

Comment on Kannemann's "The Exact Evaluation of 2-way Cross-classifications"

PIETER M. KROONENBERG and ALBERT VERBEEK¹

University of Leiden and Netherlands Central Bureau of Statistics

Abstract

KANNEMANN (1982 a, b) is criticized on several points, primarily using results available in the literature prior to his papers.

In two papers KANNEMANN (1982 a, b) presents algorithms for the exact distribution of the probability and test statistics in 2-way contingency tables with fixed margins. In his first paper he states that he will not review the literature on the subject as he purports it to be of a purely theoretical nature and not relevant for applied statisticians, researchers, and experimenters.

This disregard for the literature is, however, unfortunate as an essentially identical algorithm to KANNEMANN (1982b) has been proposed by HANCOCK (1975). HOWELL & GORDON (1976) and CANTOR (1979) made a number of small improvements on Hancock's algorithm. More efficient algorithms were, however, already available at that time, viz. BOULTON (1974), AGRESTI & WACKERLY (1977), and BAKER (1977). The latter has also provisions for handling one or two fixed margins.

At present the state of the art has even been further developed by PAGANO & TAYLOR-HALVORSON (1981), MEHTA & PATEL (1983), and VERBEEK, KROONENBERG, and KROONENBERG (1983). A virtually complete survey of the field is given by VERBEEK & KROONENBERG (1985).

With respect to closed form formulae for the number of possible tables given the row and column margins, they still do not exist, but good approximation formulae have been developed by BOULTON & WALLACE (1973), GOOD (1976), and especially GAIL & MANTEL (1977).

As to the use of the two-parametric gamma distribution as a better approximation than the one parametric chi-square, it should be noted that YARNOLD (1970) presents extensive proof against the superiority of the gamma distribution.

KANNEMANN's claim that he can analyse 50×50 -tables and three-dimensional

¹ The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics (CBS).

tables by complete enumeration seems untested and unlikely. The smallest family of 50×50 -tables with fixed margins is the one with 50 observations, each marginal frequency being 1. It contains $50! > 10^{64}$ tables, and with current processing speeds this defies enumeration. For three-dimensional tables the smallest families are larger than the third power of families of two-dimensional tables. Only for very small tables and small numbers of observations, this procedure is practical.

References

- AGRESTI, A. and D. WACKERLY, 1977: Some Exact Conditional Tests of Independence for $R \times C$ Cross-classification Tables. *Psychometrika* **42**, 111–125.
- BAKER, R. J., 1977: Algorithm AS 112. Exact Distributions Derived from Two-way Tables. *Appl. Statist.* **26**, 199–206; *Corr.* 1978: **27**, 109; *Corr.* 1981: **30**, 106–107.
- BOULTON, D. M., 1974: Remark on Algorithm 434. *Commun. ACM* **17**, 326.
- BOULTON, D. M., and C. S. WALLACE, 1973: Occupancy of a Rectangular Array. *Computer J.* **16**, 57–63.
- CANTOR, A. B., 1979: A Computer Algorithm for Testing Significance in $M \times K$ Contingency Tables. Fifth Proceedings of the Statistical Computing Section of the American Statistical Association. ASA, Washington, DC.
- GAIL, M., and N. MANTEL, 1977: Counting the Number of $r \times c$ Contingency Tables. *J. Amer. Statist. Assoc.* **72**, 859–862.
- GOOD, I. J., 1976: On the Application of Symmetric Dirichlet Distributions and their Mixtures to Contingency Tables. *Ann. Statist.* **4**, 1159–1189.
- HANCOCK, T. W., 1975: Remark on Algorithm 434. *Commun. ACM* **18**, 117–119.
- HOWELL, D. C., and L. R. GORDON, 1976: Computing the Exact Probability of an r by c Contingency Table with Fixed Marginal Totals. *Behavior Res. Meth. Instrum.* **8**, 317.
- KANNEMANN, K., 1982a: The Exact Evaluation of 2-way Cross-classifications: An Algorithmic Solution. *Biom. J.* **24**, 157–169.
- KANNEMANN, K., 1982b: The Exact Evaluation of 2-way Cross-classifications: Sequel. A Fugal Algorithm. *Biom. J.* **24**, 679–684.
- MEHTA, C. R., and N. R. PATEL, 1983: A Network Algorithm for Performing Fisher's Exact Test in $r \times c$ Contingency Tables. *J. Amer. Statist. Assoc.* **78**, 427–434.
- PAGANO, M., and K. TAYLOR-HALVORSON, 1981: An Algorithm for Finding the Exact Significance Levels of $r \times c$ Contingency Tables. *J. Amer. Statist. Assoc.* **76**, 931–934.
- VERBEEK, A., and P. M. KROONENBERG, 1985: A Survey of Algorithms for Exact Distributions of Test Statistics in $R \times C$ Contingency Tables with Fixed Margins. *Computational Statistics and Data Analysis* (in press).
- VERBEEK, A., P. M. KROONENBERG, and S. KROONENBERG, 1983: User's Manual to Fisher. WEP Reeks, WR 83–49-RP. Department of Education, University of Leiden, The Netherlands.
- YARNOLD, J. K., 1970: The Minimum Expectation in X^2 Goodness of Fit Tests and the Accuracy of Approximations for the Null Distribution. *J. Amer. Statist. Assoc.* **65**, 864–886.

Manuscript received: June, 20, 1984

Authors' address:

Dr. P. M. KROONENBERG
 Dr. A. VERBEEK
 University of Leiden
 Department of Education
 Stationsplein 12
 P. O. Box 9507
 2300 RA Leiden
 The Netherlands