Eurandom

Workshop Centre in the area of Stochastics

Young European Statisticians Series

Causal Inference

March 13-15, 2023, Eindhoven

Organisers
Stéphanie van der Pas (Amsterdam UMC)
Richard Post (Eindhoven University of Technology)
Joris Mooij (University of Amsterdam)

Workshop officer
Marianne de Bruin (Eurandom)

Workshop schedule

Monday

09.00 - 09.30	Registration	
09.30 - 09.45	Opening	
09.45 - 10.45	Rhian Daniel I	
10.45 - 11.00	*BREAK*	
11.00 - 12.00	Rhian Daniel II	
12.00 - 13.30	*LUNCH*	
13.30 - 14.30	Leonard Henckel I	
14.30 - 14.45	*BREAK*	
14.45 - 15.45	Leonard Henckel II	
15.45 - 16.00	*BREAK*	
16.00 - 16.45	Rhian Daniel III	
16.45 - 17.00	*BREAK*	
17.00 - 17.45	Leonard Henckel III	
17.45 - 19.00	Poster session I	

Tuesday

09.30 - 10.30	Mats Stensrud	Optimal decision rules assisted by algorithms
10.30 - 10.45	*BREAK*	
10.45 - 11.15	Matias Janvin	A Positivity Robust Strategy to Study Effects of Switching Treatments Based on an Early Treatment Response
11.15 - 11.45	Mirthe van Diepen	Unraveling the Causal Mechanism Behind Aortic Arch Surgery: a Guide to Causal Discovery for Health
11.45 - 13.30	*LUNCH*	
13.30 - 14.00	Phillip Bach	Practical Aspects of Double/Debiased Machine Learning
14.00 - 14.30	Patrick Klösel	Equivalence, Which Equivalence? The Case of Structural Causal Models and Potential Outcomes
14.30 - 14.45	*BREAK*	
14.45 - 15.45	Sonja Swanson	Nature as a trialist?
15.45 - 16.00	*BREAK*	

16.00 – 16.30	Jens Klooster	Outlier Robust Inference in (Weak) Linear Instrumental Variable Models
16.30 - 17.00	Heather Hufstedler	Quasi-experimental methods with pooled individual-level observational, longitudinal data
17.00 - 18.15	Poster session II	
19.00 - 22.00	Conference dinner (included with registration)	*Gezana (Willemstraat 37, 5611 HC Eindhoven)*

Wednesday

09.30 - 10.30	Sara Magliacane	Causality-inspired ML: what can causality do for ML? The domain adaptation case
10.30 - 10.45	*BREAK*	
10.45 - 11.15	Matej Zecevic	On the Tractability of Inference for the Spectrum of Causal Models
11.15 - 11.45	Zachary Jones	Compact Nonlinear Maps and HSIC regularized regression ameliorates unsupervised covariate shift
11.45 - 12.15	Jakob Zeitler	Fundamentals of Partial Identification
12.15 - 12.30	Closing	
12.30 - 13.30	*LUNCH*	

Tutorials & Keynote talks

An introduction to causal inference using potential outcomes
Rhian Daniel (Cardiff University)

In this tutorial we start by introducing potential outcomes and counterfactuals and how they are used to frame research questions in applied research. We will follow Judea Pearl and others in making a distinction between "what might happen in the future if...?" questions and "what would have happened in the past if...?" questions. We will discuss how these concepts, and their accompanying notations, help to make research questions clearer, and what further aspects they invite us to explore to achieve this clarity. We discuss how the key assumptions needed for making causal inferences are expressed using potential outcomes, and how these in turn lead to statistical causal models and methods for estimating their parameters.

