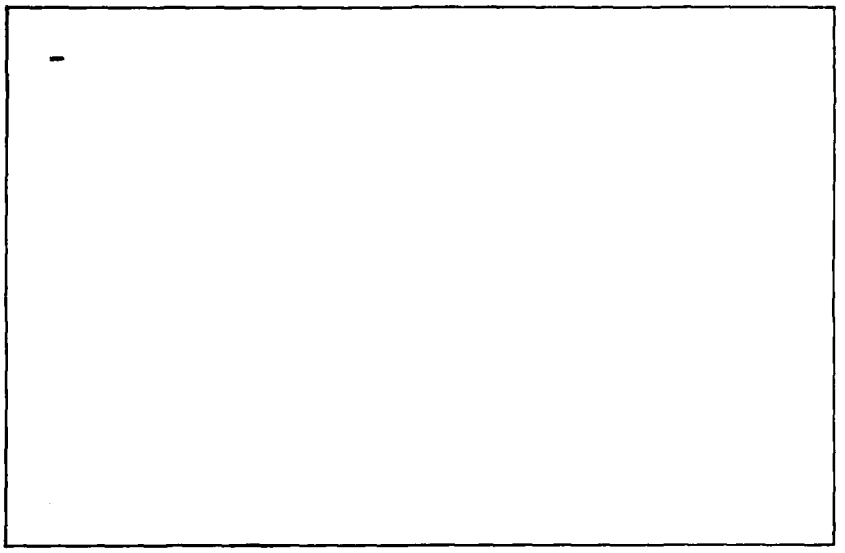
AD-A102 618 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/1
ITERATIVE PROPORTIONAL FITTING.(U)
JUN 81 S E FIENBERG, M M MEYER N00014-80-C-0637
UNCLASSIFIED TR-270 NL

LEVEL II

12



DEPARTMENT

OF

STATISTICS

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

(12

ITERATIVE PROPORTIONAL FITTING

by

Stephen E. Fienberg

and

Michael M. Meyer

Technical Report No. 270
Department of Statistics
Carnegie-Mellon University
June, 1981

This is a draft of an article prepared for the Encyclopedia of Statistical Sciences, to be published by John Wiley & Sons, Inc.

The preparation of this article was partially supported by the Office of Naval Research Contract N00014-80-C-0637 at Carnegie-Mellon University. Reproduction in whole or part is permitted for any purpose of the United States Government.

1. Introduction and Historical Remarks

The Iterative Proportional Fitting Procedure (IPFP) is a commonly used algorithm for maximum likelihood estimation in loglinear models. The simplicity of the algorithm and its relation to the theory of loglinear models make it a useful tool, especially for the analysis of cross-classified *categorical data* (q.v.) or *contingency tables* (q.v.).

To illustrate the algorithm we consider a three-way table of independent Poisson counts, $x = \{x_{ijk}\}$. Suppose we wish to fit the loglinear model of no-three-factor interaction for the mean $\mathbf{m}$, i.e. the model

$$\ln(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}. \tag{1}$$

The basic IPFP takes an initial table $\mathbf{m}^{(0)}$, such that $\ln(\mathbf{m}^{(0)})$ satistfies the model (typically we would use $m^{(0)}_{ijk} = 1$ for all i,j, and k) and sequentially scales the current fitted table to satisfy the three sets of the two-way margins of the observed table, x. The v'th iteration consists of three steps which form:

$$\begin{aligned} m^{(v,1)}_{ijk} &= m^{(v-1,3)}_{ijk} \cdot x_{ij+} / m^{(v-1,3)}_{ij+} \\ m^{(v,2)}_{ijk} &= m^{(v,1)}_{ijk} \cdot x_{i+k} / m^{(v,1)}_{i+k} \\ m^{(v,3)}_{ijk} &= m^{(v,2)}_{ijk} \cdot x_{+jk} / m^{(v,2)}_{+jk}. \end{aligned} \tag{2}$$

(The first superscript refers to the iteration number, and the second to the step number within iterations). The algorithm continues until the observed and fitted margins are sufficiently close. For a detailed discussion of convergence and some of the other properties of the algorithm, see Bishop, Fienberg and Holland (1975) or Haberman (1974). A FORTRAN implementation of the algorithm is given in Haberman (1972 and 1973). (See also the discussion of computer programs for loglinear models in the entry *Contingency Tables*, by Fienberg).

