Inference from empty contingency tables

Gerd Rippin, Marcus Kutschmann

If only the marginal counts of a contingency table are known, inference regarding a statistic e.g. like the log odds ratio is still possible. In the following a 2*2 contingency table and the log odds ratio as the interesting statistic are assumed, although the methodology can be extended to higher dimensions and to various types of statistics. Given the marginal counts all possible contingency tables are determined, such that for every possible table the estimated log odds ratio can be calculated. All resulting estimates per contingency table can be averaged in a first step e.g. by using the hypergeometric weighting under H0 (which is also used for Fisher's exact test), such that a first overall estimate of the log odds ratio together with its standard error are available. This estimate is used for assigning new weights to the estimates per contingency table corresponding to the normal density of the overall estimated log odds ratio. After these initial steps a MCMC method is applied, by drawing in a first step from the distinct distributions of the log odds ratios of the contingency tables. The averaging takes place corresponding to the weights of the last step, such that a new overall estimate is calculated. Then, secondly, also the new weights are calculated by the density of the normal distribution corresponding to the new overall estimate as described before. The MCMC algorithm has to run sufficiently long in order to obtain accurate estimates and confidence limits.

The results from the procedure are consistent with common sense. If e.g. an exposition is very common but the disease is very rare, it is unlikely that the exposition is a strong risk factor for the disease. But although this procedure can be judged to yield estimates of association in an elegant way, it has also a serious disadvantage: Its misuse can occur easily as all one has to obtain are percentages of two variables and a claim of an association can be made without even bothering having sampled data. Thus, ambiguity and multiple testing become significant issues. If there is no way to sample the corresponding contingency table, the described MCMC approach may be sensible to apply, but if there is a way, it cannot replace at all the sound scientific work of sampling data and to conduct well-designed studies.