Instrumental Variable-based Identification for Causal Effects using Covariate Information

Yuta Kawakami

Department of Mathematics, Physics, Electrical Engineering and Computer Science
Graduate School of Engineering Science, Yokohama National University
79-5 Tokiwadai, Hodogaya-ku, Yokohama 240-8501 JAPAN
kawakami-yuta-yd@ynu.jp

Abstract

This paper deals with the identification problem of causal effects in randomized trials with noncompliance. In this problem, generally, causal effects are not identifiable and thus have been evaluated under some strict assumptions, or through the bounds. Different from existing studies, we propose a novel identification condition of joint probabilities of potential outcomes, which allows us to derive a consistent estimator of the causal effect. Regarding the identification conditions of joint probabilities of potential outcomes, the assumptions of monotonicity (Pearl 2009), independence between potential outcomes (Robins and Richardson 2011), gain equality (Li and Pearl 2019) and specific functional relationships between cause and effect (Pearl 2009) have been utilized. In contrast, without such assumptions, the proposed condition enables us to evaluate joint probabilities of potential outcomes using an instrumental variable and a proxy variable of potential outcomes. The result of the present paper extends the range of solvable identification problems in causal inference.

Introduction

Practical Background

Evaluation of causal effects from randomized trials is a central aim in practical sciences. Randomized trials have been regarded as a more reliable and powerful tool for evaluating causal effects, compared with observational studies, where confounding, information bias and selection bias all hinder the evaluation of causal effects from observational data. When compliance with treatment assignment is perfect, the standard intention-to-treat (ITT) analysis provides valid estimation of the causal effect, by simply computing the effect of assignment on the outcome. However, it is well known that noncompliance is often present in randomized trials. In such cases, ITT analysis may introduce a biased causal effect (Sheiner and Rubin 1995).

The limitation of an analysis which ignores noncompliance has received growing concern in the practical sciences, and many researchers and practitioners have proposed various approaches for dealing with noncompliance. Some of these include point estimators for the causal effect (Mark and Robins 1993; Angrist, Imbens, and Rubin 1996; Wang
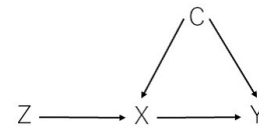
Fig. 1: Graphical representation of causal dependencies in a randomized trial with noncompliance

and Tchetgen 2018). However, these approaches generally require various strict assumptions. An alternative approach is to construct nonparametric bounds for the causal effect (Robins 1989; Manski 1990; Balke and Pearl 1997). Since this approach is based on the simple assumptions of the instrumental variable (Greenland 2000), it is applicable to most randomized trials where the treatment assignment can serve as a perfect instrumental variable. However, such parsimonious assumptions may sometimes result in rather wide bounds, which might be useless for decision making. One reason may be that these bounds are calculated using only the information on the randomized assignment, the treatment received and the outcome. When we examine most randomized trials in the practical sciences, we notice that not only the measurements of the randomized assignment, the treatment received and the outcome are available, but also some measurements of covariates such as age, gender and race (Cai, Kuroki, and Sato 2007; Lui 2011). When we are in possession of such covariate measurements, our purpose is to find out when and how we can use this information to identify causal effects in randomized trials with noncompliance.

Problem Description

To motivate our dicussion, we consider randomized clinical trials with the purpose to examine causal effects of an experimental treatment in comparison with a controlled treatment. We describe the problem with the graph shown in Fig. 1. For the graph-theoretic terminology and basic theory of the structural causal models used in the present paper, we refer readers to Pearl (2009).

Intuitively, in Fig. 1, a directed edge from X to Y indicates that X could have a direct effect on Y. In addition, the absence of a directed edge from Y to X indicates that Y cannot be a direct cause of X, and a directed path from C to Y

through X indicates that some elements of $\boldsymbol{C}$ could have an effect on Y mediated by X. This situation often appears in the medical and statistical literature, for example, (Multiple Risk Factor Intervention Trial Research Group 1982), (Cai, Kuroki, and Sato 2007), (Yan et al. 2011) and (Lui 2011).

Following Balke and Pearl (1997), for the case shown in Fig. 1, we assume that Z, X, and Y are observed dichotomous variables, where Z represents the randomized treatment assignment, X represents the treatment actually received, and Y represents the observed outcome. In addition, we let z, x, and y represent the values taken by the variables Z, X, and Y, with the following meanings: $z \in \{z_0, z_1\}$, z_1 indicates subjects randomized to the experimental treatment, while z_0 indicates subjects randomized to the controlled treatment; $x \in \{x_0, x_1\}$, x_1 indicates taking the experimental treatment, while x_0 indicates taking the controlled treatment; $y \in \{y_0, y_1\}$, y_0 indicates disease, while y_1 indicates non-disease. $\boldsymbol{C}$ represents the set of all discrete and continuous variables, both observed and unobserved, that are not affected by X or Y. In this situation, the randomized assignment Z satisfies the instrumental variable assumptions: (i) Z is associated with X; (ii) Z is independent of $\boldsymbol{C}$; and (iii) Z is conditionally independent of Y given $\boldsymbol{C}$ and $\{X\}$ (Greenland 2000). Here, regarding X, Y and Z, it is straightforward to extend our results from the case of dichotomous observed variables to the case of multivalued observed variables. In particular, as Balke and Pearl (1997) stated, multivalued or continuous outcome can be accommodated in the model using the event $Y < y$ as an outcome variable.

