Estimating causal (intervention) effects for binary data

In [1], the estimation of causal (intervention) effects for Gaussian data was discussed and fast algorithms were proposed. In this text, we adapt these results for binary data. For the notation concerning causal effects, refer to [2].

Let \( X_1, \ldots, X_p \) be \( p \) binary explanatory variables and let \( Y \) be a binary dependent variable. Without loss of generality, all levels are assumed to be in \( \{0, 1\} \). We define the causal effect of \( X_i \) on \( Y \) by

\[
E(Y|\text{do}(X_i = 1)) - E(Y|\text{do}(X_i = 0)).
\]

Let \( G \) be the directed acyclic graph on \( \{X_1, \ldots, X_p, Y\} \) describing the true underlying causal structure. (In practice, this graph is estimated.) Let \( Pa_i \) denote the parents of \( X_i \) in \( G \).

For \( Pa_i = \emptyset \), back-door adjustment (see Theorem 3.3.2 in [2]) implies that the causal effect of \( X_i \) on \( Y \) is

\[
E(Y|\text{do}(X_i = 1)) - E(Y|\text{do}(X_i = 0)) = E(Y|X_i = 1) - E(Y|X_i = 0),
\]

as illustrated in the following coordinate system.

Since the horizontal distance in Figure (1) is equal to 1, the causal effect is exactly the same as the slope of the straight line through the points \( (0, E(Y|\text{do}(X_i = 0))) \) and \( (1, E(Y|\text{do}(X_i = 1))) \). For any \( S \subseteq \{X_1, \ldots, X_p\} \setminus \{X_i\} \), let \( \beta_{i|S} \) denote the
regression coefficient of $X_i$ in the regression of $Y$ on $X_i$ and $S$. Then, the causal effect defined in equation (1) is given by $\beta_{i|\emptyset}$, i.e., the regression coefficient of $X_i$ in the regression of $Y$ on $X_i$.

For $Pa_i = \{X_j\}$, the causal effect of $X_i$ on $Y$ is

$$E(Y|do(X_i = 1)) - E(Y|do(X_i = 0))$$

$$= \sum_{x_j} E(Y|X_i = 1, X_j = x_j)P(X_j = x_j) - \sum_{x_j} E(Y|X_i = 0, X_j = x_j)P(X_j = x_j)$$

$$= [E(Y|X_i = 1, X_j = 1)P(X_j = 1) + E(Y|X_i = 1, X_j = 0)P(X_j = 0)]$$

$$- [E(Y|X_i = 0, X_j = 1)P(X_j = 1) + E(Y|X_i = 0, X_j = 0)P(X_j = 0)]$$

$$= [E(Y|X_i = 1, X_j = 1) - E(Y|X_i = 0, X_j = 1)]P(X_j = 1)$$

$$+ [E(Y|X_i = 1, X_j = 0) - E(Y|X_i = 0, X_j = 0)]P(X_j = 0).$$

If there is no interaction between $X_i$ and $X_j$, we have

$$E(Y|X_i = 1, X_j = 1) - E(Y|X_i = 0, X_j = 1) = E(Y|X_i = 1, X_j = 0) - E(Y|X_i = 0, X_j = 0) = \beta_{i|X_j}.$$

Plugging this into equation (2) yields that the causal effect of $X_i$ on $Y$ is

$$\beta_{i|X_j}P(X_j = 1) + \beta_{i|X_j}P(X_j = 0) = \beta_{i|X_j}.$$

If $Pa_i$ contains several variables, one can show analogously that the causal effect of $X_i$ on $Y$ is given by $\beta_{i|Pa_i}$. If there is interaction, this should be taken into account in the regression. However, in the case with interaction, many more parameters have to be estimated and this might lead to accuracy problems. Thus, one has to weigh the advantage of using interactions against the disadvantage of badly estimated parameters.

**References**
