An Exponential Family of Probability Distributions for Directed Graphs: Comment

Stephen E. Fienberg; Stanley Wasserman

Comment

STEPHEN E. FIENBERG and STANLEY WASSERMAN*

1. INTRODUCTION

Statistical approaches for modeling social network data have been distinguished primarily by their absence from much of the social network research literature. Holland and Leinhardt's work in this area is a notable exception, and thus it is a special pleasure for us to comment on the present paper. We believe that Holland and Leinhardt's development of exponential family models for directed graph data is a major advance that will rapidly be applied in a variety of substantive areas.

In February 1978 we first attempted to develop statistical methods for a class of log-linear models for the analysis of a social network brought to us by a Minnesota colleague, J. Galaskiewicz. We accomplished little until we had an opportunity to read an early version of Holland and Leinhardt's paper. Our progress immediately accelerated, and their exponential family of probability distributions and bivariate generalizations have served as a basis for much of our subsequent work with the Galaskiewicz data (Fienberg and Wasserman 1980, 1981; Fienberg, Meyer, and Wasserman 1981). Here we describe additional results and observations regarding Holland and Leinhardt's p_1 model, which help link their work to the extensive literature on log-linear models for contingency table analysis. The order of presentation of our comments parallels that of their paper. We restrict our comments to results for single relation data and refer the interested reader to Fienberg, Meyer, and Wasserman (1981) for a partial treatment of multivariate p_1-like models applied to the Galaskiewicz data, which involve three relations.

2. FITTING THE p_1-DISTRIBUTION

Holland and Leinhardt present a version of generalized iterative scaling (Darroch and Ratcliff 1972)—GIS—for solving the likelihood equations for the parameters of p_1. While GIS is known to converge, Holland and Leinhardt provide us with no information on how rapidly convergence takes place in practice, and specifically for the example of Section 4. Many authors (e.g., see Fienberg 1980) have observed that the presence of fractional pow-

ers of the GIS adjustment factors (i.e., the powers of $\frac{1}{2}$ in (31) and (34) tends to make convergence excrutiatingly slow relative to the standard iterative scaling procedure; however, there is no guarantee that an alteration to GIS to speed up convergence will "work" in general (i.e., converge or maximize the likelihood). Thus it is sometimes profitable to explore alternative versions of GIS and generate alternative data representations that allow the use of simpler and perhaps more rapidly converging versions of iterative scaling algorithms.

Our approach to this problem has been to begin with an alternative notation that is evocative of that used in the analysis of multidimensional contingency tables. Rather than working directly with x, we define a new array, y, with elements

$$y_{ijkl} = \begin{cases} 1 & \text{if } D_{ij} = (x_{ij}, x_{ji}) = (k, l) \\ 0 & \text{otherwise.} \end{cases}$$

Note that y_{iikl} is defined to be equal to zero and that $y_{ijkl} = y_{jilk}$. Thus we need only consider one "triangle" of y, in which $i > j$. By retaining all $g^2 \cdot 2^2$ cells in the array, however, we are able to express the sufficient statistics for the parameters of p_1 given by (27) as margins of y:

$\frac{1}{2}y_{++11} = M$, number of mutuals,

$y_{i+1+} = x_{i+}$, out-degree of node i,

$y_{+j1+} = x_{+j}$, in-degree of node j,

$y_{++1+} = x_{++}$, total number of choices.

Through the use of the redundant representation of the full y array, one can show that fitting p_1 to the x array is equivalent to fitting the "no three-factor" interaction log-linear model to the incomplete four-dimesional array, y. Technical details are given in Meyer (1980). Thus we can fit p_1 to the data by using standard iterative scaling applied to y (Fienberg and Wasserman 1981).

The no three-factor interaction log-linear model, in the common notation of Fienberg (1980) and others, simultaneously fits the following margins of y to a four-way array of estimated probabilities:

[12] [13] [24] [14] [23] [34] .

The [12] margin must be included in any iterative scaling algorithm to ensure that the fitted values satisfy the constraints on the dyadic probabilities, (12). By exploiting

the symmetries in y, we can straightforwardly reexpress the remaining two-way margins as linear combinations of [12] and the sufficient statistics, y_{++11}, $\{y_{i+1+}\}$, and $\{y_{+j1+}\}$. We compare the convergence of the standard iterative scaling algorithm with Holland and Leinhardt's algorithm on the Sampson data in a later section.