Let $\mathrm{pr}(x, y)$ be the joint probability of $(X, Y) = (x, y)$, $\mathrm{pr}(y \mid x)$ be the conditional probability of $Y = y$ given $X = x$, and $\mathrm{pr}(x)$ be the marginal probability of $X = x$. Similar notation is used for other probabilities. Then, according to (Pearl 2009), the causal effect of $X = x$ on $Y = y$ is represented as

$$\mathrm{pr}(Y_x = y) = \sum_c \mathrm{pr}(y|x, \boldsymbol{c})\mathrm{pr}(\boldsymbol{c}), \qquad (1)$$

where $Y_x = y$ denotes "Y takes the value y when X is experimentally set to x", or the counterfactual sentence: "Y would be y, had X been x." Here, summation signs are replaced by integrals whenever the summed variables are continuous. Equation (1) would be obtained if an ideal randomized trial with X were feasible. Contrarily, even when such a randomized trial is not successful, if a set of observed covariates satisfies the back-door criterion relative to (X, Y) (Pearl 2009), the causal effect is identifiable.

When compliance is not perfect and a set of observed covariates is insufficient for identification, the causal effect is not identifiable without any further assumption. To solve this problem, using the data about the randomized assignment, the treatment received, and the outcome from a randomized trial, Balke and Pearl (1997) provided the sharp bounds for causal effects. Balke & Pearl's bounds provide the range within which the causal effect must lie. In addition, noting that a set $\boldsymbol{C}$ of covariates includes observed variables, (Cai, Kuroki, and Sato 2007) derived narrower bounds than Balke & Pearl's bounds using the information on observed

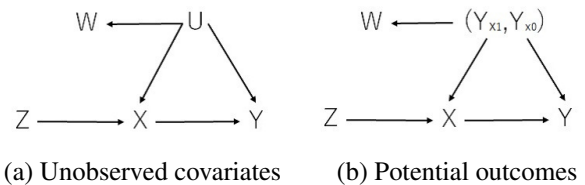(a) Unobserved covariates   (b) Potential outcomes

Fig. 2: Graphical representation of causal dependencies in a randomized trial with a proxy variable

covariates. According to (Cai, Kuroki, and Sato 2007), covariate information need not include confounding factors in order to narrow the bounds. This observation motivates the present paper, which provides a new look on how to use an instrumental variable Z together with a proxy variable W, depicted in Fig. 2 (a), to identify causal effects from studies with non-compliance.

Theoretical Background

Different from existing studies, our solution of the problem is to provide an identification condition of "joint probabilities of potential outcomes", which also allows us to identify causal effects. One of representative examples of "joint probabilities of potential outcomes" is "probabilities of causation". Tian and Pearl (2000) and Pearl (2009) developed formal semantics for probabilities of causation based on structural models of counterfactuals. The probabilities of causation are formulated based on the joint probabilities of two potential outcomes. Since one cannot simultaneously observe the results of the same subjects receiving the experimental treatment and the controlled treatment in reality, these quantities are not identifiable even for successful randomized experiments (Pearl 2009, pp.284–285). To solve the problem, Tian and Pearl (2000) showed how to bound these quantities from data obtained in experimental and observational studies. Tian & Pearl's bounds also provide the range within which the probabilities of causation must lie. In addition, Kuroki and Cai (2011) derived narrower bounds of probabilities of causation than Tian & Pearl's bounds using covariate information. However, it has been pointed out that these bounds are too wide to evaluate the probabilities of causation.

To overcome this difficulty, Tian and Pearl (2000) also noted that the probabilities of causation are identifiable if the monotonicity can be assumed and the causal effects are identifiable, and Pearl (2009) showed that specific functional relationships between cause and effect lead to the identification of the probabilities of causation. In the context of natural direct and indirect effects (Pearl 2001), under the assumption of no unmeasured confounding, Robins and Richardson (2011) stated that the joint probabilities of potential outcomes are identifiable if (i) two potential outcomes are independent or (ii) one potential outcome can be deterministically formulated as a function of the other potential outcome. In addition, in the context of the unit selection problems, Li and Pearl (2019) showed that a linear combination of joint probabilities of potential outcomes is identifiable under gain equality. These prior research studies show

that the joint probabilities of potential outcomes play an important role in solving various problems of causal inference. However, when the present assumptions are violated, there has been much less discussion of how to identify the joint probabilities of potential outcomes.

In the present paper, we provide a novel identification condition of joint probabilities of potential outcomes using an instrumental variable and a proxy variable. The proposed condition enables us to derive consistent estimators of joint probabilities of potential outcomes, without relying on the previously used assumptions. Thus, the proposed condition also allows us to identify the causal effects. Here, different from the effect restoration proposed by Kuroki and Pearl (2014), the proposed condition does not require that the number of categories of an unmeasured confounder is known. In addition, in some situations, the present result is applicable to the identification of probabilities of causation (Tian and Pearl 2000), the unit selection problems (Li and Pearl 2019), the impact evaluation problem of social programs (Heckman, Smith, and Clements 1997), the non-compliance problem of causal effects (Angrist, Imbens, and Rubin 1996; Balke and Pearl 1997), the identification problems of natural direct and indirect effects (Pearl 2001), prevented and preventable proportions (Yamada and Kuroki 2017) and counterfactual based traffic conflicts (Yamada and Kuroki 2019). Thus, the results of the present paper extend the range of solvable identification problems in causal inference.

Preliminaries

This section introduces the potential outcomes used to discuss our problems. In the present paper, we will assume that readers are familiar with the basic theory of causal inference (Pearl 2009; Imbens and Rubin 2015).

In principle, for $x \in \{x_0, x_1\}$, the i-th of the N subjects has a potential outcome $Y_x(i)$ that would have resulted if X had been x, denoted as $X(i) = x$. Here, note that the subject ensures a deterministic relationship between two variables X and Y in the semantics of structural causal models (Pearl 2009). To address the problem, the present paper assumes the stable unit treatment value assumption, which can be summarized as follows: (i) the treatment status of any subject does not affect the outcomes of the other subjects (no interference) and (ii) the treatments of all subjects are comparable (no variation in treatment). Thus, when the N subjects in the study are considered as random samples from the population of interest, $X(i)$ and $Y_x(i)$ are referred to as the values of random variables X and Y_x respectively, and thus the causal effect of $X = x$ on $Y = y$ is defined as $\mathrm{pr}(Y_x = y)$. Similar notation is used for other potential outcomes.

