A Procedure on Parameter Estimation in Latent Class Analysis

Doke, Hideyuki
Faculty of Engineering, Kyushu Tokai University

Asano, Choichiro
Department of Information Systems, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

https://doi.org/10.15017/17727

出版情報：九州大学大学院総合理工学報告. 10 (2), pp.223-229, 1988-09-30. 九州大学大学院総合理工学研究科
バージョン：
権利関係：

A Procedure on Parameter Estimation in Latent Class Analysis

Hideyuki DOUKE* and Chooichiro ASANO**

(Received May 25, 1988)

The present paper is to give a new procedure of parameter estimation in a latent class model by means of maximum likelihood. Since current methods of parameter estimation have occurred frequently improper solutions, the new procedure depends on the result of Gibson's method and, if the solution is not satisfied, the methods of maximum likelihood estimation are applied. In order to realize such an improved procedure, several functions are newly developed, i. e. contour map of the likelihood function, determination of direction vectors.

1. Introduction

Regarding the analysis of a latent class model firstly proposed by Lazarsfeld, several methods of numerical analysis have studied by Lazarsfeld [1], Green [2], Anderson [3], Gibson [4], Lazarsfeld and Henry [5], Goodman [6] and so on. Most of them have proposed the methods of Moment Equation or Maximum Likelihood (M. L.) based on Newton-Raphson, Fletcher-Powell [7] and recently EM algorithm and so on. However, robust method at any time does not exist in the general stiuations and there exists still an influential problem on the initial values.

The proposed procedure in this paper is firstly to apply Gibson's method based on pseudo-frequencies. Then, if the result is adequate, naturally any method does not applied. But if not, secondly the method of M. L. based on Newton-Raphson and some descent methods are applied as a non-linear programming problem. During such investigation in this paper, many examinations were done on computer simulation.

Thus it makes clear that the following procedure shows good performance.

That is to say, the recommended procedure is as follows,

(1) to obtain elements of direction vector from likelihood function,

(2) to find a direction basing on direct search method,

(3) to make a contour map of the likelihood function and find a search direction to the top and the approximate spot of the maximum.

2. A latent class model and Gibson's method

Let N individuals show two-value response, yes or no, for each of n items, let individuals contain exclusively m latent classes with the mixed rate $V^{(k)}$ for k-th latent class, where $k=1,2,\cdots,m$ and let $P_i^{(k)}$ and $P_{ij}^{(k)}$ be probabilities to respond positively for i-th item and to respond positively for both item, i-th and j-th in k-th class, respectively, where $i,j=$

*Faculty of Engineering, Kyushu Tokai University
**Department of Information Systems

$1,2,\cdots,n$ and $i \neq j$. Naturally, the following relations hold.

$$\sum_{k=1}^{m} V^{(k)} = 1,$$

$$P_i = \sum_{k=1}^{m} V^{(k)} P_i^{(k)}, \qquad for \quad i=1,2,\cdots,n, \qquad (2 \cdot 1)$$

$$P_{ij} = \sum_{k=1}^{m} V^{(k)} P_i^{(k)} P_j^{(k)}, \qquad for \quad i,j=1,2,\cdots,n, \qquad i \neq j,$$

where P_i and P_{ij} are manifest probabilities responded positively for i-th item and to respond positively for both items i-th and j-th, respectively, $i, j = 1,2, \cdots, n$ and $i \neq j$. Then the problem is to estimate the latent parameters, $V^{(k)}$, $P_i^{(k)}$ and $P_{ij}^{(k)}$ for all k, i and j. Gibson [4] has given an algebraic solution to estimate the latent parameters by matrix manipulation on $(2 \cdot 1)$. However, it is known that the method presents sometime improper solutions, although sometime it presents proper solutions, because his method tends too deterministic to fit observations with random errors.

Since a unique solution exists for the case of $m=2$ and $n=3$, let a numerical example be shown to investigate Gibson solution. First of all, $N \times 4$ uniform random numbers are generated by computer.

Thus we can obtain the frequencies N_1, N_2, $\cdots$, N_8 to each response pattern of three dichotomous items. Doing 1000 examinations of Gibson method based on these psuedo data, the result presents 536 proper solutions and 437 improper solutions.

Table 1 shows the sample means and variances of 536 proper solutions and shows Gibson's solution gives still many improper solutions.

Table 1.

