EUROPEAN CAUSAL INFERENCE MEETING (EuroCIM - 2021)

Wednesday 19th May 2021

Theme: Methods, Hosted by the Centre for Statistical Methodology at LSHTM Free online event on Zoom;

Time	Session / talk information
13h -14h10 BST	Session 1 Welcome to Day 1 Karla DiazOrdaz **Short-talks Session** (7 min talks + joint Q&A at the end) Chairs: Clemence Leyrat and Emily Granger
	Margarita Moreno-Betancur, University of Melbourne, Australia *Data-adaptive methods for high-dimensional mediation analysis*
	Apostolos Gkatzionis. Bristol University, UK *Using instruments for selection to adjust for selection bias in instrumental variable analysis*
	Jouni Heske, University of Jyväskylä, Finland *Estimation of causal effects with small data in the presence of trapdoor variables*
	Rohit Bhattacharya, Johns Hopkins University *Semiparametric Inference For Causal Effects In Graphical Models With Hidden Variables*
	12 min Q&A for first 3 speakers
	Daniel Rodriguez Duque, McGill University Canada *Estimation of Optimal Dynamic Treatment Regimes via Gaussian Process Emulation*
	Falco Bargagli-Stoffi, Harvard University USA *Causal Rule Ensemble: Interpretable Inference of Heterogeneous Treatment Effects*
	Michael Knaus, University of St. Gallen, Switzerland *Decomposing Causal Effect Heterogeneity under Multiple Treatment Versions*
	10 min Q&A for last 3 speakers
14:10 – 14:25	15 min break
14:25 – 15:45	Session 2, Contributed talks (20 min each including Q&A) Chair Ruth Keogh
14:25 –14-45	**Oliver Hines**, London School of Hygiene and Tropical Medicine, UK *Parameterising and inferring the effect of a continuous exposure using average derivative effects*
14:45 –15-05	**Ivana Malenica**, University of California, Berkeley, USA *Adaptive Sequential Design for a Single Time-Series*
15:05 –15-25	**Nima Hejazi**, University of California, Berkeley, USA *Nonparametric estimation of the generalized propensity score based on the highly adaptive lasso*
15:25 –15-45	**Daniel Scharfstein**, University of Utah, USA *Semiparametric Sensitivity Analysis: Unmeasured Confounding in Observational Studies*
15:45 – 16:00	15 min break
16:00 – 16:55	Session 3: Chair Stijn Vansteelandt Keynote Speaker **Edward Kennedy** *Infinite-dimensional parameter estimation in causal inference*

Abstracts

Margarita Moreno-Betancur; Nicole L Messina; Kaya Gardiner; Nigel Curtis; Stijn Vansteelandt
Data-adaptive methods for high-dimensional mediation analysis: Application to a randomised trial of tuberculosis vaccination

Statistical methods for causal mediation analysis are useful for understanding the pathways by which a treatment or exposure impacts health outcomes. While there have been many methodological developments in the past decades, there is still a scarcity of feasible and flexible, data-adaptive methods for mediation analysis with respect to high-dimensional mediators (e.g., biomarkers) and confounders. Existing methods necessitate modelling of the distribution of the mediators, which quickly becomes infeasible when mediators are high-dimensional. To avoid such high-dimensional modelling, we propose causal machine learning methods for estimating the indirect effect of a randomised treatment that acts via a pathway represented by a high-dimensional set of measurements. The proposed methods are doubly robust, enabling (uniformly) valid statistical inference when using machine learning algorithms for the two required models. This work was motivated by the Melbourne Infant Study: BCG for Allergy and Infection Reduction (MIS BAIR), a randomised controlled trial investigating the effect of neonatal Bacillus Calmette Guerin (BCG) (tuberculosis) vaccination on allergy and infection outcomes in the first years of life. The hypothesis of the trial was that the heterologous effects of BCG on innate immunity have beneficial effects on the developing immune system, resulting in improved outcomes. We illustrate the methods in the investigation of this hypothesis, where immune pathways are represented by a high-dimensional vector of cytokine responses under various stimulants. We moreover study the performance of the methods in an extensive simulation study that is closely based on this example to empirically evaluate the methods in a realistic setting.

