On the Identifiability and Estimation of Causal Effects

Machine Learning Coffee Seminar

Antti Hyttinen, University Researcher

Jussi Viinikka, Doctoral Student

Sums of Products research group
HIIT, Department of Computer Science, University of Helsinki

26.10.2020

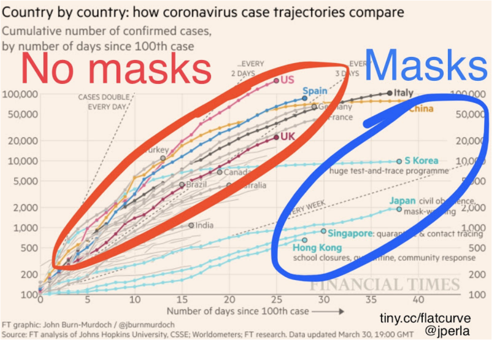
1. Background: Causal Effects and Non-parametric Identifiability
2. Identifying causal effects via CSI relations
3. Background: Linear Causal Effect Estimation from Data
4. Bayesian Posteriors for Linear Causal Effects
5. Gadget: Scalable MCMC Sampling of DAGs
6. Conclusion

Section 2:
S. Tikka, A. Hyttinen, and J. Karvanen.
Identifying causal effects via context-specific independence relations. In NeurIPS, 2019.

Sections 4 & 5:
J. Viinikka, A. Hyttinen, J. Pensar, and M. Koivisto.
Towards scalable Bayesian learning of causal DAGs.
In NeurIPS, 2020.

Background: Causal Effects and Non-parametric Identifiability

The Need for Quantifying Causal Effects



- Correlation does not imply causation.
- How large is the causal effect?
- Lack of evidence vs. evidence for an insignificant effect.

Causal Effects

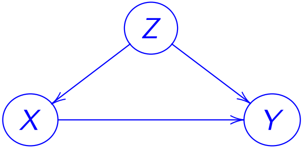
- **Causal effects** are probability distributions e.g:

  $$P(Y|\text{do}(X)),$$

  where $\text{do}(X)$ intervenes at X and sets it to e.g. x.
- E.g. $P(\text{cancer}|do(\text{smoke}))$, $P(\text{Infection}|do(\text{wear a mask}))$.
- Randomized controlled trials (RCTs) are a direct way to obtain $P(Y|do(X))$.
  - Infeasible? Unethical? Expensive? Sample size? Population?

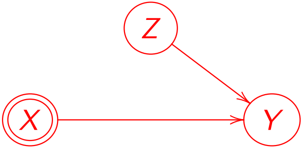
Causal Bayesian Networks

Passive observation



$P(X, Y, Z) = P(Z)P(X|Z)P(Y|Z, X)$

do(X)



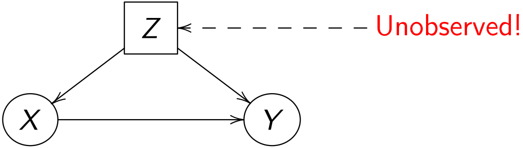
$P(Y, Z|do(X)) = P(Z)P(Y|Z, X)$

- Edges in the DAG denote direct causal relationship.
- CPDs define **stochastically** how each variable gets its value based on its direct causes.
- Dependence corresponds to reachability in the graph.
- Intervention corresponds to edge breaking, or dropping from the factorization.
- Generally $P(Y|do(X)) \neq P(Y|X)$ (doing vs. seeing).

Determining Causal Effects?

1. What data do you have?
   - interventions (RCTs)?
   - missing data?
2. What background knowledge do you have?
   - causal graph?
3. Which assumptions you are willing to make?
   - acyclicity?
   - Causally sufficient or latent confounders?
   - parametric restrictions, e.g., linearity?
   - selection bias?
4. Which output do you want?
   - Identifiability?
   - estimation?
   - bounds?
   - Average causal effect?

