Properties of calibration estimators of the average causal effect - a comparative study of balancing approaches

Ingeborg Waernbaum[1]

[1] *Uppsala University, Sweden, ingeborg.waernbaum@statistik.uu.se*

Causal analyses with observational data require adjustment for confounding variables. Properties of semi-parametric estimators using fitted propensity scores, conditional outcomes and a combination thereof with different degrees of flexibility of parametric models have been in focus in the causal literature in recent years. Early guidance to model selection suggested that model specification, fitting and balance checking could be performed in an iterative procedure. This was followed by proposals of, now standard, doubly robust AIPW estimators that fit parametric models for the propensity score and conditional outcomes given covariates.

More recently, a class of weighting estimators have been proposed that directly aim at incorporating covariate balance in the estimation process through calibration/entropy maximization.

Since covariate balance is not a sufficient condition for identification of the true propensity score the general calibration estimator, using finite constraints, has an asymptotic error which depends on the covariance of the error of an implicit propensity score fit and the conditional outcomes. Although here, as for the AIPW estimators, robustness properties are implicit in the estimation procedure.

In this talk we describe weighting estimators within the more recent calibration/entropy balancing proposals (Tan, 2020, Chan et al. 2016) and other alternatives to propensity score estimation such as RKHS (Wong and Chan, 2018) and CBPS (Imai and Ratkovic 2014, Fan et al. 2018). We describe and compare asymptotic properties for calibration/entropy balancing estimators using Kullback-Leibler and quadratic Rényi divergence (Källberg and Waernbaum, 2020) with the related calibration estimator proposed by Tan and the CBPS estimator.

The estimators are applied to data from the Swedish Childhood Diabetes Register in a study of the effect of school achievements on complications Type 1 Diabetes Mellitus. The finite-sample properties of the estimators are investigated in a simulation study also including an evaluation of proposed variance estimators.

Joint with David Källberg and Emma Persson, Umeå University.

Graphical tools for selecting efficient conditional instrumental sets

Leonard Henckel[1], Martin Buttenschön and Marloes H. Maathuis[2]

[1] *ETH Zurich, henckel@stat.math.ethz.ch*
[2] *ETH Zurich, maathuis@stat.math.ethz.ch*

The two-stage least squares (2SLS) estimator is a popular tool for total causal effect estimation in the presence of unmeasured or latent confounding. It is well known that the 2SLS estimator is only consistent if the covariates used to compute it are appropriately chosen, in which case we refer to them as a conditional instrumental set. The choice of conditional instrumental set also impacts the estimator's asymptotic variance.

In this talk we consider the problem of how to choose conditional instrumental sets to obtain a 2SLS estimator that is not just consistent but also efficient. We do so in the setting of a Gaussian causal linear model described by a known acyclic directed mixed graph. We derive a graphical criterion that allows for qualitative asymptotic variance comparisons between certain pairs of conditional instrumental sets and gives interesting insights. Building on this we provide two easy to use graphical tools for efficient conditional instrumental set selection. First, a greedy asymptotic variance decreasing growth procedure that can be applied to any conditional instrumental set and relies only on validity checks. Second, we show that a graphically identifiable asymptotically optimal set does not exist in general. Instead we provide a conditional instrumental set guaranteed to be as close to optimal as it possible with graphical information alone; a property we refer to as graphical optimality. In particular this set is asymptotically optimal whenever a graphically identifiable asymptotically optimal set exists.

Combining the partial copula with quantile regression to test conditional independence

Lasse Petersen[1] and Niels Richard Hansen[2]

[1] *University of Copenhagen, lp@math.ku.dk*
[2] *University of Copenhagen, niels.r.hansen@math.ku.dk*

Conditional independence testing lies at the heart of causal graphical structure learning due to its use in constraint-based structure learning algorithms such as the PC and FCI algorithms. One of the most common ways of testing conditional independence of X and Y given Z is by testing for vanishing partial correlation, where the partial correlation can be computed as the correlation between residuals after performing conditional mean regression.

In this talk we consider testing conditional independence by using a different residualization approach. More specifically, we residualize by transforming the variables by their conditional distribution functions, $U_1 = F(X \mid Z)$ and $U_2 = F(Y \mid Z)$, and then test for independence between U_1 and U_2 using a generalized correlation measure. Carrying out the test in practice requires estimation of the conditional distribution functions, and we present an estimator of F based on conditional quantile regression for which we can quantify the consistency rate. Furthermore, we show how the consistency rate of the conditional distribution function estimator can be transferred to level and power properties of the conditional independence test.

Lastly, we discuss the benefits of using this residualization approach over conventional residuals, and we present simulations which demonstrates that our test has superior power in cases with conditional variance heterogeneity.

References

[1] Petersen, L., & Hansen, N. R. (2021). Testing Conditional Independence via Quantile Regression Based Partial Copulas. *Journal of Machine Learning Research*, 22(70), 1-47.