The Annals of Statistics
1983, Vol. 11, No. 2, 370–371

H. MALCOLM HUDSON

Macquarie University

I shall restrict my comments to multiparameter estimation of Poisson means. The discussion is then directly applicable to Monte Carlo simulation (for an example, see Efron and Morris, 1975) and to contingency table data.

Log-linear models may be used in tables to smooth cell counts. Consider, for instance, a two-dimensional contingency table in which the counts X_{ij} have independent Poisson distributions with means μ_{ij}, for $i = 1, \cdots, I$, $j = 1, \cdots, J$. Any simple model constraining means, such as $\mu_{ij} = \alpha_i\beta_j$, for $i = 1, \cdots, I$, $j = 1, \cdots, J$, with $\beta_1 = 1$, allows one to estimate the cell means $\{\mu_{ij}\}$ with increased precision, gained through the reduction in the number of unknown parameters, in this case from IJ to $I + J - 1$. When cell means do not conform exactly to a constrained model, compound Bayes estimators based on the ideas of Robbins (1955) and Maritz (1969) may prove more suitable. Related methods evolve from Stein's 1961 result concerning estimation of means of Normal random variables and from his subsequent introduction of the identity which underlies the authors' work. Stein's methods are effective when means do not depart substantially from a simple model. The estimators adapt automatically to such deviations. Thus the theory advanced in this paper has spinoffs related to resistance of estimation procedures to model misspecification.

The criterion of risk minimaxity adopted by the authors is seen in this light as an extreme form of resistance, but the efficiency of the estimator when the model is appropriate is another important consideration. The development of efficient, resistant procedures for Poisson data is difficult, even when the corresponding methods for Normally distributed data are well understood. The subject remains a stimulating area for research.

The authors are thus to be complimented on formulating a comprehensive theory which both unifies many previous results for differing loss structures and simplifies the determination of risk properties of a stipulated estimator. Theorems 3.1 and 3.2 appear to be the key results for the Poisson distribution (though, as noted in the paper, they apply also to other cases). The ability to assess estimators which shrink towards a set of prior values is important, and gives insight into the behaviour of similar estimators in which the "prior" values are data determined. (The risk properties may not change much when one substitutes stable data determined values for the constant lambdas). Therefore, the shrinkage towards the very special origin, $\lambda_1 = \cdots = \lambda_p = X_{(n)}$, in Theorems 4.1 and 4.2 makes the results of Section 4 appear less generally applicable.

Although Theorems 3.1 and 3.2 add to our toolkit, the paper fails, in my opinion, to deliver much of substance in the way of applications. There seems to be little evidence that guesswork has been eliminated. The simulation results of Table 2 reveal that the estimators considered have very little to offer. If variables X_i represent the number of events of different kinds occurring in parallel Poisson processes in a unit interval of time, the risk reduction is equivalent to no more than an extension by 20% of the time period in which counts accumulate. This is especially poor when it is realized that shrinking is being directed to almost perfect prior estimates when δ_1 and δ_3 are being used. The other estimator δ_2 also shows little improvement and is not suited to shrinkage towards an arbitrary set of initial estimates. The failure of δ_2 to achieve material risk reduction can be illustrated by applying it to the data of Hudson and Tsui (1981, Section 4). Whilst a transformation of the data permits normal based methods to be used to achieve a squared error reduction of 46%, the estimator δ_2 achieves a reduction near 0%! By contrast, Hudson and Tsui introduced an estimator which is uniformly superior to X in the approximate sense of Hudson (1981)—in which $\mathscr{D}(X)$ is replaced by $\mathscr{D}^*(X)$—and showed that it achieves a squared error reduction of 54%. This approximation is satisfactory in many circumstances, and particularly when the means are large. It then provides a flexible, and much more efficient, class of estimators than those considered by the authors.

Received August 1982.