Direct and Indirect Treatment Effects with Social Interaction

Siwon Ryu

August 23, 2024

Abstract

This study analyzes identifications and estimations in the presence of social interactions. The potential outcomes are functions of own treatment status and exposures, and the exposure is a function of neighbors' treatment status. If the distribution of exposures is determined by own treatment status, the treatment effect can be decomposed into direct and indirect effects using an approach mediation model. Suppose the exposure distribution is from the underlying random graph of the social network. Then, it can be interpreted as the treatment has indirect effects by changing the underlying network structure. Therefore, the exposures play the role of mediator, and the variation of the mediator due to the treatment status identifies the direct and indirect treatment effects separately. Monte-Carlo simulation studies and the empirical application of the impact of co-educated high school on academic performance show the proposed estimators and decompositions work.

Keywords: Causal inference; Network change; Mediation Effects

1 Introduction

Identification and estimation of causal effects of a program or a policy have been of great interest in economic analysis. Rubin's causal model is a popular approach to dealing with causal effects. A fundamental assumption is the Stable Unit Treatment Value Assumption (SUTVA) that each individual's potential outcomes are entirely determined by their own treatment status. However, as in Kline and Tamer (2020)'s review, social interaction can provide another channel of a program affecting outcomes. In such cases, potential outcomes can be expressed as a function of the other one's treatment status in addition to the own treatment status.

For instance, consider a society consisting of N individuals. Each individual has M neighbors and interacts with others within their neighborhood. The potential outcome of each individual can be written as

$$Y_i(d, \tilde{\boldsymbol{d}}) = m(d, \boldsymbol{d}, \varepsilon_i),$$

where the first argument d is the potential value of individual i's treatment assignment, and the second argument $\tilde{\boldsymbol{d}} = (d_1, ..., d_M)$ is a $M \times 1$ vector of individual i's neighbors' treatment asssignments. The vector $\tilde{\boldsymbol{d}}$ could have an order, so that d_j is the i's jth neighbor. This model violates SUTVA if $Y_i(d, \tilde{\boldsymbol{d}}) \neq Y_i(d, \hat{\boldsymbol{d}})$ for $\tilde{\boldsymbol{d}} \neq \hat{\boldsymbol{d}}$. Moreover, note that each individual has 2^{M+1} potential outcomes. If the number of neighbors M increases with the population in the society, then the number of potential outcomes becomes large when the sample size is large.

However, the number of *effective* treatment effects is likely to be less than 2^{M+1}. For example, the potential outcome could depend on the own treatment status, and the number of treated neighbors. In this case, the domain of potential outcomes is $\{0, 1\} \times \mathbb{Z}_+$, which is 2 dimensional. The effective exposure from the neighbors' treatment status can be summarized by a function $\psi : \{0, 1\}^M \to \Psi$. And that, there exists $\tilde{Y} : \{0, 1\} \times \Psi \to \mathbb{R}$ such that

$$\tilde{Y}_i(d, \psi(\tilde{\boldsymbol{d}})) = Y_i(d, \tilde{\boldsymbol{d}}), \quad \forall (d, \tilde{\boldsymbol{d}}) \in \{0, 1\}^M.$$

This function ψ is called an exposure map or treatment rule in the literature. The treatment effect is the effect of changes in one's treatment status on the outcome when their exposures are fixed. And the spillover effects or exposure effects are defined as the effect of changes in the exposures on the outcome when their own treatment status is fixed.

The treatment and exposure effects are causal effects in that they are differences in the potential outcomes. If the own treatment assignment and the exposures are independent, then the treatment effect is a pure causal effect of the own treatment. However, if the own treatment status determines the distribution of exposure ψ, then the treatment effect can be divided into direct and indirect causal effects. Let ψ_{id} be the exposure of i when $D_i = d$. Then, the observed exposure of individual i is $\psi_i = D_i\psi_{i1} + (1 - D_i)\psi_{i0}$. The treatment effect can be written as

$$Y(1, \psi_1) - Y(0, \psi_0) = [Y(1, \psi_1) - Y(1, \psi_0)] + [Y(1, \psi_0) - Y(0, \psi_0)].$$

The first term on the right-hand side represents an indirect effect, which means that the effect of treatment when is only changing the exposure distribution. And the second term is a direct effect, which means that the effect of treatment when their distribution of exposure is not changed.

If the exposure is the number of treated neighbors as the previous example, suppose that $q(d)$ is the expected value of ψ_{id} for $d \in \{0, 1\}$, where $q(1) > q(0)$. This implies that a treated individual has more treated neighbors on average. The direct treatment effect measures the difference between the potential outcomes of a treated and untreated individual when their distributions of the number of treated neighbors are the same. On the other hand, the indirect treatment effects measure the difference between potential outcomes if only the distribution of the number of the treated neighbor is changed.

The main contribution of this paper is identifying and estimating the direct and indirect treatment effects separately using a mediation model approach. The decomposition of the overall treatment effects can be used to figure out the mechanisms of a policy affecting the outcomes. If data consists of the full information of the adjacency matrix of the social network, then we can construct various exposure maps, which will give rich information. But this framework can be applied when we don't have complete information on social networks because estimators need data on observed outcomes, treatment status, and some exposures, which is an advantage of this model.

For the identification of decomposition, I use a mediation model by regarding the exposure as a mediator of treatment effects on the potential outcomes. Corresponding frequency estimators are proposed, and their asymptotic properties are derived. Compared to the usual mediation model, one difficulty is that the exposure ψ_i may not be independent across individuals. For instance, ψ_i and ψ_j are correlated if i) i and j are connected; ii) i and j have common neighbors. To come up with this dependency, I assume a boundedness condition of the data dependency graph. Intuitively, it assumes that this dependency will likely be ignorable when the sample is sufficiently large.

The outline of this paper is as follows. Section 2, the settings and model used in this paper are introduced. Section 3 discusses identification and estimation. Section 4 shows results from Monte Carlo studies to verify the performance of estimators. Section 5 is an empirical application to illustrate the proposed method. Section 6 concludes.

Related Literature

Rubin causal model (Rubin (1974), Imbens and Rubin (2010)) based on the potential outcome framework is widely used in economic analysis for identifying and estimating treatment effects. The model assumes SUTVA, which rules out interactions between individuals on their potential outcomes. In this respect, Manski (2013) refer SUTVA as the individualistic treatment response (ITR). However, in the presence of social interaction, the treatment response of each individual may depends on the entire treatment vector in that society in general. Manski (2013) identify distributions of potential outcomes under the constant treatment response (CTR) assumption. Let $Y_i(\boldsymbol{d})$ be the potential outcome of individual i when he faces $\boldsymbol{d}$ as the entire treatment vector. Then, CTR assumption states that there exists a function c_i such that $c_i(\boldsymbol{d}) = c_i(\boldsymbol{d}')$ implies $Y_i(\boldsymbol{d}) = Y_i(\boldsymbol{d}')$. Manski call the image of this function as the effective treatments. The function c_i is called the exposure map or the treatment rule in the literature. CTR assumption is a generalization of ITR or SUTVA because they are the special case $c_i(\boldsymbol{d}) = d_i$, and c_i is the identity function for all i.

Because the general unrestricted model, in which $c_i(\boldsymbol{d}) = \boldsymbol{d}$ is difficult to analyze because of dimensionality problem of domain of potential outcomes, the concept of effective treatment is a current convention in the literatrue. Studies usually assume the effective treatment is finite dimensional space, so it reduces the dimension of the treatment and the number of potential outcomes. Forastiere, Airoldi, and Mealli (2021) postulate the Stable Unit Treatment on Neighborhood Value Assumption (SUTNVA) in that potential outcomes are functions of treatment vector of neighborhood only. This is a natural extension of SUTVA to neighborhood interaction. Aronow and Samii (2017), Vazquez-Bare (2022) assume that the size of neighborhoods are fixed, so there are fixed number of potential outcomes for each individuals.

As such studies, once we fix the dimension of effective treatment, or the domain of potential outcomes, then identifying treatment effect parameters are similar to those of multiple treatment model. Corresponding independence assumptions or ignorabilities imply the identification. As related works, Leung (2020) focus on the number of treated neighbors and the number of neighbors (i.e., degree) as the exposures. The model is therefore a linear in means model without endogenous peer effect. The author identifies the average and quantile treatment, exposure effects, and derive their asymptotic normality by using a new version of asymptotic theories. Vazquez-Bare (2022) derive general identification argument of this type of model and propose some estimators. The author assume there are independent groups in data, and the interaction arises within group. By controling the size of group and the number of group, he show asymptotic normality of the frequency estimator.

One important problem is the possibility of misspecification of the exposure map. If the exposure map is incorrect, then the identification and corresponding estimators would be misleading. In this respect, Vazquez-Bare (2022) also shows that if the true exposure map is coarser then the exposure map used in the estimator, then the average potential outcomes are identified as the usualy frequcney estimands, and some weighted average of those frequency estimands otherwise. Intuitively, the less coarsity of true exposure map means the information contained in the exposure map is more rich than that from the true exposure map. Leung (2022) assumes *approximately neighborhood interaction (ANI)* assumption to identify the treatment parameteres when the exposure map is possibly misspecified. This assumption is that the potential outcome distributions are primarily determined by the neighbors within some close distance.

However, I assume we have correctly specified exposure map in this study, to focus on the dependency between exposure and the treatment assignments.

Most studies in this literature assume the underlying social network is fixed or independent of treatment assignment. These assumptions exclude the possibility that a policy can change the underlying random graphs of social networks. Some studies consider this possibility. Comola and Prina (2021) suggests a two-period model in that treatments are assigned in the first period, and the network can change in the second period. The outcome in each period follows the linear in means model as in Bramoullé, Djebbari, and Fortin (2009), so they use the similar identifying assumption that there is no intransitive triad in networks of both periods. To apply this model, researchers need to have full information about network adjacency matrices in both periods. However, It would sometimes be hard to get such descent information.

If potential outcomes are functions of the own treatment status and the exposures, the exposures are determined by the underlying social networks. In this study, I assume that the distributions of exposures are influenced by their own treatment status. That is, the own treatment status could change the functional form of the exposure map, or change the random graph of social networks. In both cases, exposure can be thought of as a mediator of own treatment on potential outcomes. In the literature on mediation models, the main purpose is to figure out the mechanism of how treatment influences potential outcomes by decomposing the total treatment effect into direct and indirect effects. Suppose $M(d)$ is the potential mediator when own treatment status is given by $d \in \{0, 1\}$. Then, the observed mediator is $M = dM(1) + (1 - d)M(0)$. The direct treatment effect is the effect of the own treatment status when the mediator is fixed, and the indirect treatment effect is the effect of the own treatment only through changes in the mediator. If both treatment assignment and the distributions of mediators are independent of potential outcomes, then it is straightforward to identify distributions of direct and indirect treatment effects separately. Huber (2014) suggests sequential ignorability assumptions, which are weaker conditions than the full independence to identify the direct and indirect effects, and I follow these assumptions and identification arguments in this study.