True latent parameters, N=100

Class	Mixed Rate	Item 1	Item 2	Item 3
1	0.6000	0.8000	0.3000	0.8000
2	0.4000	0.2000	0.7000	0.3000

Sample means of estimated prameters

Class	Mixed Rate	Item 1	Item 2	Item 3
1	0.5793	0.8033	0.2752	0.8193
2	0.4207	0.2137	0.7252	0.2774

Sample variances of estimated prameters

Class	Mixed Rate	Item 1	Item 2	Item 3
1	0.0133	0.0074	0.0084	0.0067
2	0.0133	0.0114	0.0126	0.0141

3. M. L. method and N. R. method

Let $P_{(s)}$ be a manifest probability to show s-th response pattern and let N_s be the sample frequency for s-th response pattern, where

$\sum_{S=1}^{t} P_{(s)}=1$, $\sum_{S=1}^{t} N_s=N$, $t=2^n$ and where n is the number of dicotomous items. For the respective response patterns, N_1, N_2, $\cdots$, N_t are distributed by a multinominal distribution. When the sample values, N_1, N_2, $\cdots$, N_t are given, the M. L. E. of the latent parameters are the solution obtained by maximizing the following Log likelihood function

$$log\ L = \sum_{S=1}^{t} N_s\ log\ P_{(s)}. \qquad (3 \cdot 1)$$

On the other hand, the manifest probabilities are written as

$$P_{(s)}= \sum_{k=1}^{m} V^{(k)}\ P_{(s)}^{(k)} = \sum_{k=1}^{m-1} V^{(k)} (P_{(s)}^{(k)} - P_{(s)}^{(m)}) + P_{(s)}^{(m)}, \qquad S=1,2,\cdots, t, \qquad (3 \cdot 2)$$

and the derivatives of $P_{(s)}$ with respect to $V^{(k)}$ and $P_i^{(k)}$ are as follows,

$$\partial P_{(s)} / \partial V^{(k)} = P_{(s)}^{(k)} - P_{(s)}^{(m)} \qquad k=1,2,\cdots, m-1 \qquad (3 \cdot 3)$$

$$\partial P_{(s)} / \partial P_i^{(k)} = \begin{cases} V^{(k)} P_{(s)}^{(k)} / P_i^{(k)}, & \text{if } i \text{ is positive response,} \\ -V^{(k)} P_{(s)}^{(k)} / (1-P_i^{(k)}), & \text{if } i \text{ is negative response.} \end{cases} \qquad (3 \cdot 4)$$

Let a parameter vector $\boldsymbol{\theta}$ be newly defined by

$$\boldsymbol{\theta}' = (V^{(1)}, \cdots, V^{(m-1)}, P_1^{(1)}, P_2^{(1)}, \cdots, P_{n-1}^{(m)}, P_n^{(m)}).$$

Then the partial derivative of Log L with respect to $\boldsymbol{\theta}$ is shown as

$$\partial\, log\ L / \partial\, \boldsymbol{\theta} = \sum_{S=1}^{t} N_s / P_{(s)} \cdot \partial P_{(s)} / \partial\, \boldsymbol{\theta}. \qquad (3 \cdot 5)$$

Thus the following simultaneous equations are obtained.

$$\partial\, log\ L / \partial V^{(k)} = \sum_{S=1}^{t} N_s / P_{(s)} \cdot (P_{(s)}^{(k)} - P_{(s)}^{(m)}) = 0, \qquad (3 \cdot 6)$$

$$\partial\, log\ L / \partial P_i^{(k)} = \sum_{S_i} N_s / P_{(s)} \cdot V^{(k)} P_{(s)}^{(k)} / P_i^{(k)}$$
$$- \sum_{S\bar{i}} N_s / P_{(s)} \cdot V^{(k)} P_{(s)}^{(k)} / (1-P_i^{(k)}), \qquad k=1,2,\cdots,m-1,\ i=1,2,\cdots, n, \qquad (3 \cdot 7)$$

where S_i stands for the set of all responses which have positive response to i-th item, $S_{\bar{i}}$ stands for the set of all responses which have negative response to i-th item. The M. L. E.

is obtained by solving the simultaneous equations, however, these equations are too complex to be solved explicitly. Now, let (3 · 6) and (3 · 7) be rewritten as

$$\partial \log L / \partial \theta = 0, \tag{3 · 8}$$

and

$$\partial \log L / \partial \theta \fallingdotseq \mathbf{g}(\theta^i) + \mathbf{H}(\theta^i)(\theta - \theta^i), \tag{3 · 9}$$

where $\mathbf{g}(\theta^i)$ is the gradient vector and, $\mathbf{H}(\theta^i)$ is Hessian matrix. If $\mathbf{H}(\theta^i)$ is nonsingular, N. R. method is applied as

$$\theta^{i+1} = \theta^i - \mathbf{H}(\theta^i)^{-1}\mathbf{g}(\theta^i), \quad i=1,2, \cdots. \tag{3 · 10}$$

The examination by N. R. method is done for both cases of proper solution and improper solution given by Gibson's method (1000 trials). Table 2 shows that adopting Gibson's solution as initial value, the iteration converges after few steps, but adopting arbitorary initial value for improper solution, occurs improper solution at the first step.

