Causal Mediation Analysis with Spatial Interference

Chiara Di Maria and Giada Adelfio

Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze, building 13, Palermo 90128, Italy
chiara.dimaria@unipa.it, giada.adelfio@unipa.it

Abstract. In the causal inference framework, estimating causal effects requires specific assumptions, among which that of independence of each unit's outcome from the treatment assigned to other units. Although it can be reasonable in some settings, this is not the case when dealing with spatial data. In this paper, we address causal mediation analysis in the presence of spatial interference: we discuss assumptions for the estimation of direct and indirect effects and provide an applied example.

Keywords: spillover effects, geostatistical data, counterfactual framework

1 Introduction

Causal inference is employed in many fields to assess causal connections between variables of interest, like exposure to pollution and cancer occurrence, or parental attachment and depression in adulthood. Recently, some scholars tried to integrate causal inference and spatial statistics, since in many real-world settings, it is of interest to assess the effect of an exposure on a certain response at specific locations. Unfortunately, in spatial settings, one of the key assumptions of the causal paradigm, i.e. the absence of interference between observations, does not hold. This led to the development of a new framework for spatial causal inference, to take into account the mutual influence of units at different locations. Some studies focused on areal data to evaluate the effect of government policies: Huber and Steinmayr [2] proposed a new difference-in-difference approach to separate single and spillover effects of unemployment benefits on the unemployment rate in some regions of Austria, and Verbitski-Savitz and Radeunbush [9] analysed the spillover effects of a community policy in Chicago to reduce neighbourhood crime rates. Other scholars worked on geostatistical data, like Giffin et al. [1], who showed how to use generalised propensity scores to address spatial confounding and estimate causal spillover effects; they used a Bayesian approach, as Papadogeorgou and Samanta [4], who proposed estimands of local and spillover effects and identifiability assumptions. To the best of our knowledge, none of the studies that had focused on this topic have addressed the issue of mediation, related to how the effect of an exposure affects the response of interest directly and also indirectly through a third variable, which is very frequent in applied contexts. In this paper, we formalise causal mediated effects in spatial settings and propose an estimation method, showing its application to real data.

2 Traditional and Spatial Mediation Effects

Let X denote an exposure, Y the outcome of interest and M an intermediate variable conveying all or part of the effect of X on Y. In traditional mediation settings, one of the most widely spread definitions of causal effects is given by natural direct and indirect effects (NDE and NIE, respectively), proposed by Pearl [5]. Denoting by $M(x)$ and $Y(x)$ the counterfactual values of the mediator and the outcome, respectively, if X were, possibly contrary to fact, set to x, and by $Y(x, M(x^*))$ the outcome value if X were set to x and the mediator to the natural value it would assume under $X = x^*$, with $x \neq x^*$, the natural effects are defined as

$$NDE = \mathbb{E}[Y(x, M(x^*)) - Y(x^*, M(x^*))] \tag{1}$$

$$NIE = \mathbb{E}[Y(x, M(x)) - Y(x, M(x^*))]. \tag{2}$$

In other words, the direct effect is the expected difference in the outcome if the mediator were fixed to its natural value under the control condition and the exposure changed from x^* to x; similarly, the indirect effect is the expected difference in the outcome if the value of the exposure were fixed and the mediator changed from its natural value in the scenario where $X = x^*$ to the scenario $X = x$.

In the absence of unobserved (post-treatment) confounders between the exposure, the mediator and the outcome, and if the so-called *Stable Unit Treatment Value Assumption* (SUTVA, Rubin [6]) holds, i.e. if the counterfactuals of each unit are independent of the treatment assigned to other units, natural effects are identifiable, that is, they can be expressed in terms of observed data [7]. This assumption is unlikely to hold in many settings, for example, in the context of vaccinations, where the probability of being infected depends not only on one's own vaccination status, but also on the status of other individuals. When data are collected in space, it is reasonable to assume that closer units are likely to influence each other and that the strength of mutual influence between units decreases with distance. In this paper, we focus on geostatistical data, however, the notation and definitions introduced can be easily extended to areal data.

