DISCUSSION

NORMAN BRESLOW AND DIRK MOORE

University of Washington

In our first reading of this stimulating paper, we were struck by the apparent close correspondence between the proposed random effects model for contingency table analysis and the theory of generalized linear models and quasi-likelihood developed by Nelder and Wedderburn (1972) and Wedderburn (1974) and investigated further in the recent monograph by McCullagh and Nelder (1983). For some simple parametric models, the one-parameter family of densities (5.6) appears identical to the corresponding quasi-likelihood functions with $\nu^{-1} = (n\theta)^{-1}$ playing the role of the dispersion parameter that multiplies the variance function. Further, parallels between the two developments are seen in the fact that residual deviance or chi-square goodness-of-fit measures, divided by their degrees of freedom, are used to estimate ν (cf. equations (4.8) and (5.20)). Furthermore, $\hat{\nu}$ acts as a scaling factor for the asymptotic distribution of the vector of sufficient statistics for the β parameters of primary interest (5.23), as it does for the parameter estimates under quasi-likelihood. It is quite interesting that the relatively simple asymptotic results, already known to hold unconditionally from quasi-likelihood theory, apply even to the complicated conditional distributions considered in the paper. Since we suspect that such matters will receive more thorough discussion from others with greater knowledge of quasi-likelihood techniques, however, we turn our attention to a possible alternative measure of the degree to which a given table conforms to the hypothesis of independence.

As is well known (Bishop, Fienberg and Holland, 1975), the hypothesis of independence of row and column classifications in a contingency table may be expressed as a log-linear model for the expected cell frequencies $E(m_{ij})$. The likelihood calculations are simplest when the m_{ij} have independent Poisson distributions, and we keep to this in order to make the discussion as transparent as possible. Conditioning on the grand total leads to the hypothesis of independence for the multinomial distribution considered by Diaconis and Efron.

In order to accommodate the idea of a sample size that increases while the number of cells IJ remain fixed, we suppose that the Poisson means have the form $E(m_{ij}) = N\lambda_{ij}$. We further suppose that the λ_{ij} are sampled independently from distributions with means $\zeta_{ij} = \exp(\mu_0 + \alpha_i + \beta_j)$, the hypothesis of independence, and variances that represent the degree of departure from this hypothesis. The usual quasi-likelihood generalization of the Poisson model results from the assumption $\mathrm{Var}(\lambda_{ij}) = \sigma^2\zeta_{ij}$. It follows that $E(m_{ij}) = N\zeta_{ij} = \mu_{ij}$ and $\mathrm{Var}(m_{ij}) = (1 + N\sigma^2)\mu_{ij}$. This generalized linear model with dispersion parameter $\theta^{-1} = (1 + N\sigma^2)$ corresponds closely to the situation considered by Diaconis and Efron. Imposition of the parameter constraint $\sum_{i,j} \zeta_{ij} = 1$ via an appropriate choice of μ_0 ensures that N is estimated by the grand total n.

An alternative random effects model that seems to us more natural in the context of log-linear theory expresses the random effects on the same scale as the fixed row and column effects α_i and β_j. The model becomes $m_{ij} \mid \lambda_{ij} \sim$ Poisson $(n\lambda_{ij})$ and $\lambda_{ij} = \exp(\mu_0 + \alpha_i + \beta_j + \varepsilon_{ij})$ where the ε_{ij} are iid with mean 0 and a common variance τ^2. This leads approximately to a generalized linear model of the form $E(m_{ij}) = \mu_{ij}$ and $\mathrm{Var}(m_{ij}) = \mu_{ij} + \tau^2\mu_{ij}^2$, which equations hold exactly if one assumes the λ_{ij} to have appropriate gamma densities. Following Williams (1982) work with the binomial distribution, Breslow (1984) suggested a method of moments estimation procedure for the unknown parameters in this model such that the chi-square criterion

$$\sum_{i,j} \frac{(m_{ij} - \mu_{ij})^2}{(\mu_{ij} + \tau^2\mu_{ij}^2)} = (I - 1)(J - 1)$$

and the quasi-likelihood equations for fixed τ^2, which in this case are

$$\sum_i \frac{m_{ij}}{(1 + \mu_{ij}\tau^2)} = \sum_i \frac{\mu_{ij}}{(1 + \mu_{ij}\tau^2)}$$

$$\sum_j \frac{m_{ij}}{(1 + \mu_{ij}\tau^2)} = \sum_j \frac{\mu_{ij}}{(1 + \mu_{ij}\tau^2)},$$

are satisfied simultaneously. Approximate confidence bounds on τ^2 are obtained by solving the same set of equations with $(I - 1)(J - 1)$ replaced by percentiles of chi-square distributions having $(I - 1)(J - 1)$ degrees of freedom; however, further work is needed to establish conditions under which the Pearson statistic actually has an approximate chi-square distribution (McCullagh, 1985).

