our simplified model; the posterior mean of τ was now 0.60. This enabled us to obtain a simple analysis for the full $12 \times 8 \times 5$ table and to investigate the associations between entry levels and final grades.

My overall conclusion is that most observed contingency tables possess intrinsically individualistic structures which should not be concealed by unduly constraining alternative hypotheses in advance. Diaconis and Efron are, however, taking us in a good perceptive direction which should yield fresh advances in the future.

Acknowledgements. The collaboration of Melvin R. Novick, Stephen B. Dunbar, and Ming-Mei Wang on the analysis of the Marine Corps data is gratefully acknowledged.

REFERENCES

BATEMAN, G. I. (1950). The power of the χ^2 index of dispersion test when Neyman's contagious distribution is the alternative hypothesis. *Biometrika* **37** 59–63.

LEONARD, T. (1977). A Bayesian approach to some multinomial estimation and pre-testing problems. *J. Amer. Statist. Assoc.* **72** 869–874.

LEONARD, T. (1982). Comment on the paper by Lejeune and Faulkenberry. *J. Amer. Statist. Assoc.* **77** 657–658.

LEONARD, T. and NOVICK, M. R. (1985). Bayesian full rank marginalization for two-way contingency tables. To appear in *J. of Educ. Statist.*

PAUL, S. R. and PLACKETT, R. L. (1978). Inference insensitivity for Poisson Mixtures. *Biometrika* **65** 591–601.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706

PETER MCCULLAGH AND DARYL PREGIBON

AT&T Bell Laboratories

Some years ago, a well-known Irish newspaper carried a series of advertisements by an eccentric entrepreneur known as "The Brother," offering correspondence courses in the art of "periastral peregrinations on the aes ductile," more commonly, but less accurately, known as tight-rope walking. Despite the incentive of generous course credit, the University of the Air, as it was known, had few registered students and no known graduates. In the present paper, Diaconis and Efron give a superb exhibition of The Brother's singular art in its metaphorical form, by attempting to dance on two ropes at once—and almost succeeding!

Diaconis and Efron have chosen, quite sensibly, not to argue against modelling departures from independence, noting that such models can often give deeper insights into the data. Instead, they emphasize the common χ^2 statistic, here denoted by X^2, as "an effective device for preliminary data analysis, particularly when the statistician has many two-way tables under review." This point of view seems difficult to comprehend because the most common and compelling objection to the use of X^2 in applications is that it gives no information regarding the nature or direction of departures from independence. In Table 2, for example, the observed value of $X^2 = 568.6$ gives no indication that nearly all the variation can be accounted for by the observation that the approximate average numbers of children per family in the four income groups are 0.89, 0.94, 0.76 and 0.60, i.e., that the higher income groups have substantially fewer children per family than the lower two income groups. Similarly, in Table 1, the observed value of $X^2 = 138.3$ gives no indication that brown-eyed people tend to have darker hair, hazel- and green-eyed people tend to have intermediate shades and blue-eyed people tend to have fair hair. These are the kinds of natural systematic effects we are all familiar with and it would be an extraordinary analysis indeed that chose to ignore them, particularly at the exploratory stage.

Just when is a statistician likely to be overwhelmed with two-way tables and to have no alternative to using X^2? The most likely scenario we envisage occurs with a large multi-way table, where the statistician chooses to analyse the data SPSS/CROSSTABS style, displaying two-way tables of a pair of variables stratified by all possible combinations of the remaining variables. Putting aside our strong preference for fitting *structural* models to data of this type, we question the relevance of the authors' proposals, which rely heavily on asymptotic arguments for justification. In particular, an originally large sample size quickly disappears as one carves a multi-way table into many two-way tables. The only way large sample sizes are preserved is by *collapsing* the multi-way table into two-way tables. This will preserve the original sample size for each two-way table, but as illustrated by Simpson's Paradox, at the risk of obtaining misleading conclusions.

Our view is that, in applications, it is the departures from independence that are of most interest and it is therefore essential first to identify and isolate the major systematic effects, and only then, to assign the remainder to residual or unexplained variation. Even then, the residual variation is likely to exceed that predicted by Poisson or multinomial sampling. In such cases, the common practice is to introduce a dispersion parameter σ^2 estimated by

$$\tilde{\sigma}^2 = (\text{residual } X^2)/(\text{residual d.f.})$$

(Finney, 1971, page 471), and this factor plays the same role as the residual mean square in ordinary linear models. The practice of accommodating extra variability by treating σ^2 as a variance inflation factor can be justified quite generally, not just for two-way contingency tables, but also for arbitrary log-linear models and even for continuous responses (Wedderburn, 1974; McCullagh and Nelder, 1983, Section 6.3). This justification presupposes that all major systematic effects have been accounted for by the model. Thus, to the extent that the authors consider only models of independence, their choice of examples seems particularly unfortunate. The advantage of the parameterization σ^2 over the authors' "effective sample size" $\nu = n/\sigma^2$ is that the interpretation of σ^2, as the ratio of the variance of the data to the corresponding Poisson or multinomial variance, is independent of the sample size. Alternatively, one could use "the effective sample ratio," $\nu/n = \sigma^{-2}$.

Another point concerns the dichotomy between the authors' usage of Pearson's statistic on the one hand, and the Kullback-Leibler distance or deviance statistic, D, on the other. At this point, the authors sit rather uncomfortably on the fence, giving all probability calculations in terms of X^2 and all justifications in terms of D. Readers are invited to think of the two statistics as "close cousins," and as equivalent to a first order of approximation. This equivalence holds only if all cell counts are sufficiently large, but, in fact, overdispersion can and does occur even when the data are sparse. For this reason, the so-called improved formulae (5.20) and (5.22) seem to be a backward step. The principal justification for (5.20) is that it is the maximum likelihood estimate in the family (5.19). However, this is a circular justification because the family (5.19) was constructed to have D as a component of its sufficient statistic. The estimate based on X^2, on the other hand, can be justified independently of (5.19), behaves sensibly even where the data are sparse but extensive, and has the additional property that it is approximately independent of all estimated structural parameters. In this sense, $\tilde{\sigma}^2$ is the natural generalization of the residual mean square in ordinary linear models.

Finally, to describe the volume test as a test for independence, as the authors have done at the start of Section 2, is surely a gross abuse of terminology. Probability calculations for this test are made under the mathematically attractive but statistically uninteresting hypothesis of equi-probable lattice points in $\mathscr{L}$. The "acceptance region" based on small values of Pearson's statistic is concentrated near the independence surface, $\mathscr{I}$, for which $X^2 = 0$, but this alone does not make the volume test a test of independence. On a more technical note, if we were to take the uniformity hypothesis (2.13) seriously, this would imply that the marginal frequencies were approximately equal and the most powerful tests of uniformity with independence as alternative are based on the marginal totals alone. To avoid this embarassment, the authors have chosen to use X^2 as test statistic for testing unconditional uniformity, justifying their choice on the grounds that X^2 is appropriate for the entirely different hypothesis of conditional uniformity.

One must at least admire the careful balance required for two statisticians to sit, however precariously, on so many fences at once. The Brother would indeed be pleased!

REFERENCES

FINNEY, D. J. (1971). *Probit Analysis*, 3rd ed., Cambridge University Press.

MCCULLAGH, P. and NELDER, J. A. (1983). *Generalized Linear Models*. Chapman and Hall, London.

WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61** 439–447.

AT&T BELL LABORATORIES
600 MOUNTAIN AVENUE
MURRAY HILL, NEW JERSEY 07974