Causal Effect Identifiability [Pearl, 2000]



Problem (Causal Effect Identifiability)

Input: *A DAG over V,*
passively observed $P(W)$ for $W \subseteq V$,
a query $P(Y|do(X))$.

Task: *Output a formula for $P(Y|do(X))$ over $P(W)$, or decide that it is non-identifiable.*

- Can the effect $P(Y|do(X))$ be uniquely computed from $P(\cdot)$?
- Or, are there two different parameterizations that yield same $P(\cdot)$ but different $P(Y|do(X))$?
- Aim for a general and complete theory!

Do-Calculus [Pearl, 1995]

Rule 1 (Insertion/deletion of observations):

$P(Y|do(X), Z, W) = P(Y|do(X), W)$ if $Y \perp\!\!\!\perp Z \,|X, W$ in $G_{\overline{X}}$
(edges into X removed)

Rule 2 (Action/observation exchange):

$P(Y|do(X), do(Z), W) = P(Y|do(X), Z, W)$ if $Y \perp\!\!\!\perp Z \,|X, W$ in $G_{\overline{X},\underline{Z}}$
(edges into X removed,
edges out of Z removed)

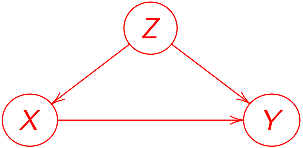
Rule 3 (Insertion/deletion of actions):

$P(Y|do(X), do(Z), W) = P(Y|do(X), W)$ if $Y \perp\!\!\!\perp Z \,|Z, W$ in $G_{\overline{X},\overline{Z(W)}}$
(edges into X removed,
and in it edges into Zs that
are not ancestors of W removed)

- Together with probability calculus!

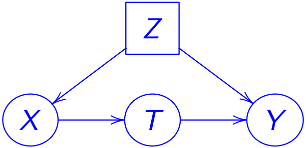
Identifiability and the ID-algorithm [Tian and Pearl, 2002, Shpitser and Pearl, 2006a]

Backdoor



$$P(Y|do(X)) = $$
$$= \sum_Z P(Y|Z,X)P(Z)$$

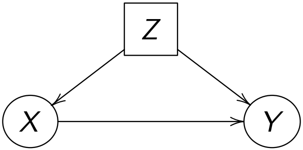
Front door



$$P(Y|do(X)) = $$
$$\sum_T P(T|X) \sum_{X'} P(C|T,X')P(X')$$

- The ID-algorithm can find the expressions in polynomial time.
- Use probabilistic modelling (e.g. BN) to calculate the terms.

Non-identifiability and Completeness



- Impossible to untangle the dependence through $X \rightarrow Y$ from the dependence through $X \leftarrow Z \rightarrow Y$.
- A graphical object called **hedge** witnesses **non-identifiability**.

#1	#2
$P(Z=1)=0.5$ $P(X=1 \mid Z)=0.5$ $Y \mid X,Z = X$	$P(Z=1)=0.5$ $X \mid Z = Z$ $Y \mid X,Z = Z$
$P(X, Y=X)=0.5$	$P(X, Y=X)=0.5$
$P(Y=1 \mid do(X=1))=1$	$P(Y=1 \mid do(X=1))=0.5$

- ID and do-calculus are **complete**. [Shpitser and Pearl, 2006a, Huang and Valtorta, 2006]

