DONALD A. PIERCE[1]

Oregon State University

As I understand Section 5, it provides strong evidence that the χ^2 statistic is a suitable measure of overdispersion only when its approximate effect is to increase $V(\mathbf{x})$ to $(1 + c)V(\mathbf{x})$. This is an important result which is helpful in judging the general usefulness of such methodology, which I will make some attempt to do here. From the viewpoint of models for overdispersion that appeal to me, I will question the appropriateness of using χ^2 as a measure of overdispersion in contingency tables when the marginal totals are quite heterogeneous, as in Table 2 of this paper.

The authors are careful to point out that this is being suggested only for a rough preliminary analysis. In view of this, I am probably scrutinizing the method rather severely. My aim is not to be critical of their results, but rather to complement them with a way of thinking about the "roughness" question.

Two-stage models, such as that described near (4.15), are very appealing to me. Since these are only mentioned in the preliminary heuristics, I may be putting more emphasis on such a formulation than the authors intend. However, it is hard for me to get down to what overdispersion really means without explicit models such as this. Further, I believe that they are not necessarily Bayesian in nature, any more than is the ordinary randomized block model. It may be helpful to emphasize briefly the approximate connection between the results of Section 5 and the two-stage models. If overdispersion is thought of as the result of a random perturbation on the mean parameter, as exemplified by the discussion near (4.15), and if its marginal result is to rescale $V(\mathbf{x})$, as noted by (5.23), then the covariance matrix of the perturbation in the mean parameter must also be a multiple of $V(\mathbf{x})$.

For the contingency table case, the implicit assumption in measuring overdispersion by χ^2 is given by (4.15); that $\boldsymbol{\pi}$ is randomly perturbed from $\hat{\boldsymbol{\pi}}$, with covariance matrix Σ. Here Σ is the covariance matrix of the Fisher-Yates distribution of $\mathbf{x}$ given the marginal totals. Consider a case where the marginal totals are as in Table 1 of this comment, which is not intended to be special other than that the totals are quite heterogeneous. The nonparenthetical cell entries are $\hat{\pi}_{ij}$ and the parenthetical ones are the variance elements of Σ.

In my judgement this is a strange enough model for dispersion in $\boldsymbol{\pi}$ to warrant concern. The (1, 2) and (2, 1) elements have the same $\hat{\pi}_{ij}$ but very different variances. The (2, 1) and (2, 2) elements have the same variance but very different $\hat{\pi}_{ij}$. There is, of course, no theoretical reason for a simple model for the coefficient of variation of π, but for a rough and ready model I think that assuming it to be constant would have considerable merit. Unfortunately, this would apparently lead to a more complex analysis.

Another example, which requires a generalization of the analysis of Section 5 which the authors might or might not endorse, is given by a typical binomial

[1] This research was supported by NSF Grant DMS 8403665.

TABLE 1
Scaling of overdispersion for a table with quite heterogeneous margins

.01 (.008)	.04 (.008)	50
.04 (.026)	.16 (.026)	200
.15 (.030)	.10 (.030)	750
200	800	1000

regression model. (The generalization corresponds to incorporation of the possibly unequal binomial sample sizes.) Finney (1971, page 70) suggested use of the χ^2 statistic to measure overdispersion in the probit model case, and perhaps more importantly, the rescaling of the binomial-based variance of parameter estimates by the mean chi-squared, to allow for overdispersion. Writing $\mathbf{r}_i \sim \text{Binomial}(m_i, p_i)$, analysis like that given here suggests that this corresponds, in terms of a two-stage model, to assuming that $V(p_i) \propto \hat{p}_i(1 - \hat{p}_i)/m_i$. This may be reasonable as a rough model when the m_i are about the same size, which was what Finney had in mind, but otherwise it seems questionable. Even when the m_i are equal and the p_i are all less than 0.5, it is different than a constant coefficient of variation model, $V(p_i) \propto \hat{p}_i^2$. Some methodology following up on this is given by Pierce and Sands (1975), Williams (1982), and for the Poisson setting by Breslow (1984).

A one-parameter model of overdispersion necessarily requires a rather strong model. It is unlikely that from enumerative data, without enormous sample sizes, one will be able to verify this model in any detail from the data. The undisputed virtue of what I take to be the implied model of this paper is that it leads to a very simple analysis, and I put high priority on this. For the highly unbalanced situations on which I have focused, it would be nice to understand more clearly the end results of measuring heterogeneity by the chi-squared statistic. For more balanced situations, I think the method may be very useful. However, I think that this is to some extent because the choice of a model for overdispersion then becomes largely irrelevant; most models, whether more generally realistic or not, converge to the same thing.

REFERENCES

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

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

PIERCE, D. A. and SANDS, B. R. (1975). Extra-Bernoulli variation in analysis of binary data. Technical Report No. 46, Dept. of Statistics, Oregon State University.

WILLIAMS, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31** 144–148.

DEPARTMENT OF STATISTICS
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331