For the i-th subject, the potential outcome Y_x is observed only if X is x. This property is called the consistency (Robins 1989; Pearl 2009), which is formulated as

$$X(i) = x \Longrightarrow Y_x(i) = Y(i)$$

. When a randomized experiment is conducted and compliance is perfect, since X is independent of (Y_{x_0}, Y_{x_1}), the

causal effect is identifiable and is given by

$$\mathrm{pr}(Y_x = y) = \mathrm{pr}(y \,|\, x).$$

Here, "identifiable" means that the causal quantities, such as $\mathrm{pr}(Y_x = y)$, can be estimated consistently from a joint probability of observed variables. In contrast, when it is difficult to conduct a randomized experiment and only observational data are available, we can evaluate the causal effects according to the conditionally-ignorable-treatment-assignment condition (Rosenbaum and Rubin 1983), or graphically, the back-door criterion (Pearl 2009). In other words, for the treatment X, if there exists such a set $\boldsymbol{S}$ of observed covariates that X is conditionally independent of (Y_{x_0}, Y_{x_1}) given $\boldsymbol{S}$, we can say that treatment assignment is conditionally ignorable given $\boldsymbol{S}$. Then, the causal effects are estimable using $\boldsymbol{S}$ as

$$\mathrm{pr}(Y_x = y) = E_s[\mathrm{pr}(y \,|\, x, \boldsymbol{S})].$$

Here, $E_s[\mathrm{pr}(y \,|\, x, \boldsymbol{S})]$ is the expectation of $\mathrm{pr}(y \,|\, x, \boldsymbol{S})$ regarding $\boldsymbol{S}$. Although there are other identification conditions that can be used to solve our problem (e.g., Tian and Pearl 2002; Pearl 2009), the present paper does not cover them due to space constraints.

Main Result

Identification

According to Fig. 2 (a), letting $\boldsymbol{U}$ be a subset of variables from the set $\boldsymbol{C}$ in Fig. 1, consider the problem of evaluating the joint probabilities of potential outcomes. Here, W (possibly, a subset of $\boldsymbol{C}$) in Fig. 2 (a) is assumed to be measured as a proxy variable of $\boldsymbol{U}$. Fig. 2 (a) also provides the graphical representation of the data generating process

$$Y = g_y(X, \boldsymbol{U}, \epsilon_y), \ X = g_x(\boldsymbol{U}, Z, \epsilon_x),$$
$$W = g_w(\boldsymbol{U}, \epsilon_w), \ Z = g_z(\epsilon_z), \ \boldsymbol{U} = \boldsymbol{g}_u(\boldsymbol{\epsilon}_u),$$

where $\epsilon_x, \epsilon_y, \epsilon_z, \epsilon_w,$ and $\boldsymbol{\epsilon}_u$ are independent random disturbances, and $\boldsymbol{g}_u(\boldsymbol{\epsilon}_u)$ is a set of functions of $\boldsymbol{\epsilon}_u$. Such a situation is also discussed in Pearl (2010) and Kuroki and Pearl (2014). However, different from Kuroki and Pearl (2014), $\boldsymbol{U}$ can include an arbitrary number of variables in the present paper. In addition, according to the previous section, we also consider a situation where $\boldsymbol{U}$ is the set of all discrete and continuous variables that influence the way a subject responds to treatments. Thus, in many situations, it is reasonable to assume the existence of a proxy variable W that is independent of $\{X, Y, Z\}$ given $\boldsymbol{U}$, and thus, it would not be difficult to observe such a proxy variable that satisfies the condition. Here, note that the independence assumptions between two observed variables would be affected by partitioning the states of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y\}$.

In the situation shown in Fig. 2 (a), irrespective of the complexity of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y\}$, the impact of $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y\}$ on Y cannot amount to more than a modification of the functional relationship between X and Y. Thus, there are exactly four functions regarding two dichotomous variables X and Y, so the values taken by $\boldsymbol{U} \cup \{\epsilon_x, \epsilon_y\}$ select one of these four functions (Pearl 2009). Considering these observations, according to Rothman, Greenland, and Lash (2008, p.59), the

states of $U \cup \{\epsilon_x, \epsilon_y\}$ are divided into the following four potential outcome types:

$u_1 = (Y_{x_0} = y_0, Y_{x_1} = y_0)$ represents the 'doomed' situation where the treatment received is irrelevant because disease occurs with the experimental or controlled treatment.

$u_2 = (Y_{x_0} = y_0, Y_{x_1} = y_1)$ represents the 'causative' situation where disease occurs if and only if subjects receive the controlled treatment.

$u_3 = (Y_{x_0} = y_1, Y_{x_1} = y_0)$ represents the 'preventive' situation where disease occurs if and only if subjects receive the experimental treatment.

$u_4 = (Y_{x_0} = y_1, Y_{x_1} = y_1)$ represents the 'immune' situation where the treatment received is again irrelevant because disease does not occur, with the experimental or controlled treatment.

According to this partition of the states of $U \cup \{\epsilon_x, \epsilon_y\}$, $U \cup \{\epsilon_x, \epsilon_y\}$ is re-defined as a variable U that takes a value u ($u \in \{u_1, u_2, u_3, u_4\}$). Here, counterfactually, an instrumental variable Z satisfies (i) X_z is a non-trivial function of z, (ii) $Y_{x,z} = Y_x$ holds for any x and z (exclusion restriction), and (iii) Z is independent of $(X_{z_0}, X_{z_1}, Y_{x_0}, Y_{x_1})$. Then, similarly to the setting of the non-compliance problem by Pearl (2009, ch.8), throughout the paper, for any x, y, z and w, we assume that Fig. 2 (a) can be re-described as in Fig. 2 (b) and the corresponding recursive factorization of the conditional probabilities of Y, Z and W given X, $\mathrm{pr}(y, z, w|x)$, is given as

$$\mathrm{pr}(y, z, w \mid x)$$
$$= \sum_{i=1}^{4} \mathrm{pr}(y \mid x, u_i)\,\mathrm{pr}(z \mid x, u_i)\,\mathrm{pr}(w \mid u_i)\,\mathrm{pr}(u_i \mid x)$$

according to Fig. 2 (b). Here, note that the association between X and (Y_{x_0}, Y_{x_1}) is generated through (X_{z_0}, X_{z_1}) (i.e.,the states of $U \cup \{\epsilon_x, \epsilon_y\}$ are divided into sixteen potential outcome types in Balke and Pearl (1997)), but we do not describe (X_{z_0}, X_{z_1}) in Fig. 2 (b), because the result of this paper can be derived by marginalizing over (X_{z_0}, X_{z_1}).

