Identification and Estimation of Causal Effects from Dependent Data

Eli Sherman
esherman@jhu.edu

with Ilya Shpitser

Johns Hopkins Computer Science

12/6/2018

Causal Inference Problems in Networks

- Goal: learn about causality from data on interacting agents:
  - Online social networks, cluster randomized trials of villages or households, infectious diseases
- Major difficulty: units are dependent
- Example (Shalizi and Thomas[1]): "If your friend Sam jumped off a bridge...[2]

[1]Shalizi and Thomas 2011.

[2]Shutterfly ID 210011107

Causal Inference Problems in Networks

- Goal: learn about causality from data on interacting agents:
  - Online social networks, cluster randomized trials of villages or households, infectious diseases
- Major difficulty: units are dependent
- Example (Shalizi and Thomas[1]): "If your friend Sam jumped off a bridge...[2]

...would you jump too?"

[1] Shalizi and Thomas 2011.
[2] Shutterfly ID 210011107

Causal Inference Problems in Networks

- Example (Shalizi and Thomas[3]): "If your friend Sam jumped off a bridge, would you jump too?"
  - yes: want to imitate Sam because they're cool (social contagion)
  - yes: Sam infected you with a judgement-suppressing parasite (physical contagion)
  - yes: known shared interest in dangerous hobbies (observed homophily)
  - yes: unknown to analyst, both you and Sam are daredevils (latent homophily)
  - yes: you and Sam were both on the bridge as it started collapsing (external causation)
- In general, not possible to disentangle these

[3] Shalizi and Thomas 2011.

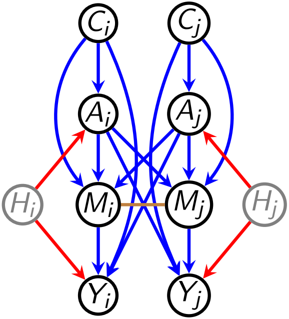
Causal Inference Problems in Networks

- Example (Shalizi and Thomas[3]): "If your friend Sam jumped off a bridge, would you jump too?"
  - yes: want to imitate Sam because they're cool (social contagion)
  - yes: Sam infected you with a judgement-suppressing parasite (physical contagion)
  - yes: known shared interest in dangerous hobbies (observed homophily)
  - yes: unknown to analyst, both you and Sam are daredevils (latent homophily)
  - yes: you and Sam were both on the bridge as it started collapsing (external causation)
- In general, not possible to disentangle these
  - Nevertheless, under some assumptions causal inference is possible!

[3] Shalizi and Thomas 2011.

A Motivating Social Networking Example

- Subject i spends time online A_i, leading to purchasing behavior Y_i
- This is mediated by participation in a social network M_i - entangled with participation of i's friends j
- Personal characteristics C_i act as confounders; unobserved confounding by H_i
- Counterfactual: if we artificially set i's online time, how would this influence j's behavior?



- This counterfactual query is complicated by:
  - *Interference* via $A_i \to M_j$, $A_i \to Y_j$
  - *Symmetric dependence* via $M_i - M_j$ edge; all Ms marginally correlated so we have one sample
  - Y_i and A_i confounded by H_i ($\to$)

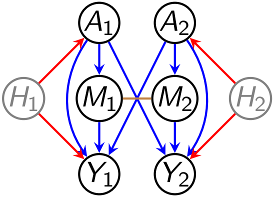
A Light Intro to Causal Inference[5]

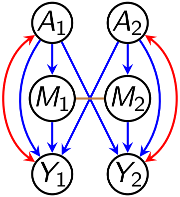
- Wish to simulate randomized control trials
  - Compare hypothetical cases ($A \leftarrow 1$) and controls ($A \leftarrow 0$)
  - Often interested in mean difference $\beta = E[Y(1)] - E[Y(0)]$
- Identification: is parameter β a function of observations?
  - Fundamental problem of causal inference: - only observe *assigned* treatment for each unit[4]
  - Sometimes identification is possible, for example:
    - $E[Y(1)] - E[Y(0)] = E[E[Y|A = 1, W] - E[Y|A = 0, W]]$
    - Identified if W is observed and encapsulates all confounders of A and Y
  - Non-identification $\implies$ ill-posed problem, even as $n \to \infty$
- Need models and assumptions for identification; we use graphical models

[4] Rubin 1976.

[5] Pearl 2009.

Chain Graphs and their Segregated Projections[7]





- Chain graphs represent models with
  - $A \rightarrow B$ - directed causal relationship from A to B
  - $A - B$ - feedback process at equilibrium between A and B
- Segregated graphs represent chain graph with latent variables
  - $A \leftrightarrow B$ - unmeasured confounding between A and B
- Complete identification algorithm
  - IN: segregated graph; OUT: estimable functional or 'failure'
  - Above demonstrates non-ID; can't disentangle effect $A_i \rightarrow Y_i$ from confounding $A_i \leftrightarrow Y_i$.
  - Algorithm extends ID algorithm for LV-DAGs[6]

[6] Tian and Pearl 2002; Shpitser and Pearl 2006.

[7] Lauritzen and Richardson 2002; Shpitser 2015.

Contributions

- Complete identification algorithm
  - Causal influence of i's online time on behavior of friend j identified:

$$\sum_{\{C_1, C_2, M_1, M_2\}} \Big[p(M_1, M_2 | a_1, a_2, C_1, C_2) \times \Big[\sum_{A_2} p(Y_2 | a_1, A_2, M_2, C_2) p(A_2 | C_2) p(C_1) p(C_2) \Big] \Big]$$

  - Failure means β is not identifiable in the model by *any* method
- Single sample inference with hidden variables
  - Gibbs sampling-based algorithm, 'Auto-G Computation'[8]
  - Experiments demonstrate consistency under correctly specified model
- The devil is in the details!
  - Come see our poster in 10 minutes: 10:45 AM - 12:45 PM in Room 210 & 230 AB #13
  - Read the paper: Identification and Estimation of Causal Effects from Dependent Data

[8] Tchetgen Tchetgen, Fulcher, and Shpitser 2017.

Works Cited I

Lauritzen, Steffen L. and Thomas S. Richardson (2002). "Chain graph models and their causal interpretations (with discussion)". In: *Journal of the Royal Statistical Society: Series B* 64, pp. 321–361.

Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press. ISBN: 978-0521895606.

Rubin, D. B. (1976). "Causal Inference and Missing Data (with discussion)". In: *Biometrika* 63, pp. 581–592.

Shalizi, Cosma Rohilla and Andrew C Thomas (2011). "Homophily and contagion are generically confounded in observational social network studies". In: *Sociological methods & research* 40.2, pp. 211–239.

Shpitser, Ilya (2015). "Segregated Graphs and Marginals of Chain Graph Models". In: *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc.

Works Cited II

Shpitser, Ilya and Judea Pearl (2006). "Identification of Joint Interventional Distributions in Recursive Semi-Markovian Causal Models". In: *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06)*. AAAI Press, Palo Alto.

Tchetgen Tchetgen, Eric J., Isabel Fulcher, and Ilya Shpitser (2017). *Auto-G-Computation of Causal Effects on a Network*. hhttps://arxiv.org/abs/1709.01577. Working paper.

Tian, Jin and Judea Pearl (2002). "A General Identification Condition for Causal Effects". In: *Eighteenth National Conference on Artificial Intelligence*, pp. 567–573. ISBN: 0-262-51129-0.