Qualifying Exam

Casey Breen

2020-12-27

Contents

Chapter 1

Summary

This set of memos covers two reading lists:

- Causal Inference and Population Studies
- Social Networks

The general format of the memos will be as follows:

- Key Background

- Data / Methods

- Research Question

- Argument / contribution

- Open Questions

Chapter 2

Causal Inference and Population Studies

2.1 Foundations

Adler, Nancy E., and David H. Rehkopf. 2008

Citation

Adler, Nancy E., and David H. Rehkopf. 2008. "U.S. Disparities in Health: Descriptions, Causes, and Mechanisms." Annual Review of Public Health.

Key Background

- Eliminating health disparities in a fundamental goal of public health research and practice.
- Studying health disparities is challenging due the (i) definition of a disparity and (ii) ability to attribute cause from association.

Methods

- Extensive literature review of both descriptive + causal methods for identifying substantice applications and Annual Review Article — summarizes key studies related to health disparities.

Research Question

- What are the key resear has been conducted on the causes and mechanisms of health disparities in the US, and how is this research conducted?
- What are the causes of health disparities in the US and why what are the specific pathways and mechanisms driving these disparities?

Argument / contribution

7

- A health disparity is a broad term loosely defined. The broadest definition refers to health differences that occur with respect to gender, race or ethnicity, education, income, geographic location, or sexual orientation. Health disparities result from both biological differences and social disparities, but the latter has a greater effect and in avoidable.
- Data limitations often preclude the study of SES and health—particularly the intersection of SES and Race/Ethnicity. Further, SES indicators may have different meanings for different groups (e.g., Blacks and Hispanics have lower wealth than non-Hispanic Whites and Asians at a given income level), further complicating "controlling" for a covariate.
- Descriptive understandings are important for (i) understanding short and long-term trends in mortality disparities, (ii) sparking causal investigations of health disparities, (iii) allocating resources to reduce disparities in specific diseases, and (iv) increasing public awareness.
- Analytic approaches for establishing causality include propensity score matching, instrumental variables, time-series analysis, causal structural equation modeling, and marginal structural models.

- Identifying the specific pathways and mechanisms by which SES and race/ethnicity affect health can strengthen causal claims and help target health interventions. For example, differential exposure to stress, particularly repeated exposure, is one of multiple pathways identified in the literature.

Unanswered Questions

- How do we collect data with adequate measures of SES, demographic covariates, and health outcomes to identify understand the mechanisms and pathways?

Xie, Yu. 2013.

Citation

Xie, Y. 2013. "Population Heterogeneity and Causal Inference." Proceedings of the National Academy of Sciences 110(16):6262–68. doi: 10.1073/pnas.1303102110.

Key Background

- The core objective of social science research is to understand the (ubiquitous) population heterogeneity. (Not identify abstract and universal laws).
- Causal inference with observational data is only possible with strong assumptions.

Methods

- Conceptualize selection into treatment as a dynamic process.
- Composition bias is generated by a dynamic process when the treatment proportion changes. Demonstrated via simulation.

Research Question

- What is the relationship between population heterogeneity and causal inference? How can composition bias arise from population heterogeneity?

Argument / contribution

- It is impossible to draw causal inference at the individual level due to heterogeneity.
- Composition bias, a form of selection bias, arises often in the real world. For example, if the administration of a medical treatment or social intervention is done on a graduated schedule, where participation is need based. Individuals selected ata later states will have lower ATEs because they are coming from a less responsive sub-population.

Key Findings

- Composition bias occurs because units with a higher intrinsic propensity of treatment are more likely to be over-represented when the treatment proportion is small.
- Researchers should be mindful of the specific subgroups of interest when deriving and interpreting average causal estimates from potentially heterogeneous subgroups.

Unanswered Questions

- When can we establish external validity for research results in?

Gangl, Markus. 2010

Citation

Gangl, Markus. 2010. "Causal Inference in Sociological Research." Annual Review of Sociology 36(1):21–47. doi: 10.1146/annurev.soc.012809.102702.

Key Background

- The counterfactual model provides a natural framework for clarifying the requirements for valid causal inference

Methods

- Summarizes recent literature on causal inference literature (estimands, quasi-experimental designs, diff-in-diff, instrumental variable estimation, semi and non-parametric methods)

Research Question

- What's the role of causal inference in sociological research?

