ON NON-HIERARCHICAL LOG-LINEAR MODELS AND THE ITERATIVE PROPORTIONAL FITTING ALGORITHM[1]

Jay Magidson, Abt Associates Inc.

Abstract

Recent literature on log-linear models gives the impression that the Interative Proportional Fitting (IPF) algorithm yields maximum likelihood estimates only for hierarchical (not non-hierarchical) models. While it is true that hierarchical models are often more easily interpreted than non-hierarchical models, it is shown here that the IPF algorithm (and existing computer programs designed for hierarchical models) can be used to estimate any non-hierarchical model. This follows directly from the symmetry between qualitative/categorical indicator variables and appropriately defined "interaction variables." The general approach is illustrated here using data from the study of "The American Soldier," Stouffer et al. (1949). We also illustrate how a qualitative analogue to the R^2 in quantitative regression analysis can be used to partition "qualitative variance" in 14 logit models.

Introduction

In recent years, new statistical methods involving log-linear models have become available for analyzing the relationships among qualitative/categorical variables. The approaches recommended by Goodman (1970), Bishop (1969), Grizzle, Starmer and Koch (1969), Ku and Kullback (1974) and others differ in certain respects but they all formulate the same multiplicative analogue to the additive analysis of variance (ANOVA) model. The development of log-linear models has led to major advances in the statistical analysis of qualitative data.

Some of the recent literature in this area conveys the impression that in order to estimate non-hierarchical log-linear models (i.e., models which hypothesize some higher order interaction terms but which exclude certain lower order terms), one must use some algorithm other than the Iterative Proportional Fitting (IPF) algorithm. The Deming-Stephan (1940) IPF algorithm is recommended by Bishop (1969) and Goodman (1970) for estimating hierarchical models. The purpose of this paper is to point out that the IPF algorithm can also be used to estimate non-hierarchical models. In this paper we illustrate the general estimation approach and also recommend the use of a qualitative analogue to R^2.

For concreteness, we use the data from the study of "The American Soldier" by Stouffer et al. (1949) to motivate the discussion and illustrate the approach. We show how non-hierarchical models which can be transformed into hierarchical models as well as non-hierarchical models which cannot be transformed into hierarchical models can all be estimated using Goodman's ECTA computer program, a program designed to estimate hierarchical models using the IPF algorithm.

We will use data from 8,036 soldiers to predict (D) Camp Preference (North or South) based on knowledge of the three explanatory variables (A) Race (Black or White), (B) Region of Origin (North or South) and (C) Present Location (North or South). Although this paper is limited to a subset of log-linear models (the logit model) and involves only four variables which are all dichotomous, the logic presented here can easily be generalized to log-linear models other than logit models involving any number of polytomous (not necessarily dichotomous) variables (see Magidson, 1976).

We also present a comparison of the likelihood ratio chi-square, the goodness of fit chi-square and the correlation ratio for each of the logit models fitted to the data. The two chi-square statistics indicate how well a model fits the data while the correlation ratio η^2 measures the proportion of variance explained. Beginning with a model of complete independence where $\eta^2_{D.ABC} = 0$, the correlation ratio steadily increases to .35 for the saturated model while the corresponding chi-square values steadily decline indicating that the models which explain the most variance also fit the data best. This gives empirical support to the meaningfulness of η^2, a qualitative analogue to R^2 which is seldom reported for logit models.

Preliminary Analysis of the Data

Table 1 displays the data in the form of a 2-way table where the rows are associated with the explanatory variables Race, Region of Origin and Present Location and the columns refer to the levels (categories) of the dependent variable Camp Preference. The conditional proportions and conditional odds in favor of a northern (and a southern) camp preference are also given in this table.

Thus, for example we see that 91.5% of the 423 black northerners in northern camps prefer a northern camp while only 9.5% of the 960 white southerners in southern camps prefer a northern camp. An equivalent way of looking at these figures is in terms of the odds in favor of a northern camp. For black northerners in northern camps the odds are 387:36 (or 10.75:1) in favor of a northern camp preference while the corresponding odds for white southerners in southern camps is 91:869 (or 0.105:1).

The single best predictor of Camp Preference is (B) Region of Origin. This can be seen directly from Table 1 by noting that a higher proportion of northern-born soldiers prefers the north than southern-born soldiers in every case regardless of Race and Present Location (i.e., even the group of northern-born soldiers least likely to prefer the north, is more likely to prefer the north than any group of southern-born soldiers). Similarly, it is seen that the next best predictor is (C) Present Location while the weakest predictor is (A) Race.

