AD-A089 411 MINNESOTA UNIV ST PAUL DEPT OF APPLIED STATISTICS F/G 12/1

ALTERNATIVE COMPUTATIONAL METHODS FOR ESTIMATION IN MULTINOMIAL--ETC(U)

NOV 79 S E FIENBERG, M M MEYER, G W STEWART N00014-78-C-0600

UNCLASSIFIED TR-348-REV NL

1 OF 1

AD A089411

END

DATE FILMED

10-80

DTIC

ALTERNATIVE COMPUTATIONAL METHODS FOR ESTIMATION
IN MULTINOMIAL LOGIT RESPONSE MODELS

by

Stephen E. Fienberg and Michael M. Meyer
University of Minnesota
and
G.W. Stewart
University of Maryland

Technical Report No. 348

June 1979
Revised November 1979

SUMMARY

Several algorithms have been proposed for the computation of maximum likelihood estimates for contingency tables. Since multinomial logit response models can be treated as special versions of loglinear models, many of these techniques can be used for logit models as well. In this paper we compare, in a qualitative fashion, the relative merits of (i) two variants of Newton's method developed by Fienberg and Stewart (ii) GLIM, as developed by Nelder and Wedderburn (iii) the BMDP program for stepwise logistic regression, and (iv) the widely used method of iterative proportional fitting.

1. INTRODUCTION

The analysis of cross-classified categorical data involves statistical problems where both the explanatory variables (or factors) and response variables are categorical. Loglinear, logit and multinomial logit response models are often recommended for the analysis of such data (e.g. see Bishop, Fienberg and Holland, 1975; Bock, 1975, Fienberg, 1977, Haberman, 1974, 1978). Since logit models are also loglinear models, the computational methods associated with fitting loglinear models can also be used for logit models; however, it is often more efficient to use techniques which are specially designed for analyses using logit or multinomial logit response models.

In this paper we compare four different computational approaches for maximum likelihood estimation of parameters and expected cell values in logit models;

(a) a variant on Newtons's method, as developed by Fienberg and Stewart (1979), applied in somewhat different forms for the loglinear and logit formulations,

(b) iteratively reweighted least squares, as implemented in GLIM (see Nelder and Wedderburn, 1972),

(c) Newton's method, for the logit model, as used in BMDPLR (see Jennrich and Moore, 1975),

(d) iterative proportional fitting (IPF).

Additional comparisons could be made with the Newton-Raphson formulation of Bock (1975) or Haberman (1978), but are not included here.

2. LOGLINEAR, LOGIT AND MULTINOMIAL LOGIT RESPONSE MODELS

Consider a problem involving a response variable with K categories and two explanatory variables, with I and J categories respectively.

The data are counts in the form of an $I \times J \times K$ table where the totals in the $I \times J$ margin, corresponding to the explanatory variables, are taken as fixed. We assume the sampling model for the counts is product-multinomial (Bishop, Fienberg and Holland, 1975). Multinomial logit response models, involving $K-1$ simultaneous logit equations, are equivalent to loglinear models that treat the three variables as responses but include in the model all terms corresponding to main effects and interactions for the explanatory variables (Fienberg, 1977, Chapter 6). The special case of $K=2$ yields the familiar logit model. By using this correspondence, any algorithm for fitting loglinear models can be used for the logit problem. The disadvantage to such an approach is the added number of (unnecessary) parameters in the model. In the following sections we discuss some of the properties of various algorithms and their implementation. The discussion is briefly summarized in Table 1.

2.1 NEWTON'S METHOD

Fienberg and Stewart (1979) have used a variant of Newton's method for analyzing both loglinear and multinomial logit response models. Their first algorithm treats the logit problem as a loglinear model. The estimated covariance matrix for the parameters is adjusted for the required conditioning at the end of the computations using formulae available from Haberman (1974). Their second algorithm proceeds by initially conditioning on the explanatory variables, and then using a somewhat different, but closely related, set of computations. The choice between the algorithms depends upon the number of categories and structure of the response variables. The second approach should be more efficient when K is large or there are many response variables. In other words, if the multinomial logit response model has considerably fewer parameters than its loglinear equivalent, then one should initially condition on the explanatory variables.

Both of these algorithms involve the construction of the upper half of a $p \times p$ weighted cross product matrix, where p is the dimension of the design manifold. Note that p is not the same for the two approaches. In both algorithms a sparse $n \times p$ design matrix (where n corresponds to the number of cells in the cross-classification) is generated internally, but is never actually stored during the computations. The calculations proceed via Newton's method with variable step length, using a Cholesky decomposition with pivoting. Extrinsic aliasing (i.e., non-identifiability of certain parameters) is detected by small pivot elements during the decomposition. As the internal parameterization is not readily interpreted, u-terms or other desired parameterizations are estimated by direct computations on the table of fitted values. In this way it is not necessary to address the statistical consideration of marginality (see discussions in Nelder, 1976 and 1977, and in Fienberg, 1977) until the end of the computations.