Even if the treatments are randomly assigned, the data of outcomes and exposures could be dependent due to social interactions. This dependency makes it tricky to derive asymptotic properties. In the case of the sum of independent random variables, Esseen's method (Esseen (1945)) is convenient to approximate normal distribution. Vazquez-Bare (2022) uses the Berry-Esseen bound to derive the asymptotic normality by assuming exposures are independent across groups. However, it is difficult to apply Esseens' method to dependent data. Instead, Stein's method (Stein (1972)) is widely used to deal with dependent data. As an example, Chen and Shao (2004) provides a version of the central limit theorem with a bounded maximum degree of the dependency graph. Leung (2020) derived conditions on the moment of dependency graph instead of directly applying Stein's method, but in this study, I assume boundedness of the maximum degree of dependency graph and use Stein's method.

2 Model

2.1 Notation and the exposure map

Suppose we have a random sample consisting of N individuals. Let $Y_i \in \mathbb{R}$ be outcome and $D_i \in \{0,1\}$ be a binary treatment indicator for individual i. The potential outcome for individual i can be written as a function of the entire treatment vector: $Y_i(\boldsymbol{D})$, for $\boldsymbol{D} \in \{0,1\}^N$. Each individual can interact with others and let M be the number of others who interact with individual i.[1]. Define $\tilde{\boldsymbol{D}}_i = (D_{1i}, ..., D_{Mi})$, where D_{ji} denotes the treatment status of i's jth neighbor. This setting is similar to that of *SUTNVA* in Forastiere, Airoldi, and Mealli (2021). That is, there exists $\tilde{Y}_i : \{0,1\}^M \to \mathbb{R}$ such that

$$Y_i(\boldsymbol{D}) = \tilde{Y}_i(D_i, \tilde{\boldsymbol{D}}_i).$$

By abusing the notation, let $Y_i(D_i, \tilde{\boldsymbol{D}}_i) \equiv \tilde{Y}_i(D_i, \tilde{\boldsymbol{D}}_i)$ be the potential outcome.

Next, as similar to Vazquez-Bare (2022), assume there is a known exposure map ψ that satisfies $Y_i(d, \tilde{\boldsymbol{d}}) = Y_i(d, \psi(\tilde{\boldsymbol{d}}))$ for all $\tilde{\boldsymbol{d}} \in \{0,1\}^M$, where $\psi : \{0,1\}^M \to \Psi$. Because $\tilde{\boldsymbol{D}}_i$ has finite support, Ψ is also finite. The dimension of Ψ should be less than

[1] Hence, M is assumed to be the same for all individuals. When each individual has different number of neighbors, i.e., M_i, then the asymptotic argument could be modified conditional on M_i

N. The exposure map is a treatment rule about how the treatment vector of neighbors is related to the potential outcome. In this study, the exposure map is assumed to be correctly specified.[2,3]

In this setting, the treatment effect is defined as the effect of change in D_i on the outcome, and the spillover effect or the exposure effect is defined as the effect of change in $\psi(\tilde{\boldsymbol{D}}_i)$ on the outcome. For example, Leung (2022) use $\psi(\tilde{\boldsymbol{D}}_i) = \mathbb{1}\left\{\sum_j A_{ij} D_j > 0\right\}$ in his empirical applications. In this case, the exposure effect is the difference in potential outcomes between when an individual has treated neighbors and when there is no treated neighbor.

Compared to the previous studies in literature, I assume that the distribution of $\tilde{\boldsymbol{D}}_i$ can be different according to the own treatment status. $\tilde{\boldsymbol{D}}_i$ may include some information about the underlying social network. Therefore, this allows individual i's own treatment status to affect his network formation or link status to others. Let $\tilde{\boldsymbol{D}}_i(d)$ be the potential vector of neighbors' treatment of individual i, when $D_i = d$. Define $\psi_{id} = \psi(\tilde{\boldsymbol{D}}_i(d))$. Then, $\psi_i = \psi(\tilde{\boldsymbol{D}}_i) = D_i \psi_{i1} + (1 - D_i)\psi_{i0}$. Therefore, ψ_i is the observed exposure value, and ψ_{id} are potential exposure values following some potential distributions.

Note that the number of potential outcomes is $2|\Psi|$. Therefore, if distributions of ψ_{i1}, ψ_{i0} are the same, or treatment does not affect exposure distributions,[4] the model becomes a potential outcome model with multiple treatments. And then, as in Vazquez-Bare (2022), Leung (2020), identification follows corresponding independence assumptions. The average potential outcomes can be estimated by the frequency estimator computed in each cell.

Because of the dependence between treatment and exposure map, the treatment effect should be redefined as the effect of an *exogenous change* in the own treatment status on the potential outcome, to be interpreted as a causal effect. Otherwise, the treatment status changes the exposure distribution, so the usual treatment effect would

[2] Aronow and Samii (2017) discuss about when the exposure map is misspecified. Leung (2022) proposed a solution when the exposure map is misspecified by using the concept of *approximated neighborhood inference*.

[3] Constructing a model using an exposure map is convenient when only limited information of the network structure is available in data instead of the full information of the adjacency matrix.

[4] If exposure is determined by underlying network structure, this is the case when the network is fixed or independent of treatment.

be a mixed effect of direct and indirect effects. The main contribution of this study is to identify and estimate such direct and indirect effects separately. Also, this setting is a generalization of the model used in the literature because the cases when the underlying network is fixed or independent of treatment are special cases of this setting.

This model is compatible for the models in the literature. When $\psi(\cdot) = c$ for some constant c for $d = 0, 1$, then the model satisfies SUTVA. When $\psi(\tilde{\boldsymbol{d}}) = \tilde{\boldsymbol{d}}$, then this is the model with unrestricted interaction. When $\tilde{\boldsymbol{D}}_i(1) = \tilde{\boldsymbol{D}}_i(0)$, then the model is the same as the treatment effect with social interactions and correcly specified exposure map as Vazquez-Bare (2022). When $\psi(\cdot)$ is defined as Leung (2020), i.e., $\psi = \frac{1}{\sum_j A_{ij}} \sum_j A_{ij} D_j$, the model becomes a linear-in-means model without endogenous peer effect.

If the full adjacency matrix representing the social network is available in data, then the exposure map used in Leung (2020), Leung (2022) can be extended as

$$\psi(\tilde{\boldsymbol{D}}_i(d)) = \left(\sum_j A_{ij}(d) D_j, \sum_j A_{ij}(d) \right), \quad \text{or,} \quad \psi(\tilde{\boldsymbol{D}}_i(d)) = \mathbb{1}\left\{ \sum_j A_{ij}(d) D_j > 0 \right\},$$

where A_{ij} is the $(i, j)-$th element of the adjacency matrix A, which depends on the own treatment status d. The potential vector of neighbor's treatments is determined by $A_i(d)$. If A represents an undirected network, then it is hard to say that the distribution of $(A_{i1}, ...A_{iN})$ only depends on i's own treatment status. Therefore, I assume the underlying network is directed. Therefore, if the underlying network represent a friendship network, then A_{ij} is 1 if i *thinks* j as his friend.

2.2 Parameters of interest

Let $\psi_i \equiv \psi(\tilde{\boldsymbol{D}}_i)$ and $\psi_{id} \equiv \psi(\tilde{\boldsymbol{D}}_i(d))$ for $d = 0, 1$. Then, the observed value of ψ_i is $\psi_i = D_i \psi_{i1} + (1 - D_i)\psi_{i0}$, and the observed outcome can be written as

$$\begin{aligned} Y_i &= D_i Y_i(1, \psi_{i1}) + (1 - D_i) Y_i(0, \psi_{i0}) \\ &= \sum_{d \in \{0,1\}} \sum_{s \in \Psi} \mathbb{1}\{\psi_i = s, D_i = d\} Y_i(d, s) \end{aligned} \tag{1}$$

Data consists of $\{(Y_i, \psi_i, D_i) : 1 \leq i \leq N\}$. Thus, one of $Y_i(1, \psi_{1i})$ and $Y_i(0, \psi_{i0})$ is observed, while $Y_i(d, \psi_{id'})$ for $d \neq d'$ are never observed.

Once D_i is given, it has a direct effect on the potential outcome, and it determines the distribution of exposure ψ_i. And then, the distribution of exposure affects the potential outcome. Therefore, this model can be thought of as a mediation model. The exposure ψ_i is a mediator of treatment effect on the potential outcome. Hence, following the mediation model literature, we can decompose the overall treatment effects as follows.

The *average overall treatment effect (ATE)* is defined as the mean difference between potential outcomes when the own treatment is exogenously changed:

$$\Delta \equiv E[Y_i(1, \psi_{i1}) - Y_i(0, \psi_{i0})].$$

The *average direct treatment effect (ADTE)* can be defined as the average difference between potential outcomes when the own treatment is exogenously changed, but the mediator is fixed (network structure is fixed) at its potential distribution for given $d \in \{0, 1\}$:

$$\theta(d) \equiv E[Y_i(1, \psi_{id}) - Y_i(0, \psi_{id})].$$

Similarly, *the average indirect treatment effect (AITE)* is defined as the average difference between potential outcomes when the distribution of the mediator is exogenously changed. Still, the own treatment status is fixed at $d \in \{0, 1\}$:

$$\delta(d) \equiv E[Y_i(d, \psi_{i1}) - Y_i(d, \psi_{i0})].$$

By construction, the ATE is decomposed by the sum of DTE and ITE:

$$\Delta = \theta(0) + \delta(1) = \theta(1) + \delta(0).$$

The spillover effects or the exposure effects are the effects on potential outcomes when the other's treatment status is exogenously changed. Compared to the treatment effect, there is no indirect spillover effect because the other's treatments do not affect the link status[5]. Direct exposure effects are the difference between potential outcomes for two exposure map values. Still, the treatment status is fixed at d, so that the exposure distribution is fixed at ψ_{id}. Let s, s' be two different values of ψ_{id}. The *(direct) spillover effect* is defined as $\tau(d, s, s') \equiv E[Y_i(d, s) - Y_i(d, s')]$.