An introduction to causal inference with graphical models
Leonard Henckel (University of Copenhagen)

Questions of cause and effect are central to many research areas. It is therefore important to understand when and how we can use statistics to infer causal effects. Causal graphical models provide a framework in which we can try to answer these two questions in a principled way. In this tutorial, I will provide a basic introduction to causal graphs and how we can use them as a practical tool. Specifically, I will first define causal graphical models and discuss in what sense they go beyond probabilistic graphical models. Second, I will show how we can use causal graphs to identify statistical models that are valid for causal inference with mathematical rigor. Finally, I will discuss how we can sometimes learn causal graphs from observational data and what the practical implications of these results are.

Optimal decision rules assisted by algorithms
Mats Stensrud (École polytechnique fédérale de Lausanne)

Decision makers desire to implement decision rules that, when applied to individuals in the population of interest, yield the best possible outcomes. For example, the current focus on precision medicine reflects the search for individualized decision rules, adapted to a patient's characteristics. In this presentation, I will introduce superoptimal decision rules, which are guaranteed to outperform conventional optimal decision rules. Importantly, identification of superoptimal rules and their values require exactly the same assumptions as identification of conventional optimal rules in several common settings, including common instrumental variable settings. The superoptimal rules can also be identified in data fusion contexts, in which experimental data and (possibly confounded) observational data are available. We study superoptimal rules in two examples that have been presented in the optimal decision rule literature, illustrating that the superoptimal rules perform better than conventional optimal rules.

Nature as a trialist?
Sonja Swanson (University of Pittsburgh)

Mendelian randomization (MR) approaches have been increasingly used to estimate causal effects, with recent years seeing hundreds of applications published annually in the medical literature. MR studies are often described as naturally occurring randomized trials in which genetic variants are randomly assigned by nature. In this talk, we deconstruct this oft-made analogy between MR and randomized trials and describe its implications for the design, conduct, reporting, and interpretation of MR studies. By precisely defining the causal effects being estimated, we underscore that MR estimates are often vaguely analogous to per-protocol effects in randomized trials, and then discuss opportunities and challenges to addressing that vagueness.

Causality-inspired ML: what can causality do for ML? The domain adaptation case
Sara Magliacane (University of Amsterdam)

Applying machine learning to real-world cases often requires methods that are robust w.r.t. heterogeneity, missing not at random or corrupt data, selection bias, non i.i.d. data etc. and that can generalize across different domains. Moreover, many tasks are inherently trying to answer causal questions and gather actionable insights, a task for which correlations are usually not enough. Several of these issues are addressed in the rich causal inference literature. On the other hand, often classical causal inference methods require either a complete knowledge of a causal graph or enough experimental data (interventions) to estimate it accurately. Recently, a new line of research has focused on causality-inspired machine learning, i.e. on the application ideas from causal inference to machine learning methods without necessarily knowing or even trying to estimate the complete causal graph. In this talk, I will present an example of this line of research in the unsupervised domain adaptation case, in which we have labelled data in a set of source domains and unlabeled data in a target domain ("zero-shot"), for which we want to predict the labels. In particular, given certain assumptions, our approach is able to select a set of provably "stable" features (a separating set), for which the generalization error can be bound, even in case of arbitrarily large distribution shifts. As opposed to other works, it also exploits the information in the unlabeled target data, allowing for some unseen shifts w.r.t. to the source domains. While using ideas from causal inference, our method never aims at reconstructing the causal graph or even the Markov equivalence class, showing that causal inference ideas can help machine learning even in this more relaxed setting.

Contributed talks

A Positivity Robust Strategy to Study Effects of Switching Treatments Based on an Early Treatment Response
Matias Janvin (École polytechnique fédérale de Lausanne)

In studies of medical treatments, individuals often experience post-treatment events that predict their future outcomes. In this work, we study how to use initial observations of a recurrent event -- a type of post-treatment event -- to offer updated treatment recommendations in settings where no, or few, individuals are observed to switch between treatment arms. Specifically, we formulate an estimand quantifying the average effect of treatment-switching on subsequent events. Furthermore, we derive sharp bounds on its value under plausible conditions, and provide consistent non-parametric estimators of the bounds. Next, we define a value and regret function for a dynamic treatment-switching regime, and we use these to determine three optimal regimes under partial identification: the pessimist (maximin value), optimist (maximax value) and opportunist (minimax regret). The pessimist regime is guaranteed to perform at least as well as the standard of care. We present simulations to illustrate properties of our proposed method and its relation to previous work. Finally, we apply our methods to data from the Systolic Blood Pressure Intervention Trial.