From the consistency property, we have

$$\mathrm{pr}(y_0, z, w, u_i \mid x_0) = \mathrm{pr}(z, w, u_i \mid x_0), \quad i = 1, 2,$$

$$\mathrm{pr}(y_0, z, w, u_i \mid x_1) = \mathrm{pr}(z, w, u_i \mid x_1), \quad i = 1, 3,$$

$$\mathrm{pr}(y_0, z, w, u_i \mid x_0) = 0, \quad i = 3, 4,$$

$$\mathrm{pr}(y_0, z, w, u_i \mid x_1) = 0, \quad i = 2, 4.$$

From these equations, $\mathrm{pr}(y, z, w \mid x)$ and $\mathrm{pr}(z, w \mid x)$ can be rewritten as

$$\mathrm{pr}(y_0, z, w \mid x_0)$$
$$= \sum_{i=1,2} \mathrm{pr}(w \mid u_i)\,\mathrm{pr}(z \mid x_0, u_i)\,\mathrm{pr}(u_i \mid x_0),$$

$$\mathrm{pr}(y_0, z, w \mid x_1)$$
$$= \sum_{i=1,3} \mathrm{pr}(w \mid u_i)\,\mathrm{pr}(z \mid x_1, u_i)\,\mathrm{pr}(u_i \mid x_1),$$

$$\mathrm{pr}(z, w \mid x_0) = \sum_{i=1}^{4} \mathrm{pr}(z \mid x_0, u_i)\,\mathrm{pr}(w \mid u_i)\,\mathrm{pr}(u_i \mid x_0),$$

$$\mathrm{pr}(z, w \mid x_1) = \sum_{i=1}^{4} \mathrm{pr}(z \mid x_1, u_i)\,\mathrm{pr}(w \mid u_i)\,\mathrm{pr}(u_i \mid x_1),$$

and other joint and marginal probabilities of Y, Z and W can be rewritten in a similar way. In addition, when W is a variable with the number of values $k \geq 4$, say w_1, w_2, w_3, w_4, for $x \in \{x_0, x_1\}$, let

$$P_x = \begin{pmatrix} 1 & \mathrm{pr}(z_0 \mid x) \\ \mathrm{pr}(w_1 \mid x) & \mathrm{pr}(w_1, z_0 \mid x) \\ \mathrm{pr}(w_2 \mid x) & \mathrm{pr}(w_2, z_0 \mid x) \\ \mathrm{pr}(w_3 \mid x) & \mathrm{pr}(w_3, z_0 \mid x) \end{pmatrix},$$

$$Q_x = \begin{pmatrix} \mathrm{pr}(y_0 \mid x) & \mathrm{pr}(y_0, z_0 \mid x) \\ \mathrm{pr}(y_0, w_1 \mid x) & \mathrm{pr}(y_0, w_1, z_0 \mid x) \\ \mathrm{pr}(y_0, w_2 \mid x) & \mathrm{pr}(y_0, w_2, z_0 \mid x) \\ \mathrm{pr}(y_0, w_3 \mid x) & \mathrm{pr}(y_0, w_3, z_0 \mid x) \end{pmatrix},$$

$$A_{x_0} = \begin{pmatrix} 1 & \mathrm{pr}(w_1 \mid u_1) & \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_3 \mid u_1) \\ 1 & \mathrm{pr}(w_1 \mid u_2) & \mathrm{pr}(w_2 \mid u_2) & \mathrm{pr}(w_3 \mid u_2) \\ 1 & \mathrm{pr}(w_1 \mid u_3) & \mathrm{pr}(w_2 \mid u_3) & \mathrm{pr}(w_3 \mid u_3) \\ 1 & \mathrm{pr}(w_1 \mid u_4) & \mathrm{pr}(w_2 \mid u_4) & \mathrm{pr}(w_3 \mid u_4) \end{pmatrix},$$

$$A_{x_1} = \begin{pmatrix} 1 & \mathrm{pr}(w_1 \mid u_1) & \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_3 \mid u_1) \\ 1 & \mathrm{pr}(w_1 \mid u_3) & \mathrm{pr}(w_2 \mid u_3) & \mathrm{pr}(w_3 \mid u_3) \\ 1 & \mathrm{pr}(w_1 \mid u_2) & \mathrm{pr}(w_2 \mid u_2) & \mathrm{pr}(w_3 \mid u_2) \\ 1 & \mathrm{pr}(w_1 \mid u_4) & \mathrm{pr}(w_2 \mid u_4) & \mathrm{pr}(w_3 \mid u_4) \end{pmatrix},$$

$$B_{x_0} = \begin{pmatrix} 1 & \mathrm{pr}(z_0 \mid x_0, u_1) \\ 1 & \mathrm{pr}(z_0 \mid x_0, u_2) \\ 1 & \mathrm{pr}(z_0 \mid x_0, u_3) \\ 1 & \mathrm{pr}(z_0 \mid x_0, u_4) \end{pmatrix}, B_{x_1} = \begin{pmatrix} 1 & \mathrm{pr}(z_0 \mid x_1, u_1) \\ 1 & \mathrm{pr}(z_0 \mid x_1, u_3) \\ 1 & \mathrm{pr}(z_0 \mid x_1, u_2) \\ 1 & \mathrm{pr}(z_0 \mid x_1, u_4) \end{pmatrix},$$

