

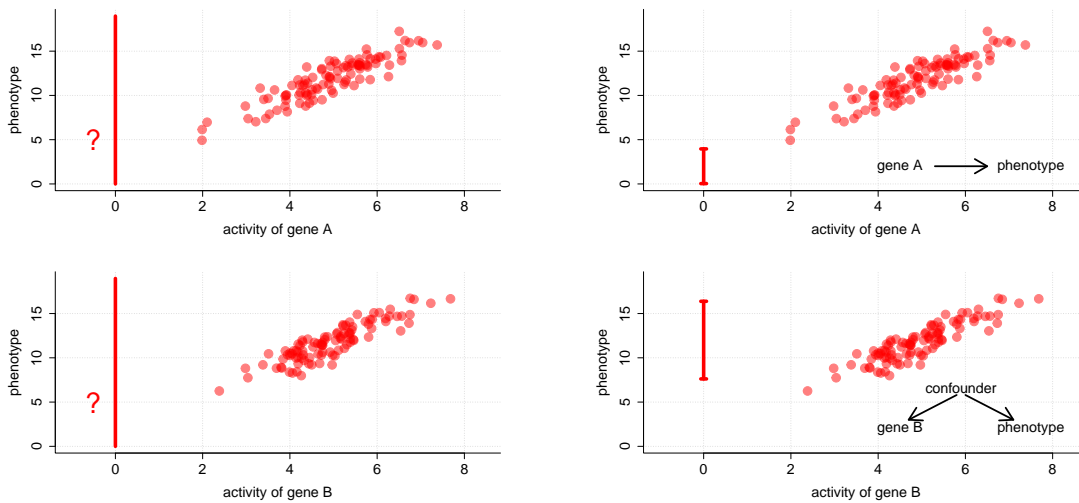
Causality

Jonas Peters

Copenhagen, December 2018

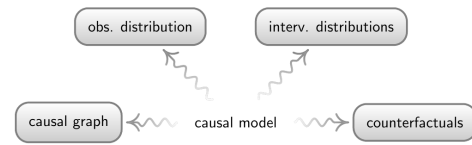
In recent years, we have seen many successful applications of statistical and machine learning methods in various domains. In these applications, the aim is usually to predict a random variable given the observation of some covariates (also called “predictor variables”). In speech recognition the spoken word is predicted from a regularly sampled amplitude of an audio signal. In computer vision or medical imaging one predicts the presence of a pedestrian or localizes a tumor from a vector of pixel values. Often, however, we want to actively influence the system and we are thus interested in predicting a system’s behavior under intervention.

Causality is more than statistics. Assume, we are given data showing expression level of a gene and the value of some phenotype; the figure below (top left) shows a toy data set.



Clearly, both variables strongly correlate with each other and the data are well explained by linearly regressing the phenotype on the activity of the gene. This can be exploited for classical prediction: if we *observe* the activity of gene A to be 6, we expect the phenotype to lie between 12 and 16 with high probability. The same holds for a gene B (bottom left). Assume, however, that we are interested in predicting a phenotype after deleting (not observing) a gene, that is, after *setting* its activity to zero (neglecting measurement noise); this is called an intervention. The following argument shows that we then require knowledge of the underlying causal structure. Gene A is a cause of the phenotype, and we thus expect a low value of the phenotype after deleting gene A, see the above figure (top right). In fact, we may exploit the linear model that we have learned from the observational data: it is invariant under the intervention. The case for gene B is different: there is a common cause (a “confounder”) gene C, causally effecting both gene B and the phenotype, see the above figure (bottom right). Then, the intervention on gene B has no effect on the phenotype and we predict the phenotype to lie in its “usual range”. Thus, if we are posed with the task of predicting a random variable after an intervention on some of the covariates, and we do not take causality into consideration, there is no alternative than to answer “I do not know”.

If we want to describe the above situation properly, we require a so-called causal model that models both observational data and interventional data (e.g., the distribution that arises after the gene deletion). Many causal models also output a graph, see the figure on the right. Structural causal models (also called structural equation models) are one class of such models, but potential outcomes can be used, too.



Causal discovery. In practice, we are usually not given the underlying causal model, and we need to learn it from data instead. Randomized experiments constitute a “gold standard” for this task. If a randomized treatment is significantly correlated with the recovery rate, there must be a causal link from treatment to recovery. Due to ethical, financial or physical reasons randomization is often impossible. The goal of causal discovery is to infer the underlying causal structure from observational and/or interventional data. A lot of my research has focused on the question of when and how this is possible. As for prediction problems, nothing is possible without assumptions – there is no such thing as a free lunch. For example, researchers have used the Markov assumption and faithfulness to connect properties of the causal structure (d -separation) with properties of the observational distribution (conditional independence) to make causal discovery possible. In current research, scientists aim to either exploit a set of existing assumptions (e.g., they construct new algorithms for causal discovery and analyze their statistical properties) or they explore which other assumptions may yield causal identifiability.

For example, we have proved that causal discovery can be based on simplicity assumptions of the data generating process (“additive noise models”) or on invariance properties of the causal model if the system is observed under different environments (“invariant causal prediction”). These methods can be extended to include the notion of time and to deal with the existence of hidden variables.

We begin to see more and more promising results on real data sets, but in my opinion, we still need to understand better how the algorithms behave under model misspecification. This may help us to approach the important causal questions that are omnipresent in climate science, the social sciences or health.

Distributional robustness. Many researchers believe that causal ideas have the potential of playing a major role in more classical problems of computational statistics or machine learning, too. The reasoning is as follows: while we are seeing many successful applications in areas, where a lot of data are obtained under comparable conditions (cf. the “i.i.d.” assumption), machines are less successful in transferring knowledge. For example, autonomous cars recognize objects by generalizing from a huge amount of labeled training data. Assume that a few years after the car’s production, more and more people start to wear green jackets. We require the cars to reliably recognize pedestrians, even after such a distributional shift. In particular, cars have to make decisions in situations of high risk, which usually look very different from the vast majority of data that the cars were trained on. Causal ideas may give us a hint on which parts of the models we may assume to be invariant between training and testing and which parts of the model we should not trust after a distributional shift. The connection between causality and distributional robustness is mainly unexplored, but there are a few first attempts, such as “anchor regression”, that aim to establish a more formal link.

Further details and references. If you are interested in more details, see, e.g., J. Peters, D. Janzing, and B. Schölkopf: *Elements of Causal Inference: Foundations and Learning Algorithms*, MIT Press, Cambridge, MA, USA, 2018. It is available as open access and can be downloaded for free from the MIT Press website <https://mitpress.mit.edu/books/elements-causal-inference>. The book further contains pointers to relevant literature that are omitted here.