A Bayes view on Simpson's paradox

Gerhard Nehmiz

Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
Email: gerhard.nehmiz.ext@boehringer-ingelheim.com

Given 3 classifications A, B and C, we investigate the association between A and B conditionally for each value of C. Simpson's paradox[1,2,3] is said to be present if the association is reversed when investigated marginally over C, i.e. when the corresponding frequency table is collapsed w.r.t. C.

As a special case, we consider 2x2x2 tables and generate random numbers $p_{1,\ldots,8}$ from a prior uniform distribution on the simplex with $p_{1,\ldots,8} \geq 0$ and ≤ 1 and sum=1, which is a special case of a Dirichlet distribution[6]. The probability for achieving a table with association reversal is then 1/60. Also, we overlay the prior distribution with the multinomial likelihood[4,8] and obtain as posterior again a Dirichlet distribution. We can then calculate a Bayes factor to quantify the evidence that association reversal is in fact present.

To resolve the paradox on a subject-matter basis, it becomes necessary to look not only at the prior and data but also at the nature of the classifications A, B and C. Embedding in a hierarchical model, as displayed by a Directed Acyclic Graph[9], provides a criterion to decide which of the 2 contrary propositions, i.e. whether the association goes in one or the other direction, is better (not to say "is correct"). A nice example from meta-analysis[5] is shown. As another observation, Simpson phenomena can be induced through the familiar "continuity correction"[7].

References:

[1]Yule GU: Notes on the theory of association of attributes in statistics. Biometrika 1903; 2: 121-134.
[2]Simpson EH: The Interpretation of Interaction in Contingency Tables. J.R.S.S.B 1951; 13: 238-241.
[3]Samuels ML: Simpson's Paradox and Related Phenomena. J.A.S.A. 88 (1993), 81-88.
[4]Julious SA, Mullee MA: Confounding and Simpson's paradox. British Medical Journal 1994; 309: 1480-1481.
[5]Rücker G, Schumacher M: Simpson's paradox visualized: The example of the Rosiglitazone meta-analysis. BMC Medical Research Methodology 2008; 8(34): 1-8.
[6]Pavlides MG, Perlman MD: How Likely Is Simpson's Paradox? American Statistician 2009; 63: 226-233.
[7]Greenland S: Simpson's Paradox From Adding Constants in Contingency Tables as an Example of Bayesian Noncollapsibility. American Statistician 2010; 64: 340-344.
[8]Chuang-Stein C, Beltangady M: Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. Pharmaceutical Statistics 2011; 10: 3-7.
[9]Armistead TW: Resurrecting the Third Variable: A Critique of Pearl's Causal Analysis of Simpson's Paradox. (With comments by Pearl J, Christensen R, Liu K/Meng X-L, and concluding remarks.) American Statistician 2014; 68: 1-31. Correction: American Statistician 2014; 68: 132.