The submodels described by Holland and Leinhardt as special cases of p_1 and listed in their Table 1 are all related to equivalent log-linear models for y, and thus they too can be fit by using standard iterative scaling computer programs (see Fienberg and Wasserman 1981 for further details). For example, fitting the version of p_1 with $\rho = 0$ is equivalent to fitting the model given by [12] [13] [24] [14] [23] to y, and fitting the version of p_1 with $\{\beta_j = 0\}$ is equivalent to fitting the model given by [12] [13] [24] [34].

3. ASYMPTOTICS FOR TESTS WITHIN p_1

A noteworthy feature of Holland and Leinhardt's work is their pilot Monte Carlo study of the "small-sample" distribution of likelihood ratio test statistics for comparing hypotheses within p_1; for example, testing for $\rho = 0$ by testing H_1 against H_3. The usual asymptotic chi-squared distributions associated with such tests are based on a *fixed* parametric structure and let the sample sizes tend to infinity. Such an asymptotic setup is inapplicable in the present context because the sample size is $g(g - 1)$, which is a function of g, the number of nodes. But as we let g grow (so that the sample size can get large), we get a sequence of probability structures whose dimensions are also growing as a function of g; that is, there are $2g$ parameters for p_1 and $2g - 1$ for p_1 with $\rho = 0$.

For the test of $\rho = 0$ using H_1 against H_3, the special large sparse asymptotics of Haberman (1977) appear, at least on the surface, to be applicable since the difference in dimensionality is fixed; that is, the difference equals one. In this case, Haberman's results suggest the use of the standard χ_1^2 approximation to the null distribution under H_1 for large samples. The Holland and Leinhardt simulation provides some support for this conjecture. The approach to this asymptotic distribution as g increases in their simulation appears to be from the direction of distributions with slightly heavier upper tails than χ_1^2.

For similar likelihood ratio tests of $\{\alpha_i = 0\}$ or $\{\beta_j = 0\}$, for example, H_2 against H_3, Haberman's results are inapplicable as the difference in dimensionality, $g - 1$, increases with the sample size. All may not be lost here, since the dimensionality is of order the square root of the sample size, $g(g - 1)$, but new asymptotic expansions for such situations need to be developed. Even if the resulting asymptotic null distribution as g tends to infinity is not χ_{g-1}^2, normal approximations similar to those developed by Koehler (1977) are likely to be applicable.

4. TESTING THE FIT OF p_1

To check on the fit of p_1, Holland and Leinhardt suggest using a probability distribution p_2 that contains p_1

as a special case. One approach to developing such distributions or models is to search for more complicated log-linear models for y, which contain all six two-factor effects, and hence p_1. These models should be substantively interesting and conceivably provide a better fit than p_1.

One such model could be based on the notion of differential reciprocity. Rather than setting the dyadic reciprocity effects equal to a constant as with p_1 in (19), we could take

$$\rho_{ij} = \rho + \rho_i + \rho_j , \quad i > j ,$$

where the $\{\rho_i\}$ are normalized to sum to zero. This model allows the effect of reciprocity to depend in a linear manner on the two actors in a dyad. It can be shown that fitting this model of differential reciprocity to the x array is equivalent to fitting the following log-linear model to the y array:

$$[12] \quad [134] \quad [234] .$$

There is no closed-form solution available for fitting estimated probabilities under this model to the y array, and thus some form of iterative solution, such as iterative scaling, is required.

Unfortunately, there are no other standard log-linear models that can be fit to y that contain p_1 as a special case. For example, one model, mentioned by Holland and Leinhardt, replaces $\alpha_i + \beta_j + \theta$ by θ_{ij} where $\theta_{ij} \neq \theta_{ji}$. This model corresponds to fitting the log-linear model to y specified by

$$[123] \quad [124] \quad [34] .$$

The sufficient statistics [123] and [124] are the margins of each 2×2 subtable of y. Because they consist of $0 - 1$ entries, however, the margins completely determine the interior of the 2×2 tables, and thus this model is equivalent to the saturated model for the y array.