State of the Art in Non-parametric Identifiability

	Problem (Reference)	Target	Input (assumptions)	Missing data pattern	Method (complete)
1	CE identifiability [Shpitser and Pearl, 2006a]	$P(Y \mid do(X))$	$P(W)$	None	ID (Yes)
2	CE identifiability [Shpitser and Pearl, 2006b]	$P(Y \mid do(X), Z)$	$P(W)$	None	IDC (Yes)
3	z-identifiability [Bareinboim and Pearl, 2012]	$P(Y \mid do(X), Z)$	$P(W)$, $P(W \setminus B \mid do(B))$ (NE, ED)	None	zID (Yes)
4	g-identifiability [Lee et al., 2019]	$P(Y \mid do(X))$	$\{P(W \setminus B_i \mid do(B_i)\}$ (ED)	None	gID (Yes)
5	Surrogate outcome [Tikka and Karvanen, 2019]	$P(Y \mid do(X), Z)$	$\{P(A_i \mid do(B_i), C_i)\}$ (NE, SO)	None	TRSO (No)
6	mz-transportability [Bareinboim and Pearl, 2014]	$P(Y \mid do(X), Z)$	$\{P(W \setminus (B_i \cup T_i) \mid do(B_i), T_i)\}$ (NEDD, ED)	None	$\mathrm{TR}^{\mathrm{mz}}$ (Yes)
7	Selection bias [Bareinboim and Tian, 2015]	$P(Y \mid do(X), Z)$	$P(W \setminus S \mid S)$	Selection	RC (Unknown)
8	***Gen. identifiability*** [Tikka et al., 2020]	$P(Y \mid do(X), Z)$	$\{P(A_i \mid do(B_i), C_i)\}$	***None***	***do-search*** (Unknown)
9	Missing data [Mohan et al., 2013]	$P(W)$	$P(W^*)$	Restricted	– (Yes)
10	Missing data [Bhattacharya et al., 2019]	$P(W)$	$P(W^*)$	Arbitrary	– (Unknown)
11	***Gen. identifiability*** [Tikka et al., 2020]	$P(Y \mid do(X), Z)$	$\{P(A_i^* \mid do(B_i), C_i^*)\}$	***Arbitrary***	***do-search*** (No)

Identifying causal effects via CSI relations

Context-specific Independence [Boutilier et al., 1996]

$$X \perp\!\!\!\perp Y | Z = 0$$

i.e. $P(X|Y, Z = 0) = P(X|Z = 0)$

but $X \not\perp\!\!\!\perp Y | Z = 1$ (possibly)

- A very natural independence restriction, for example:

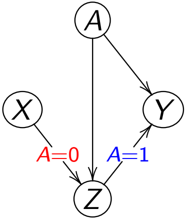
  $$\text{INCOME} \perp\!\!\!\perp \text{WEATHER}|\text{JOB} = \text{clerk}$$

  $$\text{INCOME} \not\perp\!\!\!\perp \text{WEATHER}|\text{JOB} = \text{farmer}$$

- CSIs have been extensively exploited in BN inference, but only recently been used to make novel causal inferences. [Hyttinen et al., 2018, Mooij et al., 2020]

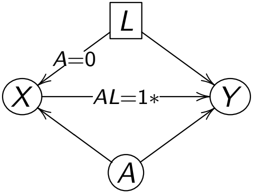
Labeled DAGs [Pensar, Nyman, Koski, and Corander, 2015]

$P(Z \mid A, X)$	$Z = 0$	$Z = 1$
$AX = 00$	0.1	0.9
$AX = 01$	0.1	0.9
$AX = 10$	0.5	0.5
$AX = 11$	0.6	0.4



- A label on an edge encodes contexts where the edge is absent.
- Any assignment in a label denotes a local CSI:
  e.g. $X \perp\!\!\!\perp Z \mid A = 0$.
- Labels allow for representation, theory on equivalence classes, and separation criteria.

Causal Effect Identifiability via CSIs [Tikka, Hyttinen, and Karvanen, 2019]



Problem (Causal Effect Identifiability via CSIs)

Input: *An LDAG over V,*
passively observed $P(W)$ for $W \subseteq V$,
a query $P(Y|do(X))$.

Task: *Output a formula for $P(Y|do(X))$ over $P(W)$,*
or decide that it is non-identifiable.