$$\Delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$M_{x_0} = \begin{pmatrix} \mathrm{pr}(u_1 \mid x_0) & 0 & 0 & 0 \\ 0 & \mathrm{pr}(u_2 \mid x_0) & 0 & 0 \\ 0 & 0 & \mathrm{pr}(u_3 \mid x_0) & 0 \\ 0 & 0 & 0 & \mathrm{pr}(u_4 \mid x_0) \end{pmatrix},$$

$$M_{x_1} = \begin{pmatrix} \mathrm{pr}(u_1 \mid x_1) & 0 & 0 & 0 \\ 0 & \mathrm{pr}(u_3 \mid x_1) & 0 & 0 \\ 0 & 0 & \mathrm{pr}(u_2 \mid x_1) & 0 \\ 0 & 0 & 0 & \mathrm{pr}(u_4 \mid x_1) \end{pmatrix}.$$

Then, for $x \in \{x_0, x_1\}$, we have

$$P_x = A_x^T M_x B_x, \quad Q_x = A_x^T \Delta M_x B_x, \qquad (2)$$

where the notation "T" indicates a transposed vector/matrix.

Here, for $x \in \{x_0, x_1\}$, let $P_{1,x}$ and $Q_{1,x}$ be 2×2 matrices constructed from the first two rows of P_x and Q_x, respectively. In addition, letting $A_{11,x}$ be a 2×2 block matrix which is constructed by the first and second rows and columns of A_x^T, and $A_{22,x}$ be a 2×2 block matrix which is constructed by the third and fourth rows and columns of A_x^T, suppose that $A_{11,x}$, $A_{22,x}$ and A_x are invertible for

$x \in \{x_0, x_1\}$. Although they include the probabilities of "observed" variable W given "unobserved" potential outcomes U, the violation of the assumptions is testable by checking whether $Q_{1,x}$ and $P_{1,x} - Q_{1,x}$ are invertible or not in some situations. Then, we can derive the following theorem:

Theorem 1 *When Z is an instrumental variable relative to (X, Y) and statistical dependencies among X, Y, Z, W and U are as described by Fig. 2 (b) and the corresponding recursive factorization of $pr(x, y, z, w, u)$, if (i) $A_{11,x}$, $A_{22,x}$, A_x, $Q_{1,x}$ and $P_{1,x} - Q_{1,x}$ are invertible for any x, (ii) the second column vectors of $Q_{2,x}Q_{1,x}^{-1}$ and $(P_{2,x} - Q_{2,x})(P_{1,x} - Q_{1,x})^{-1}$ are different from those of $Q_{2,x'}Q_{1,x'}^{-1}$ and $(P_{2,x'} - Q_{2,x'})(P_{1,x'} - Q_{1,x'})^{-1}$ for $x \neq x'$, and (iii) both $Y_{x,z,w} = Y_x$ and $X_{z,w} = X_z$ hold for any x, z and w, the joint probabilities of potential outcomes $pr(u_i)$ are identifiable $(i = 1, 2, 3, 4)$.*

The proof of Theorem 1 is given in Appendix. Condition (ii), which is testable, is used to provide the unique solution of the simultaneous linear equation. From Theorem 1, the causal effects are identifiable for any x and y and are given by, e.g.,

$$pr(Y_{x_0} = y_0) = \sum_{i=1,2} pr(u_i), pr(Y_{x_1} = y_0) = \sum_{i=1,3} pr(u_i).$$

Here, in the context of "probabilities of causation" (e.g., Tian and Pearl 2002; Cai and Kuroki 2005; Pearl 2009), $pr(u_2)$ is called a "probability of necessity and sufficiency". In addition, for $x \in \{x_0, x_1\}$, $pr(u_2 | x)$ are used to evaluate "probability of necessity" and "probability of sufficiency".

Some Remarks

Here, we would like to state some remarks. First, note that Theorem 1 provides no guarantee of identifying the joint probabilities of the four-dimensional potential outcomes, i.e., $pr(Y_{x_0} = y_i, Y_{x_1} = y_j, X_{z_0} = x_k, X_{z_1} = x_l)$ $(i, j, k, l = 0, 1)$. Second, different from the effect restoration method based on the eigenvalue decomposition (Kuroki and Pearl 2014), note that the proof of Theorem 1 comes down to solving the simultaneous linear equation of $pr(u_1 | x), pr(u_2 | x), pr(u_3 | x), pr(u_4 | x)$ for $x \in \{x_0, x_1\}$. Third, violation of the assumptions is testable: if at least one of $Q_{1,x}$ and $P_{1,x} - Q_{1,x}$ is not invertible, W may not be appropriate for identifying the causal effects or there may be certain functional relationships between potential outcomes. Note especially, Theorem 1 implies that (a) all potential outcome types exist and thus the monotonicity assumption (e.g., no-prevention assumption; $pr(u_3) = 0$) does not hold, and (b) X is not independent of U, and thus X is not randomly assigned. Here, when both an instrumental variable and a proxy variable are available, under similar assumptions, the identification under such monotonicity can be achieved through a slight revision of the proof of Theorem 1.