If we denote this model of differential reciprocity by

$$H_4: \theta, \rho, \{\alpha_i\}, \{\beta_i\}, \{\rho_i\} \quad \text{all unconstrained,}$$

then a test for the adequacy of p_1 can be based on testing H_3 (i.e., p_1) against H_4. The corresponding likelihood ratio statistic has $g - 1$ degrees of freedom, and our earlier discussion about possible asymptotic distributions for tests of $\{\alpha_i = 0\}$ or $\{\beta_j = 0\}$ is relevant here as well.

Holland and Leinhardt's approach to testing the fit of p_1 is to develop p_2 with a single additional parameter, δ, yielding a new sufficient statistic, $z(\mathbf{x})$. Then they suggest using a normal approximation to the null distribution of the test statistic τ in (62). The version of τ that they use in practice is based on the triad census of $\mathbf{x}$. The null distribution is developed using a series of approximations (see Holland and Leinhardt 1978), only some of which are made completely explicit in this paper. For example, they recognize but choose to ignore dependencies in the triad counts, and they do not use μ and Σ, the conditional moments of τ given M, $\{X_{i+}\}$, and $\{X_{+j}\}$. Instead, they approximate μ and Σ by the conditional moments of τ

given M, $\bar{X}$, V(in), V(out), and COR(out,in). The effects of such approximations on the distribution of τ are very unclear (see Wasserman 1977 for a discussion of these conditional distributions). Finally, Holland and Leinhardt suggest that the asymptotic distribution of their statistic τ^2(max), given in (64), is χ_{10}^2. This appears to be based on a further assumption, not discussed either in the present paper or in Holland and Leinhardt (1978), regarding the nonzero eigenvalues of the conditional covariance matrix. Given the problematic nature of many of these assumptions, we would be extremely cautious in using the suggested χ_{10}^2 reference distribution, even as an approximation. While Holland and Leinhardt's work on the distribution of τ^2(max) demonstrates a very high level of technical virtuosity, clearly further work needs to be done.

5. THE EXAMPLE

In our earlier discussion of fitting p_1, we discussed three different algorithms: (a) Holland and Leinhardt's, (b) an alternative generalized iterative scaling algorithm without the powers of $\frac{1}{2}$, (c) regular iterative scaling applied to y. The speed of convergence of all three algorithms is illustrated in our Figure 1, using as the conver-
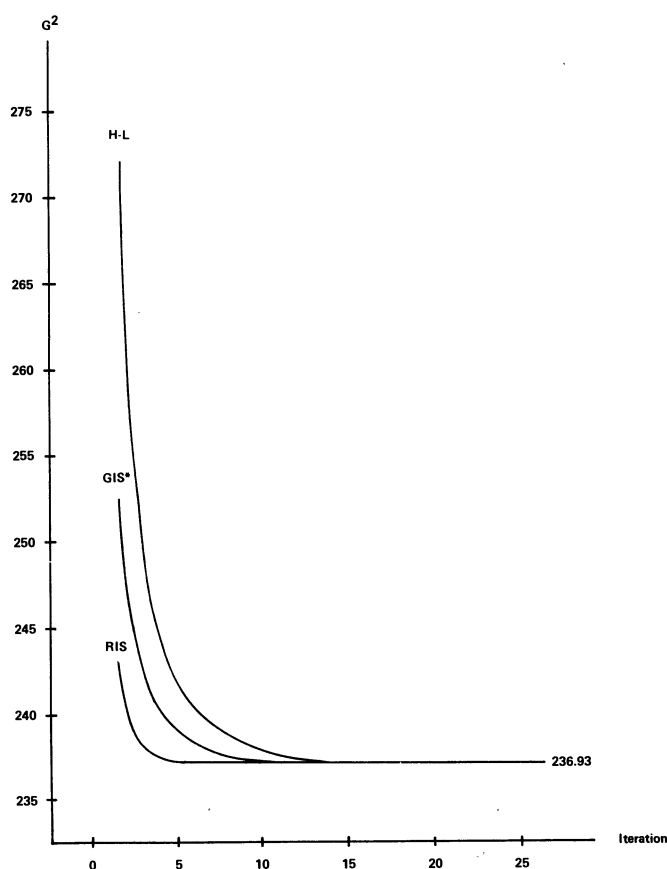

Figure 1. Comparison of the Rates of Convergence for Various Algorithms Applied to Sampson's Data: H-L = Holland-Leinhardt Algorithm; GIS* = Generalized Iterative Scaling Algorithm Without Powers of 1/2; RIS = Regular Iterative Scaling Algorithm

gence criterion the likelihood ratio statistic of p_1 against unrestricted alternatives, referred to here as G^2. In this example, regular iterative scaling applied to y was far superior to our version of generalized iterative scaling without the powers of $\frac{1}{2}$, and GIS without the powers of $\frac{1}{2}$ was in turn superior to Holland and Leinhardt's GIS algorithm. These results are in accord with our experience in applying such algorithms to other problems.

