) \;=\; \frac{1}{1+\exp(-D(\theta_i - b_j))}$$

where D on the right side is called a scale factor, and $D = 1.7$. A larger value indicates that the examinee's ability is higher. If b_j is large, then item j is difficult, and if it is low then j is easy.

There are the 2 parameter logistic model (2PLM) and the 3 parameter logistic model (3PLM). In addition to the item parameters, 2PLM includes the discrimination parameter a_j for item j. 3PLM includes the guessing parameter c_j. In this paper, we consider 1PLM, so we only handle difficulty parameter b_j of item j.

III. Parameter Estimation

Here, we describe the likelihood estimation of IRT. The ability parameter is estimated by the likelihood estimation.

A. Maximum Likelihood Estimation

When examinee $i = 1, 2, \cdots, I$ take the test of the number of questions, $j = 1, 2, \cdots, J$, the realization probability (Likelihood) for the test score is as follows .

$$\mathrm{L}(\theta_i, b_j|\delta_{ij}) = \prod_{i=1}^{I}\prod_{j=1}^{J} \mathrm{P}_j(\theta_i)^{\delta_{ij}}(1-\mathrm{P}_j(\theta_i))^{1-\delta_{ij}}$$

θ_i and b_j are all unknown. To find the item parameter b_j and the ability parameter θ_i that maximize this Likelihood $\mathrm{L}(\theta_i, b_j|\delta_{ij})$. This method is called maximum likelihood estimation. At this time, δ_{ij} is the correctness (reaction) that the examinee i answered to the question j,

$$\delta_{ij} = \begin{cases} 1 & \text{(Correct)} \\ 0 & \text{(False)} \end{cases}$$

B. IRT Optimization Problem

The maximum likelihood estimation method in IRT is considered as a maximization problem of the likelihood $\mathrm{L}(\theta_i, b_j|\delta_{ij})$. It can be expressed as follows.

$$\begin{aligned} \max \quad & \mathrm{L}(\theta_i, b_j|\delta_{ij}) \\ s.t. \quad & \theta_i, b_j \in \mathbf{R}^n \end{aligned}$$

Parameter estimation is performed by solving this maximization problem.

C. Log Likelihood Function

About the likelihood $\mathrm{L}(\theta_i, b_j|\delta_{i,j})$

$$\mathrm{L}(\theta_i, b_j|\delta_{i,j}) = \prod_{i=1}^{I}\prod_{j=1}^{J} \mathrm{P}_j(\theta_i)^{\delta_{ij}}(1-\mathrm{P}_j(\theta_i))^{1-\delta_{ij}}$$

taking natural logarithms on both sides, then we obtain the log likelihood function

$$\begin{aligned} \ln \mathrm{L}(\theta_i, b_j|\delta_{i,j}) &= \ln \prod_{i=1}^{I}\prod_{j=1}^{J} \mathrm{P}_j(\theta_i)^{\delta_{ij}}(1-\mathrm{P}_j(\theta_i))^{1-\delta_{ij}} \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\{\delta_{ij}\ln \mathrm{P}_j(\theta_i, b_j) \\ &\qquad + (1-\delta_{ij})\ln(1-\mathrm{P}_j(\theta_i, b_j s))\}. \end{aligned}$$

In other words, finding the maximum likelihood is equal to estimation of $\hat{\theta}_i$ and $\hat{b_j}$ which satisfies

$$\begin{cases} \dfrac{\partial \ln \mathrm{L}(\theta_i, b_j|\delta_{i,j})}{\partial \theta_i} = 0, \ i = 1, \ldots, I \\ \dfrac{\partial \ln \mathrm{L}(\theta_i, b_j|\delta_{i,j})}{\partial b_j} = 0, \ j = 1, \ldots, J. \end{cases}$$

We find the parameter θ_i and b_j to maximize the likelihood $\mathrm{L}(\theta_i, b_j|\delta_{i,j})$.

