What does Simpson's Paradox have to do with Causality?

Prasanta S. Bandyopadhyay
Department of History & Philosophy
Montana State University
Bozeman, MT
USA
email: psb@montana.edu

Palash Sarkar
Applied Statistics Unit
Indian Statistical Institute
203, B.T. Road, Kolkata
India 700108
email: palash@isical.ac.in

April 20, 2020

Simpson's Paradox (SP) is the cessation of the association of data when data are combined to constitute a whole. Causal theorists such as Pearl contest that SP is solely explainable causally. However, we will argue that this is not the case. We will address the above question by discussing the logic behind SP based on the concepts of odd-ratio, homogeneity, and collapsed tables. Let T be a 2×2 contingency table of the form:

$$T = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}\,.$$

Good and Mittal (1987) define a measure of assocation α for the table T as a function $\alpha(T)$ of T satisfying *homogeneity of degree zero*, symmetry, and row-scale and column-scale invariance. A particular measure of association considered by Good and Mittal is the Odds (or, Cross-Product) Ratio $\kappa(T)$ which is defined as follows: $\kappa(T) = ad/(bc)$.

Let T_1, T_2 and the collapsed table $T_1 + T_2$ be 2×2 tables defined as below.

$$T_1 = \begin{array}{|c|c|} \hline a_1 & b_1 \\ \hline c_1 & d_1 \\ \hline \end{array}\,, \qquad T_2 = \begin{array}{|c|c|} \hline a_2 & b_2 \\ \hline c_2 & d_2 \\ \hline \end{array}\,, \qquad T_1 + T_2 = \begin{array}{|c|c|} \hline a_1 + a_2 & b_1 + b_2 \\ \hline c_1 + c_2 & d_1 + d_2 \\ \hline \end{array}\,.$$

According to Good and Mittal, the sub-populations T_1 and T_2 satisfy Odds Ratio Homogeneity (ORH) if $\kappa(T_1) = \kappa(T_2) = \kappa(T_1 + T_2)$. We introduce the notion of Weak Odds Ratio Homogeneity (WORH): The subpopulations T_1 and T_2 are weakly homogeneous with respect to the Odds Ratio if *either* $\kappa(T_1) = \kappa(T_1 + T_2)$ *or* $\kappa(T_2) = \kappa(T_1 + T_2)$.

Simpson's paradox (SP) is defined as being present in the contingency tables T_1 and T_2 if

$$\left(\frac{a_1}{a_1 + b_1} > \frac{c_1}{c_1 + d_1}\right) \wedge \left(\frac{a_2}{a_2 + b_2} > \frac{c_2}{c_2 + d_2}\right) \wedge \sim \left(\frac{a_1 + a_2}{a_1 + a_2 + b_1 + b_2} > \frac{c_1 + c_2}{c_1 + c_2 + d_1 + d_2}\right).$$

We prove the following three theorems (informally stated) connecting Odds Ratio and SP:

Theorem 1: *SP is logically equivalent to the following condition:* $(\kappa(T_1) > 1) \wedge (\kappa(T_2) > 1) \wedge \sim (\kappa(T_1 + T_2) > 1)$.

Theorem 2: *SP does not hold if and only if* $(\kappa(T_1) > 1) \wedge (\kappa(T_2) > 1)$ *implies* $\kappa(T_1 + T_2) > 1$.

Theorem 3: *If WORH holds for T_1 and T_2, then SP does not hold for T_1 and T_2.*

The above discussion provides a strong case against the well-entrenched causal resolution of SP. The above three theorems provide a logical foundation for SP where none of the theorems in which SP holds (or does not) contain a causal reasoning which is central to the causal resolution of the paradox. Bollen and Pearl (2013, see also Pearl 2009) contend that homogeneity is a causal notion. For the sake of argument if we assume that the latter is correct, then a natural question would be, "how could a causal notion like homogeneity generate an odds ratio, as well as a collapsed table (central to the idea of the collapsibility principle) which is admittedly non-causal?" So, the worry is, "how does a causal notion such as homogeneity produce a non-causal principle like the collapsibility principle contained in a collapsed table?"