Bayesian inference of causal effects with incorrectly measured interference network

Bar Weinstein Daniel Nevo

Department of Statistics and Operations Research, Tel Aviv University

Introduction

- Interference occurs when the potential outcomes of a unit depend on treatments assigned to other units.
- The interference structure can be represented by a network. Nodes are units and edges indicate whether interference is possible between pairs of units.
- An observed network is often assumed to correctly specify the interference structure, and analysis is conditioned on it [4].
- However, accurately measuring the interference network is challenging.
- A misspecified network leads to biased estimation [5].
- Observing an incorrect network may result from:
  - Constraints in data acquisition methods (e.g., measurement errors, edge censoring).
  - Sampling a sub-network from the population network.

Setup and assumptions

- $\boldsymbol{Z} \in \{0,1\}^n$ binary treatments; $Y_i(\boldsymbol{z})$ potential outcomes; $\boldsymbol{Y}$ observed outcomes; $\boldsymbol{X}$ covariates; $\boldsymbol{A}$ observed network.
- Interference via true network $\boldsymbol{A}^*$. Assumed to be undirected and unweighted.
- A.1 (Consistency) *If* $\boldsymbol{Z} = \boldsymbol{z}$, *then* $Y_i = Y_i(\boldsymbol{z})$.
- Exposure mapping $f(\boldsymbol{Z}_{-i}, \boldsymbol{A}_i^*)$ with image space $\mathcal{C} \subseteq \mathbb{R}$.
- A.2 (Neighborhood interference) *For any* $\boldsymbol{z}, \boldsymbol{z}'$, *if* $z_i = z_i'$ *and* $f(\boldsymbol{z}_{-i}, \boldsymbol{A}_i^*) = f(\boldsymbol{z}'_{-i}, \boldsymbol{A}_i^*)$, *then* $Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}')$ *w.p.1.*
- A.3 (Positivity) $0 < \Pr(\boldsymbol{Z}_i = z, f(\boldsymbol{Z}_{-i}, \boldsymbol{A}_i^*) = c | \boldsymbol{X} = \boldsymbol{x}) < 1$, $\forall z \in \{0,1\}, c \in \mathcal{C}, \forall \boldsymbol{x}$.

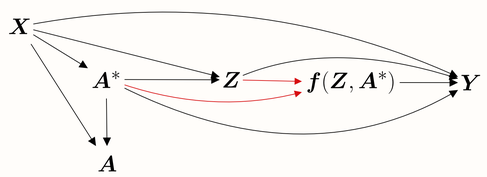

Figure 1. DAG representing the assumed causal structure. Red arrows are deterministic.

Estimands and identification

$\widetilde{Y}_i(z,c), z \in \{0,1\}, c \in \mathcal{C}$ potential outcomes expressed in term of exposure values. From A.2, $\widetilde{Y}_i(z,c)$ is equivalent to all $Y_i(\boldsymbol{z})$ with the same effective treatments.

Causal estimands

Comparisons of the following.

1. *Exposure levels.* $\mathbb{E}\left[\widetilde{Y}_i(z,c)\right]$.
2. *Treatment intervention.* $\mathbb{E}\left[Y_i(\boldsymbol{z})\right]$.

The first estimand reveals how changes in treatment and exposure values influence outcomes, while the second depicts the effect of setting $\boldsymbol{Z}$ to $\boldsymbol{z}$. Alternatively, conditional estimands, e.g., $\mathbb{E}\left[Y_i(\boldsymbol{z}) | \boldsymbol{A}^*\right]$, can be used instead.

Identification

1. $\mathbb{E}\left[\widetilde{Y}_i(z,c)\right] = \mathbb{E}_{\boldsymbol{X}}\mathbb{E}_{\boldsymbol{A}^*|\boldsymbol{X}}\mathbb{E}\left[Y_i | \boldsymbol{Z}_i = z, f(\boldsymbol{Z}_{-i}, \boldsymbol{A}_i^*) = c, \boldsymbol{A}^*, \boldsymbol{X}\right]$.
2. $\mathbb{E}\left[Y_i(\boldsymbol{z})\right] = \mathbb{E}_{\boldsymbol{X}}\mathbb{E}_{\boldsymbol{A}^*|\boldsymbol{X}}\mathbb{E}\left[Y_i | \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{A}^*, \boldsymbol{X}\right]$

- Both requires obtaining $\boldsymbol{A}^*|\boldsymbol{X}$ distribution.
- Can be extended to stochastic and network interventions [4].