2.2 GLIM

The GLIM algorithm, as developed by Nelder and his colleagues, is designed for analyzing Generalised Linear Models (see Nelder and Wedderburn, 1972). Logit, multinomial logit response, and loglinear models are all encompassed in the family of generalised linear models, however only logit and loglinear models are easily fitted in GLIM. In order to fit a $K > 2$ level response variable in GLIM, the user can either fit the associated loglinear model or treat the problem in an asymmetric fashion, e.g., by the use of continuation ratios (see Fienberg, 1977, Chapter 6). It is also possible to fit the multinomial logit response model directly by using GLIM's macro and user-defined model capabilities. This approach requires considerable storage and numerical sophistication on the part of the user. For many problems fitting the loglinear model is perfectly satisfactory, but for some problems the number of parameters becomes too large for the program to handle. The asymptotic covariance matrix is not adjusted for the appropriate conditioning in this approach for multinomial logits.

GLIM uses Gaussian elimination to sweep out the rows of a weighted cross-product matrix. In order to preserve the marginality constraints (i.e., due to the restriction to hierarchial models) there is no pivoting during the Gaussian elimination. As indicated in the previous sub-section, marginality considerations can be satisfactorily addressed at the end of the computational problem, and the good numerical properites of pivoting could have been utilized during the Gaussian elimination. GLIM does not do this. Extrinsic aliasing of parameters is detected when a diagonal element of the weighted cross-product matrix drops by more than 10^{-6} in two successive iterations. This procedure may encounter numerical instabilities, particularly when the program is used for logistic regression problems.

2.3 BMDP

There are at least two methods of fitting logit models in BMDP. If the response is binary the stepwise logistic regression program BMDPLR may be used. Otherwise the iterative proportional fitting (IPF) algorithm for loglinear models, in BMDP3F, is a possibility. We will confine our comments here to the logistic regression program, defering a discussion of IPF to the next section.

The algorithm of BMDPLR is a specialisation of the BMDP nonlinear regression program, and corresponds to iteratively reweighted least squares. The description by Jennvich and Moore (1975) indicates that the algorithm proceeds by explicitly inverting the weighted cross-product matrix, using the method of Gauss-Jordam elimination. Aliasing is then determined by small diagonal pivots; however, it appears that marginality may be violated here. As with GLIM it is possible to fit user defined models with BMDP. In our multinomial logit response case, this would involve using BMDP3R, the nonlinear regression program, together with some Fortran subroutines. Again, considerable expertise on the part of the user is required.

Although it is not our direct concern here, we comment briefly on the stepwise aspects of this algorithm. The BMDPLR program is the only one we reviewed that contains an automated selection procedure. However, such a structure could easily be implemented in the other algorithms, particularly GLIM, which has a macro facility. Two stepping procedures for BMDPLR are outlined in Dixon and Brown (1979). One is based on the likelihood ratio and uses standard asymptotic results for its justification. The other is based on an F-statistic approximation, using an estimate of the asymptotic covariance matrix, which may be of practical use but seems to lack a theoretical basis. The stepping procedures do preserve marginality, although this option may be overridden.

2.4 ITERATIVE PROPORTIONAL FITTING

The Iterative Proportional Fitting algorithm (IPF) is especially useful for fitting loglinear models because of the elementary nature of the iterations. The algorithm does not directly fit logit or multinomial logit response models; these must be fitted in their equivalent loglinear forms. This is not as severe a restriction as it seems, since the major storage requirement for IPF is the table itself. Extra model parameters cause only a very slight increase in storage needs. Of course, the major disadvantages of IPF are (a) its sometimes very slow rate of convergence and (b) the lack of covariance estimates for the fitted values. One advantage of IPF is that when maximum likelihood estimates for the fitted values have a closed form (i.e., for decomposable models), a version of IPF will obtain these in one iteration (see Bishop, Fienberg and Holland, 1975, and Haberman, 1974, p. 191). This is not true for any of the Newton algorithms.

3. COMPARISONS

Some of the qualitative aspects of the algorithms in question are summarized in Table 1. The GLIM and BMDP programs have similar properties. As GLIM is an interactive program, it is easier to use but pays the price by being able to analyze only smaller problems. (Depending on the implementation, the BMDP package can also suffer from a lack of storage space.) Both algorithms consider the logit response problem as a special case of logistic regression.

The Fienberg-Stewart algorithms have opted for economy of storage over efficiency of operation. Thus with similar front-end programs, be they interactively or batch oriented, one would expect these algorithms to be able to handle larger problems than either GLIM or BMDP. The Fienberg-Stewart algorithms have the advantage that they are able to directly fit multinomial logit response models, whereas the other approaches require that the loglinear equivalent of the model be fitted. Whether this is an advantage or not depends upon the size of the marginal array corresponding to the explanatory variables. When this margin is small, some advantages may accrue to the loglinear approach.

