VVSOR Conference 2020

Causality & Complexity

&

Annual Meeting of the Netherlands Society for Statistics and Operations Research (VVSOR)

Thursday March 12, 2020

LOCATION

Gertrudiskapel – In de Driehoek congres- en vergadercentrum

Willemsplantsoen 1 C, 3511 LA Utrecht

Letter from the President

As has become usual over the last years, I direct myself to the VVSOR members in the first *STAtOR* issue of the new year. The main reason to do so is to make publicity for the annual meeting of the VVSOR and to inform our members about what the board of the VVSOR has in mind on how to maintain and develop our society.

The central theme for an VVSOR annual meeting should comply with two requirements. Firstly, the theme should be of interest to the statistical and OR community for being a hot research topic. Secondly, the theme should illustrate the impact of statistics and OR on science, technology and society. In that spirit, we had 'Climate Change' as central theme in 2018 and 'Big Data & Privacy' in 2019. For 2020, a special committee within the VVSOR has been created that fully dedicates itself to the organisation of the annual meeting. This committee chose the topic 'Causality and Complexity'.

VVSOR Conference 2020 – Causality & Complexity

The following eminent researchers have agreed to speak on this topic: Marloes Maathuis, Stijn Vansteelandt, Joris Mooij, Rick Quax and Jonas Peters. Their talks look at various aspects of causal modelling and modelling of complex systems. Adjustment for confounders is a major issue in causal modelling. Both Marloes Maathuis and Stijn Vansteelandt will elaborate their vision. Joris Mooij will address a new constraint-based method of causal discovery from multiple data sets. Rick Quax will look at causal interpretations of information flows among stochastic dynamical variables. Jonas Peters will introduce a new technique called Anchor Regression that presents a trade-off between predictive and causal modelling. For more detail on the scientific program, I refer to the abstracts that are published elsewhere in this issue of *STAtOR* and to the VVSOR webpages

To me, the scientific program for the VVSOR annual meeting looks very exciting. Beyond the scientific program, the annual meeting also comprises the announcement of the winners of the best MSc thesis (Jan Hemelrijk Award), PhD thesis (Willem R. van Zwet Award) in statistics and operations research. The Hemelrijk and Van Zwet awards are annual awards. This year, we will also present the winner of the five-yearly Van Dantzig Award, an award for outstanding work during the career of a young statistician or operations research scientist.

Agenda points for Algemene Ledenvergadering

Around noon, we will celebrate our annual members' meeting (Algemene Ledenvergadering). There are some serious issues to be discussed. Last year, the board has created a discussion group to think about the future structure of the VVSOR. Some of the ideas that have turned up for restructuring our society will be discussed with the members. And just as we did last year, we again ask for volunteers to assist the VVSOR board in elaborating new policies and ideas to keep our society viable. The areas in which we need new people remain organisation of the annual meeting, communication, and information on statistics and operations research to journalists and media. The committees on Education and on Certification (re-)started their activities after last year's appeal for members. The role of the Sections in central activities, like organizing the Annual Meeting, is another topic that deserves attention.

Last year I also asked for VVSOR members to reflect on a publication strategy for *Statistica Neerlandica*. This remains an issue that requires further input. Good news is that Miroslav Ristic has joined the chief editorial team, replacing Eric Cator (thanks Eric for all the good work). He runs the journal now together with Marijtje van Duijn and Nan van Geloven. Furthermore, Wiley has presented new data on *Statistica Neerlandica* that show that its impact has increased considerably.

Recently an important discussion took place around the plans on the intended restructuring of the curriculum of basic and secondary schools and the place of Statistics in that curriculum. Several members of our society participated very intensively in that discussion and their resulting recommendations seem to have had a considerable influence on the final plans.

We have a very attractive scientific program for the annual meeting on the 12th of March in Utrecht, therefore we count on your presence to discuss issues related to our society. And of course, the 12th of March will also provide an excellent opportunity to meet your fellow society members. Therefore, I am confident to see many of you in In de Driehoek in Utrecht.

Fred van Eeuwijk, President VVSOR

VVSOR Conference 2020

Causality & Complexity

DATE
Thursday, March 12, 2020

VENUE
Gertrudiskapel – In de Driehoek congres- en vergadercentrum, Willemsplantsoen 1C, 3511 LA Utrecht (very close to Utrecht CS).

REGISTRATION
Registration for the conference is mandatory at www.vvsor.nl/articles/vvsor-annual-meeting. Detailed information can be found on our website.

LANGUAGE
The talks at the annual meeting will be in English, the Annual General Meeting (ALV) will be in Dutch.