The Saturated and Unsaturated Logit Models

The saturated logit model for predicting Camp Preference as a function of (A) Race, (B) Region of Origin and (C) Present Location is

$$\log \Omega = \beta + \beta^{A}X^{A} + \beta^{B}X^{B} + \beta^{C}X^{C} + \beta^{AB}X^{AB} + \beta^{AC}X^{AC} + \beta^{BC}X^{BC} + \beta^{ABC}X^{ABC} \quad (1)$$

where Ω denotes the expected value of the conditional odds in favor of a northern camp preference and the X's are indicators associated with the explanatory variables. The X's are defined in Table 2. We will refer to X^A, X^B, and X^C as "main variables" and to the other X's as "interaction variables." We will also refer to the X's as vectors as displayed in Table 2.

Model 1 is a saturated or full rank model because the X-vectors (together with a vector of ones) form a basis for the entire 8-dimensional space. Thus, improved prediction is not possible by including additional variables into the model because any additional variables can be expressed as linear combinations of the X's and absorbed into model 1. The basis vectors are displayed in the form of a design matrix in Table 2. It can easily be verified that the basis is orthogonal (although the estimated β-parameters will not be orthogonal).

Unsaturated or restricted models can be formed from model 1 by omitting some of the X's (i.e., setting some of the β's to zero). Each unsaturated model therefore corresponds to a hypothesis that the vector of expected odds of preferring the north is located in the subspace spanned by the X-vectors included in the model. For example, the main-effects-only model hypothesizes that the vector of odds is located in the subspace spanned by X^A, X^B and X^C (and the constant vector). We will now distinguish between hierarchical hypotheses (models) and non-hierarchical hypotheses (models).

A model including one or more interaction vectors is said to be hierarchical if all lower order X-variables having the same superscripts are also included in the model. Thus, the model including X^{BC} is hierarchical if it also includes X^B and X^C, otherwise it is non-hierarchical. It follows that the saturated model is the only hierarchical model containing the X^{ABC} vector. A model which excludes all interaction vectors is also said to be hierarchical. Thus, the main-effects-only model is hierarchical, the model which includes only X^A (and the constant) is hierarchical and the total independence model which omits all of the X's is also hierarchical.

Any model which is not hierarchical is said to be non-hierarchical. Thus, for example, the model which omits all X-vectors except for X^A, X^B and X^{BC} is non-hierarchical because it excludes X^C.

Results from 14 Logit Models

Table 3 summarizes the results for 13 unsaturated logit models (and the saturated model). Models H_1-H_{10} are hierarchical models estimated earlier by Goodman (1972a) using the ECTA computer program. Models H_{11}-H_{13} are non-hierarchical models also estimated using ECTA. Model H_0 is the saturated model.

The main-effects-only model is designated as model H_2. There are 4 degrees of freedom associated with this model corresponding to the 4 interaction terms omitted. The large chi-square value is significant at well beyond the .01 level so we reject the main-effects-only model in favor of a model postulating interaction.

Model H_1 fits the data exceptionally well as indicated by a chi-square value of only 1.5 with 3 degrees of freedom. This parsimonious model hypothesizes only one interaction term, the (BC) Region of Origin/Present Location term. This model is accepted by Goodman (1972a) for this data. It states that black soldiers are about 2.1 times more likely to prefer the north than white soldiers having the same region of origin and the same present location. (Since there are no interaction terms associated with Race in model H_1, this number is constant over the four joint categories of Region of Origin/ Present Location.)[2]

Table 4 compares the estimates of the parameters in model H_1 with those of the saturated model. The estimates are almost identical to two decimal places. Notice that the estimated parameters associated with the interaction variables are smaller in magnitude than those associated with the main variables. Also notice that these estimates are consistent with our preliminary analysis which concluded that (B) Region of Origin was the most important predictor, (C) Present Location was next in importance while (A) Race was the least important explanatory variable for the prediction of (D) Camp Preference as indicated by the correlation ratio. We discuss these correlation ratios in more detail in a later section.