D. Newton-Raphson Method

Newton-Raphson method is used for maximum likelihood estimation for parameter estimation of IRT. The Newton-Raphson method is a method to solve $\mathbf{x} = (x_1, \ldots, x_n)$ which maximizes the value of the function $f(\mathbf{x})$. The solution derived from the initial value $\mathbf{x}_0 = (x_{10}, \ldots, x_{n0})$ and derived at the k times as x_k by the following updating expression.

$$\mathbf{x}_{k+1} := \mathbf{x}_k - \mathbf{H}^{-1}(\mathbf{x}_k)\nabla_{\mathbf{x}_k}(\mathbf{x}_k)$$

$\mathbf{H}(\mathbf{x}_k)$ is a Hessian matrix.

To apply Newton-Raphson method for the log-likelihood function, the update expression of θ_i and b_j is $\ln \mathrm{L}(\theta_i, b_j|\delta_{ij})$ as the target function $f(\mathbf{x})$.

The IRT parameter is solved by updating these equations by the following procedure.

1) Given the initial value of θ_i and b_j
2) Update the value of θ_i under the given b_j
3) Update the value of b_j under the given θ_i
4) Repeat step 2 and 3until the solution converges.

IV. Holonomic Approach

This section explains the holonomic gradient method. This method is able to apply to numerical calculation of probability distribution functions.

A. Holonomic Function

A function $f(x)$ is a holonomic function if the function $f(x)$ becomes the solution of the following linear differential equation in the multivariate $x \in \mathbb{R}^n$.

$$\partial_x^r f(x) + \cdots + a_k(x)\partial_x^k f(x) + \cdots + a_1(x) f(x) = 0$$

At this time,

$$\partial_x^k := \frac{d^k}{dx^k}.$$

where, $a_k(x)$ is a function for x.

B. Holonomic Approach

As we mentioned before subsection, Holonomic Approach is used as a method of numerical calculation of $f(x)$, where $f(x)$ is a holonomic function. The numerical calculation of the holonomic function is archived by the following procedure.

1) Find a differential equation which the solution is the function $f(x)$.
2) Determine the initial value of x.
3) Find $f(x)$ as a solution to the obtained differential equation.

V. IRT Parameter Estimation by Holonomic Approach

In this section, we describe an optimization method by holonomic approach to maximum likelihood estimation of IRT. Confirm that differential equations of the holonomic function of likelihood can be generated for the parameters θ_i and b_j.

A. IRT Model by Holonomic Approach

In order to use the holonomic gradient method of IRT model, we transform the likelihood function, and set the transformed one as the objective function. The likelihood function $\mathrm{L}(\theta_i)$ is transformed as follows.

$$\begin{aligned}
\mathrm{L}(\theta_i, b_j|\delta_{ij}) &= \prod_{i=1}^{I}\prod_{j=1}^{J} \mathrm{P}_j(\theta_i)^{\delta_{ij}}(1-\mathrm{P}_j(\theta_i))^{1-\delta_{ij}} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{\exp(D(\theta_i-b_j)\delta_{ij})}{1+\exp(D(\theta_i-b_j))} \\
&= \frac{\prod_{i=1}^{I}\prod_{j=1}^{J}\exp(D(\theta_i-b_j)\delta_{ij})}{\prod_{i=1}^{I}\prod_{j=1}^{J}\{1+\exp(D(\theta_i-b_j))\}} \\
&= \frac{\exp\left(D\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_i-b_j)\delta_{ij}\right)}{\prod_{i=1}^{I}\prod_{j=1}^{J}\{1+\exp(D(\theta_i-b_j))\}}
\end{aligned}$$

We take the reciprocal of the likelihood function $\mathrm{L}(\theta_i, b_j|\delta_{ij})$ and represent it as $f(\theta_i, b_j)$.

$$\begin{aligned}
f(\theta_i, b_j) &= \{\mathrm{L}(\theta_i, b_j|\delta_{ij})\}^{-1} \\
&= \exp\left(-D\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_i-b_j)\delta_{ij}\right) \\
&\quad \times \prod_{i=1}^{I}\prod_{j=1}^{J}\{1+\exp(D(\theta_i-b_j))\}
\end{aligned}$$

Here we check whether the function $f(\theta_i)$ is a holonomic function or not. It can be established if a differential equation having the function $f(\theta_i)$ is generated. The function $f = f(\theta_i)$ is

$$f = gh.$$

Each part is expressed as

$$\begin{aligned}
g &= \exp\left(-D\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_i-b_j)\delta_{ij}\right) \\
h &= \prod_{i=1}^{I}\prod_{j=1}^{J}\{1+\exp(D(\theta_i-b_j))\}.
\end{aligned}$$

