Causal Data Science for Business Decision Making

Transportability & External Validity

Paul Hünermund

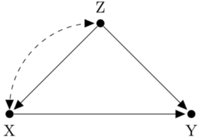
COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

The Data Fusion Process

(1) Query:

Q = Causal effect at target population

(2) Model:



(3) Available Data:

Observational:	$P(v)$
Experimental:	$P(v \mid do(z))$
Selection-biased:	$P(v \mid S = 1) +$ $P(v \mid do(x), S = 1)$
From different populations:	$P^{(source)}(v \mid do(x)) +$ observational studies

Causal Inference Engine:

Three inference rules of *do-calculus*

Solution exists? Yes

Estimable expression of Q

No

Assumptions need to be strengthened (imposing shape restrictions, distributional assumptions, etc.)

Motivating Example: Banerjee et al. (2007)

- Banerjee et al. (2007) study the effect of a randomized remedial education program for third and fourth graders in two Indian cities: Mumbai and Vadodara
  - They find similar effects on math skills, but effect positive impact on language proficiency is much smaller in Mumbai compared to Vadodara
- Banerjee et al. (2007) explain this result by baseline reading skills that were higher in Mumbai, because families are wealthier there and schools are better equipped
- What do we do if we do not have a second experiment to validate our results?
  - Naïve extrapolation (also called *direct transportability*) clearly would have gotten them into trouble

External Validity of A/B Tests

A/B Testing and Covid-19: Data-Driven Decisions in Times of Uncertainty

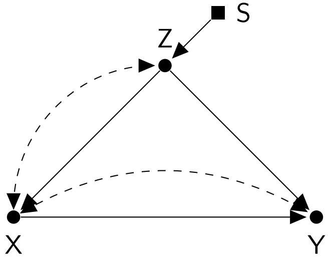
June 26, 2020

I ran an A/B test: will the results be valid once "normal" life resumes?

We can't know. User behavior always changes over time, so it is never certain whether the impact of a feature will persist 2, 3, or 12 months from now. A/B tests help us to optimize features for current usage patterns. But rarely have people's activities, work style, and concerns changed so drastically. Don't assume that usage will stabilize in the same patterns they follow now, or in the patterns you observed pre-Covid. If you run an A/B test now and decide to fully deploy, consider re-running the A/B test comparing the old variant again as circumstances evolve. Re-running the A/B test will allow you to see if the treatment effect has substantially changed. Consider this especially if:

Selection Diagram

- We can incorporate knowledge about structural differences across domains by a selection node (■) in a causal diagram
- Captures the notion that domains differ either in the distribution of background factors $P(U_i)$ or causal mechanisms f_i in the underlying structural causal model
  - Differences across domains can be in arbitrary ways (akin to the nonparametric nature of DAGs)



Selection Diagram (II)

Definition: Selection Diagram (Pearl and Bareinboim, 2011)

Let $\langle M, M^* \rangle$ be a pair of structural causal models relative to domains $\langle \Pi, \Pi^* \rangle$, sharing a causal diagram G. $\langle M, M^* \rangle$ is said to induce a selection diagram D if D is constructed as follows:

(i) Every edge in G is also an edge in D.

(ii) D contains an extra edge $S_i \rightarrow V_i$ whenever there might exist a discrepancy $f_i \neq f_i^*$ or $P(U_i) \neq P^*(U_i)$ between M and M^*.

- Switching across domains Π and Π^* is denoted by conditioning on different values of S
  - Note: $P^*(V) = P(V|S = 1)$
- Compared to selection bias case, now S points into other variables

Transportability Task

- The transportability problem: We have experimental results from a source domain Π, how can we transport them to a target Π^* where we only have passive observations?
  - I.e., we know $P(y|do(x))$ but would like to know $P^*(y|do(x))$
- Note that Π and Π^* share the same causal diagram G. Thus, if the causal effect in Π would be identified from observational data alone (i.e., no experimental data needed) then it would also be identified in Π^* and there would be no need for transportation ("trivial transportability", Pearl and Bareinboim, 2011)
- So transportability theory is (mainly) concerned with transporting experimental results across domains
  - Although observational / statistical transportability can be useful to economize on data collection efforts (Pearl and Bareinboim, 2011)

S-Admissibility

Theorem 2 in Pearl and Bareinboim (2011)

Let D be the selection diagram characterizing two populations, Π and Π^*, and S the set of selection variables in D. The strata-specific causal effect $P^*(y|do(x), z)$ is transportable from Π to Π^* if Z d-separates Y from S in the X-manipulated version of D, that is, Z satisfies $(Y \perp\!\!\!\perp S|Z)_{D_{\overline{X}}}$.