CSI-do-calculus [Tikka, Hyttinen, and Karvanen, 2019]

Rule 1 (Insertion/Deletion of observations):

$$P(Y_1, y_2|Z_1, z_2, X_1, x_2) = P(Y_1, y_2|X_1, x_2) \text{ if } Y_1, Y_2 \perp\!\!\!\perp Z_1, Z_2|X_1, x_2$$

Rule 2 (Marginalization/Sum-rule): $P(Y_1, y_2|X_1, x_2) = \sum_Z P(Y_1, y_2, Z|X_1, x_2)$

Rule 3 (Conditioning): $P(Y_1|Z_1, z_2, X_1, x_2) = \dfrac{P(Y_1, Z_1, z_2|X_1, x_2)}{\sum_{Y_1} P(Y_1, Z_1, z_2|X_1, x_2)}$

Rule 4 (Product-rule): $P(Y_1, y_2, Z_1, z_2|X_1, x_2) = P(Y_1, y_2|Z_1, z_2, X_1, x_2)P(Z_1, z_2|X_1, x_2)$

Rule 5 (General-by-case): $P(Y_1, y_2, 1 - z|X_1, x_2) = P(Y_1, y_2|X_1, x_2) - P(Y_1, y_2, z|X_1, x_2)$

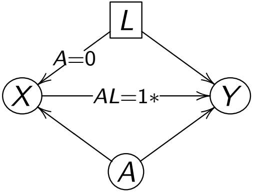
Rule 6 (Case-by-case): $P(Y_1, y_2, Z|X_1, x_2) = \begin{cases} P(Y_1, y_2, Z = 0|X_1, x_2) \\ P(Y_1, y_2, Z = 1|X_1, x_2) \end{cases}$

Rule 7 (Case-by-general (a)): $P(Y_1, y_2, z|X_1, x_2) = P(Y_1, y_2, Z|X_1, x_2)\big|_{Z=z}$

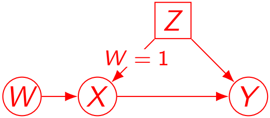
Rule 8 (Case-by-general (b)): $P(Y_1, y_2|X_1, x_2, z) = P(Y_1, y_2|X_1, x_2, Z)\big|_{Z=z}$

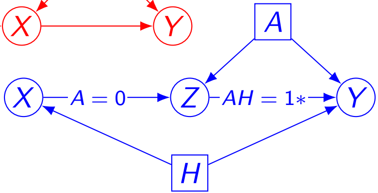
- These rules subsume do-calculus, do-operator not needed.
- Deciding non-identifiability is NP-hard: we need case by case reasoning.

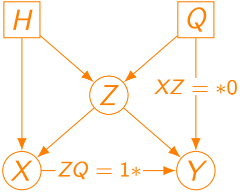
Extended Identifiability for $P(Y|do(X))$ via CSIs

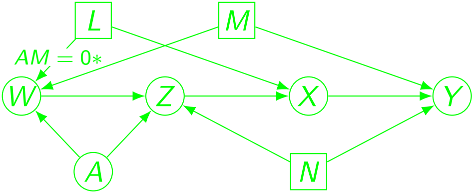


$$P(Y|A=0,X)P(A=0)$$
$$+P(Y|A=1)P(A=1)$$









→ **Few CSIs may be sufficient to turn a previously non-identifiable instance into identifiable.**

Background: Linear Causal Effect Estimation from Data

Linear Causal Effect Estimation from Data

Problem (Linear Causal Effect Estimation from Data)

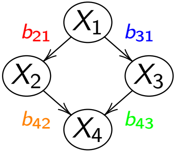
Input: *Passively observed* ***causally sufficient*** *data* **D**.

Task: *Estimate the* ***linear*** *causal effect*

$$\pi_{ji} = \frac{\partial}{\partial x_i} E\big(X_j \mid do(X_i = x_i)\big)$$

.