Applying this procedure to the data in Table 1 leads to the estimate $\hat{\tau}^2 = 0.415$ with 90% confidence bounds (0.178, 1.476). For the data in Table 2, we have $\hat{\tau}^2 = 0.072$ with 90% bounds (0.043, 0.342). This confirms the finding of Diaconis and Efron that the Table 2 data are closer to independence than those in Table 1, in spite of the fact that Table 2 has the larger χ^2 statistic. However, judging from the overlap in the confidence intervals for τ^2, the two tables are more comparable in terms of their degree of departure from independence with the alternative procedure we suggest. A disadvantage of our approach, of course, is the need for iterative calculations to estimate τ^2.

Diaconis and Efron's χ^2 based procedure and the alternative random effects model suggested here both fail to account for the obvious structure that is present in Table 2 but lacking from Table 1. A more meaningful appraisal of the degree of departure from independence in Table 2 could be made by comparing the mean number of children for families of different income levels, for example, by fitting separate Poisson distributions to each column. We underscore the authors' comment that "more elaborate structural models $\cdots$ often can give deeper insight into the data," and would suggest further that such structure should generally be accounted for even in routine approaches to data analysis.

REFERENCES

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

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Appl. Statist.* **33** 38–44.

McCullagh, P. (1985). On the asymptotic distribution of Pearson's statistic in linear exponential-family models. *Int. Stat. Rev.* **53** 61–67.
McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.
Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.
Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61** 439–447.
Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31** 144–148.

Department of Biostatistics
University of Washington
Seattle, Washington 98195

James M. Dickey[1]

State University of New York at Albany

1. General. The paper under discussion by Diaconis and Efron (1985) (DE) is impressive and stimulating. I would like to bring forward here a few general questions to which it gives rise and then take a brief look at coherent inference for the models employed by DE.

Toward the end of Section 1, DE state that their "goal is to extend the usefulness of χ^2." I would wish to ask first, how should χ^2 be used? On the one hand, inferences based on tail areas, rather than probability densities or masses, are not coherent. On the other hand, tail areas are naturally interesting facts about the data (and about other nonoccurring data values). I do not know the best answer to this question and I would personally prefer to keep both kinds of tools in our kit.

Recall that the coherent inference in favor of a hypothesis H versus its alternative $\bar{H}$ is given by the Bayes factor $B(H, \bar{H})$ (Jeffreys, 1939; Good, 1950; Edwards, Lindman and Savage, 1963; Dickey and Lientz, 1968). This is the ratio of the coherent posterior odds $P(H \mid \mathbf{x})/[1 - P(H \mid \mathbf{x})]$ to the prior odds $P(H)/[1 - P(H)] > 0$. This ratio depends on the data $\mathbf{x}$, but not on the prior odds, so it serves as a sufficient report of the data for inference regarding H. The Bayes factor also equals the ratio of predictive densities, $B(H, \bar{H}) = p(\mathbf{x} \mid H)/p(\mathbf{x} \mid \bar{H})$, each a function of the respective conditional prior distribution, $p(\mathbf{x} \mid J) = \int p(\mathbf{x} \mid \pi)\, \mathrm{d}P(\pi \mid J)$, $J = H, \bar{H}$. The dependence on conditional uncertainty may necessitate a tabular or graphical report of the Bayes factor (Dickey, 1973).

Technical point. In the case of a sharp hypothesis defined by a point value of a constraining parameter, H: $\eta = \mathbf{0}$, where $\eta \equiv \eta(\pi)$, it is tempting to use a single joint density $g(\pi)$ to specify both of the conditional prior distributions, $p(\pi \mid \bar{H}) = g(\pi)$ and $p(\pi \mid H) = g(\pi \mid \eta = \mathbf{0})$, where $g(\pi \mid \eta)$ is a lower-dimensional density obtained in the usual way by conditioning in $g(\pi)$. For one thing, Savage's

[1] Research supported by U.S. National Science Foundation Grant MCS-8301335.