The statistical significance of the BC term in model H_1 can be tested by subtracting the likelihood ratio chi-square for model H_1 from the likelihood ratio chi-square for the main-effects-only model H_2. This difference is asymptotically distributed as a chi-square statistic with one degree of freedom under the null hypothesis that the main-effects-only model is correct (i.e., the null hypothesis is that $\beta^{BC} = 0$ in model H_1). The number of degrees of freedom is the difference in degrees of freedom between these two models. This difference (24.96 - 1.45 = 23.51) is highly significant so we reject the null hypothesis (model H_2) and accept model H_1.

The significance of A in model H_1 can be similarly tested by subtracting the chi-square value for model H_1 from the chi-square value for the hierarchical model H_3. Similarly, the significance of B and C can be tested using the non-hierarchical models H_{11} and H_{12} respectively. All parameter estimates in model H_1 are statistically significant at well beyond the .01 level.

Model H_{13} is similar to model H_1. The only difference is that it includes the highest order interaction term ABC instead of the BC term. Model H_1 fits the data exceptionally well but model H_{13} does not fit well at all. In the next section we show that model H_{13} cannot be transformed into a hierarchical model by simple transformations while the other non-hierarchical models H_{11} and H_{12} can be so transformed. We also show how these three non-hierarchical models were all estimated using the ECTA computer program.

The proportion of variance explained by these 14 logit models is given in the rightmost column of Table 3. The correlation ratios are discussed in a later section.

The general approach is to convert any model to a main-effects-only model by viewing all variables as main variables whether they are in fact main variables or interaction variables. This will generally involve inputting a larger number of variables into ECTA than is really the case and some (or many) of the frequencies will be structural zeros.

For purposes of illustration, let us first consider the 4 models H_1, H_2, H_{11} and H_{12}. These models include only 4 of the X-variables in their formulation. They include the dependent variable D, and the X-variables, X^A, X^B, X^C and X^{BC}. The coercion approach to estimating these models is to input a 5-way table of frequencies rather than a 4-way table despite the fact that there are really only the four dichotomous variables A, B, C and D. Table 10 displays the 32 frequencies input for these models, 16 of which are structural zeros.[3]

Table 11 gives the marginal tables which are fit for each of these models based on the inputted frequencies of Table 10. The $\{\overline{BC}D\}$ table is the 2x2 table which crossclassifies the D dichotomy with the X^{BC} dichotomy. It is different from the 2x2x2 {BCD} table which crossclassifies the three dichotomies B, C and D.

Model H_1 can be estimated based on the inputted frequencies given in Table 10 by specifying that the {B } 3-way table be fit instead of specifying that the three 2-way tables {BD} , { CD} , $\{\overline{BC}D\}$ be fit. The fact that these alternative specifications are equivalent is shown in Magidson (1976).

Model H_{13} can be estimated in a similar fashion by inputting the 32 frequencies given in Table 12. Or all 5 unsaturated models can be estimated from a single set of frequencies if the 64 frequencies (with 48 structural zeros) corresponding to the 6-way table formed by the X-variables X^A, X^B, X^C, X^{BC}, X^{ABC} and D are input. Taking this logic to the extreme, any model can be estimated based on an 8-way table which also includes the X^{AB} and X^{AC} terms.

Thus, we have shown how any non-hierarchical model can be estimated using ECTA, a program designed for hierarchical models. For occasional estimation of non-hierarchical models, the ECTA program should suffice. For extensive estimation of non-hierarchical models, ECTA can easily be modified to include an option so that one need not input any structural zeros. In any case, it is the IPF algorithm which can be used to calculate ML estimates for the expected frequencies under any hierarchical or non-hierarchical model of the kind usually considered.[4]

References

Birch, M.W., "Maximum Likelihood in Three-Way Contingency Tables," Journal of the Royal Statistical Society, Series B, 25 (1963), 220-233.

Bishop, Y.M.M., "Full Contingency Tables, Logits and Split Contingency Tables," Biometrics, 25 (1969), 383-400.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W., Discrete Multivariate Analysis: Theory and Practice, Cambridge, Mass.: MIT Press, 1975.

Bloomfield, P., "Linear Transformation for Multivariate Binary Data," Biometrics, 30 (1974), 609-617.

Deming, W.E. and Stephan, F.F., "On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Tables Are Known," Annals of Mathematical Statistics, 11 (1940), 427-444.

Evers, M. and Namboodiri, N.K., "Weighted Least Squares Versus Maximum Likelihood Estimates of Non-Hierarchical Models," paper presented at the annual meeting of the American Statistical Association, August 1976.