Let us assume that data are collected at n spatial locations $s \in \{s_1, \ldots, s_n\} \in \mathcal{D} \subset \mathbb{R}^2$, and let a variable with the s subscript denote the value of that variable at location s, while the subscript $-s$ denotes all locations in $\mathcal{D}$ but s. To define mediational effects in a spatial setting, we need to extend the counterfactual notation: let $Y_s(x_s, \mathbf{x}_{-s}, m_s, \mathbf{m}_{-s})$ be the value of the outcome at location s if X and M at s were x_s and m_s, respectively, and in all other locations they were the vectors $\mathbf{x}_{-s}$ and $\mathbf{m}_{-s}$. The former two express the exposure and mediator effect on the outcome at location s, while the latter two indicate the spillover effect of all other units on the outcome at s. Although, in principle, it is possible to deal with this kind of counterfactuals, it becomes more and more difficult as the number of locations increases. For this reason, in line with recent literature about spatial causal inference [1, 4], we propose to summarise the spillover effects as weighted means of the exposure and the mediator, where the weights are the relative distances (or functions of them) between locations. Let $\mathbf{D}$ denote the distance matrix, with $D_{ss'}$ the distance between locations s and s', and

$$\tilde{X}_{-s} = \sum_{s' \in \mathcal{D} \setminus \{s\}} X_{s'} D_{ss'} / \sum_{s' \in \mathcal{D} \setminus \{s\}} D_{ss'}, \quad \tilde{M}_{-s} = \sum_{s' \in \mathcal{D} \setminus \{s\}} M_{s'} D_{ss'} / \sum_{s' \in \mathcal{D} \setminus \{s\}} D_{ss'}. \tag{3}$$

For the sake of simplicity, we consider a binary exposure; local and spillover effects, denoted by the subscripts s and $-s$, respectively, can be defined as follows:

$$\begin{aligned}
NDE_s &= \mathbb{E}[Y(x_s, \tilde{x}^*_{-s}, M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s})) - Y(x^*_s, \tilde{x}^*_{-s}, M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s}))] \\
NDE_{-s} &= \mathbb{E}[Y(x_s, \tilde{x}_{-s}, M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s})) - Y(x_s, \tilde{x}^*_{-s}, M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s}))] \\
NIE_s &= \mathbb{E}[Y(x_s, \tilde{x}_{-s}, M_s(x_s), \tilde{M}_{-s}(x^*_{-s})) - Y(x_s, \tilde{x}_{-s}, M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s}))] \\
NIE_{-s} &= \mathbb{E}[Y(x_s, \tilde{x}_{-s}, M_s(x_s), \tilde{M}_{-s}(x_{-s})) - Y(x_s, \tilde{x}_{-s}, M_s(x_s), \tilde{M}_{-s}(x^*_{-s}))]
\end{aligned}$$

The local direct effect NDE_s compares the expected outcome if location s were assigned the treatment to the expected outcome if s were not treated, in the scenario where all the other units are not treated and the mediator is as if all the other units are not treated. The local indirect effect NIE_s is the expected difference in the outcome moving from a scenario where the mediator is as if no units were treated to the scenario where the mediator assumes the value it would assume if only s were treated. The spillover effects compare instead the expected counterfactuals changing the values of X_{-s} and M_{-s} from the values they would assume if all units except s were not treated to the values they would take if all other units were treated. The sum of the local and the spillover direct effect equals the direct effect, and the sum of the local and the spillover indirect effect is equal to the indirect effect; in turn, the sum of the direct and the indirect effect yields the total effect, i.e. the expected difference in the outcome if all units were treated compared to the case where all units were not treated.