ALGEMENE LEDENVERGADERING (ALV)
The Annual General Meeting (ALV) takes place on March 12, 12:15 – 13:00. The relevant documents will be provided on the website two weeks before the meeting. You can also get them by e-mail if you send a request to <admin@vvsor.nl>.

LUNCH, COFFEE, TEA AND DRINKS
Lunch, coffee and tea during the breaks and drinks afterwards are offered by the Society.

DINNER
Dinner will take place on March 12, 19:00 – 21:00 at xxxxxxxx

ORGANIZING COMMITTEE
The annual meeting is organized by a special committee in cooperation with the board of the VVSOR. For questions, contact the administration by e-mail at <admin@vvsor.nl>.

REGISTRATION BEFORE MARCH 10!

PROGRAMME

09:15 – 09:45 Registration and coffee/tea

09:45 – 09:50 **Welcome** by Fred van Eeuwijk, President of the VVSOR

09:50 – 10:35 **Causality and Covariate Adjustment**
Marloes Maathuis, Eidgenössische Technische Hochschule (ETH), Zürich

10:35 – 11:20 **Machine Learning for the Evaluation of Treatment Effects: Challenges, Solutions and Improvements**
Stijn Vansteelandt, Ghent University & London School of Hygiene and Tropical Medicine

11:20 – 11:40 Break

11:20 – 12:15 **Ceremony of the Willem R. van Zwet Award, Jan Hemelrijk Award & Van Dantzig Award**
Prize winners will be presented by the juries, followed by a short presentation by the laureates

12:15 – 13:00 **Annual General Meeting (ALV)**

13:00 – 13:45 Lunch

13:45 – 14:30 **How to Learn Causal Relations from Data?**
Joris Mooij, University of Amsterdam

14:30 – 15:15 **Defining and Discovering (Causal) Information Flows: Nudge Causality**
Rick Quax, University of Amsterdam

15:15 – 15:45 Break

15:45 – 16:30 **Causality and Robust Prediction**
Jonas Peters, University of Copenhagen

16:30 – 18:00 Snacks and Drinks

19:00 – 21:00 **Dinner**
At The Colour Kitchen (Oudegracht 214, 3511 NS Utrecht)

09:50 – 10:35

CAUSALITY AND COVARIATE ADJUSTMENT

Marloes Maathuis
Eidgenössische Technische Hochschule (ETH), Zürich

Covariate adjustment is commonly used in practice to adjust for confounders. In this talk, we argue that the covariate set should be chosen with care. We will discuss graphical criteria that identify covariate sets that give consistent and efficient estimates of total effects in causal linear models. We will illustrate the concepts in examples and discuss various generalizations.

Marloes Maathuis is Professor of Statistics at the Seminar for Statistics at ETH Zürich. She studied Applied Mathematics at Delft University of Technology and obtained a PhD in Statistics from the University of Washington, Seattle. Her research interests include causality, graphical models, high-dimensional models, and applications of statistics. In 2004 she was the recipient of the VVSOR prize for the best MSc thesis. Since 2014 this prize features as the Jan Hemelrijk Award. For more information, please see https://stat.ethz.ch/~maathuis/.

10:35 – 11:20

MACHINE LEARNING FOR THE EVALUATION OF TREATMENT EFFECTS: CHALLENGES, SOLUTIONS AND IMPROVEMENTS

Stijn Vansteelandt
Ghent University & London School of Hygiene and Tropical Medicine

The evaluation of treatment effects from observational studies typically requires adjustment for high-dimensional confounding. This is the result of a lack of comparability between treated and untreated subjects in possibly many (pre-treatment) factors that are also related to outcome. While such adjustment is routinely achieved via parametric modelling, it is not entirely satisfactory as model misspecification is likely, and even relatively minor misspecifications over the observed data range may induce large bias in the treatment effect estimate. Over the past 2 decades, there has therefore been growing interest in the use of machine learning methods to assist this task. This is not surprising if one considers the enormous contributions that the machine learning literature has offered on how to predict outcomes based on possibly high-dimensional predictors or features. In this talk, I will therefore focus on the use of machine learning for the evaluation of (causal) treatment effects. This turns out to be a challenging task: while the prediction performance of a given machine learning algorithm can be measured by contrasting observed and predicted outcomes, such evaluation becomes impossible when machine learning is used for treatment effect estimation since the 'true' treatment effect is always unknown. In this talk, I will demonstrate that naive use of existing machine learning algorithms is problematic for treatment evaluation and explain why that is the case. I will next give a gentle introduction to pioneering work on Targeted Learning and on Double Machine Learning, and will discuss improvements that we have made to these techniques. Throughout the talk, 'machine learning' will be considered in the broad sense as any algorithm that uses data to 'learn' a proper model for the data, thus including (though not being limited to) routine variable

selection procedures. The talk is based on joint work with Oliver Dukes (Ghent University) and will be accessible to attendees without a detailed understanding of machine learning algorithms.

