this clear, I will express the same recursions in network terms.

The network consists of nodes and arcs over $k + 1$ stages labeled $0, 1, \ldots, k$. The nodes at stage j are ordered pairs of the form (j, m_j), where $m_j \in \Lambda_j$. The set $S(j, m_j)$ defines the successor nodes to the node (j, m_j), while the set $P(j, m_j)$ defines its parent nodes. If we start with an initial node $(0, 0)$ at stage 0, and apply (2.1) systematically to all the nodes created at each stage, we end up with a single terminal node (k, m_k) at stage k. Each successor to node (j, m_j) is a node of the form $(j + 1, m_{j+1})$. It is connected to (j, m_j) by an arc of length, $w_{j+1}(m_{j+1} - m_j)$ and probability $n_{j+1}!/(m_{j+1} - m_j)!(n_{j+1} - m_{j+1} + m_j)!$. A path through the network is a sequency of directed arcs connecting the initial node $(0, 0)$ to the terminal node (k, m_k). Its length is the sum of lengths, and its probability the product of probabilities, of the arcs constituting the path. Through this specification, each path through the network represents one and only one table $x \in \Gamma$. Its length is given by (1.1) and its probability is given by (1.2). The problem of generating the distribution (1.6) is now equivalent to generating the distribution of the lengths of all paths through the network. The set $\Omega(j, m_j)$ represents the distribution of the lengths of all the paths from node $(0, 0)$ to node (j, m_j). The recursions (2.4) and (2.5) amount to expressing the distribution of lengths at node (j, m_j) in terms of the distributions of lengths at its parent nodes $P(j, m_j)$. Also, computing $SP(j, m_j)$ and $LP(j, m_j)$ amounts to computing the lengths of the shortest and longest paths, respectively, from node (j, m_j) to the terminal node (k, m_k). These may be obtained by backward induction on the network (Mehta, Patel and Senchaudhuri, 1992) or by more formal integer programming theorems (Joe, 1988; Agresti, Mehta and Patel, 1991).

In our research papers, although not in this discussion, the network representation of a computational problem has always preceded its algebraic representation. It is certainly elegant to express the computational problem directly in terms of recursions like (2.4) and (2.5). However, it is not so easy to gain the necessary insight to write out the recursions in the first place. Nor is it clear how one implements them on a computer once they are written down. We regard the network approach as a general technique for deriving these recursions, guiding us in selecting appropriate data structures for computer implementation, and solving the necessary integer programming problems. We have used this approach for $2 \times k$ tables, stratified $2 \times k$ tables, $r \times c$ tables and logistic regression.

In summary, this discussion has attempted to show, through a detailed dissection of the $2 \times k$ problem, that the basic ingredients of an efficient numerical algorithm for permutational inference comprise of, recursive generation of the distribution of the test statistic, good data structures for storing intermediate distributions through all stages of the recursion and the use of integer programming to generate a truncated distribution. The network paradigm is a useful aid for carrying out these steps. In particular, forward processing of the network is a general way to conceptualize and implement complicated recursions, whereas backward induction on the network is a general way to solve the integer programming problem.

Comment

Samy Suissa

Professor Agresti must be congratulated for this long-awaited review of the principal issues and methodology surrounding exact inference in contingency tables. Since Fisher proposed his exact method of analysis for the 2×2 table in 1934, the amount of literature produced on the subject and the resulting debates have reached immeasurable proportions. (Yes, this pun intended!) Whether dealing with the accuracy of various asymptotic techniques in small sample situations, the diverse possible factors of correction for continuity, or the conditional, unconditional and Bayesian alternatives, the ensuing research has definitely contributed to our increased knowledge of the situation and has motivated imaginative developments in computing algorithms. Professor Agresti pre-

Samy Suissa is an Associate Professor, Departments of Epidemiology and Biostatistics and Medicine, McGill University, and Research Scholar in the Division of Clinical Epidemiology at the Montreal General Hospital. Mailing address: Purvis Hall, 1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada.

sents us with an elegant synthesis of this literature and guides us masterfully through the maze of research generated by the topic.