Differentiate g with θ_i,

$$\begin{aligned}
\partial_{\theta_i} g &= -\left(D\sum_{j=1}^{J}\delta_{ij}\right)\exp\left(-D\sum_{j=1}^{J}(\theta_i-b_j)\delta_{ij}\right) \\
&= -\left(D\sum_{j=1}^{J}\delta_{ij}\right) g.
\end{aligned}$$

When differentiating g with b_j,

$$\begin{aligned}
\partial_{b_j} g &= \left(D\sum_{i=1}^{I}\delta_{ij}\right)\exp\left(-D\sum_{i=1}^{I}\sum_{j=1}^{J}(\theta_i-b_j)\delta_{ij}\right) \\
&= \left(D\sum_{i=1}^{I}\delta_{ij}\right) g
\end{aligned}$$

Also,

$$\begin{aligned}
\sum_{i=1}^{I}(\partial_{\theta_i} g) &= \left(\sum_{i=1}^{I}\partial_{\theta_i}\right) g \\
&= -\left(D\sum_{i=1}^{I}\sum_{j=1}^{J}\delta_{ij}\right) g, \\
\sum_{j=1}^{J}(\partial_{b_j} g) &= \left(\sum_{j=1}^{J}\partial_{b_j}\right) g \\
&= \left(D\sum_{i=1}^{I}\sum_{j=1}^{J}\delta_{ij}\right) g.
\end{aligned}$$

And we take

$$\left(\sum_{i=1}^{I}\partial_{\theta_i}\right) g + \left(\sum_{j=1}^{J}\partial_{b_j}\right) g = 0.$$

Now we obtain a differential equation with solution g. That is

$$\left(\sum_{i=1}^{I}\partial_{\theta_i} + \sum_{j=1}^{J}\partial_{b_j}\right)\cdot g = 0.$$

That is, the function g is a holonomic function.

Next, we consider differentiating of h.

$$\begin{aligned}
h &= \prod_{i=1}^{I}\prod_{j=1}^{J} h_j \\
h_{ij} &= 1+\exp(D(\theta_i-b_j)).
\end{aligned}$$

When differentiating h_{ij} with θ_i,

$$\begin{aligned}
\partial_{\theta_i} h_{ij} &= D\exp(D(\theta_i-b_j)) \\
&= D(h_{ij}-1).
\end{aligned}$$

When differentiating h_j with b_j,

$$\begin{aligned}
\partial_{b_j} h_{ij} &= -D\exp(D(\theta_i-b_j)) \\
&= -D(h_{ij}-1).
\end{aligned}$$

And we take

$$\partial_{\theta_i} h_{ij} + \partial_{b_j} h_{ij} = 0.$$

Now we obtain a differential equation with solution h_{ij}.

$$(\partial_{\theta_i} + \partial_{b_j})\cdot h_{ij} = 0.$$

Therefore, the function h_{ij} is a holonomic function. And

$$\begin{aligned} h &= \prod_{i=1}^{I}\prod_{j=1}^{J} h_{ij} \\ &= \prod_{i=1}^{I}\prod_{j=1}^{J}\{1+\exp(D(\theta_i - b_j)\}. \end{aligned}$$

We get

$$\left(\sum_{i=1}^{I}\partial_{\theta_i} + \sum_{j=1}^{J}\partial_{b_j}\right)\cdot h = 0.$$

The function h is a holonomic function. Differential equations with the function f as a solution can be obtained by differentiating the function f with θ_j

$$\begin{aligned} \partial_{\theta_i} f &= \partial_{\theta_i}(gh) \\ &= (\partial_{\theta_i} g)h + g(\partial_{\theta_i} h). \end{aligned}$$

When differentiating f with b_j,

$$\begin{aligned} \partial_{b_j} f &= \partial_{b_j}(gh) \\ &= (\partial_{b_j} g)h + g(\partial_{b_j} h). \end{aligned}$$

