MAXMUM LIKELIHOOD ESTIMATION FOR NON-HIERARCHICAL LOG-LINEAR MODELS

by Sung-Ho Kim
School of Humanities and General Sciences
Korea Advanced Institute of Science and Technology

The statistical concepts of sufficiency and conditional independence have been attracting increasing attention in applied statistics, in particular for representing a relation among a group of variables in a graphical format. Methods of fitting graphical models of continuous variables such as structural equation models are well described in Bollen (1989). Graphical models of categorical or finitely discrete variables are representable in the form of undirected graph, directed acyclic graph, or mixture of them (see Whittaker (1990), Chapter 3.) The graphical models include the graphical log-linear model (Fienberg (1980); Darroch, Lauritzen, and Speed (1980)), recursive models for contingency tables (Wermuth and Lauritzen (1983)), Bayesian networks (Pearl (1988)), and influence diagrams (Howard and Matheson (1981); Olmsted (1983); Shachter (1986); Smith (1989)). Among them, recursive models, Bayesian networks, and influence diagrams of finitely discrete random variables share a common feature that the joint probability of the variables involved in each of them is expressible as a product of marginal and conditional probabilities. Statistical modelling of such models mostly rests on the methods developed for log-linear modelling (Bishop, Fienberg, and Holland (1975)).

We will say that a set of variables is fully interactive if the conditional probability of any one of the variables depends upon the rest of the variables. We say that the log-linear model of a set of variables is hierarchical if there is a subset of the variables which is fully interactive and every subset of which is also fully interactive. We call such log-linear model a hierarchical log-linear model.

The Iterative Proportional Fitting (IPF) algorithm and the Newton-Raphson algorithm are well known for fitting hierarchical log-linear models. Among the hierarchical log-linear models, graphical log-linear models as defined by Darroch, Lauritzen, and Speed (1980) have received lots of attention since they are accompanied by a graphic representation which depicts the relation among the variables involved in the model, the relation being interpretable in the context of Markov property. While undirected graphs are used for graphical log-linear models, we use directed acyclic graphs for recursive models (Wermuth and Lauritzen (1983)) and influence diagrams (Smith (1989)). We will abbreviate "influence diagram" into "ID" from now on.

The log-linear model for the variables involved in an ID is not hierarchical in general. In other words, IDs are not hierarchical in general. In this paper, we will consider non-hierarchical log-linear models in the form of ID, and present a new approach to estimating the parameters involved in the model, that is developed by the author of the paper.

A brief review of the methodology for modelling IDs or the like follows. Birch (1963) considered maximum likelihood estimation for an ID (although he did not use this terminology) of three variables by dealing with its log-linear model as a combination of a log-linear model of a marginal probability and that of a conditional probability, where the whole joint probability is given by the product of the marginal and the conditional probabilities. In the laguage of latent class model, Goodman (1974a, 1974b) related the latent class models to log-linear models and gave a general algorithm for maximum likelihood estimation for latent class models (also see Haberman (1977, 1979)). We can view the latent class model as an ID with the unobserved variables as the parent nodes of the observed variables. Rebane and Pearl (1987) improved upon Chow and Liu (1968)'s algorithm toward an algorithm by which one can recover the relation among a set of variables, under the assumption that the relation is representable via a tree-like directed graph where the parent nodes of each node in the graph are not connected to each other. Glymour, Scheines, Spirtes, and Kelly (1987) developed an algorithm which suggests a set of possible causal structures for statistical data together with their measures of fit, but the algorithm does not provide estimates of the marginal or conditional probabilities of the nodes in a given structure. Their algorithm makes use of the property of the correlation coefficients of the pair of variables on a path in a tree-structure.

A good side of the ID is that its probability model is multiplicative in some order of the variables. The other side of it is that the log-linear model of the variables is not necessarily hierarchical. This point was well addressed by Birch (1963) and Wermuth and Lauritzen (1983). Methods of fitting ID models to statistical data would include Birch (1963)'s multiplicative model approach and the Newton-Raphson approach as described in Haberman (1979). A drawback of these approaches is that the complexity of the model increases exponentially. The new approach to be presented in this paper is an alternative approach as a generalised version of the iterative proportinal fitting (IPF) algorithm for hierarchical log-linear models. We will call the approach Generalized IPF or GIPF approach.

The idea behind the GIPF approach follows. First we transform a given ID into an undirected graph by connecting the parents of each node to each other. So each child-parents set becomes a clique in the undirected graph. As in Lauritzen and Spiegelhalter (1988) we will call such an undirected graph a moral graph. The moral graph may not necessarily be decomposable. Then we proceed to do the ordinary IPF for the log-linear model corresponding to the undirected graph with the following structural constraint imposed in the IPF process. The structural constraint applies to the sets of the newly connected nodes and their parent nodes. Such a set actually consists of all the newly connected parent nodes of a node and all the parent nodes of the newly connected parent nodes. We will call the set a constraint set. In the ordinary IPF process, we impose the marginal structure among the variables in each constraint set so that the structural constraint may be realized in the ordinary IPF process.

The GIPF approach is easy to apply for fitting non-hierarchical models in the form of ID whether they contain unobserved variables or not. When a non-hierarchical model involves unobservable variables, we can employ the idea behind the EM algorithm (Goodman (1974), Haberman (1979)).

We applied the GIPF approach to a couple of simulated data sets, each involving 14 binary variables. The fitting result strongly suggests that the approach is very useful for fitting non-hierarchical models of categorical variables in the form of ID.

Keywords and phrases; Iterative proportional fitting, Log-linear model; Maximum likelihod estimation, Multiplicative model; Structural constraint.

REFERENCES

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. R. Statist. Soc. B.* **25**, 220-233.

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

Chow, C. K. and Liu, C. N. (1968). Approximating discrete probability distributions with dependent trees. *IEEE Trans. on Info. Theory* IT-14, 462-67.

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**, 3, 522-539.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data,* 2nd edn. Cambridge, MA: MIT Press.

Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. (1987). *Discovering causal structure.* New York: Academic Press.

Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach. *American Journal of Sociology.* **79**. 1179-1259.

Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika.* **61**, 215-31.

Haberman, S. J. (1977). Product models for frequency tables involving indirect observation. *The Annals of Statistics.* **5**. 1124-47.

Haberman, S. J. (1979). *Analysis of Qualitative Data,* Vol. 2. New York: Academic Press.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc. B,* **50**, 2, 157-224.

Rebane, G. and Pearl, J. (1987). The recovery of causal poly-trees from statistical data. *Proc., 3rd Worksop on Uncertainty in AI,* Seattle. 222-28.

Smith, J. Q. (1989). Influence diagrams for statistical modelling. *The Annals of Statistics,* **17**, 2, 654-72.

Wermuth, N. and Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika,* **70**, 3, 537-52.