Finally, to derive narrower bounds than Balke & Pearl's bounds, Cai, Kuroki, and Sato (2007) divided the observed variable W into the six cases shown in Fig. 3. Fig. 3 (a) represents the case that W is a confounder, where W has an
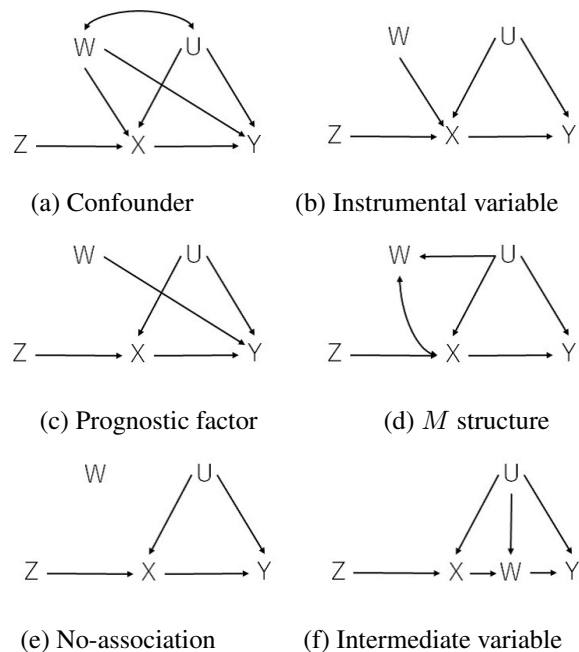

Fig. 3: Graphical representation of causal dependencies in a randomized trial with observed and unobserved covariates (Cai, Kuroki, and Sato 2007)

effect on both X and Y; Fig. 3 (b) represents the case that W satisfies the instrumental variable conditions, where W has an effect on X; Fig. 3 (c) represents the case that W is a prognostic factor, where W has an effect on Y; Fig. 3 (d) represents the case that W is in an "M" structure, where W is a proxy variable of U and associated with X; Fig. 3 (e) represents the case that W is associated with neither X nor Y; finally, Fig. 3 (f) represents the case that W is an intermediate variable between X and Y. Based on Fig. 3, for the case that some minimal causal knowledge is available regarding the relationship between the measured covariates and other variables, for example, when we know that W occurs before X, Cai, Kuroki, and Sato (2007) stated that narrower bounds could be obtained using covariate information. However, they did not take the relationship in Fig. 2 (a) into account when deriving narrower bounds. The results of the present paper demonstrate that, when we observe one proxy variable associated with potential outcomes in the instrumental variable framework, the joint probabilities of potential outcomes can be constructed from the proxy and instrumental variable without the monotonicity assumption, far from bounding the causal quantities, in some situations.

Discussion

Joint probabilities of potential outcomes are of interest in epidemiology, risk analysis, legal reasoning, artificial intelligence, and policy analysis. Therefore, the identification problem of these probabilities has been an important topic in causal inference. To solve the problem, the paper provided a novel identification condition for the probabilities of potential outcomes by using an instrumental variable and

a proxy variable. The proposed condition also enables us to derive consistent estimators of causal effects. To the best of our knowledge, the possibility of using proxies to evaluate the joint probability of potential outcomes in nonparametric systems has not previously appeared in the literature.

Finally, we conducted an asymptotic analysis of bias removal, but inferential aspects were not discussed. When the joint probabilities of potential outcomes are identifiable, although the present paper implicitly assumes that the method of moments is utilized to estimate these quantities, it would be necessary to develop a more efficient estimation method based on singular models. We also assumed that observed variables are dichotomous in the present paper. This assumption can be relaxed by allowing them to have more than two categories, and thus, we can easily extend our results to multicategorical variables, which makes our results applicable to broader variety of situations. However, in such cases, it may be difficult to obtain reliable statistics of the recovered probabilities, due to data sparseness. This is left as future work.

Acknowledgements

Appendix: Proof of Theorem 1

For $x \in \{x_0, x_1\}$, from equation (2) and the assumption, note that we can derive

$$Q_x = A_x^T \Delta A_x^{T-1} P_x.$$

Here, for $x \in \{x_0, x_1\}$, we let

$$A_x^T = \begin{pmatrix} A_{11,x} & A_{12,x} \\ A_{21,x} & A_{22,x} \end{pmatrix},$$

$$P_x = \begin{pmatrix} P_{1,x} \\ P_{2,x} \end{pmatrix}, \quad Q_x = \begin{pmatrix} Q_{1,x} \\ Q_{2,x} \end{pmatrix},$$

where

$$P_{1,x} = \begin{pmatrix} 1 & \mathrm{pr}(z_0 \mid x) \\ \mathrm{pr}(w_1 \mid x) & \mathrm{pr}(w_1, z_0 \mid x) \end{pmatrix},$$

$$P_{2,x} = \begin{pmatrix} \mathrm{pr}(w_2 \mid x) & \mathrm{pr}(w_2, z_0 \mid x) \\ \mathrm{pr}(w_3 \mid x) & \mathrm{pr}(w_3, z_0 \mid x) \end{pmatrix},$$

$$Q_{1,x} = \begin{pmatrix} \mathrm{pr}(y_0 \mid x) & \mathrm{pr}(y_0, z_0 \mid x) \\ \mathrm{pr}(y_0, w_1 \mid x) & \mathrm{pr}(y_0, w_1, z_0 \mid x) \end{pmatrix},$$

$$Q_{2,x} = \begin{pmatrix} \mathrm{pr}(y_0, w_2 \mid x) & \mathrm{pr}(y_0, w_2, z_0 \mid x) \\ \mathrm{pr}(y_0, w_3 \mid x) & \mathrm{pr}(y_0, w_3, z_0 \mid x) \end{pmatrix}.$$