Argument / contribution

- Traditional setup in sociology papers is insufficient—comparing regression specifications and testing competing hypotheses is unlikely to identify any causal effect of interest.
- While the benefits of RCTs are widely understood, regression using observational data is the workhorse of sociology research. Lit review provides a sobering view of the efficacy of drawing causal conclusions from regression analysis, as practiced in sociology.
- Counterfactual framework doesn't only apply to explicitly manipulable treatments—it can also apply to non-manipulable factors such as gender, race, or class.
- Standard treatment of effect estimates are local — historically and situationally contingent.

Key Findings

- The availability of longitudinal data and informative natural experiments will allow for causal identification within an area plagued by confounders.

Petersen, Maya L., Sandra E. Sinisi, and Mark J. van der Laan. 2006.

Citation

Petersen, Maya L., Sandra E. Sinisi, and Mark J. van der Laan. 2006. "Estimation of Direct Causal Effects." Epidemiology 17(3):276–84.

Key Background

- Most common problems in epidemiology (and social science more broadly) involve estimating the effect of an exposure on an outcome while blocking the exposure's effect on an intermediate variable. Estimation of direct effects is typically the goal of research attempting to estimate the causal pathways for which a treatment causes an outcome.
- Controlled direct Effect: hold intermediate variable at a fixed level. In contrast, a natural direct effect would measure the effect of the exposure, blocking the exposure's effect on intermediate background but allowing the intermediate to vary among individuals.

Methods

Research Question

What's the difference between a controlled and a natural direct effect and how can the natural direct effect be estimated (and with which underlying assumptions)?

Argument / contribution

- Natural direct effect of an exposure on an individual is defined as the difference in counterfactual outcome if an individual was unexposed vs. exposed allowing the intermediate to remain at counterfactual level under no exposure. For a controlled direct effect, the intermediate would be set to a fixed value for all persons.
- To estimate the direct effect in the whole population, one can take the weighted average of subgroup-specific direct effects with the weight fort a given subgroup determined by the relative size of the subgroup in comparison with the population.
- Formal assumptions for estimating controlled and natural effects

$$A \perp Y_{AZ}|W Z \perp Y_{AZ}|A, W \tag{2.1}$$

In words, this means no unmeasured confounders of either the effect of the exposure on the outcome of the effect of the intermediate on the outcome.

Additionally, to identify natural direct effects, we need:

$$A \perp Z_a|W \tag{2.2}$$

which in words says there are no unmeasured confounders of the effect of the intermediate on the treatment.

Second, within subgroups defined by covariates, the level of the intermediate variable in the absense of exposure does not tell us anything about the expected magnitude of exposure's effect at a controlled level of the intermediate variable. (Direct effect assumption)

$$E(Y_{az} - Y_{0z}|Z_0 = z, W) = E(Y_{az} - Y_{0z}|W) \tag{2.3}$$

Key Findings

- The barriers to estimation of direct effects are not as great as have been previously suggested. Researchers are encouraged to estimate direct effects which giving appropriate consideration to relevant assumptions and interpretations of their estimates.

Imai, Kosuke, Luke Keele, Dustin Tingley, and Teppei Yamamoto. 2011.

Key Background - Many empirical studies focus on establishing whether on variable affects another and fails to explain how such a causal relationship arises. This "black box" approach to causality has been criticized across disciplines. - A causal mechanism is a process in which a causal variable of interest influences an outcome. - Quantitative investigation of causal mechanisms is based on the estimation of causal mediation effects.

Methods - Review of causal inference literature, particularly as it applies to the estimation of causal mediation effects.

Research Question
How can researchers better identify the causal mechanisms behind observed causal effects.

Argument / contribution - Consistency assumption—potential outcomes must take the same values as long as the treatment and mediator values are the same—is especially important in the analysis of causal mechanisms. Even if experimental designs involve the manipulation of mediators, one must assume that subjects would respond in the same way if those values of mediators were spontaneously chosen by the subjects themselves. - Instrumental variables: assume that there is no direct effect on the outcome and effects all units in one direction. However, this often leads to the "black box" approach to causal inference, where insufficient attenuation is paid to causal mechanisms.
- Given the difficult of studying causal mechanisms, some researchers believe that a focus should be place on identifiation of causal effects and not causal mechanisms.

Key Finding / Conclusion

- Paper demonstrates three ways to move forward in research into causal mechanisms. First, the potential outcomes model of causal inference. Second, a sensitivity analysis to evaluate robustness of conclusions. Finally, new research designs for experimental and observational studies can reduce need for untestable conclusions. These new methods allow the black box of causality to be unpacked, going beyond the estimation of causal effects.

Unanswered Questions

--- **Key   Background   Methods   Research Question**
Argument / contribution Key Findings Unanswered Questions

Chapter 3

Social Networks

We describe our methods in this chapter.