⟨Research Paper⟩

Journal of the Korean
Statistical Society
Vol. 16, No. 2, 1987

Simpson's Paradox and Randomization

Myung-Hoe Huh*

ABSTRACT

The roie of randomization is examined with regard to the Simpson's paradox. When the sample size n is large, it is known that the randomization is powerful in preventing the Simpson's paradox. In the present study, the question is whether it performs well for small n.

1. Introduction

Lindley and Novick(1981) consider the following medical experiment. Each patient is one of two sex types, male M or female $\bar{M}$, and receives either treatment T or control $\bar{T}$. Afterwards the experimenter observes each patient recovered R or not $\bar{R}$. It is possible that

$$P(R|\bar{T}M)<P(R|TM),\ P(R|\bar{T}\bar{M})<P(R|T\bar{M}), \qquad (1)$$

and at the same time

$$P(R|\bar{T})>P(R|T). \qquad (2)$$

It is called Simpson's (1951) paradox. This happens when $P(M|T)$ and $P(M|\bar{T})$ (or, $P(\bar{M}|T)$ and $P(\bar{M}|\bar{T})$) are significantly different (Blyth, 1972), since $P(R|T)$ ($P(R|\bar{T})$) is the weighted average of $P(R|TM)$ and $P(R|T\bar{M})$ ($P(R|\bar{T}M)$ and $P(R|\bar{T}\bar{M})$) with weights $P(M|T)$ and $P(\bar{M}|T)$ ($P(M|\bar{T})$ and $P(\bar{M}|\bar{T})$). See Figure 1, which is originally due to Lindley and Novick (1981). In this case, the sex type is the confounding factor. Observational studies often do not have control of confounding factors.

If the patients are assigned randomly to the treatments, and if the number of patients

* Dept. of Statistics, Korea University

is very large, then $P(M|T)$ and $P(M|\bar{T})$ are nearly the same. Consequently, in such a case, Simpson's paradox would not happen. The question is whether the machinery of randomization prevents Simpson's paradox effectively when the number of patients is not large.

2. The Chance of Simpson's Paradox(α, δ)

Assume that, among $n(=n_1+n_2)$ patients, n_1 patients are male (M) and n_2 patients are female $(\bar{M})$. Also assume that the sex type of each patient is not measured (or cannot be measured) for some reasons. Each patient is assigned independently to the treatment T with probability 1/2 and to the control $\bar{T}$ with probability 1/2.

Let B_1 be the number of male patients assigned to T and B_2 be the number of female patients assigned to T. B_1 and B_2 are random variables and distributed as Binomial $(n_1, 1/2)$ and Binomial $(n_2, 1/2)$, respectively.

We are interested in the probability that (2) holds given

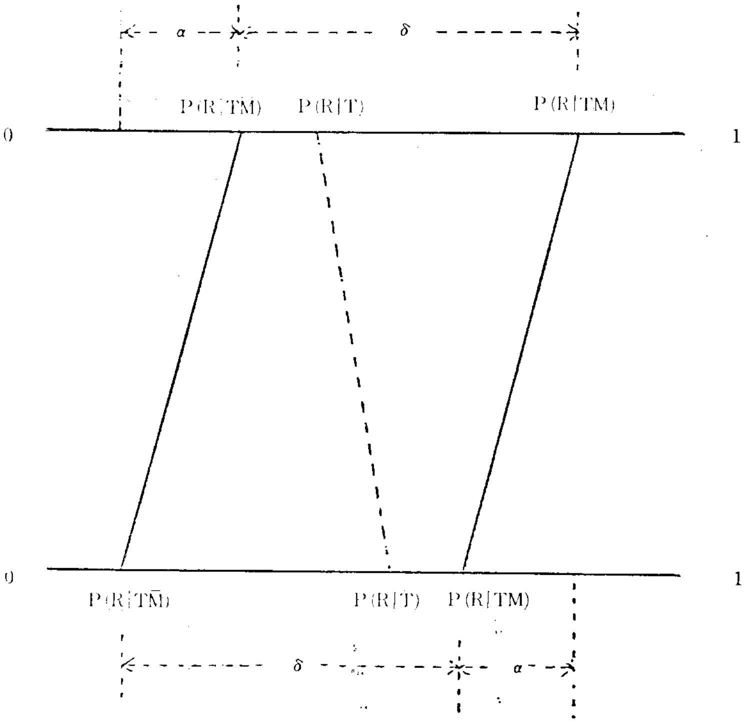

Figure 1. Simpson's Paradox (α, δ)