In addition,

$$A_{11,x_0} = \begin{pmatrix} 1 & 1 \\ \mathrm{pr}(w_1 \mid u_1) & \mathrm{pr}(w_1 \mid u_2) \end{pmatrix},$$

$$A_{12,x_0} = \begin{pmatrix} 1 & 1 \\ \mathrm{pr}(w_1 \mid u_3) & \mathrm{pr}(w_1 \mid u_4) \end{pmatrix},$$

$$A_{21,x_0} = \begin{pmatrix} \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_2 \mid u_2) \\ \mathrm{pr}(w_3 \mid u_1) & \mathrm{pr}(w_3 \mid u_2) \end{pmatrix},$$

$$A_{22,x_0} = \begin{pmatrix} \mathrm{pr}(w_2 \mid u_3) & \mathrm{pr}(w_2 \mid u_4) \\ \mathrm{pr}(w_3 \mid u_3) & \mathrm{pr}(w_3 \mid u_4) \end{pmatrix},$$

$$A_{11,x_1} = \begin{pmatrix} 1 & 1 \\ \mathrm{pr}(w_1 \mid u_1) & \mathrm{pr}(w_1 \mid u_3) \end{pmatrix},$$

$$A_{12,x_1} = \begin{pmatrix} 1 & 1 \\ \mathrm{pr}(w_1 \mid u_2) & \mathrm{pr}(w_1 \mid u_4) \end{pmatrix},$$

$$A_{21,x_1} = \begin{pmatrix} \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_2 \mid u_3) \\ \mathrm{pr}(w_3 \mid u_1) & \mathrm{pr}(w_3 \mid u_3) \end{pmatrix},$$

$$A_{22,x_1} = \begin{pmatrix} \mathrm{pr}(w_2 \mid u_2) & \mathrm{pr}(w_2 \mid u_4) \\ \mathrm{pr}(w_3 \mid u_2) & \mathrm{pr}(w_3 \mid u_4) \end{pmatrix}.$$

Then, for $x \in \{x_0, x_1\}$, noting that A_x^{T-1} can be represented as

$$A_x^{T-1} = \begin{pmatrix} A_x^{11} & -A_x^{11} A_{12,x} A_{22,x}^{-1} \\ -A_{22,x}^{-1} A_{21,x} A_x^{11} & A_x^{22} \end{pmatrix},$$

where

$$A_x^{11} = (A_{11,x} - A_{12,x} A_{22,x}^{-1} A_{21,x})^{-1},$$

$$A_x^{22} = (A_{22,x} - A_{21,x} A_{11,x}^{-1} A_{12,x})^{-1},$$

we can derive

$$A_x^T \Delta A_x^{T-1} = \begin{pmatrix} A_{11,x} A_x^{11} \\ A_{21,x} A_x^{11} \end{pmatrix} \left(I_{2,2}; -A_{12,x} A_{22,x}^{-1} \right),$$

where $I_{2,2}$ is a 2×2 identity matrix. Thus, we can derive

$$Q_{1,x} = A_{11,x} A_x^{11} (P_{1,x} - A_{12,x} A_{22,x}^{-1} P_{2,x}),$$

$$Q_{2,x} = A_{21,x} A_x^{11} (P_{1,x} - A_{12,x} A_{22,x}^{-1} P_{2,x}).$$

From these equations, we have $Q_{2,x} = A_{21,x} A_{11,x}^{-1} \times Q_{1,x}$, i.e.,

$$A_{21,x} = Q_{2,x} Q_{1,x}^{-1} A_{11,x} \tag{3}$$

and

$$A_x^{11,-1} A_{11,x}^{-1} Q_{1,x} = Q_{1,x} - A_{12,x} A_{22,x}^{-1} A_{21,x} A_{11,x}^{-1} Q_{1,x}$$
$$= P_{1,x} - A_{12,x} A_{22,x}^{-1} P_{2,x},$$

which gives

$$Q_{1,x} - P_{1,x} = A_{12,x} A_{22,x}^{-1} (A_{21,x} A_{11,x}^{-1} Q_{1,x} - P_{2,x})$$
$$= A_{12,x} A_{22,x}^{-1} (Q_{2,x} - P_{2,x}),$$

i.e.,

$$A_{22,x} = (Q_{2,x} - P_{2,x})(Q_{1,x} - P_{1,x})^{-1} A_{12,x}. \tag{4}$$

Here, for $x \in \{x_0, x_1\}$, letting a_{ij}^x and b_{ij}^x be the (i,j) elements of observed matrices $Q_{2,x} Q_{1,x}^{-1}$ and $(Q_{2,x} - P_{2,x})(Q_{1,x} - P_{1,x})^{-1}$, respectively $(i, j = 1, 2)$, from equation (3), we have

$$Q_{2,x_0} Q_{1,x_0}^{-1} A_{11,x_0}$$
$$= \begin{pmatrix} a_{11}^{x_0} + a_{12}^{x_0} \mathrm{pr}(w_1 \mid u_1) & a_{11}^{x_0} + a_{12}^{x_0} \mathrm{pr}(w_1 \mid u_2) \\ a_{21}^{x_0} + a_{22}^{x_0} \mathrm{pr}(w_1 \mid u_1) & a_{21}^{x_0} + a_{22}^{x_0} \mathrm{pr}(w_1 \mid u_2) \end{pmatrix}$$
$$= A_{21,x_0} = \begin{pmatrix} \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_2 \mid u_2) \\ \mathrm{pr}(w_3 \mid u_1) & \mathrm{pr}(w_3 \mid u_2) \end{pmatrix},$$

$$Q_{2,x_1} Q_{1,x_1}^{-1} A_{11,x_1}$$

$$= \left(\begin{array}{cc} a_{11}^{x_1} + a_{12}^{x_1} \mathrm{pr}(w_1 \mid u_1) & a_{11}^{x_1} + a_{12}^{x_1} \mathrm{pr}(w_1 \mid u_3) \\ a_{21}^{x_1} + a_{22}^{x_1} \mathrm{pr}(w_1 \mid u_1) & a_{21}^{x_1} + a_{22}^{x_1} \mathrm{pr}(w_1 \mid u_3) \end{array} \right)$$

$$= A_{21,x_1} = \left(\begin{array}{cc} \mathrm{pr}(w_2 \mid u_1) & \mathrm{pr}(w_2 \mid u_3) \\ \mathrm{pr}(w_3 \mid u_1) & \mathrm{pr}(w_3 \mid u_3) \end{array} \right),$$

$$A_{22,x_0} = \left(\begin{array}{cc} \mathrm{pr}(w_2 \mid u_3) & \mathrm{pr}(w_2 \mid u_4) \\ \mathrm{pr}(w_3 \mid u_3) & \mathrm{pr}(w_3 \mid u_4) \end{array} \right)$$