In order to obtain adeguate solution for arbitorary initial value, M. L. method can be deal with as nonlinear optimization problem. It become clear that even if descent method on the optimization problem, e. g. Fletcher-Powell method [7], steepest descent method [8] are used, these solutions occur similarly the improper solution.

Table 2.

(1) Iterations by the use of Gibson's solution

Iteration	Parameters				−Log L
0	0.6915	0.7774	0.2768	0.8151	191.193
	0.3085	0.1359	0.7410	0.2799	
1	0.6924	0.7773	0.2768	0.8144	191.192
	0.3076	0.1359	0.7424	0.2799	
2	0.6924	0.7773	0.2768	0.8144	191.192
	0.3076	0.1359	0.7424	0.2799	

(2) Iterations by a given intial value

Iteration	Parameters				−Log L
0	0.5000	0.4000	0.4000	0.4000	207.710
	0.5000	0.6000	0.6000	0.6000	
1	−0.1529	0.4421	0.3200	0.4142	
	1.1529	0.2930	0.4326	0.3384	

4. The proposed procedure

According to the above examinations, it becomes clear that the N. R. method and descent method depend on how to adopt initial values. Now, the proposed optimization procedure which searches sequentialy minimum point is not so influenced by the initial values and is useful to apply the shape of likelihood function surface by a contour map of the likelihood function, where other parameters are fixed. Also, the L. F. map is made by writing number 0, 1, ⋯ , 9 which are assigned to the values of likelihood functions. Furthermore, we can show the search direction in the map from the direction vector obtained in process of the procedure. Thus let the optimization procedure be proposed in the following way.

Step 1. Give firstiy initial values in the prameter vector $\theta^{1\prime}=(\theta_1, \theta_2, \cdots, \theta_l)$, evaluate L. F. $f(\theta^1)$ and $g(\theta^1)$ and set $i=1$.
Step 2. If $\|g(\theta^i)\| < \varepsilon$, stop the procedure,
if $\|g(\theta^i)\| \geqq \varepsilon$, select θ_j, θ_k, $j \neq k$ among all parameters and go to step 3.
Step 3. Make L. F. table on θ_j, θ_k and give interval of grid 0.1.
Step 4. Calculate α_j, α_k which are the displacement along θ_j, θ_k axises from the present point θ^i to minimum point in the table.
Step 5. If L. F. table about all parameters are made, go to step 6, and otherwise select θ_j, θ_k among parameters which have not selected until now and go to step 3.
Step 6. Make $L=2^{l}$'s direction vectors which have the element of displacement obtained for all parameters as follows,

$$\mathbf{d}^{1\prime}=(\alpha_1, \alpha_2, \cdots, \alpha_l),\ \mathbf{d}^{2\prime}=(0, \alpha_2, \cdots, \alpha_l),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots,\ \mathbf{d}^{L\prime}=(0, 0, \cdots\cdots, 0).$$

Step 7. Decide derection vector $\mathbf{d}^i$ from direct search method which minimize $f(\theta^i+\mathbf{d}^k)$, $k=1, 2, \cdots, L$ and set $\theta^{i+1}=\theta^i+\mathbf{d}^i$,
if $f(\theta^{i+1}) < f(\theta^i)$, go to step 8,
if $f(\theta^{i+1}) \geqq f(\theta^i)$, stop the procedure.
Step 8. Make the L. F. map and show the direction of $\mathbf{d}^i$ in the map, set $i=i+1$ and go to 2

Table 3.

An illustrative results based on the proposed procedure

Iteration	Parameters				−Log L
0	0.5	0.7	0.7	0.7	209.52
	0.5	0.3	0.3	0.3	
1	0.5	0.8	0.4	0.8	198.89
	0.5	0.3	0.4	0.3	
2	0.6	0.8	0.4	0.8	197.50
	0.4	0.1	0.5	0.3	
3	0.6	0.8	0.3	0.8	196.63
	0.4	0.1	0.5	0.2	
4	0.6	0.8	0.3	0.8	196.39
	0.4	0.1	0.6	0.2	

Table 3 shows the process of search in the procedure, thus we can know that the solutions obtain estimates near to the true latent parameters. Figure 1 shows only L. F. maps of $P_1^{(1)}$, $P_2^{(2)}$ according to L. F. values.