As a computational technique for adjusting tables of counts, the IPFP appears to have been first described by Kruithof (1937) (see also Krupp (1979)) and then independently formulated by Deming and Stephan (1940). They considered the problem of adjusting (or raking) a table, $\mathbf{n} = \{n_{ijk}\}$, of counts to satisfy some external information about the margins of the table. Deming (1943, p.107) gives an example of a cross-classification, by age and by state, of white persons attending school in New England. The population, $N = \{N_{ijk}\}$, cross-classification is unknown but the marginal totals are known. In addition a sample, $\mathbf{n}$, from the population is available. Deming and Stephan's aim was to find an estimate N which satisfies the marginal constraints and minimizes the χ^2-like distance,

$$\Sigma(N_{ij} - n_{ij})^2/n_{ij} . \tag{3}$$

Their erroneous solution (see Stephan (1942)) was the IPFP. Although the N produced by the IPFP need not minimize (3), it does provide an approximate and easily calculated solution.

Over twenty years after the work of Deming and Stephan, Darroch (1962) implicity used a version of the IPFP to find the maximum likelihood estimates in a contingency table but left the details of the general algorithm unclear. Bishop (1967) was the first to show how the IPFP could be used to solve the maximum likelihood estimation problem in multidimensional tables. Some further history and other uses of the algorithm, including applications to *doubly-stochastic matrices* (q.v.), are discussed in Fienberg (1970).

2. A Coordinate-free Version of the IPFP

The basic IPFP is applicable to a class of models much more general than those described soley in terms of margins of a multiway table. Consider an index set J with J elements and let $\mathbf{x}$ be a table of observed counts which are realizations of independent Poisson random variables with mean $\mathbf{m}$. Further let M be a linear subspace of $\mathbf{R}^J$ with a spanning set $\{\mathbf{f}_k : k = 1,2,...,K\}$ where each $\mathbf{f}_k$ is a vector of zeros and ones. The calculation of the maximum likelihood estimate $\mathbf{m}$ for the loglinear model

$$\ln(\mathbf{m}) \in M ,$$

begins by taking a starting table $\mathbf{m}^{(0)}$ with $\ln(\mathbf{m}^{(0)}) \in M$ ($\mathbf{m}^{(0)} = \mathbf{1}$ will always work), and sequentially adjusts the table to satisfy the "margins", i.e. $\langle \mathbf{f}_k , \mathbf{x} \rangle$ for k=1,2,....K , the inner products of the data with the spanning vectors. The v'th cycle of the procedure takes the current estimate $\mathbf{m}^{(v-1,K)} = \mathbf{m}^{(v,0)}$ and forms

$$\mathbf{m}^{(v,k)} = \mathbf{m}^{(v,k-1)} \frac{\langle \mathbf{f}_k , \mathbf{x} \rangle \, \mathbf{f}_k}{\langle \mathbf{f}_k , \mathbf{m}^{(v,k-1)} \rangle} + \mathbf{m}^{(v,k-1)} \cdot (\mathbf{1} - \mathbf{f}_k) , \quad k=1,2,...K , \tag{4}$$

(i.e. adjusts the current fitted table so that the margin corresponding to $\mathbf{f}_k$ is correct) to yield $\mathbf{m}^{(v)} = \mathbf{m}^{(v,K)}$. The maximum likelihood estimate is $\lim_v \mathbf{m}^{(v)}$. If one wished to fit the log-affine model

$$\ln(\mathbf{m}) \in \mathbf{t} + M ,$$

which is just the translation by $\mathbf{t}$ of the loglinear model M , then using the IPFP with starting values which satisfy this model (e.g. $\mathbf{m}^{(0)} = \exp(\mathbf{t})$) leads to the MLE.

There are many ways to view this basic algorithm and many problems for which the IPFP is of especial use. Although, the basic algorithm is limited to linear manifolds, M , with zero-one spanning sets, it is possible to generalize the method to work with any linear manifold. We now look at some topics which relate to the algorithm or its generalizations.

3. Some Computational Properites

Common alternatives to the IPFP are versions of Newton's method or other algorithms which use information about the second derivative of the likelihood function, or *Hessian* (q.v.). While such methods have quadratic convergence properties compared to the linear properties of the IPFP, and are often quite efficient (see e.g. Chambers (1977), Haberman (1974), or Fienberg, Meyer and Stewart (1979)) they are of limited use for models of high dimensionality. For example, the model of no-three-factor interaction in a 10 x 10 x 10 table has 271 parameters and this requires ½ x 271 x 272 = 36.856 numbers to represent the Hessian. In contrast the IPFP requires only about 300 numbers (i.e. the 3 marginal totals) in addition to the table itself. For many large contingency table problems the IPFP is the most reasonable computational method in use. Of course, for problems with only a small number of parameters Newton's method may be preferable, especially when the model is such that the basic IPFP is not applicable. Newton's method also automatically produces an estimate of the variance-covariance matrix of the parameters, but this is what requires all of the storage space.