Generative model and posterior distribution

- True network generation $p(\boldsymbol{A}^*|\boldsymbol{X}, \theta)$.
- Observed network (network misspecification model) $p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X}, \gamma)$.
- Outcome model $p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*, \eta)$.
- Possible to augment outcome model with propensity scores.
- Parameters space can be finite or infinite.
- Assume prior independence $\pi(\eta, \theta, \gamma) = \pi(\eta)\pi(\theta)\pi(\gamma)$.
- Denote observed data by $\boldsymbol{O} = (\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A})$.

Posterior distribution

$$\begin{aligned}\pi(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) \propto\ & p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*, \eta)\pi(\eta) \\ & \cdot p(\boldsymbol{A}|\boldsymbol{X}, \boldsymbol{A}^*, \gamma)\pi(\gamma) \\ & \cdot p(\boldsymbol{A}^*|\boldsymbol{X}, \theta)\pi(\theta)\end{aligned}$$

Examples of network misspecification models

Measurement error

$p(\boldsymbol{A}|\boldsymbol{A}^*, \boldsymbol{X}, \gamma)$ can be differential or non-differential measurement error model. For instance,

- *Random noise.* Edges in $\boldsymbol{A}$ are observed with true positive rate $1 - \gamma_1$ and false positive rate γ_0.
- *Censoring.* Edges between units with degrees larger than a censoring threshold are missing w.p. γ.

Sampled network

- Study on a sample from a population $n < N$.
- $\boldsymbol{A}$ is obtained via a network sampling procedure, such as random node, egocentric, or link-tracing sampling.
- Write $\boldsymbol{A}_o, \boldsymbol{A}_m$, as the observed and missing parts, respectively.
- Posterior can be written as

$$\begin{aligned}\pi(\eta, \theta, \boldsymbol{A}_m|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}_o) \propto\ & \pi(\eta)p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}_o, \boldsymbol{A}_m, \eta) \\ & \cdot \pi(\theta)p(\boldsymbol{A}_o, \boldsymbol{A}_m|\boldsymbol{X}, \theta)\end{aligned}$$

- In this scenario, further restrictions are required:
  - Missingness mechanism of the network sampling design (*ignorability*).
  - Interference between recruited and non-recruited units. RCTs are possible with further restrictions on exposure mappings. Observational studies are tricky.
  - Projective network models [2].

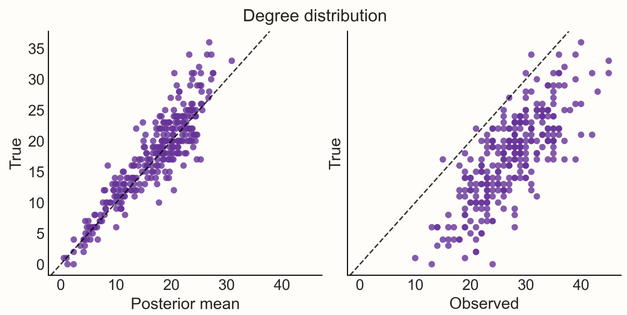

Figure 2. Reconstructing $\boldsymbol{A}^*$ from $\boldsymbol{A}$. Degrees $d_i = \sum_{j \neq i} A_{ij}$ in the observed network (right) and the posterior mean degrees (left) versus the true degrees.

Sampling from the posterior

- The full posterior is a mixed space of continuous (η, θ, γ) and discrete $(\boldsymbol{A}^*)$ latent variables. The discrete space has $\mathcal{O}\left(2^{n^2}\right)$ terms.
- MCMC methods such as MH or modified HMC/NUTS [6, 7] do not scale well. Marginalizing over $\boldsymbol{A}^*$ is problematic since Y_i depends on $\boldsymbol{A}_i^*$.

