Assumption-Lean (Causal) Modeling

Stijn Vansteelandt

(Ghent University)

11/12/2024, 4:15 pm

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 631 1190 7291; Password: StatsCol)

Traditional inference in (semi-)parametric models, such as generalized linear models, typically assumes that the model is correctly specified and pre-determined. However, this approach is increasingly unsatisfactory because models are often selected adaptively based on the data, introducing unacknowledged uncertainty. Moreover, these models rarely represent a known ground truth, making standard inferences prone to model misspecification bias and limiting their ability to fully leverage the information in the data. While significant progress has been made in recent decades through advancements in debiased machine learning and targeted learning, these methods are inherently model-free, which can limit their applicability and interpretability in complex settings.

Assumption-lean modeling, introduced by Vansteelandt and Dukes (2022) in a discussion paper in the Journal of the Royal Statistical Society – Series B, rethinks this trade-off between model correctness and parsimony. This framework starts with data-adaptive outcome predictions and then projects these onto specific model parameters. The projection is designed to ensure that the resulting parameters are interpretable and meaningful, even when the underlying model is misspecified. By moreover leveraging debiased machine learning techniques, assumption-lean modeling achieves minimal bias, maximal interpretability, and valid confidence intervals that account for both model misspecification and uncertainty.

In this talk, I will introduce the core ideas behind assumption-lean modeling, focusing on generalized linear models to ensure accessibility. Additionally, I will explore its extensions to causal inference and time-to-event data, highlighting recent developments aimed at balancing efficiency with interpretability.

About the Speaker:
Prof. Stijn Vansteelandt is a biostatistician in the Department of Applied Mathematics, Computer Science, and Statistics of Ghent University, whose primary expertise is in causal inference, debiased machine learning and semiparametric statistics. His current research focuses on assumption-lean modeling and estimation, striving to combine the strengths of the three mainstream modeling cultures – traditional (descriptive) statistical modeling, (prescriptive) causal modeling and machine learning – with the aim to learn causal effects or associations.

References:
Vansteelandt, S. (2021). Statistical Modelling in the Age of Data Science. *Observational Studies*, 7(1), 217-228.

Vansteelandt, S and Dukes, O. (2022) Assumption-lean inference for generalised linear model parameters (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(3), 657– 685.

Vansteelandt, S., Dukes, O., Van Lancker, K., & Martinussen, T. (2024). Assumption-lean Cox regression. *Journal of the American Statistical Association*, 119(545), 475-484.