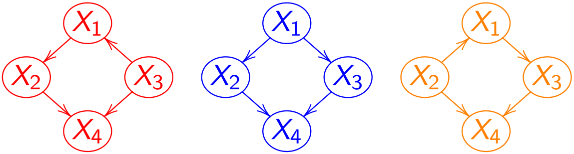
- Sum-product of edge coefficients on directed paths.



$$\pi_{41} = b_{42} \cdot b_{21} + b_{43} \cdot b_{31}$$
$$= -1.03 \cdot .78 + .74 \cdot .60 = -.36$$

IDA [Maathuis, Kalisch, and Bühlmann, 2009]

- Gaussian DAGs only identifiable up to the Markov eq. class.



- For a fixed DAG, the causal effect is identifiable via linear regression over the cause X_1 and its parents (backdoor adj.):

$$X_4 \sim X_1 + X_3, \quad X_4 \sim X_1, \quad X_4 \sim X_1 + X_2$$

- Output is a set of possible causal effects:

$$\pi_{41} \in \{-.74, . - .29, .49\}$$

- Need only the possible parent sets of the cause X_1 via e.g. PC.

Bayesian Posteriors for Linear Causal Effects

BIDA [Pensar, Talvitie, Hyttinen, and Koivisto, 2020, AAAI]

Posterior of π_{ji}

Bayesian lin. regression of X_j over X_i,$pa(X_i)$.

computed **exactly** via dynamic prog.

$$f(\pi_{ji}|\mathbf{D}) = \sum_{pa(X_i)} f(\pi_{ji}|\mathbf{D}, pa(X_i))P(pa(X_i)|\mathbf{D})$$

Bayesian model averaging over DAGs

Beeps: Bayesian Effect Estimation by Posterior Sampling [Viinikka, Hyttinen, Pensar, and Koivisto, 2020, NeurIPS]

Posterior of π_{ji}

Bayesian model averaging over DAGs

Sample G using MCMC

$$f(\pi_{ji}|\mathbf{D}) = \sum_{G,\mathbf{B}} f(\pi_{ji}|\mathbf{B}) f(\mathbf{B}|\mathbf{D}, G) P(G|\mathbf{D})$$

Sample coefficients $\mathbf{B}$ (BGe prior)

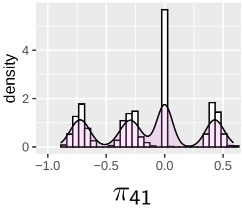
$$\pi_{ji} = ((\mathbf{I} - \mathbf{B})^{-1})[j, i]$$

IDA example continued...

IDA

$\pi_{41} \in \{-.74, -.29, .49\}$

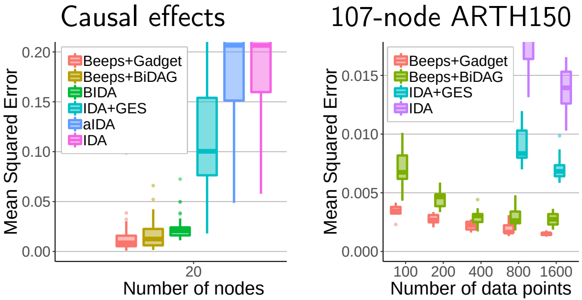
Beeps



- Uncertainty in the estimates due to low sample size.
- Uncertainty over the causal structure.

→ Beeps catches both.

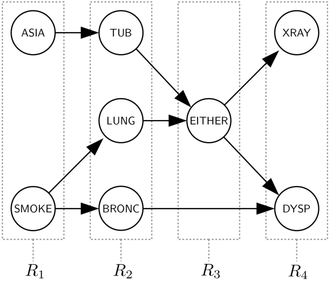
Simulation Results



- Better accuracy than IDA-based methods.
- We can scale up to over 100 nodes and our MCMC outperforms BiDAG.
- Concurrent similar suggestions by Kuipers et al. [2019] and Castelletti and Consonni [2020].