Apostolos Gkatzionis, Kate Tilling
Using instruments for selection to adjust for selection bias in instrumental variable analysis

Selection bias (due to participation, dropout or missing data) can bias causal analyses. Common approaches to address selection, such as inverse probability weighting, usually assume that data are missing at random (MAR) and can yield biased results if this assumption is violated. In observational studies with a non-MAR outcome, an instrumental variable for missingness approach has been proposed to adjust for selection bias. An instrumental variable for missingness should affect selection into the study, but should not affect the outcome, conditional on observed covariates. After reviewing the approach for observational studies, we extend it to instrumental variable analyses using one or more instruments for inference and a separate instrument for selection. This includes Mendelian randomization analyses using individual-level data, with selection affected by either the risk factor or outcome or both. We review identification conditions and describe how to obtain missingness-adjusted two-stage least squares estimates in this setting. The proposed method is then evaluated in a simulation study, with results suggesting that it can mitigate selection bias but may yield parameter estimates with large standard errors when the instrument for selection is weak.

Jouni Helske; Santtu Tikka; Juha Karvanen
Estimation of causal effects with small data in the presence of trapdoor variables

We consider the problem of estimating causal effects of interventions from observational data when well-known back-door and front-door adjustments are not applicable. We show that when an identifiable causal effect is subject to an implicit functional constraint that is not deducible from conditional independence relations, the estimator of the causal effect can exhibit bias in small samples (where parameter estimation exhibits non-negligible uncertainty). This bias is related to variables that we call trapdoor variables. We use simulated data to study different strategies to account for trapdoor variables and suggest how the related trapdoor bias might be

minimized. The importance of trapdoor variables in causal effect estimation is illustrated with real data from the Life Course 1971-2002 study. Using this dataset, we estimate the causal effect of education on income in the Finnish context. Using the Bayesian modelling approach allows us to take the parameter uncertainty into account and gives us the full interventional distribution instead of only average causal effect estimates.

Rohit Bhattacharya; Razieh Nabi; Ilya Shpitser
Semiparametric Inference For Causal Effects In Graphical Models With Hidden Variables

The last decade witnessed the development of algorithms that completely solve the identifiability problem for causal effects in hidden variable causal models associated with directed acyclic graphs (DAGs.) However, much of this machinery remains underused in practice, owing to the complexity of estimating the identifying functionals yielded by these algorithms. In this work, we describe simple graphical criteria and semiparametric estimators that bridge the gap between identification and estimation for causal effects involving a single treatment and a single outcome. We first provide influence function based doubly robust estimators that cover a broad class of causal DAGs with unmeasured confounders where the effect is identifiable. We then provide an algorithm that can be used to determine precisely when the hidden variable causal DAG imposes no restrictions on the observed data distribution, implying that this doubly robust estimator is also the most efficient one in the class of regular and asymptotically linear estimators that we considered. For an important subset of hidden variable causal DAGs that do impose restrictions on the observed data distribution, we describe how these restrictions can be incorporated to derive asymptotically efficient estimators. These results allow us to extend the use of flexible machine learning methods for causal effect estimation to a wide variety of settings where the conditional ignorability assumption does not hold. The implementation of these semiparametric estimators are available through a Python software package called Ananke.

Daniel Rodriguez Duque; Erica E.M. Moodie; David A. Stephens
Estimation of Optimal Dynamic Treatment Regimes via Gaussian Process Emulation

In precision medicine, researchers often aim to infer about dynamic treatment regimes (DTRs) and to identify optimal DTRs. Conventional methods of inference involve sophisticated semi-parametric estimators. However, these are not without their strong assumptions. Dynamic Marginal Structural Models (MSMs) are one semi-parametric approach used to infer about optimal DTRs in a family of regimes. To achieve this, investigators are forced to model the expected outcome under adherence to a DTR in the family; reasonable models may still lead to bias in the optimum. One way to obviate this difficulty is to use estimators for the value of a DTR and to perform a grid search for the optimum. Unfortunately, this approach can be computationally prohibitive as the complexity of regimes increases. In recently developed Bayesian methods for Dynamic MSMs, computational challenges compound because at each grid point, a posterior mean must be calculated. We propose a manner by which to alleviate modelling detriments by using Gaussian process optimization. This optimization approach emulates expensive-to-evaluate surfaces while limiting the number of evaluation points. We pair this optimization technique with estimators for the value of a DTR to identify optimal DTRs, including in multi-modal settings. Furthermore, the surfaces we emulate exhibit a noisy quality, as they are the result of point-wise evaluations of estimators. We examine possible sources of this variation, establish that it may be heteroskedastic, and present a modelling strategy that accounts for these features.