As shown by VanderWeele et al. [8] in the context of cluster interference, these effects can be identified if the following assumptions hold

$$Y(x_s, \tilde{x}_{-s}, m_s, \tilde{m}_{-s}) \perp\!\!\!\perp X \tag{4}$$

$$Y(x_s, \tilde{x}_{-s}, m_s, \tilde{m}_{-s}) \perp\!\!\!\perp M_s, \tilde{M}_{-s} | X \tag{5}$$

$$M_s(x_s), \tilde{M}_{-s}(x_{-s}) \perp\!\!\!\perp X \tag{6}$$

$$Y(x_s, \tilde{x}_{-s}, m_s, \tilde{m}_{-s}) \perp\!\!\!\perp M_s(x^*_s), \tilde{M}_{-s}(x^*_{-s}) \tag{7}$$

$$M_s(x_s) \perp\!\!\!\perp \tilde{M}_{-s}(x^*_{-s}) \tag{8}$$

Assumptions (4)-(6) concern the absence of exposure-mediator, exposure-outcome and mediator-outcome confounders, while assumption (7) requires the absence of post-treatment mediator-outcome confounders. All of them are quite standard in mediation literature. Assumption (8) is quite strong: it basically requires that the effect of the exposure at s does not affect the mediator values of the other locations. In other words, the spillover effects are only on the outcome and do not involve the mediator.

3 Estimation

To estimate the proposed effects, we adopt the approach introduced by Imai et al. [3] and implemented in the `mediation` R package. The idea is to fit parametric or semi/nonparametric models for the mediator and the outcome and simulate counterfactual values, combining them to derive causal effects. In particular, the key steps of this approach are:

1. Create J bootstrap samples from the original one and for each of them repeat the following steps:
   1.1 Fit models for the mediator and the outcome
   1.2 Obtain K predictions for the mediator under each treatment value, i.e. generate K counterfactuals
   1.3 Obtain K predictions for the outcome, substituting each value of the treatment and the counterfactual mediator obtained at the previous step
   1.4 Compute the causal effects by averaging the simulated counterfactuals over K and n
2. Obtain summary statistics of the causal effects from the bootstrap distribution (mean, median, quantiles...)

4 Application

We analysed a well-known data set in the context of geostatistics, related to the prevalence of malaria among children in 65 villages of The Gambia. The data contain information about the coordinates of each village where children's blood samples were collected, the percentage of vegetation in the village area (derived from satellite measures), the presence or absence of a health centre in the village, the age of each child, if he/she was found positive to malaria, if they regularly sleep under a bed-net and if it is treated against mosquitoes. The data set includes 2035 observations, living in 65 villages: this means that analysing data at the village level markedly reduces the sample size. To fully exploit the total number of observations, we applied spatial jittering with a displacement of 5 km to the coordinates, in order to add random noise. The original and the newly obtained locations are shown in Figure 1.

Our aim is to assess the local and the spillover effects of the presence of a health center on the probability that a child has malaria, mediated by the use of a bednet. First, we computed X_{-s} and M_{-s} as in Eq. (3), using a Gaussian kernel for the distances, so that $D_{ss'} = \exp(-||s - s'||^2/\tau)$, where $||s - s'||$ is the Euclidean distance between s and s'. To estimate τ, we considered a grid of eight

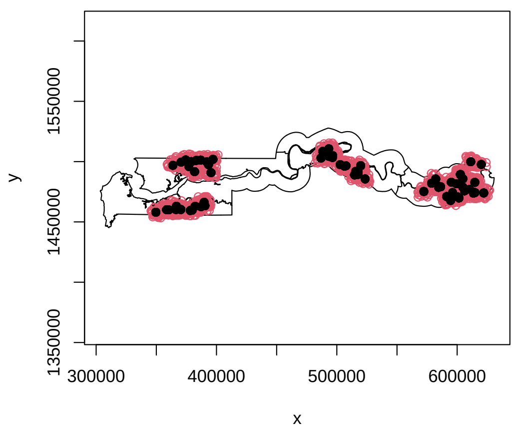

Fig. 1. Original locations (in black) and new locations obtained by applying spatial jittering with a 5 km displacement (in red).