3 Identification and Estimation

In this subsection, the identification results are discussed. First, the followings are identifying assumptions.

Assumption 1. $\{Y_i(1, s), Y_i(0, s), \psi_{i1}, \psi_{i0} : s \in \Psi\}$ *are independent of* D_i.

Assumption 2. $\{Y_i(1, s), Y_i(0, s) : s \in \Psi\}$ *are independent of* ψ_i *conditional on* D_i.

These assumptions are similar to identifying assumptions in Huber (2014), called the sequential independence assumptions. Because $\tilde{\boldsymbol{D}}_i(d)$ consists of neighbors' treatment assignments which are independent with D_i, Assumption 1 is satisfied when treatments are randomly assigned. Once $D_i = d$ is given, $\psi_i = \psi_{id}$, and hence the second assumption is that potential outcomes are independent of distribution of potential exposures after treatment is assigned. Assumption 2 fails if there is a common factor in determining both potential outcomes and exposures. This is the case that, for example, there exists a random variable X_i such that $Y_i(d, \tilde{\boldsymbol{d}}) = m(d, \tilde{\boldsymbol{d}}, X_i, \varepsilon_i)$ and $\psi_i = \psi(D_i, \tilde{\boldsymbol{d}}, X_i)$.

Assumption 3. *For each* $d \in \{0, 1\}$ *and* $s \in \Psi$, $P(d, s) \equiv P(D_i = d, \psi_i = s) > 0$.

Assumption 3 is the usual overlap assumption. If $P(d, s)$ are zero for some d, s, then distribution of $Y(d, s)$ are not identified because the conditioning event has zero probability. The following lemma states that the distribution of interests is identified.

[5] ***

Lemma 1 (Identification of distributions)**.** *Under Assumptions 1, and 2,*

$$\begin{aligned}
G^{d,s}(y) &\equiv \Pr\left(Y_i(d,s) \leq y\right) = \Pr\left(Y_i \leq y | D_i = d, \psi_i = s\right), \\
F^{d,d'}(y) &\equiv \Pr(Y_i(d, \psi_{id'}) \leq y) = \sum_{s \in \Psi} \Pr(Y_i \leq y | D_i = d, \psi_i = s) \Pr(\psi_i = s | D_i = d'), \\
F^{d}(y) &\equiv F^{d,d}(y) = \Pr(Y_i(d, \psi_{id}) \leq y) = \Pr(Y_i \leq y | D_i = d).
\end{aligned}$$

Proof. Notice that Assumption 1 and Assumption 3 implies

$$\Pr(\psi_{id} = s) = \Pr(\psi_{id} = s | D_i = d) = \Pr(\psi_i = s | D_i = d) > 0.$$

Also, Assumption 1 implies $\Pr(Y_i(d,s) \leq y) = \Pr(Y_i(d,s) \leq y | D_i = d)$. It follows that

$$\begin{aligned}
G^{d,s}(y) &\equiv \Pr\left(Y_i(d,s) \leq y\right) \\
&= \frac{\Pr\left(Y_i(d,s) \leq y | D_i = d\right) \Pr(\psi_i = s | D_i = d)}{\Pr(\psi_i = s | D_i = d)} \\
&= \frac{\Pr\left(Y_i(d,s) \leq y, \psi_i = s | D_i = d\right)}{\Pr(\psi_i = s | D_i = d)} && \text{by Assumption 2} \\
&= \Pr\left(Y_i(d,s) \leq y | \psi_i = s, D_i = d\right) \\
&= \Pr\left(Y_i \leq y | \psi_i = s, D_i = d\right). && \text{by (1)}
\end{aligned}$$

The distributions of potential outcomes $Y_i(d, \psi_{id'})$ are identifed as

$$\begin{aligned}
F^{d,d'}(y) &\equiv \Pr(Y_i(d, \psi_{id'}) \leq y) \\
&= \sum_{s \in \Psi} \Pr(Y_i(d,s) \leq y | \psi_{id'} = s) \Pr(\psi_{id'} = s) && \text{by L.I.E.} \\
&= \sum_{s \in \Psi} \Pr(Y_i(d,s) \leq y | \psi_i = s, D_i = d) \Pr(\psi_i = s | D_i = d') && \text{by Assumption 1} \\
& && (3) \\
&= \sum_{s \in \Psi} \Pr(Y_i \leq y | D_i = d, \psi_i = s) \Pr(\psi_i = s | D_i = d') && \text{by (1)} \\
&= \sum_{s \in \Psi} G^{d,s}(y) \Pr(\psi_i = s | D_i = d').
\end{aligned}$$

Therefore,

$$\begin{aligned}
F^d(y) = F^{d,d}(y) &\equiv \Pr(Y_i(d, \psi_{id}) \leq y) \\
&= \sum_{s \in \Psi} \Pr(Y_i \leq y | D_i = d, \psi_i = s) \Pr(\psi_i = s | D_i = d) \\
&= \Pr(Y_i \leq y | D_i = d). && \text{by L.I.E.}
\end{aligned}$$

(3) is because

$$\begin{aligned}
\Pr(Y_i(d,s) \leq y | \psi_{id'} = s) &= \frac{\Pr(Y_i(d,s) \leq y, \psi_{id'} = s | D_i = d')}{\Pr(\psi_{id'} = s | D_i = d')} && \text{by Assumption 1} \\
&= \frac{\Pr(Y_i(d,s) \leq y, \psi_i = s | D_i = d')}{\Pr(\psi_i = s | D_i = d')} \\
&= \frac{\Pr(Y_i(d,s) \leq y | D_i = d') \Pr(\psi_i = s | D_i = d')}{\Pr(\psi_i = s | D_i = d')} && \text{by Assumption 2} \\
&= \Pr(Y_i(d,s) \leq y | D_i = d') \\
&= \Pr(Y_i(d,s) \leq y | D_i = d) && \text{by Assumption 1} \\
&= \Pr(Y_i(d,s) \leq y | D_i = d) \frac{\Pr(\psi_i = s | D_i = d)}{\Pr(\psi_i = s | D_i = d)} \\
&= \frac{\Pr(Y_i(d,s) \leq y, \psi_i = s | D_i = d)}{\Pr(\psi_i = s | D_i = d)} \\
&= \Pr(Y_i(d,s) \leq y | \psi_i = s, D_i = d).
\end{aligned}$$

□

Note that the distributions of $Y(d, s)$ and $Y(d)$ are identified in the usual way from the independence assumptions. Identification of $Y(d, \psi_{id'})$ for $d \neq d'$ requires that the support of ψ_{i1} and ψ_{i0} are the same. If $\text{Supp}(\psi_{i1}) \subsetneq \text{Supp}(\psi_{i0})$, then only $F^{0,1}(y)$ is identified, but not $F^{1,0}(y)$. By Lemma 1, we have the following result.

Proposition 1 (Identification of Averages)**.** *Under Assumptions 1, 2,*

$$\begin{aligned}
\theta(0) &= \sum_{s\in\Psi} E[Y_i|D_i=1,\psi_i=s]\Pr(\psi_i=s|D_i=0) - E[Y_i|D_i=0],\\
\delta(1) &= E[Y_i|D_i=1] - \sum_{s\in\Psi} E[Y_i|D_i=1,\psi_i=s]\Pr(\psi_i=s|D_i=0),\\
\Delta &= \delta(1)+\theta(0) = E[Y_i|D_i=1]-E[Y_i|D_i=0].
\end{aligned}$$

$\theta(1)$ and $\delta(0)$ are identified similarly with $\Delta = \theta(1)+\delta(0)$. For each $d\in\{0,1\}$ and for $s,s'\in\Psi$, the exposure effects are identified as

$$\tau(d,s,s') = E[Y_i|D_i=d,\psi_i=s'] - E[Y_i|D_i=d,\psi_i=s].$$

Proof. By Lemma 1, expectations are identified as follows

$$\begin{aligned}
E[Y_i(d,s)] &= \int_{\mathbb{R}} ydG^{d,s}(y) = E[Y_i|\psi_i=s,D_i=d],\\
E[Y_i(d,\psi_{id'})] &= \int_{\mathbb{R}} ydF^{d,d'}(y) = \sum_{s\in\Psi} E\left[Y_i|D_i=d,\psi_i=s\right]\Pr(\psi_i=s|D_i=d'),\\
E[Y_i(d,\psi_{id})] &= \int_{\mathbb{R}} ydF^{d}(y) = E\left[Y_i|D_i=d\right].
\end{aligned}$$

□

3.1 Estimator

From the identification results in Proposition 1, we can construct estimators for the averages of potential outcomes and treatment parameters. For notational simplicity, let $\mathbb{1}_i(d,s) = \mathbb{1}\{D_i=d,\psi_i=s\}$. Define $N(d,s)\equiv\sum_{i=1}^N \mathbb{1}_i(d,s), N(d)\equiv\sum_{i=1}^N \mathbb{1}_i(d)$, and

$$\begin{aligned}
\nu(d,s) &= E[Y(d,s)] = E[Y_i|D_i=d,\psi_i=s] \quad d\in\{0,1\}, s\in\Psi,\\
\mu(d,d') &= E[Y(d,\psi_{di'})] \quad d,d'\in\{0,1\},\\
\mu(d) &= \mu(d,d).
\end{aligned}$$

These average outcomes can be estimated by the following frequency estimators.

$$
\begin{aligned}
\hat{\nu}(d,s) &= \frac{1}{N(d,s)}\sum_{i=1}^{N}\mathbb{1}_i(d,s)Y_i,\\
\hat{\mu}(d) &= \frac{1}{N(d)}\sum_{i=1}^{N}\mathbb{1}\{D_i=d\}Y_i,\\
\hat{\mu}(d,d') &= \frac{1}{N(d')}\sum_{j=1}^{N}\hat{\nu}(d,\psi_j)\mathbb{1}\{D_j=d'\}\\
&= \frac{1}{N(d')}\sum_{s\in\Psi}\sum_{j=1}^{N}\hat{\nu}(d,s)\mathbb{1}_j(d',s) = \sum_{s\in\Psi}\hat{\nu}(d,s)\frac{N(d',s)}{N(d')}.
\end{aligned}
$$

The $\nu(d,s)$ is the sample average of observed outcome on the subsample with $D_i = d, \psi_i = s$. Note that these estimators are undefined when the corresponding cells are empty. Similarly, $\hat{\mu}(d)$ is the sample average of outcome on the subsample with $D_i = d$. $\hat{\mu}(d,d')$ is the weighted average of average potential outcomes $Y(d,s)$, in which the weights are sample analog of $\Pr(\psi_i = s|D_i = d')$. Using these estimators, the overall, direct, and indirect treatment effects can be estimated by

$$
\begin{aligned}
\hat{\Delta} &= \hat{\mu}(1) - \hat{\mu}(0),\\
\hat{\theta}(d) &= \hat{\mu}(1,d) - \hat{\mu}(0,d),\\
\hat{\delta}(d) &= \hat{\mu}(d,1) - \hat{\mu}(d,0),\\
\hat{\tau}(d,s,s') &= \hat{\nu}(d,s) - \hat{\nu}(d,s').
\end{aligned}
$$

3.2 Asymptotic Properties

In this subsection, the consistency and asymptotic normality of estimators of average outcomes are derived. As aforementioned, Y_i would be dependent across individuals. However, because the treatment is randomly assigned, the potential outcome could be identically and independently distributed. Thus, assume following

Assumption 4. *For all $d \in \{0,1\}$ and $s \in \Psi$, $\{Y_i(d,s) : 1 \leq i \leq N\}$ are i.i.d.*

Assumption 5. *For all $d \in \{0,1\}$ and $s \in \Psi$, suppose $E[Y_i(d,s)^3] < \infty$.*

The following results use the fact that $Y_i = Y_i(d,s)$ conditional on $D_i = d, \psi_i = s$, so that the outcomes are conditionally independent.

