GRAPHICAL CHAIN MODELS

By

Wan Fauzi Wan Mamat

ABSTRACT

This article is to show the GLIM (Statistical package) approach in use, as viewed by the working statistician. One modern approach is to consider initially only those models which can be understood purely in terms of conditional independence relationship. This is made attractive through the simple graphical characterisation of such models given by Darroch, Lauritzen and Speed (1980). Graphical models are easily interpreted in terms of conditional independence relation.

1.0 INTRODUCTION

We illustrate our introduction to a modern approach for models selection in multiway contingency tables by considering some well discussed data on Convention Delegates and Exhibitors survey in Malaysia. In this survey we have chosen three variables such as Category of participant (PART), Visitors (VIS) and Age (AGE) of the respondents. We used graphical chain model to fit the best model for these variables.

Example of the 3-way contingency table.

Consider 3 variables, Category of participants (Question 3 - PART) By Visitors (Question 6a - VIS) By AGE (Question 13a -AGE).

PART consists of five categories -

1. Delegates
2. Participant in Exhibition
3. Exhibitors
4. Observers
5. Others

VIS consists of two categories-

1. International Participants
2. Domestic participants

AGE consists of six categories -

1. 18 - 24 yrs
2. 25 - 34 yrs
3. 35 - 44 yrs
4. 45 - 54 yrs
5. 55 - 64 yrs
6. 65 yrs and over

We can think of this as a 3 - way contingency table: PART * VIS * AGE

2.0 FITTING GRAPHICAL CHAIN MODELS

Follow the step given above, firstly, we analyse the Left Hand Block of Explanatory Variates i.e. PART and VIS. Consider for this case a Two-way Table and select a well fitting model for these variables. Treating the next block i.e. AGE as a block of 'RESPONSE VARIATE' explained by the Left Hand Block and find a well fitting model. Note that this fit must always 'FIX THE MARGINS' in the Left Hand Block, i.e. all fits must contain PART *VIS (so the analysis is equivalent to Logistic Regression).

Explanatory Variates are variables which can explain the next block and Response variates are variables which is explained by the left hand block of Explanatory variates.

Possible Chain Models

1. VIS o o AGE — All 3 Independent
   PART o

2. VIS o o AGE — PART, VIS related but do not explain variation in AGE of the respondents.
   PART o

3. VIS o → o AGE — VIS explain some variation in AGE of the respondents. PART unrelated
   PART o

4. VIS o — PART explain some variation in AGE of the respondents.
   PART o → o AGE — VIS unrelated.

5. VIS o → o AGE — VIS, PART related. VIS explain AGE. or AGE is independent
   PART o — of PART given VIS

6. VIS o — VIS, PART related. PART explains AGE. OR
   PART o → o AGE — VIS is independent of AGE given PART.

7. VIS o — VIS, PART related. Both explain some variability in AGE.
   PART o → o AGE — To predict AGE, we have to know both VIS and PART.

3.0 DISCUSSION

1. In the Graphical Chain Model, we have a 3-way contingency table with 2 blocks. The Left HandBlock consists of Two Explanatory Variates (PART and VIS)

   Analyse the Explanatory Variates by a 2-way Table and fit the best model between the two variables. From here, we get the value of the residual deviance of the best fitting model and deviance of the independence model.

   Using GLIM 4 (Refer APPENDIX 1)

   $c Analysis of VIS2; PART2 2-way table.

   $fit PART2 * VIS2 : PART2 + VIS2 $

   From the GLIM output, we get the results as follows:-

   The deviance (Do) of PART2 + VIS2 is 21.420 on 4 df (degree of freedom), which is significant. The only possible model is PART2 * VIS2, with zero deviance (D1) and zero df. PART2 and VIS2 are the category of participants and visitors of the respondent when we analyse by the 2-way Table respectively.

   Example of Calculation

   Ho : PART2 + VIS2 =======> Do = 21.42, dfo = 4

   Hi : PART2 * VIS2 ========> D1 = 0, df1 =0

   C=Do - D1 =21.42 ============> C = calculated Chi-squared

   D= dfo - df1 = 4

   Since C > Chi-squared table

   Therefore we accept Hi, Model (PART2 * VIS2)

2. Return to the 3-way Table, treating AGE as a RESPONSE. Eliminate (margin fixed) PART * VIS, fitting the models and select the best model by forward selection, comparing models with Chi-squared tests.

AGE is greatly improved by AGE * PART or AGE * VIS.

Biggest improvement is AGE * PART (difference in scale is 58.0 with 20 df).

So we arrived at AGE * (PART + VIS).

The residual (scaled deviance) of this 'best model' is 30.57 with 20 df.

The total deviance for the independence of AGE from PART and VIS (obtained from fit of AGE alone) is 102.53 with 45 df.

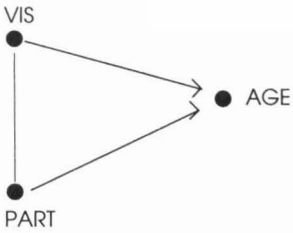
3. The (scaled) deviance for the independence of all 3 variates obtained from $fit PART + VIS + AGE is 123.95 on 49 df

4. Analysis of deviance table obtained from GLIM 4 output:- (Refer APPENDIX 1.)

Model	Residual Dev	df	Total Dev	Independence df	explained
PART * VIS	0.0	0	21.42	4	PART+VIS in PART*VIS
PART*VIS+AGE (PART+VIS)	30.57	20	102.53	45	AGE from PART*VIS
TOTAL	30.57	20	123.95	49	AGE+PART+VIS in AGE*PART* VIS

5. The Chain Model explained (123.95 - 30.57)/123.95 = 75.34% of the total variation.

Therefore the best model is AGE* (PART + VIS)



VIS and PART are related. Both explain some variability in AGE of the respondents. In other words, to predict age of the respondents (AGE), we have to know both the category of participants (PART) and the type of visitors (VIS).

REFERENCES

Aitkin, M.,Anderson, D.,Francis, B. and Hinde, J. 1989. *Statistical Modelling in GLIM.* Oxford : Clarendon Press.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. 1980. *Markov Fields and Log Linear Interaction Models for Contingency Tables.*

Edwards, D. and Kreiner, S. 1983. *The Analysis of Contingency Tables by Graphical Models.*

Lindsey, J.K. .undated. *The Analysis of Categorical Data Using GLIM.* Springer Verlag Lecture Notes in Statistics, 56. New York: Springer Verlag.

Robert Gilchrist. 1990. *Multiway Contingency Tables : Model Selection Using Graphical Models.*UNL

The Analysis of Survey Data. Model Fitting. Vol 2. London : Wiley.

Whittaker, J. 1990 *Graphical Models,* London : Wiley