Bayesian modularization

The posterior can be written as a composition of modules

$$\pi(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) \propto \underbrace{\pi(\eta|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*)}_{\text{Outcome module}} \underbrace{\pi(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A})}_{\text{Network module}} \underbrace{p(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*)}_{\text{Feedback term}}$$

Consequently, sampling from the 'cut' posterior [1, 3]

$$\pi_{cut}(\eta, \theta, \gamma, \boldsymbol{A}^*|\boldsymbol{O}) \propto \pi(\eta|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{A}^*)\pi(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}),$$

is attractive since

$$\pi(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}) = \pi(\boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}, \theta, \gamma) \sum_{\boldsymbol{A}^*} \pi(\theta, \gamma, \boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}),$$

can be simplified tremendously. Sampling from the 'cut' posterior by [1, 3]:

1. Generate (θ_m, γ_m) samples. Then, sample multiple $\boldsymbol{A}_m^*$. For each $\boldsymbol{A}_m^*$, sample η from the outcome module. Networks can be sampled via either
   - i. **"Three-stage"**. For each (θ_m, γ_m) sample one network from $\boldsymbol{A}_m^* \sim \pi(\boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}, \theta_m, \gamma_m)$.
   - ii. **"Two-stage"**. Compute $\mathbb{E}[\theta, \gamma|\cdot]$ and sample $\boldsymbol{A}_m^* \sim \pi(\boldsymbol{A}^*|\boldsymbol{X}, \boldsymbol{A}, \mathbb{E}[\theta, \gamma|\cdot])$.
2. **"Plug-in"**. Sample multiple $\boldsymbol{A}^*$, estimate sufficient statistics of $p(\boldsymbol{Y}|\cdot)$ (e.g., exposure values), and sample η from the outcome module.

Numerical illustration

1. Outcome model $Y_i = \eta_0 + \eta_1 Z_i + \eta_2 \sum_{j \neq i} Z_j A_{ij}^* + \eta_3 X_i + \varepsilon_i$
2. Network generation $logit(\Pr(A_{ij}^* = 1)) = \theta_0 + \theta_1 |X_i - X_j|$.
3. Observed network from a random noise measurement error model.

- Implemented in the probabilistic programming language *NumPyro*. Accelerated cut-posterior sampling with *JAX* JIT compilation.
- MCMC sampling via NUTS.

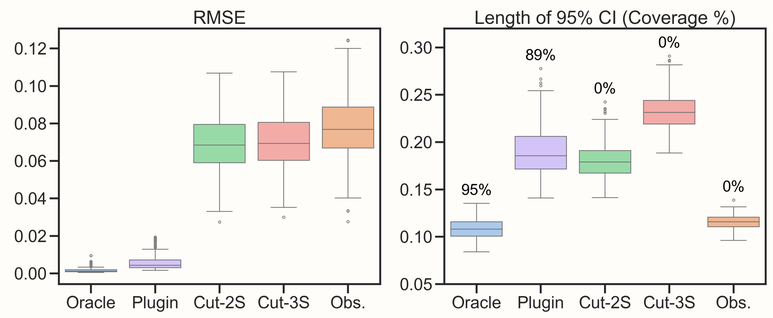

Figure 3. Distribution of η_2 RMSE and 95% credible intervals (coverage) for $n = 300$ and 300 replications. 'Oracle' and 'Obs.' display results using $\boldsymbol{A}^*, \boldsymbol{A}$, respectively. 'Cut-2S' is "Two-stage" sampling, and similarly for "Cut-3S".

References

[1] M. J. Bayarri, J. O. Berger, and F. Liu. Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150, Mar. 2009. Publisher: International Society for Bayesian Analysis.

[2] H. Crane and W. Dempsey. A Statistical Framework for Modern Network Science. *Statistical Science*, 36(1), 2021.

[3] P. E. Jacob, L. M. Murray, C. C. Holmes, and C. P. Robert. Better together? Statistical learning in models made of modules. Aug. 2017. arXiv:1708.08719 [stat].

[4] E. L. Ogburn, O. Sofrygin, I. Díaz, and M. J. van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 2022.

[5] B. Weinstein and D. Nevo. Causal inference with misspecified network interference structure. *arXiv preprint arXiv:2302.11322*, 2024.

[6] Y. Zhang, Z. Ghahramani, A. J. Storkey, and C. Sutton. Continuous Relaxations for Discrete Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[7] G. Zhou. Mixed Hamiltonian Monte Carlo for Mixed Discrete and Continuous Variables. In *Advances in Neural Information Processing Systems*, volume 33, pages 17094–17104. Curran Associates, Inc., 2020.