It is well known that the IPFP can often be slow to converge. Our experience is that it is generally restrictions on storage rather than computational time which limit an algorithm's usefulness. Thus slow convergence, while disturbing in some contexts, is not necessarily a crucial property.

As we have seen, the basic IPFP is very simple and can require little more than hand calculation. The simplicity of the algorithm allows one to understand and use the mechanics of the calculations to show theoretical results. A good example of this is the theory of decomposable models (models with closed-form estimates) as developed by Bishop, Fienberg and Holland (1975) or Haberman (1974). These models are closely related to the IPFP ; a fundamental theorem (Haberman (1974, p. 191)) says that for every decomposable model there is an ordering of the margins such that the simple IPFP converges in one iteration.

One of the ideas underlying the IPFP is to sequentially equate a vector of expected values with the sufficient statistics of the model. The IPFP does this one dimension at a time but there is no reason why several dimensions can not be simultaneously adjusted. This idea underlies the estimation scheme for partially decomposable graphical models outlined in Darroch, Lauritzen and Speed (1980). They show that for many models it is possible to fit certain subsets of the marginal totals and to combine the resulting partial estimates using a direct formula. Their approach helps answer the question which asks for the cyclic order that the IPFP should use in satisfying the marginal totals. The results of Darroch, Lauritzen and Speed show that certain groupings and orderings are particularly advantageous.

4. Generalizations of the IPFP

A limitation of the basic IPFP is that only certain types of models can be fit. We now consider several methods for extending the IPFP to cover any loglinear model. For multinomial and Poisson data the problems of maximizing the likelihood function and minimizing the *Kullback-Leibler information number* (q.v.) can be considered as dual problems which lead to the same estimates (see the entry, *Contingency Tables* by Fienberg). We now consider generalizations of the IPFP from both these points of view.

Haberman (1974) shows that, when viewed from the likelihood perspective, the IPFP is just a version of the cyclic coordinate ascent method of functional maximization. To illustrate Haberman's approach, we choose a fixed set of vectors which span the model space, M, and then we maximize the likelihood along each of these directions in turn. Specifically, we consider a set of vectors $= \{\mathbf{f}_k : k = 1,2,....K\}$ which span M. If we denote the log-likelihood by $\ell(\mathbf{m} \mid \mathbf{x})$ and consider an initial estimate $\mathbf{m}^{(0)}$ with $\ln(\mathbf{m}^{(0)})$ in M, then the algorithm proceeds by finding $\mathbf{m}^{(i)}$ such that

$$\ln(\mathbf{m}^{(i)}) = \ln(\mathbf{m}^{(i-1)}) + \alpha_i \mathbf{f}_k \; ; \; i = k \bmod |K| \; .$$

where α_i is determined so as to increase the likelihood sufficiently. When $\mathbf{f}_k$ is a vector of zeros and ones

$$\alpha_i = \ln(\langle \mathbf{f}_k, \mathbf{x} \rangle / \langle \mathbf{f}_k, \mathbf{m}^{(i-1)} \rangle)$$

(i.e. the α_i corresponding to the IPFP adjustment maximizes the likelihood in this direction). For arbitrary $\mathbf{f}_k$ there is no direct estimate of α_i and we are left with a one dimensional maximization problem.

Csiszar (1975) considers the IPFP as a method for maximizing the Kullback-Leibler information between two probability distributions. When specialized to distributions on finite sets, Csiszar's methods yield a generalized IPFP. The class of algorithms which result from Csiszar's work are dual algorithms to the cyclic ascent methods except now maximization can be over entire subspaces of M rather than just vectors. These methods yield powerful theoretical tools and have been instrumental in finding new algorithms which combine some of the advantages of both Newton's method and the IPFP (see Meyer (1981)).

The third generalization of the IPFP we consider is due to Darroch and Ratcliff (1972). This algorithm, known as Generalized Iterative Scaling, was also developed from the information theory perspective, but is not closely related to Csiszar's method. The calculations are similar to those of the basic IPFP: a set of vectors F which span M is chosen and the likelihood is increased (but not maximized) in each of these directions in turn. Each iteration can require that the scaling factors be raised to arbitrary powers. These features combine to

make the algorithm expensive as it often takes many iterations to converge and each iteration is complicated.