Assumption 6. *Let $C_i = (D_i, \psi_i)$, and $G = (g_{ij}) \in \mathbb{R}^{N \times N}$ with*

$$g_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } C_i \perp C_j, \\ 0 & \text{if } i = j \text{ or } C_i \not\perp C_j. \end{cases}$$

Then, $\sum_{j=1}^N g_{ij} = O(N^\delta)$ for some $0 < \delta < 1$.

3.2.1 Consistency and asymptotic normality

Proposition 2 proves the average treatment effect estimators are consistent.

Proposition 2. *Under Assumptions 1-6, for each $d \in \{0,1\}$ and $s \in \Psi$,*

$$\begin{aligned} \sqrt{N}(\hat{\nu}(d,s) - \nu(d,s)) &\xrightarrow{p} N(0, V(d,s)), \\ \sqrt{N}(\hat{\mu}(d,d') - \mu(d,d')) &\xrightarrow{d} N\left(0, V_\mu(d,d')\right), \end{aligned}$$

where

$$\begin{aligned} V(d,s) &= \frac{\mathrm{Var}(Y_i(d,s))}{P(d,s)}, \\ V_\mu(d,d') &= \sum_{s \in \Psi} \frac{P(d',s)^2}{P(d')^2} V(d,s) + \sum_{s \neq s' \in \Psi} \frac{P(d',s)}{P(d')} \frac{P(d',s')}{P(d')} E[Y(d,s)]E[Y(d,s')]. \end{aligned}$$

3.2.2 Variance estimator

Let $\sigma(d,s)^2 = \mathrm{Var}(Y(d,s))$. Consider an estimator of $\sigma(d,s)^2$ as

$$\hat{\sigma}^2(d,s) = \frac{1}{N(d,s)} \sum_{i=1}^N \mathbb{1}_i(d,s)[Y_i - \hat{\nu}(d,s)]^2.$$

Then,

Assumption 7. *For all $d \in \{0,1\}$ and $s \in \Psi$, suppose $E[Y_i(d,s)^6] < \infty$.*

Proposition 3. *Under Assumptions 1-4, 6, and 7, for all $d \in \{0,1\}$ and $s \in \Psi$,*

$$|\hat{\sigma}^2(d,s) - \sigma^2(d,s)| \xrightarrow{p} 0.$$

The next Proposition is about the asymptotic disribution of estimators of treatment effects.

Proposition 4. [6] *Under Assumptions 1-7, for all $(c,c'),(d,d') \in \{0,1\}^2$,*

$$\sqrt{N}(\hat{\delta}(d) - \delta(d)) \to N(0, V_\delta),$$
$$\sqrt{N}(\hat{\theta}(d) - \theta(d)) \to N(0, V_\theta).$$

4 Simulation

This section illustrates the asymptotic properties derived in Section 3 by Monte Carlo simulations. In particular, the means squared errors, and the coverage rates of estimators provide simulation evidence of the asymptotic normality. The data for simulation consists of N units. Each unit has a binary treatment assignment D_i, which is drawn from the Bernoulli experiment with probability q. Each unit's exposure map is defined as

$$\psi_{id}(\boldsymbol{D}) = \left(\sum_j A_{ij}(d) D_j, \sum_j A_{ij}(d)(1 - D_j) \right) = (M_i(d), M - M_i(d)),$$

where $M_i(d)$ is the number of treated neighbors when $D_i = d$ is given. Here, $M_i(d)$ is drawn from a truncated normal distribution on $[0, M]$ with mean $Mp(d)$ and variance σ^2, where $p(d) = p_1^d p_0^{1-d}$. Potential outcomes are generated by the following DGP:

$$Y_i(d,s) = \begin{pmatrix} 1 & d & s' \end{pmatrix} \theta + \varepsilon, \quad \varepsilon \sim N(0,1).$$

This model is a sort of linear-in-sums model without endogenous peer effect. Parameters are set by $\theta = (1,2,3,4)', M = 10, \sigma = 5, p_1 = 0.26298, p_2 = 0.73701, q = 0.5$. The

[6] This proposition is not yet proved

choice of θ, p_1, p_2 makes the true ATE as 6, DTE as 2, and ITE as 4. $\sigma = 5$ is for overlapping assumption. Table 1, and Table 2 show mean squared errors of each estimator, which are defined as

$$MSE(\hat{\theta}, \theta) = \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta)^2.$$

The number of replication is $S = 10,000$.

Table 1: Mean Squared Errors of Average Potential Outcomes

Design	N	$\mu(0)$	$\mu(0,1)$	$\mu(1,0)$	$\mu(1,1)$
1	500	0.032	0.0325	0.0439	0.0447
	1,000	0.0164	0.0165	0.0176	0.017
	5,000	0.0033	0.0032	0.0033	0.0034
	10,000	0.0016	0.0016	0.0017	0.0017

Table 2: Mean Squared Errors of Treatment Effects

Design	N	Δ	$\theta(1)$	$\theta(0)$	$\delta(1)$	$\delta(0)$
1	500	0.0648	0.0189	0.0201	0.0686	0.0701
	1,000	0.0329	0.0051	0.0047	0.0297	0.0299
	5,000	0.0064	0.0009	0.0009	0.0058	0.0058
	10,000	0.0033	0.0005	0.0005	0.003	0.003

Notes. MSEs are computed by 10,000 simulations. $MSE = \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta)^2$, where θ is the true value of parameters from the design.

The derived rate of convergence in Proposition 2 is $O_p(1/\sqrt{N})$. This implies $N \times MSE(\hat{\theta}, \theta) = O_p(1)$ for all estimators $\hat{\theta}$. The simulation results coincide with the theory because $N \times MSE(\hat{\theta}, \theta)$ are stable.

Table 3, and Table 4 show the coverage rates of each estimator, which are defined as

$$C(\hat{\theta},\theta)=\frac{1}{S}\sum_{s=1}^{S}\mathbb{1}\{\theta\in CI(\hat{\theta}_s)\},$$

where $CI(\hat{\theta}_s)=[\hat{\theta}_s-1.96\mathrm{se}(\hat{\theta}),\hat{\theta}_s+1.96\mathrm{se}(\hat{\theta})]$.

Table 3: 95% Coverage Rates of Average Potential Outcomes

Design	N	$\mu(0)$	$\mu(0,1)$	$\mu(1,0)$	$\mu(1,1)$
1	500	0.9692	0.9813	0.9788	0.9721
	1,000	0.9759	0.986	0.9845	0.9744
	5,000	0.9773	0.9843	0.987	0.9765
	10,000	0.9774	0.9841	0.9847	0.9754

Table 4: 95% Coverage Rates of Treatment Effects

Design	N	Δ	$\theta(1)$	$\theta(0)$	$\delta(1)$	$\delta(0)$
1	500	0.9944	0.9959	0.9958	0.9971	0.9976
	1,000	0.9987	0.9999	0.9999	0.9997	0.9998
	5,000	0.9983	1	1	0.9998	0.9996
	10,000	0.9978	1	1	0.9993	0.9994

Notes. Coverage probabilities are computed by 10,000 replications. For the treatment parameters, the confidence intervals are computed by ignoring the asymptotic covariances of average potential outcomes. So now it is conservative. I will fix this after deriving the exact asymptotic distributions.

5 Empirical Application

This section describes a simple empirical analysis to show how the decomposition proposed in Section 3 can be applied to real data. To estimate the treatment effects and decompose them, we need data consisting of a random experiment, outcome, and exposure map. High schools are almost randomly assigned when students graduate from middle school in Korea. I use this random assignment to treat and estimate the impact of entering both-gender high school on their academic performance. The results in this section are preliminary and need to be completed later.

5.1 Data and Institutional Background

The data used in this application is from the Korean Education and Employment Panel II (KEEP II) from Korean Research Institute for Vocational Education and Training (KRIVET). The population was the junior students in high school (2nd-grade students) in 2016. The initial sample consists of 10,558 students in 416 schools.

When students graduate from middle school, they choose which type of school they want to apply to. Once they choose the type, the high schools are almost randomly assigned within the type, and region according to the student's address. I exploit the exogenous variations of this random assignment of high school in this application.

There are 5 types of high schools in Korea. First, a general high school is the most common school in Korea. Most students in general high school are likely to enter a university after graduation. Engineering high schools are for students to get a job after graduation. To enter Special, Science, and Foreign Language schools, students need to pass the entrance exam so they are not randomly assigned. In this application, I focus on the general high school only.

5.2 Outcomes

I set two outcomes for students' academic performances. First outcome is the relative grade within each school. For each high school, students are graded by 9 grades, where grade 1 is the top 4%, and grade 9 is the bottom 4%.[7] The second outcome is the indicator if the student enters the university in Seoul. Because most highly rated schools are located in Seoul, this outcome is expected to measure students' academic performances.[8] [9]

[7] The raw scores in the first wave data are available.

[8] Because almost all students enter a university after graduation from general high schools, entering university could not measure their performance well. Also, the exact ranking of each university can be measured in data so that I will use this information later.

[9] The outcomes are roughly defined for now, but they can be defined more rigorously later.