Unraveling the Causal Mechanism Behind Aortic Arch Surgery: a Guide to Causal Discovery for Health
Mirthe van Diepen (Radboud University)

Understanding the causal relationships between demographic information and biomarkers can be extremely useful to get a better understanding of causal risk factors in healthcare. It can motivate future studies to search for an intervention that lowers the risk or the search for possible treatment alternatives that can improve quality of life expectations. Using random controlled trials (RCTs) we can try to infer specific causal relationships. However, it is not always possible to directly intervene on (proxy) variables due to ethical reasons or it is just impossible in practice. Causal discovery algorithms try to address this problem, by searching for the causal structure between variables in an observational data set instead of using interventions on the variables. However, currently, in medical journals, the methods to analyze data are usually not based on causal discovery methods due to the assumptions made which are difficult to test for, and the non-intuitive definitions which are required for this field. Here we show how to handle these using a specific case study that exhibits many of these challenges. This study is motivated by a data set containing subjects that had aortic surgery at the St. Antonius hospital in Nieuwegein. We use this data set to demonstrate what important steps are needed for the analysis. Challenges of this aortic surgery data set are (1) small sample size, (2) consisting of a complex combination of very different variables, both discrete and continuous, (3) unknown causal structure (there might be unknown confounders or cycles in the causal structure), (4) context variables and time-dependent variables (variables from the different phases in the perioperative period), and (5) missing values. We will show what to consider when choosing a causal discovery method and the impact of different choices for the hyperparameters for it. Moreover, we suggest how one can combine the outputs of a causal discovery method with bootstrapping to make it more robust for small data sets, how to deal with context variables, and how to deal with mixed data.

Practical Aspects of Double/Debiased Machine Learning
Philip Bach (University of Hamburg)

The DML framework by Chernozhukov et al. (2018) is a general approach to perform inference on causal parameters based on machine learning methods. When researchers want to apply the DML approach in empirical studies, they are faced with many practical choices, for example: Which machine learning method should they choose? How should they tune the hyperparameters of the ML learners? They may also wonder which sample splitting schedule and which causal model to use in case multiple options are available. With this project, we want to shed light on practical aspects of Double/Debiased Machine Learning (DML). We perform extensive simulation experiments and evaluate various empirical benchmarks to derive some guidance on the use of DML in practice. First, we briefly review the DML approach with its three key ingredients: (1) Neyman orthogonality, (2) high-quality machine learning estimation and (3) sample splitting. Second, we provide results from an extensive simulation study and compare the performance of different approaches, mainly regarding different sample splitting schemes and hyperparameter optimization techniques. Third, we evaluate the performance of these approaches in semi-synthetic and real benchmark data sets. Finally, we summarize our findings and derive recommendations for the use of DML in practice. We investigate the use of various diagnostics tools that are intended to help applied researchers specify Double Machine Learning models in empirical studies.

Equivalence, Which Equivalence? The Case of Structural Causal Models and Potential Outcomes
Patrick Klösel (Potsdam institute for Climate Impact Research)