We have not made direct comparisons here on speed of convergence, but we note that, when IPF needs to iterate (i.e., for non-decomposable models), it has linear convergence, while the other algorithms enjoy quadratic convergence properties. We expect that the "special features" in the Fienberg-Stewart algorithms should allow for slightly faster convergence than do the GLIM and BMD algorithms, but rate of convergence should not be a serious distinction among these algorithms.

ACKNOWLEDGEMENT

This research was supported by Office of Naval Research Contracts N00014-78-C-0151 and N00014-78-C-0600 to the University of Minnesota. Reproduction in whole or in part is permitted for any purpose of the United States Government.

REFERENCES

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

Bock, R.D. (1975). Multivariate Statistical Methods in Behavioral Research. McGraw-Hill.

Dixon, W.J. and Brown, M.G. (eds.)(1979). BMDP-79. Berkeley, California, University of California Press.

Fienberg, S.E. (1977). The Analysis of Cross-classified Categorical Data. Cambridge, Mass., The MIT Press.

Fienberg, S.E. and Stewart, G.W. (1979). "The numerical analysis of contingence tables." Unpublished manuscript.

Haberman, S.J. (1974). The Analysis of Frequency Data. Chicago, Ill., University of Chicago Press.

Haberman, S.J. (1978). Analysis of Qualitative Data, Volume 1, Introductory Topics. New York, Academic Press.

Jennrich, R.I. and Moore, R.H. (1975). "Maximum Likelihood Estimation by Means of Nonlinear Least Squares." Proceedings of the Statistical Computation Section, American Statistical Association.

Nelder, J.A. (1976). Hypothesis Testing in Linear Models (Letter to the Editor). Amer. Statist. 30, 101.

Nelder, J.A. (1977). A Reformulation of Linear Models. J. Roy. Statist. Soc. A 140, 48-63.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. J. Roy. Statist. Soc. A 135, 370-384.

Table 1: Qualitative Comparison of Methods

Criterion	(i) Fienberg-Stewart Loglinear	(ii) Fienberg-Stewart Logit	(iii) GLIM Logit	(iv) BMDPLR	(v) IPF
1. Utilizes conditioning implied by logit model	No	Yes	Yes	Yes	No
2. Handles polytomous response structures directly	Yes	Yes	No[3]	No[3]	Yes
3. Can easily be extended to general logistic regression	No	Yes	Yes	Yes	No
4. Storage requirements	SSP matrix[1]	SSP matrix[2]	SSP matrix[2]	SSP matrix[2]	Data array plus margins
5. Detects non-existence of MLE's	Yes	Yes	No, only by non-convergence	No only by by slow convergence	No, only by slow convergence to zero
6. Handles structural *zeros*	Yes	Yes	Yes	Yes	Yes
7. Produces parameter estimates	Yes	Yes	Yes	Yes	Only for complete tables
8. Produces covariance estimates	Yes	Yes	Yes	Yes	No
9. Detects Aliasing	Yes, via pivoted Cholesky decomposition	Yes, via pivoted Cholesky decomposition	Yes, via small diagonal elements in non-pivoting Gaussian elimination	Yes, by (pivoted) Gauss-Jordan elimination	Notapplicable

[1] SSP matrix for all loglinear parameters

[2] SSP matrix for logit parameters only

[3] Both BMDP and GLIM can handle this problem by treating it as a loglinear model.

12 (14) TR-348-REV

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: (9) Technical Report No. 348, Revised	2. GOVT ACCESSION NO.: AD-A089 411	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): (6) ALTERNATIVE COMPUTATIONAL METHODS FOR ESTIMATION IN MULTINOMIAL LOGIT RESPONSE MODELS. Revision		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) Stephen E. Fienberg, Michael M. Meyer, G.W. Stewart		8. CONTRACT OR GRANT NUMBER(s): (15) N00014-78-C-0600
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Applied Statistics, School of Statistics, University of Minnesota, 1994 Buford Avenue, St. Paul, MN 55108		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, 800 N. Quincy Street, Arlington, VA 22217		12. REPORT DATE: (11) Nov ~~ember~~ 1979
		13. NUMBER OF PAGES: 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 13		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Categorical data; logit models; loglinear models; iterative proportional fitting; GLIM; Newton's method.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several algorithms have been proposed for the computation of maximum likelihood estimates for contingency tables. Since multinomial logit response models can be treated as special versions of loglinear models, many of these techniques can be used for logit models as well. In this paper we compare, in a qualitative fashion, the relative merits of (i) two variants of Newton's method developed by Fienberg and Stewart (ii) GLIM, as developed by Nelder and Wedderburn (iii) the BMDP program for stepwise logistic regression, and (iv) the widely used method of iterative proportional fitting.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

411265 JW