Comment: Exact Inference in Multidimensional Tables

Svend Kreiner

1. INTRODUCTION

Professor Agresti's authoritative and stimulating paper is filling a lacuna in the literature on exact conditional inference, and I find that he has done statistics, in general, and this reader, in particular, a great service in presenting this comprehensive survey on exact methods. I would just like to bring up a few additional points on the topic of exact inference in multidimensional tables.

It is correct that many problems remain to be solved before we have real exact inference in multidimensional tables. The situation is not completely hopeless, however. Limited exact inference is a practical possibility today, and it may have an important role to play in serious loglinear model building. To that end, however, it is necessary that we change our ideas about how strategies for loglinear modeling should work.

We have to distinguish between two different problems: (1) tests for higher order interactions and (2) tests for conditional independence. There seems to be no practical solution to the first problem in the predictable future. For the second problem, however, we have some solutions through Monte Carlo sampling. In connection with collapsibility, exact conditional tests may first of all be used for a larger number of situations than one perhaps would think possible at first glance. Second, they generalize without special problems to exact goodness of fits for decomposable loglinear models.

2. TESTING CONDITIONAL INDEPENDENCE IN MULTIDIMENSIONAL TABLES

Consider first a five-dimensional table, n_{ABCDE}, and the problem of testing conditional independence of two variables given the rest: $A \perp E \mid BCD$.

If nothing is assumed on a model for the five-way table, this is equivalent to fitting the loglinear ($ABCD$, $BCDE$) model against a saturated alternative. Under this assumption we may, without loss of information, combine the CDE variables into one stacked variable ($Z = B * C * D$) and treat the problem as if it were a problem of inference in a three-way table. The hypothesis we consider is the third of the hypotheses discussed by Agresti for $I \times J \times K$ tables. Exact conditional tests by Monte Carlo sampling of tables from the product multiple hypergeometric distribution is no problem with today's computing power. It works even for very large tables on IBM-PC compatible microcomputers.

We assumed nothing on a model for the five-way table. We may, therefore, think of the test of conditional independence as a non-parametric test. One is well-advised to be concerned with the power of chi-squared statistics in this case and use of ordinal statistics (e.g., the partial Goodman–Kruskall coefficient is strongly recommended whenever ordinal statistics are appropriate). We notice also, however, that there is no practical obstacle for tests of (AZ, ZE) against parametric alternatives for ordinal variables or models assuming constant AB-association (AE, AZ, ZE). These tests will, of course, be considerably more time-consuming than tests against the saturated alternative, because calculation of tests statistics requires iteration for each sampled table. That, however, is a problem where we can count on computer science for a solution, and it should not concern us here.

If a test against a nonsaturated model is required, we have to rely on collapsibility to guide us toward exact conditional tests. Assume for instance that we want to test conditional independence of A and E against a loglinear model (AC, CD, BD, ABE). The conditional distribution of the complete table given the AB, AC, CD, BD and BE marginals may seem inaccessible at first sight. Collapsibility properties implies that the test of (AB, AC, CD, BD, BE) against (AC, CD, BD, ABE) is equivalent to a test of (AB, BE) against the saturated alternative in the ABE marginal, because both models are collapsible in the sense discussed by Asmussen and Edwards (1983). Not only are the interaction parameters of interest, the same in the complete and marginal model, but estimates and test statistics will be exactly the same whether or not we calculate them in the complete or the marginal table. And, finally, the conditional probability of a test statistic defined on the ABE marginal, $T(n_{ABE})$, given the sufficient marginals,

Svend Kreiner is a Statistician and Senior Researcher, The Danish Institute for Educational Research, 28 Hermodsgade, DK-2200 Copenhagen North, Denmark.