Table 5: Distribution of friendships over types of high schools

			Y_1		Y_2	
Type	Gender	N	Mean	SD	Mean	SD
Single	Male	736	4.11	1.55	11.35	31.74
Single	Female	1,053	3.92	1.52	18.01	38.45
Both	Male	1,092	4.31	1.65	11.08	31.4
Both	Female	1,262	3.81	1.47	16.18	36.84
Total		4208	4.02	1.55	14.44	35.16

notes.

The average relative grade of female students are 3.92 (female schools), 3.81 (combined schools), and those of male students are 4.11 (male schools), 4.31 (combined schools). The percentages of entering universities in Seoul of female students are 18.01% (female schools), 16.18% (combined schools), and those of male students are 11.35% (male schools), 11.08% (combined schools). Therefore, according to Table 5, female students perform better than male students in both outcomes.

5.3 Exposures

The distribution of this exposure is likely different between when a student enters a single-gender school and when the student enters a both-gender school. For example, a female student would have more female students when she has been assigned to a female school, while there would be more opportunities to make a male student in a both-gender school. Actually, the Table 6 show this phenomenon. Students who said they have only same-gender friends was 43% in male school, 48% in female school, and 27.9% in both-gender school. The difference in this distribution seems significant between single-gender schools and both-gender schools.

Table 6: Distribution of friendships over types of high schools

	No friends (%)	Only same gender friends (%)	Both (%)
Male school	3.16	43.57	53.28
Female school	4.51	48.33	47.17
Combined	4.42	27.98	67.59
Total	4.22	34.88	60.9

notes.

This implies that the distribution of number of same/opposite gender friends would be significantly different according to the own treatment status. According to Table 6, I defined the following exposure map.

$$\psi_{id} = \begin{cases} 1 & \text{if } i \text{ has no friends when } D_i = d, \\ 2 & \text{if } i \text{ has only friends with same gender when } D_i = d, \\ 3 & \text{if } i \text{ has friends with both genders when } D_i = d, \end{cases}, d \in \{0, 1\}.$$

Note that Table 6 guarantees all the estimators are well defined.

5.4 Estimation

After cleaning the data by removing no response and errors, the final data has 216 schools consisting of male schools (40), female schools (53), and combined schools (123), and 4208 students consisting of male (1850) and female (2358) students.

Table 7 shows the estimated treatment effects. Y_1 is the relative grades, and Y_2 is the indicator of entering universities located in Seoul. For Y_1, it seems that there are no indirect average treatment effects, but for Y_2, most direct and indirect effects are statistically significant.

Table 7: Estimation of Direct and Indirect Treatment effects

	Y_1			Y_2		
	Total	Male	Female	Total	Male	Female
ATE	0.08**	0.25**	-0.12**	-1.46**	-0.27**	-1.83**
$\delta(1)$	0.05**	0.24**	-0.16**	-1.27**	1-1.05**	-1.33**
$\delta(0)$	0.11**	0.27**	-0.1**	-1.44**	1-0.23**	-2.04**
$\theta(1)$	-0.03	-0.02	-0.02	-0.02	-0.04**	0.21
$\theta(0)$	0.03	0.01	0.05	-0.19**	0.78**	-0.51**

notes.

5.5 Discussion

In this section, an empirical application illustrates how treatment effects can be decomposed into direct and indirect effects using real data. In this application, the exposure map is the number of same-gender friends and the number of opposite-gender friends. The distribution of the exposure map could be partly determined by the underlying friendship networks and the gender of students. But then, Assumption 2 fails in this case because even conditional on the treatment assignment, the potential outcomes, and the exposures would be correlated if potential outcomes are also affected by gender. This is a limitation of the current settings in the model, and it needs to incorporate covariates into the model.

6 Conclusion

In this study, I proposed a way to decompose the treatment effect into direct and indirect effects using a potential outcome framework in the presence of social interactions, and treatments are randomly assigned. Neighbors' treatment status affects one's potential outcome through a correctly specified exposure map. Also, the underlying social network determining the neighborhood of each individual is assumed to be influenced by their own treatment status; hence the distribution of exposures would be different in different treatment statuses. Under the sequential ignorability assumption from the mediation model literature, the distributions of potential counterfactual outcomes are identified, and corresponding frequency estimators are proposed. The asymptotic normality of the proposed estimators is derived.

A contribution of this study in the literature is to identify and estimate the treatment effects in the presence of social interactions into direct and indirect effects separately. Also, an advantage of this model is that the model does not need the knowledge of network formation or the exact adjacency matrix representing the network structures. Identifying indirect effects needs the difference in the distribution of the value of exposure status for different own treatment statuses.

However, the model is not complete yet because the model does not consider covariates. As mentioned in Section 5, Assumption 2 could be violated if there is a common factor determining both potential outcome and exposure. Also, the calculation of estimators needs overlapping exposure values. Therefore, we need to define the exposure map carefully to apply this model. Another limitation of this setting is that this setting may not be plausible if the underlying network is undirected.

7 Future Direction and Extensions

- $M(d)$ can be different for each individual, i.e., $M_i(d)$.

- Incorporate covariates: Sequential conditional independence, linear, or kernel estimators.

- Supports of exposure map could be different for $d \in \{0, 1\}$. e.g., Ψ_0 and Ψ_1. Then, assumptions like $\Psi_0 \subseteq \Psi_1$ are required to identify $F^{10}(y)$. Both $F^{10}(y)$ and $F^{01}(y)$ are identified if $\Psi_0 = \Psi_1$.

- Endogenous treatment: Treatment is $D_{ig} = \mathbb{1}\{\chi(\cdot) > 0\}$, and apply mediation model as in Huber (2019).

- When exposure map is misspecified?

- When network is undirected?

- Testing distributional (direct/indirect) treatment effects and spillover effects: The distribution is identified, and $\sqrt{n}\sup_y |\hat{F}^{d,d'}(y) - F^{d,d'}(y)| \Rightarrow \mathcal{B} \circ F^{d,d'}$.. We can test hypothesis $H_0 : F^{d,d'}(y) \geq F^d(y) \quad \forall y$

References

Aronow, Peter M and Cyrus Samii (2017). “Estimating average causal effects under general interference, with application to a social network experiment”. In: *The Annals of Applied Statistics* 11.4, pp. 1912–1947.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). “Identification of peer effects through social networks”. In: *Journal of econometrics* 150.1, pp. 41–55.

Chen, Louis HY and Qi-Man Shao (2004). “Normal approximation under local dependence”. In: *The Annals of Probability* 32.3, pp. 1985–2028.

Comola, Margherita and Silvia Prina (2021). “Treatment effect accounting for network changes”. In: *Review of Economics and Statistics* 103.3, pp. 597–604.

Esseen, Carl-Gustav (1945). “Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law”. In: *Acta mathematica* 77, pp. 1–125.

Forastiere, Laura, Edoardo M Airoldi, and Fabrizia Mealli (2021). “Identification and estimation of treatment and interference effects in observational studies on networks”. In: *Journal of the American Statistical Association* 116.534, pp. 901–918.

Huber, Martin (2014). “Identifying causal mechanisms (primarily) based on inverse probability weighting”. In: *Journal of Applied Econometrics* 29.6, pp. 920–943.

— (2019). “A review of causal mediation analysis for assessing direct and indirect treatment effects”. In.

Imbens, Guido W and Donald B Rubin (2010). “Rubin causal model”. In: *Microeconometrics*. Springer, pp. 229–241.

Kline, Brendan and Elie Tamer (2020). “Econometric analysis of models with social interactions”. In: *The Econometric Analysis of Network Data*. Elsevier, pp. 149–181.

Leung, Michael P (2020). “Treatment and spillover effects under network interference”. In: *Review of Economics and Statistics* 102.2, pp. 368–380.

Leung, Michael P (2022). “Causal inference under approximate neighborhood interference”. In: *Econometrica* 90.1, pp. 267–293.

Manski, Charles F (2013). “Identification of treatment response with social interactions”. In: *The Econometrics Journal* 16.1, S1–S23.

Rubin, Donald B (1974). “Estimating causal effects of treatments in randomized and nonrandomized studies.” In: *Journal of educational Psychology* 66.5, p. 688.

Stein, Charles (1972). “A bound for the error in the normal approximation to the distribution of a sum of dependent random variables”. In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: Probability theory*. Vol. 6. University of California Press, pp. 583–603.

Vazquez-Bare, Gonzalo (2022). “Identification and estimation of spillover effects in randomized experiments”. In: *Journal of Econometrics*.

Appendix

A Technical Proofs in Ch3

Lemma 2 (Proposition 1 in Sourav Chatterjee (WP 2014))**.** *Let W be a $\mathbb{R}$-valued random variable with distribution function G, Φ be the distribution function of the standard normal distribution. Then,*

$$\sup_{x\in\mathbb{R}} |G(x) - \Phi(x)| \leq 2 \left(C_1 \sup_{f\in\mathcal{D}} |E(f'(W) - Wf(W))| \right)^{\frac{1}{2}},$$

$$\sup_{x\in\mathbb{R}} |G(x) - \Phi(x)| \leq 2 \left(C_2 d_W(G,\Phi) \right)^{\frac{1}{2}} \leq 2 \left(C_3 \sup_{f\in\mathcal{D}} |E(f'(W) - Wf(W))| \right)^{\frac{1}{2}},$$

where

$$\mathcal{D} \equiv \left\{ f \in \mathbb{R}^{\sharp} : |f(x)| \leq 1, |f'(x)| \leq 1, |f''(x)| \leq 1, \forall x \in \mathbb{R} \right\},$$

$$\mathcal{D}' \equiv \left\{ f \in \mathbb{R}^{\sharp} : |f(x) - f(y)| \leq |x-y|, \forall x, y \in \mathbb{R} \right\},$$

$$d_W(F,G) = \sup_{f\in\mathcal{D}'} \left| \int f dF - \int f dG \right|.$$

Moreover, if $\{X_i\}_{i=1}^n$ is a sequence of real random variables with distribution function G_i, and $d_W(G_n,\Phi) \to 0$, then $X_n \Rightarrow N(0,1)$.