Gini, C.W., "Variability and Mutability," contribution to the study of statistical distributions and relations, Studi Economico - Guiridici della R. Universita de Cagliari, 1912.

Goodman, L.A., "The Multivariate Analysis of Qualitative Data: Interactions Among Multiple Classification," Journal of the American Statistical Association, 65 (1970), 225-256.

Goodman, L.A., "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications," Technometrics, 13 (1971), 33-61.

Goodman, L.A., "A Modified Multiple Regression Approach to the Analysis of Dichotomous Variables," American Sociological Review, 37 (1972a), 28-46.

Goodman, L.A., "A General Model for the Analysis of Surveys," American Journal of Sociology, 77 (1972b), 1035-1086.

Goodman, L.A., "Causal Analysis from Panel Studies and Other Kinds of Surveys," American Journal of Sociology, 78, No. 5 (1973), 1135-1191.

Goodman, L.A., "The Relationship Between Modified and the More Usual Multiple Regression Approach to the Analysis of Dichotomous Variables," Sociological Methodology, 1974.

Goodman, L.A. and Kruskal, W.H., "Measures of Association for Cross Classifications," Journal of the American Statistical Association, 49 (1954), 732-764.

Grizzle, J.E., Starmer, C.F. and Koch, G.G., "Analysis of Categorical Data by Linear Models," Biometrics, 25 (1969), 489-504.

Haberman, S., book review of Bishop, Feinberg and Holland, Discrete Multivariate Analysis: Theory and Practice in Annals of Statistics, 4, No. 4 (1976), 817-820.

Ku, H.H. and Kullback, S., "Loglinear Models in Contingency Table Analysis," The American Statistician, 28, No. 4 (1974), 115-122.

Light, R.J. and Margolin, B.H., "An Analysis of Variance for Categorical Data," Journal of the American Statistical Association, 66 (1971), 534-544.

Magidson, J., "Qualitative Variables and Simultaneous Equation Econometric Models," Discussion Paper No. 142, Center for Mathematical Studies, Northwestern University, Evanston, Ill., 1975.

Magidson, J., "The Multivariate Analysis of Qualitative Variance: Analyzing the Probability of an Event as a Function of Observable Variables," unpublished doctoral thesis, Northwestern University, June 1976.

Nelder, J.A., "Hypothesis Testing in Linear Models (letters to the editor)," The American Statistician, 30, No. 2 (1976), 103.

Stouffer, S.A., Suchman, E.A., Devinney, L.C., Star, S.A. and Williams, R.M., Jr., The American Soldier: Adjustment During Army Life: Studies in Social Psychology in World War II, Vol. 1, Princeton, N.J.: Princeton University Press, 1949.

Swan, J., Magidson, J. and Berk, R., "Estimation of Non-Hierarchical Log-Linear Models by Circulation, Subversion and Coercion," 1976.

Footnotes

[1]This is an abbreviated version of the original paper, prepared especially for these proceedings. Copies of the complete paper are available upon request from the author at Abt Associates Inc., 55 Wheeler Street, Cambridge, Massachusetts 02138.

[2]The number 2.1 is twice the estimate of β^A expressed in units of odds. By taking the logarithm of 2.1 we convert back to the logit formulation where the parameters are expressed in logarithms of odds. (See Goodman, 1972b.)

[3]ECTA has an option which allows the user to specify which frequencies correspond to structural zeros.

[4]We can conceive of models formed by other kinds of restrictions of course, but these other models are beyond the scope of this paper.

Table 1 Cross-classification of Soldiers with Respect to 4 Dichotomized Variables: (A) Race, (B) Region of Origin, (C) Location of Present Camp, and (D) Preference as to Camp Location

(B) Region of Origin	(C) Location of Present Camp	(A) Race	Number of Soldiers Preferring Camp: In North Freq.	In North Prob.	In North Odds	In South Freq.	In South Prob.	In South Odds	Total
North	North	Black	387	.915	10.750	36	.085	0.093	423
North	North	White	955	.855	5.895	162	.145	0.170	1117
North	South	Black	876	.778	3.504	250	.222	0.285	1126
North	South	White	874	.632	1.714	510	.368	0.584	1384
South	North	Black	383	.587	1.419	270	.413	0.705	653
South	North	White	104	.371	0.591	176	.629	1.692	280
South	South	Black	381	.182	0.223	1712	.818	4.493	2093
South	South	White	91	.095	0.105	869	.905	9.549	960
			4051	.504	1.017	3985	.496	0.984	8036

