I. J. Good[1]

Virginia Polytechnical Institute

One can interpret one aspect of this very interesting paper as the testing of two null hypotheses. These may be called $H(\infty)$ ("independence" of rows and columns, called H_1 by the authors) and $H(1)$ (called H_0 by the authors) where $H(k)$ corresponds to a symmetric Dirichlet prior with parameter k. In the model adopted, for example, in Good (1983) and in my joint work with Crook, a mixture of all these symmetric Dirichlet prior was used, the mixture being regarded as a hyperprior density $\phi(k)$. (The idea dates back in part to Good, 1965, but was not developed much there.) Thus $H(k_0)$ can be regarded as the hypothesis for which $\phi(k)$ is the Dirac delta function $\delta(k - k_0)$. For multinomials, tests of $H(k)$ for $k = 1$, ½ (Jeffreys, 1946) and $1/t$ (Perks, 1947) were given in that paper, as well as a test for $U_k H(k)$ (the hypothesis that *some* symmetric prior is appropriate). Applications were made by Good (1983) to the problem of scientific induction, a point that could easily be overlooked because induction was not mentioned in the title of that paper.

The authors also interpolate a continuous infinity of hypotheses between $H(1)$ and $H(\infty)$, but in a manner different from that used by Crook and myself. Our interpolation was effected by using hypotheses $H(k)$ $(k > 0)$. For $k < 1$ these extrapolate beyond $H(1)$. It should be possible, but perhaps difficult, to determine which interpolation seems to give a better description of a given contingency table. Perhaps a very large sample size would usually be necessary to make this discrimination.

As a spin-off from the statistical work, the authors propose a new approximate formula for the famous old combinatorial problem of enumerating rectangular arrays with given margins. It might not be easy to decide under what circumstances their approximation is better than the ones proposed by Good (1976) and by Good and Crook (1977). My "ad hoc" improvement that the authors mention was an empirical adjustment to allow for the "roughness" of row or column totals. A really satisfactory formula should be algebraically symmetrical with respect to rows and columns thus avoiding giving two different answers. I wonder whether the authors have any further comments on this point.

The question of how much evidence regarding "independence" is contained in the marginal totals alone is of both logical and historical interest because Fisher's "exact test" for 2×2 tables ignored this evidence by conditioning on the marginal totals. The authors say, towards the end of their Section 3, that Crook and I "do not observe a very large difference between conditional and unconditional inferences (for this problem). This is because they use a variant of the uniform prior (2.11) which eliminates most of the supposed information in the margins." I think this comment requires some elaboration if it is not to be nearly tautological. We certainly selected a Bayesian model (the simplest usable one?) in which the row totals *alone*, or the column totals alone, convey no evidence against the null

[1] Supported by NIH Grant GM 18770.

hypothesis (of independence). (The history of this aspect of the analysis is given by Good, 1983, page 199.) This seems to me to be essential in principle for a general-purpose test. But in what other nontautological sense is the information (or evidence) largely eliminated? We thought we had made a substantive theoretical contribution by showing that a natural Bayesian model did lead to an approximate justification of Fisher's judgement. Conversely, we thought that our numerical results would also be regarded as support for our Bayesian model by anyone who regards Fisher as "Big-Daddy" (to quote Oscar Kempthorne's affectionate appelation). It is useful to decrease the gap between Bayesian and non-Bayesian approaches in some circumstances although it cannot always be done. Our work was intended to exemplify this gap-closing activity.

The authors cite the unpublished note by Good and Crook (1980) which gave examples of "FRACT," the Bayes factor arising from row and column totals, for the especially important but modest 2×2 contingency table. Some of the results were reprinted in Good (1983, page 200) which is more accessible.

I agree very much with the authors' warning at the end of their Section 4, against over-reliance on the model if the hypothesis of independence is rejected. In fact, quite generally, whether an assumed model is Bayesian or not, if a null hypothesis is rejected it is advisable to switch gear either to a data-processing or a scientific mode, especially when the model has many parameters, and in particular for large contingency tables. Even for the 2×2 tables, a peculiar-looking diagonal might be suggestive.

I think the author's θ might be regarded as a hyperparameter, analogous to our k. If so, it would be possible to use a mixture of the authors' hyperparametrized family corresponding to our use of a mixture of Dirichlet priors. In this manner, a more thoroughly Bayesian version of their work might be attained.

REFERENCES

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* MIT Press, Cambridge, MA.

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Ser. A* **136** 453–461.

PERKS, W. (1947). Some observations on inverse probability including a new indifference rule. *J. Inst. Actuar.* **73** 285–312.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNICAL INSTITUTE
AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA 24061