Falco J. Bargagli Stoffi; Kwonsang Lee; Francesca Dominici
Causal Rule Ensemble: Interpretable Inference of Heterogeneous Treatment Effects

In social and health sciences, it is critically important to identify subgroups of the study population where a treatment (or exposure) has a larger or smaller causal effect on an outcome compared to

the population average. In recent years, there have been many methodological developments for addressing heterogeneity of causal effects. A common approach is to estimate the conditional average treatment effect (CATE) given a pre-specified set of covariates. However, this approach does not allow to discover new subgroups, but only to estimate causal effects on subgroups that have been specified a priori by the researchers. Recent causal machine learning (ML) approaches estimate the CATE at an individual level in presence of large number of observations and covariates, with great accuracy. However, because of their complex parametrization of the feature space, these ML approaches do not provide an interpretable characterization of the heterogeneous subgroups. In this paper, we propose a new causal rule ensemble (CRE) method that: 1) discovers de novo subgroups (i.e., causal rules) with heterogeneous treatment effects; 2) ensures interpretability of these subgroups because they are defined in terms of decision rules; and 3) estimates the CATE for each of these newly discovered subgroups with small bias and high statistical precision. We provide theoretical results that guarantee consistency of the estimated causal effects for the newly discovered causal rules. A nice feature of CRE is that it is agnostic to the choices of (i) the ML algorithms that can be used to discover the causal rules, and (ii) the estimation methods for the causal effects with the discovered rules. Via simulations, we show that the CRE method has competitive performance as compared to existing approaches while providing enhanced interpretability. We also introduce a new sensitivity analysis to unmeasured confounding bias. We apply the CRE method to discover subgroups that are more vulnerable (or resilient) to the causal effects of long-term exposure to air pollution on mortality.

Phillip Heiler; **Michael Knaus**
Decomposing Causal Effect Heterogeneity under Multiple Treatment Versions

This paper develops a method to decompose treatment effect heterogeneity when the treatment is not homogeneous and can have multiple versions. It disentangles observed aggregated treatment effect heterogeneity into true effect heterogeneity and heterogeneity due to selection into versions. This allows (i) to avoid spurious discovery of heterogeneous effects, (ii) to detect actual hidden heterogeneity in versions, and (iii) to evaluate the underlying version assignment mechanism.
We propose a semiparametric method for estimation and statistical inference for the decomposition parameters. Our framework allows for the use of modern machine learning techniques in the estimation of the underlying causal effects. It can be used to conduct simple joint hypothesis tests that consider all treatment versions simultaneously. This alleviates the need for multiple testing procedures when deciding on the aggregation level of the treatment variable in empirical applications.
We analyze heterogeneity due to different types of academic or vocational training in the large scale training program for the disadvantaged youth Job Corps. We find that often curricula are not better allocated than random and only specific age and income groups benefit from the actual allocation.

Oliver Hines, Stijn Vansteelandt; Karla Diaz-Ordaz
Parameterising and inferring the effect of a continuous exposure using average derivative effects

The causal inference literature has increased awareness of the importance of non-parametrically defining the causal effect measure of interest prior to any modelling of the data. This is relatively easily done when the exposure is binary, but much less straightforward when continuous exposures are considered. With clear interventions in mind, one may focus on estimands that express the effect of shifting the exposure distribution in a certain, pre-specified way. However, in an exploratory phase of the study, or when shift interventions are not immediately intended to be rolled out, there is a need for estimands that capture the effect of exposure on outcome in a more generic way. In this talk, we therefore propose to summarise conditional exposure effects based on weighted average derivative effects. These express the effect of a small change in the exposure

and, depending on the chosen weights, encompass several estimands that have been previously proposed in the literature. The proposed estimands have several appealing properties when the conditional mean outcome is smooth in the exposure (in the sense of being differentiable with respect to the exposure): they can be causally expressed in terms of counterfactuals when standard exchangeability assumptions are met and can moreover be linked to parameters indexing a novel class of measured Taylor series expansion models. However, they remain well-defined when the conditional mean outcome lacks sufficient smoothness. We develop data-adaptive inference for the considered estimands by deriving their efficient influence function and using it as the basis for a one-step estimator. We moreover show that the recently proposed R-learner delivers an estimator of an (optimal) exposure effect in our class. Simulation studies confirm the desirable performance of our proposal in finite samples.