For some problems it is possible to avoid the complications of the generalized IPFP's by transforming the contingency table into a form where the basic IPFP can be used (see Meyer (1981) for details and Fienberg, Meyer, and Wasserman (1981) for some examples). This can result in a significant saving in the computational effort and recognition of some of the theoretical advantages (e.g. closed-form estimates) associated with the IPFP. Fienberg and Wasserman (1981, Fig. 1) present an example where the convergence rate can be substantially improved by taking advantage of this transformation technique.

Related Entries Categorical Data, Contingency Tables, Kullback-Leibler Information.

References

Bishop, Y.M.M. (1967). *Multidimensional Contingency Tables: Cell Estimates.* PhD thesis, Harvard University .

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

Chambers, J.M. (1977). *Computational Methods for Data Analysis.* New York: John Wiley & Sons, Inc.

Csiszar, I. (1975). I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability, 3,* 146-158.

Darroch, J.N. (1962). Interaction in Multi-Factor Contingency Tables. *Journal of the Royal Statistical Society, Series B, 24,* 251-263.

Darroch, J.N., Lauritzen, S.L., and Speed, T.P. (1980). Markov Fields and Log-linear Interaction Models for Contingency Tables. *The Annals of Statistics, 8,* 522-539.

Darroch, J.N. and Ratcliff, D. (1972). Generalized Iterative Scaling for Log-linear Models. *The Annals of Mathematical Statistics, 43,* 1470-1480.

Deming, W.E. (1943). *Statistical Adjustment of Data.* New York: John Wiley & Sons, Inc.

Deming, W.E. and Stephan, F.F. (1940). On a Least Squares Adjustment of a Sample Frequency Table when the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics, 11,* 427-444.

Fienberg, S.E. (1970). An Iterative Procedure for Estimation in Contingency Tables. *The Annals of Mathematical Statistics, 41,* 907-917.

Fienberg, S.E., Meyer, M.M., and Stewart, G.W. (1979). *Alternative Computational Methods for Estimation in Multinomial Logit Response Models.* Technical Report 348, School of Statistics, University of Minnesota .

Fienberg, S.E., Meyer, M.M., and Wasserman, S.S. (1981). Analyzing Data from Multivariate Directed Graphs: An Application to Social Networks. In Vic Barnett (Ed.), *Looking at Multivariate Data,* Chichester, England: John Wiley & Sons., to appear.

Fienberg, S.E. and Wasserman, S.S. (1981). Comment on "An Exponential Family of Probability Distributions". *Journal of the American Statistical Association, 76(373),* 54-57.

Haberman, S.J. (1972). Loglinear Fit for Contingency Tables (Algorithm AS 51). *Applied Statistics, 21,* 218-225.

Haberman, S.J. (1973). Printing Multidimensional Tables (Algorithm AS 57). *Applied Statistics, 22,* 118-126.

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

Kruithof, R. (1937). Telefoonverkeersrekening. *De Ingenieur, 52*, E15-E25.

Krupp, R.S. (1979). Properties of Kruithof's Projection Method. *The Bell System Technical Journal, 58(2)*, 517-538.

Meyer, M.M. (1981). *Applications and Generalizations of the Iterative Proportional Fitting Procedure.* PhD thesis, School of Statistics, University of Minnesota .

Stephan, F.F. (1942). An Iterative Method of Adjusting Frequency Tables when Expected Marginal Totals are Known. *The Annals of Mathematical Statistics, 13*, 166-178.

Table of Contents

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER: Technical Report No. 270	2. GOVT ACCESSION NO: AD-A102 618
3 RECIPIENT'S CATALOG NUMBER	
4 TITLE (and Subtitle): Iterative Proportional Fitting	5 TYPE OF REPORT & PERIOD COVERED
6 PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Stephen E. Fienberg Michael M. Meyer	8 CONTRACT OR GRANT NUMBER(s): [illegible]
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics Carnegie-Mellon University Pittsburgh, PA 15213	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Contracts Office Carnegie-Mellon University Pittsburgh, PA 15213	12. REPORT DATE: June, 1981 13 NUMBER OF PAGES
14 MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	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)

loglinear models, categorical data, generalized iterative scali[illegible]

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

This paper considers the uses of the Iterative Proportional Fitting Procedures for fitting loglinear models. Some historical remarks are followed by a discussion of generalizations of the algorithm and some applications.

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)

END

DATE
FILMED

9-81

DTIC