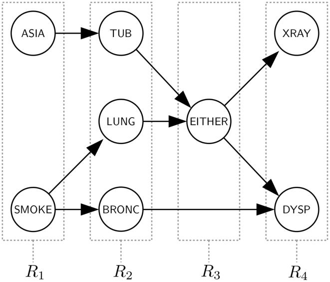
Gadget: Scalable MCMC Sampling of DAGs

Background – Root-partitions (Kuipers & Moffa, 2017)



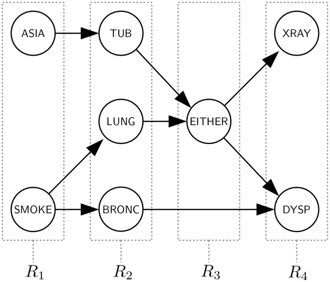
- Root-partition R: ordered set partition of DAG nodes V
- Iterate: root-nodes to the next part, remove them from the DAG
- For $v \in R_i$ parents valid if $pa(v) \subseteq \underbrace{R_{1,i-1}}_{U},\ pa(v) \cap \underbrace{R_{i-1}}_{T} \neq \emptyset$
- $\pi(R) = \prod_{t=1}^{k} \prod_{i \in R_t} \tau_i(R_{1,t-1}, R_{t-1})$, with $\tau_i(U,T) := \sum_{S \subseteq U : S \cap T \neq \emptyset} \pi_i(S)$

Background – *K* candidate parents (Kuipers et al., 2020)



- For each node $i \in V$ select K candidate parents C_i
- Now $\tau_i(U, T) = \tau_i(U \cap C_i, T \cap C_i)$, $O(3^K)$ space per node

Preprocessing – Scoring nodes by subtraction

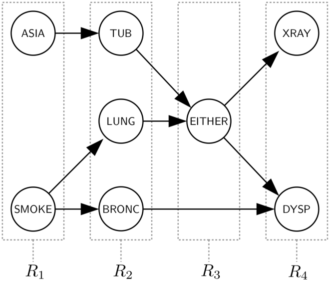


- For each node $i \in V$ select K candidate parents C_i
- Now $\tau_i(U,T) = \tau_i(U \cap C_i, T \cap C_i)$, $O(3^K)$ space per node
- For any $i \in V$ and $J \subseteq V \setminus \{i\}$, let

$$\tau_i(J) := \sum_{S \subseteq J \cap C_i} \pi_i(S)$$

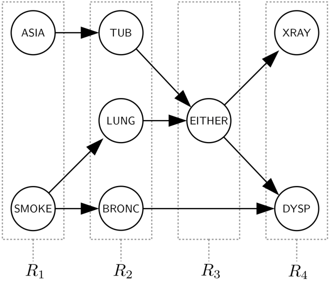
- Now $\tau_i(U,T) = \tau_i(U) - \tau_i(U \setminus T)$, $O(2^K)$ space per node

Preprocessing – Parents outside of the candidates



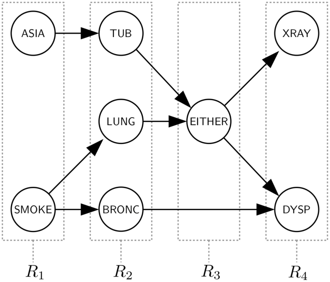
- Need to allow small number of parents outside of candidates too
  i. C_i may not be optimal or large enough for all $i \in V$
  ii. Posterior landscape may contain large zero-probability regions making transition between root-partitions inefficient for the MCMC
- We allow parent sets of maximum indegree d not contained in C_i
- Scores sorted, only need to accumulate certain amount to reach acceptable error