Ivana Malenica; Mark van der Laan
Adaptive Sequential Design for a Single Time-Series

The current work is motivated by the need for robust statistical methods for precision medicine; as such, we address the need for statistical methods that provide actionable inference for a single unit at any point in time. We aim to learn an optimal, unknown choice of the controlled components of the design in order to optimize the expected outcome; with that, we adapt the randomization mechanism for future time-point experiments based on the data collected on the individual over time. Our results demonstrate that one can learn the optimal rule based on a single sample, and thereby adjust the design at any point t with valid inference for the mean target parameter. This work provides several contributions to the field of statistical precision medicine. First, we define a general class of averages of conditional causal parameters defined by the current context for the single unit time-series data. We define a nonparametric model for the probability distribution of the time-series under few assumptions, and aim to fully utilize the sequential randomization in the estimation procedure via the double robust structure of the efficient influence curve of the proposed target parameter. We present multiple exploration-exploitation strategies for assigning treatment, and methods for estimating the optimal rule. Lastly, we present the study of the data-adaptive inference on the mean under the optimal treatment rule, where the target parameter adapts over time in response to the observed context of the individual. Our target parameter is pathwise differentiable with an efficient influence function that is doubly robust – which makes it easier to estimate than previously proposed variations. We characterize the limit distribution of our estimator under a Donsker condition expressed in terms of a notion of bracketing entropy adapted to martingale settings.

Nima S. Hejazi, David C. Benkeser, Ivan Diaz, Mark J. van der Laan
Nonparametric estimation of the generalized propensity score based on the highly adaptive lasso

Continuous treatment variables have posed a significant challenge for causal inference, both in the formulation and identification of causal effects and in their robust estimation. Traditionally, focus has been placed on techniques applicable to binary or categorical treatments with few levels, settings allowing the application of propensity score-based methodology with relative ease. Efforts to accommodate continuous treatments introduced the generalized propensity score (the conditional density of treatment given covariates), a nuisance parameter required for the estimation of scientifically informative parameters like the causal dose-response curve and the causal effects of stochastic interventions that shift the value of treatment received (modified treatment policies). Unfortunately, the vast majority of generalized propensity score estimators impose restrictive 5odelling assumptions, sharply limiting the real-world applicability of classical and doubly robust estimators alike. We present several novel estimators of the generalized propensity score, all based on the highly adaptive lasso, a recently developed nonparametric regression function demonstrated to achieve a convergence rate suitably fast for the formulation of functional parameter estimators with desirable properties, including asymptotic linearity and

variance converging to the nonparametric efficiency bound. Using a class of causal effect estimands tailored to modified treatment policies, we demonstrate the construction of nonparametric-efficient inverse probability weighted estimators in which the highly adaptive lasso generalized propensity score estimator is undersmoothed. Through numerical experiments, we compare the relative performance of our efficient inverse probability weighted estimators to variants of doubly robust estimators utilizing either our nonparametric generalized propensity score estimators or adaptations of common but restrictive semiparametric alternatives.

Daniel Scharfstein; Razieh Nabi; Edward Kennedy; Ming-Yueh Huang; Matteo Bonvini; Marcela Smid

Semiparametric Sensitivity Analysis: Unmeasured Confounding in Observational Studies

Establishing cause-effect relationships from observational data often relies on untestable assumptions. It is crucial to know whether, and to what extent, the conclusions drawn from non-experimental studies are robust to potential unmeasured confounding. In this paper, we focus on the average causal effect (ACE) as our target of inference. We build on the work of Franks et al. (2019) and Robins (2000) by specifying non-identified sensitivity parameters that govern a contrast between the conditional (on measured covariates) distributions of the outcome under treatment (control) between treated and untreated individuals. We use semiparametric theory to derive the non-parametric efficient influence function of the ACE, for fixed sensitivity parameters. We use this influence function to construct a one-step bias-corrected estimator of the ACE. Our estimator depends on semiparametric models for the distribution of the observed data; importantly, these models do not impose any restrictions on the values of sensitivity analysis parameters. We establish sufficient conditions ensuring that our estimator has root-n asymptotics. We use our methodology to evaluate the causal effect of smoking during pregnancy on birth weight. We also evaluate the performance of estimation procedure in a simulation study.

Keynote Speaker Edward Kennedy

Infinite-dimensional parameter estimation in causal inference

Standard target parameters in causal inference are typically finite-dimensional - for example, the average treatment effect is a single number representing the difference in mean outcomes if all versus none were treated. For these kinds of parameters, now-standard tools from semiparametric theory can be used to characterize minimax optimality and construct nonparametric efficient estimators. However, many targets of interest are in fact infinite-dimensional, e.g., curves living in an infinite-dimensional function space. Some common examples include (i) effects of continuous treatments (e.g., dose response curves), (ii) heterogeneous effects (e.g., conditional treatment effects), and (iii) counterfactual densities. Optimality for these kinds of targets is not well understood, as parametric root-n rates are unachievable even if all potential outcomes were observed, and estimator construction can be highly nontrivial. In this talk I detail some recent results in the three aforementioned examples. Discussion will center on: estimating the target itself versus a projection, nonparametric efficiency and optimality, doubly robust estimation, and applications in public policy.