A noticeable observation from this survey is that the conditional approach is the preferred method of inference among investigators, and for very valid reasons. Indeed, it is based on rigorous theoretical considerations and results in well-behaved manageable and computationally feasible distributions, thus making the approach increasingly practical, for even complex designs. In addition, it has been the recipent of recent remarkable advances in computing algorithms that have expanded the scope of application of exact inference and highlighted its importance vis-à-vis its relatively inaccurate asymptotic counterparts. Nevertheless, in spite of these tremendous theoretical properties and practical possibilities, the conditional approach leaves open some considerations of a significant practical nature, some of which are at times disturbing by the dilemmas they pose. I will address three of these; namely, the questions of power, two-sided significance testing and interval estimation.

Although the attained significance level (p-value) is unquestionably preferable to a nominal significance level, such as the prominent and at times infamous "$\alpha = 0.05$," it is unfortunately not possible to use it when designing a study on the basis of its power. At the design phase, maximum acceptable levels of the magnitude of type I and type II errors must be set a priori to determine the sample size necessary to conduct this investigation. The common practice is to use α at the design stage and to report the p-value at the data analysis step. Thus, α becomes essential when power is discussed.

In evaluating the power of the conditional test, the unconditional power function is used, since the observed value of the statistic on which conditioning is performed is not known at this point of a trial. A feature of this unconditional power function is that it is averaged over all conditional critical regions that are inherently highly discrete, most particularly for the small sample situations that call for exact methods. This feature, compounded with the fact that exact unconditional p-values are generally lower than the corresponding conditional ones, has led to important differences in power comparisons between the two. These power evaluations and exact sample size determinations have been generated for the unconditional approach in both the independent binomial and matched pairs 2×2 tables (Suissa and Shuster, 1985, 1991), and contrasted with the corresponding conditional ones. A result of these investigations is the greater power of the exact unconditional approach, uniformly across all situations considered.

Although the differences found between the conditional and unconditional sample sizes do not at first appear to be of much practical significance, they turn out to be very much so in real life studies, particularly in the small sample situations for which exact inference is indicated. As an example, the McGill Cancer Centre was planning a trial evaluating a special feeding formulation designed to reduce the risk of acute radiation damage to the small bowel in patients with bladder cancer. This damage from radiation greatly complicates the conventional subsequent surgical treatment of this cancer, thus the need for testing this dietary preventive measure. The logistical difficulties involved in making this a multicenter trial mitigated against this strategy, and the study was therefore conducted in a single center. Of course, the inconvenience of this latter approach is that the pool of study subjects is now smaller, thus highlighting the sample size determination operation. To detect the large anticipated effect in this study, namely a rate decrease of 0.40 from the current damage rate of 0.45, with $\alpha = 0.05$ and 80% power, the trial would require 17 subjects per arm under the exact conditional approach or 13 subjects per arm under the exact unconditional approach. The additional eight required subjects resulting from selecting the conditional method over the unconditional one represent a significant 30% increase in the number of eligible subjects, accrual time, human resources, etc.; basically, a 30% increase in all facets of the study. This was deemed excessive by all investigators involved in the trial. Such power benefits of the exact unconditional approach have also been noted in the context of the matched-pairs design.

The second consideration of this commentary deals with two-sided significance levels. As briefly noted by Agresti in the context of Fisher's exact test, there are different ways of forming two-sided p-values, three of which are presented. He comments on the fact that various techniques may lead to diverse results and illustrates this point with a numerical example. It may be useful to complement the presentation with some results and observations noted in the context of research on the unconditional approach to exact inference. For the example used by Agresti, namely that of contrasting 10/100 with 20/100, the exact unconditional two-sided p-value under the Pearson chi-square statistic is 0.054, in contrast with 0.073 under all three approaches for the conditional test. When the data are slightly modified to 10/101 versus 20/100, the exact unconditional p-value becomes 0.048, as compared with the reported 0.069 and 0.050. As expected, the unconditional method leads to lower p-values than the conditional ones. The fact,

however, that the three conditional approaches to two-sided significance testing may lead to differing p-values is not surprising. Indeed, each will lead to different orderings of the sample space and, in turn, will likely produce different p-values. These variations are, therefore, understandable.

