Statistical Science
2020, Vol. 35, No. 3, 427–429
https://doi.org/10.1214/20-STS768
Main article: https://doi.org/10.1214/19-STS721

Comment: Invariance, Causality and Robustness

Vanessa Didelez

I would like to congratulate Peter Bühlmann on the honor of being invited to give the Neyman Lecture. Jointly with a number of co-authors of recent papers, he has produced a substantial and thought-provoking body of work in recent years around the concept of invariance. His achievement is two-fold: He extends causal reasoning to involve prediction under new environments; after several decades of existing research in the field of causal inference; see early work in the 1970s and 1980s by Rubin, Robins, Pearl, Spirtes and colleagues and the ensuing explosion of work on this topic in bio-medical statistics, epidemiology, computer science, sociology and political science—this is a *novel angle* on causal inference, using data in a different way with an original target of inference so far undervalued in the causal inference literature. Vice versa, he demonstrates how causal reasoning is important to predictive modelling. It is a particular achievement of Bühlmann to have brought key ideas and concepts of causality and causal inference to the attention of mainstream statistics. This is not least due to linking causal ideas, such as invariance (also known as *stability* (Dawid and Didelez, 2010)), with fundamental concepts of traditional statistical inference, such as worst-case risk optimization.

In the following, I will review the differences and similarities of ‘classical’ causal inference and Bühlmann’s approach.

CAUSAL INFERENCE, BIAS AMPLIFICATION AND PREDICTION

I would like to discuss some of the ideas in Bühlmann’s paper by attempting to relate them to a phenomenon known in the bio-medical/causal inference literature as ‘bias amplification’ (Pearl, 2010, Middleton et al., 2015, Ding, VanderWeele and Robins, 2017). Consider a simple linear SEM where

$$Y = \beta X + \alpha H + \epsilon,$$

and where A is a valid instrumental variable for the effect β of X on Y (as in Bühlmann’s Figure 6 with no $A \to H$ and no $A \to Y$ edges, see Figure 1(a) in this commentary). The classical aim of causal inference is to to estimate β: we may be interested in β because under the above SEM this parameter represents the effect on Y of fixing X at x versus fixing it at $x+1$, that is, the average causal effect $\beta = \mathbb{E}(Y|\mathrm{do}(X=x+1)) - \mathbb{E}(Y|\mathrm{do}(X=x))$ (due to linearity and no interaction, the marginal and the conditional average causal effects are the same in this special case; but we must not forget that this does not hold for more general models[1]).

In the above model, we know that (i) a linear regression of Y on X results in a biased estimator for β due to the hidden confounder H, unless the $H \to X$ or $H \to Y$ relations vanish; (ii) using A as an instrument to perform two-stage least squares (2SLS) yields a consistent estimator of β; (iii) regressing Y on both X and A (or partialling out A first) typically results in *even more* bias than approach (i). This last phenomenon is known as ‘bias amplification’ (Pearl, 2010). Intuitively, the amplification occurs because including the IV A as additional regressor explains away some of the ‘free’ variability in X, with the variability due to H remaining, and hence amplifying the bias due to unobserved confounding by H (Greenland and Pearl, 2011).

When using anchor regression, (i) corresponds to $\gamma = 1$, (ii) to $\gamma = \infty$, and (iii) to $\gamma = 0$. Hence, when A is a valid instrument, we can roughly say the larger γ the less bias we have in estimating β; at the same time (due to Theorem 4.1) for large γ we minimise worst-case prediction risk under large shift perturbations but not under small shift perturbations.

Consider now the case where A is not a valid instrument (see Figure 1(b) in this commentary) and

$$Y = \beta X + \alpha H + \xi A + \epsilon,$$

with $\xi \neq 0$ (note that under the shift perturbations ξA is replaced by the shift v). In this case, we know (i′) a linear regression of Y on X results in a ‘doubly biased’ estimator for β due to the hidden confounder H and the hidden (unused) confounder A; (ii′) using A as an instrument to perform two-stage least-squares will also yield a biased estimator of β as A is now not a valid IV anymore; (iii′)

Vanessa Didelez is Professor, Leibniz Institute for Prevention Research and Epidemiology—BIPS, and Faculty of Mathematics and Computer Science, University of Bremen, Germany (e-mail: didelez@leibniz-bips.de).

[1]Much of the causal inference literature is concerned with robust estimation of a marginal causal effects under *considerably weaker* parametric assumptions than a linear SEM.