Proof. Let $\varepsilon > 0$ be given. Let f be a solution of the following differential equation:

$$f'(x) - xf(x) = g(x) - E(g(Z)),$$

where $Z \sim N(0,1)$. Consider a pairwise linear function g on $\mathbb{R}$:

$$g(x) = \begin{cases} 1 & x \leq t \\ 1 - \frac{x-t}{\varepsilon} & x \in [t, t+\varepsilon] \\ 0x \geq t + \varepsilon \end{cases}$$

Then, by Stein (1972), there exists a bounded solution f that satisfies

$$|f(x)| \leq \frac{2}{\varepsilon},\ |f'(x)| \leq \frac{1}{\varepsilon}\sqrt{\frac{2}{\pi}} \leq \frac{1}{\varepsilon},\ |f''(x)| \leq \frac{2}{\varepsilon}.$$

Observe that $\frac{\varepsilon}{2}f \in \mathcal{D}$, $\mathbb{1}\{W \leq t\} \leq g(W)$, and $E[g(Z)] \leq E[\mathbb{1}\{Z \leq t\}] + \varepsilon A$, where $A = \max_{x\in\mathbb{R}} \phi(x)$. It follows that

$$\begin{aligned} G(t) &= E[\mathbb{1}\{W \leq t\}] \\ &\leq E[g(W)] \\ &= E[g(Z) + f'(W) - Wf(W)] \\ &\leq \Phi(t) + \varepsilon A + E[f'(W) - Wf(W)] \\ &= \Phi(t) + \varepsilon A + \frac{2}{\varepsilon}E[(\frac{\varepsilon}{2}f)'(W) - W(\frac{\varepsilon}{2}f)(W)] \\ &\leq \Phi(t) + \varepsilon A + \frac{2}{\varepsilon}\sup_{h\in\mathcal{D}} E[h'(W) - Wh(W)]. \end{aligned}$$

Conversely, by consideing

$$g(x) = \begin{cases} 1 & x \leq t - \varepsilon \\ 1 - \frac{x-t-\varepsilon}{\varepsilon} & x \in [t-\varepsilon, t] \\ 0x \geq t \end{cases}$$

, we have

$$G(t) \geq \Phi(t) - \varepsilon A - \frac{2}{\varepsilon}\sup_{h\in\mathcal{D}} E[h'(W) - Wh(W)].$$

Hence,

$$|G(t) - \Phi(t)| \leq \varepsilon A + \frac{B}{\varepsilon}.$$

Notice that the minimizer of the RHS is $\varepsilon = \sqrt{B/A}$, and $A = \frac{1}{\sqrt{2\pi}}$. Hence,

$$|G(t) - \Phi(t)| \leq 2\sqrt{AB} = 2\sqrt{\sqrt{\frac{2}{\pi}} \sup_{h \in \mathcal{D}} E[h'(W) - Wh(W)]}.$$

Next, in both cases, $|g'(x)| \leq \frac{1}{\varepsilon}$ for all differentiable points. Hence, g is $\frac{1}{\varepsilon}$-Lipschitz, and hence εg is 1-Lipschitz. Also, $Eg(Z) - \Phi(t) \leq \varepsilon A$ as noted above, where $A = \frac{1}{\sqrt{2\pi}}$. It follows that

$$\begin{aligned}|G(t) - \Phi(t)| &\leq |Eg(W) - Eg(Z) + Eg(Z) - \Phi(t)| \\ &\leq \frac{1}{\varepsilon} \sup_{g \in \mathcal{D}'} |Eg(W) - Eg(Z)| + |Eg(Z) - \Phi(t)| \\ &\leq \frac{1}{\varepsilon} d_W(G, \Phi) + \varepsilon A.\end{aligned}$$

The optimizer is $\varepsilon = \sqrt{\sqrt{2\pi} d_W(G, \Phi)}$. Thus,

$$\sup_{t \in \mathbb{R}} |G(t) - \Phi(t)| \leq 2\sqrt{\sqrt{2\pi} d_W(G, \Phi)}.$$

Next, consider the following differential equation:

$$f'(x) - xf(x) = g(x) - E[g(Z)],$$

where $Z \sim N(0, 1)$. Then, by Stein (1972), there exists a bounded solution f, that satisfies

$$\begin{aligned}|f|_\infty &\leq \sqrt{\frac{\pi}{2}} |g - E(g(Z))|_\infty, \\ |f'|_\infty &\leq 2|g - E(g(Z))|_\infty.\end{aligned}$$

If g is Lipschitz, (not necessarily bounded) then

$$\begin{aligned}|f|_\infty &\leq |g'|_\infty,\\ |f'|_\infty &\leq \sqrt{\frac{2}{\pi}}|g'|_\infty,\\ |f''|_\infty &\leq 2|g'|_\infty.\end{aligned}$$

.

Therefore, whenever g is Lipschitz, the solution satisfies

$$\begin{aligned}|f|_\infty &\leq \frac{1}{\varepsilon},\\ |f'|_\infty &\leq \sqrt{\frac{2}{\pi}}\frac{1}{\varepsilon} \leq \frac{1}{\varepsilon},\\ |f''|_\infty &\leq \frac{2}{\varepsilon},\end{aligned}$$

and hence, $\frac{\varepsilon}{2}f \in \mathcal{D}$.

Therefore, whenever g is 1-Lipschitz, we have

$$|E[g(W)] - E[g(Z)]| \leq \sup_{f\in\mathcal{D}} |E[f'(W)] - E[Wf(W)]|,$$

or,

$$d_W(G, \Phi) \leq \sup_{f\in\mathcal{D}} |E[f'(W)] - E[Wf(W)]|$$

□

Lemma 3. [10] *Let* $\{X_i\}_{i=1}^N$ *be a random variables with*

- $E(X_i) = 0$,
- $E(|X_i|^3) < \infty$.

Let $G = (g_{ij}) \in \{0,1\}^{N \times N}$ *be a dependency graph for* $\{X_i\}$*, that is if for all disjoint interval* $I_1, I_2 \subset \{1, ..., N\}$*, we have* $\{X_k : k \in I_1\} \perp \{X_\ell : \ell \in I_2\}$ *whenever* $G_{ij} = 0$ *for all* $i \in I_1$ *and* $j \in I_2$*. Define* $D_N = \max_{1 \le i \le N} \sum_{j=1}^N g_{ij} = \max_{1 \le i \le N} |N_i|$*, the maximum degree of the dependency graph, where* $N_i = \{j : g_{ij} = 1\}$.

Next, define

$$\sigma_N^2 = \mathrm{Var}\left(\sum_{i=1}^N X_i\right),$$
$$Z_N = \frac{1}{\sigma_N} \sum_{i=1}^N X_i.$$

Let F_N *be distribution function for* Z_N*, and* Φ *be the distribution function of the standard normal distribution. Then,*

$$d_W(F_N, \Phi) \le \frac{7D_N^2}{\sigma_N^3} \sum_{i=1}^N E|X_i|^3.$$

[10] This lemma is not yet proved

Proof. Let $[N] \equiv \{1, ..., N\}$. Fix $i \in [N]$, and $j \in N_i = \{j : g_{ij} = 1\}$.

Then, $[N] = I_1 \cup I_2 \cup I_3$, where $I_1, I_2, I_3 \subset [N]$ are disjoint,

$$\begin{aligned} I_1 &= N_i, \\ I_2 &= (N_i \cup N_j)^c, \\ I_3 &= (N_j \backslash N_i)^c. \end{aligned}$$

Define $Z_N^k = \frac{1}{\sigma_N} \sum_{\ell \in I_k} X_\ell$, and observe that

(a) $Z_N = Z_N^1 + Z_N^2 + Z_N^3$.

(b) $X_i \perp (Z_N^2 + Z_N^3)$.

(c) $(X_i, X_j) \perp Z_N^2$.

(d) $E[Z_N^2] = \frac{1}{\sigma_N^2} E\left[\sum_{i=1}^N X_i\right] = \frac{1}{\sigma_N^2} \operatorname{Var}\left(\sum_{i=1}^N X_i\right) = 1$ because $E[X_i] = 0$. And it follows that

$$\begin{aligned} 1 = E[Z_N^2] &= \frac{1}{\sigma_N} E\left[Z_N \sum_{i=1}^N X_i\right] \\ &= \frac{1}{\sigma_N} \sum_{i=1}^N E\left[Z_N X_i\right] \\ &= \frac{1}{\sigma_N} \sum_{i=1}^N E\left[Z_N^1 X_i\right] & \because (b) \\ &= \frac{1}{\sigma_N^2} \sum_{i=1}^N \sum_{j \in I_1} E\left[X_i X_j\right] \end{aligned}$$

Then, by Lemma (??), it suffices to show that for any real bounded functional f with bounded first and second moments,

$$|E[Z_N f(Z_N) - f'(Z_N)]| \leq \frac{7 D_N^2}{\sigma_N^3} \sum_{i=1}^N E[|X_i|^3].$$

By triangle inequality, the left-hand side can be upper bounded by

$$
\begin{aligned}
&|E[Z_N f(Z_N) - f'(Z_N)]| \\
&\leq \left| E[Z_N f(Z_N)] - \frac{1}{\sigma_N} \sum_{i=1}^{N} E[f'(W_N) X_i Z_N^1] \right| \\
&+ \left| \frac{1}{\sigma_N} \sum_{i=1}^{N} E[f'(W_N) X_i Z_N^1] - \frac{1}{\sigma_N^2} \sum_{i=1}^{N} \sum_{k \in N_i} E[f'(Z_N^2) X_i X_k] \right| \\
&+ \left| \frac{1}{\sigma_N^2} \sum_{i=1}^{N} \sum_{k \in N_i} E[f'(Z_N^2) X_i X_k] - E[f'(Z_N)] \right|
\end{aligned}
$$

Let f be a real functional with $|f''(x)| \leq 2$ for all $x \in \mathbb{R}$, and $W_N = Z_N^2 + Z_N^3$ (hence $Z_N = Z_N^1 + W_N$). Then, by mean value expansion,

$$
f(Z_N) = f(W_N) + f'(W_N) Z_N^1 + \frac{1}{2} f''(\overline{W_N})(Z_N^1)^2
$$

for some intermediate value $\overline{W_N}$ between Z_N and W_N. By using this expansion,

$$
\begin{aligned}
Z_N f(Z_N) &= \frac{1}{\sigma_N} \sum_{i=1}^{N} X_i f(Z_N) \\
&= \frac{1}{\sigma_N} \sum_{i=1}^{N} X_i \left(f(W_N) + f'(W_N) Z_N^1 + \frac{1}{2} f''(\overline{W_N})(Z_N^1)^2 \right) \\
&= \frac{1}{\sigma_N} \sum_{i=1}^{N} f(W_N) X_i + \frac{1}{\sigma_N} \sum_{i=1}^{N} f'(W_N) X_i Z_N^1 + \frac{1}{2\sigma_N} \sum_{i=1}^{N} f''(\overline{W_N}) X_i (Z_N^1)^2
\end{aligned}
$$

