Simpson's paradox for Cox model

Clelia Di Serio　Yosef Rinott　Marco Scarsini

Blyth (1973) has given a very simple and suggestive description of Simpson's paradox in terms of conditional probabilities. Given three events A, B, C, the paradox is the simultaneous appearing of the three following inequalities

$$\begin{aligned} P(A|B \cap C) &\geq P(A|B^c \cap C), \\ P(A|B \cap C^c) &\geq P(A|B^c \cap C^c), \\ P(A|B) &< P(A|B^c). \end{aligned} \tag{1}$$

Samuels (1993) has extended the consideration of the paradox from events to random variables and has explained it as a particular case of the *association reversal* or of the *association distortion* phenomena. This idea has been further extended by Scarsini and Spizzichino (1999) who have determined necessary conditions for the paradox when different notions of positive dependence are considered, and by Rinott and Tam (2003), who have used a condition of monotone regression.

We consider the proportional hazard model due to Cox (1972), i.e. we assume that the conditional survival function of a random lifetime T, given covariates X, Y is

$$P(T > t+s | T > t, X = x, Y = y) = \exp\left\{\left(-\int_t^{t+s} h_0(u)\ \mathrm{d}u\right)\exp\{\beta_1 x + \beta_2 y\}\right\},$$

where h_0 is a positive function such that $\int_t^\tau h_0(u)\ \mathrm{d}u$ is finite if and only if $\tau < \infty$ for all $t > 0$.

We say that there is a Simpson's paradox for (t, s) if the following two conditions hold:

$$P(T > t+s | T > t, X = x, Y = y) \text{ is decreasing in } x \text{ for all } y, \tag{2}$$

$$P(T > t+s | T > t, X = x) \text{ is increasing in } x. \tag{3}$$

Heuristically, (2) means that conditionally on every value of the covariate Y, higher values of X stochastically reduce the conditional survival time (i.e., X is *detrimental* for T given Y), whereas (3) means that unconditionally the opposite is true (i.e., X is *protective* for T).

We find conditions for the paradox in terms of the coefficients β_1, β_2 and the joint distribution of the covariates (X, Y), and we study the range of values of (t, s) for which the paradox obtains (in some circumstances it holds for all $s, t > 0$).

It is particularly remarkable that it is possible to have the paradox even when the covariates are independent. This is due to the fact that independence is not preserved under conditioning, hence covariates that are independent at time 0 become dependent as time goes by, and they may do so in a way that generates the paradox.