Markov chain – Metropolis Coupled Markov Chain Monte Carlo



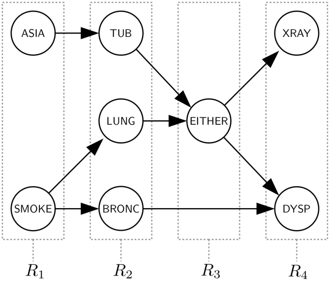
- Moves in root-partition space:
  - split R_i
  - merge R_i and R_{i+1}
  - swap nodes between R_i and R_j
- *M* parallel “heated” chains with the stationary distribution of *k*th chain proportional to $\pi^{k/M}$, i.e. Metropolis coupling
- On every other step a randomly chosen chain and neighbour swap states with certain probability

Post processing – Sampling DAGs



- DAG is sampled for each root-partition stored during Markov chain simulation
- Datastructure allowing fast parent set sampling after time and space invested in precomputing
- Sampling proceeds one node at a time for all DAGs generated

Gadget – Summary



1. Preprocessing
   - Candidate parent selection
   - Precomputing data structures for scoring root-partitions
2. Markov chain simulation
3. Postprocessing
   - Generating DAGs from root-partitions

Conclusion

Research directions

- Unknown graph and latent variables. [Hyttinen et al., 2015, Malinsky and Spirtes, 2017, Jaber et al., 2019]
- Soft interventions. [Correa and Bareinboim, 2020]
- Cyclic causal graphs. [Forré and Mooij, 2019]
- Intervals and bounds. [Malinsky and Spirtes, 2017, Peters et al., 2016]
- Path specific effects. [Malinsky et al., 2019]
- Linear identifiability. [Kumor et al., 2020]
- Counterfactuals. [Kusner et al., 2017, Shpitser and Pearl, 2008]

Conclusion

- Need causal effects from the data and knowledge we have.
- CSIs allow for identifiability beyond do-calculus.
- Bayesian posteriors for linear causal effects are more accurate and characterize the remaining uncertainty.
- State of the art MCMC posterior sampling for DAGs.
- Future work:
  - Completeness of CE identification via CSIs?
  - Relax the assumptions for Bayesian posteriors?
  - How to select candidate parents? How to scale up further?

Collaborators: Mikko Koivisto (UH), Johan Pensar (University of Oslo), Santtu Tikka (University of Jyväskylä), Juha Karvanen (University of Jyväskylä), Topi Talvitie (UH)

References I

E. Bareinboim and J. Pearl. Causal inference by surrogate experiments: z-identifiability. In N. de Freitas and K. Murphy, editors, UAI, pages 113–120. AUAI Press, 2012.

E. Bareinboim and J. Pearl. Transportability from multiple environments with limited experiments: Completeness results. In NeurIPS, pages 280–288, 2014.

E. Bareinboim and J. Tian. Recovering causal effects from selection bias. In AAAI, pages 3475–3481, 2015.

Rohit Bhattacharya, Razieh Nabi, Ilya Shpitser, and James M. Robins. Identification in missing data models represented by directed acyclic graphs. In UAI, 2019.

Craig Boutilier, Nir Friedman, Moises Goldszmidt, and Daphne Koller. Context-specific independence in Bayesian networks. In Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence, UAI'96, pages 115–123, San Francisco, CA, USA, 1996. Morgan Kaufmann Publishers Inc. ISBN 1-55860-412-X. URL http://dl.acm.org/citation.cfm?id=2074284.2074298.

Federico Castelletti and Guido Consonni. Bayesian inference of causal effects from observational data in Gaussian graphical models. Biometrics, 2020.

J. Correa and E. Bareinboim. General transportability of soft interventions: Completeness results. In NeurIPS, 2020.

Patrick Forré and Joris M. Mooij. Causal calculus in the presence of cycles, latent confounders and selection bias. In Proceedings of the 35th Annual Conference on Uncertainty in Artificial Intelligence (UAI-19), 2019.

Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In UAI, pages 217–224. AUAI Press, 2006.

A. Hyttinen, F. Eberhardt, and M. Järvisalo. Do-calculus when the true graph is unknown. In UAI, pages 395–404. AUAI Press, 2015.