Hence,

$$\begin{aligned}
&\left|E[Z_N f(Z_N)] - \frac{1}{\sigma_N}\sum_{i=1}^{N} E[f'(W_N)X_i Z_N^1]\right| \\
&= \left|\frac{1}{\sigma_N}\sum_{i=1}^{N} E[f(W_N)X_i] + \frac{1}{2\sigma_N}\sum_{i=1}^{N} E[f''(\overline{W_N})X_i(Z_N^1)^2]\right| \\
&= \left|\frac{1}{2\sigma_N}\sum_{i=1}^{N} E[f''(\overline{W_N})X_i(Z_N^1)^2]\right| && \text{by (b), } E[X_i f(W_N)] = 0 \\
&\leq \frac{|f''|_\infty}{2\sigma_N}\sum_{i=1}^{N} E[|X_i|(Z_N^1)^2] \\
&= \frac{|f''|_\infty}{2\sigma_N}\sum_{i=1}^{N} E\left[|X_i|\left(\frac{1}{\sigma_N}\sum_{k\in N_i} X_k\right)^2\right] \\
&= \frac{|f''|_\infty}{2\sigma_N^3}\sum_{i=1}^{N}\sum_{k,\ell\in N_i} E\left[|X_i X_k X_\ell|\right] \\
&\leq \frac{|f''|_\infty}{2\sigma_N^3}\sum_{i=1}^{N}\sum_{k,\ell\in N_i} \max_{1\leq i\leq N} E[|X_i|^3] \\
&\leq \frac{|f''|_\infty N D_N^2}{2\sigma_N^3} \max_{1\leq i\leq N} E[|X_i|^3] && \text{(A)}
\end{aligned}$$

Next, consider

$$\begin{aligned}
f'(W_N)X_i Z_N^1 &= f'(Z_N^2 + Z_N^3)X_i\,(X_k) \\
&= \frac{1}{\sigma_N}\sum_{k\in N_i}\left(f'(Z_N^2) + f''(\overline{Z_N^2})Z_N^3\right)X_i X_k \\
&= \frac{1}{\sigma_N}\sum_{k\in N_i} f'(Z_N^2)X_i X_k + \frac{1}{\sigma_N}\sum_{k\in N_i} f''(\overline{Z_N^2})Z_N^3 X_i X_k,
\end{aligned}$$

by Mean value expansion for some $\overline{Z_N^2}$ between Z_N^2 and $Z_N^2 + Z_N^3$. It follows

$$\begin{aligned}
&\left|\frac{1}{\sigma_N}\sum_{i=1}^N E[f'(W_N)X_i Z_N^1] - \frac{1}{\sigma_N^2}\sum_{i=1}^N \sum_{k\in N_i} E[f'(Z_N^2)X_i X_k]\right| \\
&= \left|\frac{1}{\sigma_N}\sum_{i=1}^N \sum_{k\in N_i} E[f''(\overline{Z_N^2})Z_N^3 X_i X_k]\right| \\
&\le \frac{|f''|_\infty}{\sigma_N}\sum_{i=1}^N \sum_{k\in N_i} E[|Z_N^3 X_i X_k|] \\
&\le \frac{|f''|_\infty}{\sigma_N^3}\sum_{i=1}^N \sum_{k\in N_i} \sum_{\ell\in I_3} E[|X_i X_k X_\ell|] \\
&\le \frac{|f''|_\infty}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3]\sum_{i=1}^N \sum_{k\in N_i} |I_3| \\
&\le \frac{|f''|_\infty}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3]\sum_{i=1}^N |N_i| D_N \\
&\le \frac{|f''|_\infty N D_N^2}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3] \qquad \text{(B)}
\end{aligned}$$

Lastly, by observation (d),

$$\begin{aligned}
E[f'(Z_N)] &= E[f'(Z_N)]\frac{1}{\sigma_N^2}\sum_{i=1}^N \sum_{k\in N_i} E[X_i X_k] \\
&= \frac{1}{\sigma_N^2}\sum_{i=1}^N \sum_{k\in N_i} E[f'(Z_N)]E[X_i X_k]. \\
&= \frac{1}{\sigma_N^2}\sum_{i=1}^N \sum_{k\in N_i} \left[E[f'(Z_N^2)] + E[f''(\widetilde{Z_N^2})(Z_N^1 + Z_N^3)]\right] E[X_i X_k]. \qquad \text{(C)}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left|\frac{1}{\sigma_N^2}\sum_{i=1}^{N}\sum_{k\in N_i} E[f'(Z_N^2)X_iX_k] - E[f'(Z_N)]\right| \\
&= \left|\frac{1}{\sigma_N^2}\sum_{i=1}^{N}\sum_{k\in N_i} E[f'(Z_N^2)]E[X_iX_k] - \frac{1}{\sigma_N^2}\sum_{i=1}^{N}\sum_{k\in N_i}\left[E[f'(Z_N^2)] + E[f''(\widetilde{Z_N^2})(Z_N^1+Z_N^3)]\right]E[X_iX_k].\right| \qquad \text{by (} \\
&= \left|\frac{1}{\sigma_N^2}\sum_{i=1}^{N}\sum_{k\in N_i} E[f''(\widetilde{Z_N^2})(Z_N^1+Z_N^3)E[X_iX_k]\right| \\
&\le \frac{|f''|_\infty}{\sigma_N^2}\sum_{i=1}^{N}\sum_{k\in N_i} E[|Z_N^1+Z_N^3|]E[|X_iX_k|] \\
&= \frac{|f''|_\infty}{\sigma_N^3}\sum_{i=1}^{N}\sum_{k\in N_i}\sum_{\ell\in N_i\cup I_3} E[|X_\ell|]E[|X_iX_k|] \\
&\le \frac{|f''|_\infty}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3]\sum_{i=1}^{N}\sum_{k\in N_i}|N_i\cup N_j| \\
&\le \frac{|f''|_\infty}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3]\left[\sum_{i=1}^{N}|N_i|^2 + \sum_{i=1}^{N}\sum_{k\in N_i}|N_j|\right] \\
&\le \frac{2|f''|_\infty N D_N^2}{\sigma_N^3}\max_{1\le i\le N} E[|X_i|^3]. \qquad \text{(D)}
\end{aligned}$$

Combining (**??**),(**??**),and (D) and by using triange inequaliy, we have the desired result.

□

Proof of proposition 2. Define

$$
\begin{aligned}
P(c) &= \Pr(C_i = c) = \Pr(D_i = d, \psi_i = s), \\
\hat{P}(c) &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{C_i = c\}, \\
m(c) &= E[Y_i(c)] = E[Y_i(d, s)], \\
\hat{m}(c) &= \frac{1}{N} \sum_{i=1}^{n} \mathbb{1}\{C_i = c\} Y_i \\
&= \frac{1}{N} \sum_{i=1}^{n} \mathbb{1}\{C_i = c\} Y_i(c).
\end{aligned}
$$

Let

- $X_i = \frac{V_i}{\sqrt{N}}$
- $V_i = \mathbb{1}_i(d, s) Y_i(d, s) - P(d, s) E[Y(d, s)]$
- $\sigma_N^2(d, s) = \mathrm{Var}\left(\sum_{i=1}^N X_i\right)$
- $Z_N(d, s) = \frac{1}{\sigma_N^2(d,s)} \sum_{i=1}^N X_i$

Then, $E(X_i) = 0$. Assume $E|X_i|^3 < \infty$. Then,

$$
\begin{aligned}
\sigma_N^2(d, s) &= \frac{1}{N} \sum_{i=1}^{N} \mathrm{Var}(V_i) + \frac{1}{N} \sum_{i \neq j} \mathrm{Cov}(V_i, V_j) \\
&\leq \max \mathrm{Var}(V_i) + \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in N_i} \mathrm{Cov}(V_i, V_j) \\
&= \max \mathrm{Var}(V_i) + D_N \max \mathrm{Cov}(V_i, V_j)
\end{aligned}
$$

, , and $Z_N = \sum_{i=1}^{N} X_i$. Then, we have $D = O(N^{\delta})$ by assumption 6. Thus,

$$\begin{aligned}
E[X_i] &= \frac{1}{\sqrt{N}} E[V_i] = 0 \\
\sigma^2 &= \mathrm{Var}(Z_N) \\
&= \sum_{i=1}^{N} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \\
&= \mathrm{Var}(V_i) + \frac{1}{N} \sum_{i \neq j} \mathrm{Cov}(V_i, V_j) \\
&= \mathrm{Var}(V_i) + O(N^{\delta}) = O(N^{\delta}).
\end{aligned}$$

Also, by the lemma,

$$\begin{aligned}
\sup_f |Ef(Z_N/\sigma) - Ef(Z)| &\leq \frac{7D^2}{\sigma^3} \sum_{i=1}^{N} E|X_i|^3 \\
&\leq 7O(N^{2\delta})O(N^{-\frac{3}{2}\delta})O(N^{-\frac{3}{2}})N \\
&= 7O(N^{\frac{\delta}{2}-\frac{1}{2}}) \to 0
\end{aligned}$$

Let $\mathrm{Var}(V_i) = \sigma_m^2$. Then, $\left|\frac{\sigma^2}{N} - \frac{\sigma_m^2}{N}\right| \to 0$. Therefore,

$$\begin{aligned}
\left|\frac{Z_N}{\sigma} - \frac{Z_N}{\sigma_m}\right| &= \left|\frac{\frac{1}{N}\sum_{i=1}^{N} V_i}{\sigma/\sqrt{N}} - \frac{\frac{1}{N}\sum_{i=1}^{N} V_i}{\sigma_m/\sqrt{N}}\right| \\
&= \left|\frac{1}{N}\sum_{i=1}^{N} V_i\right| \left|\frac{1}{\sigma/\sqrt{N}} - \frac{1}{\sigma_m/\sqrt{N}}\right| \to 0
\end{aligned}$$

Hence, for any 1-Lipschitz function f, we have

$$\left|Ef\left(\frac{Z_N}{\sigma}\right) - Ef\left(\frac{Z_N}{\sigma_m}\right)\right| \to 0.$$

By triangle inequality,

$$\begin{aligned}
\frac{1}{\sigma_m}\sqrt{N}\left(\hat{m}(d,s) - m(d,s)\right) &= \frac{1}{\sqrt{N}\sigma_m} \sum_{i=1}^{N} \left(\mathbb{1}_i(d,s)Y_i(d,s) - P(d,s)E[Y(d,s)]\right) \\
&\xrightarrow{d} N(0,1),
\end{aligned}$$

where

$$\sigma_m^2 = \mathrm{Var}(V_i) = \mathrm{Var}(\mathbb{1}_i(d,s)Y_i(d,s)) = P(d,s)E[Y_i(d,s)^2] - P(d,s)^2E[Y_i(d,s)]^2.$$