- The set of variables Z is then called *s-admissible*
- Note that $D_{\overline{X}}$ is the result of experimentally manipulating X in the source domain
  - Since S has to be d-separated from Y in $D_{\overline{X}}$, it follows that every S pointing into X does not threaten transportability and can be ignored

S-Admissibility (II)

Corollary 1 in Pearl and Bareinboim (2011)

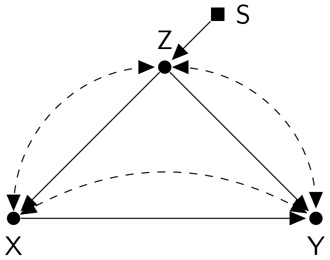
The causal effect $P^*(y|do(x))$ is transportable from Π to Π^* if there exists a set Z of observed pretreatment covariates that is s-admissible. Moreover, the transport formula is given by the weighting

$$P^*(y|do(x)) = \sum_z P(y|do(x), z)P^*(z).$$

- This *transport formula* says that we reweight the z-specific causal effect in the source domain by the distribution of z in the target domains
  - E.g., find experimental results for several income levels in Mumbai and weight by income distribution in Vadodara

S-Admissibility (III)

- Consider this selection diagram, which is the same as before expect for the added edge $Z \dashleftarrow\dashrightarrow Y$
- Is the causal effect transportable in this case?



Transportability – The General Case

- Transportability formula is well-known in economics (Hotz et al., 2005; Andrews and Oster, 2018), but these papers focus on s-admissible *pretreatment* variables. Solutions based on do-calculus are more general

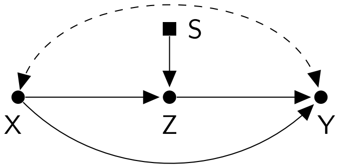
Theorem 1 in Pearl and Bareinboim (2011)

Let D be the selection diagram characterizing two populations, Π and Π^*, and S as set of selection variables in D. The relation $R = P^*(y|do(x))$ is transportable from Π to Π^* if the expression $P(y|do(x), s)$ is reducible, using the rules of do-calculus, to an expression in which S appears only as a conditioning variable in do-free terms.

- Bareinboim and Pearl (2013a) develop a complete algorithm for automatization

Example: Selection Affecting Post-Treatment Variables

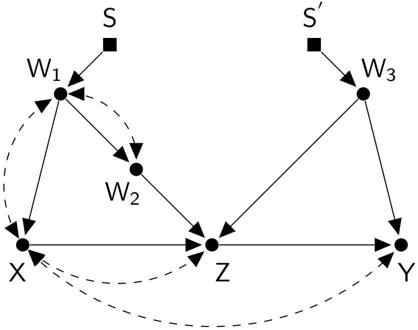
- Gordon et al. (2019) discuss an example in which the effectiveness of a social media advertising campaign, X, is mediated by users' exposure to ads, Z



- Imagine a company that runs advertising campaigns on the desktop version of a social media platform
- Exposure to ads differs across the desktop and mobile version of the platform
- How can experimental results be transported from desktop to mobile?
  - With the help of the transportability algorithm developed by Bareinboim and Pearl (2013a) we find the transport formula

$$P^*(y|do(x)) = \sum_z P(y|do(x), z)P^*(z|x)$$

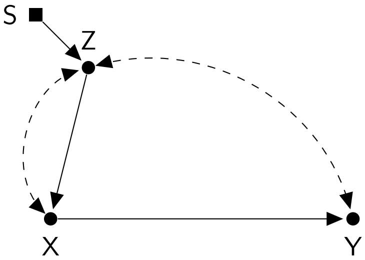
Example: A Complex Selection Diagram



- Here the relevant transport formula is found to be

$$P^*(y|do(x)) = \sum_{z,w_2,w_3} P(y|do(x), z, w_2, w_3) P(z|do(x), w_2, w_3) P^*(w_2, w_3)$$

z-Transportability



- What if we do not have the possibility to run experiments on the treatment X, but we can conduct a surrogate experiment on Z instead (as in the encouragement design from development economics discussed previously)
- This gives rise to the idea of z-transportability

z-Transportability (II)