The potential outcome framework (Rubin, 1974) and structural causal models (Pearl, 2009) are claimed to be 'logically equivalent' (ibid.). Their (formal) equivalence is usually taken for granted, despite an ongoing debate about the (pragmatic) benefits of either of the two frameworks in empirical sciences like epidemiology or economics. Markus (2021) has recently argued that the two frameworks are only weakly equivalent, albeit without reference to any formal notion of equivalence. The thesis defended in this presentation is that the best available explication of the proof of equivalence of the two frameworks is in terms of 'equivalence as intertranslatability', a well-established and intuitive notion of equivalence that builds on inverse translations (Barrett and Halvorson, 2016). There is a growing literature on theoretical equivalence in the philosophy of physics that has developed and explicated several different conceptions of what 'theoretical equivalence' could mean (for an overview see Weatherall, 2019). Most of the authors start from their intuitions on which physical theories should be regarded as equivalent in some relevant sense. They then develop formal criteria of theoretical equivalence that conform to these intuitions. In arguing for 'equivalence as intertranslatability', I demonstrate why competing formal notions of equivalence, such as logical equivalence, categorical equivalence, and definitional equivalence, are unfit for explicating the equivalence between the potential outcome framework and structural causal models. This presentation extends the literature in two distinct ways. Firstly, the discussion on theoretical equivalence in philosophy of science has mostly focused on examples from physics, with some exceptions; I will focus on two frameworks from the causal inference literature. The second contribution of this paper is a service to the empirical sciences engaged in causal inference. In clarifying the notion of equivalence relevant for this methodological discussion, I hope to assist empirical researchers in choosing the framework that best fits their research agenda.

Outlier Robust Inference in (Weak) Linear Instrumental Variable Models
Jens Klooster (Erasmus University Rotterdam)

We propose a general robust framework to construct weak instrument robust testing procedures that are also robust to outliers in the linear instrumental variable model. The framework is constructed upon M-estimators and we show that classical weak instrument robust tests, such as the Anderson and Rubin (1949) test and the Moreira (2003) conditional likelihood ratio (CLR) test can be obtained by specifying the M-estimators to be the Least Squares estimators. As it turns out that the classical testing procedures are not robust to outliers, we show how to construct robust alternatives. In particular, we show how to construct a robust CLR statistic based on Mallows type M-estimators and show that its asymptotic distribution is the same as the (classical) CLR statistic. The theoretical results are corroborated by a simulation study. Finally, we revisit three empirical studies affected by outliers and apply the robust CLR test to re-evaluate their results.

Quasi-experimental methods with pooled individual-level observational, longitudinal data
Heather Hufstedler (University Hospital Heidelberg)

Difference-in-difference (DiD), or marginal structural models (MSM), e.g., are potentially much more effective at dealing with time-varying confounding and controlling for measured and unmeasured confounding than standard statistical methods or regression-based adjustments. Unfortunately, according to our findings in two recent systematic reviews, such causal inference methods are not widely implemented by infectious disease researchers who pool individual-level data, despite their utility. We apply some of these methods to newly assembled pooled individual-level patient data collected during the 2014-2016 Ebola outbreak in West Africa, and discuss challenges and opportunities that the data presents. The focus of our investigation is to highlight viable sources of exogenous variation for identifying the causal effect of treatment on patient survival.

On the Tractability of Inference for the Spectrum of Causal Models
Matej Zecevic (TU Darmstadt)

Neurally-parameterized Structural Causal Models in the Pearlian notion to causality, referred to as NCM, were recently introduced as a step to- wards next-generation learning systems. However, said NCM are only concerned with the learning aspect of causal inference but totally miss out on the architecture aspect. That is, actual causal inference within NCM is intractable in that the NCM won't return an answer to a query in polynomial time. This insight follows as corollary to the more general statement on the intractability of arbitrary SCM parameterizations, which we prove in this work through classical 3-SAT reduction. Since future learning algorithms will be required to deal with both high dimensional data and highly complex mechanisms governing the data, we ultimately believe work on tractable inference for causality to be decisive. To this end, we further define a spectrum of causal models to classify and provide a perspective from standard correlation-based models up to Pearlian SCM. By investigating specific representative models within said spectrum we recover discriminative criteria in terms of their tractability properties that become neatly representable in tabular form. To conclude our work, we also provide ideas on how to cope or even overcome intractability of causal inference. We also propose a new model with which we further provide an empirical highlight onto the importance of tractability and an idea on the coping aspect when facing intractability.