Next, by the same argument for $V_i = \mathbb{1}_i(d,s) - P(d,s)$, we have

$$\begin{aligned}\frac{1}{\sigma_p}\sqrt{N}\left(\hat{P}(d,s) - P(d,s)\right) &= \frac{1}{\sqrt{N}\sigma_p}\sum_{i=1}^{N}(\mathbb{1}_i(d,s) - P(d,s)) \\ &\xrightarrow{d} N(0,1),\end{aligned}$$

where

$$\sigma_p^2 = \mathrm{Var}(V_i) = \mathrm{Var}(\mathbb{1}_i(d,s)) = P(d,s)(1 - P(d,s)).$$

Let $\boldsymbol{a} = (a_1, a_2) \in \mathbb{R}^2$. Then,

$$a_1\sqrt{N}\left(\hat{m}(d,s) - m(d,s)\right) + a_2\sqrt{N}\left(\hat{P}(d,s) - P(d,s)\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} a_1 V_{1i} + a_2 V_{2i},$$

where $V_{1i} = \mathbb{1}_i(d,s)Y_i(d,s) - P(d,s)E[Y(d,s)]$, $V_{2i} = \mathbb{1}_i(d,s) - P(d,s)$. Also note that $E[a_1V_{1i} + a_2V_{2i}] = 0$ and $E[|a_1V_{1i} + a_2V_{2i}|^3] < \infty$ and

$$\begin{aligned}
\sigma_{mp} &\equiv E[a_1 a_2 V_{1i} V_{2i}] \\
&= a_1 a_2 \operatorname{Cov}(\mathbb{1}_i(d,s)Y_i(d,s), \mathbb{1}_i(d,s)) \\
&= a_1 a_2 P(d,s)E[Y(d,s)] - P(d,s)^2 E[Y(d,s)] \\
&= a_1 a_2 P(d,s)(1 - P(d,s))E[Y(d,s)].
\end{aligned}$$

Therefore, by Cramer-Wold device, we have

$$\sqrt{N}\begin{pmatrix} \hat{m}(d,s) - m(d,s) \\ \hat{P}(d,s) - P(d,s) \end{pmatrix} \to N(0, \boldsymbol{V}),$$

where

$$\begin{aligned}
\boldsymbol{V} &= \begin{pmatrix} \sigma_m^2 & \sigma_{mp} \\ \sigma_{mp} & \sigma_P^2 \end{pmatrix} \\
&= \begin{pmatrix} P(d,s)E[Y_i(d,s)^2] - P(d,s)^2 E[Y_i(d,s)]^2 & P(d,s)(1 - P(d,s))E[Y(d,s)] \\ P(d,s)(1 - P(d,s))E[Y(d,s)] & P(d,s)(1 - P(d,s)) \end{pmatrix}.
\end{aligned}$$

By MVT,

$$\begin{aligned}
\sqrt{N}(\hat{\nu}(d,s) - \nu(d,s)) &= \sqrt{N}\left(\frac{\hat{m}(d,s)}{\hat{P}(d,s)} - \frac{m(d,s)}{P(d,s)}\right) \\
&= \frac{1}{\tilde{P}(d,s)}\sqrt{N}\left(\hat{m}(d,s) - m(d,s)\right) - \frac{\tilde{m}(d,s)}{\tilde{P}(d,s)^2}\sqrt{N}\left(\hat{P}(d,s) - P(d,s)\right) \\
&\longrightarrow N(0, \Sigma),
\end{aligned}$$

where

$$\Sigma = \frac{\mathrm{Var}(Y_i(d,s))}{P(d,s)}.$$

Next, consider $\hat{\mu}(d,d')$. Let $\Psi = (s_1, ..., s_K)$, and define

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \frac{N(d',s_1)}{N(d')}\frac{N}{N(d,s_1)} \\ \vdots \\ \frac{N(d',s_K)}{N(d')}\frac{N}{N(d,s_K)} \end{pmatrix}.$$

Then, $\hat{\boldsymbol{B}} \to \boldsymbol{B}$, where $B_k = \frac{\Pr(\psi_i = s_k|D_i=d')}{\Pr(\psi_i=s_k, D_i=d)}$. By the similar argument of using lemma and Cramer-Wold device, we have

$$\sqrt{N}\begin{pmatrix} \hat{m}(d,s_1) - m(d,s_1) \\ \vdots \\ \hat{m}(d,s_K) - m(d,s_K) \end{pmatrix} \xrightarrow{d} N(0, \boldsymbol{V}_m),$$

where $(\boldsymbol{V}_m)_{kk} = \sigma_m(d,s_k)^2 = P(d,s_k)E[Y(d,s_k)^2] - P(d,s_k)^2E[Y(d,s_k)]^2$ and $(\boldsymbol{V}_{k\ell}) = -P(d,s_k)P(d,s_\ell)E[Y(d,s_k)]E[Y(d,s_\ell)]$. Therefore,

$$\sqrt{N}(\hat{\mu}(d,d') - \mu(d,d')) = \hat{\boldsymbol{B}}\sqrt{N}\begin{pmatrix} \hat{m}(d,s_1) - m(d,s_1) \\ \vdots \\ \hat{m}(d,s_K) - m(d,s_K) \end{pmatrix}\hat{\boldsymbol{B}} \xrightarrow{d} N\left(0, V_\mu(d,d')\right),$$

where

$$\begin{aligned} V_\mu(d,d') &= \boldsymbol{B}\boldsymbol{V}_m\boldsymbol{B}' \\ &= \sum_{s\in\Psi} \Pr(\psi_i = s|D_i = d')^2 \frac{\sigma(d,s)^2}{\Pr(D_i = d, \psi_i = s)} \\ &+ \sum_{s\neq s'\in\Psi} \Pr(\psi_i = s|D_i = d')\Pr(\psi_i = s'|D_i = d')E[Y(d,s)]E[Y(d,s')] \end{aligned}$$

□

Proof of proposition 3. Note that

$$\hat{\sigma}^2(d,s) = \frac{1}{N(d,s)} \sum_{i=1}^{N} \mathbb{1}_i(d,s) Y_i^2 - \hat{\nu}(d,s)^2$$

Let $\hat{P}(d,s) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_i(d,s)$. Then, in the proof of proposition 2, we have $\hat{P}(d,s) \xrightarrow{p} P(d,s)$. Next, define

$$\hat{L}(d,s) = \frac{1}{N(d,s)} \sum_{i=1}^{N} \mathbb{1}_i(d,s) Y_i^2.$$

observe that

$$\hat{\sigma}^2(d,s) = \frac{\hat{L}(d,s)}{\hat{P}(d,s)} - \hat{\nu}(d,s)^2,$$

and by the same argument of $\hat{m}(d,s)$ in the proof of proposition 2, we have

$$|\hat{L}(d,s) - L(d,s)| = O_p\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, by Slutsky's theorem and continuous mapping theorem, we have

$$\hat{\sigma}^2(d,s) = \frac{\hat{L}(d,s)}{\hat{P}(d,s)} - \hat{\nu}(d,s)^2 \xrightarrow{p} \frac{L(d,s)}{P(d,s)} - \nu(d,s)^2 = \sigma^2(d,s).$$

□

Lemma 4. [11] *Under Assumptions 1-7, for all* $(c, c'), (d, d') \in \{0, 1\}^2$,

$$\sqrt{N}\begin{pmatrix}\hat{\mu}(d, d') - \mu(d, d') \\ \hat{\mu}(c, c') - \mu(c, c')\end{pmatrix} \to N(0, \boldsymbol{V}_\mu(d, d', c, c')),$$

where $\boldsymbol{V}_\mu(d, d', c, c') =$.

Proof of Proposition 4. By Lemma 4, and delta method, we have

$$\begin{aligned} V_\delta(d) &= \begin{pmatrix}1 - 1\end{pmatrix} \boldsymbol{V}_\mu(1, d, 0, d) \begin{pmatrix}1 \\ -1\end{pmatrix}, \\ V_\theta(d) &= \begin{pmatrix}1 - 1\end{pmatrix} \boldsymbol{V}_\mu(d, 1, d, 0) \begin{pmatrix}1 \\ -1\end{pmatrix} \end{aligned}$$

□

[11] This lemma is not yet proved

Suppose there exists an independent subsample of size $N' < N$, and $\lim N'/N > 0$.

Then, under the above assumptions,

$$\sqrt{N}(\hat{\boldsymbol{G}}(d,s)(\cdot) - \boldsymbol{G}(d,s)(\cdot)) \Rightarrow \zeta(\cdot),$$

where ζ is a $|2\Psi|$ dimensional mean zero Gaussian process.

It also follows

$$\sqrt{N}(\hat{\boldsymbol{q}}(d,s)(\cdot) - \boldsymbol{q}(d,s)(\cdot)) \Rightarrow \gamma(\cdot),$$

provided that $\boldsymbol{q}(d,s)(\cdot)$ are Hadamard differentiable.

Sketch of the proof. Recall that

$$\hat{G}(d,s)(t) = \frac{1}{N(d,s)} \sum_{i=1}^{N} \mathbb{1}_i(d,s)\mathbb{1}(Y_i \leq t),$$

Let $[N']$ be a set of indices of the subsample.

Let $\tilde{\boldsymbol{G}}(d,s)(\cdot)$ be an empirical cdf computed by independent subsample:

$$\tilde{G}(d,s)(t) = \frac{1}{N'(d,s)} \sum_{i\in[N']} \mathbb{1}_i(d,s)\mathbb{1}(Y_i \leq t),$$

where

$$N'(d,s) = \sum_{i\in[N']} \mathbb{1}_i(d,s).$$

Then, first show

$$\sup_{t\in\mathbb{R}} |\sqrt{N}(\hat{G}(d,s)(t) - \tilde{G}(d,s)(t))| = o_p(1)$$

Next,

$$\sup_{t\in\mathbb{R}} \left| \sqrt{N}(\tilde{G}(d,s)(t) - G(d,s)(t)) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i(t) \right| = o_p(1),$$

where

$$\gamma_i(t) = \frac{\mathbb{1}(Y_i \leq t)}{P(d,s)},$$

Lastly show $\{\gamma_i(t) : t \in \mathbb{R}\}$ is Donsker class. Then, the empirical process $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \gamma_i(\cdot)$ converges to a mean zero Gaussian process, and so does $\sqrt{N}(\hat{G}(d,s)(\cdot) - G(d,s)(\cdot))$.

□