Also

$$\begin{aligned} \left(\sum_{i=1}^{I}\partial_{\theta_i}\right) f &= \left(\sum_{i=1}^{I}\partial_{\theta_j} g\right) h + g\left(\sum_{i=1}^{I}\partial_{\theta_j} h\right), \\ \left(\sum_{j=1}^{J}\partial_{b_j}\right) f &= \left(\sum_{j=1}^{J}\partial_{b_j} g\right) h + g\left(\sum_{j=1}^{J}\partial_{b_j} h\right). \end{aligned}$$

Therfore,

$$\begin{aligned} &\left(\sum_{i=1}^{I}\partial_{\theta_i}\right) f + \left(\sum_{j=1}^{J}\partial_{b_j}\right) f \\ &= \left(\sum_{i=1}^{I}\partial_{\theta_j} g\right) h + g\left(\sum_{i=1}^{I}\partial_{\theta_j} h\right) \\ &\quad + \left(\sum_{j=1}^{J}\partial_{b_j} g\right) h + g\left(\sum_{j=1}^{J}\partial_{b_j} h\right) \\ &= \left(\left(\sum_{i=1}^{I}\partial_{\theta_i} + \sum_{j=1}^{J}\partial_{b_j}\right)\cdot g\right) h \\ &\quad + g\left(\left(\sum_{i=1}^{I}\partial_{\theta_i} + \sum_{j=1}^{J}\partial_{b_j}\right)\cdot h\right) \\ &=0. \end{aligned}$$

Finally,

$$\left(\sum_{i=1}^{I}\partial_{\theta_i} + \sum_{j=1}^{J}\partial_{b_j}\right)\cdot f = 0.$$

It was shown that the function f is a holonomic function.

B. Holonomic Approach optimization problem

The reciprocal of the Likelihood of item response theory was shown to be a holonomic function. Therefore, the optimization problem of item response theory Parameter Estimation can be replaced with the following minimization problem.

$$\begin{aligned} \min \quad & f = \{\mathrm{L}(\theta_i, b_j|\delta_{ij})\}^{-1} \\ s.t. \quad & \theta_i, b_j \in \mathbf{R}^n \end{aligned}$$

We can generate differential equations with the function f as a solution, so we can use methods such as Euler's method and Runge-Kutta method, which are calculation methods of differential equations, so that new calculation methods are expected.

VI. Conclusion

In this study, we try to We showed that the holonomic approach is possible for the Estimation of the performance Parameter θ_i and b_j in the item response theory, which is an objective function obtained by transforming the Likelihood function of the ability value Parameter derived by the item response theory A differential equation with a solution of $f = \{\mathrm{L}(\theta_i, b_j|\delta_{ij})\}^{-1}$ was obtained.In this differential equation, the Likelihood function of the item response theory is a holonomic function In the future, we will do a holonomic approach to item Parameters.

References

[1] Baker, F, B., and Kim, S.-H. (2004). Item response theory: Parameter estimation techniques. 2nd ed. New York, NY: Marcal Dekker.

[2] Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), Statistical theories of measurement scores (pp. 395–479). Reading, MA: Addison-Wesley.

[3] Hashiguchi, H., Numata, Y., Takayama, and N., Takemura, A. (2013). Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, J= Multivariiate Anal., 117, (2013) 296- 312.

[4] Jang, E. E., and Roussos, L. A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach, Journal of Educational Measurement, 44, 1–21.

[5] Jennrich, R, I., and Sampson, P. F. (1976). Newton-Raphson and related algorithms for maximum likelihood variance component estimation, Technometrics, 18, 11–17.

[6] Koyama, T., Nakayama, H., Nishiyama, K., and Takayama, N. (2014). Holonomic gradient descent for the Fisher-Bingham distribution on the d - dimensional sphere, Computational Statistics, 29 (3), 661-683.

[7] Lord, F. M. (1952). A theory of test scores, Psychometrika Monograph, No.7, 17.

[8] Nakayama, H., Nishiyama, K., Noro, M., Ohara, K., Sei, T., Takayama, N. and Takemura, A. (2011). Holonomic gradient descent and its application to the Fisher-Bingham integral. Advances in Applied Mathematics, 47:639-658.

[9] Sei, T. and Kume, A. (2015). Calculating the normalising constant of the Bingham distribution on the sphere using the holonomic gradient method, Stat. Comput., 25, 321-332.

[10] Zeilberger, D. (1990). A holonomic systems approach to special functions identities, Journal of Computational and Applied Mathematics 32, 321-368