Topics in Causal Inference

Lorenzo Fabbri

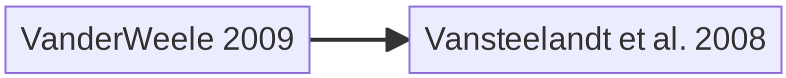
Table of contents

Preface

Part I

Assumptions

1 Interaction and Effect Modification



From the abstract of (VanderWeele 2009):

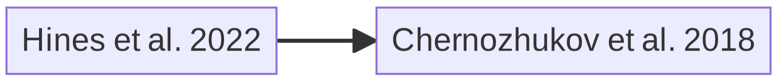
> This paper contrasts the concepts of interaction and effect modification using a series of examples. Interaction and effect modification are formally defined within the counterfactual framework. Interaction is defined in terms of the effects of 2 interventions whereas effect modification is defined in terms of the effect of one intervention varying across strata of a second variable. Effect modification can be present with no interaction; interaction can be present with no effect modification. There are settings in which it is possible to assess effect modification but not interaction, or to assess interaction but not effect modification. The analytic procedures for obtaining estimates of effect modification parameters and interaction parameters using marginal structural models are compared and contrasted. A characterization is given of the settings in which interaction and effect modification coincide.

From the abstract of (Vansteelandt et al. 2008):

Part II

Misc

2 Non-Parametric Efficiency Theory



From the abstract of (Hines et al. 2022):

> Evaluation of treatment effects and more general estimands is typically achieved via parametric modelling, which is unsatisfactory since model misspecification is likely. Data-adaptive model building (e.g. statistical/machine learning) is commonly employed to reduce the risk of misspecification. Naive use of such methods, however, delivers estimators whose bias may shrink too slowly with sample size for inferential methods to perform well, including those based on the bootstrap. Bias arises because standard data-adaptive methods are tuned towards minimal prediction error as opposed to e.g. minimal MSE in the estimator. This may cause excess variability that is difficult to acknowledge, due to the complexity of such strategies. Building on results from **nonparametric statistics**, **targeted learning** and **debiased machine learning** overcome these problems by constructing estimators using the estimand's **efficient influence function** under the nonparametric model. These increasingly popular methodologies typically assume that the efficient influence function is given, or that the reader is familiar with its derivation. In this paper, we focus on derivation of the efficient influence function and explain how it may be used to construct statistical/machine-learning-based estimators. We discuss the requisite conditions for these estimators to perform well and use diverse examples to convey the broad applicability of the theory.

3 Debiased Learning

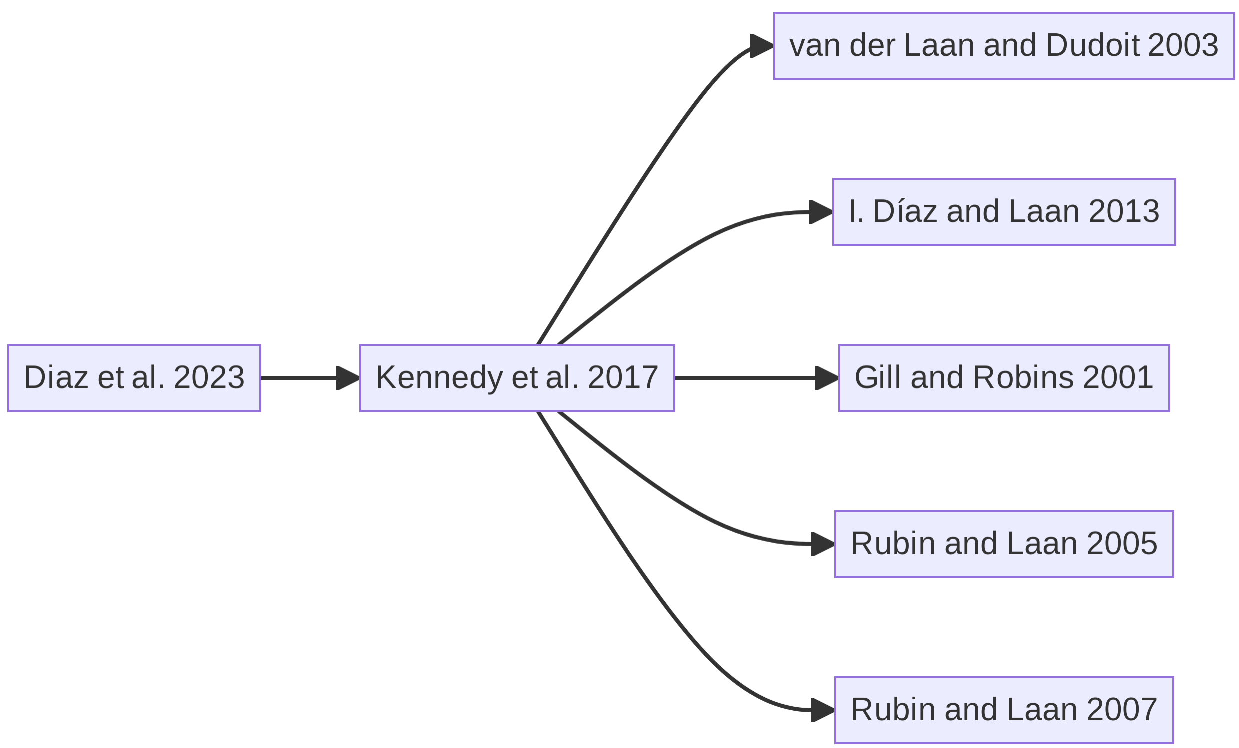
Chernozhukov et al. 2018

From the abstract of (Chernozhukov et al. 2018):