Compact Nonlinear Maps and HSIC regularized regression ameliorates unsupervised covariate shift
Zachary Jones (University of Hamburg)

It is desirable and a sign of overall robustness against unsupervised covariate shift that the residual error of a model be independent of the distribution of the input data. Approaches to incorporating causal ideas of independence have been explored using the Hilbert-Schmidt Independence Criteria between the data distribution and the model residuals as an objective function using linear models. This leaves two critical aspects to be addressed: the choice of kernel embedding and broadening the model class beyond the linear scope. This work aims to address both of these problems by reformulating the learning problem as a constrained nonconvex loss minimization problem using compact nonlinear maps to both learn a data dependent kernel and approximate a support vector machine regressor. Empirical results on simulated DAG data using different distributions for data generation and testing illustrate a significant improvement over classical models on out of sample prediction accuracy.

Fundamentals of Partial Identification

Jakob Zeitler (University College London)

Causal inference provides the fundamental causal reasoning that machine learning is missing to effectively tackle decision making problems. So far, full identification of causal effects has been the focus of the majority of research: Strong and mostly untestable assumptions, such as no unmeasured confounding, yield point estimates of how a sprint will increase my endurance by 2% or how $10k more in savings will get my loan application accepted. Ideally, we would want to make fewer strong assumptions, but still provide informative suggestions. Partial identification enables this by calculating lower and upper bounds on the true causal effect which are more trustworthy due to more realistic assumptions. Unfortunately, current partial identification methods practically do not scale due to super-exponential parameter growth in the number of variables. Hence, I am developing scalable methods that trade-off computational cost with tightness of bounds. Exact bounding approaches will be crucial to high-stake decision making problems such as AI fairness, which require provable guarantees. Approximate methods will find use in environments valuing execution cost over guarantees, such as personal exercise recommendations or prioritisation of user experience experiments. Both approaches will become fundamental building blocks for trustworthy, causal machine learning.

Poster sessions

Day 1

Lingjie Shen (Tilburg University/ IKNL)	RCTrep: An R package for validation of estimators for conditional average treatment effect
Nils Sturma (TU Munich)	Parameter Identifiability in Latent Variable Models
Marcel Wienöbst (University of Lübeck)	Algorithms for Markov Equivalent DAGs
Jiawei Zhang (University of copenhagen)	Long-term exposure to air pollution and morbidity and mortality from COVID-19: a causal modeling approach
Zhigao Guo (University of Manchester)	Causal Assistant: Human-in-Loop
Caglar Hizli (Aalto University)	Joint Point Process Model for Counterfactual Treatment-Outcome Trajectories Under Policy Interventions
Lincen Yang (Leiden University)	Estimating Conditional Mutual Information for Discrete-Continuous Mixtures using Multi-Dimensional Adaptive Histograms
Rickard Karlsson (TU Delft)	Detecting hidden confounding in observational data using multiple environments

Day 2

Annet Dijkzeul (Erasmus MC)	Using inverse probability weighting to address potential selection bias in studying the relation between ADHD symptoms and brain structure in the general population
Fan Feng (University of Amsterdam)	Factored Adaptation for Non-Stationary Reinforcement Learning
Lorenzo Gasparollo (EPFL)	Population surveillance parameters defined by counterfactual treatment regimes: identification and estimation
Florian Busch (TU Darmstadt)	Causality in Sum-Product Networks
Alexander Mey (TU Eindhoven)	Causal Discovery in Time Series Data Using Causally Invariant Locally Linear Models
Máté Kormos (TU Delft)	Asymptotics of Caliper Matching for Average Treatment Effects
Oliver Schacht (University of Hamburg)	Practical Aspects of Double/Debiased Machine Learning - Insights from an Extensive Simulation Study
Daniele Tramontano (TU Munich)	Learning Linear Gaussian Polytree Models with Interventions
Amit Sawant *(EPFL)*	A Nationwide Lockdown and Deaths due to COVID-19 in the Indian Subcontinent