This comparison of the three algorithms is for illustrative purposes only and is certainly not complete. We have not directly examined the number of operations performed, the extent of central memory and storage used, execution time, and so on. Note that the y array is much larger than the x array, and thus our approach, which requires a *single* large array, could rapidly exhaust the storage capacity of even the largest computers as g increases.

The likelihood ratio statistic for testing H_3 against H_4 ($\rho_1 = \rho_2 = \cdots = \rho_g = 0$) for Sampson's network is the difference in G^2's for the two models: $\Delta G^2 = 236.93 - 205.96 = 30.97$. If χ_{17}^2 is used as an approximate reference distribution, the significance level for this test is $p \approx .02$, partially indicative of lack of fit for p_1. We see that in addition to tendencies toward *general* reciprocity and transitivity, there is evidence that the monks enter into reciprocal relationships at *different* rates.

In their dicussion of the example, Holland and Leinhardt note the alternative estimate of ρ suggested by Davis (1968). Davis's 2 × 2 table in (68) is similar to the 2 × 2 *intraclass* contingency table that often arises in genetics applications such as twins studies (e.g., see Plackett 1974, pp. 122–123; and Altham 1971). Perhaps specializations of p_1, using nodal characteristics as in Fienberg and Wasserman (1981), may have some application to this completely different class of practical problems.

Finally, as Holland and Leinhardt note, the version of the Sampson data that they analyze in Section 4 involves only one of eight different relations in the original study. These relations are of two kinds: positive feelings—like, esteem, influence, and praise; and negative feelings—antagonism, disesteem, negative influence, and blame. How similar are the four positive relations? Are the negative relations antithetical to the positive? Earlier analyses of this group of 17 monks indicated that the individuals can be placed into three subgroups: "loyal opposition," "young Turks," and "outcasts." Once we control for membership in these subgroups, are there any individual differences? A complete analysis of these data, using the multivariate versions of p_1 in Section 5 of Holland and Leinhardt and the methods and models in Fienberg, Meyer, and Wasserman (1981), may yield answers to such questions.

ADDITIONAL REFERENCES

ALTHAM, P.M.E. (1971), "Exact Bayesian Analysis of an Intraclass 2 × 2 Table," *Biometrika*, 58, 679–680.

FIENBERG, S.E. (1980), *The Analysis of Cross-Classified Categorical Data* (2nd ed.), Cambridge, Mass.: MIT Press.

FIENBERG, S.E.. MEYER, M.M., and WASSERMAN, S. (1981), "Analyzing Data From Multivariate Directed Graphs: An Application to Social Networks," in *Interpreting Multivariate Data*, ed. V. Barnett, London: John Wiley & Sons, in press.

FIENBERG, S.E., and WASSERMAN, S. (1980), "Methods for the Analysis of Data From Multivariate Directed Graphs," in *Proceedings of the Conference on Recent Developments in Statistical Methods and Applications*, Taipei, Taiwan: Institute of Mathematics, Academia Sinica, 137–161.

——— (1981), "Categorical Data Analysis of Single Sociometric Relations," in *Sociological Methodology 1981*, ed. S. Leinhardt, San Francisco: Jossey-Bass, 156–192.

HABERMAN, S. (1977), "Log-linear Models and Frequency Tables With Small Expected Counts," *Annals of Statistics*, 5, 1148–1169.

KOEHLER, K.J. (1977), "Goddness of Fit Statistics for Large Sparse Multinomials," PhD dissertation, University of Minnesota, School of Statistics.

MEYER, M.M. (1980), "Generalizing the Iterative Proportional Fitting Procedure," Technical Report #371, University of Minnesota, School of Statistics.

PLACKETT, R.L. (1974), *The Analysis of Categorical Data*, London: Griffin's Statistical Monographs.

WASSERMAN, S. (1977), "Random Directed Graph Distributions and the Triad Census in Social Networks," *Journal of Mathematical Sociology*, 5, 61–86.