> We revisit the classic semi-parametric problem of inference on a low-dimensional parameter θ_0 in the presence of high-dimensional nuisance parameters η_0. We depart from the classical setting by allowing for η_0 to be so high-dimensional that the traditional assumptions (e.g. Donsker properties) that limit complexity of the parameter space for this object break down. To estimate η_0, we consider the use of statistical or machine learning (ML) methods, which are particularly well suited to estimation in modern, very high-dimensional cases. ML methods perform well by employing regularization to reduce variance and trading off regularization bias with overfitting in practice. However, both regularization bias and overfitting in estimating η_0 cause a heavy bias in estimators of θ_0 that are obtained by naively plugging ML estimators of η_0 into estimating equations for θ_0. This bias results in the naive estimator failing to be $N^{-1/2}$ consistent, where N is the sample size. We show that the impact of regularization bias and overfitting on estimation of the parameter of interest θ_0 can be removed by using two simple, yet critical, ingredients: (1) using Neyman-orthogonal moments/scores that have reduced sensitivity with respect to nuisance parameters to estimate θ_0; (2) making use of cross-fitting, which provides an efficient form of data-splitting. We call the resulting set of methods double or debiased ML (DML). We verify that DML delivers point estimators that concentrate in an $N^{-1/2}$-neighbourhood of the true parameter values and are approximately unbiased and normally distributed, which allows construction of valid confidence statements. The generic statistical theory of DML is elementary and simultaneously relies on only weak theoretical requirements, which will admit the use of a broad array of modern ML methods for estimating the nuisance parameters, such as random forests, lasso, ridge, deep neural nets, boosted trees, and various hybrids and ensembles of these methods. We illustrate the general theory by applying it to provide theoretical properties of the following: DML applied to learn the main regression parameter in a partially linear regression model; DML

applied to learn the coefficient on an endogenous variable in a partially linear instrumental variables model; DML applied to learn the average treatment effect and the average treatment effect on the treated under unconfoundedness; DML applied to learn the local average treatment effect in an instrumental variables setting. In addition to these theoretical applications, we also illustrate the use of DML in three empirical examples.

4 Modified Treatment Policy



4.1 Estimation

From the abstract of (van der Laan and Dudoit 2003):

From the abstract of (Rubin and Laan 2005):

From the abstract of (Rubin and Laan 2007):

4.2 Continuous treatments

From the abstract of (Gill and Robins 2001):

From the abstract of (Kennedy et al. 2017):

> **Continuous treatments** (e.g. doses) arise often in practice, but many available causal effect estimators are limited by either requiring parametric models for the effect curve, or by not allowing doubly robust covariate adjustment. We develop a novel **kernel smoothing approach** that requires only mild smoothness assumptions on the effect curve and still allows for misspecification of either the treatment density or outcome regression. We derive asymptotic properties and give a procedure for data-driven bandwidth selection. The methods are illustrated via simulation and in a study of the effect of nurse staffing on hospital readmissions penalties.

From the abstract of (I. Díaz and Laan 2013):

From the abstract of (Iván Díaz et al. 2023):

References

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21: 1–68. https://doi.org/10.1111/ectj.12097.

Díaz, I., and M. J. Laan. 2013. "Targeted Data Adaptive Estimation of the Causal Dose–Response Curve." *Journal of Causal Inference* 1: 171–92.

Díaz, Iván, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. 2023. "Nonparametric Causal Effects Based on Longitudinal Modified Treatment Policies." *Journal of the American Statistical Association* 118 (542): 846–57. https://doi.org/10.1080/01621459.2021.1955691.

Gill, R. D., and J. M. Robins. 2001. "Causal Inference for Complex Longitudinal Data: The Continuous Case." *Ann. Statist* 29: 1785–1811.

Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. "Demystifying Statistical Learning Based on Efficient Influence Functions." *The American Statistician* 76 (3): 292–304. https://doi.org/10.1080/00031305.2021.2021984.

Kennedy, Edward H., Zongming Ma, Matthew D. McHugh, and Dylan S. Small. 2017. "Non-Parametric Methods for Doubly Robust Estimation of Continuous Treatment Effects." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 79 (4): 1229–45. https://doi.org/10.1111/rssb.12212.

Rubin, D., and M. J. Laan. 2005. "A General Imputation Methodology for Nonparametric Regression with Censored Data." Technical Report 194,. Berkeley Division of Biostatistics Working Paper Series. U.C.

———. 2007. "A Doubly Robust Censoring Unbiased Transformation." *The International Journal of Biostatistics* 3.

van der Laan, Mark, and Sandrine Dudoit. 2003. "Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples." *U.C. Berkeley Division of Biostatistics Working Paper Series*, November.

VanderWeele, Tyler J. 2009. "On the Distinction Between Interaction and Effect Modification." *Epidemiology* 20 (6): 863. https://doi.org/10.1097/EDE.0b013e3181ba333c.

Vansteelandt, Stijn, Tyler J. VanderWeele, Eric J. Tchetgen, and James M. Robins. 2008. "Multiply Robust Inference for Statistical Interactions." *Journal of the American Statistical Association* 103 (484): 1693–1704. https://doi.org/10.1198/016214508000001084.