There are, however, other discrepancies in two-sided conditional testing that may occasionally lead to disturbing results. For all subsequent comparisons, we only use Agresti's second approach to two-sided p-values, namely the one based on ordering the sample space according to the hypergeometric probability of each table. The anomalies discussed have been indicated in Cormack (1986) and Shuster and Suissa (1990). The first illustration contrasts 4/35 with 0/35, which produces a conditional p-value of 0.114, whereas the unconditional p-value is 0.042. When the second rate is modified to 0/36, the conditional p-value drops to a "significant" 0.054, whereas the unconditional one becomes 0.041. Another numerical example of this situation is in the contrasts of 2/171 versus 9/172 and 2/172 versus 9/171, where the conditional p-values are 0.061 and 0.035, respectively, whereas the unconditional ones are very similar at 0.036 and 0.032. In both examples, the declines in the conditional p-value associated with such practically insignificant changes in the data are a source of concern among users of the method. Indeed, such minute alterations in the sample size should not affect the p-values to this extent. The discreteness imposed by the conditional approach and the magnitude of the jumps in the p-value are often significant, most particularly in the small sample situations it is specifically indicated for. This phenomenon is not observed with the exact unconditional approach.

The second illustration contrasts 7/14 versus 1/14 and 6/12 versus 1/14. The odds ratios of these tables are equal to each other (= 13), as are all other measures of association. Yet, the corresponding exact conditional p-values are 0.033 and 0.026, whereas the unconditional ones are both 0.014. The drop in p-value under the conditional approach is counterintuitive as a reasonable rule of the "evidence" in these data should diminish (it does remain the same for the unconditional approach); it should not, however, increase.

The third consideration deals with the question of interval estimation. A limitation of the exact conditional approach in 2×2 tables is its inability to estimate anything but measures of association that are functions of the odds ratio. Consequently, the conditional approach becomes ineffective in a field like epidemiology, where relevant measures of association are often the rate ratio and rate difference (ratio and difference of two binomial proportions or Poisson rates). Although some attempts have been made to estimate these, they have proven futile because of this limitation of the exact conditional approach. This obstacle could form the basis of research on alternative approaches to the restrictive conditional technique in resolving the problem of exact estimation in this context.

In summary, Professor Agresti's presents us with a superb synthesis of the extensive research conducted on the exact analysis of data from contingency tables. The review particularly highlights the extremely popular exact conditional approach that is unquestionably a highly potent technique. Moreover, the spectacular advances in computing algorithms have made the conditional approach attractive from the practical standpoint. We presented, in the context of the 2×2 table, three aspects of the conditional approach that may put in perspective its practical effectiveness. First, conditional tests are found to be significantly less powerful than their exact unconditional counterparts. Second, exact conditional two-sided p-values display an inefficient discrete behavior and, at times, lead to inconsistent results, thus rousing suspicion from the users of this approach. These two aspects are accentuated when the sample size is particularly small, the precise circumstance where exact methods are indicated. The final concern lies with the inability of the exact conditional approach to estimate relevant measures of association other than functions of the odds ratio.

In essence, although the conditional approach is a formidable tool for the exact analysis of categorical data, it has limitations that are not always apparent nor understood. Consequently, research on exact alternatives must continue, not only to offer a wider range of possibly more efficient techniques, but also to alleviate the concerns and doubts raised by the inquisitive consumer of these statistical tools.

ACKNOWLEDGMENT

Samy Suissa is a Research Scholar supported by the Fonds de la recherche en santé du Quebec.