$$P(R|\bar{T}M)+\alpha=P(R|TM),\quad P(R|\bar{T}\bar{M})+\alpha=P(R|T\bar{M}), \tag{1'}$$

where α is a positive constant and can be interpreted as the treatment effect. Note that

$$P(R|T)-P(R|\bar{T})$$
$$=P(R|TM)P(M|T)+P(R|T\bar{M})P(\bar{M}|T)-P(R|\bar{T}M)P(M|\bar{T})-P(R|\bar{T}\bar{M})P(\bar{M}|\bar{T})$$
$$=\{P(R|\bar{T}M)-P(R|\bar{T}\bar{M})\}\{P(M|T)-P(M|\bar{T})\}+\alpha.$$

Hence (2) holds given (1′), if and only if,

$$P(M|T)-P(M|\bar{T})<-(\alpha/\delta), \tag{3}$$

where

$$\delta\equiv P(R|\bar{T}M)-P(R|\bar{T}\bar{M})=P(R|TM)-P(R|T\bar{M}),$$

which can be interpreted as the sex effect on the recovery rate. Here, α and δ satisfy $0<\alpha<\delta<1$ and $\alpha+\delta<1$. See Figure 1.

Since $P(M|T)$ and $P(M|\bar{T})$ are given by $B_1/(B_1+B_2)$ and $(n_1-B_1)/(n-B_1-B_2)$ respectively, we can calculate the probability that (3) holds; that is,

$$P\{B_1/(B_1+B_2)-(n_1-B_1)/(n-B_1-B_2)<-(\alpha/\delta)\}. \tag{3'}$$

We call (3′) the chance of Simpson's paradox (α,δ). Table 1 gives chances of Simpson's paradox (α,δ) for $\delta/\alpha=2(2)8$, $n=20(20)100$, and $n_1/n_2=1$. Jeon, Chung, and Bae (1987) computed (3′) more extensively, including the cases $n_1\neq n_2$.

Table 1. The Chance of Simpson's Paradox (α,δ)*

n	n_1	n_2	$\delta/\alpha=2$	$\delta/\alpha=4$	$\delta/\alpha=6$	$\delta/\alpha=8$
20	10	10	0.021	0.134	0.253	0.263
40	20	20	0.001	0.077	0.138	0.215
60	30	30	0.000	0.027	0.123	0.183
80	40	40	0.000	0.015	0.073	0.157
100	50	50	0.000	0.006	0.045	0.099
∞**	∞	∞	0	0	0	0

* This table is obtained by summing cell probabilities from exact joint distribution of B_1 and B_2, for the cases of finite n_1 and n_2.

** $n_1\to\infty$, $n_2\to\infty$, and $n_1=n_2$. The results are immediate consequences of strong law of large numbers.

3. Conclusion

From Table 1, note that the chance of Simpson's paradox decreases as n increases for each δ/α. But it decreases slowly for large δ/α. For instance, when $\alpha=0.1$, $\delta=0.8$ (thus $\delta/\alpha=8$), and $n_1=n_2=40(50)$, the chance of Simpson's paradox is still 0.16 (0.10).

In sum, the machinery of randomization works efficiently for the given sample size only when δ/α, the ratio of the sex (confounding factor) effect over the treatment effect, is not large. We emphasize that it is important to find out the confounding factor and incorporate it into the experiment even for the randomized experiment. We should not rely on the randomization too much.

References

(1) Blyth, C.R. (1972). On Simpson's paradox and sure-thing principle. *Journal of the American Statistical Association*, 67, 364~366.

(2) Jeon, J.W., Chung, H.Y., and Bae, J.S. (1987). Chances of Simpson's paradox. *Journal of the Korean Statistical Society*, 16, 117~127.

(3) Lindly, D.V., and Novick, M.R. (1981). The role of exchangeability in inference. *The Annals of Statistics*, 9, 45~58.

(4) Simpson, E.H. (1951). The interpretation of interaction contingency tables. *Journal of the Royal Statistcal Society, Series B*, 13, 238~241.