Stijn Vansteelandt is Professor of Statistics at Ghent University (Belgium) and Professor of Statistical Methodology at the London School of Hygiene and Tropical Medicine (UK). As a causal inference expert, he primarily develops methods for causal machine learning, mediation analysis, time-varying confounding control, and for handling intercurrent events in randomised experiments. He has authored over 150 peer-reviewed publications in international journals on a variety of topics in biostatistics, epidemiology and medicine, such as the analysis of longitudinal and clustered data, missing data, mediation and moderation/interaction, instrumental variables, family-based genetic association studies, analysis of outcome-dependent samples, phylogenetic inference, meta-analysis, post-selection inference and interim analysis. He is currently Associate Editor of the Journal of the Royal Statistical Society (Series B) and has previously served as Co-Editor of Biometrics, the leading flagship journal of the International Biometrics Society, and as Associate Editor for the journals Biometrics, Biostatistics, Epidemiology, Epidemiologic Methods and the Journal of Causal Inference.

13:45 – 14:30

HOW TO LEARN CAUSAL RELATIONS FROM DATA?

Joris Mooij
University of Amsterdam

Many questions in science, policy making and everyday life are of a causal nature: how would a change of A affect B? Causal inference, a branch of statistics and machine learning, studies how cause-effect relationships can be discovered from data and how these can be used for making predictions in situations where a system has been perturbed by an external intervention. In this talk, I will introduce the basics of two, apparently quite different, approaches to causal discovery. I will discuss how both approaches can be elegantly combined in Joint Causal Inference (JCI), a novel constraint-based approach to causal discovery from multiple data sets. This approach leads to a significant increase in the accuracy and identifiability of the predicted causal relations. One of the remaining big challenges is how to scale up the current algorithms such that large-scale causal discovery becomes feasible.

Joris Mooij is Associate Professor at the University of Amsterdam, the Netherlands. He studied mathematics and physics and received his PhD degree with honours from the Radboud University Nijmegen, the Netherlands, in 2007 on the topic of approximate inference in graphical models. Afterwards, he was a research scientist at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. In the following years, he has obtained an NWO VENI grant, an NWO VIDI grant and an ERC Starting grant on several topics in the area of causal inference. The research topics addressed by his group span the entire spectrum from causal modeling, discovery, prediction, validation and application and combine mathematical, algorithmic, statistical and modeling aspects. He has won several awards for his work.

March 12 | VVSOR Conference 2020

14:30 – 15:15

DEFINING AND DISCOVERING (CAUSAL) INFORMATION FLOWS: NUDGE CAUSALITY

Rick Quax
University of Amsterdam

Seemingly very different systems may actually display the same kind of (complex) emergent behaviour. We argue that (Shannon's) information theory – or an appropriately generalized version – may bring a revolution in the form of a unified theoretical framework in complexity science. A crucial concept in this framework is that of 'information flows' among (stochastic) dynamical variables. In this talk I will explore causal interpretations of information flows, a possible way to detect causal information flows, and present an outlook of challenges in the multivariate setting. The work presented is part of an ongoing effort to develop a novel viewpoint on causality based on small interventions.

Rick Quax is Assistant Professor in the Computational Science Lab (CSL) at the University of Amsterdam. His interests include studying emergent behaviours exhibited by complex adaptive systems such as phase transitions, tipping points, and controllability, using information theory and computational modelling. The application domains are diverse and include financial markets, social systems, ecosystems, and marine and cell biology.

15:45 – 16:30

CAUSALITY AND ROBUST PREDICTION

JONAS PETERS
University of Copenhagen

Purely predictive methods have do not perform well when the test distribution changes too much from the training distribution. Causal models are known to be stable with respect to distributional shifts such as arbitrarily strong interventions on the covariates, but do not perform well when the test distribution differs only mildly from the training distribution. We introduce Anchor Regression, a framework that provides a trade off between causal and predictive models. The method poses different (convex and non-convex) optimization problems and relates to methods that are tailored for instrumental variable settings. If time allows, we show how similar principles can be used for inferring metabolic networks. No prior knowledge about causality is required.

Jonas Peters is a professor in statistics at the Department of Mathematical Sciences at the University of Copenhagen. Previously, he has been at the MPI for Intelligent Systems in Tübingen and was a Marie Curie fellow at the Seminar for Statistics, ETH Zürich. He studied Mathematics at the University of Heidelberg and the University of Cambridge. In his research, Jonas is interested in inferring causal relationships from different types of data and in building statistical methods that are robust with respect to distributional shifts. He seeks to combine theory, methodology, and applications. For more information, please see http://web.math.ku.dk/~peters/index.html