$$= (Q_{2,x_0} - P_{2,x_0})(Q_{1,x} - P_{1,x_0})^{-1} A_{12,x_0}$$

$$= \left(\begin{array}{cc} b_{11}^{x_0} + b_{12}^{x_0} \mathrm{pr}(w_1 \mid u_3) & b_{11}^{x_0} + b_{12}^{x_0} \mathrm{pr}(w_1 \mid u_4) \\ b_{21}^{x_0} + b_{22}^{x_0} \mathrm{pr}(w_1 \mid u_3) & b_{21}^{x_0} + b_{22}^{x_0} \mathrm{pr}(w_1 \mid u_4). \end{array} \right)$$

$$A_{22,x_1} = \left(\begin{array}{cc} \mathrm{pr}(w_2 \mid u_2) & \mathrm{pr}(w_2 \mid u_4) \\ \mathrm{pr}(w_3 \mid u_2) & \mathrm{pr}(w_3 \mid u_4) \end{array} \right)$$

$$= (Q_{2,x_1} - P_{2,x_1})(Q_{1,x_1} - P_{1,x_1})^{-1} A_{12,x_1}$$

$$= \left(\begin{array}{cc} b_{11}^{x_1} + b_{12}^{x_1} \mathrm{pr}(w_1 \mid u_2) & b_{11}^{x_1} + b_{12}^{x_1} \mathrm{pr}(w_1 \mid u_4) \\ b_{21}^{x_1} + b_{22}^{x_1} \mathrm{pr}(w_1 \mid u_2) & b_{21}^{x_1} + b_{22}^{x_1} \mathrm{pr}(w_1 \mid u_4) \end{array} \right)$$

From these equations, $\mathrm{pr}(w|u)$ is identifiable. In addition, for $x \in \{x_0, x_1\}$, since A_x is identifiable, comparing the first columns of the left- and right-hand sides of

$$A_x^{T-1} P_x = M_x B_x,$$

both M_{x_0} and M_{x_1} are identifiable. Furthermore, since both $\mathrm{pr}(x_0)$ and $\mathrm{pr}(x_1)$ are estimable from observed probabilities, the joint probabilities of potential outcomes are identifiable and the causal effects are also identifiable.

References

Angrist, J. D.; Imbens, G. W.; and Rubin, D. B. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91(434): 444–455.

Balke, A.; and Pearl, J. 1997. Bounds on Treatment Effects From Studies With Imperfect Compliance. *Journal of the American Statistical Association* 92(439): 1171–1176.

Cai, Z.; and Kuroki, M. 2005. Variance Estimators for Three "Probabilities of Causation". *Risk Analysis* 25(6): 1611–1620.

Cai, Z.; Kuroki, M.; and Sato, T. 2007. Non-parametric bounds on treatment effects with non-compliance by covariate adjustment. *Statistics in Medicine* 26(16): 3188–3204.

Greenland, S. 2000. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 29(4): 722–729.

Heckman, J. J.; Smith, J.; and Clements, N. 1997. Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies* 64(4): 487–535.

Imbens, G. W.; and Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Kuroki, M.; and Cai, Z. 2011. Statistical analysis of 'probabilities of causation' Using covariate information. *Scandinavian Journal of Statistics* 38(3): 564–577.

Kuroki, M.; and Pearl, J. 2014. Measurement bias and effect restoration in causal inference. *Biometrika* 101(2): 423–437.

Li, A.; and Pearl, J. 2019. Unit Selection Based on Counterfactual Logic. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* 1793–1799.

Lui, K.-J. 2011. Binary Data Analysis of Randomized Clinical Trials with Noncompliance. *John Wiley & Sons* 1–20.

Manski, C. F. 1990. Nonparametric bounds on treatment effects. *American Economic Review, Papers and Proceedings* 80(2): 319–323.

Mark, S. D.; and Robins, J. M. 1993. A method for the analysis of randomized trials with compliance information; An application to the Multiple Risk Factor Intervention Trial. *Controlled Clinical Trials* 14(1): 79–97.

Multiple Risk Factor Intervention Trial Research Group. 1982. Multiple Risk Factor Intervention Trial: risk factor changes and mortality results. *Journal of the American Medical Association* 248: 1465–1477.

Pearl, J. 2001. Direct and Indirect Effects. *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* 411–420.

Pearl, J. 2009. Causality: Models, Reasoning and Inference. *Cambridge University Press* 283–308.

Pearl, J. 2010. On Measurement Bias in Causal Inference. *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence* 425–432.

Robins, J. 1989. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health Service Research Methodology: A Focus on AIDS* 113–159.

Robins, J.; and Richardson, T. S. 2011. Alternative graphical causal models and the identification of direct effects. *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures* 103–158.

Rosenbaum, P. R.; and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41–55.

Rothman, K.; Greenland, S.; and Lash, T. 2008. Modern Epidemiology. *Lippincott Williams & Wilkins* 59.

Sheiner, L. B.; and Rubin, D. B. 1995. Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics* 57(1): 6–15.

Tian, J.; and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* 28(1): 287–313.

Tian, J.; and Pearl, J. 2002. A General Identification Condition for Causal Effects. *Proceedings of the 18th National Conference on Artificial Intelligence* 567–573.

Wang, L.; and Tchetgen, E. T. 2018. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society. Series B, Statistical methodology* 80(3): 531.

Yamada, K.; and Kuroki, M. 2017. Counterfactual-Based Prevented and Preventable Proportions. *Journal of Causal Inference* 5: 20160020, eISSN 2193–3685.

Yamada, K.; and Kuroki, M. 2019. New Traffic Conflict Measure Based on a Potential Outcome Model. *Journal of Causal Inference* 7: 20180001, eISSN 2193–3685.

Yan, W.; Ding, P.; Geng, Z.; and Zhou, X. 2011. Identifiability of causal effects on a binary outcome within principal strata. *Frontiers of Mathematics in China* 6(6): 1249–1263.