Theorem 1 in Bareinboim and Pearl (2013b)

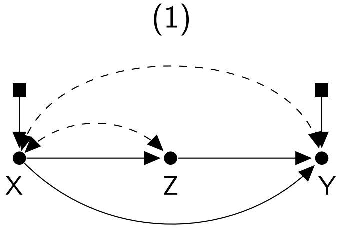
Let D be the selection diagram characterizing two populations, Π and Π^*, and S as set of selection variables in D. The relation $R = P^*(y|do(x))$ is z-transportable from Π to Π^* in D if the expression $P(y|do(x), s)$ is reducible, using the rules of do-calculus, to an expression in which all do-operators apply to subsets of Z, and the S-variables are separated from these do-operators.

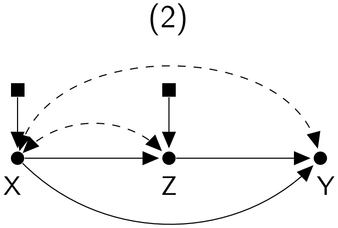
- Again, this theorem is only procedural. Bareinboim and Pearl (2013b) develop a complete algorithm for the z-transportability case

Meta-Transportability

- Transportability techniques are particularly valuable we can combine results from several source domains
- This strategy is generally known under the rubric of “meta-analysis”
  - Meta-analyses become increasingly popular in economics (Card et al., 2010; Dehejia et al., 2015)
- The problem with standard meta-analytic tools is that they do not take domian heterogeneity into account but instead aim to “average out” differences across populations
- Bareinboim and Pearl (2013c) extend the transportability idea, which captures domain-specific heterogeneity by the ■-nodes in the causal diagram, to the case with multiple source domains

Meta-Transportability (II)





- Both selection diagrams D_1 and D_2 depict how domains π_1 and π_2 differ from the target domain π^*
- Here, the causal effect would not be individually transportable from a single source domain. But it is transportable using combined information from both domains as

$$P^*(y|do(x)) = \sum_z P^{(2)}(y|do(x), do(z))P^{(1)}(z|do(x)).$$

Meta-Transportability (III)

- Bareinboim and Pearl (2013c) develop a complete algorithm for deciding about meta-transportability
- Bareinboim and Pearl (2014) combine the idea of meta-transportability with z-transportability to what they call *mz*-transportability
  - Bareinboim and Pearl (2014) develop a complete algorithm for *mz*-transportability
- These results will hopefully allow to combine more flexibly and make more effective use of results from a whole body of empirical literature

Thank you

Personal Website: p-hunermund.com

Twitter: @PHuenermund

Email: phu.si@cbs.dk

References I

Isaiah Andrews and Emily Oster. Weighting for external validity. NBER Working Paper No. 23826, 2018.

Abhijit V. Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden. Remedying education: Evidence from two randomized experiments in india. *The Quartely Journal of Economics*, 122(3):1235–1264, 2007.

Elias Bareinboim and Judea Pearl. A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134, 2013a.

Elias Bareinboim and Judea Pearl. Causal transportability with limited experiments. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 95–101, 2013b.

Elias Bareinboim and Judea Pearl. Meta-transportability of causal effects: A formal approach. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 31, Scottsdale, AZ, 2013c.

Elias Bareinboim and Judea Pearl. Transportability from multiple environments with limited experiments: Completeness results. In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances of Neural Information Processing Systems*, volume 27, pages 280–288, November 2014.

David Card, Jochen Kluve, and Andrea Weber. Active labour market policy evaluations: A meta-analysis. *The Economic Journal*, 120:452–477, 2010.

Rajeev Dehejia, Cristian Pop-Eleches, and Cyrus Samii. From local to global: External validity in a fertility natural experiment. NBER Working Paper No. 21459, 2015. URL `https://www.nber.org/papers/w21459`.

Brett R. Gordon, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38:193–364, 2019. URL `https://www.kellogg.northwestern.edu/faculty/gordon_b/files/fb_comparison.pdf`.

References II

V. Joseph Hotz, Guido W. Imbens, and Julie H. Mortimer. Predicting the efficacy of future training programs using past experiences at other locations. *Journal of Econometrics*, 125:241–270, 2005.

Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.