A. Hyttinen, J. Pensar, J. Kontinen, and J. Corander. Structure learning for Bayesian networks over labeled DAGs. In PGM, 2018.

References II

Amin Jaber, Jiji Zhang, and Elias Bareinboim. Causal identification under markov equivalence: Completeness results. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA, volume 97 of Proceedings of Machine Learning Research, pages 2981–2989. PMLR, 2019. URL http://proceedings.mlr.press/v97/jaber19a.html.

Jack Kuipers, Giusi Moffa, Elizabeth Kuipers, Daniel Freeman, and Paul Bebbington. Links between psychotic and neurotic symptoms in the general population: an analysis of longitudinal british national survey data using directed acyclic graphs. Psychological Medicine, 49(3):388–395, 2019. doi: 10.1017/S0033291718000879.

D. Kumor, C. Cinelli, and E. Bareinboim. Efficient identification in linear structural causal models with auxiliary cutsets. In Proceedings of the 37th International Conference on Machine Learning, volume 119. PMLR, 2020.

Matt J. Kusner, Joshua R. Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. In NIPS, 2017.

S. Lee, J. Correa, and E. Bareinboim. General identifiability with arbitrary surrogate experiments. In Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence, 2019.

M. H. Maathuis, M. Kalisch, and P. Bühlmann. Estimating high-dimensional intervention effects from observational data. The Annals of Statistics, 37(6A):3133–3164, 2009.

D. Malinsky and P. Spirtes. Estimating bounds on causal effects in high-dimensional and possibly confounded systems. International Journal of Approximate Reasoning, 88:371–384, 2017.

Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. In The 22nd International Conference on Artificial Intelligence and Statistics, pages 3080–3088, 2019.

K. Mohan, J. Pearl, and J. Tian. Graphical models for inference with missing data. In NIPS, pages 1277–1285, 2013.

Joris M. Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. Journal of Machine Learning Research, 21(99):1–108, 2020. URL http://jmlr.org/papers/v21/17-123.html.

References III

J. Pearl. Causal diagrams for empirical research. Biometrika, 82(4):669–688, 1995.

J. Pearl. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2000.

J. Pearl and D. Mackenzie. The Book of Why. Basic Books, 2018.

Johan Pensar, Henrik J. Nyman, Timo Koski, and Jukka Corander. Labeled directed acyclic graphs: a generalization of context-specific independence in directed graphical models. Data Min. Knowl. Discov., 29(2):503–533, 2015. doi: 10.1007/s10618-014-0355-0. URL http://dx.doi.org/10.1007/s10618-014-0355-0.

Johan Pensar, Topi Talvitie, Antti Hyttinen, and Mikko Koivisto. A Bayesian approach for estimating causal effects from observational data. In AAAI. AAAI Press, 2020.

J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference using invariant prediction: identification and confidence intervals. J. of the Royal Statistical Society, B, 78(5):947–1012, 2016.

I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In Proceedings of the 21st National Conference on Artificial Intelligence – Volume 2, pages 1219–1226. AAAI Press, 2006a.

I. Shpitser and J. Pearl. Identification of conditional interventional distributions. In UAI, pages 437–444. AUAI Press, 2006b.

Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. Journal of Machine Learning Research, 9(64):1941–1979, 2008. URL http://jmlr.org/papers/v9/shpitser08a.html.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In Proc. AAAI, pages 567–573. AAAI Press, 2002.

S. Tikka and J. Karvanen. Surrogate outcomes and transportability. International Journal of Approximate Reasoning, 108:21–37, 2019.

S. Tikka, A. Hyttinen, and J. Karvanen. Identifying causal effects via context-specific independence relations. In NeurIPS, 2019.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Causal effect identification from multiple incomplete data sources. Journal of Statistical Software, 2020.

J. Viinikka, A. Hyttinen, J. Pensar, and M. Koivisto. Towards scalable Bayesian learning of causal DAGs. In NeurIPS, 2020.