values between 100 thousands and 800 millions and fitted the following generalised additive model for the expectation of the outcome Y, i.e. the probability that a child at location s has malaria,

$$logit(\mathbb{E}[Y_s|X,M,Z]) = \beta_0+\beta_1 X_s+\beta_2 M_s+\beta_3\tilde{X}_{-s}+\beta_4\tilde{M}_{-s}+\boldsymbol{\beta}_5 Z+spl(lat,long)$$

where X is the presence of a health center at location s, M_s is a binary variable denoting if child at s regularly uses a bed-net and Z includes additional covariates like children's age and the vegetation coverage in the area; finally, the term $spl(lat,long)$ is a smoothing term of latitude and longitude of each location to take into account the spatial structure of the data. For each value of τ, we fitted this model and evaluated its log-likelihood and the Root Mean Squared error comparing the observed and the predicted values of the outcome, selecting the value of τ presenting the best performance. This led us to $\tau = 1e8$. Once chosen it, we estimated the mediation effects using also the mediator model

$$logit(\mathbb{E}[M_s|X,Z]) = \gamma_0 + \gamma_1 X_s + \gamma_2 Z + spl(lat,long) \tag{9}$$

and performing the steps described in Section 3 using $J = 200$ bootstrap samples and drawing $K = 100$ copies of each counterfactual. Results are shown in Table 1. We can notice that all estimates are negative, but only the total and the direct effect are significant, the latter due to the significance of the local term. This means that the presence of a health center at location s reduces the probability that a child develops malaria, but it does not have a spillover effect on nearby locations, nor an indirect effect through the use of a bed net.

Table 1. Estimates, standard errors and 95% confidence intervals of the causal effects for the Gambia data set. Bold figures denote significant estimates.

Effect	Estimate and s.e.	Confidence interval
Total	**-0.381** (0.125)	(-0.603, -0.124)
Direct	**-0.321** (0.169)	(-0.586, -0.024)
Indirect	-0.060 (0.137)	(-0.487, 0.004)
Direct local	**-0.062** (0.029)	(-0.116, -0.009)
Indirect local	-0.001 (0.002)	(-0.006, 0.003)
Direct spillover	-0.259 (0.178)	(-0.558, 0.001)
Indirect spillover	-0.059 (0.137)	(-0.488,0.001)

5 Conclusions

In this work, we proposed the introduction of spatial mediation effects for geostatistical data in the presence of interference. This represents a novelty, since so far there has been no research in the context of causal spatial mediation. We proposed estimands and identification assumptions, which can be difficult to hold and are untestable. In the future, different kind of causal effects not relying on nested counterfactuals may be proposed.

References

1. Giffin, A., Reich, B.J., Yang, S., Rappold, A.G.: Generalized Propensity Score Approach to Causal Inference with Spatial Interference. Biom **79**(3), 2220–2231 (2023)
2. Huber, M., Steinmayr, A.: A framework for separating individual-level treatment effects from spillover effects. J Bus Econ Stat **39**(2), 422–436 (2021)
3. Imai, K., Keele, L., Tingley, D.: A general approach to causal mediation analysis. Psychol Methods **15**(4), 309–334 (2010)
4. Papadogeorgou, G., Samanta, S.: Spatial causal inference in the presence of unmeasured confounding and interference. arXiv preprint arXiv:2303.08218 (2023)
5. Pearl, J.: Direct and Indirect Effects. In: J. Breese, D. Koller (eds.) Uncertainty in Artificial Intelligence, Proceedings of the Seventeenth Conference, pp. 411–420. Morgan Kaufmann, San Francisco (2001)
6. Rubin, D.B.: Bayesian inference for causal effects: The role of randomization. Ann Stat **6**(1), 34–58 (1978)
7. VanderWeele, T.J.: Explanation in causal inference: methods for mediation and interaction. Oxford University Press (2015)
8. VanderWeele, T.J., Hong, G., Jones, S.M., Brown, J.L.: Mediation and spillover effects in group-randomized trials: a case study of the 4Rs educational intervention. J Am Stat Assoc **108**(502), 469–482 (2013)
9. Verbitsky-Savitz, N., Raudenbush, S.W.: Causal inference under interference in spatial settings: a case study evaluating community policing program in Chicago. Epidemiol Methods **1**(1), 107–130 (2012)