Table 3 The Results from Fourteen Logit Models for the Prediction of (D) Location Preference Based On the Explanatory Variables (A) Race, (B) Region of Origin and (C) Present Location

Model	Explanatory Variables Included in the Model	Degrees of Freedom	Likelihood Ratio Chi-Square	Goodness of Fit Chi-Square	Proportion of Variance Explained $\hat{\eta}_{D.ABC}$
H_0	ALL (A,B,C,AB,AC,BC,ABC)	0	0	0	.350
H_9	A,B,C,AB,BC	2	0.68	0.69	.349
H_8	A,B,C,AC,BC	2	1.32	1.34	.349
H_1	A,B,C,BC	3	1.45	1.46	.349
H_{10}	A,B,C,AB,AC	2	17.29	18.73	.347
H_{13}	A,B,C,ABC	3	24.80	25.48	.345
H_2	A,B,C	4	24.96	25.73	.345
H_3	B,C,BC	4	152.65	147.59	.336
H_4	B,C	5	186.36	180.26	.329
H_{12}	A,B,BC	4	674.78	675.74	.285
H_6	A,B	5	695.01	727.16	.282
H_{11}	A,C,BC	4	1604.57	1905.35	.176
H_5	A,C	5	2286.83	2187.71	.099
H_7	NONE	7	3111.47	2812.64	0

Table 2 The Orthogonal Basis Vectors for the Saturated Logit Model

i j k	constant	X^A	X^B	X^C	X^{AB}	X^{AC}	X^{BC}	X^{ABC}
1 1 1	+1	+1	+1	+1	+1	+1	+1	+1
1 1 2	+1	+1	+1	-1	+1	-1	-1	-1
1 2 1	+1	+1	-1	+1	-1	+1	-1	-1
1 2 2	+1	+1	-1	-1	-1	-1	+1	+1
2 1 1	+1	-1	+1	+1	-1	-1	+1	-1
2 1 2	+1	-1	+1	-1	-1	+1	-1	+1
2 2 1	+1	-1	-1	+1	+1	-1	-1	+1
2 2 2	+1	-1	-1	-1	+1	+1	+1	-1

Table 10 The Frequencies Input to ECTA to Estimate Models H_1, H_2, H_{11} and H_{12} by the Coercion Approach

X^A	X^B	X^C	X^{BC}	North	South
1	1	1	1	387	36
1	1	1	-1	0	0
1	1	-1	1	0	0
1	1	-1	-1	876	250
1	-1	1	1	0	0
1	-1	1	-1	383	270
1	-1	-1	1	381	1,712
1	-1	-1	-1	0	0
-1	1	1	1	955	162
-1	1	1	-1	0	0
-1	1	-1	1	0	0
-1	1	-1	-1	874	510
-1	-1	1	1	0	0
-1	-1	-1	-1	104	176
-1	-1	-1	1	91	869
-1	-1	1	-1	0	0

Table 11 Four Logit Models and the Marginal Tables Fit in Order to Estimate These Models by the Coercion Approach Using the Input Data from Table 10

Model	Explanatory Variables Included	Marginal Tables Fit
H_2	A,B,C	{ABC},{AD},{BD},{CD}
H_1	A,B,C,BC	{ABC},{AD},{BD},{CD},$\{\overline{BC}D\}$ or {ABC},{AD},{BCD}
H_{11}	A,C,BC	{ABC},{AD},{CD},$\{\overline{BC}D\}$
H_{12}	A,B,BC	{ABC},{AD},{BD},$\{\overline{BC}D\}$

Table 12 The Frequencies Input to ECTA to Estimate Model H_{13} by the Coercion Approach

X^A	X^B	X^C	X^{ABC}	North	South
1	1	1	1	387	36
1	1	1	-1	0	0
1	1	-1	1	0	0
1	1	-1	-1	876	250
1	-1	1	1	0	0
1	-1	1	-1	383	270
1	-1	-1	1	381	1,712
1	-1	-1	-1	0	0
-1	1	1	1	0	0
-1	1	1	-1	955	162
-1	1	-1	1	874	510
-1	1	-1	-1	0	0
-1	-1	1	1	104	176
-1	-1	-1	-1	0	0
-1	-1	-1	1	0	0
-1	-1	1	-1	91	869