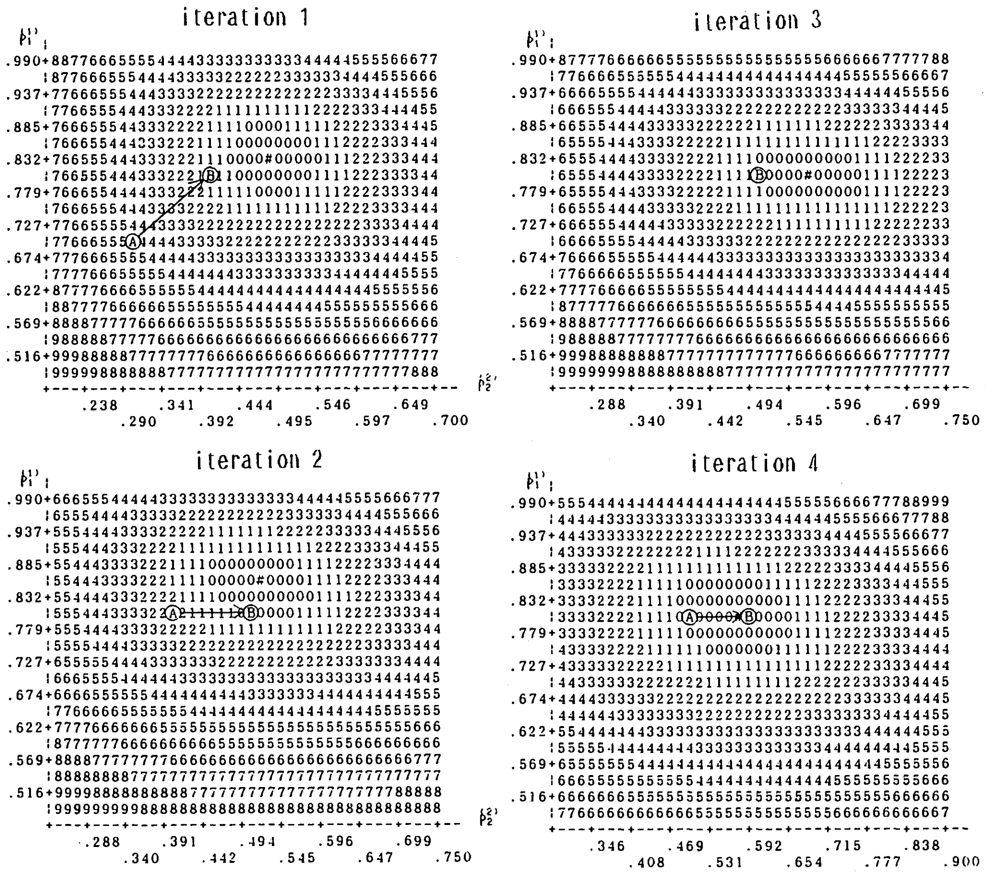

Figure 1 Contour Maps

5. Conclusion

According to the detailed investigation in the previous sections, it becames clear that the proposed procedure is superior to the single use of N. R. method or descent method. Namely, the utility and the procedure may be mentioned as follows, (1) no improper solution, (2) no linear search, (3) no gradient vector and Hessian matrix, (4) usefullness of the contour map of the likelihood function.

Since, the procedure can sequentialy do the optimization calculation without having improper solution or stopping at local maximum, the procedure is useful for obtaining M. L. E. in practice.

References

1) Lazarsfeld, P. F. The logical and mathematical function of latent structure analysis, chapter 10, Measurement and Prediction, S. A. Stouffer et al Princetion University Press, 1950.

2) Green, B. F. A general solution for the latent class model of latent structure analysis, Psychometrika vol. **16**, No. 2 (1951).

3) Anderson, T. W. On estimation of parameters in latent structure, Psychometrika vol. **19**, No. 1 (1954).

4) Gibson, W. A. An extension of Andersons solution for the latent structure equations, Psychometrika, vol. **20**, 1955.

5) Lazarsfeld, P. F. & Henry, N. W. Latent Structure analysis, Boston, Houghton Mifflin, 1968.

6) Goodman, L. A. Exploratory latent structure analysis using both identifiable and unidentifiable models, Biometrika, **62**, 2 (1974).

7) Fletcher, R & Powell, M. J. D. A rapidly convergent descent method for minimization, Comp. J. 6, 1963.

8) Householder, A. S. Principles of Numerical Analysis, New York, Mcgraw-Hill, 1953.