UNIVERSITY OF COPENHAGEN

Further perspectives

Anne Helby Petersen

Applications of TPC and TFCI

TPC:

- Petersen, Osler & Ekstrøm (2021): Data-driven model building for life-course epidemiology. *American Journal of Epidemiology*.
- Andersen et al. (2023). Nighttime smartphone use, sleep quality, and mental health: investigating a complex relationship. *Sleep*.
- Foraita et al. (2024): A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents. *Scientific Reports*.

TFCI:

- Lee et al. (2023): Causal determinants of postoperative length of stay in cardiac surgery using causal graphical learning. *The Journal of Thoracic and Cardiovascular Surgery*.

On practical usefulness of causal discovery

- Didelez (2024). Invited Commentary: Where Do the Causal DAGs Come From?. *American Journal of Epidemiology*.
- Petersen, Ekstrøm, Spirtes & Osler (2023). Constructing Causal Life-Course Models: Comparative Study of Data-Driven and Theory-Driven Approaches. *American Journal of Epidemiology*.
- Talk @ EuroCIM: Anne Helby Petersen - "Comparing theory-driven and data-driven approaches to constructing causal models for life course epidemiology" (Thursday)

Other causal discovery algorithms

Today, we have considered two algorithms for constraint-based causal discovery (PC and FCI). Alternative approaches include:

- Score-based methods: Heuristic search through possible (CP)DAGs, score each model according to fit. Examples: GES (Chickering 2002), GRASP (Lam et al. 2022), BOSS (Andrews et al. 2023).
- Optimization-based methods: Try to estimate the global DAG structure at once. Examples: LiNGAM (Shimizu et al. 2006), NOTEARS (Zheng et al. 2018), DAG-GNN (Yu et al. 2019).

Fast-paced research field with many new algorithms being proposed each year.

Finite-sample and realistic real-world data performance receives much less focus.

Additional software for causal discovery

- R packages: `pcalg` (backend for `causalDisco`'s PC + FCI implementations), `bnlearn`, `tpc` (also uses `pcalg` as backend), `micd`
- Python: `causal-learn`
- TETRAD: Point-and-click software (also available in commandline (in Java), or via Python (`py-tetrad`)

Causal inference after causal discovery

Dealing with **non-identified causal models**:

- IDA: Intervention calculus when the DAG is Absent. Bounding causal effects from a CPDAG (Maathuis, Kalisch & Bühlmann 2009) or MPDAG (Fang & He 2020)
- Talk @ EuroCIM 2024: Emilja Perkovic - "Causal effects in a Markov equivalence class: Identification and efficient estimation" (Thursday)

Dealing with **post-selection issues**:

- Estimating model (CPDAG/TMPDAG) *and* estimating causal effects using the same data will generally lead to overfitting.
- Classical machine learning approaches may be used (data splitting, bootstrapping), although the topic is under-studied
- Talk @ EuroCIM 2024: Daniel Malinsky - "Post-selection inference for causal effects after causal discovery" (Thursday)
- Talk @ EuroCIM 2024: David Strieder - "Confidence in Causal Inference under Structure Uncertainty" (Thursday)

More causal discovery at EuroCIM 2024

- Jeroen Uleman: "Evidence triangulation for causal loop diagrams: constructing biopsychosocial models using group model building, literature review, and causal discovery" (Wednesday poster)
- Jesus Renero: "Explainability in Causal Discovery: A Novel Approach for Inferring Causal Graphs from Observational Data" (Wednesday poster)
- Kai Teh: "A general framework for Causal Learning algorithms" (Wednesday poster)
- Jaime Maldonado: "Causal Modeling of Visual Fixations" (Wednesday poster)
- Luca Bergen: "Causal discovery for nonlinear mixed-scale data" (Thursday poster)
- Luka Kovacevic: "PerturbSCM: Benchmarking Causal Structure Learning for Gene Perturbation Screens" (Thursday poster)

Evaluation

Please fill out this short course evaluation:
https://biostatistics.dk/teaching/eurocim/evaluation

References (1/3)

Andersen, Sejling, Jensen, Drews, Ritz, Varga & Rod (2023). Nighttime smartphone use, sleep quality, and mental health: investigating a complex relationship. *Sleep*

Andrews, Ramsey, Romero, Camchong & Kummerfeld (2023). Fast Scalable and Accurate Discovery of DAGs Using the Best Order Score Search and Grow Shrink Trees. *Advances in Neural Information Processing Systems*.

Chickering (2002): Optimal structure identification with greedy search. *Journal of machine learning research*.

Didelez (2024). Invited Commentary: Where Do the Causal DAGs Come From?. *American Journal of Epidemiology*.

Fang & He (2020). IDA with Background Knowledge. In *Proceeding for Uncertainty in Artificial Intelligence.*

Foraita, Witte, Börnhorst, Gwozdz, Pala, Lissner, Lauria, Reisch, Molnár, De Henauw, Moreno, Veidebaum, Tornaritis, Pigeot & Didelez (2024): A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents. *Scientific Reports*.

References (2/3)

Lam, Andrews & Ramsey (2022). Greedy relaxations of the sparsest permutation algorithm. In *Proceedings for Uncertainty in Artificial Intelligence*.

Lee, Srinivasan, Ong, Alejo, Schena, Shpitser, Sussman, Whitman & Malinsky (2023): Causal determinants of postoperative length of stay in cardiac surgery using causal graphical learning. *The Journal of Thoracic and Cardiovascular Surgery*.

Maathuis, Kalisch & Bühlmann (2009). Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*.

Petersen, Ekstrøm, Spirtes & Osler (2023). Constructing Causal Life-Course Models: Comparative Study of Data-Driven and Theory-Driven Approaches. *American Journal of Epidemiology*.

Shimizu, Hoyer, Hyvarine, Kermine & Jordan (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*.

References (3/3)

Yu, Chen, Gao & Yu (2019). DAG-GNN: DAG structure learning with graph neural networks. In *Proceeding of International Conference on Machine Learning*.

Zheng,Aragam, Ravikumar & Xing (2018). DAGs with no tears: Continuous optimization for structure learning. *Advances in neural information processing systems*.